

REPORT

# AUTOMATING SOCIETY

2020



ALGORITHM  
WATCH

| BertelsmannStiftung

## **IMPRINT**

### **Automating Society Report 2020**

October 2020

Available online at <https://automatingsociety.algorithmwatch.org>

#### **Publishers**

AlgorithmWatch gGmbH  
Linienstr. 13  
10178 Berlin  
Germany

Bertelsmann Stiftung  
Carl-Bertelsmann-Str. 256  
33311 Gütersloh  
Germany

#### **Editors**

Fabio Chiusi  
Sarah Fischer  
Nicolas Kayser-Bril  
Matthias Spielkamp

#### **Project manager**

Fabio Chiusi

#### **Publication coordinator**

Marc Thümmeler

#### **Comic artists**

Samuel Daveti  
Lorenzo Palloni  
Alessio Ravazzani

#### **Copy editing**

Graham Holliday

#### **Layout**

Beate Autering  
Beate Stangl

#### **Additional editing**

Leonard Haas

Editorial deadline: 30 September 2020



This publication is licensed under a Creative Commons Attribution 4.0 International License  
<https://creativecommons.org/licenses/by/4.0/legalcode>

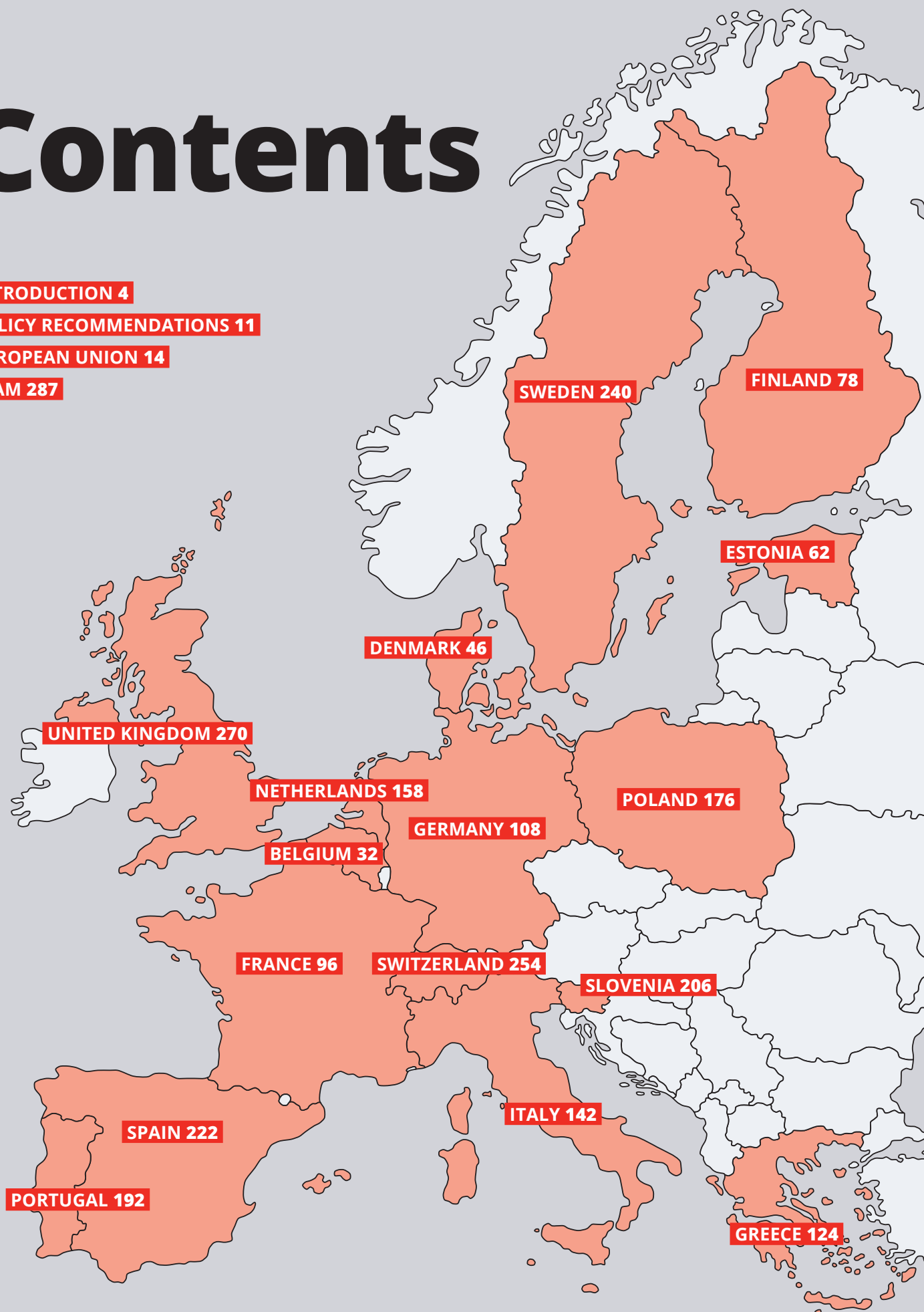
# Contents

**INTRODUCTION 4**

**POLICY RECOMMENDATIONS 11**

**EUROPEAN UNION 14**

**TEAM 287**



**Key of colours:**




European countries  
covered in this report



European countries  
not covered





# **Life in the automated society: How automated decision-making systems became mainstream, and **what to do about it****

By Fabio Chiusi

The editorial deadline for this report was 30 September 2020.  
Later developments could not be included.



On a cloudy August day in London, students were angry. They **flocked** to Parliament Square by the hundreds, in protest – their placards emblazoned with support for unusual allies: their teachers, and an even more unusual target: an algorithm.

Due to the COVID-19 pandemic, schools closed in March in the United Kingdom. With the virus still raging throughout Europe over the summer of 2020, students knew that their final exams would have to be canceled, and their assessments – somehow – changed. What they could not have imagined, however, was that thousands of them would end up with **lower** than expected grades as a result.

Students protesting knew what was to blame, as apparent by their signs and chants: the automated decision-making (ADM) system deployed by the Office of Qualifications and Examinations Regulation (Ofqual). It **planned** to produce the best data-based assessment for both General Certificates of Secondary Education and A-level results, in such a way that “the distribution of grades follows a similar pattern to that in other years, so that this year’s students do not face a systemic disadvantage as a consequence of circumstances this year”.

The government wanted to avoid the excess of optimism<sup>1</sup> that would have resulted from human judgment alone, according to its own **estimates**: compared to the historical series, grades would have been too high. But this attempt to be “as far as possible, fair to students who had been unable to sit their exams this summer” failed spectacularly, and, on that grey August day of protest, the students kept on coming, performing chants, and holding signs to express an urgent need for social justice. Some were desperate, some broke down and cried.

“Stop stealing our future”, read one placard, echoing the Fridays for Future protests of climate activists. Others, however, were more specifically tailored to the flaws of the ADM grading system: “Grade my work, not my postcode”, we’re “students, not stats”, they read, denouncing the discriminatory outcomes of the system<sup>2</sup>.

Finally, a chant erupted from the crowd, one that has come

to the future of protest: “Fuck the algorithm”. Scared that the government was casually – and opaquely – automating their future, no matter how inconsistent with their skills and efforts, students screamed for the right not to have their life chances unduly affected by bad code. They wanted to have a say, and what they said should be heard.

Algorithms are neither “neutral” nor “objective” even though we tend to think that they are. They replicate the assumptions and beliefs of those who decide to deploy them and program them. Humans, therefore, are, or should be, responsible for both good and bad algorithmic choices, not “algorithms” or ADM systems. The machine may be scary, but **the ghost within it** is always human. And humans are complicated, even more so than algorithms.

The protesting students were not as naive as to believe that their woes were solely the fault of an algorithm, anyway. In fact, they were not chanting against “the algorithm” in an outburst of technological determinism; they were motivated by an urge to protect and promote social justice. In this respect, their protest more closely resembles that of the Luddites. Just as the labor movement that crushed mechanized looms and knitting frames in the 19th Century, they know that ADM systems are about power, and should not be mistaken for being an allegedly objective technology. So, they chanted “justice for the working class”, asked for the resignation of the Health Secretary, portrayed the ADM system as “classism at its finest”, “blatant classism”.

Eventually, the students succeeded in abolishing the system which put their educational career and chances in life at risk: in a spectacular U-turn, the UK government **scrapped** the error-prone ADM system and utilized the grades predicted by teachers.

But there’s more to this story than the fact that the protesters won in the end. This example highlights how poorly designed, implemented, and overseen systems that reproduce human bias and discrimination fail to make use of the potential that ADM systems have, such as leveraging comparability and fairness.

More clearly than many struggles in the past, this protest reveals that we’re no longer just automating society. We have automated it already – and, finally, somebody noticed.

1 “The research literature suggests that, in estimating the grades students are likely to achieve, teachers tend to be optimistic (although not in all cases)”, writes Ofqual, [cfr. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/909035/6656-2\\_-\\_Executive\\_summary.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/909035/6656-2_-_Executive_summary.pdf)

2 Cfr. the UK chapter for details.

## / From Automating Society to the automated society

When launching the first edition of this report, we decided to call it “Automating Society”, as ADM systems in Europe were mostly new, experimental, and unmapped – and, above all, the exception rather than the norm.

This situation has changed rapidly. As clearly shown by the many cases gathered in this report through our outstanding network of researchers, the deployment of ADM systems has vastly increased in just over a year. ADM systems now affect almost all kinds of human activities, and, most notably, the distribution of services to millions of European citizens – and their access to their rights.

The stubborn opacity surrounding the ever-increasing use of ADM systems has made it all the more urgent that we continue to increase our efforts. Therefore, we have added four countries (Estonia, Greece, Portugal, and Switzerland) to the 12 we already analyzed in the previous edition of this report, bringing the total to 16 countries. While far from exhaustive, this allows us to provide a broader picture of the ADM scenario in Europe. Considering the impact these systems may have on everyday life, and how profoundly they challenge our intuitions – if not our norms and rules – about the relationship between democratic governance and automation, we believe this is an essential endeavor.

This is especially true during the COVID-19 pandemic, a time in which we have witnessed the (mostly rushed) adoption of a plethora of ADM systems that aim to contribute to securing public health through data-based tools and automation. We deemed this development to be so important that we decided to dedicate a “preview report” to it, published<sup>3</sup> in August 2020 within the scope of the ‘Automating Society’ project.

Even in Europe, when it comes to the deployment of ADM systems, the sky is the limit. Just think of some of the cases introduced in this report, adding to the many – from welfare to education, the health system, to the judiciary – that we already reported on in the [previous edition](#). In the following pages, and for the first time, we provide updates on the development of these cases in three ways. Firstly, through journalistic stories, then, through research-based sections cataloging different examples, and, finally, with graphic novels. We felt that these ADM systems are – and

increasingly will be – so crucial in everyone’s lives that we needed to try and communicate how they work, and what they actually *do to us*, in both rigorous and new ways, to reach all kinds of audiences. After all, ADM systems have an impact on all of us.

Or at least they should. We’ve seen, for example, how a new, automated, proactive service distributes family benefits in Estonia. Parents no longer even need to apply for benefits: from birth, the state collects all the information about each newborn and their parents and collates it in databases. As a result, the parents automatically receive benefits if they are entitled to them.

In Finland, the identification of individual risk factors related to social exclusion in young adults is automated through a tool developed by the Japanese giant, Fujitsu. In France, data from social networks can be scraped to feed machine learning algorithms that are employed to detect tax fraud.

Italy is experimenting with “predictive jurisprudence”. This uses automation to help judges understand trends from previous court rulings on the subject at hand. And, in Denmark, the government tried to monitor every keyboard and mouse click on students’ computers during exams, causing – again – massive student protests that led to the withdrawal of the system, for the time being.

## / Time to put ADM wrongs to right

In principle, ADM systems have the potential to benefit people’s lives – by processing huge amounts of data, supporting people in decision-making processes, and providing tailored applications.

In practice, however, we found very few cases that convincingly demonstrated such a positive impact.

For example, the VioGén system, deployed in Spain since 2007 to assess risk in cases of domestic violence, while far from perfect, [shows](#) “reasonable performance indexes” and has helped protect many women from violence.

In Portugal, a centralized, automated system deployed to deter fraud associated with medical prescriptions has [re-reportedly](#) reduced fraud by 80% in a single year. A similar system, in Slovenia, used to combat tax fraud has proved useful for inspectors, according to tax authorities<sup>4</sup>.

3 ‘Automated Decision-Making Systems in the COVID-19 Pandemic: A European Perspective’, <https://algorithmwatch.org/en/project/automating-society-2020-covid19/>

4 Cfr. the chapter on Slovenia for details.

When looking at the current state of ADM systems in Europe, positive examples with clear benefits are rare. Throughout the report, we describe how the vast majority of uses tend to put people at risk rather than help them. But, to truly judge the actual positive and negative impact, we need more transparency about goals and more data about the workings of ADM systems that are tested and deployed.

The message for policy-makers couldn't be clearer. If we truly want to make the most of their potential, while at the same time respecting human rights and democracy, the time to step up, make those systems transparent, and put ADM wrongs right, is now.

## **/ Face recognition, face recognition, everywhere**

Different tools are being adopted in different countries. One technology, however, is now common to most: face recognition. This is arguably the newest, quickest, and most concerning development highlighted in this report. Face recognition, nearly absent from the 2019 edition, is being trialed and deployed at an alarming rate throughout Europe. In just over a year since our last report, face recognition is present in schools, stadiums, airports, and even in casinos. It is also used for predictive policing, to apprehend criminals, against [racism](#), and, regarding the COVID-19 pandemic, to enforce social distancing, both in apps and through “smart” video-surveillance.

**Face recognition, nearly absent from the 2019 edition, is being trialed and deployed at an alarming rate throughout Europe.**

New ADM deployments continue, even in the face of [mounting evidence](#) of their lack of [accuracy](#). And when challenges emerge, proponents of these systems simply try and find their way around them. In Belgium, a face recognition system used by the police is still “partially active”, even though a temporary ban has been issued by the Oversight Body for Police Information. And, in Slovenia, the use of face recognition technology by the police was legalized five years after they first started using it.

This trend, if not challenged, risks normalizing the idea of being constantly – and opaquely – watched, thus crystallizing a new status quo of pervasive mass surveillance. This is why many from the civil liberties community would have welcomed a much more aggressive policy response by EU institutions to this<sup>5</sup>.

Even the act of smiling is now part of an ADM system piloted in banks in Poland: the more an employee smiles, the better the reward. And it's not just faces that are being monitored. In Italy, a sound surveillance system was proposed as an anti-racism tool to be used in all football stadiums.

## **/ Black boxes are still black boxes**

A startling finding in this report is that, while change happened rapidly regarding the deployment of ADM systems, the same is not true when it comes to the transparency of these systems. In 2015, Brooklyn Law School professor, Frank Pasquale, famously called a networked society based on opaque algorithmic systems a “[black box society](#)”. Five years later, and the metaphor, unfortunately, still holds – and applies to all the countries we studied for this report, across the board: there is not enough transparency concerning ADM systems – neither in the public, nor the private sector. Poland even mandates opacity, with the law that introduced its automated system to detect bank accounts used for illegal activities (“STIR”). The law states that the disclosure of adopted algorithms and risk indicators may result in up to 5 years in jail.

While we firmly reject the idea that all such systems are inherently bad – we embrace an evidence-based perspective instead – it is undoubtedly bad to be unable to assess their functioning and impact based on accurate and factual knowledge. If only because opacity severely impedes the gathering of evidence that is necessary to come to an informed judgment on the deployment of an ADM system in the first place.

<sup>5</sup> As detailed in the EU chapter.



When coupled with the difficulty both our researchers and journalists found in accessing any meaningful data on these systems, this paints a troubling scenario for whoever wishes to keep them in check and guarantee that their deployment is compatible with fundamental rights, the rule of law, and democracy.

### **/ Challenging the algorithmic status quo**

What is the European Union doing about this? Even though the strategic documents produced by the EU Commission, under the guidance of Ursula Von der Leyen, refer to “artificial intelligence” rather than ADM systems directly, they do state laudable intentions: promoting and realizing a “trustworthy AI” that puts “people first”<sup>6</sup>.

However, as described in the EU chapter, the EU’s overall approach prioritizes the commercial and geopolitical imperative to lead the “AI revolution” over making sure that its products are consistent with democratic safeguards, once adopted as policy tools.

This lack of political courage, which is most apparent in the decision to ditch any suggestion of a moratorium on live face recognition technologies in public places in its AI regulation package, is surprising. Especially at a time when many Member States are witnessing an increasing number of legal challenges – and defeats – over hastily deployed ADM systems that have negatively impacted the rights of citizens.

A landmark case comes from the Netherlands, where civil rights activists took an invasive and opaque automated system, supposed to detect welfare fraud (SyRI), to court and won. Not only was the system found in violation of the European Convention on Human Rights by the court of The Hague in February, and therefore halted. The case also set a precedent: according to the ruling, governments have a “special responsibility” to safeguard human rights when implementing such ADM systems. Providing much-needed transparency is considered a crucial part of this.

Since our first report, media and civil society activists have established themselves as a driving force for accountability in ADM systems. In Sweden, for example, journalists managed to force the release of the code behind the Trelleborg

system for fully automated decisions related to social benefit applications. In Berlin, the Südkreuz train station face recognition pilot project failed to lead to the implementation of the system anywhere in Germany. This was thanks to the loud opposition of activists, so loud that they managed to influence party positions and, ultimately, the government’s political agenda.

Greek activists from Homo Digitalis showed that no real traveler participated in the Greek pilot trials of a system called ‘iBorderCtrl’, an EU-funded project that aimed to use ADM to patrol borders, thus revealing that the capabilities of many such systems are frequently oversold. Meanwhile, in Denmark, a profiling system for the early detection of risks associated with vulnerable families and children (the so-called “Gladaxe model”) was put on hold thanks to the work of academics, journalists, and the Data Protection Authority (DPA).

DPAs themselves played an important role in other countries too. In France, the national privacy authority ruled that both a sound surveillance project and one for face recognition in high schools were illegal. In Portugal, the DPA refused to approve the deployment of video surveillance systems by the police in the municipalities of Leiria and Portimão as it was deemed disproportionate and would have amounted to “large-scale systematic monitoring and tracking of people and their habits and behavior, as well as identifying people from data relating to physical characteristics”. And, in the Netherlands, the Dutch DPA asked for more transparency in predictive algorithms used by government agencies.

Lastly, some countries have referred to an ombudsperson for advice. In Denmark, this advice helped to develop strategies and ethical guidelines for the use of ADM systems in the public sector. In Finland, the deputy parliamentary ombudsperson considered automated tax assessments unlawful.

And yet, given the continued deployment of such systems throughout Europe, one is left wondering: is this level of oversight enough? When the Polish ombudsperson questioned the legality of the smile detection system used in a bank (and mentioned above), the decision did not prevent a later pilot in the city of Sopot, nor did it stop several companies from showing an interest in adopting the system.

---

<sup>6</sup> Cfr. the EU chapter, and in particular the section on the EU Commission’s ‘White Paper on AI’

# A startling finding in this report is that, while change happened rapidly regarding the deployment of ADM systems, the same is not true when it comes to the transparency of these systems.

## **/ Lack of adequate auditing, enforcement, skills, and explanations**

Activism is mostly a reactive endeavor. Most of the time, activists can only react if an ADM system is being trialed or if one has already been deployed. By the time citizens can organize a response, their rights may have been infringed upon unnecessarily. This can happen even with the protections that should be granted, in most cases, by EU and Member States' law. This is why proactive measures to safeguard rights – before pilots and deployments take place – are so important.

And yet, even in countries where protective legislation is in place, enforcement is just not happening. In Spain, for example, “automated administrative action” is legally codified, mandating specific requirements in terms of quality control and supervision, together with the audit of the information system and its source code. Spain also has a Freedom of Information law. However, even with these laws, only rarely, our researcher writes, do public bodies share detailed information about the ADM systems they use. Similarly, in France, a 2016 law exists that mandates algorithmic transparency, but again, to no avail.

Even bringing an algorithm to court, according to the specific provisions of an algorithmic transparency law, may not be enough to enforce and protect users' rights. As the case of the Parcoursup algorithm to sort university applicants in France shows<sup>7</sup>, exceptions can be carved out at will to shield an administration from accountability.

This is especially troubling when coupled with the endemic lack of skills and competences around ADM systems in the public sector lamented by many researchers. How could public officials explain or provide transparency of any kind around systems they don't understand?

Recently, some countries tried to address this issue. Estonia, for example, set up a competence center suited to ADM systems to better look into how they could be used to develop public services and, more specifically, to inform the operations of the Ministry of Economic Affairs and Communications and the State Chancellery for the development of e-government. Switzerland also called for the creation of a “competence network” within the broader framing of the “Digital Switzerland” national strategy.

And yet, the lack of digital literacy is a well-known issue affecting a large proportion of the population in several European countries. Besides, it is tough to call for the enforcement of rights you don't know you have. Protests in the UK and elsewhere, together with high profile scandals based on ADM systems<sup>8</sup>, have certainly raised awareness of both the risks and opportunities of automating society. But while on the rise, this awareness is still in its early stages in many countries.

The results from our research are clear: while ADM systems already affect all sorts of activities and judgments, they are still mainly deployed without any meaningful democratic debate. Also, it is the norm, rather than the exception, that

<sup>7</sup> Cfr. the chapter on France

<sup>8</sup> Think of the “Buona Scuola” algorithm debacle in Italy, cfr. the chapter on Italy.

enforcement and oversight mechanisms – if they even exist – lag behind deployment.

Even the purpose of these systems is not commonly justified or explained to affected populations, not to mention the benefits they are supposed to gain. Think of the “AuroraAI” proactive service in Finland: it is supposed to automatically identify “life events”, as our Finnish researchers report, and in the minds of proponents, it should work as “a nanny” that helps citizens meet particular public service needs that may arise in conjunction with certain life circumstances, e.g., moving to a new place, changing family relations, etc. “Nudging” could be at work here, our researchers write, meaning that instead of empowering individuals, the system might end up doing just the opposite, suggesting certain decisions or limiting an individual’s options through its own design and architecture.

It is then all the more important to know what it is that is being “optimized” in terms of public services: “is service usage maximized, are costs minimized, or is citizen well-being improved?”, ask the researchers. “What set of criteria are these decisions based on and who chooses them?” The mere fact that we don’t have an answer to these fundamental questions speaks volumes about the degree of participation and transparency that is allowed, even for such a potentially invasive ADM system.

### / The techno-solutionist trap

There is an overarching ideological justification for all this. It is called “technological solutionism”, and it still severely affects the way in which many of the ADM systems we studied are developed. Even if the term has been long-denounced as a flawed ideology that conceives of every social problem as a “bug” in need of a “fix” through technology<sup>9</sup>, this rhetoric is still widely adopted – both in the media and in policy circles – to justify the uncritical adoption of automated technologies in public life.

When touted as “solutions”, ADM systems immediately veer into the territory described in Arthur C. Clarke’s Third Law: magic. And it is difficult, if not impossible, to regulate magic, and even more so to provide transparency and explanations around it. One can see the hand reaching inside the

hat, and a bunny appears as a result, but the process is and *should remain* a “black box”.

Many researchers involved in the ‘Automating Society’ project denounced this as the fundamental flaw in the reasoning behind many of the ADM systems they describe. This also implies, as shown in the chapter on Germany, that most critiques of such systems are framed as an all-out rejection of “innovation”, portraying digital rights advocates as “neo-luddites”. This not only ignores the historical reality of the Luddite movement, which dealt in labor policies and not technologies *per se*, but also, and more fundamentally, threatens the effectiveness of hypothesized oversight and enforcement mechanisms.

At a time when the “AI” industry is witnessing the emergence of a “lively” lobbying sector, most notably in the UK, this might result in “**ethics-washing**” guidelines and other policy responses that are ineffective and structurally inadequate to address the human rights implications of ADM systems. This view ultimately amounts to the assumption that we humans should adapt to ADM systems, much more than ADM systems should be adapted to democratic societies.

To counter this narrative, we should not refrain from foundational questions: whether ADM systems can be compatible with democracy and deployed for the benefit of society at large, and not just for parts of it. It might be the case, for example, that certain human activities – e.g., those concerning social welfare – should not be subject to automation, or that certain technologies – namely, live face recognition in public spaces – should not be promoted in an endless quest for “AI leadership”, but banned altogether instead.

Even more importantly, we should reject any ideological framing that prevents us from posing such questions. On the contrary: what we need to see now is actual policies changing – in order to allow greater scrutiny of these systems. In the following section we list the key demands that result from our findings. We hope that they will be widely discussed, and ultimately implemented.

Only through an informed, inclusive, and evidence-based democratic debate can we find the right balance between the benefits that ADM systems can – and do – provide in terms of speed, efficiency, fairness, better prevention, and access to public services, and the challenges they pose to the rights of us all.

<sup>9</sup> See Evgeny Morozov (2014), *To Save Everything, Click Here. The Folly of Technological Solutionism*, Public Affairs, <https://www.publicaffairsbooks.com/titles/evgeny-morozov/to-save-everything-click-here/9781610393706/>



# Policy Recommendations

In light of the findings detailed in the 2020 edition of the Automating Society report, we recommend the following set of policy interventions to policymakers in the EU parliament and Member States' parliaments, the EU Commission, national governments, researchers, civil society organizations (advocacy organizations, foundations, labor unions, etc.), and the private sector (companies and business associations). The recommendations aim to better ensure that ADM systems currently being deployed and those about to be implemented throughout Europe are effectively consistent with human rights and democracy:

## 1. Increase the transparency of ADM systems

Without the ability to know precisely how, why, and to what end ADM systems are deployed, all other efforts for the reconciliation of fundamental rights and ADM systems are doomed to fail.

### / Establish public registers for ADM systems used within the public sector

We, therefore, ask for legislation to be enacted at the EU level to mandate that Member States establish public registers of ADM systems used by the public sector.

They should come with the legal obligation for those responsible for the ADM system to disclose and document the purpose of the system, an explanation of the model (logic involved), and information about who developed the system. This information has to be made available in an easily readable and accessible manner, including structured digital data based on a standardized protocol.

Public authorities have a particular responsibility to make the operational features of ADM systems deployed in public administration transparent. This was underlined by a recent administrative complaint in Spain, that argues that “any ADM system used by the public administration should be made public by default”. If upheld, the ruling could become precedent in Europe.

Whereas disclosure schemes on ADM systems should be mandatory for the public sector in all cases, these trans-

parency requirements should also apply to the use of ADM systems by private entities when an AI/ADM system has a significant impact on an individual, a specific group, or society at large.

### / Introduce legally-binding data access frameworks to support and enable public interest research

Increasing transparency not only requires disclosing information about a system's purpose, logic, and creator, as well as the ability to thoroughly analyze, and test a system's inputs and outputs. It also requires making training data and data results accessible to independent researchers, journalists, and civil society organizations for public interest research.

That's why we suggest the introduction of robust, legally-binding data access frameworks, focused explicitly on supporting and enabling public interest research and in full respect of data protection and privacy law.

Learning from existing best practices at the national and EU levels, such tiered frameworks should include systems of sanctions, checks and balances as well as regular reviews. As private data-sharing partnerships have illustrated, there are legitimate concerns regarding user privacy and the possible de-anonymization of certain kinds of data.

Policymakers should learn from health data sharing frameworks to facilitate privileged access to certain kinds of more granular data, while ensuring that personal data is adequately protected (e.g., through secure operating environments).

While an effective accountability framework will require transparent access to platform data, this is a requirement for many auditing approaches to be effective as well.

## 2. Create a meaningful accountability framework for ADM systems

As findings from Spain and France have shown, even if transparency of an ADM system is required by law and/or information has been disclosed, this does not necessarily result in accountability. Further steps are needed to ensure that laws and requirements are actually enforceable.

### **/ Develop and establish approaches to effectively audit algorithmic systems**

To ensure that transparency is meaningful, we need to complement the first step of establishing a public register by processes that effectively audit algorithmic systems.

The term “auditing” is widely used, but there is no common understanding of the definition. We understand auditing in this context in accordance with ISO’s definition as a “systematic, independent and documented process for obtaining objective evidence and evaluating it objectively to determine the extent to which the audit criteria are fulfilled.”<sup>10</sup>

We do not have satisfying answers to the complex questions<sup>11</sup> raised by the auditing of algorithmic systems yet; however, our findings clearly indicate the need to find answers in a broad, stakeholder engagement process and through thorough and dedicated research.

Both audit criteria and appropriate processes of auditing should be developed, following a multi-stakeholder approach that actively takes into consideration the disproportionate effect ADM systems have on vulnerable groups and solicits their participation.

We, therefore, ask policymakers to initiate such stakeholder processes in order to clarify the outlined questions, and to make available sources of funding aimed at enabling the participation by stakeholders who have so far been inadequately represented.

10 <https://www.iso.org/obp/ui/#iso:std:iso:19011:ed-3:v1:en>

11 Thinking of potential models of algorithmic auditing, several questions emerge. 1) Who/what (services/platforms/products) should be audited? How to customize the auditing systems to the type of platform/type of service? 2) When should an audit be undertaken by a public institution (at EU level, national level, local level), and when can it be done by private entities/experts (business, civil society, researchers)? 3) How to clarify the distinction between assessing impact ex-ante (i.e. in the design phase) and ex-post (i.e. in operation) and the respective challenges? 4) How to assess trade-offs in the different virtues and vices of auditability? (e.g., simplicity, generality, applicability, precision, flexibility, interpretability, privacy, efficacy of an auditing procedure may be in tension). 5) Which information needs to be available for an audit to be effective and reliable (e.g., source code, training data, documentation)? Do auditors need to have physical access to systems during operation in order to audit effectively? 6) What obligation to produce proof is necessary and proportionate for vendors/service providers? 7) How can we ensure the auditing is possible? Do auditing requirements need to be considered in the design of algorithmic systems (“auditable by construction”)? 8) Rules for publicity: When an audit is negative, and the problems are not solved, what should be the behavior of the auditor, in what way can it be made public that a failure occurred? 9) Who audits the auditors? How to make sure the auditors are held accountable?

We furthermore demand the provision of adequate resources to support/fund research projects on developing models to effectively audit algorithmic systems.

### **/ Support civil society organizations as watchdogs of ADM systems**

Our findings clearly indicate that the work of civil society organizations is crucial in effectively challenging opaque ADM systems. Through research and advocacy, and, often, in co-operation with academia and journalists, they repeatedly intervened in policy debates around those systems over recent years, in several cases effectively making sure that the public interest and fundamental rights are duly considered both before and after their deployment in many European countries.

Civil society actors should, therefore, be supported as watchdogs of the “automating society”. As such, they are an integral component of any effective accountability framework for ADM systems.

### **/ Ban face recognition that might amount to mass surveillance**

Not all ADM systems are equally dangerous, and a risk-based approach to regulation, such as Germany’s and the EU’s, correctly reflects this. But in order to provide workable accountability for systems that are identified as risky, effective oversight and enforcement mechanisms must be put in place. This is all the more important for those deemed at “high risk” of infringing on users’ rights.

A crucial example that emerged from our findings is face recognition. ADM systems that are based on biometric technologies, including face recognition, have been shown to pose a particularly serious threat to the public interest and fundamental rights, as they clear the path to indiscriminate mass surveillance – and especially as they are widely, and opaquely, deployed nonetheless.

We demand that public uses of face recognition that might amount to mass surveillance are decisively banned until further notice, and urgently, at the EU level.

Such technologies may even be considered as already illegal in the EU, at least for certain uses, if deployed without “specific consent” of the scanned subjects. This legal reading has been suggested by the authorities in Belgium, who issued a landmark fine for face recognition deployments in the country.

### 3. Enhance algorithmic literacy and strengthen public debate on ADM systems

More transparency of ADM systems can only be truly useful if those confronted with them, such as regulators, government, and industry bodies, can deal with those systems and their impact in a responsible and prudent manner. In addition, those affected by these systems need to be able to understand, where, why, and how these systems are deployed. This is why we need to enhance algorithmic literacy at all levels, with important stakeholders as well as the general public, and to reinforce more diverse public debates about ADM systems and their impact on society.

#### / Establish independent centers of expertise on ADM

Together with our demand for algorithmic auditing and supporting research, we call for the establishment of independent centers of expertise on ADM at the national level to monitor, assess, conduct research, report on, and provide advice to government and industry in coordination with regulators, civil society, and academia about the societal and human rights implications of the use of ADM systems. The overall role of these centers is to create a meaningful accountability system and to build capacity.

The national centers of expertise should involve civil society organizations, stakeholder groups, and existing enforcement bodies such as DPAs and national human rights bodies to benefit all aspects of the ecosystem and build trust, transparency, and cooperation between all actors.

As independent statutory bodies, the centers of expertise would have a central role in coordinating policy development and national strategies relating to ADM and in helping to build the capacity (competence/skills) of existing regulators, government, and industry bodies to respond to the increased use of ADM systems.

These centers should not have regulatory powers, but provide essential expertise on how to protect individual human rights and prevent collective and societal harm. They should, for instance, support small and medium-sized enterprises (SMEs) in fulfilling their obligations under human rights due diligence, including conducting human rights assessments or algorithmic impact assessments, and by registering ADM systems in the public register discussed above.

#### / Promote an inclusive and diverse democratic debate around ADM systems

Next to strengthening capacities and competencies with those deploying ADM systems, it is also vital to advance algorithmic literacy in the general public through broader debate and diverse programs.

Our findings suggest that ADM systems not only remain non-transparent to the wider public when they are in use, but that even the decision whether or not to deploy an ADM system in the first place is usually taken without either the knowledge or participation of the public.

There is, therefore, an urgent need to include the public (interest) in the decision-making on ADM systems from the very beginning.

More generally, we need a more diverse public debate about the impact of ADM. We need to move beyond exclusively addressing expert groups and make the issue more accessible to the wider public. That means speaking a language other than the techno-judicial to engage the public and spark interest.

In order to do so, detailed programs – to build and advance digital literacy – should also be put in place. If we aim at enhancing an informed public debate and creating digital autonomy for citizens in Europe, we have to start by building and advancing digital literacy, with a specific focus on the social, ethical, and political consequences of adopting ADM systems.



# European Union: Setting the stage for the **future** of ADM in Europe



**As automated decision-making systems take center stage for distributing rights and services within Europe, institutions across the region increasingly recognize their role in public life, both in terms of opportunities and challenges.**

By Kristina Penner and Fabio Chiusi



Since our first report in January 2019 – and even though the EU is still mired in the broader debate around “trustworthy” artificial intelligence – several bodies, from the EU Parliament to the Council of Europe, have published documents aimed at setting the EU and Europe on a course to deal with ADM over the coming years, if not decades.

In summer 2019, newly elected Commission President, Ursula von der Leyen, a self-stated “tech optimist”, [pledged](#) to put forward “legislation for a coordinated European approach on the human and ethical implications of artificial intelligence” and to “regulate Artificial Intelligence (AI)” within 100 days of taking office. Instead, in February 2020, the European Commission published a [‘White Paper’ on AI](#) containing “ideas and actions” – a strategy package that aims to inform citizens and pave the way for future legislative action. It also makes the case for European “technological sovereignty”: in Von der Leyen’s [own terms](#), this translates into “the capability that Europe must have to make its own choices, based on its own values, respecting its own rules”, and should “help make tech optimists of us all”.

A second fundamental endeavor affecting ADM in Europe is the Digital Services Act (DSA), announced in Von der Leyen’s ‘Agenda for Europe’ and supposed to replace the E-Commerce Directive that has been in place since 2000. It aims to “upgrade our liability and safety rules for digital platforms,

services and products, and complete our Digital Single Market” – thus leading to foundational debates around the role of ADM in content moderation policies, intermediary liability, and freedom of expression more generally<sup>1</sup>.

An explicit focus on ADM systems can be found in a Resolution [approved](#) by the EU Parliament’s Internal Market and Consumer Protection Committee, and in a [Recommendation](#) “on the human rights impacts of algorithmic systems” by the Council of Europe’s Committee of Ministers.

The Council of Europe (CoE), in particular, was found to be playing an increasingly important role in the policy debate on AI over the last year, and even though its actual impact on regulatory efforts remains to be seen, a case can be made for it to serve as the “guardian” of human rights. This is most apparent in the Recommendation, [‘Unboxing Artificial Intelligence: 10 steps to protect Human Rights’](#), by the CoE’s Commissioner on Human Rights, Dunja Mijatović, and in the work of the Ad Hoc Committee on AI (CAHAI) founded in September 2019.

Many observers see a fundamental tension between business and rights imperatives in how EU institutions, and es-

---

<sup>1</sup> Detailed remarks and recommendations around ADM systems in the context of the DSA can be found in the outputs of AlgorithmWatch’s ‘Governing Platforms’ project.

***MANY OBSERVERS SEE A  
FUNDAMENTAL TENSION BETWEEN  
BUSINESS AND RIGHTS IMPERATIVES  
IN HOW EU INSTITUTIONS, AND  
ESPECIALLY THE COMMISSION, ARE  
FRAMING THEIR REFLECTIONS AND  
PROPOSALS ON AI AND ADM.***



pecially the Commission, are framing their reflections and proposals on AI and ADM. On the one hand, Europe wants “to increase the use of, and demand for, data and data-enabled products and services throughout the Single Market”; thereby becoming a “leader” in business applications of AI, and boosting the competitiveness of EU firms in the face of mounting pressure from rivals in the US and China. This is all the more important for ADM, the assumption being that, through this “data-agile” economy, the EU “can become a leading role model for a society empowered by data to make better decisions – in business and the public sector”. As the White Paper on AI puts it, “Data is the lifeblood of economic development”.

Whereas on the other hand, the automatic processing of data about a citizen’s health, job, and welfare can form decisions with discriminatory and unfair results. This “dark side” of algorithms in decision-making processes is tackled in the EU toolbox through a series of principles. In the case of high-risk systems, rules should guarantee that automated decision-making processes are compatible with human rights and meaningful democratic checks and balances. This is an approach that EU institutions label as “human-centric” and unique, and as fundamentally opposed to those applied in the US (led by profit) and China (led by national security and mass surveillance).

However, doubts have emerged as to whether Europe can attain both goals at the same time. Face recognition is a case in point: even though, as this report shows, we now have plenty of evidence of unchecked and opaque deployments in most member countries, the EU Commission has failed to act swiftly and decisively to protect the rights of European citizens. As leaked drafts of the EC White Paper on AI revealed<sup>2</sup>, the EU was about to ban “remote biometric identification” in public places, before shying away at the last minute and promoting a “broad debate” around the issue instead.

In the meantime, controversial applications of ADM for border controls, even including face recognition, are still being pushed in EU-funded projects.

**“WE WANT TO ENCOURAGE OUR BUSINESSES, OUR RESEARCHERS, THE INNOVATORS, THE ENTREPRENEURS, TO DEVELOP ARTIFICIAL INTELLIGENCE. AND WE WANT TO ENCOURAGE OUR CITIZENS TO FEEL CONFIDENT TO USE IT. WE HAVE TO UNLEASH THIS POTENTIAL.”**  
URSULA VON DER LEYEN

## Policies and political debates

### / The European Data Strategy Package and the White Paper on AI

While the promised comprehensive legislation “for a coordinated European approach on the human and ethical implications of artificial intelligence”, announced in Von der Leyen’s “Agenda for Europe”, has not been put forward within her “first 100 days in office”, the EU Commission published a series of documents that provide a set of principles and ideas to inform it.

On February 19, 2020, a “[European Strategy for Data](#)” and a “[White Paper on Artificial Intelligence](#)” were jointly published, laying out the main principles of the EU’s strategic approach to AI (including ADM systems, even though they are not explicitly mentioned). These principles include, putting “people first” (“technology that works for the people”), technological neutrality (no technology is good or bad per se; this is to be determined by its use only) and, of course, sovereignty, and optimism. As Von der Leyen [puts it](#): “We want to encourage our businesses, our researchers, the innovators, the entrepreneurs,

to develop Artificial Intelligence. And we want to encourage our citizens to feel confident to use it. We have to unleash this potential”.

The underlying idea is that new technologies should not come with new values. The “new digital world” envisioned by the Von der Leyen administration should fully protect human and civil rights. “Excellence” and “trust”, highlighted in the very title of the White Paper, are considered the twin pillars upon which a European model of AI can and should stand, differentiating it from the strategies of both the US and China.

However, this ambition is lacking in the detail of the White Paper. For example, the White Paper lays out a risk-based approach to AI regulation, in which regulation is proportional to the impact of “AI” systems on citizen’s lives. “For high-risk cases, such as in health, polic-

<sup>2</sup> <https://www.politico.eu/article/eu-considers-temporary-ban-on-facial-recognition-in-public-spaces/>

ing, or transport”, it reads, “AI systems should be transparent, traceable and guarantee human oversight”. Testing and certification of adopted algorithms are also included among the safeguards that should be put in place, and should become as widespread as for “cosmetics, cars or toys”. Whereas, “less risky systems” only have to follow voluntary labelling schemes instead: “The economic operators concerned would then be awarded a quality label for their AI applications”.

But critics [noted](#) that the very definition of “risk” in the Paper is both circular and too vague, allowing for several impactful ADM systems to fall through the cracks of the proposed framework<sup>3</sup>.

Comments<sup>4</sup> gathered in the public consultation, between February and June 2020, highlight how controversial this idea is. 42.5% of respondents agreed that “compulsory requirements” should be limited to “high-risk AI applications”, while 30.6% doubted such a limitation.

Moreover, there is no description of a clear mechanism for the enforcement of such requirements. Neither is there a description of a process to move towards one.

The consequences are immediately visible for biometric technologies, and face recognition in particular. On this, the White Paper proposed a distinction between biometric “authentication”, which is seen as non-controversial (e.g., face recognition to unlock a smartphone), and remote biometric “identification” (such as deployment in public squares to identify protesters), which could arouse serious human rights and privacy concerns.

Only cases in the latter category would be problematic under the scheme proposed by the EU. The [FAQ](#) in the White Paper states: “this is the most intrusive form of facial recog-

nition and in principle prohibited in the EU”, unless there is “substantial public interest” in its deployment.

The explanatory document claims that “allowing facial recognition is currently the exception”, but findings in this report arguably contradict that view: face recognition seems to be rapidly becoming the norm. A leaked draft version of the White Paper seemed to recognize the urgency of the problem, by including the idea of a three-to-five-year moratorium on live uses of face recognition in public places, until – and if – a way to reconcile them with democratic checks and balances could be found.

Just before the official release of the White Paper, even EU Commissioner, Margrethe Vestager, [called](#) for a “pause” on these uses.

However, immediately after Vestager’s call, Commission officials added that this “pause” would not prevent national governments from using face recognition according to the existing rules. Ultimately, the final draft of the White Paper scrapped any mention of a moratorium, and called for “a broad European debate on the specific circumstances, if any, which might justify” its use for live biometric identification purposes instead. Among them, the White Paper includes justification, proportionality, the existence of democratic safeguards, and respect for human rights.

Throughout the whole document, risks associated with AI-based technologies are more generally labeled as “potential”, while the benefits are portrayed as very real and immediate. This led many<sup>5</sup> in the human rights community to claim that the overall narrative of the White Paper suggests a worrisome reversal of EU priorities, putting global competitiveness ahead of the protection of fundamental rights.

Some foundational issues are, however, raised in the documents. For example, the interoperability of such solutions and the creation of a network of research centers focused on applications of AI aimed at “excellence” and competence-building.

The objective is “to attract over €20 billion of total investments in the EU per year in AI over the next decade”.

3 “To give two examples: VioGén, an ADM system to forecast gender-based violence, and Ghostwriter, an application to detect exam fraud, would most likely fall between the cracks of regulation, even though they come with tremendous risks” (<https://algorithmwatch.org/en/response-european-commission-ai-consultation/>)

4 “In total, 1215 contributions were received, of which 352 were on behalf of a company or business organisations/associations, 406 from citizens (92% EU citizens), 152 on behalf of academic/research institutions, and 73 from public authorities. Civil society voices were represented by 160 respondents (among which 9 consumer’s organisations, 129 non-governmental organisations and 22 trade unions). 72 respondents contributed as “others””. Comments arrived “from all over the world”, including countries such as “India, China, Japan, Syria, Iraq, Brazil, Mexico, Canada, the US and the UK”. (from the Consultation’s Summary Report, linked here: <https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence>)

5 Among them: Access Now ([https://www.accessnow.org/cms/assets/uploads/2020/05/EU-white-paper-consultation\\_AccessNow\\_May2020.pdf](https://www.accessnow.org/cms/assets/uploads/2020/05/EU-white-paper-consultation_AccessNow_May2020.pdf)), AI Now (<https://ainowinstitute.org/ai-now-comments-to-eu-whitepaper-on-ai.pdf>), EDRI (<https://edri.org/can-the-eu-make-ai-trustworthy-no-but-they-can-make-it-just/>) – and AlgorithmWatch (<https://algorithmwatch.org/en/response-european-commission-ai-consultation/>).

A certain technological determinism seems to also affect the White Paper. “It is essential”, it reads, “that public administrations, hospitals, utility and transport services, financial supervisors, and other areas of public interest rapidly begin to deploy products and services that rely on AI in their activities. A specific focus will be in the areas of healthcare and transport where technology is mature for large-scale deployment.”

However, it remains to be seen whether suggesting a rushed deployment of ADM solutions in all spheres of human activity is compatible with the EU Commission’s efforts in addressing the structural challenges brought about by ADM systems to rights and fairness.

## / EU Parliament’s Resolution on ADM and consumer protection

A [Resolution](#), passed by the EU Parliament in February 2020, more specifically tackled ADM systems in the context of consumer protection. The Resolution correctly pointed out that “complex algorithmic-based systems and automated decision-making processes are being made at a rapid pace”, and that “opportunities and challenges presented by these technologies are numerous and affect virtually all sectors”. The text also highlights the need for “an examination of the current EU legal framework”, to assess whether “it is able to respond to the emergence of AI and automated decision-making”.

Calling for a “common EU approach to the development of automated decision-making processes”, the Resolution details several conditions that any such systems should possess to remain consistent with European values. Consumers should be “properly informed” about how algorithms affect their lives, and they should have access to a human with decision-making power so that decisions can be checked and corrected if needed. They should also be informed, “when prices of goods or services have been personalized on the basis of automated decision-making and profiling of consumer behavior”.

In reminding the EU Commission that a carefully drafted risk-based approach is needed, the Resolution points out that safeguards need to take into consideration that ADM

THROUGHOUT THE WHOLE DOCUMENT, RISKS ASSOCIATED WITH AI-BASED TECHNOLOGIES ARE MORE GENERALLY LABELED AS “POTENTIAL”, WHILE THE BENEFITS ARE PORTRAYED AS VERY REAL AND IMMEDIATE.

systems “may evolve and act in ways not envisaged when first placed on the market”, and that liability is not always easy to attribute when harm comes as a result of the deployment of an ADM system.

The Resolution echoes [art. 22 of the GDPR](#) when it notes that a human subject must always be in the loop when “legitimate public interests are at stake”, and should always be ultimately responsible for decisions in the “medical, legal and accounting professions, and for the banking sector”. In particular, a “proper” risk assessment should precede any automation of professional services.

Finally, the Resolution lists detailed requirements for quality and transparency in data governance: among them, “the importance of using only high-quality and unbiased data sets in order to improve the output of algorithmic systems and boost consumer trust and acceptance”; using “explainable and unbiased algorithms”; and the need for a “review structure” that allows affected consumers “to seek human review of, and redress for, automated decisions that are final and permanent”.

## / Making the most of the EU Parliament’s “right of initiative”

In her inaugural address, von der Leyen clearly [expressed](#) her support for a “right of initiative” for the European Parliament. “When this House, acting by majority of its Members, adopts Resolutions requesting the Commission to submit legislative proposals”, she said, “I commit to responding with a legislative act in full respect of the proportionality, subsidiarity, and better law-making principles”.

If “AI” is indeed a revolution requiring a dedicated legislative package, allegedly coming over the first quarter of 2021, elected representatives want to have a say about it. This, coupled with von der Leyen’s stated intent of empowering their legislative capabilities, could even result in what Politico [labeled](#) a “Parliament moment”, with parliamentary committees starting to draft several different reports as a result.

Each report investigates specific aspects of automation in public policy that, even though they are meant to shape upcoming “AI” legislation, are relevant for ADM.

# *IF "AI" IS INDEED A REVOLUTION REQUIRING A DEDICATED LEGISLATIVE PACKAGE, ELECTED REPRESENTATIVES WANT TO HAVE A SAY ABOUT IT.*

For example, through its "Framework of ethical aspects of artificial intelligence, robotics and related technologies", the Committee for Legal Affairs [calls](#) for the constitution of a "European Agency for Artificial Intelligence" and, at the same time, for a network of national supervisory authorities in each Member State to make sure that ethical decisions involving automation are and remain ethical.

In "[Intellectual property rights for the development of artificial intelligence technologies](#)", the same committee [lays out](#) its view for the future of intellectual property and automation. For one, the draft report states that "mathematical methods are excluded from patentability unless they constitute inventions of a technical nature", while at the same time claiming that, regarding algorithmic transparency, "reverse engineering is an exception to the trade secrets rule".

The report goes as far as considering how to protect "technical and artistic creations generated by AI, in order to encourage this form of creation", imagining that "certain works generated by AI can be regarded as equivalent to intellectual works and could therefore be protected by copyright".

Lastly, in a third [document](#) ("Artificial Intelligence and civil liability"), the Committee details a "Risk Management Approach" for the civil liability of AI technologies. According to it, "the party who is best capable of controlling and managing a technology-related risk is held strictly liable, as a single entry point for litigation".

Important principles concerning the use of ADM in the criminal justice system can be found in the Committee of Civil Liberties, Justice and Home Affairs' [report](#) on "Artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters". After a detailed list of actual and current uses of "AI" – these are, actually, ADM systems – by police forces<sup>6</sup>, the Committee "considers it necessary to create a clear and fair regime for assigning legal responsibility for the potential adverse consequences produced by these advanced digital technologies".

It then goes about detailing some of its features: no fully automated decisions<sup>7</sup>, algorithmic explainability that is "intelligible to users", a "compulsory fundamental rights impact assessment (...) of any AI systems for law enforcement or judiciary" prior to its deployment or adoption, plus "periodic mandatory auditing of all AI systems used by law enforcement and the judiciary to test and evaluate algorithmic systems once they are in operation".

6 On p. 5, the report states: "AI applications in use by law enforcement include applications such as facial recognition technologies, automated number plate recognition, speaker identification, speech identification, lip-reading technologies, aural surveillance (i.e. gunshot detection algorithms), autonomous research and analysis of identified databases, forecasting (predictive policing and crime hotspot analytics), behaviour detection tools, autonomous tools to identify financial fraud and terrorist financing, social media monitoring (scraping and data harvesting for mining connections), international mobile subscriber identity (IMSI) catchers, and automated surveillance systems incorporating different detection capabilities (such as heartbeat detection and thermal cameras)".

7 "in judicial and law enforcement contexts, the final decision always needs to be taken by a human" (p. 6)



A moratorium on face recognition technologies for law enforcement is also called for in the report, “until the technical standards can be considered fully fundamental rights compliant, results derived are non-discriminatory, and there is public trust in the necessity and proportionality for the deployment of such technologies”.

The aim is to eventually boost the overall transparency of such systems, and advising Member States to provide a “comprehensive understanding” of the AI systems adopted by law enforcement and the judiciary, and – along the lines of a “[public register](#)” – to detail “the type of tool in use, the types of crime they are applied to, and the companies whose tools are being used”.

The Culture and Education Committee and the Industrial Policy Committee were also [working](#) on their own reports at the time of writing.

All these initiatives led to the [creation](#) of a Special Committee on “Artificial Intelligence in a Digital Age” (AIDA) on June 18, 2020. Composed of 33 members, and with an initial duration of 12 months, it will “analyse the future impact” of AI on the EU economy, and “in particular on skills, employment, fintech, education, health, transport, tourism, agriculture, environment, defence, industry, energy and e-government”.

## **/ High-Level Expert Group on AI & AI Alliance**

In 2018, the High-Level Expert Group (HLEG) on AI, an expert committee made up of 52 experts, was set up by the European Commission to support the implementation of the European strategy on AI, to identify principles that should be observed in order to achieve “trustworthy AI”, and, as the steering committee of the supporting AI Alliance, to create an open multi-stakeholder platform (consisting of more than 4000 members at the time of writing) to provide broader input for the work of the AI high-level expert group.

After the publication of the first draft of the AI Ethics Guidelines for Trustworthy AI in December 2018, followed by feedback from more than 500 contributors, a revised version was published in April 2019. It puts forward a “human-centric approach” to achieve legal, ethical, and robust AI throughout the system’s entire life cycle. However, it remains a voluntary framework without concrete and applicable recommendations on operationalization, implementation, and enforcement.

Civil society, consumer protection, and rights organizations commented and called for the translation of the guidelines into tangible rights for people<sup>8</sup>. For example, digital rights non-profit Access Now, a member of the HLEG, urged the EC to clarify how different stakeholders can test, apply, improve, endorse, and enforce “Trustworthy AI” as a next step, while at the same time recognizing the need to determine Europe’s red lines.

In [an op-ed](#), two other members of the HLEG claimed that the group had “worked for one-and-a-half years, only for its detailed proposals to be mostly ignored or mentioned only in passing” by the European Commission who drafted the final version.<sup>9</sup> They also argued that, because the group was initially tasked with identifying risks and “red lines” for AI, members of the group pointed to autonomous weapon systems, citizen scoring, and automated identification of individuals by using face recognition as implementations of AI that should be avoided. However, representatives of industry, who dominate the committee<sup>10</sup>, succeeded in getting these principles deleted before the draft was published.

This imbalance towards highlighting the potentials of ADM, compared to the risks, can also be observed throughout its second deliverable. In the HLEG’s “Policy and investment recommendations for trustworthy AI in Europe”, made [public](#) in June 2019, there are 33 recommendations meant to “guide Trustworthy AI towards sustainability, growth and competitiveness, as well as inclusion – while empowering, benefiting and protecting human beings”. The document is predominantly a call to boost the uptake and scaling of AI in the private and public sector by investing in tools and applications “to help vulnerable demographics” and “to leave no one behind”.

Nevertheless, and despite all legitimate criticism, both guidelines still express critical concerns and demands regarding automated decision-making systems. For example, the ethics guidelines [formulate](#) “seven key requirements that AI systems should meet in order to be trustworthy”. These guidelines go on to provide guidance for the prac-

8 E.g., the European Consumer Protection Organization (BEUC): [https://www.beuc.eu/publications/beuc-x-2020-049\\_response\\_to\\_the\\_ecs\\_white\\_paper\\_on\\_artificial\\_intelligence.pdf](https://www.beuc.eu/publications/beuc-x-2020-049_response_to_the_ecs_white_paper_on_artificial_intelligence.pdf)

9 Mark Coeckelbergh and Thomas Metzinger: Europe needs more guts when it comes to AI ethics, <https://background.tagesspiegel.de/digitalisierung/europe-needs-more-guts-when-it-comes-to-ai-ethics>

10 The group was composed of 24 business representatives, 17 academics, five civil society organizations and six other members, like the European Union Agency for Fundamental Rights

tical implementation of each requirement: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability.

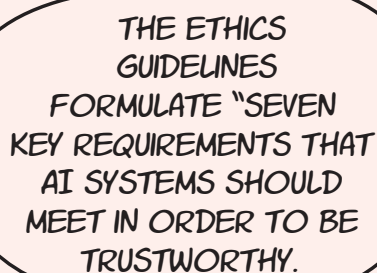
The guidelines also provide a concrete pilot, called the “Trustworthy AI assessment list”, which is aimed at making those high-level principles operational. The goal is to have it adopted “when developing, deploying or using AI systems”, and adapted “to the specific use case in which the system is being applied”.

The list includes many issues that are associated with the risk of infringing on human rights through ADM systems. These include the lack of human agency and oversight, technical robustness and safety issues, the inability to avoid unfair bias or provide equal and universal access to such systems, and the lack of meaningful access to data fed into them.

Contextually, the pilot list included in the guidelines provides useful questions to help those who deploy ADM systems. For example, it calls for “a fundamental rights impact assessment where there could be a negative impact on fundamental rights”. It also asks whether “specific mechanisms of control and oversight” have been put in place in the cases of “self-learning or autonomous AI” systems, and whether processes exist “to ensure the quality and integrity of your data”.

Detailed remarks also concern foundational issues for ADM systems, such as their transparency and explainability. Questions include “to what extent the decisions and hence the outcome made by the AI system can be understood?” and “to what degree the system’s decision influences the organisation’s decision-making processes?” These questions are highly relevant to assess the risks posed by deploying such systems.

To avoid bias and discriminatory outcomes, the guidelines point to “oversight processes to analyze and address the system’s purpose, constraints, requirements and decisions in a clear and transparent manner”, while at the same time demanding stakeholder participation throughout the whole process of implementing AI systems.



THE ETHICS  
GUIDELINES  
FORMULATE “SEVEN  
KEY REQUIREMENTS THAT  
AI SYSTEMS SHOULD  
MEET IN ORDER TO BE  
TRUSTWORTHY.”

Added to that, the recommendations on policy and investment foresee the determination of red lines through an institutionalized “dialogue on AI policy with affected stakeholders”, including experts in civil society. Furthermore, they urge to “ban AI-enabled mass scale scoring of individuals as defined in [the] Ethics Guidelines, and [to] set very clear and strict rules for surveillance for national security purposes and other purposes claimed to be in the public or national interest”. This ban would include biometric identification technologies and profiling.

Relevant to automated decision-making systems, the document also states that “clearly defin[ing] if, when and how AI can be used (...) will be crucial for the achievement of Trustworthy AI”, warning that “any form of citizen scoring can lead to the loss of [the citizen’s] autonomy and endanger the principle of non-discrimination”, and “therefore should only be used if there is a clear justification, under proportionate and fair measures”. It further stresses that “transparency cannot prevent non-discrimination or ensure fairness”. This means that the possibility of opting out of a scoring mechanism should be provided, ideally without any detriment to the individual citizen.

On the one hand, the document acknowledges “that, while bringing substantial benefits to individuals and society, AI systems also pose certain risks and may have a negative impact, including impacts which may be difficult to anticipate, identify or measure (e.g. on democracy, the rule of law and distributive justice, or on the human mind itself.)”. On the other, however, the group claims that “unnecessarily prescriptive regulation should be avoided”.

In July 2020, the AI HLEG also presented their final [Assessment List for Trustworthy Artificial Intelligence \(ALTAI\)](#), compiled after a piloting process together with 350 stakeholders.

The list, which is entirely voluntary and devoid of any regulatory implications, aims to translate the seven requirements laid out in the AI HLEG’s Ethics Guidelines into action. The intention is to provide whoever wants to implement AI solutions that are compatible with EU values – for example, designers and developers of AI systems, data scientists, procurement officers or specialists, and legal/compliance officers – with a self-assessment toolkit.

## / Council of Europe: how to safeguard human rights in ADM

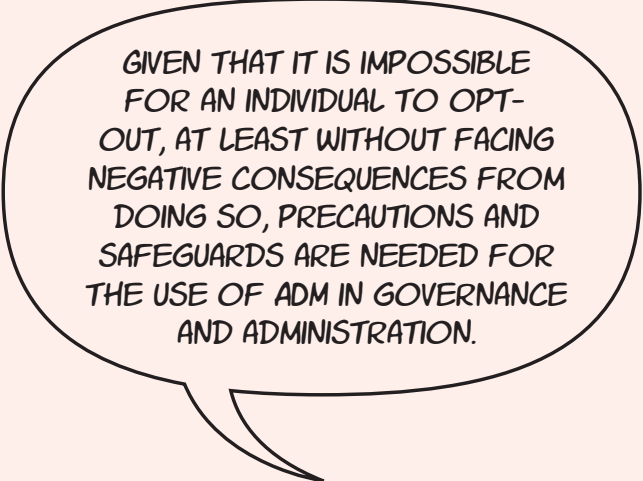
In addition to the Ad hoc Committee on Artificial Intelligence (CAHAI), set up in September 2019, the Committee of Ministers of the Council of Europe<sup>11</sup> has published a substantial and persuasive framework.

Envisioned as a standard-setting instrument, its “[Recommendation to Member states on the human rights impacts of algorithmic systems](#)” describes “significant challenges”<sup>12</sup> that arise with the emergence and our “increasing reliance” on such systems, and that are relevant “to democratic societies and the rule of law”.

The framework, which underwent a public consultation period with detailed [comments](#) from civil society organizations, goes beyond the EU Commission’s White Paper when it comes to safeguarding values and human rights.

The Recommendation thoroughly analyzes the effects and evolving configurations of algorithmic systems (Appendix A) by focusing on all stages of the process that go into making an algorithm, i.e., procurement, design, development, and ongoing deployment.

While generally following the ‘human-centric AI approach’ of the HLEG guidelines, the Recommendation outlines actionable “obligations of States” (Appendix B) as well as responsibilities for private sector actors (Appendix C). In addition, the Recommendation adds principles such as “in-



GIVEN THAT IT IS IMPOSSIBLE FOR AN INDIVIDUAL TO OPT-OUT, AT LEAST WITHOUT FACING NEGATIVE CONSEQUENCES FROM DOING SO, PRECAUTIONS AND SAFEGUARDS ARE NEEDED FOR THE USE OF ADM IN GOVERNANCE AND ADMINISTRATION.

formational self-determination”<sup>13</sup>, lists detailed suggestions for accountability mechanisms and effective remedies, and demands human rights impact assessments.

Even though the document clearly acknowledges the “significant potential of digital technologies to address societal challenges and to socially beneficial innovation and economic development”, it also urges caution. This is to ensure that those systems do not deliberately or accidentally perpetuate “racial, gender and other societal and labor force imbalances that have not yet been eliminated from our societies”.

On the contrary, algorithmic systems should be used proactively and sensitively to address these imbalances, and pay “attention to the needs and voices of vulnerable groups”.

Most significantly, however, the Recommendation identifies the potentially higher risk to human rights when algorithmic systems are used by Member States to deliver public services and policy. Given that it is impossible for an individual to opt-out, at least without facing negative consequences from doing so, precautions and safeguards are needed for the use of ADM in governance and administration.

11 The CoE is both “a governmental body where national approaches to European problems are discussed on an equal footing and a forum to find collective responses to these challenges.” Its work includes “the political aspects of European integration, safeguarding democratic institutions and the rule of law and protecting human rights – in other words, all problems which require concerted pan-European solutions.” Although the recommendations to the governments of members are not binding, in some cases the Committee may request the governments of members to inform it of the action taken by them with regard to such recommendations. (Art. 15b Statute). Relations between the Council of Europe and the European Union are set out in the (1) Compendium of Texts governing the relations between the Council of Europe and the European Union and in the (2) Memorandum of Understanding between the Council of Europe and the European Union.

12 Under the supervision of the Steering Committee on Media and Information Society (CDMSI) and prepared by the Committee of Experts on Human Rights Dimensions of Automated Data Processing and Different Forms of Artificial Intelligence (MSI-AUT).

13 “States should ensure that all design, development and ongoing deployment of algorithmic systems provide an avenue for individuals to be informed in advance about the related data processing (including its purposes and possible outcomes) and to control their data, including through interoperability”, reads Section 2.1 of Appendix B.

The Recommendation also addresses the conflicts and challenges arising from public-private-partnerships (“neither clearly public nor clearly private”) in a wide range of uses.

Recommendations for Member State governments include abandoning processes and refusing to use ADM systems, if “human control and oversight become impractical” or human rights are put at risk; deploying ADM systems if and only if transparency, accountability, legality, and the protection of human rights can be guaranteed “at all stages of the process”. Furthermore, the monitoring and evaluation of these systems should be “constant”, “inclusive and transparent”, comprised of a dialogue with all relevant stakeholders, as well as an analysis of the environmental impact and further potential externalities on “populations and environments”.

In Appendix A, the COE also defines “high-risk” algorithms for other bodies to draw inspiration from. More specifically, the Recommendation states that “the term “high-risk” is applied when referring to the use of algorithmic systems in processes or decisions that can produce serious consequences for individuals or in situations where the lack of alternatives prompts a particularly high probability of infringement of human rights, including by introducing or amplifying distributive injustice”.

The document, which did not require the unanimity of members to be adopted, is non-binding.

### **/ Regulation of terrorist content online**

After a long period of sluggish progress, the [regulation](#) to prevent the dissemination of terrorist content gained momentum in 2020. Should the adopted regulation still include automated and proactive tools for recognizing and removing content online, these would likely fall under Art. 22 of the GDPR.

As the European Data Protection Supervisor (EDPS) [puts it](#): “since the automated tools, as envisaged by the Proposal, could lead not only to the removal and retention of content (and related data) concerning the uploader, but also, ultimately, to criminal investigations on him or her, these tools would significantly affect this person, impacting on his or her right to freedom of expression and posing significant risks for him or her rights and freedoms”, and, therefore, fall under Art. 22(2).

Also, and crucially, it would require more substantive safeguards compared to those that the Commission currently foresees. As the advocacy group, European Digital Rights (EDRi), explains: “the proposed Terrorist Content Regulation needs substantive reform to live up to the Union’s values, and to safeguard the fundamental rights and freedoms of its citizens”.

An early stream of strong criticism on the initial proposal from civil society groups, European Parliament (EP) committees, including opinions and analysis by the European Union Agency for Fundamental Rights (FRA), EDRi, as well as a critical joint report by three UN Special Rapporteurs, highlighted threats to the right of freedom of expression and information, the right to freedom and pluralism of the media, the freedom to conduct a business and the rights to privacy and protection of personal data.

Critical aspects include an insufficient definition of terrorist content, the scope of the regulation (at present, this includes content for educational and journalistic purposes), the aforementioned call for “proactive measures”, a lack of effective judicial supervision, insufficient reporting obligations for law enforcement agencies, and missing safeguards for “cases where there are reasonable grounds to believe that fundamental rights are impacted” (EDRi 2019).

The EDPS stresses that such “suitable safeguards” should include the right to obtain human intervention and the right to an explanation of the decision reached through automated means (EDRi 2019).

Although safeguards that were suggested or demanded found their way into the EP’s draft report on the proposal, it is yet to be seen who can hold their breath longer going into the last round before the final vote. During closed-door dialogues between the EP, the new EC, and the EU Council (which began in October 2019), only minor changes are still possible, according to a leaked document.



# Oversight and regulation

## / First decisions on the compliance of ADM systems with the GDPR

“Although there was no great debate on facial recognition during the passage of negotiations on the GDPR and the law enforcement data protection directive, the legislation was designed so that it could adapt over time as technologies evolved. [...] Now is the moment for the EU, as it discusses the ethics of AI and the need for regulation, to determine whether – if ever – facial recognition technology can be permitted in a democratic society. If the answer is yes, only then do we turn [to] questions of how and safeguards and accountability to be put in place.” – EDPS, Wojciech Wiewiórowski.

“Facial recognition processing is an especially intrusive biometric mechanism, which bears important risks of privacy or civil liberties invasions for the people affected” – (CNIL 2019).

Since the last Automating Society report, we have seen the first cases of fines and decisions related to breaches of the regulation issued by national Data Protection Authorities (DPAs) based on the GDPR. The following case studies, however, show the limits of the GDPR in practice when it comes to Art. 22 relating to ADM systems and how it is leaving the privacy regulators to make assessments on a case-by-case basis.

In Sweden, a face recognition test project, conducted in one school class for a limited period of time, was found to violate several obligations of Data Protection Regulation (esp. GDPR Art. 2(14), Art. 9 (2)). (European Data Protection Board 2019).

A similar case is on hold after the French Commission Nationale de l'Informatique et des Libertés (CNIL raised concerns when two high schools planned to introduce face recognition technology in partnership with the US tech firm, Cisco. The opinion is non-binding, and the filed suit is ongoing<sup>14</sup>.

The ex-ante authorization by data regulators is not required to conduct such trials as the consent of the users is generally considered to be sufficient to process biometric data. And yet, in the Swedish case, it wasn't. This was due

to power imbalances between the data controller and the data subjects. Instead, an adequate impact assessment and prior consultation with the DPA were deemed necessary.

The European Data Protection Supervisor (EDPS) [confirmed this](#):

“Consent would need to be explicit as well as freely-given, informed and specific. Yet unquestionably a person cannot opt out, still less opt in, when they need access to public spaces that are covered by facial recognition surveillance. [...] Finally, the compliance of the technology with principles like data minimization and the data protection by design obligation is highly doubtful. Facial recognition technology has never been fully accurate, and this has serious consequences for individuals being falsely identified whether as criminals or otherwise. [...] It would be a mistake, however, to focus only on privacy issues. This is fundamentally an ethical question for a democratic society.” (EDPS 2019)

Access Now commented:

“As more facial recognition projects develop, we already see that the GDPR provides useful human rights safeguards that can be enforced against unlawful collection and use of sensitive data such as biometrics. But the irresponsible and often unfounded hype around the efficiency of such technologies and the underlying economic interest could lead to attempts by central and local governments and private companies to circumvent the law.”

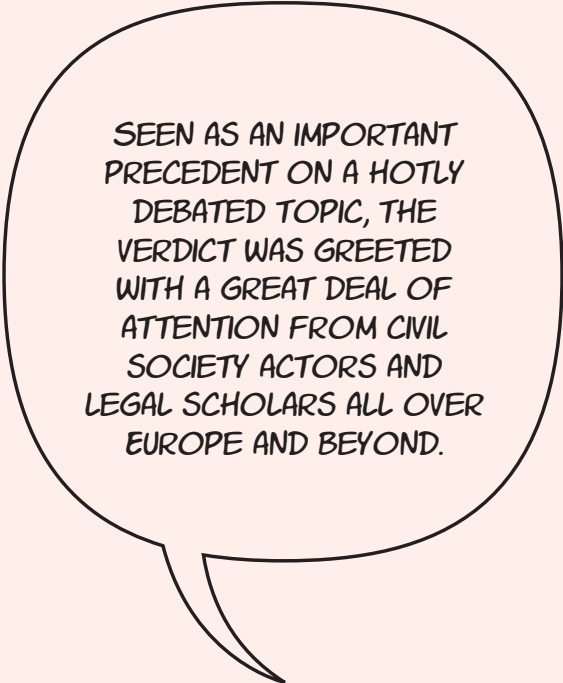
## / Automated Face Recognition in use by South Wales Police ruled unlawful

Over the course of 2020, the UK witnessed the first high profile application of the Law Enforcement Directive<sup>15</sup> concerning the use of face recognition technologies in public spaces by the police. Seen as an important precedent on a hotly debated topic, the verdict was greeted with a great deal of attention from civil society actors and legal scholars all over Europe and beyond <sup>16</sup>.

<sup>15</sup> The Law Enforcement Directive, in effect from May 2018, “deals with the processing of personal data by data controllers for ‘law enforcement purposes’ – which falls outside of the scope of the GDPR”. <https://www.dataprotection.ie/en/organisations/law-enforcement-directive>

<sup>16</sup> The decision was rendered on September, 4th, 2019 by the High Court sitting in Cardiff in the case *Bridges v. the South Wales Police* (High Court of Justice 2019).

<sup>14</sup> See France chapter and (Kayalki 2019)



SEEN AS AN IMPORTANT PRECEDENT ON A HOTLY DEBATED TOPIC, THE VERDICT WAS GREETED WITH A GREAT DEAL OF ATTENTION FROM CIVIL SOCIETY ACTORS AND LEGAL SCHOLARS ALL OVER EUROPE AND BEYOND.

The case was brought to court by Ed Bridges, a 37 years old man from Cardiff, who [claimed](#) his face was scanned without his consent both during Christmas shopping in 2017, and at a peaceful anti-arms protest a year later.

The court initially upheld the use of [Automated Facial Recognition technology](#) ("AFR") by South Wales Police, declaring it lawful and proportionate. But the decision was appealed by Liberty, a civil rights group, and the Court of Appeal of England and Wales decided to overturn the High Court's dismissal and [ruled it unlawful](#) on August 11, 2020.

In ruling against South Wales Police on three out of five grounds, the Court of Appeal [found](#) "fundamental deficiencies" in the existing normative framework around the use of AFR, that its deployment did not meet the principle of "proportionality", and, also, that an adequate Data Protection Impact Assessment (DPIA) had not been performed, lacking multiple crucial steps.

The court did not, however, rule that the system was producing discriminatory results, based either on sex or race,

as South Wales Police had not gathered sufficient evidence to make a judgment on that<sup>17</sup>. However, the court felt it was worth adding a noticeable remark: "We would hope that, as AFR is a novel and controversial technology, all police forces that intend to use it in the future would wish to satisfy themselves that everything reasonable which could be done had been done in order to make sure that the software used does not have a racial or gender bias."

After the ruling, Liberty [called](#) for South Wales Police and other police forces to withdraw the use of face recognition technologies.

## ADM in practice: border management and surveillance

While the EU Commission and its stakeholders debated whether to regulate or ban face recognition technologies, extensive trials of the systems were already underway all over Europe.

This section highlights a crucial and often overlooked link between biometrics and the EU's border management systems, clearly showing how technologies that can produce discriminatory results might be applied to individuals – e.g., migrants – who already suffer the most from discrimination.

### / Face recognition and the use of biometrics data in EU policies and practice

Over the last year, face recognition and other kinds of biometrics identification technologies garnered a lot of attention from governments, the EU, civil society, and rights organizations, especially concerning law enforcement and border management.

---

<sup>17</sup> The Police claimed it had no access to the demographic composition of the training dataset for the adopted algorithm, "Neoface". The Court notes that "the fact remains, however, that SWP have never sought to satisfy themselves, either directly or by way of independent verification, that the software program in this case does not have an unacceptable bias on grounds of race or sex".

Over 2019, the EC tasked a consortium of public agencies to “map the current situation of facial recognition in criminal investigations in all EU Member States,” to move “towards the possible exchange of facial data”. They commissioned the consultancy firm Deloitte to perform a feasibility study on the expansion of the Prüm system of face images. [Prüm](#) is an EU-wide system that connects DNA, fingerprint, and vehicle registration databases to allow mutual searching. The concern is that a pan-European, database of faces could be used for pervasive, unjustified, or illegal surveillance.

## **/ Border management systems without borders**

As reported in the previous edition of Automating Society, the implementation of an overarching interoperable smart border management system in the EU, initially proposed by the Commission back in 2013, is on its way. Although the new systems that have been announced (EES, ETIAS<sup>18</sup>, ECRIS-TCN<sup>19</sup>) will only start operating in 2022, the Entry/Exit System (EES) regulation has already introduced face images as biometric identifiers and the use of face recognition technology for verification purposes for the first time in EU law<sup>20</sup>.

The European Fundamental Rights Agency (FRA) confirmed the changes: “the processing of facial images is expected to be introduced more systematically in large-scale EU-level IT systems used for asylum, migration and security purposes [...] once the necessary legal and technical steps are completed”.

According to Ana Maria Ruginis Andrei, from the European Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice (eu-LISA), this expanded new interoperability architecture was “assembled in order to forge the perfect engine to

successfully fight against the threats to internal security, to effectively control migration and to overcome blind spots regarding identity management”. In practice, this means to “hold the fingerprints, facial images, and other personal data of up to 300 million non-EU nationals, merging data from five separate systems.” (Campbell 2020)

## **/ ETIAS: automated border security screenings**

The [European Travel Information and Authorization System](#) (ETIAS), which is still not in operation at the time of writing, will use different databases to automate the digital security screening of non-EU travelers (who do not need a visa, or “visa-waiver”) before they arrive in Europe.

This system is going to gather and analyze data for the advanced “verification of potential security or irregular migration risks” (ETIAS 2020). The aim is to “facilitate border checks; avoid bureaucracy and delays for travelers when presenting themselves at the borders; ensure a coordinated and harmonised risk assessment of third-country nationals” (ETIAS 2020).

Ann-Charlotte Nygård, head of FRA’s Technical Assistance and Capacity Building unit, sees two specific risks concerning ETIAS: “first, the use of data that could lead to the unintentional discrimination of certain groups, for instance if an applicant is from a particular ethnic group with a high in-migration risk; the second relates to a security risk assessed on the basis of past convictions in the country of origin. Some such earlier convictions could be considered unreasonable by Europeans, such as LGBT convictions in certain countries. To avoid this, [...] algorithms need to be audited to ensure that they do not discriminate and this kind of auditing would involve experts from interdisciplinary areas” (Nygård 2019).

## **/ iBorderCtrl: face recognition and risk scoring at the borders**

iBorderCtrl was a project that involved security agencies from Hungary, Latvia, and Greece that [aimed](#) to “enable faster and thorough border control for third country nationals crossing the land borders of EU Member States”. iBorderCtrl used face recognition technology, a lie detector, and a scoring system to prompt a human border policeman if it deemed someone dangerous or if it deemed their right to entry was questionable.

18 ETIAS (EU Travel Information and Authorisation System), is the new “visa waiver” system for EU border management developed by eu-LISA. “The information submitted during the application will be automatically processed against existing EU databases (Eurodac, SIS and VIS), future systems EES and ECRIS-TCN, and relevant Interpol databases. This will enable advance verification of potential security, irregular migration and public health risks” (ETIAS 2019).

19 The European Criminal Records Information System – Third Country Nationals (ECRIS-TCN), to be developed by eu-LISA, will be a centralized hit/no-hit system to supplement the existing EU criminal records database (ECRIS) on non-EU nationals convicted in the European Union.

20 EES will enter into operation in the first quarter of 2022, ETIAS will follow by the end of 2022 – set out to be “game changers in the European Justice and Home Affairs (JHA) area”.

The iBorderCtrl project came to an end in August 2019, and the results – for any potential EU-wide implementation of the system – were contradictory.

Although it “will have to be defined, how far the system or parts of it will be used”, the project’s “Outcomes” page sees “the possibility of integrating the similar functionalities of the new ETIAS system and extending the capabilities of taking the border crossing procedure to where the travellers are (bus, car, train, etc.)”.

However, the modules this refers to were not specified, and the ADM-related tools that were tested have not been publicly evaluated.

At the same time, the project’s [FAQ](#) page confirmed that the system that was tested is not considered to be “currently suitable for deployment at the border (...) due to its nature as a prototype and the technological infrastructure on EU level”. This means that “further development and an integration in the existing EU systems would be required for a use by border authorities.”

In particular, while the iBorderCtrl Consortium was able to show, in principle, the functionality of such technology for border checks, it is also clear that ethical, legal, and societal constraints need to be addressed prior to any actual deployment.

### **/ Related Horizon2020 projects**

Several follow-up projects focused on testing and developing new systems and technologies for Border Management and Surveillance, under the Horizon2020 program. They are listed on the European Commission’s CORDIS website, which provides information on all EU-supported research activities related to it.

The site [shows](#) that 38 projects are currently running under the “H2020-EU.3.7.3. – Strengthen security through border management” program/topic of the European Union. Its parent program – “Secure societies – Protecting freedom and security of Europe and its citizens”, boasts an overall budget of almost 1.7 billion euros and funds 350 projects – claims to tackle “insecurity, whether from crime, violence, terrorism, natural or man-made disasters, cyber attacks or privacy abuses, and other forms of social and economic disorders increasingly affect[ing] citizens” through projects mainly developing new technological systems based on AI and ADM.

Some projects have already finished and/or their applications are already in use – for example, FastPass, ABC4EU, MOBILEPASS, and EFFISEC – all of which looked into requirements for “integrated, interoperable Automated Border Control (ABC)”, identification systems, and “smart” gates at different border crossings.

TRESSPASS is an ongoing project that started in June 2018 and will finish in November 2021. The EU [contributes](#) almost eight million euros to the project, and the coordinators of iBorderCRL (as well as FLYSEC and XP-DITE) are aiming to “leverage the results and concepts implemented and tested” by iBorderCRL and “expand[ing] them into a multi-modal border crossing risk-based security solution within a strong legal and ethics framework.” (Horizon2020 2019)

The project has the stated goal of turning security checks at border crossing points from the old and outdated “Rule-based” to a new “Risk-based” strategy. This includes applying biometric and sensing technologies, a risk-based management system, and relevant models to assess identity, possessions, capability, and intent. It aims to enable checks through “links to legacy systems and external databases such as VIS/SIS/PNR” and is collecting data from all the above data sources for security purposes.

Another pilot project, FOLDOUT, started in September 2018 and will finish in February 2022. The EU contributes €8,199,387,75 to the project to develop “improved methods for border surveillance” to counter irregular migration with a focus on “detecting people through dense foliage in extreme climates” [...] by combining “various sensors and technologies and intelligently fuse[ing] these into an effective and robust intelligent detection platform” to suggest reaction scenarios. Pilots are underway in Bulgaria, with demonstration models in Greece, Finland, and French Guiana.

MIRROR, or Migration-Related Risks caused by misconceptions of Opportunities and Requirement, started in June 2019 and will finish in May 2022. The EU contributes just over five million euros to the project, which aims to “understand how Europe is perceived abroad, detect discrepancies between image and reality, spot instances of media manipulation, and develop their abilities for counteracting such misconceptions and the security threats resulting from them”. Based on “perception-specific threat analysis, the MIRROR project will combine methods of automated text, multimedia and social network analysis for various types of media (including social media) with empirical studies” to develop “technology and actionable insights, [...]



thoroughly validated with border agencies and policy makers, e.g. via pilots”.

Other projects that are already closed, but that get a mention, include Trusted Biometrics under Spoofing Attacks (TABULA RASA), which started in November 2010 and finished in April 2014. It analyzed “the weaknesses of biometric identification process software in scope of its vulnerability to spoofing, diminishing efficiency of biometric devices”. Another project, Bodega, which started in June 2015 and finished in October 2018, looked into how to use “human factor expertise” when it comes to the “introduction of smarter border control systems like automated gates and self-service systems based on biometrics”.

***THE EU  
CONTRIBUTES  
8,199,387,75 EUROS  
TO THE PROJECT TO  
DEVELOP “IMPROVED  
METHODS  
FOR BORDER  
SURVEILLANCE” TO  
COUNTER IRREGULAR  
MIGRATION.***

## References:

Access Now (2019): Comments on the draft recommendation of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems <https://www.accessnow.org/cms/assets/uploads/2019/10/Submission-on-CoE-recommendation-on-the-human-rights-impacts-of-algorithmic-systems-21.pdf>

AlgorithmWatch (2020): Our response to the European Commission’s consultation on AI <https://algorithmwatch.org/en/response-european-commission-ai-consultation/>

Campbell, Zach/Jones, Chris (2020): Leaked Reports Show EU Police Are Planning a Pan-European Network of Facial Recognition Databases <https://theintercept.com/2020/02/21/eu-facial-recognition-database/>

CNIL (2019): French privacy regulator finds facial recognition gates in schools illegal <https://www.biometricupdate.com/201910/french-privacy-regulator-finds-facial-recognition-gates-in-schools-illegal>

Coeckelbergh, Mark/Metzinger, Thomas (2020): Europe needs more guts when it comes to AI ethics <https://background.tagesspiegel.de/digitalisierung/europe-needs-more-guts-when-it-comes-to-ai-ethics>

Committee of Ministers (2020): Recommendation CM/Rec (2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems [https://search.coe.int/cm/pages/result\\_details.aspx?objectid=09000016809e1154](https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154)

Commissioner for Human Rights (2020): Unboxing artificial intelligence: 10 steps to protect human rights <https://www.coe.int/en/web/commissioner/-/unboxing-artificial-intelligence-10-steps-to-protect-human-rights>

Committee on Legal Affairs (2020): Draft Report: With recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies [https://www.europarl.europa.eu/doceo/document/JURI-PR-650508\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-PR-650508_EN.pdf)

Committee on Legal Affairs (2020): Artificial Intelligence and Civil Liability [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL\\_STU\(2020\)621926\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf)

Committee on Legal Affairs (2020): Draft Report: On intellectual property rights for the development of artificial intelligence technologies [https://www.europarl.europa.eu/doceo/document/JURI-PR-650527\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-PR-650527_EN.pdf)

Committee on Civil Liberties, Justice and Home Affairs (2020): Draft Report: On artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters [https://www.europarl.europa.eu/doceo/document/LIBE-PR-652625\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/LIBE-PR-652625_EN.pdf)

Delcker, Janosch(2020): Decoded: Drawing the battle lines – Ghost work – Parliament’s moment [https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-drawing-the-battle-lines-ghost-work-parliaments-moment/?utm\\_source=POLITICO.EU&utm\\_campaign=5a7d137f82-EMAIL-CAMPAIGN\\_2020\\_09\\_09\\_08\\_59&utm\\_medium=email&utm\\_term=0\\_10959edeb5-5a7d137f82-190607820](https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-drawing-the-battle-lines-ghost-work-parliaments-moment/?utm_source=POLITICO.EU&utm_campaign=5a7d137f82-EMAIL-CAMPAIGN_2020_09_09_08_59&utm_medium=email&utm_term=0_10959edeb5-5a7d137f82-190607820)

Data Protection Commission(2020): Law enforcement directive <https://www.dataprotection.ie/en/organisations/law-enforcement-directive>

EDRI (2019): FRA and EDPS: Terrorist Content Regulation requires improvement for fundamental rights <https://edri.org/our-work/fra-edps-terrorist-content-regulation-fundamental-rights-terreg/>

GDPR (Art 22): Automated individual decision-making, including profiling <https://gdpr-info.eu/art-22-gdpr/>

European Commission (2018): White paper: On Artificial Intelligence – A European approach to excellence and trust [https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020\\_en.pdf](https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf)

European Commission (2020): A European data strategy [https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy\\_en](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en)

European Commission (2020): Shaping Europe’s digital future – Questions and Answers [https://ec.europa.eu/commission/presscorner/detail/en/qanda\\_20\\_264](https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_264)

European Commission (2020): White Paper on Artificial Intelligence: Public consultation towards a European approach for excellence and trust <https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence>

European Commission (2018): Security Union: A European Travel Information and Authorisation System – Questions & Answers [https://ec.europa.eu/commission/presscorner/detail/en/MEMO\\_18\\_4362](https://ec.europa.eu/commission/presscorner/detail/en/MEMO_18_4362)

European Data Protection Board (2019): Facial recognition in school renders Sweden’s first GDPR fine [https://edpb.europa.eu/news/national-news/2019/facial-recognition-school-renders-swedens-first-gdpr-fine\\_en](https://edpb.europa.eu/news/national-news/2019/facial-recognition-school-renders-swedens-first-gdpr-fine_en)

European Parliament (2020): Artificial intelligence: EU must ensure a fair and safe use for consumers <https://www.europarl.europa.eu/news/en/press-room/20200120IPR70622/artificial-intelligence-eu-must-ensure-a-fair-and-safe-use-for-consumers>

European Parliament (2020): On automated decision-making processes: ensuring consumer protection and free movement of goods and services [https://www.europarl.europa.eu/doceo/document/B-9-2020-0094\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/B-9-2020-0094_EN.pdf)

European Data Protection Supervisor (2019): Facial recognition: A solution in search of a problem? [https://edps.europa.eu/press-publications/press-news/blog/facial-recognition-solution-search-problem\\_de](https://edps.europa.eu/press-publications/press-news/blog/facial-recognition-solution-search-problem_de)

ETIAS (2020): European Travel Information and Authorisation System (ETIAS) [https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/smart-borders/etias\\_en](https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/smart-borders/etias_en)

ETIAS (2019): European Travel Information and Authorisation System (ETIAS) <https://www.eulisa.europa.eu/Publications/Information%20Material/Leaflet%20ETIAS.pdf>

Horizon2020 (2019): robust Risk based Screening and alert System for PASSengers and luggage <https://cordis.europa.eu/project/id/787120/reporting>

High Court of Justice (2019): *Bridges v. the South Wales Police* <https://www.judiciary.uk/wp-content/uploads/2019/09/bridges-swp-judgment-Final03-09-19-1.pdf>

High-Level Expert Group on Artificial Intelligence (2020): Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>

Hunton Andrew Kurth (2020): UK Court of Appeal Finds Automated Facial Recognition Technology Unlawful in *Bridges v South Wales Police* <https://www.huntonprivacyblog.com/2020/08/12/uk-court-of-appeal-finds-automated-facial-recognition-technology-unlawful-in-bridges-v-south-wales-police/>

Kayalki, Laura (2019): French privacy watchdog says facial recognition trial in high schools is illegal <https://www.politico.eu/article/french-privacy-watchdog-says-facial-recognition-trial-in-high-schools-is-illegal-privacy/>

Kayser-Bril, Nicolas (2020): EU Commission publishes white paper on AI regulation 20 days before schedule, forgets regulation <https://algorithmwatch.org/en/story/ai-white-paper/>

Leyen, Ursula von der (2019): A Union that strives for more – My agenda for Europe [https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission\\_en.pdf](https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf)

Leyen, Ursula von der (2020): Paving the road to a technologically sovereign Europe <https://delano.lu/d/detail/news/paving-road-technologically-sovereign-europe/209497>

Leyen, Ursula von der (2020): Shaping Europe's digital future [https://twitter.com/eu\\_commission/status/1230216379002970112?s=11](https://twitter.com/eu_commission/status/1230216379002970112?s=11)

Leyen, Ursula von der (2019): Opening Statement in the European Parliament Plenary Session by Ursula von der Leyen, Candidate for President of the European Commission [https://ec.europa.eu/commission/presscorner/detail/en/SPEECH\\_19\\_4230](https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_19_4230)

Ann-Charlotte Nygård, (2019): The New Information Architecture as a Driver for Efficiency and Effectiveness in Internal Security <https://www.eulisa.europa.eu/Publications/Reports/eu-LISA%20Annual%20Conference%20Report%202019.pdf>

Sabbagh, Dan (2020): South Wales police lose landmark facial recognition case <https://www.theguardian.com/technology/2020/aug/11/south-wales-police-lose-landmark-facial-recognition-case>

South Wales Police (2020): Automated Facial Recognition <https://afr.south-wales.police.uk/>

Valero, Jorge (2020): Vestager: Facial recognition tech breaches EU data protection rules <https://www.euractiv.com/section/digital/news/vestager-facial-recognition-tech-breaches-eu-data-protection-rules/>



# BELGIUM STORY

PAGE 34

# RESEARCH

PAGE 38

DENMARK

ESTONIA

FINLAND

FRANCE

GERMANY

GREECE

ITALY

NETHERLANDS

POLAND

PORTUGAL

SLOVENIA

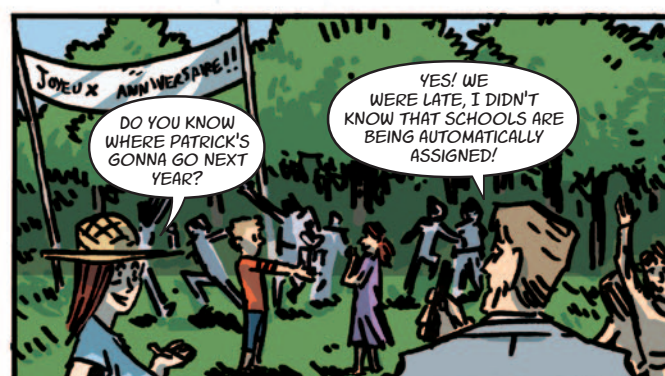
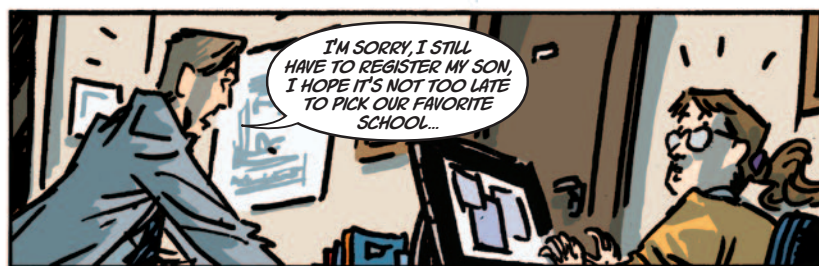
SPAIN

SWEDEN

SWITZERLAND

UNITED KINGD





Find out more in the story on the next page.



# In Flanders, an algorithm attempts to make **school choice** fairer

**In Belgium, some schools don't have enough capacity for all students who want to go there. In the Flemish part of the country, the government introduced an algorithm to assign places in schools, in the hope of giving every student the same chances.**

By Koen Vervloesem



Belgium has a long tradition of free choice of school, but in recent years this has met its limits. Some schools are more popular than others. Because class sizes are limited, this is a problem if too many parents want to enroll their children in the same school. As a result, in many neighborhoods, parents were spending several nights, sometimes up to a week, in tents outside their preferred school. This way, they could be the first to enroll their children, because the rule was “First come, first served”. In bigger cities such as Brussels, Leuven, Ghent, and Antwerp, you could see such a long line of parents camping in tents.

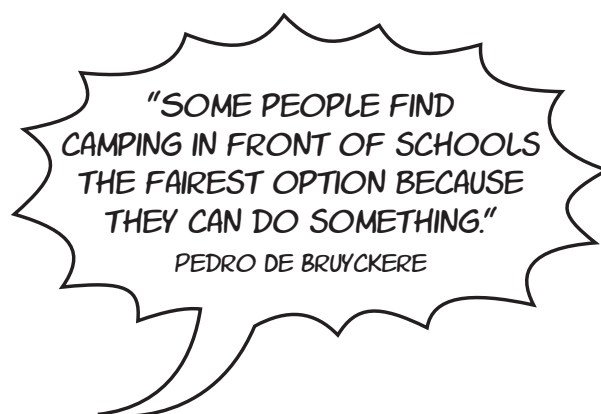
This doesn't seem to be a fair situation: not every parent can take a week off from work to get their children in their preferred school. This forces parents who don't have this opportunity to enroll their children in schools that they like less, such as a school that is farther from home or a school that has a lesser reputation.

Another solution would be to assign each student to a school using an algorithm. But it is not obvious to everybody whether that would be fairer, says Pedro De Bruyckere, an educator at Artevelde University College, Ghent and a postdoctoral researcher at Leiden University: “Some people find camping in front of schools the fairest option because they can do something, while you don't have any impact on an algorithm or a lottery. Others prefer an algorithm or a lottery for the same reason: no one can have impact on the system.”

## / A central online system

In 2018, in response to this situation, the Flemish government implemented a mandatory central online system for regions or cities that have a capacity problem in their schools. This system uses an algorithm that decides in which school a child can enroll. After the decision, a child gets a ‘ticket’ to the school, and she or he can then enroll in the school by using this ticket. This practice should eliminate camping lines: if a school does not use the central online system, it cannot refuse a student; if it uses the system, it has to follow the algorithm.

But how would one algorithmically assign students to a school in a fair way? According to Bart Mesuere, a postdoctoral researcher on algorithms and data visualization at Ghent University, one evident prerequisite for a school enrollment algorithm is that it should assign as many children as possible to a school of their highest possible preference. On the other hand, the system should disappoint as



few children as possible, meaning that the number of children that cannot enroll in any school in their list of choices should be minimized.

Mr. Mesuere adds another requirement: children (or their parents) shouldn't be able to ‘game’ the system with a clever strategy. “The algorithm should be maximally strategy-neutral. For instance, if the algorithm was programmed to avoid solutions where a student doesn't get any of her preferred schools, someone who puts just one school on his list would benefit. That's not OK.”


## / Family preference

Mr. Mesuere computed some numbers from the city of Ghent's use of the system. In 2018, their central school registration system had 3,081 registrations for secondary schools. The city encouraged people to choose multiple preferred schools, and 720 pupils selected only one preferred school. Some children were not in these figures: children who have a brother or sister in their school of preference or a parent working in their school of preference automatically get a ticket for the school.

In 2018, 78.1% of the students in Ghent could enroll in their school of preference, 13.1% got assigned the second school on their list, 2.7% the third school of choice, and 0.2% their fourth choice, while 6.0% of the students didn't receive a ticket for any school (they can still apply to schools with free places).

## / Exchanging tickets

This algorithm was strategy-neutral, but it had one big issue: one child could get the first choice of a second child while the latter also received the first choice of the former. The solution seemed simple: both could exchange their ticket and they would be happier then.



This situation actually occurred in Ghent, but the rules didn't allow for exchanging tickets. This was a sensible rule: the whole point of the central registration system was to make the system fairer, but if students could exchange tickets this would give a strong advantage to families with a big social network.

However, the fact that these exchanges were possible and thus could improve the number of students assigned to their school of preference proved that the system could be improved without any disadvantages: the algorithm could automatically find these situations and randomly pick two children with the other's preferred school and exchange the tickets before the results would be published. In 2019, these automatic exchanges were added to the algorithm, and the results were indeed better.

## **/ Disadvantaged students**

There is another issue related to school choice: social segregation. In some neighborhoods, there are schools with more disadvantaged students and schools with more students of a privileged social background.

The Flemish government decided to lessen the social segregation by making classes more heterogeneous. The government originally devised a system of "double quota": schools should delineate free places in two groups, one for disadvantaged students and one for privileged students. This social correction should lead to a more diverse social mix.

The delineation is quite simple: students are considered disadvantaged if their mother has not completed secondary education or if the student receives a bursary. Each school must have a given percentage of disadvantaged children, which they set in collaboration with other schools from a district. The school's available places are assigned independently to disadvantaged and non-disadvantaged children. Only when all children of a group are assigned a place, and if the school still has free places reserved for that group, are students of the other group allowed to take these places.

The decision was controversial. Some privileged parents felt discriminated against, while others found that freedom of choice of the school was more important than a better social mix. This controversy resulted in a political compromise. In an agreement on 7 September 2018, the Flemish government decided that elementary schools would keep the double quota, but secondary schools would not have to.

## **/ A standard algorithm**

The algorithm goes roughly like this. First, for each school in their list of preferred schools, students receive a random rank. After this, each school has an ordered list of students, numbered from low (they get the place) to high (they only get a place in the school if there are enough places). The school gives as many tickets to students as there are available places.

Now, if a student gets a ticket for more than one school, she keeps only the ticket for her first school of preference. She will lose her tickets for the other schools, which means that these schools now have an extra free ticket that they assign to the next student on their waiting list.

***THE FLEMISH  
GOVERNMENT  
DECIDED TO  
LESSEN THE  
SOCIAL  
SEGREGATION  
BY MAKING  
CLASSES MORE  
HETERO-  
GENEOUS.***

When there are no more students with multiple tickets, there is another step: optimization. This is done by exchanging tickets between students that have each other's better choice. A similar optimization happens on the waiting list, to prevent students from getting on the waiting list of a school of each other's higher preference.

Only then are the tickets communicated to the parents, together with the place on the waiting list of their higher choices. With these tickets, they can enroll in their school.

According to Mr. Mesuere, the algorithm that Ghent and Antwerp are using seems to work fine, but there are always other solutions possible. "However, the alternatives aren't perfect either. It's just a matter of weighing pros and cons, and this is a political choice. For some experts, strategy-neutrality is holy, while others prefer the best possible result for each student."

In 2019, the decree that obligated the central online system for schools was revoked and the Flemish government is working on a new version. Mr. De Bruyckere points out that the current situation could lead to problems: "Because there's no legal requirement anymore, some schools will keep using the system, while others won't."

## / School commute

For secondary schools, one important factor that the algorithm doesn't take into account when assigning a school to a student is the distance from home to school. This is an issue in bigger cities that attract students from far away. Leuven is such a city, and its mayor, Mohamed Ridouani, pointed out this problem when, in 2018, he announced that the city would start with the central system in the 2019-2020 school year.

"If the capacity of our schools can't be improved, I'm in favor of including the distance between home and the school as a criterion in the algorithm," Mr. Ridouani said. This means that students from Leuven and the neighboring municipalities would take precedence over students who commute to Leuven from much farther away. "I just can't explain to a mom from Heverlee [a part of Leuven] that her child cannot go to her first school of preference, while a student from Vilvoorde [30 km away] can go to school here, although that student from Vilvoorde could also go to school closer to his home." With the current system, it is indeed possible that some students from Leuven cannot get a place in a school in their own city, while students from farther away can.

**"THE ALTERNATIVES  
AREN'T PERFECT EITHER.  
IT'S JUST A MATTER  
OF WEIGHING PROS  
AND CONS, AND THIS  
IS A POLITICAL CHOICE.  
FOR SOME EXPERTS,  
STRATEGY-NEUTRALITY  
IS HOLY, WHILE OTHERS  
PREFER THE BEST  
POSSIBLE RESULT FOR  
EACH STUDENT."**

BART MESUERE

Pedro De Bruyckere adds that the problem is even more severe for schools that provide special secondary education (for students with learning or physical disabilities), which are even more distributed geographically: "Distance should be an important factor for these schools."

## / Pursuing dreams

Another problem is that a preference is not always 'just' a preference, especially in secondary schools. Mr. De Bruyckere gives an example: "In Ghent, there's only one school where you can study sports, more specifically football. If a talented student who wants to become a top footballer doesn't get the first choice, he or she can't pursue this dream."

Bart Mesuere adds that you have to draw a line somewhere, and this is a political choice too: "Most people will understand the importance of being able to choose a sports school, but what about other schools that are the only ones with a specific field of study? What about the only school in your neighborhood with maths/sciences/informatics if your dream is to become a scientist? Or what about the only school in town with a specific learning method?"



# Research

By Rosamunde van Brakel

## Contextualization

As a result of the different governments, and the different levels of government, in Belgium (Federal and Regional), several different strategies dealing with digitization emerged in 2018. In Flanders, this strategy is called *Vlaanderen Radicaal Digitaal*, while in the Walloon region, it is known as *Plan Numérique*. In 2015, the federal government launched a strategy called Digital Belgium, but this covered more than just ADM. In 2018, the Flemish government launched an action plan with a budget of €30 million to deal specifically with artificial intelligence. The main targets of this investment are fundamental research, applications for industry, framing policy concerning education, sensitization, and ethics. While in Flanders, the biggest opportunities for AI are in personalized healthcare, smart mobility, and industry 4.0. Concerning regulation, the Belgian national law on data protection, which is based on the GDPR, came into force on 5 September 2018. In the same year, the Belgian Chamber of Representatives adopted a resolution to have a preventative ban on fully automated weapons.

## A catalog of ADM cases

### / Face Recognition at Brussels Airport

In July 2015, Brussels Airport installed 'e-gates' with face recognition technology that checks the identity of EU-citizens from outside the Schengen zone. It electronically checks to see if identity papers are real and if a passenger is flagged. In addition, the face recognition software compares the picture on the passport chip in the identity card with the digital picture the software renders at the e-gate (De Morgen, 2019). In early 2020, the airport scrapped the e-gates, which cost €2.4 million, because they were constantly defective, according to the police (The Bulletin, 2020).

In July 2019, the Belgian news magazine KNACK published an interview with the Commissioner-General of the Federal Police in which he proposed the implementation of face recognition software at Brussels airport (KNACK, 2019). However, the Belgian Oversight Body for Police Information (COC) was never contacted for advice about this, and a Data Protection Impact Assessment had not been conducted. As a result, the COC started an investigation and found the following. In early 2017, the Federal Police began piloting a new face recognition system with four cameras. The system functioned in two phases. When the software that looks at video images is activated, so-called snapshots are made. The system constantly individualizes the video images through a biometric template of people whereby the snapshots are created. These snapshots are stored and linked to the blacklists of people who are suspected of a crime. When there is a positive link, a hit is created. During the test phase, the error rate (number of false positives) was very high; as a result, in March 2017, testing of the system was stopped. However, during a visit by the COC it was found that the system was still partially active, in the sense that snapshots of all passengers in the airport were still being taken, but without any comparison to a blacklist being performed. To be able to keep testing the system, it had to store the images for a minimum amount of time (COC, 2019).

IN EARLY 2020, THE AIRPORT  
SCRAPPED THE E-GATES, WHICH  
COST 2.4 MILLION EUROS,  
BECAUSE THEY WERE CONSTANTLY  
DEFECTIVE, ACCORDING TO  
THE POLICE.

The COC identified three problems with this pilot:

- 1) The police did not conduct an evaluation of the risks related to privacy and did not inform the COC about the pilot project as they are obliged to do.
- 2) A database containing data associated with a hundred thousand faces of travelers was created and stored temporarily, and that is not allowed.
- 3) The system gave a lot of false alarms (Heymans & Vanrenterghem, 2019).

In their public report, the COC concluded that there was too little information about the implementation and risks of the technology as there was no clear policy or data protection impact assessment conducted to come to a conclusion or offer advice. As a result, they enforced a temporary corrective measure in asking for a temporary ban of the pilot project (COC, 2020).

### **/ Smart video surveillance by the local police: Briefcam**

Since 2019, the local police in the cities of Kortrijk, Kuurne, and Lendelde (VLAS) have been using a 'smart' video surveillance system developed by an American company called Briefcam. According to the police, the goals of the system are as follows:

- 1) To aid investigations. For instance, if a person with a backpack and wearing a blue coat has fled in a certain direction, this person can be found easily by the algorithm. The person is then traced on other cameras to get a more complete picture of the route he or she took and to detect crimes that might have been committed. The system can also help search for lost children and solve bicycle theft more easily.
- 2) The generation of live alerts. For instance, at the beginning and end of the school day, trucks are not permitted to drive near schools. If a truck is spotted driving through an area it is not allowed to, the system sends an alert to the police.
- 3) Collect statistical information to support policy.

The system stores all 'objects' (for example: people, small and large vehicles, and animals) that appear in the video images. An algorithm then decides what category each object belongs to, and, after this first categorization, a sub-categorization is made. For example, when it comes to people, the system categorizes each person as either a man, woman, or child before further categorizing by clothing: short/long

sleeves, short/long trousers, color of clothing, and things such as hats, handbags, backpacks, etc. The same system of categorization is used for vehicles. For example, if the vehicle has two wheels: Is it a bicycle or a motorcycle? For different vehicles: Is it a car, pick-up, van, truck, or a bus? In addition, the direction, size, and velocity of the object are also registered. The system is also able to implement face recognition (PZVlas, 2019), although, at the moment, there is no legal framework to regulate use by the police. However, as the spokesperson of the VLAS police zone indicated in an interview: in exceptional circumstances, when requested by an investigatory judge, it is possible to use face recognition (Verrycken, 2019).

### **/ Face recognition at football matches**

Since early 2019, the RWD Molenbeek football club (RWDM) has been experimenting with face recognition software at football matches. The system allows people with season

***TWO CAMERAS  
AT THE STADIUM  
ENTRANCE DETECT  
THE FACES OF  
PEOPLE COMING  
IN, AND, IN REAL-  
TIME, THE SYSTEM  
VERIFIES IF THE  
PERSON HAS  
A SEASON TICKET  
OR NOT.***

tickets to enter Molenbeek football stadium without having to show their tickets. Two cameras at the stadium entrance detect the faces of people coming in, and, in real-time, the system verifies if the person has a season ticket or not. The aim of the system is to create a “fast lane” whereby season ticket holders, who have given their consent and have uploaded a picture to the database, do not have to queue. According to a representative of the biometric company that delivered the software for the football club, after a one-year pilot, the system is deemed sufficiently reliable, even in difficult circumstances; such as rain, or when people are wearing hats or glasses. (Verrycken, 2019).

## / Algorithmic school registrations

In many Belgian cities, there are problems when registering children for schools, especially when a lot of parents want to register their children at the same school. This has led to parents spending several nights in tents outside schools so that they can be the first to register their children. In response to this situation, which was considered unfair, several cities started organizing school registration via a central online system that uses an algorithm to decide in which school a child can be registered. In the system for primary schools, that is used in Leuven, a city east of Brussels, a number of variables are taken into account. On the basis of the answers given by the parents to a series of questions relating to the education level of the mother and whether or not the student receives a grant, students are then divided into ‘indicator students’ and ‘non-indicator students’. An indicator student is a student who is defined as having fewer life chances and is based upon the level of education of the mother and if the student is eligible for an education allowance. Further categorization happens on the basis of two criteria: distance from home to school and preference of school. The schools can decide how much weight they assign to distance and preference i.e. between 30%, 50%, or 70% for each (Meldjeaan, 2020). This weight can differ between cities, for instance, in Antwerp the schools have to assign a minimum of 50% to the distance criterium.

For secondary schools, some cities also use an online registration system, which uses an algorithm to decide which school the child will go to. However, the system is not the same as for primary schools. As not every municipality has a secondary school, the distance criterium is considered discriminatory and, therefore, it is not taken up (Dierickx and Berlanger, 2019). The system does take into account priority students (i.e., children of the staff of the school) and indicator-students (see above). The algorithm works ran-

domly on the basis of the first choice in three rounds (Aanmelden school, 2020). As the system does not take distance into account, this has led to situations where students from Brussels took up places in Flemish secondary schools outside of Brussels, and children, living close to the school not getting in and having to be registered in schools much farther away (Dierickx and Berlanger, 2019).

## / Predictive ADM systems for healthcare and upbringing

A Brussels start-up called Kantify and the AI Lab at the Université Libre Bruxelles (Iridia) has—for the first time—developed a model for predicting atrial fibrillation, a heart rhythm disorder, which causes heart attacks (Het Nieuwsblad, 2019).

According to an interview in the Flemish newspaper De Standaard, Lauren Van Parys, a Flemish member of Parliament, said a newly formed government agency of Flanders called the Agency for Upbringing (Agentschap Opgroeien), will use AI to better predict whether problems will emerge with a particular child, based upon existing information (De Standaard, 2019). However, when asked about this further, the Agency for Upbringing indicated that there are no concrete plans yet, but that they are preparing a policy note about it.

## / New developments in predictive policing and algorithmic work activation

### / Predictive policing

In the 2019 AlgorithmWatch report, it was suggested that the Belgian police were planning to roll out predictive policing nationally based on the iPolice system, which will be operational by 2020. The iPolice system aims to centralize all police data in the cloud. It is a cooperation between the Ministry of Home Affairs, the Digital Agenda, and the Ministry of Justice. However, up until now, there has been no news about iPolice. Local police are increasingly experimenting with predictive policing applications, for instance, in the police zone of Zennevallei (Beersel, Halle, and Sint-Pieters-Leeuw), the police are very active in exploring how to implement predictive policing in response to a research proposal by the University of Ghent. A PhD student at the university will research the police zone for two years. The student will try to collect as much data as possible, including times and locations of crimes, but also weather condi-

tions, and demographic characteristics, among other factors, according to the spokesperson of the police zone (Van Liefveringe, 2019).

## / Algorithmic work activation

In an interview, the head of the public employment service of Flanders, VDAB, argued that artificial intelligence can predict which citizens need more support and that this will be a positive outcome for vulnerable groups (De Cort, 2019). He indicated that they could do even more if they could make use of more data from partner organizations. However, questions have been raised in the Flemish Parliament about what the VDAB is doing as the algorithms they use are not transparent, and there is a worry that the algorithms will work in a discriminatory way (El Kaouakibi, 2020).

The Innovation Lab of the Flemish public employment service, VDAB, is currently developing and experimenting with two projects. Firstly, a prediction model that predicts—based on an individual file and click data—the chance that a job seeker will still be unemployed after six months, and what the reasons for that are. And secondly, a prediction model—also based on individual file and click data—that predicts which jobseekers will look less actively for work in the coming months. The hope is that the system will help quickly activate these people and avoid more drastic measures such as referral to the control service of the VDAB and possible sanctions. At the moment, two different models have finished the proof-of-concept phase. One was developed by Accenture and the other by Deloitte (VDAB, 2019).

# Policy, oversight and debate

## / Flanders

The Flemish Minister of Innovation launched the Flemish Action Plan of Artificial Intelligence (AI) in March 2019. The plan includes an annual investment of €32 million and is focused on research, implementation, ethics, and training. €12 million will go specifically to research (Vlaamse Regering, 2019a), and €5 million will go to ethics and training. Part of the plan is to establish a knowledge center that will reflect upon the ethical challenges associated with AI.

This knowledge center—which will focus on juridical, ethical, and societal aspects of artificial intelligence, together with data-driven applications, launched on 9 December 2019 in Brussels (Kenniscentrum Data & Maatschappij, 2019). The center is not a new organization, but a cooperation between three academic research groups: Centre for IT and IP Law (CITIP), Katholieke Universiteit Leuven, Studies in Media, Innovation and Technology (SMIT), Vrije Universiteit Brussel/imec and Research Group for Media, Innovation, and Communication Studies (MICT), Universiteit van Gent/imec.

The main aims of the center are a widespread co-design approach to delineate AI policy, to stimulate societal debate about the acceptance of new technology, deliver thought leadership about both societal and economically acceptable developmental trajectories for AI, identify and valorize important human factors in the development of AI and finally develop legal frameworks and guidelines for policy-makers and companies.

Activities of the center include:

- The translation of general ethical principles into concrete guidelines, recommendations and rules.
- Removing risk from AI-development by companies and innovators.
- Stimulate current and future regulatory and policy reforms with an impact on AI (for instance, directive product liability, Machinery directive reform, e-Privacy directive, etc.).
- To gain clarity about the implications of current and new regulatory instruments (such as the GDPR, regulation on free flow of non-personal data, etc.).
- To create space for experiments relating to “ethical validation” and “regulatory sandboxing”.
- Increase societal awareness and create more support. (Departement Economie, Wetenschap & Innovatie, 2019).

The main driver for investment in AI in Flanders is economic. The “Quarter nota to the Flemish Government about the Flemish AI Plan” (Vlaamse Regering, 2019b), proposed by the Minister, states that—considering the technological and international developments with regards to AI—“if Flanders wants to remain competitive investments in AI are necessary” (Vlaamse Regering, 2019b: 4). The nota goes on to mention social and ethical consequences, but in the same sentence, it questions the validity of these concerns as these discussions block thinking relating to the oppor-

tunities of AI. It also does not mention much about these potential consequences, suggesting that all will be solved, if these consequences are included in the design phase. In addition, the opportunities and uses of AI are generalized in a way that they are considered the same across all sectors.

## / Federal Government

Under the direction of the former deputy prime ministers, Philip De Backer and Alexander De Croo, the Federal government launched the “AI 4 Belgium strategy” on 18 March 2019, with the aim to position Belgium as a central player in the European AI landscape (AI4 Belgium, 2019). The strategy is the result of consultation with 40 experts from the technology sector in Belgium. According to vice-premier Alexander De Croo: “through lack of knowledge about AI it is possible that Belgium will miss chances for economic development and prosperity. Agoria estimates that digitization and AI can create 860.000 jobs by 2030. Instead of lagging behind, we want to switch a gear higher. We have the ambition to develop the right knowledge and skills, so it is possible to reap the societal and economical fruits of AI.” (Van Loon, 2019).

As a result of the implementation of the GDPR in May 2018, the Belgian Privacy Commission was reformed and became the Belgian Data Protection Authority. In addition, the oversight body which looks at how the police use information (*Controleorgaan op politionele informatie, COC*) was reformed to function as an independent data protection body. This

body will be specifically responsible for overseeing how the police use information.

In answer to a question in the Belgian Chamber of Representatives on 25 October 2019, regarding the use of face recognition systems by the police, the Ministry of Security and Home Affairs indicated that the legal basis for using face recognition is currently being investigated (De Kamer, 2019).

## / Walloon government

In 2019, Wallonia launched its 2019-2024 digital strategy. According to the website of [digitalwallonia.be](https://digitalwallonia.be), the strategy has five major themes:

- 1) Digital sector: A strong digital sector with cutting-edge research, able to capture and maintain the added value of the digital transformation on our territory.
- 2) Digital business: Boosting the Walloon economy depends on a strong and rapid increase in the digital maturity of our companies.
- 3) Skills and education: Each Walloon citizen must become an actor in the digital transformation process by acquiring strong technological skills and adopting an entrepreneurial attitude.
- 4) Public services: A new generation of open public services, acting transparently and being, by themselves, an example of digital transformation.

***IN GENERAL, VERY LITTLE ATTENTION  
IN POLICY DISCOURSE FOCUSES ON  
THE ETHICAL QUESTIONS OR THE  
SOCIETAL IMPACT OF AI.***



- 5) Digital Territory: Smart and connected to broadband networks. Our territory must offer unlimited access to digital innovation and act as a driver for industrial and economic development. (Digital Wallonia, 2019)

Wallonia invested €900,000 in a one-year action plan Digital Wallonia4.ai between July 2019 and July 2020 supported by l'Agence du Numérique, Agoria, l'Infopôle Cluster TIC, and le Réseau IA. The goal was to accelerate the adoption of AI in enterprise and demystify the technology for the general public. Wallonia, which is lagging behind Flanders, wants to develop its potential and see the birth of AI champions (Samain, 2019). The strategy was centered around four axes: 1) AI and society, 2) AI and enterprises, 3) AI and training and 4) AI and partnerships (Digital Wallonia, 2020). There were two main actions of this program. Firstly, the creation of Start AI that aimed to help companies learn about AI through a three-day coaching session run by one of the members of the Digital Wallonia AI expert pool. Secondly, the creation of Tremplin IA to help with Proof of Concepts (PoCs) of AI in the Walloon Region. In all, 19 projects for individual PoCs and nine collective projects joined the program. Of these, the jury selected five individual PoC projects and four group projects. Agoria set up an online course, while Numeria created AI training courses. Meanwhile, a study on the social and ethical impact of AI will be conducted by the Research Centre in Information Law and Society (CRIDS) of the University of Namur (Digital Wallonia, 2020).

IN BELGIUM,  
BELIEF IN AI  
PROGRESS,  
TOGETHER  
WITH A FEAR OF  
MISSING OUT, AND  
FALLING BEHIND  
IN COMPARISON  
TO OTHER  
COUNTRIES, ARE  
STRONG DRIVERS.

consequences of AI. And, even there, the emphasis has been on legal compliance. In general, very little attention in policy discourse focuses on the ethical questions or the societal impact of AI. The only place where attention is given to the risks of AI is in an advisory report compiled by the independent Advice Commission of the Flemish Government for Innovation and Entrepreneurship (VARIO, 2019). However, the overview is limited, and there is a lack of knowledge regarding the research that is being done in Belgium on this topic, and of academic research in this area in general. No concrete solutions have been proposed to deal with the ethical questions or the societal impact of AI or in making the development and implementation of AI more transparent.

Furthermore, it seems that the advice about the dangers of AI by this agency are not taken that seriously when it comes to Flemish policy. Only a very small amount of the foreseen budget will be invested in the knowledge center for data and society (mentioned previously). In addition, if one looks at the goals and activities of the center, this does not include supporting and financing research on social and ethical issues or value-sensitive design (design that takes into account certain ethical values), for instance. As much of the development of these technologies is still in the design phase, there are opportunities for the new Knowledge Center to advise on data protection, ethical issues, and social impact, which could make a difference.

At the level of the Federal government, nothing has been found in the AI 4 Belgium strategy about the legal, ethical, and social consequences of AI. It is presented as if these issues do not exist. Nothing was found regarding the establishment of oversight bodies or ethical commissions that would oversee the technological developments in AI. There seems to be a lot of trust in the researchers and companies developing new AI technologies. Likewise, in the Digital Wallonia strategy, nothing can be found that addresses legal, ethical, and social concerns.

To summarize, there is a big challenge for policymakers when it comes to ADM and AI in Belgium. They need to take legal, ethical, and social concerns seriously and to ensure these systems are transparent, and implemented with the necessary democratic oversight, have a positive impact on Belgian society, and empower Belgian citizens.

## Key takeaways

In 2018, most stories about AI were still very theoretical. Whereas, in 2019, it is clear that there has been an increase in the use of ADM. The main drivers for investment in AI in Belgium are predominantly economic. Belief in AI progress, together with a fear of missing out, and falling behind in comparison to other countries, are strong drivers.

The biggest investments have been in the development of new AI technologies, whereas, in comparison, very little has been invested in research into the ethical, legal, and social

# References

Aanmelden School (2020) Toewijzing, <https://www.aanmelden.school/toewijzing.html#>.

AI 4 Belgium (2019), Website AI 4 Belgium, <https://www.ai4belgium.be/nl/>.

COC (2020) Executive summary Tussentijds rapport met corrigerende maatregel betreffende de visitatie bij de Federale Politie van de Luchthaven Zaventem door het Controleorgaan op de Politie informatie met betrekking tot het gebruik van gezichtsherkenning op de nationale luchthaven van Zaventem, [https://www.controleorgaan.be/files/DIO19005\\_Onderzoek\\_LPABRUNAT\\_Gezichtsherkenning\\_Publiek\\_N.PDF](https://www.controleorgaan.be/files/DIO19005_Onderzoek_LPABRUNAT_Gezichtsherkenning_Publiek_N.PDF).

De Cort, G. (2019) Kwetsbare werklozen bereiken met AI, 10 May 2019, *De Standaard*, [https://www.standaard.be/cnt/dmf20190509\\_04388637](https://www.standaard.be/cnt/dmf20190509_04388637).

De Morgen (2019) Topman federale politie: 'We gaan camera's met gezichtsherkenning inzetten in Zaventem', *De Morgen*, 9 July 2019, <https://www.demorgen.be/nieuws/topman-federale-politie-we-gaan-camera-s-met-gezichtsherkenning-inzetten-in-zaventem~b1ff4514/>.

De Kamer (2019) Schriftelijk Vraag Tim Vandeput, Wettelijk kader voor AI, <https://www.dekamer.be/kvvcr/showpage.cfm?section=qrva&language=nl&cfm=qrvaXml.cfm?legislat=55&dossierID=55-B004-1120-0223-2019202000677.xml>.

De Standaard (2019) Aanranding, diefstal en drugs: 144 kleuters plegen misdrijf, 28 November 2019, [https://www.standaard.be/cnt/dmf20191128\\_04742377](https://www.standaard.be/cnt/dmf20191128_04742377).

Departement Economie, Wetenschap & Innovatie (2019) Artificiële Intelligentie: Oprichting Kenniscentrum Data & Maatschappij in kader Vlaams Beleidsplan AI, [https://ewi-vlaanderen.be/nieuws/artificiele-intelligentie-oprichting-kenniscentrum-data-maatschappij-data-society?utm\\_medium=email&utm\\_campaign=Nieuwsbrief%20EWI%20%2028%20mei%202019&utm\\_content=Nieuwsbrief%20EWI%20%2028%20mei%202019+CID\\_7e6b04f41738782bd12cd7efbfc0efb4&utm\\_source=Email%20marketing%20software&utm\\_term=Artificile%20intelligentie%20oprichting%20kenniscentrum](https://ewi-vlaanderen.be/nieuws/artificiele-intelligentie-oprichting-kenniscentrum-data-maatschappij-data-society?utm_medium=email&utm_campaign=Nieuwsbrief%20EWI%20%2028%20mei%202019&utm_content=Nieuwsbrief%20EWI%20%2028%20mei%202019+CID_7e6b04f41738782bd12cd7efbfc0efb4&utm_source=Email%20marketing%20software&utm_term=Artificile%20intelligentie%20oprichting%20kenniscentrum).

Dierickx, R. and D. Berlinger (2019) Tientallen ouders ongerust: kinderen vinden geen school in eigen gemeente, *Het Laatste Nieuws*, 4 april 2019, <https://www.hln.be/in-de-buurt/kapelle-op-den-bos/tientallen-ouders-ongerust-kinderen-vinden-geen-school-in-eigen-gemeente~a394ab24/>.

Digital Wallonia (2019) Digital Strategy 2019-2024, <https://www.digitalwallonia.be/en/digital-strategy>.

Digital Wallonia (2020) DigitalWallonia4.ai: bilan et actions, <https://www.digitalwallonia.be/fr/publications/digitalwallonia4-ai-home>.

El Kaouakibi, S. (2020) Parliamentary question about use of artificial intelligence by VDAB, 29 January 2020, <https://www.vlaamsparlement.be/plenaire-vergaderingen/1357705/verslag/1364381>.

Het Nieuwsblad (2019) Belgische artificiële intelligentie kan hartaanvallen voorspellen: "Een wereldprimeur", *Het Nieuwsblad*, 30 October 2019, [https://www.nieuwsblad.be/cnt/dmf20191030\\_04692793](https://www.nieuwsblad.be/cnt/dmf20191030_04692793).

Heymans, P. & A. Vanrenterghem (2019) Politie mag geen automatische gezichtsherkenning meer gebruiken op de luchthaven, VRT NWS 20 September 2019, <https://www.vrt.be/vrtnws/nl/2019/09/20/politie-mag-geen-automatische-gezichtsherkenning-gebruiken-op-de-app/>.

Kenniscentrum Data & Maatschappij (2019) <https://www.data-en-maatschappij.ai/#launch-event-anchor>.

Knack (2019) Federale politie moet project rond gezichtsherkenning op Zaventem stopzetten, 20 September 2019, <https://www.knack.be/nieuws/belgie/federale-politie-moet-project-rond-gezichtsherkenning-op-zaventem-stopzetten/article-news-1511927.html>.

Meldjeaan (2020) FAQ, Onderwijs Leuven, <https://meldjeaan.leuven.be/faq>.

Mesuer, B. (2018) 'Hoe het algoritme achter het aanmeldsysteem voor het onderwijs verbeterd kan worden', KNACK, <https://www.knack.be/nieuws/belgie/hoe-het-algoritme-achter-het-aanmeldsysteem-voor-het-onderwijs-verbeterd-kan-worden/article-opinion-1138733.html>.

PZVlas (2019) Persnota – woensdag 22 mei 2019. Primeur - Veiligheidscamera's PZ Vlas worden eerste online en offline slimme camera's van Vlaanderen, [https://www.pzvlas.be/fileadmin/MEDIA/website/documenten/nieuwsberichten/20190522\\_-\\_Persnota\\_slimme\\_camera\\_s\\_-\\_BriefCam.pdf](https://www.pzvlas.be/fileadmin/MEDIA/website/documenten/nieuwsberichten/20190522_-_Persnota_slimme_camera_s_-_BriefCam.pdf).

Samain, M. (2019) La Wallonie fait (enfin) de l'intelligence artificielle une priorité, *L'Echo*, 27 November 2019, <https://www.lecho.be/economie-politique/belgique/wallonie/la-wallonie-fait-enfin-de-l-intelligence-artificielle-une-priorite/10185999.html>

Van Liefferinge, J. (2019) Deze politiezone wil binnenkort kunnen voorspellen waar inbrekers zullen toeslaan, en daar hebben ze geen 'Madame Soleil' voor nodig, Het Nieuwsblad, 15 November 2019, [https://www.nieuwsblad.be/cnt/dmf20191114\\_04716625](https://www.nieuwsblad.be/cnt/dmf20191114_04716625).

Van Loon, C. (2019) Belgische AI-experts stellen plan voor samen met Belgische overheid, Techzine 18 March 2019, <https://www.techzine.be/nieuws/trends/30817/belgische-ai-experts-stellen-plan-voor-samen-met-belgische-overheid/>.

VARIO (2018) Vlaamse Beleidsagenda Artificiële Intelligentie, November 2018, [https://www.vario.be/sites/default/files/documents/2018\\_NOVEMBER\\_VARIO%20ADVIES\\_NR%205\\_VLAAMSE%20BELEIDSAGENDA%20AI\\_.pdf](https://www.vario.be/sites/default/files/documents/2018_NOVEMBER_VARIO%20ADVIES_NR%205_VLAAMSE%20BELEIDSAGENDA%20AI_.pdf).

VDAB (2019) VDAB Innovation Lab, <https://partners.vdab.be/innovatielab>.

Verrycken, R. (2019) U wordt herkend, De Tijd, 25 May 2019, <https://www.tijd.be/tech-media/technologie/u-wordt-herkend/10130429.html>.

Vlaamse regering (2019a) Besluit van de Vlaamse Regering houdende de herverdeling van begrotingsartikelen van de algemene uitgavenbegroting van de Vlaamse Gemeenschap voor het begrotingsjaar 2019 in het kader van het Vlaams beleidsplan Artificiële Intelligentie, [https://www.ewi-vlaanderen.be/sites/default/files/bsluit\\_van\\_de\\_vlaamse\\_regering\\_-\\_vlaams\\_beleidsplan\\_artificiele\\_intelligentie.pdf](https://www.ewi-vlaanderen.be/sites/default/files/bsluit_van_de_vlaamse_regering_-_vlaams_beleidsplan_artificiele_intelligentie.pdf).

Vlaamse regering (2019b) Quaternota aan de Vlaamse Regering. Betreft: Vlaams Beleidsplan Artificiële Intelligentie, [https://www.ewi-vlaanderen.be/sites/default/files/quaternota\\_aan\\_de\\_vlaamse\\_regering\\_-\\_vlaams\\_beleidsplan\\_artificiele\\_intelligentie.pdf](https://www.ewi-vlaanderen.be/sites/default/files/quaternota_aan_de_vlaamse_regering_-_vlaams_beleidsplan_artificiele_intelligentie.pdf).



BELGIUM  
**DENMARK**  
**STORY**

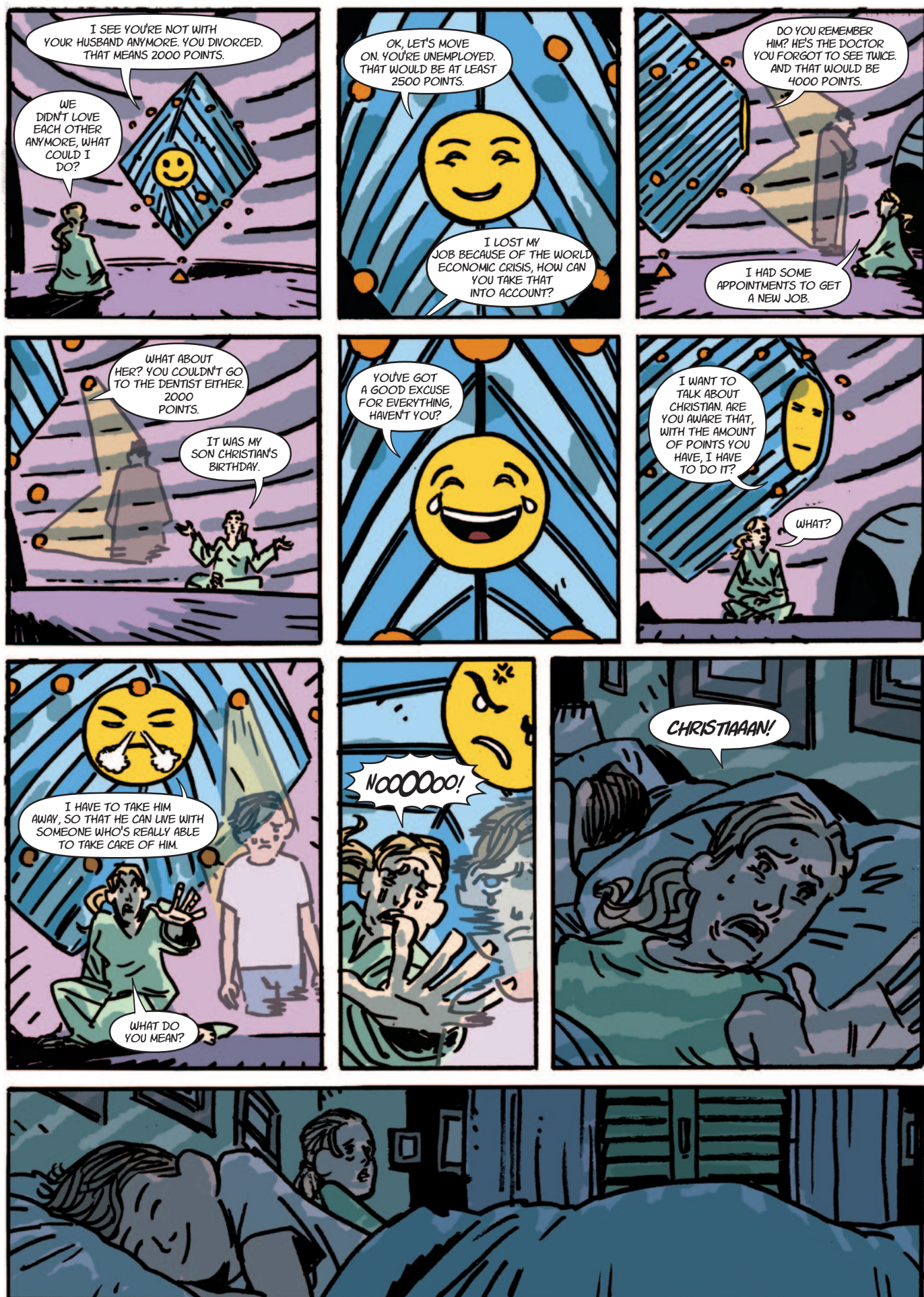
PAGE 48

**RESEARCH**

PAGE 52

ESTONIA  
FINLAND  
FRANCE  
GERMANY  
GREECE  
ITALY  
NETHERLANDS  
POLAND  
PORTUGAL  
SLOVENIA  
SPAIN  
SWEDEN  
SWITZERLAND  
UNITED KINGDOM





Find out more on page 52 in the research chapter under "The Gladsaxe Model - profiling families for early detection of vulnerable children"



# // In a quest to optimize **welfare** **management,** Denmark built a surveillance behemoth

**Udbetaling Danmark was created in 2012 to streamline the payment of welfare benefits. Its fraud control algorithms can access the personal data of millions of citizens, not all of whom receive welfare payments.**

By Nicolas Kayser-Bril

In the mid-2000s, Denmark's government embarked on an ambitious reorganization of local administration. The number of municipalities was cut from 271 to 98. The mergers were expected to create economies of scale and increase efficiency in public services.

Part of the economies of scale took the form of Udbetaling Danmark (Danish for "Payout Denmark", abbreviated UDK). The organization, which is run by the agency for the digitization of the ministry of finance, was set up to centralize the payment of welfare benefits overseen by municipalities: pensions and benefits related to housing, family, disability, and maternity leave. At the time, the government claimed that the move would save 35% in administrative costs (officials from several municipalities disputed this figure).

### **/ Call centers**

UDK started operations in 2012, and service was cut to the bare minimum from the start. In-person meetings became impossible, as all communication with beneficiaries was to be done over the phone. This did not go down well with Danes. In a 2017 survey, one-third of respondents claimed that public service deteriorated since the reforms, with only one in ten seeing an improvement.

The caseworkers who used to work for municipalities were re-hired by UDK when it was created. However, the nature of their work shifted from a focus on beneficiaries to a focus on data. In 2013, the organization started to link data on its two million beneficiaries with external databases from other administrations such as the tax and employment authorities, and the business registers (municipalities have had the same powers since 2011, but it is unclear if they ever made use of them). This data network allowed UDK to automate the checks required before benefits are granted to a requester, such as whether their income level or their wealth were low enough.

This process also allows UDK to perform controls after a benefit has been granted to verify that a beneficiary's situation has not changed. Since 2015, UDK can access data about the housemates or the family of a beneficiary in order to spot irregularities. The data comes from the civil register, the housing register, or the tax authorities.

A 2019 report by the Danish public service broadcaster gave an example of such checks. In it, a member of UDK's "Joint Data Unit" explained that they looked for cases where women were employed by the company of a family mem-

***FOR CASE-  
WORKERS, THE  
NATURE OF  
WORK SHIFTED  
FROM A FOCUS  
ON BENEFICIARIES TO A  
FOCUS ON  
DATA.***

ber just long enough to be eligible for maternity leave. Such a case could be a sign of fictitious employment, they said. (UDK did not disclose how many false-positives this method produced.)

### **/ 91 referrals to the police**

In 2017 and 2018, controls by UDK and municipalities unveiled half a billion Danish crowns in erroneous payment (approximately 70 million euros). It is unclear how much was recovered, what was not recovered due to the beneficiary's incapacity to repay, and what was an accounting trick (the figure includes future payments that were averted because of the control).

UDK does not split the amount between honest errors and actual cheats. In 2017, UDK and municipalities reported a total of 91 individuals to the police for welfare fraud.

In 2018, ostensibly to improve UDK's fraud-detection capabilities, the Danish government attempted to give UDK access to the electricity consumption of welfare beneficiaries and people close to them. The bill was eventually withdrawn after a public outcry.

## / State fraud

It remains unclear if UDK's automated checks focus exclusively on fraud from beneficiaries or if they also verify that beneficiaries receive the amounts they are entitled to. In 2013, for instance, it was found that 300,000 pensioners were cheated of several hundred Danish crowns a month, in a scheme that totaled 700 million crowns (94 million euros) a year. While the government argued the mistake was not deliberate, others called it "legalized social fraud". In 2015, another irregularity, this time directly attributable to UDK, led to 325,000 households receiving lower housing benefits (the error was fixed in subsequent months). UDK had implemented a provision found in a bill before the Danish parliament had made it into law (the bill never passed).

It is unclear if UDK verifies the quality of the data it processes, and whether its programs are audited. In November 2019, UDK provided the tax authorities with erroneous information about 111,000 households. Tax authorities subsequently sent a letter to each taxpayer to inform them of the mistake.

The agency for the digitization of the ministry of finance, which runs UDK, did not answer our requests for comment.

## / Systematic surveillance

In Denmark, the conversation about UDK's powers was revived in July 2019, when Justitia, a think-tank, released a report that detailed its activities, which they called "systematic surveillance". UDK does not provide the actual number of individuals receiving payments, but estimates range between 2.1 and 2.7 million, in a country of 5.8 million inhabitants. Because UDK also collects data on other household members and the immediate family, Justitia considers it likely that UDK processes data on the whole population of the country, sometimes pulling information at daily intervals.

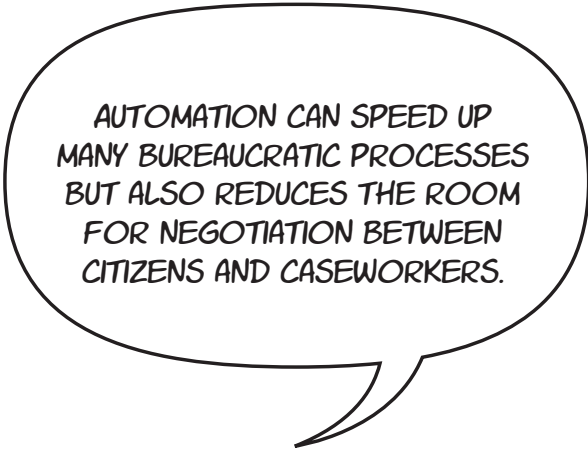
Birgitte Arent Eiriksson, who wrote the Justitia report, is now part of the Danish Data Ethics Council, which has advised the government since 2018. She chairs a working group on data linkage for public authorities. (While UDK is not named in the description of the working group, there is little doubt that the issue under scrutiny is related to UDK's appetite for merging databases). They will provide the government with "a concrete tool that the authorities can use to incorporate data ethics when they want to link personal data and use new technology," to be delivered later in 2020, Ms. Arent Eiriksson told AlgorithmWatch.

***UDK PROCESSES DATA ON THE  
WHOLE POPULATION OF THE  
COUNTRY, SOMETIMES PULLING  
INFORMATION AT DAILY INTERVALS.***

## / The DPA strikes

In early 2019, in a case related to housing benefits, a Danish man found out that UDK's database held a lot of information about him, although he was not a beneficiary. He subsequently filed a complaint to the Danish Data Protection Authority (DPA). The DPA ruled that the blanket collection of data on relatives of beneficiaries infringed the General Data Protection Regulation, which prohibits the collection of unnecessary data. UDK refused to comply, explaining that their mission of fraud control made the data collection necessary and proportionate.

The DPA reopened the case in 2020 and made clear to UDK that they had to delete data on individuals who were not beneficiaries nor suspected of fraud. In March 2020, UDK told the DPA that they had complied with the decision and removed the data they had collected illegally. Whether or not UDK applied the change just to housing benefits or to their whole data management system is unclear. (The case referred to the DPA was about housing benefits only, so that the DPA could not give an opinion on UDK's other activities).



*AUTOMATION CAN SPEED UP  
MANY BUREAUCRATIC PROCESSES  
BUT ALSO REDUCES THE ROOM  
FOR NEGOTIATION BETWEEN  
CITIZENS AND CASEWORKERS.*

## / Limiting negotiations

Whether or not UDK limits the scope of its surveillance, and whether or not the Data Ethics Council succeeds in tooling up the government for its ethical challenges, the automation of welfare management still fundamentally changes the relationship between the state and its citizens.

Søren Skaarup, the former head of citizen services in Albertslund Municipality and now a post-doctoral researcher at the IT University of Copenhagen, published his PhD on the mediation of authority in 2013. In it, he warned that automation, while it can speed up many bureaucratic processes, also reduces the room for negotiation between citizens and caseworkers. The recognition of a beneficiary's individuality, and the construction of trust, are mostly born of face-to-face interactions, Mr. Skaarup wrote. Removing these interactions risks weakening the sense of justice and fairness between citizens and the state.

# Research

By Brigitte Alfter

## Contextualization

Denmark has stated that it wishes to be a front-runner in digitalization (Denmark, 2018). As a result, since January 2018, all new legislation must be “ready to be digitised” (Finansministeriet, 2018), which implies the option of partially or fully automated decisions. Although Denmark is a small country with less than six million inhabitants, the fact that digital solutions are already widespread means that this ambition does not appear to be completely unrealistic. Online solutions have long replaced annoying queues at car registration, tax, land registry offices, and other places, and today public administration is digital by default. It is true that, sometimes, cases surface about citizens who get into trouble – like a couple who did not register their baby’s name due to difficulties logging in to the birth registration system (Ritzau, 2019) – but, generally speaking, Denmark is well-positioned to become one of the world’s digital front-runners.

When looking at Denmark, experts who work internationally point to the widespread public trust in institutions when compared to other countries. In Denmark, distrust against surveillance concepts are rare, data are widely available, including a unique individual identifier for every registered citizen in the form of a social security number in Det Centrale Personregister (The Central Person Registry, CPR). This unique identifier is required not only to register with the authorities but also for health care services, banks, insurance, education grants, work contracts, social security services, driving licenses, and so forth. These systems allow access to a wide variety of information on each individual, for that information to be analyzed, and – increasingly in some cases – used for profiling, predictions, or even for fully automated decisions.

The public discourse, particularly the critical discourse, on these developments appears to be limited to circles of specialists and elite media, except for very few cases about profiling or predictions – especially concerning children – or indeed, when it comes to platform business models (see for example Sorgenfri Kjær, 2018 or Bernsen, 2019-1).

However, discourse does exist. Friction lines in such discussions go between a wish to digitalize swiftly based upon an intention to serve the citizens as well as to create growth – as suggested by the national strategies (Denmark 2016, Denmark 2018) – and on the other side, there is a wish to balance this with citizens’ rights as expressed by academics, civil society, and civil rights groups.

## A catalog of ADM cases

Denmark is not alone in using automated predictions, assessments, and decisions. For example, major football clubs introduced face recognition systems to refuse entry to known trouble-makers (Andersen 2020-3) in a similar way to developments in other countries. In this chapter, we focus on cases of profiling, decision support, and automated decisions in public administration.

### / The Gladsaxe model – profiling families for early detection of vulnerable children

The “Gladsaxe-model” was a profiling system based upon combinations of ‘risk indicators’ for the early detection of children in vulnerable families. The model started as a pilot project led by the municipality of Gladsaxe, on the outskirts of Copenhagen, along with two other municipalities. The model consisted of a points-based analytical system with parameters to estimate risk. These included such things as parental mental health (3000 points), unemployment (500 points), missed medical appointment (1000 points), missed dental appointment (300 points), and divorce (Sorgenfri Kjær, 2018). Although the overall purpose of the project was laudable, the way the task was carried out was heavily criticized and resulted in the model being put on hold. This meant that Gladsaxe and its partner municipalities were not allowed to roll out the system. However, in late 2018, the municipality stated that they



had continued work on the algorithm, and had not only included data from the municipality itself but also statistical data on children who had already received special support as well as their families (Andersen, 2018). In 2019, after the data protection authorities denied permission, and critical coverage in the press, work stopped without any further explanation (Andersen, 2019). In 2020, a new research project at the University of Århus announced that it was developing “an algorithmic tool for decision support” to detect particularly vulnerable children (Andersen 2020). This research project was also criticized for following the same line of thought as the Gladsaxe Model (Andersen 2020-2).

## / Profiling the unemployed

Imagine you lose your job, and, for a while, you have to rely upon unemployment benefits. This system is a public insurance model that you contributed to while you were working. To claim your benefits, a profile of you is made to estimate how likely it is that you will end up in long-term unemployment. The profile is automatically generated, by a nationwide tool and based upon a combination of various data about you, such as your age, origin, unemployment history, previous jobs, education, and the cause of unemployment, including any health indications (Lovforslag L209, 2019. P. 212). This profile of a citizen’s likelihood to become long-term unemployed is used to assist the official dealing with the case. It is not an automated decision but a decision-supporting tool.

In May 2019, questions about this process became a full-fledged public discussion, when a new law was passed that included a passage allowing the minister of labor to develop guidelines for the use of “a digital clarification – and dialogue tool to be used by job centres and unemployment units” (Lov om aktiv beskæftigelsesindsats, 2019). The new tool was set to combine the newly unemployed person’s estimate of their working potential with other data that would then be matched against the characteristics of citizens who previously ended up in long-term unemployment. The tool would then flag those with a higher risk of long-term unemployment (Andersen, 2019-2).

THE MODEL CONSISTED OF A POINTS-BASED ANALYTICAL SYSTEM WITH PARAMETERS TO ESTIMATE RISK. THESE INCLUDED SUCH THINGS AS PARENTAL MENTAL HEALTH, UNEMPLOYMENT, MISSED MEDICAL APPOINTMENT, MISSED DENTAL APPOINTMENT, AND DIVORCE.

The public discussion concerned both the content of the model and the process of how the law was passed without consulting the relevant public data protection authority, Datatilsynet. Exceptionally, Datatilsynet reopened its evaluation of the law, emphasizing that the ministry did not alert it to the automated tool in approximately 1000 pages of background material to the draft law (Datatilsynet, 2019). In August 2019, however, Datatilsynet concluded that the plans for automated profiling of the unemployed was compliant with the rules, but it also suggested that the tool should be re-evaluated to estimate whether the “mathematical model continued to be relevant” to ensure that the rights of those registered were respected “and to avoid unobjective discrimination” (Datatilsynet 2019-2).

Two political parties – both supporters of the current social democratic minority government – announced their objections against profiling and that they would push for evaluation and less profiling. The reason given was that citizens might feel uncomfortable and would be less trustful in sharing information with officials. Although the tool was planned to be used for decision supporting only, one of the opposition politicians was reluctant: “The reality in the employment system is that officials have far too many cases on their table. Hence a profiling tool easily can be what you turn to because it goes faster than a thorough talk,” Eva Flyvholm of Enhedslisten said to the Altinget news service. Flyvholm also argued for sufficient staffing at job centers (Kaasgaard, 2019).

One commercial solution, a so-called “assistant” by the name of “Asta”, is already available and offered by a specialized commercial company. In its advertising material, the company describes Asta as being “equipped with a toolbox containing loads of technological handles such as robot technology and artificial intelligence” and that Asta can be likened to the person peeling the potatoes for a tv-chef before a broadcast on cooking (Schultz, 2019). Asta is already used by the employment agency in Copenhagen and was described positively in the online magazine of KL, the association of Danish municipalities (Romme Andersen, 2019).

## **/ Controlling social welfare recipients**

Udbetaling Danmark (Payout Denmark, UDK) is a public authority responsible for payments such as study grants, child allowances, the minimum pension, maternity, and sick support for those unemployed, and many other payments under the welfare state system. Since 2012, this centralized agency has been responsible for payments previously handled and controlled by local municipalities. Today, UDK is not only responsible for payments but also for some control over the lawfulness of payments. Some payments are made automatically, for example, Statens Uddannelsesstøtte (the State Education Support or SU), which is based upon clear criteria, such as whether or not a citizen is accepted as a student at a university.

Criticism, however, has surfaced about the control functions of UDK. A report published in July 2019 questioned the proportionality of the surveillance, which included the screening of “2,7 million non-unique beneficiaries” leading to only 705 cases in 2017 and 1094 cases in 2018 where the payments were stopped, adjusted, or paid back because of mistakes. In 2017, UDK handed 25 cases to the police for further investigation. (Eiriksson, 2019-3).

This criticism of the surveillance of large parts of the Danish population without actual suspicion is not new. The mindset of the limitless collection of personal data is exemplified in the comments on a case regarding the personal data of a representative of a beneficiary. The data of the representative, as well as close family, appear to have been collected. UDK said that it could not fulfill its legal obligations without collecting the data of the representatives of beneficiaries as well as the family members of such a representative. (Eiriksson, 2019-3, p. 17). What is new is a description of how the analysis of the controls is carried out. Data analysis is based upon a set of criteria that are being developed constantly, including “modular criteria” where data analysis is run on different questions, and all such sub-results are, then, included in the overall conclusion (Eiriksson, 2019-3, pp. 10-11).

## **/ Taxation and the automatization of property evaluation**

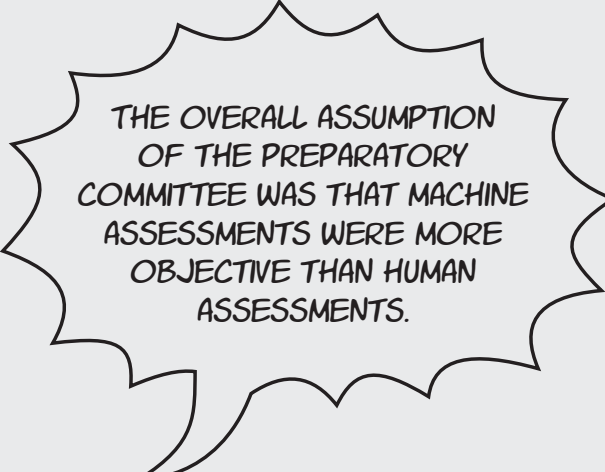
For years, and with varying degrees of success, the Danish government has tried to centralize and automate tax administration. Attempts to centralize and automate tax collection in 2005 with EFI (Et Fælles Inddrivelsessystem or One Shared Tax collection system) failed, and EFI was

shelved in 2015 (Alfter, 2019, 51). A new system, to collect billions of Danish crowns owed by taxpayers, was expected to be finalized by mid-2019, however, in September 2019, the Auditors of the Danish Parliament, Rigsrevision, published a critical report indicating that tax collection would only be fully up and running by the end of 2020, or later (Rigsrevisionen, 2019).

Missing out on collecting billions of Danish crowns in tax revenue, however, is just one of several automatization concerns on the table of every new minister of finance. One heated discussion is about automatizing the estimation of property values – particularly in a context where housing costs are growing significantly in the capital and near larger cities while they are dropping elsewhere, thus affecting property owners in a country where large parts of the population own their own home.

Passed in 2017, the Ejendomsvurderingslov (law on property assessment, 2017) states that automatic real estate valuation assessment will become standard practice. In preparation for the law, this fully automated system was described as “machine assessments” (Motzfeldt & Abkenar, 2019, p. 36-37). The system is set to include “sales prices related to the neighborhood, location and size. Sales price levels are adjusted to take into account the distance to forests and recreation areas, schools, streets, railways”. The system will also take into account whether or not there is the possibility to build on the surrounding terrain and whether or not ground pollution is an issue (Motzfeldt & Abkenar, 2019, p. 37). A total of 19 parameters will be used to automatically calculate property value (Version2 Editorial, 2019). A preparatory committee prepared an analysis to guide the government ahead of the legislation. The committee followed the overall line to move from “subjective to objective” assessments, from human to digital assessments, and from manual to automatic administration, to improve assessments and lower administration costs (Engberg-udvalget, 2014, p. 68-69). While it does not explicitly mention it, this model uses a level of machine learning to develop a “statistical predictive model” (Motzfeldt & Abkenar, 2019, p. 37). Furthermore, the overall assumption of the preparatory committee was that machine assessments were more objective than human assessments, where human assessments are considered subjective.

As with many other systems, the automated property assessment system has been significantly delayed. Along the way, fears surfaced that it did not comply with Danish administrative law (Mølsted, 2018). It also proved to



THE OVERALL ASSUMPTION  
OF THE PREPARATORY  
COMMITTEE WAS THAT MACHINE  
ASSESSMENTS WERE MORE  
OBJECTIVE THAN HUMAN  
ASSESSMENTS.

be more expensive than the estimate in the initial report regarding the cost of a property valuation calculated by a human (Motzfeldt, 2019-2). This project, and other problematic attempts by the Danish tax authorities to incorporate IT projects, have been described as “the ministers’ (...) IT-phantoms” (Version2 Editorial, 2019). The law to assess property values allows a variation of 40% in the value of a given property – 20% either side of the price calculated by the system – “and one does not need to hold a PhD in human behaviour to predict that this will lead to protests if or when the system comes into use. Worst case we will see headlines that two newly built and identical houses will be taxed very differently. Could we blame citizens if they ask critical questions why they had to spend billions for a system with such a low level of precision” (Motzfeldt, 2019-2).

## **/ Exam control in schools: fighting fraud or profiling students?**

You try to control me, I outsmart you. This was the logic of a Danish high school student who, in a post on Github, described how to disable a digital exam monitor (Schou, 2019). The move was part of massive protests by the students against The Digital Exam Monitor, developed by the Danish Ministry of Education (Den Digitale Prøvevagt), and which was set to roll out in spring 2019. Following the protests against this surveillance, which included copies of the clipboard, active websites in four browsers, a key logger, a screen dump every minute, or whenever an active window changed, detection of virtual machines in the background, a list of all active programs, various unique IDs, etc.) (Møller Kjemtrup, 2019), the ministry of education put the tool on hold (Undervisningsministeriet, 2019). Yet under the new social democrat government, students continue to ask the authorities to shelve it entirely (Mejlgaard, 2019).

Meanwhile, another tool that uses machine learning to detect exam fraud has been developed at Copenhagen

University. This move stirred new controversies and fears about student profiling in 2019.

In 2014, high school software provider Macom supplied a tool called Lectio, which operated as an early detection system to trace students in danger of dropping out (Copenhagen University, 2014). The tool was developed by university students and based upon information gathered through the administration and communication system used by 9 out of 10 Danish high schools, according to Macom and media reports. However, use of the algorithm was stopped about a week after it was introduced. The data used to develop it had been handed to a third party – the university and its master’s students – without permission from the high schools (Møllerhøj, 2015).

More recently, in May 2019, a group at the same department of Copenhagen University published a machine learning system by the name of Ghostwriter. The developers said that it could be used to trace the writing style of students in assignments and that it could detect with some certainty, whether or not a student had submitted a text written by someone else. The aim of Ghostwriter was to fight exam fraud, but (the university wrote in a press release) that the model could also be applied elsewhere to detect texts written in an unusual style (Copenhagen University, 2019).

The machine learning was based upon 130,000 student assignments held in the Lectio system by Macom. The ensuing controversy between the high schools, the university professor, and the director of Macom focused on a breach of contract related to the handling of the students’ data: “We feel unsafe when the data handling agreement is not respected,” a representative of the high schools stated to online technology news website Version2, (Bang Frederiksen, 2019-1; Bang Frederiksen, 2019-2; Bang Frederiksen, 2019-3).

## **Policy, oversight and debate**

### **/ Digitalization strategies**

Denmark follows national digitalization strategies, and, at the moment, the 2016-2020 strategy is in action (Digitaliseringsstyrelsen, 2016). It is a strategy that, ever since 2001, builds upon previous years’ strategies (Motzfeldt, 2019,

p. 20). An inherent part of this strategy requires Danish authorities at the state, regional, and local level to promise each other to push for digital solutions that are easy to use and further growth while maintaining security and trust in digital systems (Digitaliseringsstyrelsen, 2016-2). In October 2018, the center-right government introduced its digitalization policy under the title Digital service i verdensklasse (World Class Digital Service, Ministry of Finance, 2018). In 2019, this was followed, with an AI strategy “to serve all of Denmark” (Digitaliseringsstyrelsen, 2019). This strategy points out three focus areas for pilot projects, or so-called ‘signature projects’. These are health, social welfare, and labor (Motzfeldt, 2019-2, p. 35). These national-level strategies are accompanied by investment funds as well as information campaigns such as Digital Hub Denmark, which is a “public-private partnership between the Danish Government, the Confederation of Danish Industry, the Danish Chamber of Commerce and Finance Denmark” (Digital Hub Denmark, 2019).

## / Local authorities & Dare Disrupt

Strategies for digitalization are also being developed at the municipal level. In practice, social welfare and contact with citizens is normally handled by Denmark’s 98 local municipalities that have a certain level of autonomy, and are organized via Kommunernes Landsforening (National association of municipalities, KL). Municipalities are responsible for the care of the elderly, child care, primary schools, social welfare, and local infrastructure. KL has a four-year strategy 2016-2020 following the national agreement (KL, 2016).

In 2018, and on behalf of the 98 municipalities, KL asked the consultancy company, Dare Disrupt, to look into the potential of digitalization at the local level, particularly, when working with children, social welfare, the elderly, health, and other tasks, where municipalities directly serve citizens (Dare Disrupt, 2018; KL 2018).

KL’s use of this particular consultancy group was criticized in both the specialized and the general media, as were some of the elements of the report by Dare Disrupt. In the report, Dare Disrupt emphasized the competences of the scandal-ridden British political consulting company, Cambridge Analytica, and suggested local municipalities should use similar methods for citizen involvement (KL, 2018, p. 108; Sorgenfri Kjær, 2018-

2; Egedal, 2018). It is worth mentioning that the report emphasized the opportunities of profiling and prediction. For example, it suggested that it would be possible to predict how children younger than six years old would succeed in life, “who would be successful, and who would become costly” (KL 2018, p. 54). Another scenario described how municipalities can “predict destructive behavior of the individual citizen through analysis of data in the infrastructure of the municipality even before the citizen can and definitely before the act is carried out (own abuse, abuse of others or violence...)?” (KL, 2016, p. 198).

Confronted with ethical and privacy considerations, KL – who ordered the report – emphasized that the report was merely supposed to act as inspiration for possible opportunities while at the same time, KL did not address potential obstacles. Ethical considerations, a representative stated, were up to the individual municipalities themselves (Andersen, 2018-2).

## / The new government’s legal package on security

In June 2019, a new government took power. It is a social-democratic minority government supported by two left-wing parties and one liberal party. The government agreement does not mention data, digitalization, automatization, or surveillance considerations explicitly – an absence that has been criticized by actors in the field. In one of its first legal packages in the field of justice, the government considered increasing camera surveillance by private and public actors (including the police), more license plate recognition systems (including longer storage of license plate data), more surveillance of gang members, and more police drones, among other actions (Justitsministeriet, 2019). The bill was tabled in February 2020 (Folketinget 2020).

The political discussion on the government’s legal package is ongoing, and the deputy chairperson of the largest opposition party, the center-right Venstre party, has brought some suggestions to the table. These include the introduction of face recognition technology in the so-called ghetto areas, labeled as such in a legal package (previously mentioned), which defines neighborhoods with rates of high unemployment, high criminality, and high numbers of immigrants. The deputy chairperson’s suggestions were commented on by a

THE APPARENT  
ENTHUSIASM TO  
QUICKLY DIGITIZE PUBLIC  
ADMINISTRATION DOES  
MEET HIGHLY QUALIFIED  
CRITICISM.

fellow Venstre party member who said that the suggestions were not official party policy (Ahrenkilde Holm, 2019).

## / “Grab the safety buoy”

The apparent enthusiasm to quickly digitize public administration does meet highly qualified criticism. Under the headline “Do grab the safety buoy”, in the journal for research into social work, Hanne Marie Motzfeldt, associate professor at the faculty of law at the University of Copenhagen, emphasized the need to respect good administrative practices and ethics when digitizing public administration functions. She argued that a decision by the Danish Ombudsperson – an authority in both the political and the legal contexts – should also be used in the field of digitalization to ensure a “qualified and correct basis for decisions” before implementing new tools for digital administration. “It is no longer enough to ask an external consultancy report recommending this or that” (Motzfeldt, 2019-2, p. 40).

Motzfeldt has contributed significantly, not only to a qualified debate but also to the upgrading of legal competences in the field of digital administration – including various levels of automation. Most recently, she published a book targeted at lawyers working in the concept phase of new digital administration projects. In it, she suggests using clear terminology and raising awareness of the division of roles in teams preparing systems for digital administration (Motzfeldt, 2019).

While Motzfeldt’s book is worth mentioning, it is focused on the law, legal professionals, and the use of digital tools from practical assistance up to fully automated decisions. However, more popular books have been published which, more generally, address various questions related to digitalization in societies. For example, in “Digital totalitarianism”, the author – a PhD fellow of philosophy at the University of Copenhagen – takes a closer look at the much-criticized Chinese social scoring models. He argues that the surveillance and social scoring models are far more nuanced than depicted in Western media, and that very similar tools and mechanisms are applied at various levels – in both commercial and public systems – in Western countries (Vestergaard, 2019).

## / Data ethics council

In April 2019, the government announced a new Data Ethics Council (Dataetisk Råd, 2019). The purpose of the council is to create a forum to address ethical questions, on the balance between the advantages of new technologies on the

one side, and civil rights and liberties, as well as societal values, on the other. The council is explicitly tasked with creating a public debate and to constantly support “responsible and sustainable” use of data in both the business and the public sector. The members of the council are experts from public administration, universities, civil society, and business. The council began its work in the course of 2019.

## / Critical observation and information

Several civil society and professional groups take an active part in discussions. In academia, specialists in various disciplines regularly contribute to the public debate and write books and guidance in fields such as law, civil liberties, public administration, among other areas.

Two private think-tanks focus on civil liberties (Justitia 2019) and data ethics ([Dataethics.eu](https://dataethics.eu), 2019), respectively. Meanwhile, the Human Rights Institute, a government-supported research institute, also specializes in this field (Institut for Menneskrettigheder, 2019).

Media coverage of digitalization and automation appears sporadic. In 2019, a dedicated weekly radio program, Aflyttet, ended when the broadcaster that hosted it lost its financial support. However, a specialized online news website, called Version2, provides daily updates, focusing on tech in society and often includes critical voices. Version2 is published by Teknologiens Mediehus (Technology Publishing House), a private company owned by the trade union of Danish engineers and two pension funds. In addition, several online newsletters provide specialized information; however, this is hosted behind expensive paywalls and thus inaccessible to the wider public.

# Key takeaways

Digitalization, including automated decisions, automated preparation of decisions, and profiling, is quickly gaining ground in Denmark. This development is likely to continue, and potentially even speed up, as the government’s current strategies and legislation follow the explicit purpose of paving the way for future digitalization.

The speedy development of digital solutions, rather than focusing on functioning solutions, appears to have happened in far too many instances. Critics emphasize that



rapid development in several cases has been prioritized over respect for both good administrative practice and civil liberties, thus ultimately risking the population's trust in digitalization.

This development, which has been pushed by previous center-right governments, and is also under discussion by the center-left government elected in 2019. A recent comment by the social democratic minister of finance – when he stated that there is no need to wait for the new Data Ethics Council to comment on a pool of 15 publicly-funded

experimental cases on artificial intelligence in the Danish health system – indicates an eagerness in government to develop swiftly rather than prioritizing ethics (Fribo, 2019).

“We need time to breath and to seriously think about which digital solutions it makes sense to invest in,” stated two senior personalities in the field in a recent op-ed arguing that speedy implementation alone is not valuable. They also suggested more careful, case-by-case, considerations if Danish public administrators do not want to sacrifice oversight for speed (Motzfeldt & Kjellerup Hansen, 2019).

## References:

- Ahrenkilde Holm, Thue (2019): Inger Støjberg åbner for ansigtsgenkendelse i ghettolistede boligområder – men ifølge Jan E. Jørgensen er det ikke Venstres politik. Berlingske, <https://www.berlingske.dk/politik/inger-stoejberg-aabner-for-ansigtsgenkendelse-i-ghettolistede>, (12.12.2019).
- Alfter, Brigitte (2019): Denmark, in Automating Society – Taking Stock of Automated Decision-Making in the EU, page 50-51, Algorithmwatch. <https://algorithmwatch.org/en/automating-society> (29.01.2019).
- Andersen, Tania (2018): Gladsaxe arbejder videre på overvågningsalgoritme – trods nej fra ministerium, Version 2. <https://www.version2.dk/artikel/gladsaxe-arbejder-videre-paa-overvaagningsalgoritme-trods-nej-ministerium-1087097> (20.12.2018).
- Andersen, Tania (2018-2): KL: Kommunerne må selv klare etiske fremtidsproblemer. Version2, <https://www.version2.dk/artikel/kl-kommunerne-maa-selv-klare-etiske-fremtidsproblemer-1084088>, (22.01.2018).
- Andersen, Tania (2019): Gladsaxe Kommune sætter algoritme-udvikling på pause, Version 2. <https://www.version2.dk/artikel/gladsaxe-kommune-saetter-algoritme-udvikling-paa-pause-1087433> (06.02.2019)
- Andersen, Tania (2019-2): Ny lov: Arbejdsløse skal dataprofileres og samkøres i landsdækkende register, Version2. <https://www.version2.dk/artikel/ny-lov-arbejdsløse-skal-dataprofileres-samkoeres-landsdaekkende-register-1088085> (15.05.2019).
- Andersen, Tania (2019-3): Sådan kommer Horsens' bedrageri-algoritme til at se ud (This is what Horsens' fraud algorithm will look like). Version2, <https://www.version2.dk/artikel/saadan-kommer-horsens-bedrageri-algoritme-at-se-ud-1087555> (27.02.2019).
- Andersen, Tania (2020): Advokat sætter spørgsmålstegn ved lovlighed af nyt AI-værktøj om udsatte børn. <https://www.version2.dk/artikel/advokat-saetter-spoergsmaalstegn-ved-lovlighed-nyt-ai-vaerktoej-udsatte-boern-1089929>
- Andersen, Tania (2020-2): Gladsaxe-modellen spørger: Nyt AI-projekt skal forudsige mistrivsel hos børn. <https://www.version2.dk/artikel/gladsaxe-modellen-spoeger-nyt-ai-projekt-skal-forudsige-mistrivsel-hos-boern-1089919>
- Andersen, Tania (2020-3): Sådan fungerer Brøndby Stadions ansigtsgenkendelse. <https://www.version2.dk/artikel/saadan-fungerer-broendby-stadions-ansigtsgenkendelse-1090030>
- Bang Frederiksen, Rune (2019): DIKU-professor afviser problemer med dataetikken i algoritme til at opdage købte opgaver. Version2, <https://www.version2.dk/artikel/diku-professor-afviser-problemer-med-dataetikken-algoritme-at-opdage-koebte-opgaver-1088236> (11.06.2019).

Bang Frederiksen, Rune (2019): Lectio stiller 130.000 opgaver til rådighed for forskere uden at informere gymnasier eller elever. Version2, <https://www.version2.dk/artikel/lectio-stiller-130000-opgaver-raadighed-forskere-uden-at-informere-gymnasier-eller-elever> (08.07.2019).

Bang Frederiksen, Rune (2019): Lectio-direktør om kritik af forsker adgang: Gymnasieopgaver indgår kun som data til maskinlæring. Version2, <https://www.version2.dk/artikel/lectio-direktoer-kritik-forskeradgang-gymnasieopgaver-indgaar-kun-data-maskinlaering-1088414> (10.07.2019).

Bernsen, Marcus (2019-1): Høj surf på databølgen (High surf on the data wave), Weekendavisen. <https://www.weekendavisen.dk/2019-33/samfund/hoej-surf-paa-databoelgen> (16.08.2019)

Copenhagen University (2014): Big Data-værktøj forudsiger hvilke gymnasieelever, der vil droppe ud. <https://di.ku.dk/Nyheder/2014/big-data-vaerktoej-til-gymnasierne/> (13.06.2014).

Copenhagen University (2019): Fristet til at snyde med eksamensopgaven? Kunstig intelligens opdager dig med 90 procent sikkerhed (Tempted to cheat with an assignment? Artificial intelligence will detect you with 90 percent security). <https://presscloud.com/csp/vocast/message.csp?KEY=888686758981081> (29.05.2019).

Dare Disrupt (2018): DareDisrupt og KL kortlægger kommunernes teknologiske fremtid. <https://daredisrupt.com/daredisrupt-og-kl-kortlaegger-kommunernes-teknologiske-fremtid/> (16.01.2018).

DataEthics.eu (2019): DataEthics is a politically independent ThinkDoTank. <https://dataethics.eu/about/>

Dataetisk Råd (2019): Dataetisk Råd. VLAK-regeringen nedsatte i maj 2019 et nyt Dataetisk Råd. Digitaliseringsstyrelsen deltager i sekretariatsbetjeningen af rådet. <https://digst.dk/data/dataetisk-raad/> <https://digst.dk/data/dataetisk-raad/> (03.04.2019).

Datatilsynet (2019-1): Datatilsynet ser på beskæftigelseslov igen (Data protection authority to revisit employment law), press release. <https://www.datatilsynet.dk/presse-og-nyheder/nyhedsarkiv/2019/maj/datatilsynet-ser-paa-beskaeftigelseslov-igen> (15.5.2019)

Datatilsynet (2019-2): Supplerende høringssvar over forslag til lov om en aktiv beskæftigelsesindsats og forslag om ændring af lov om organisering og understøttelse af beskæftigelsesindsatsen mv., lov om aktiv socialpolitik, lov om sygedagpenge, integrationsloven og forskellige andre love (konsekvenslovforslag). <https://www.datatilsynet.dk/media/7758/fornyet-udtalelse-til-star.pdf> (5.7.2019)

Denmark, government Lars Løkke (2018): Strategi for Danmarks digitale vækst / Digital strategi: Danmark skal være en digital frontløber (Strategy for Denmark's digital growth / Digital strategy: Denmark shall be a digital frontrunner). <https://www.regeringen.dk/nyheder/2018/strategi-for-danmarks-digitale-vaekst/> (30.01.2018).

Denmark, government Lars Løkke, regional and local municipalities (2016): Den fællesoffentlige digitaliseringsstrategi 2016-2020 (the general public digitalisation strategy 2016-2020). <https://www.kl.dk/okonomi-og-administration/digitalisering-og-teknologi/digitaliseringsstrategier/den-faellesoffentlige-digitaliseringsstrategi-2016-2020/> (16.05.2016).

Digital Hub Denmark (2019). <https://digitalhubdenmark.dk/about/> (14.12.2019).

Digitaliseringsstyrelsen, (2016). Om Digitaliseringsstrategien 2016-2020. <https://digst.dk/strategier/digitaliseringsstrategien/om-digitaliseringsstrategien-2016-2020/> (11.05.2016).

Digitaliseringsstyrelsen (2016): A STRONGER AND MORE SECURE DIGITAL DENMARK. Digital Strategy 2016-2020 (in English). [https://digst.dk/media/16165/ds\\_singlepage\\_uk\\_web.pdf](https://digst.dk/media/16165/ds_singlepage_uk_web.pdf) (11.05.2016)

Digitaliseringsstyrelsen (2019): Ny strategi: Kunstig intelligens til gavn for hele Danmark. <https://digst.dk/nyheder/nyhedsarkiv/2019/marts/ny-strategi-kunstig-intelligens-til-gavn-for-hele-danmark/> (14.03.2019).

Egedal, Morten (2018): KL-rapport om digitale visioner skamroser Cambridge Analytica. Version2, <https://www.version2.dk/artikel/kl-rapport-digitale-visioner-skamroser-cambridge-analytica-1084698>, (22.03.2018).

Eiriksson, Birgitte (2019-1): Justitia: Vi frygter hovedløs digitalisering, Altinget.dk, <https://www.altinget.dk/digital/artikel/justitia-vi-frygter-hovedloes-digitalisering> (08.05.2019).

Eiriksson, Birgitte (2019-2): Justitia: Digitalisering kræver demokratisk sikring af privatlivet, Altinget.dk, <https://www.altinget.dk/digital/artikel/justitia-digitalisering-kræver-demokratisk-sikring-af-privatlivet> (22.05.2019).

Eiriksson, Birgitte (2019-3): Analyse: Udbetaling Danmarks systematiske overvågning (Analysis: Pay Out Denmark's systematic surveillance). Justitia, <http://justitia-int.org/analyse-udbetaling-danmarks-systematiske-overvaagning/> (15.07.2019).

Engberg-udvalget (2014): Forbedring af ejendomsvurderingen: Resultater og anbefalinger fra regeringens eksterne ekspertudvalg. <https://www.skm.dk/aktuelt/publikationer/rapporter/forbedring-af-ejendomsvurderingen-resultater-og-anbefalinger-fra-regeringens-eksterne-ekspertudvalg/> (11.09.2014).

Finansministeriet (2018): Bred politisk aftale skal gøre lovgivningen klar til digitalisering, <https://fm.dk/nyheder/nyhedsarkiv/2018/januar/bred-politisk-aftale-skal-goere-lovgivningen-klar-til-digitalisering/>

Folketinget (2020). Tryghed og sikkerhed i det offentlige rum. <https://www.ft.dk/da/search?q=tryghed%20og%20sikkerhed%20i%20det%20offentlige%20rum&SearchDepthFull=&sf=&msf=&as=1> (04.06.2020).

Fribo, Anders (2019): Stor AI-satsning på danske sygehuse: Finansministeren afviser at vente på Dataetisk Råd. Version2, <https://www.version2.dk/artikel/stor-ai-satsning-paa-danske-sygehuse-finansministeren-afviser-at-vente-paa-dataetisk-raad> (21.11.2019).

Hjelm Poulsen, Birgitte (2019): Konsulent til Justitia: Beslutninger kommer ikke tæt på borgerne med jeres forslag, Altinget.dk, <https://www.altinget.dk/digital/artikel/it-chef-til-justitia-beslutninger-kommer-ikke-taet-paa-borgerne-med-jeres-forslag> (15.05.2019).

Institut for Menneskerettigheder (2019): Forskning om teknologi. <https://menneskeret.dk/viden/forskning/emne/forskning-teknologi>

Justitia (2019): Civil liberties thinktank. <http://justitia-int.org/om-justitia/>

Justitsministeriet (2019): Styrket beskyttelse af grænsen mod Sverige og flere værktøjer til politiet skal sætte hårdt ind mod kriminelle. <https://www.justitsministeriet.dk/pressemeddelelse/styrket-beskyttelse-af-graensen-mod-sverige-og-flere-vaerktoejer-til-politiet-skal-saette-haardt-ind-mod-kriminelle/> (10.10.2019).

Kaasgaard, Kasper (2019): Støttepartier ser kritisk på samkørsel af data om ledige, Altinget. <https://www.altinget.dk/digital/artikel/stoettepartier-ser-kritisk-paa-samkoersel-af-data-om-ledige> (2019-08-09).

KL (2016): Ny fællesoffentlig digitaliseringsstrategi – med fokus på at skabe fundamenter for digital udvikling i den samlede offentlige sektor (New public digitalisation strategy with a focus on creating the basis for digital development in the entire public sector). <https://www.kl.dk/nyhed/2016/maj/ny-faellesoffentlig-digitaliseringsstrategi-med-fokus-paa-at-skabe-fundamenter-for-digital-udvikling-i-den-samlede-offentlige-sektor/> (17.05.2016).

KL (2018): Kommunernes teknologispring. Kommunernes Landsforening <https://www.kl.dk/politik/teknologispring/>, full report, <https://www.kl.dk/media/15341/kommunernes-teknologiske-fremtid-fuld-version.pdf>.

Lov om aktiv beskæftigelsesindsats (law about active support for employment) (2019), §8.2, <https://www.retsinformation.dk/eli/lt/2019/548#idefaafed5-e8e7-4062-abf9-c62a4c18d707> (07.05.2019).

Lovforslag L209 (2019): Forslag til Lov om en aktiv beskæftigelsesindsats (draft law about active support for employment), Danish Parliament records. Special remarks on p. 212-213, [https://www.ft.dk/ripdf/samling/20181/lovforslag/l209/20181\\_l209\\_som\\_fremsat.pdf](https://www.ft.dk/ripdf/samling/20181/lovforslag/l209/20181_l209_som_fremsat.pdf) (27.03.2019).

Lov om ejendomsvurderinger (2017): <https://www.retsinformation.dk/Forms/R0710.aspx?id=191816> (08.06.2017).

Mejlgaard, Martin (2019). For lidt over et halvt år siden formulerede @dgs\_dk 5 krav til en ny undervisningsminister. Der er nok at kæmpe for endnu, men vi det går den rigtige retning #dkpol #uddpol (Mejlgaard is the chairperson of the high school students' association <https://www.gymnasieelever.dk/bestyrelse>), <https://twitter.com/MartinMejlgaard/status/1201934884496969731> (03.12.2019).

Ministry of Finance (2018): Digital service i verdensklasse. <https://fm.dk/udgivelser/2018/oktober/digital-service-i-verdensklasse/> (23.10.2018).

Motzfeldt, Hanne & Abkenar, Azad Taheri (2019): Digital forvaltning. Udvikling af sagsbehandlende løsninger. DJØF Forlag, Copenhagen.

Motzfeldt, Hanne Marie (2019-2): Grib dog redningskransen! (Do grab the safety buoy). Uden for nummer – tidsskrift for forskning og praksis i socialt arbejde, #39/2019, <https://socialraadgiverne.dk/wp-content/uploads/2019/11/39-UdenForNummer.pdf>.

Motzfeldt, Hanne Marie & Kjellerup Hansen, Jens: Debat: Vi har mistet overblikket over digitaliseringen. Altinget, <https://www.alinget.dk/uddannelse/artikel/debat-vi-har-mistet-overblikket-over-digitaliseringen> (11.12.2019).

Møller Kjemtrup, Niels (2019). „Den Digitale Prøvevagt’: Totalovervågning af elevers computere midlertidigt trukket tilbage. <https://www.version2.dk/artikel/digitale-proevevagt-totalovervaagning-elevers-computere-midlertidigt-trukket-tilbage-1087609>

Møllerhøj, Jakob (2015): It-system til varsel af elevfrafald blev øjeblikkeligt standset af gymnasierne. Version2, <https://www.version2.dk/artikel/noejagtig-frafalds-algoritme-blankt-afvist-i-udbredt-gymnasie-it-320922> (17.08.2017).

Mølsted, Henrik (2018): Nyt ejendomsvurderingssystem udsættes: Frygt for ulovlig sagsbehandling af borgernes ejendom. Version 2, <https://www.version2.dk/artikel/nyt-ejendomsvurderingssystem-udsættes-frygt-ulovlig-sagsbehandling-borgernes-ejendom> (31.10.2019).

Rigsrevisionen (2019): Beretning om udvikling af det nye inddrivelsessystem og tilslutning af fordringshavere (Report about the new tax collection system). <https://www.rigsrevisionen.dk/publikationer/2019/222018/> (05.09.2019)

Ritzau (2019): Forældres bøl med NemID førte til sen navngivning af søn, DR.dk, 28.11.2019 <https://www.dr.dk/nyheder/indland/foraeldres-bovl-med-nemid-for-te-til-sen-navngivning-af-son>

Romme Andersen, Bettine (2019): Robotten Asta hjælper jobkonsulenten med at forberede borgersamtalen (Asta the robot helps job the consultant to prepare the meeting with the citizen). <https://vpt.dk/jobcenter/robotten-asta-hjaelper-jobkonsulenten-med-forberede-borgersamtalen> (30.08.2019).

Schou, Carl (2019): The nadir of surveillance (Den Digitale Prøvevagt). Github, <https://secret.club/2019/03/07/exam-surveillance.html> (07.03.2019).

Schultz (2019): Mød Asta – din nye assistent. Interview med Simon Kohrtz, produktchef, Schultz (Meet Asta – your new assistant. Interview with Simon Kohrtz, product chief, Schultz). Schultz, <https://schultz.dk/om-schultz/nyt-fra-schultz/moed-asta-din-nye-assistent/> (accessed 04.12.2019).

Sorgenfri Kjær, Jakob (2018): Regeringen vil overvåge alle landets børnefamilier og uddele point, Politiken, <https://politiken.dk/indland/art6365403/Regeringen-vil-overv%C3%A5ge-alle-landets-b%C3%B8rnefamilier-og-uddele-point> (02.03.2018).

Sorgenfri Kjær, Jakob (2018-2): KL-rapport hylder Cambridge Analytica. Politiken, <https://politiken.dk/viden/Tech/art6392494/KL-rapport-hylder-Cambridge-Analytica> (21.03.2018).

Sorgenfri Kjær, Jakob (2019): Regeringen vil overvåge alle landets børnefamilier og uddele point (Government wants to surveil all parents and give them points), Politiken, <https://politiken.dk/indland/art6365403/Regeringen-vil-overv%C3%A5ge-alle-landets-b%C3%B8rnefamilier-og-uddele-point> (02.03.2018).

The Central Person Registry, CPR: <https://www.borger.dk/samfund-og-rettigheder/Folkeregister-og-CPR/Det-Centrale-Personregister-CPR>

Undervisningsministeriet (2019): Den digitale prøvevagt når ikke at blive klar til sommer. <https://www.uvm.dk/aktuelt/nyheder/uvm/2019/mar/190312-den-digitale-proevevagt-bliver-ikke-et-krav-til-sommer> (07.03.2019).

Version2, editorial (2019): Leder: Skat har erstattet fagfolk med et it-fantom. Version2, <https://www.version2.dk/artikel/leder-skat-har-erstattet-fagfolk-med-it-fantom-1087337> (25.01.2019).

Vestergaard, Mads (2019): Digital totalitarisme. Informations Forlag, <https://butik.information.dk/products/digital-totalitarisme> (16.05.2019).



BELGIUM  
DENMARK

**ESTONIA**

**STORY**

PAGE 64

**RESEARCH**

PAGE 68

FINLAND  
FRANCE  
GERMANY  
GREECE  
ITALY  
NETHERLANDS  
POLAND  
PORTUGAL  
SLOVENIA  
SPAIN  
SWEDEN  
SWITZERLAND  
UNITED KINGDOM





Find out more on p. 69 in the research chapter under "Prediction model for the healthcare needs of patients with chronic illnesses".



# A city is automating homes to reduce **energy** consumption

**The city of Tartu installed automated systems in old housing blocks. Using nudges, sensors, and automated decision-making, it hopes to reduce energy consumption by two-thirds.**

By Gerda Kelly Pill

SmartEnCity is a Europe-wide project that aims to transform cities to become carbon neutral. Tartu, a small campus town in Estonia and home to around 100,000 people, is one of the “demonstrator cities” for this project, which it called Tark Tartu (Smart Tartu). As part of this project, the city introduced a public bike-sharing system, bought gas-powered buses, and installed automated LED streetlights that take into account local weather and traffic conditions and adjust lighting levels accordingly. The city authorities are also retrofitting Soviet-era “khrushchyovkas” into “smartovkas”. This means turning old concrete-panel apartment blocks into modernized, “smart” living spaces.

Khrushchyovkas dot the cityscape of most former Soviet countries. In Estonia, they are an integral part of many neighborhoods. They have housed several generations, although they were originally built to last 50 years – an estimate that was revised upwards by successive governments. The buildings are now thought to be fit for another half-century if properly renovated.

These concrete-panel apartment buildings were built between the 1950s and the early 1970s, when Nikita Khrushchev was the first secretary of the Communist Party of the Soviet Union, hence their nickname of ‘khrushchyovkas’. They were easy and cheap to build, and millions of people still call them home.

In 1991, Estonia became independent from the Soviet Union, society moved on, the economy grew, and cities developed, but the khrushchyovkas remained. They are not without problems: heating is expensive because of subpar insulation, poor ventilation, and deteriorating plumbing. Also, many of these buildings have not been fully-renovated since they were originally built.

Apartment block renovation, to bring it up to current standards, is expensive, requires much work, and the responsibility usually falls on the inhabitants. Khrushchyovkas residents come from all walks of life and ages. Some people have lived in these buildings since the beginning. However, due to a lack of funds, leadership, or, perhaps, opportunity, many khrushchyovkas are in a less-than-prime condition, even though they sit in desirable locations in Tartu.

## / Automated homes

Taking the opportunity of European funding, some of these buildings were incorporated into a pilot program that

**“THE HEATING BILLS ARE LOWER, AND I DO NOT HAVE TO REPORT MY WARM WATER AND GAS USAGE MANUALLY ANYMORE.”**

ANATOLI, TENANT

promised to make them energy-efficient, high-quality living environments with embedded, automated systems.

Upon closer inspection, the reality is far from a futuristic, ultra-modern, automated smart home, but the project is a work in progress, and the real results will only become clear in the years to come. The goal is to renovate and “smarten up” 17 apartment buildings located in the center of Tartu. The systems are still under development, however, there are already tangible results. “Of course, I’m happy. The heating bills are lower, and I do not have to report my warm water and gas usage manually anymore,” said Anatoli, whose house was renovated, as part of the project, in the summer of last year. “My home is finished, and everything works. But in some apartments, gas and electricity consumption indicators are still not showing,” he added.

## / Energy efficiency

“The goal is to get the building’s energy rating as good as possible, and, for that, the inhabitants have to contribute with their behavior,” said Tõnis Eelma, who is one of the project’s leads and the chairman of the apartment association of his building, which was the second bloc to be renovated as a part of the project.

“Every apartment has a tablet attached to the wall where residents can monitor their consumption and, we hope that, based on that information, people adjust their habits,” said Mr. Eelma. The ultimate goal is to lower the building’s yearly energy usage from the current 270 kWh/m<sup>2</sup> to 90 kWh/m<sup>2</sup>.

“One of the most revolutionary things is our demand-driven central ventilation system. This means that the carbon dioxide levels are measured, and the influx of fresh air to your apartment is regulated automatically,” said Mr. Eelma. The solution is unique, as, usually, renovated apartment buildings get a certain amount of fresh air throughout the day regardless of whether you are home or not. The other option is that there are pre-set ventilation intervals, which



only take into account people's 9-to-5 schedule, but not the ones who are home all day, such as the elderly.

Raivo Raestik is the CEO of Enlife, the company that won the tender to develop the smart home systems for the "smart-ovkas". He explained that the tablets enable two main things for the inhabitants: monitoring information about themselves and controlling their indoor climate. The system records electricity, water, and gas usage for each apartment, uploads the data to a cloud environment, and shares it with various stakeholders, such as the city of Tartu.

Users can view their usage statistics, the current and desired inside temperature for all rooms, CO<sub>2</sub> levels (measured in parts per million), and the outside air temperature. They can also check the time and greet guests through an intercom system. And if they previously had to state their hot water and gas usage to the utility company, it is now automatically read remotely.

Ten out of the 17 buildings opted to have their building upgraded with radio connections as opposed to cable connections, meaning that their sensors are all wireless. Smart home systems also allow the user to check battery levels for those remote sensors, such as the smoke alarm or thermostat.

Users can control the temperature and ventilation and monitor their statistics, using the tablet installed in the apartments, through a computer web interface or a smartphone app. Control over ventilation was given to the inhabitants in case they are not happy with the level of fresh air input chosen by the system.

## / Nudges

Seeing statistics is part of raising energy-consumption awareness. Tõnis Eelma said that in the future they would like to offer the option to compare an apartment's energy usage to the rest of the building. The building would then get an average, which in turn could be compared to the other renovated apartments in the project, which could introduce an element of competition.

But it might not be so simple, due to different ways people think and live. "We are still discussing what to compare and with whom," said Mr. Raestik. "The first comparisons were made with the building's average." Those statistics were generated by calculating the usage of all apartments. "But the apartments are very different. You can have an elderly

**USERS CAN  
CONTROL THE  
TEMPERATURE  
AND VENTILATION  
AND MONITOR  
THEIR  
STATISTICS,  
USING THE  
TABLET  
INSTALLED  
IN THE  
APARTMENTS.**

person living alone or big families – that's apples and oranges. If you see that your usage is +156% of the average, then that does not motivate you to save energy." EnLife is rather looking to develop a statistical comparison related to each resident's past behavior. "But that also needs some thought because you cannot indefinitely lower your energy consumption – you still need some level for your basic activities. We do not want people to get discouraged when they cannot lower their energy usage levels anymore."

However, the development is still in progress. Mr. Raestik said that one of the next things in the pipeline is to deliver automated scheduling for temperature levels so that people can have lower temperatures at night without having to manually change the temperature via the system's interface.

In addition to the smart home system, the apartments were fitted with solar panels, insulation for the walls and roof, a new heating and plumbing system, new windows, and outside doors.

## / Unique solution

No solution like this existed in the market before. In Estonia, smart homes have a reputation for being expensive and complex. Few people have these solutions in their homes unless already provided by the developer. Retrofitting homes to become smart is uncommon. The SmartEnCity project ordered a one-size-fits-all solution. EnLife started by developing it for about 600 apartments. They designed everything in a way that allows for retrofitting all the homes in case of future changes because the smart home system is evolving together with the SmartEnCity project.

Comparing this smart home system to Apple Home, Samsung SmartThings, and Google Nest, Mr. Eelma said that there were two main differences: those solutions would not have fit their budget, and they do not focus on energy consumption. The goal of the project is to raise awareness among people about their habits and inspire them to change their behavior to save energy.

Mr. Raestik added that when you create a smart home system at such a large scale, they must consider technical support and continuous updates for as long as the building stands. "In my experience with big players, as time goes by, tech-support for older versions gets discontinued," he said.

The city's total investment to retrofit 17 buildings is around nine million euros, plus 400,000 euros to develop the smart home solution. The exact cost of each building's renovation depends on a tender. "Our building had 32 apartments, and the renovations cost over a million euros. We had to cover around half of that by taking out a loan," said Mr. Eelma. Using Horizon 2020 funds, the city of Tartu helped cover 25% of the total cost, and the rest came from a national renovation grant.

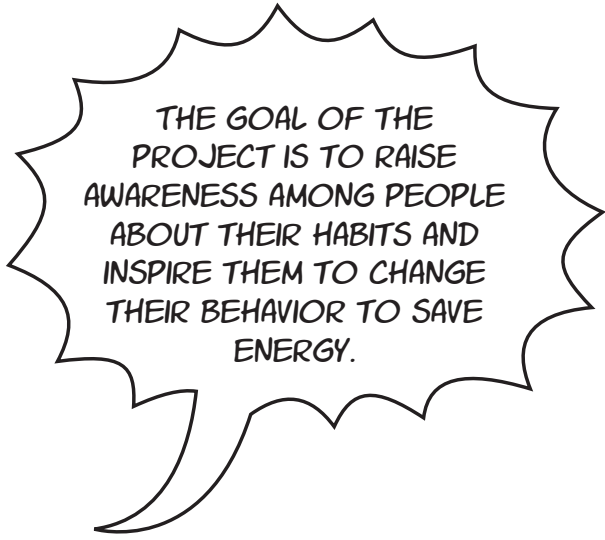
## / Reduced costs

The first results show that, on average, heating bills have been cut in half. "We were hoping that they come down to one-third of the original cost, but we must consider that, before, heat was only used to warm radiators in apartments. Now, in addition to that, it heats water and air in the central

ventilation system," said Mr. Eelma. "But the monetary win is small because while we can save from consuming less energy, the loan payments don't get any lower."

The renovations came with some conditions. For example, the installation of solar panels was mandatory for all buildings to provide additional energy to the buildings. Furthermore, while all the energy-consumption data belongs to the inhabitants, they are obliged to share it (in an aggregated format) with the city of Tartu until 2021 to measure the effectiveness of the renovation program.

"It is a great project; we are testing a lot of new things. But it is not ready yet – fine-tuning the houses' systems to lower costs is still ongoing," said Mr. Eelma. With four years and some results already under its belt, this ambitious project hopes to prove that when you give people the opportunity to monitor their energy-consumption behavior, they will change their habits to save energy in the long run. You can only improve what you can measure.



THE GOAL OF THE PROJECT IS TO RAISE AWARENESS AMONG PEOPLE ABOUT THEIR HABITS AND INSPIRE THEM TO CHANGE THEIR BEHAVIOR TO SAVE ENERGY.



# Research

By Maris Männiste

## Contextualization

In recent years, the Estonian government has taken several steps to support the implementation of artificial intelligence (AI) in both public and private sector institutions. In Estonia, the term “Kratt” is used in relation to AI, and it refers to practical applications that use AI to perform specific functions. Automation related to AI is a major focus of discussion, both in the government and in the public domain.

While there are discussions about the use of automated decision-making or AI, for the moment, the main focus is predominantly on the benefits these technologies will bring. In the public sector, the benefits of AI are seen as threefold: it increases the user-centeredness of services, improves data analysis, and makes the country work more efficiently to achieve the goal of developing an e-government. AI can also play an important role in the digital revolution of industry and to attract new investment and innovation in Estonia – technology developers are searching for a development and test environment that favors AI solutions. In this regard, the Estonian strategy seems to focus on a small, well-established, digital e-government, and on a future that sees the country become a good testing ground for AI solutions.

Several research centers in Estonia focus mainly on the technical aspects of AI or offer a testbed to pilot the solutions to possible problems. In addition, a study that will focus on the opportunities for using AI and ADM in e-governance is also planned. There are several governmental and nongovernmental organizations which have pointed out possible risks and problems associated with automated decision-making. Since 2017, Estonian residents can check the [Eesti.ee](https://eesti.ee) website to see which institutions have accessed their data and for what purposes by using the [Data tracker](#) tool. By September 2019, four major government agencies were participating in this project (The Population Register, The Health Insurance Fund, the Estonian Unemployment Insurance Fund, The Social Insurance Board). The website also indicates if data has been accessed for automatic pro-

cesses or by a specific automated service. This solution is seen as vital in fostering trust and transparency in governmental services (Plantera 2019).

## A catalog of ADM cases

### / Monitoring and profiling risk groups

### / A system to support NEET youth

Since 2018, municipal governments in Estonia have been able to use a tool called The Youth Guarantee Support System (YGSS) (Mis on noortegarantii tugisüsteem? – [Tööelu.ee](#) o. J.). This tool helps case managers (social workers, child protection officers, youth workers, etc.) identify young people aged 16-26 who are not in education, employment or training (NEET). It is used to support them when they need to return to education or enter the labor market. Case managers of the Youth Guarantee Support System or the municipal government employees can see information about young people living in their specific locality. However, up until now, not all of the municipalities have joined the program. In order to join, the municipal governments have to send an application to the Municipal Government Information System for Social Services and Benefits (STAR) (Noortegarantii | Sotsiaalministeerium o. J.). The case managers or municipal government employees are then shown a list in the STAR system of all the NEET youth who are registered as residents in a specific municipal area and who – according to other registry data – need help. This information is based on data from nine different registries, creating, therefore, a very comprehensive dataset that includes information on the young person's address, email, phone number, educational background, and if they studied in Estonian or Russian. Young people have the right to forbid the analysis by a case manager, and in cases where the data is processed, young people must give their consent.

The case manager then contacts the young people identified by the system and a letter or an SMS is sent to each person asking them to complete a survey. The Youth Guarantee Support System Request is automatically triggered twice a year (15.03 and 15.10).

## / Prediction model for the healthcare needs of patients with chronic illnesses

In 2015, the Estonian Health Insurance Fund (EHIF) started cooperating with the World Bank to develop and pilot a risk-based management model for healthcare (Enhanced Care Management: Improving Health for High Need, High Risk Patients in Estonia 2017) that would help increase the integration of health services. Risk-based treatment management enables family physicians to identify patients on their list with multiple chronic illnesses who would benefit most from additional help with prevention, counseling, and follow-up care to improve their quality of life. If these patients are neglected by family physicians, this can lead to serious problems, including unnecessary deterioration of health, which, in addition, causes unnecessary costs for the healthcare system (avoidable hospitalization, duplication of studies, etc.). The aim of the machine learning project was to improve the so-called mechanically constructed algorithm developed during the Risk Patient Pilot Project and to better identify the patients who should be included in the risk patient program.

Between 2016-2017, the Estonian Health Insurance Fund (EHIF) and the World Bank conducted a pilot project on treatment management in Estonia. The first step for this was to develop a clinical algorithm in cooperation with family physicians in order to identify patients who were potentially at risk. The algorithm combines EHIF medical billing data with the clinical opinion of family physicians.

The pilot was officially launched in January 2017 (Enhanced Care Management: Improving Health for High Need, High-Risk Patients in Estonia 2017).

Throughout the pilot, family practitioners joined a series of webinars, led by the local pilot coordinator (World Bank consultant), to reinforce and refresh the initial training. This solution contributes to the empowerment of primary care and also helps find the best algorithm to predict which

patients with selected diagnoses are likely to be admitted to the hospital. The solution first identifies certain medical conditions/diagnoses in the EHIF's medical invoice database. It then provides a practical model for family physicians to predict which patients are more likely to be admitted to the hospital or suffer other health problems. The [EHIF evaluation report from 2017](#) (p.33) also emphasizes that: "the prioritisation of patients within the registry list based on behavioural data (i.e. whether the patients have filled all their prescriptions during past months) and social patient characteristics (e.g. whether the patient may be socially vulnerable) still needs to be fully developed and then used, as currently this information is not used in the patient selection process".

The EHIF has been cooperating with the World Bank since 2014. The pilot was created by the World Bank with the involvement of the Estonian Health Insurance Fund. Costa Rica is also involved in the project by allowing access to its medical billing database (to look at clinical/socio-economic data not included in EHIF databases).

## / Profiling and matching solutions

### / Machine learning software to match job seekers with employers

Estonian start-up Skillific is developing a recruitment service (Skillific kasutuslugu | Krattide veebileht o. J.) that uses a machine learning software program to enable employers to find potential employees by analyzing the skills and knowledge they need. Skillific's search process is based on,

among other things, the European Skills, Competences, Qualifications and Occupations (ESCO) classification system, developed by the European Commission, which defines skills needed in many areas of life. Based on this, the machine learning algorithm looks for potential candidates in the same skill category as the job profile. Skillific's AI application, through the web, partner data, and public databases, can find and

profile potential candidates for different jobs, and also assess their suitability for those jobs (Eestis töötati välja tehisintellekti sugemetega talendiotsingumootor Skillific 2016).

THIS SOLUTION CONTRIBUTES TO THE EMPOWERMENT OF PRIMARY CARE AND ALSO HELPS FIND THE BEST ALGORITHM TO PREDICT WHICH PATIENTS WITH SELECTED DIAGNOSES ARE LIKELY TO BE ADMITTED TO THE HOSPITAL.

At the moment, the software is still in the testing phase as there is not enough good quality training data for the full implementation of the software. The Skillific database already has about 400,000 user-profiles and a smaller number of workplace profiles. Based on this information, the machine learning algorithm predicts the suitability of job candidates and provides better quality solutions through easier storage and collection of data. Without this system, the process is very resource-intensive and generally does not utilize previously analyzed candidate profiles.

The aim in the future is to create an environment similar to the Finnish database, MyData. With the consent of the job seeker, this system would allow companies to access more data and thus provide a better service to both the job seeker and the employer. Skillific estimates that the entire recruitment process could be fully automated within the next 5 to 10 years.

### / ADM solutions used by the Estonian Unemployment Insurance Fund

Estonia's Unemployment Insurance Fund (EUIF) increasingly uses different automated solutions in its everyday work (Kõik muutub automaatseks: töötukassa IT-osakond ei istu päevad otsa mustades ülikondades, tõstes tuimalt pabeleid ühest hunnikust teise 2019). After a citizen registers on the Unemployment Insurance Fund website as unemployed, the data is checked and, if it is correct, the citizen is then registered as unemployed. In the background, the system uses AI to check an applicant's data in different databases. It then decides which document to send to the applicant.

EUIF also uses ADM to decide what a person is entitled to, including what amount of unemployment aid or unemployment insurance aid, and for how long (Raudla 2020). According to Erik Aas, an EUIF council member, 50% of those decisions are made entirely through ADM (Raudla 2020).

In addition, the Unemployment Insurance Fund plans to start predicting how long specific individuals will remain unemployed.

The main goal of this is to make use of all of the data the Unemployment Insurance Fund has.

According to the [Estonian Unemployment Insurance Fund development plan for 2019-2022](#) – which aims to increase the impact, quality and accessibility of services – a profiling model should be used. This would help to identify the risk of long-term unemployment for each client and to choose the appropriate counseling channel, the frequency, and specific services. The Estonian [Kratt report](#) (p.18) indicates that this profiling solution is already in use.

The Unemployment Insurance Fund has the right to use automated decision-making (Millal teeb töötukassa Teie suhtes automaatse otsuse | Töötukassa o. J.) according to the Unemployment Insurance Act § 23 lg 4. The applicant is informed of this ADM process when submitting an application and the use of ADM is also displayed on the decision document.

### / Face recognition and machine vision solutions

#### / Machine vision AI solution for better traffic management

The machine vision Kratt, or AI solution of the Public Transport and Traffic Management Department in Tallinn, started in autumn 2018 in cooperation with Sifr OÜ (Kasutuslood | Krattide veebileht o. J.). The main purpose of this solution is to monitor the traffic load in Tallinn, specifically the cars driving in and out of the city every day.

This information is used to make decisions about parking problems or road construction, among other issues. To test the solution, three cameras were selected to monitor intersections in Tallinn. This helped train the software to count vehicles passing the camera. This solution uses a machine vision algorithm, which counts buses, cars, trucks, and motorcycles. In the future, it is hoped that the same solution can be used to count pedestrians. The accuracy of the results can be affected by different weather conditions, including fog, and by dirt on the cameras.

IF THE CUSTOMER  
WANTS TO BUY  
TOBACCO PRODUCTS,  
THEY HAVE TO CHOOSE  
THE ITEM FROM  
THE SELF-SERVICE  
CHECKOUT AND THEN  
PASS A PRELIMINARY  
FACE DETECTION  
PROCESS WHICH ISSUES  
A PAPER TO ALLOW  
THE CUSTOMER TO  
GO TO THE TOBACCO  
MACHINE.

## / Age-verification technology used in Keila

In Keila, a city in northern Estonia, the Keila Consumer Co-operative uses a self-service check-in desk that verifies the customer's age with special age-verification technology and is designed so that the customer can buy age-restricted products. The self-checkout ID is based on Strongpoint self-checkout software and the Yot Digital Identification Platform (Keila Tarbijate Ühistu võttis kasutusele Euroopas ainulaadse näotuvastusega iseteeninduskassa – StrongPoint o. J.). To purchase age-restricted products, the customer's age is checked at the self-checkout desk and after that the customer can then pay for the item. If the customer wants to buy tobacco products, they have to choose the item from the self-service checkout and then pass a preliminary face detection process which issues a paper to allow the customer to go to the tobacco machine. The vending machine then performs another check before issuing the tobacco. The customer does not have to provide any additional documents to verify their age.

THE SERVICE IS BASED ON THE PRINCIPLE THAT, AS THE STATE ALREADY HAS THE INFORMATION ABOUT EACH CITIZEN FROM BIRTH, ALL THE FOLLOWING SERVICES, SUCH AS FAMILY BENEFITS, CAN BE ACTIVATED AUTOMATICALLY.

This is believed to be the first solution of its kind in use in Europe.

A journalist from the digital newspaper Delfi conducted an experiment (Veltman 2019) to see if the self-service check-in desk face recognition software can distinguish adults from minors. For this purpose, he recruited 14 and 15-year-old girls and boys to try to buy age-restricted products. Initially, it seemed that the technology stopped minors from purchasing age-restricted products, correctly deducing that they were underage. However, on the second attempt, the technology permitted an underage boy to buy the products. The software developer explained that the anomaly happened due to the repeated face detections without completing the purchase the first time. According to them, the problem should be fixed now.

However, several big supermarkets are not ready to start using this software as, for them, this system is still in the development phase and may be a security risk (Belkin 2019).

## / Event-based services

### / Applying for a place at a nursery school, childcare institution, or kindergarten

Education services in the city of Tartu use a system called ARNO ([www.arno.ee](http://www.arno.ee)) for submitting and processing applications for places at nursery schools and for childcare. Tartu uses ADM systems to determine a child's municipal school based upon the child's address registered in the population register. The school placement offering is generated automatically once a year and is based upon three sets of data: the automatic queue for a child of pre-school age, the parent's preferred kindergarten and the nearest kindergarten. Parents can apply at a later stage to change school depending on specific circumstances. In the case of kindergartens, parents have to apply for a kindergarten spot after a child is born. ARNO is developed by OÜ Piksel.

### / Proactive services mean parents no longer have to apply for family benefits

In 2019, the Social Insurance Board (SKA) completed its first automatic proactive service (Sotsiaalkindlustusamet avas esimese proaktiivse avaliku teenuse | Sotsiaalkindlustusamet 2019) meaning that parents no longer need to apply for services. The service is based on the principle that, as the state already has the information about each citizen from birth, all the following services, such as family benefits, can be activated automatically. It is the first event-based service in Estonia, and the SKA aims to move fully towards application-free services.

Once a birth is registered in the population register and the child is given a name, the Social Security Agency will send an e-mail to the parents. Once the parents have confirmed receipt of the notification, they can receive family benefits.

This solution was developed by Nortal and is considered to be unique in the world. The service is based on an automated algorithmic solution, and it has also led to changes being made in some organizational processes (Sotsiaalkindlustusamet avas esimese proaktiivse avaliku teenuse | Sotsiaalkindlustusamet 2019).

## / Robot judge project in Estonia

In 2019, the Ministry of Justice officially requested that Estonia's chief data officer design a robot judge to preside over small claims disputes of less than 7,000 euros to clear a backlog of such cases (Niiler 2019). The project is still in progress, but in theory, the two parties in each case will upload all the relevant information to a database that the AI software will then analyze to render a decision based on pre-programmed algorithms and previous training. The software's decision would be legally binding, but that decision could then be appealed and looked into by a human judge. The country's economic ministry is considering granting AI and robot judges legal status to help the legislation allocate responsibility for decision-making involving AI-controlled software. This solution is still considered an aid to the courts in Estonia and not a standalone solution.

Still, there have been some changes made to Estonian procedural law which will be introduced in several stages between the years 2020-2023 (Mandri 2019).

# Policy, oversight and public debate

## / Political Debates on Aspects of Automation – Government and Parliament

### / The AI Programme – “Kratt project” and the plan for 2019-2021

The AI Programme, called the “Kratt plan” (Majandus- ja Kommunikatsiooni ministeerium/Riigikantselei 2019b), was commissioned by the Government Office and launched at the beginning of 2018. A steering group consisting of experts from public and private sector organizations and academia was established and is led by Siim Sikkut. Kratt is a magical creature from Estonian mythology and it was decided to use this character as a metaphor to facilitate communication. In Estonian mythology, Kratt was a servant who built things from hay or old household items and he needed to be looked after so that it would not go idle. AI is seen as a huge opportunity and a powerful tool for Estonia which also needs to be wielded responsibly and safely so that the Kratt project, like the mythological character, will not come to harm if left unattended.

The main tasks of the expert group were to: prepare draft legislation to ensure clarity in the judicial sector, to organize the necessary supervision, to develop an AI action plan, to notify the public about the implementation of the Kratt project, and publicize possible options (e.g. via development of the [www.kratid.ee](http://www.kratid.ee) website).

The proposals made by the expert group were compiled into a [report](#) (Majandus- ja Kommunikatsiooni ministeerium/Riigikantselei 2019b) that outlines advice and activities on how to accelerate the use of AI in both private and public sector organizations. The proposals made by the expert group in 2019 formed the national AI action plan for 2019-2021 (ajandus- ja Kommunikatsiooni ministeerium/Riigikantselei 2019). The strategy was adopted at a cabinet meeting on 25 July, 2019. According to the strategy, the government takes a leading role in accelerating and supporting the use of AI-based applications in both the public and private sectors alike. The Estonian government will be investing at least 10 million euros between 2019 and 2021 to implement the strategy.

The proposals, and Estonia's AI action plan, focus on developing the basic competencies required for implementation. Moreover, it is emphasized in the planning document and on the [e-Estonia webpage](#), that “Estonia could become the role model and testbed for the rest of the world as a place where Kratt, or AI, is put to work for the people's well-being in both the public and private sectors”.

In particular, Estonia emphasizes (Majandus- ja Kommunikatsiooni ministeerium/Riigikantselei 2019b) the objective in the AI plan should be to boost AI development in the country, starting with basic competences. The plan gives an overview of the development of AI and of the plans in other countries and, based upon this, it recommends that particular effort should be made to facilitate the implementation of Kratts in the public sector.

The 2019-2021 action plan (Majandus- ja kommunikatsiooniministeerium/Riigikantselei 2019a) not only considers the steps that should be taken to boost AI development in Estonia, but also looks at support for research, improvement of skills for public sector workers and financial support. One of the actions includes making basic Kratt components and “tools” available for reuse. This involves establishing a collaborative network in the public sector, and creating and disseminating instructional material for launching and conducting research projects. It also emphasizes the need to showcase the opportunities AI brings and introduce specific



projects in different networks and formats. Furthermore, it suggests that, at least at the ministerial level, roles for chief data officers should be established. It also emphasizes the need to launch in-depth data management workshops and create support for data audits in institutions.

The legal analysis in the report concludes that there is no need for substantial changes to the basics of the legal system and no need for a specific “Kratt act”. However, the report emphasizes that current legislation needs several changes. For example, there needs to be clarity around the issue of liability related to the implementation of Kratt. There is legislation to regulate automated decision-making and profiling. A specific institution can only use ADM or profiling by using a specific legal act, a signed contract, or if a person gives consent.

The report concludes (Majandus- ja Kommunikatsiooni ministeerium/Riigikantselei 2019b) that Estonia should start piloting Kratt projects to gather initial feedback and experience to better inform long-term planning. This is seen as an important step as pilot projects will help the government better understand the possible benefits and risks of using Kratt while also helping to form future strategies. The strategy also emphasizes the need to raise awareness, increase research and development activities, encourage innovation, and to foster the implementation of AI, or Kratt, in the business sector.

Although the Kratt project specifically does not mention ADM, many of the activities financed and supported through the AI strategy for 2019-2021, are also important in relation to ADM.

Moreover, there is a plan to create a bureaucratic Kratt, or artificial solution, as a personal digital assistant. The idea behind this AI solution coincides largely with the development of the Aurora concept in Finland. It would, therefore, be reasonable to cooperate between these developments – or at the very least, exchange ideas and experiences and to develop technical solutions together, if possible.

Among other things, this would enable interoperability of the state Kratt ecosystems of Estonia and Finland so that the virtual assistants, or Kratt sessions, could also function as cross-border services. The common digital space shared between Estonia and Finland is one of the priorities of the digital policy of Estonia and this could also continue in the Kratt age. However, this is still in the early stages of development.

As of October 2019, there were at least 23 AI solutions deployed in the Estonian public sector and that number was set to increase to at least 50 by 2020. The examples already in use show that the scope of the use of Kratt, or AI, has some limits. Kratt solutions are not used for making decisions on their own yet, but they are used to help humans make faster and better decisions.

## **/ Elements of Artificial Intelligence web course**

Estonia’s AI strategy emphasizes the need for a broader understanding of AI. Therefore, one of the aims the expert group posed in the AI strategy document was to develop a web-based training program. The Finnish “Elements of Artificial Intelligence” online course, which was developed by the Department of Computer Science at the University of Helsinki in partnership with the technology company Reaktor, was used as the basis for a suitable online course. In the middle of November 2019, an Estonian language version of the “Elements of AI” (Tasuta veebipõhine sissejuhatus tehisintellekti mitte-ekspertidele o. j.) online course was launched. The Estonian version of the course was developed by Tallinn University of Technology (TalTech). According to the strategy, the main aim of the course is to increase the interest and receptivity of AI in Estonian businesses. On the course webpage, it is also emphasized that taking part in the course will help citizens see AI as a friend, something that will offer support in life and not something to be feared.

The course introduces basic concepts and applications of AI and machine learning to increase knowledge and equip people with a good understanding so that they can participate in the debate on the subject.

The societal implications of AI, such as algorithmic bias and de-anonymization, are also introduced in the course. This helps underscore the need for policies and regulations to guarantee that society can adapt well to the changes the increasing use of AI brings. The course is free for anyone to take and close to 220,000 people have already completed it.

The aim (Liive 2019) is to spread knowledge about AI to at least 1% of Estonian citizens.

The former president of Estonia, Toomas-Hendrik Ilves, is the course patron. The course has more than 30 partners from different businesses and organizations.

To promote the course among Estonians, Tallinn University of Technology started a campaign – #Alväljakutse (#Alchallenge). More than 1200 people took the challenge when the course was first launched. The challenge to complete the course has been taken up by nearly 35 companies and organizations in Estonia who want to educate their employees about AI.

### **/ CITIS – Center of IT Impact Studies**

The Center of IT Impact Studies (CITIS) was established in 2015 and it is an integrated teaching and research center at the Johan Skytte Institute of Political Studies at the University of Tartu, Estonia. Their main aim is to use big data generated by various Estonian public services (for example, e-health, digital ID, e-residency, etc.) to estimate the impact those services have economically, politically, and socially (Center of IT Impact Studies (CITIS), n.d.). CITIS teams also contribute to the development of the e-state directly by creating and testing prototypes for new e-services, based on the cross-usage of existing data. The CITIS team, together with Nortal, one of Estonia's largest developers, are building a set of profiling and segmentation policy tools for the Estonian Unemployment Insurance Fund (EUIF) (Profiling and policy evaluation tools for EUIF (2018-2020) 2019).

### **/ STACC—Software Technology and Applications Competence Center**

Established in 2009, [STACC](#) is the leading machine learning and data science competence center in Estonia and it also develops artificial intelligence solutions. Its mission is to conduct high-level applied research in the field of data science and machine learning in cooperation with a consortium of scientific and industrial partners. STACC was founded by two universities – the University of Tartu and Tallinn University of Technology and by several leading IT companies: Cybernetica AS, Reach-U AS, Nortal AS, CGI Eesti AS, and Quretec OÜ. One purpose established in the [Estonian national AI action program](#) is to direct and support STACC activities related to AI. STACC is one of the six technology competence centers currently in operation and it is financed by Enterprise Estonia. The centers aim to motivate companies to develop innovative products through cooperation with research institutions, thereby bridging the gap between scientific and economic innovation.

### **/ GovAiLab – Government Technology and Artificial Intelligence Lab**

The [Government Technology and Artificial Intelligence Lab](#) (GovAILab) was opened on 18 November 2019 at Tallinn University of Technology (TalTech). Its main focus is to advise and help governmental institutions find new technological solutions, but it also functions as a place to pilot public sector ADM or AI projects to help decide how to proceed. The lab is seen as a partner for the organization's leaders, and it helps specialists, who are accountable for the development of technical AI solutions, to make better choices and to experiment with technology.

### **/ Political debates on aspects of automation – Civil society and academia**

#### **/ The Foundation for Future Technologies**

The Foundation for Future Technologies (FFT) (Tulevik 2018) is a non-governmental organization that aims to help maximize the positive impact of future technologies on Estonian society and to make their deployment a considered and widely understood decision (Tulevik 2018). The Foundation for Future Technologies member Kaspar Kruup spoke at the TEDxTTÜ event on 19 March 2016 about AI. The foundation has also organized panels on topics such as, the future of work, virtual reality, and AI at the 2016 Opinion Festival. In addition, foundation members also organized an AI Think Club in which 75 people took part and they maintain a Facebook Group on Artificial Intelligence and SINA (Tehisintellekt ja SINA FB page o. J.). The Facebook group points to some foreign news articles and discussions but there are no Estonia specific discussions. The page is followed by 375 people. The foundation no longer takes an active part in AI or ADM related discussions.

Two members of the foundation (Allan Aksim and Pirkko Kõnsa) led the expert group on self-driving cars (Isejuhtiva sõidukite ajastu algus – ekspertrühma lõppraport 2018) which was initiated by the Estonian Ministry of Economy and Communication and the Government Office. However, the work of this expert group has now ended.

#### **/ Estonian Human Rights Center**

The [Estonian Human Rights Center](#) is an independent non-governmental organization that stands for human rights in Estonia. In 2019, the center published a document predominantly aimed at mapping the human rights situation in relation to the information society and technology (Käsper/Rajavee 2019). Authors carried out focus group interviews

and discussions with experts, from both private and public sector organizations, third sector activists, and a journalist, who had come into contact with the topics covered in the report. Several other documents were also analyzed for this purpose. The report considers issues such as profiling by the state, GDPR Implementation, and freedom of expression on the Internet. The report indicates several risks and interviewed experts to determine the possibility of discrimination through profiling and privacy concerns and whether or not proactive services should be considered. The authors (Käsper/Rajavee 2019) suggested developing visual solutions that would make it easier for people to understand which state services use their data and how they use it, but they also emphasized the need for further analysis about how ADM impacts human rights.

## **/ Estonian Institute of Human Rights**

The Estonian Institute of Human Rights was established in 1992 and it aims to collect, systematize, analyze, and disseminate information on individual rights for public awareness. On 10 December 2019, the institute organized a human rights conference in Estonia on the topic of “Values and Interests In parallel worlds” (Inimõiguste aastakonverents 2019. 2019). One of the panels discussed how technological developments override human rights and the possible impacts and risks associated with ADM, among other things.

## **/ A study on how to best use ADM in e-governance**

In 2019, the Estonian Research Council signed a contract with a research consortium to look for opportunities where and how ADM could be used in developing public services (Teadlased asuvad uurima tehisintellekti kasutamise võimalusi avalike teenuste pakkumisel – Sihtasutus Eesti Teadusagentuur 2019). This study is funded by the RITA program “Opportunities for the Use of Artificial Intelligence in E-Governance”, jointly conducted by the University of Tartu, Tallinn University of Technology, Cybernetica AS, and STACC and will last for two years. The research results will be used by the Ministry of Economic Affairs and Communications and the State Chancellery for the development of e-government.

The overall aim of RITA is to increase the role of the state in the strategic direction of science and the capacity of research and development institutions to carry out socially relevant research.

The Estonian public sector is smaller than average, and the services provided by the state and local governments must be equivalent. At the moment, the provision of services in Estonia is not optimally organized. One way to improve the quality of services without increasing the number of officials is to implement artificial intelligence, or ‘chats’. The results of the survey are expected by the end of 2021. The budget for the study is 805,212 euros, which comes from the European Regional Development Fund and the state budget of Estonia through the RITA program.

# **Key takeaways**

In Estonia, several solutions that use ADM have already been developed and are planned under the Kratt project. This report, and the different organizations mentioned above, emphasize several issues, such as possible discrimination related to data or ethics, related to AI and ADM which should be taken into account. The overall public discussion tends to concentrate on the possible opportunities that developed solutions and AI will bring. Risks and problems are acknowledged, but they are not discussed widely.

Several of the services could be described as proactive as their main aim is to monitor or select citizens who may require additional counseling or other kinds of help. There is one particular proactive service already offered for families, and several others are under development. Services like these can be seen as positive as citizens are provided with the services, they are entitled to without the need to apply for them. Therefore, some of the groups or individuals who would have a right to those services, but who may not be digitally literate, are still offered the services proactively.

In some cases, it can be incomprehensible for a person to understand how the systems reach specific decisions. Although, for some time now, Estonians have had an opportunity to check on the [eesti.ee](https://eesti.ee) webpage to see which institutions have accessed their data and for what reasons. However, as the Human Rights Center report emphasizes, clearer, visual solutions should be developed. The solution used right now does not offer a clear understanding of what profiling is done by the state, which data is collected, how it is used, and for what purpose. Moreover, it is not always clear if data is used as a part of an automatic process or viewed by an official.

# References:

Belkin, Märt (2019): Suured poeketid näotuvastusega iseteeninduskassast: tehnoloogia pole veel päris valmis, in: Geenius portaal, [online] <https://raha.geenius.ee/rubriik/uudis/suured-poeketid-naotuvastusega-iseteeninduskassast-tehnoloogia-pole-veel-paris-valmis/> [02.12.2019].

Majandus- ja kommunikatsiooniministeerium / Riigikantselei (2019a): Eesti riiklik tehisintellekti alane tegevuskava 2019-2021, in: Majandus-ja kommunikatsiooniministeeriumi kodulehekül, [online] [https://www.mkm.ee/sites/default/files/eesti\\_kratikava\\_juuli2019.pdf](https://www.mkm.ee/sites/default/files/eesti_kratikava_juuli2019.pdf) [01.12.19 n. Chr.].

Majandus- ja Kommunikatsiooni ministeerium / Riigikantselei (2019b): Eesti tehisintellekti kasutuselevõtu eksperidirühma aruanne, in: Riigikantselei kodulehekül, [online] [https://www.riigikantselei.ee/sites/default/files/riigikantselei/strateegiaburoo/eesti\\_tehisintellekti\\_kasutuselevotu\\_eksperdiruhma\\_aruanne.pdf](https://www.riigikantselei.ee/sites/default/files/riigikantselei/strateegiaburoo/eesti_tehisintellekti_kasutuselevotu_eksperdiruhma_aruanne.pdf) [01.12.19 n. Chr.].

Eestis töötati välja tehisintellekti sugemetega talendiotsingumootor Skillific (2016): in: Personaliuudised, [online] <https://www.personaliuudised.ee/uudised/2016/03/07/eestis-tootati-valja-tehisintellekti-sugemetega-talendiotsingumootor-skillific> [02.12.2019].

eHealth Network (2020): Mobile applications to support contact tracing in the EU's fight against COVID-19. Common EU Toolbox for Member States. Version 1.0, in: eHealth Network, [online] [https://ec.europa.eu/health/sites/health/files/ehealth/docs/covid-19\\_apps\\_en.pdf](https://ec.europa.eu/health/sites/health/files/ehealth/docs/covid-19_apps_en.pdf) [10.05.2020].

Enhanced Care Management: Improving Health for High Need, High Risk Patients in Estonia (2017): in: Haigekassa kodulehekül, [online] <https://www.haigekassa.ee/sites/default/files/enhanced-care-management.pdf> [01.12.2019].

GovAiLab (o. J.): in: govailab.com, [online] <https://govailab.com/> [02.12.2019].

Isejuhtiva sõidukite ajastu algus – ekspertrühma lõppraport (2018): in: Riigikantselei kodulehekül, [online] [https://www.riigikantselei.ee/sites/default/files/riigikantselei/strateegiaburoo/isejuhtivad\\_loppraport.pdf?fbclid=IwAR27W\\_f1dJfTsBmzxncH-VwBfrjkepAluOkuZTHhpS\\_2rUOgJRSO-XvfQ](https://www.riigikantselei.ee/sites/default/files/riigikantselei/strateegiaburoo/isejuhtivad_loppraport.pdf?fbclid=IwAR27W_f1dJfTsBmzxncH-VwBfrjkepAluOkuZTHhpS_2rUOgJRSO-XvfQ) [01.12.2019].

Kasutuslood | Krattide veebileht (o. J.): in: Kratt project, [online] <https://www.kratid.ee/kasutuslood> [02.12.2019].

Keila Tarbijate Ühistu võttis kasutusele Euroopas ainulaadse näotuvastusega iseteeninduskassa – StrongPoint (o. J.): in: Strongpoint kodulehekül, [online] <https://www.strongpoint.com/ee/uudised/keila-tarbijate-uhistu-vottis-kasutusele-euroopas-ainulaadse-naotuvastusega-iseteeninduskassa/> [02.12.2019].

Kõik muutub automaatseks: töötukassa IT-osakond ei istu päevad otsa mustades ülikondades, tõstes tuumalt pabereid ühest hunnikust teise (2019): in: Rahageenius.ee, [online] <https://raha.geenius.ee/blogi/tootukassa-blogi/koik-muutub-automaatseks-tootukassa-it-osakond-ei-istu-paevad-otsa-mustades-ulikondades-tostes-tuumalt-pabereid-uhest-hunnikust-teise/> [02.12.2019].

Krusten, Mari (2019): National AI strategy for 2019-2021 gets a kick-off – e-Estonia, in: e-estonia, [online] <https://e-estonia.com/national-ai-strategy/> [02.12.2019].

Laks, Liina (2020): Nortal hakkab sakslastega koroonareisiäppi arendama, in: Tehnika, [online] [https://tehnika.postimees.ee/6970668/nortal-hakkab-sakslastega-koroonareisiappi-arendama?fbclid=IwAR0uy74fVucnHddajANoxNg0I3\\_FEy07KBD4o-KzW7Gggaj8Nuz4vOtpHXM](https://tehnika.postimees.ee/6970668/nortal-hakkab-sakslastega-koroonareisiappi-arendama?fbclid=IwAR0uy74fVucnHddajANoxNg0I3_FEy07KBD4o-KzW7Gggaj8Nuz4vOtpHXM) [15.05.2020].

Liive, Ronald (2019): Täna alustas tasuta eestikeelne tehisintellekti kursus, in: Geenius portaal, [online] <https://digi.geenius.ee/rubriik/teadus-ja-tulevik/tanasest-alustas-tasuta-eestikeelne-tehisintellekti-kursus/> [02.12.2019].

Liive, Ronald (2020): Eestis hakatakse arendama koroonaviirusesse nakatunute kontaktijälgimise äppi, in: Digi-geenius, [online] <https://digi.geenius.ee/rubriik/uudis/konsortsium-hakkab-riigile-koroonaviiruse-kontaktijalgimise-appi-arendama/> [01.05.2020].

Liive, Ronald (2020a): Eesti koroonäppi saab hakata kasutama augustis, aga selline hakkab see välja nägema, in: Digigeenius, [online] <https://digi.geenius.ee/rubriik/uudis/eesti-koroonappi-saab-hakata-kasutama-augustis-aga-selline-hakkab-see-valja-nagama/?nocache=1> [10.07.2020].

Mandri, Jan-Matthias (2019): Kohtunikud saavad robotabilised: riik otsib võimalusi kohtusüsteemis tehisintellekti rakendamiseks, in: Delfi Forte, [online] <https://forte.delfi.ee/news/digi/kohtunikud-saavad-robotabilised-riik-otsib-voimalusi-kohtususteemis-tehisintellekti-rakendamiseks?id=85777829> [02.12.2019].

Maran, Kaur (2020): Eesti e-riik valmistub WHO kaudu maailma vallutama, in: Tänapäev, [online] <https://leht.postimees.ee/6973751/eesti-e-riik-valmistub-who-kaudu-maailma-vallutama> [18.05.2020].

Millal teeb töötukassa Teie suhtes automaatse otsuse | Töötukassa (o. J.): in: Eesti Töötukassa, [online] <https://www.tootukassa.ee/content/tootukassast/millal-teeb-tootukassa-teie-suhtes-automaatse-otsuse> [02.12.2019].

Mis on noortegarantii tugisüsteem? – Tööelu.ee (o. J.): in: Tööelu, [online] <https://www.tooelu.ee/et/Tooturule-sisenejale/Noortegarantii-tugisysteem/Kohalikule-omavalitsusele/Mis-on-noortegarantii-tugisysteem> [02.12.2019].

Niiler, Eric (2019): Can AI Be a Fair Judge in Court? Estonia Thinks So, in: WIRED, [online] <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/> [02.12.2019].

Noortegarantii | Sotsiaalministeerium (o. J.): in: Euroopa Liidu noortegarantii, [online] <https://www.sm.ee/et/noortegarantii> [02.12.2019].

Raudla, Merly (2020): Töötukassa automatiseeritud infosüsteem otsustab raha jagamise sekunditega, in: Virumaa Teataja, [online] <https://virumaateataja.postimees.ee/6904529/tootukassa-automatiseeritud-infosusteem-otsustab-raha-jagamise-sekunditega> [10.08.2020].

Salter, Philip (2020): Lessons For The World On How Estonia's Digital State Is Coping With Coronavirus, in: Forbes, [online] [https://www.forbes.com/sites/philipsalter/2020/05/01/lessons-for-the-world-on-how-estonias-digital-state-is-coping-with-coronavirus/?fbclid=IwAR0YSuOsHYLisb\\_6z40ezCeFO4tDuccsHCON3wpVS2NrOffJPySb3cEtPA#24e6fb895001](https://www.forbes.com/sites/philipsalter/2020/05/01/lessons-for-the-world-on-how-estonias-digital-state-is-coping-with-coronavirus/?fbclid=IwAR0YSuOsHYLisb_6z40ezCeFO4tDuccsHCON3wpVS2NrOffJPySb3cEtPA#24e6fb895001) [14.05.2020].

Skillific kasutuslugu | Krattide veebileht (o. J.): in: Krati projekt, [online] <https://www.kratid.ee/skillific-kasutuslugu> [02.12.2019].

Tallin University of Technology (o. J.): Accept the #AIchallenge, in: TalTech webpage, [online] <https://www.ttu.ee/accept-the-aichallenge> [02.12.2019].

Tasuta veebipõhine sissejuhatus tehisintellekti mitte-ekspertidele (o. J.): in: Elements of AI, [online] <https://www.elementsofai.ee/> [02.12.2019].

Teadlased asuvad uurima tehisintellekti kasutamise võimalusi avalike teenuste pakkumisel – Sihtasutus Eesti Teadusagentuur (2019): in: Eesti Teadusagentuur, [online] <https://www.etag.ee/teadlased-asuvad-uurima-tehisintellekti-kasutamise-voimalusi-avalike-teenuste-pakkumisel/> [02.12.2019].

Tehisintellekt ja SINA FB page (o. J.): in: Tehisintellekt ja SINA FB page, [online] <https://www.facebook.com/tehisintellekt/> [01.12.19 n. Chr.].

Tulevik (2018): Mis on Tuleviku Tehnoloogiate Sihtasutus?, in: Tuleviku Tehnoloogiate Sihtasutuse tutvustus, [online] <https://medium.com/@Tulevik/mis-on-tuleviku-tehnoloogiate-sihtasutus-76410369d862> [02.12.2019].

Tuul, Harry (2020): Eesti firma teeb koostööks sakslaste ja araablastega koroonaaajastu lennupassi, in: Digigeenius, [online] <https://digi.geenius.ee/rubriik/uudis/eesti-firma-teeb-koostooos-sakslaste-ja-araablastega-koroonaaajastu-lennupassi/> [14.05.2020].

Veltman, Madis (2019): EKSPERIMENT | Müüjate hirm on kohal! Kas tuliuus näotuvastusega iseteeninduskassa müüb lapsele suitsu?, in: Delfi Kasulik, [online] <https://kasulik.delfi.ee/news/uudised/eksperiment-muujate-hirm-on-kohal-kas-tuliuus-naotuvastusega-iseteeninduskassa-muub-lapsele-suitsu?id=87848345> [02.12.2019].

Profiling and policy evaluation tools for EUIF (2018-2020) (2019): in: Center of IT Impact Studies (CITIS), [online] <http://citis.ut.ee/articles/r-and-d-projects/profiling-and-policy-evaluation-tools-for-euif-2018-2020> [04.02.2020].

Inimõiguste aastakonverents 2019. (2019): in: Inimõiguste aastakonverents. Väärtused ja huvid paralleelmaailmas. In Memoriam Mart Nutt., [online] <http://aastakonverents.humanrightsestonia.ee/2019-inimoguste-aastakonverents/2019-programm/> [02.02.2020].

Käsper, Kari / Liina Rajavee (2019): INIMÕIGUSED, INFOÜHISKOND JA EESTI: ESIALGNE KAARDISTUS, in: Inimõiguste Keskuse aruanne, 2019, [online] <https://humanrights.ee/app/uploads/2019/12/EIK-kaardistamine.pdf>.

Sotsiaalkindlustusamet avas esimese proaktiivse avaliku teenuse | Sotsiaalkindlustusamet (2019): in: Sotsiaalkindlustusamet, [online] <https://www.sotsiaalkindlustusamet.ee/et/uudised/sotsiaalkindlustusamet-avas-esimese-proaktiivse-avaliku-teenuse> [02.02.2020].





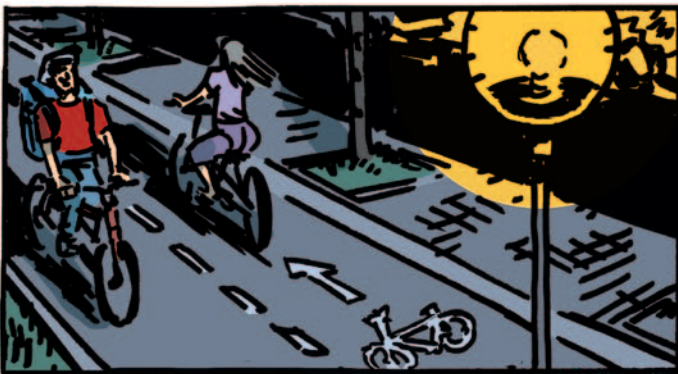
BELGIUM  
DENMARK  
ESTONIA  
**FINLAND**  
**STORY**

PAGE 80

**RESEARCH**

PAGE 84

FRANCE  
GERMANY  
GREECE  
ITALY  
NETHERLANDS  
POLAND  
PORTUGAL  
SLOVENIA  
SPAIN  
SWEDEN  
SWITZERLAND  
UNITED KINGDOM



Find out more in the story on the next page.

# // How to **unionize** when your boss is an algorithm, and you're self- employed

**A group of Finnish couriers launched the Justice4Couriers campaign in 2018. Although they are technically self-employed, they must obey the whims of their platform's algorithm. They are fighting back.**

By Aleksi Knuutila



“You like to cycle, so it must be a good job for you!” This is the response that Miran Hamidulla, a Finnish student of political history and an activist, often hears when he mentions his work. Hamidulla is a bike courier for the food delivery platform Foodora.

In a social media post, Hamidulla explained that he did indeed enjoy his job navigating the streets of Helsinki. His love of cycling, however, does not make up for the exceptionally poor conditions under which Foodora couriers have to cycle for a living.

Cycling in the central areas of Helsinki can be risky, and if Hamidulla had an accident and was to work the following day, he would not be entitled to any sick leave. For whatever reason, if Foodora came to see him as unreliable and terminated his contract or reduced his pay, Hamidulla would have no protection.

Foodora couriers have no guarantee that they can maintain the level of income they receive from one week to the next. Legally speaking, they are not even employees of Foodora, but rather self-employed entrepreneurs with their own small businesses. This excludes them from the basic rights most workers in Finland would take for granted, such as sick leave and social insurance paid by the employer.

Hamidulla works for about 20-25 hours every week. In an interview, he explained that he earned about 750 euros per month after subtracting his taxes and insurance costs (the median monthly income for full-time work in Finland is about 2200 euros after taxes). He must also pay for the maintenance of his bike. He says that the work arrangement with Foodora is passable for him because he has other sources of income.

Hamidulla is a spokesperson for the Justice4Couriers campaign.

Just taking on this role was no easy decision. He was able to do it because he does not rely entirely on Foodora for his livelihood. Not everybody is as lucky. Recent immigrants to Finland fear to speak out, as they might not speak Finnish and often work for several food delivery companies simultaneously, for lack of better employment opportunities. They are most vulnerable to the consequences of the algorithmic management of work.

**RECENT IMMIGRANTS  
TO FINLAND FEAR TO SPEAK OUT. THEY  
ARE MOST VULNERABLE TO THE  
CONSEQUENCES OF THE ALGORITHMIC  
MANAGEMENT OF WORK.**

## / Food delivery is the beginning

Foodora is a platform company, which acts as an intermediary between restaurants, consumers, and couriers. As an intermediary, its business model is relatively new to Finland. Foodora, a subsidiary of the Berlin-based Delivery Hero, came to Finland in 2014, the same year the Finnish food delivery platform Wolt, Foodora's main competitor, launched its service.

Like the day laborers of yore, platform companies hire their contractors on an as-needed basis. The assignment of individual tasks, as well as the allocation of work shifts between workers, is largely decided by algorithms. They predict the location and time of demand and decide which individuals are best placed to do the work. The job is marketed as flexible and accommodating to would-be couriers, but flexibility cuts both ways. Platforms automatically reduce the size of their workforce when there is less demand.

This way of working is being trialed in other sectors in Finland, such as taxi drivers and home cleaners. Also, in Helsinki, for example, newspapers are delivered to homes each morning with labor organized through apps, with similarly precarious work arrangements.

People whose work is directed by algorithms usually have self-employed status as independent sub-contractors. This allows platform companies to circumvent employment regulations and to have couriers themselves bear some of the costs and risks, such as sick leave.

According to Finnish legislation, a person has to be classified as an employee, with corresponding rights, if they work under the direction and supervision of an employer. Finnish platform companies have made the case that this condition is not met under app-mediated, flexible work contracts. For

instance, Foodora couriers are not required to take any specific shift or to work a minimum number of hours. Thanks to algorithmic organization, platforms can be less prescriptive than regular

employers about shifts and still have enough workers to deliver the service. From the point of view of couriers, however, the arrangement can be a far cry from the freedom some associate with being self-employed.

# PEOPLE WHOSE WORK IS DIRECTED BY ALGORITHMS USUALLY HAVE SELF- EMPLOYED STATUS AS INDEPENDENT SUB- CONTRACTORS.

## / When your boss is an algorithm

Every Wednesday, Foodora offers its couriers a choice of work shifts. The app displays a list of two- to four- hour slots with an assigned location that the couriers can choose. Getting the right shifts is important not only for one's personal life but also because some shifts are more lucrative than others. Lunch shifts, for instance, are especially prized, because every delivery counts on top of the courier's hourly rate.

Foodora workers are not all equal when it comes to choosing working hours. Shifts are released in batches. The first batch, sent out on Wednesday morning, is offered only to the couriers rated as most effective by the platform. Foodora operates a scoring algorithm, and the 10 percent who make up the highest-scoring couriers get to pick first. The next workers in line, with slightly lower scores, have to wait until Wednesday afternoon to choose from the remaining shifts.

Those in the fourth and last batch make their choice in the evening. At this point, there are often no shifts left to select. According to Hamidulla, shifts may run out even during the second batch, particularly in calm periods such as the summer. One way in which couriers circumvent this system is by operating an unofficial black market, where couriers with a higher score sell shifts to others, using the functionality of the app for assigning a replacement for oneself.

When there are fewer shifts to go around, getting a lower score on the Foodora algorithm is not unlike being suspended from work. What makes this threat particularly stressful is that the details of how Foodora's scoring algorithm works are unknown to the couriers.

Couriers are required to install Foodora's app, called Roadrunner, in order to receive directions for their work. The app collects a large range of information that it could take into account. Foodora tracks how long it takes for couriers to accept gigs and how many minutes they spend on breaks. Should a courier take too long to accept the next gig, the Roadrunner app makes a beeping sound, familiar from the Looney Tunes animation character of the same name. A human operator will send an online chat message if the breaks become too long.

Foodora does not publish the details of its scoring algorithm. Between work days, couriers receive emails telling them how their score compares to that of other couriers, encouraging them to do better. They are told what batch they are in and how many times they have been late or have not clocked in for their shifts. [According to an investigation by the Finnish newspaper, Helsingin Sanomat](#), it is these variables, the consistency of showing up to work, that affects one's score most.

Foodora does offer an opportunity for couriers to cancel a work shift, provided they have a reason Foodora considers acceptable. If couriers cancel a shift less than 24 hours in advance, they have to find a replacement themselves or risk losing points. A single missed shift can push them into a lower batch.

The investigation by Helsingin Sanomat included interviews from couriers who showed up to work the day following a serious accident. They were afraid that spending the day recovering would impact their scores. In other cases, bugs in the Roadrunner app sent their scores plummeting. Because Foodora employs only a few staff in Finland (85 people in 2017), it can be hard to get a score corrected by a human.



## / Couriers working together to improve their conditions

The Bongo Cat was a short-lived meme in 2018. One of its manifestations, Bongo Cat 4 Couriers, [spread on Finnish Facebook and Instagram](#) in October 2018. It encouraged people to give Foodora's app one-star reviews to show support for the fight for couriers' rights.

The meme was created by Justice4Couriers, a campaign group of couriers seeking to improve their working conditions. The group formed in September 2018 in response to a unilateral reduction in pay by Foodora and Wolt. The couriers participating in the campaign formulated five demands, including repealing the pay cuts, offering couriers the possibility of an employment contract, and bringing transparency to the functioning of the platforms' scoring algorithms.

Justice4Couriers used many tactics, including calling for consumer boycotts of the platforms on strategic occasions, to increase the leverage that couriers have and to force the platforms to negotiate. The campaign pushed the issue into the limelight through demonstrations, as well as by distributing flyers in Finnish cities. The campaign was featured in national news, for example, in the [Iltalehti](#) and [Iltasanomat](#) newspapers.

So far, only one of Justice4Couriers' demands has been met. Foodora now provides a physical space where couriers can take a break between and after shifts. As a reaction to the COVID-19 pandemic, Foodora gave its couriers 12 euros to buy protective equipment, which was much less than what Justice4Couriers [argued for](#).

However, the campaign has successfully pushed the issue of platform workers onto the political agenda. Established Finnish unions, such as the Service Union United (PAM), offered training to Justice4Couriers' members and lobbied for changes to the law. The new Finnish government, which took office in December 2019, plans to review the status of platform workers and close loopholes that allow employers to avoid their responsibilities. Similar updates to the legal status of platform workers have taken place or are under review in [California](#) and [Spain](#).

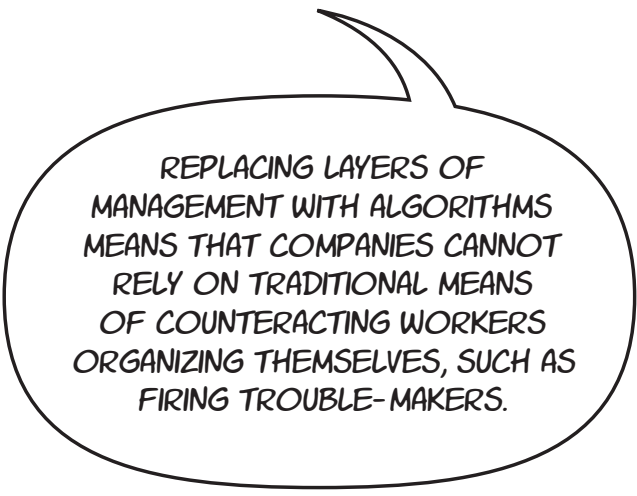
The algorithmic organization of work, with flexible working hours and location, poses challenges to collaboration between workers. It can also open up surprising new opportunities. Workers for Deliveroo, another food delivery

platform, went on [strike in Brighton \(UK\) in 2017](#), following impromptu meetings of couriers. The platform's algorithm had told large numbers of them to wait by a central square. Callum Cant, one of the participants in the Brighton strike and a scholar, has suggested that replacing layers of management with algorithms meant that companies could not rely on traditional means of counteracting workers organizing themselves, such as firing trouble-makers.

Justice4Couriers is part of an international network of groups experimenting with new ways of improving the lives of workers in the platform economy. While workers' rights are important, the groups are also looking further ahead. One of Justice4Couriers' sister organizations is RidersXDerechos (Riders 4 Rights) in Barcelona. They are building their own algorithms, apps, and technical infrastructure, that will be collectively owned and run by the couriers themselves.

Justice4Couriers is not questioning the application of algorithms as such. In a [blog post](#), they mention that "instead of making workers compete with one another over shifts and orders, the applications could be designed to enable greater co-operation between couriers". Foodora did not answer a request for comment.

Algorithms are currently deployed by platforms that seek to maximize profit by shifting risk to their workers. Other algorithms are possible, which could improve the work conditions for couriers and others, and give them greater control over their work lives.



REPLACING LAYERS OF  
MANAGEMENT WITH ALGORITHMS  
MEANS THAT COMPANIES CANNOT  
RELY ON TRADITIONAL MEANS  
OF COUNTERACTING WORKERS  
ORGANIZING THEMSELVES, SUCH AS  
FIRING TROUBLE-MAKERS.

# Research

By Minna Ruckenstein & Tuukka Lehtiniemi

## Contextualization

In Finland, both public and private actors from different sectors of society are involved in the development of automation systems and put forward plans to implement them. At the central government level, attempts to specify the debate around AI and ADM are ongoing. This debate includes technical, legal, ethical, and societal perspectives. The technical and legal perspectives tend to dominate the debate, but the ethical perspective is currently gaining more attention. The societal perspective remains marginal and it needs to be reintroduced time and again into the debate by emphasizing the socio-technical nature of ADM systems. From this perspective, ADM is not a stand-alone technology, but a framework that encompasses the decision-making model and the political, economic, and organizational context of its use.

Driven by pressures to do more with fewer funds, Finnish authorities across the board are employing and planning to deploy automated procedures in their decision-making. Yet, legal amendments and clarifications are needed to maintain and proceed with automation projects. Various public institutions are currently under scrutiny concerning the compatibility of automation with existing laws. Another area of concern is the fulfillment of good governance. Here, the issue of full automation – that is, the use of ADM systems without direct human involvement in the decision process – is particularly contested.

Commercial actors are an integral part of public sector initiatives, raising questions about the aims and outcomes of public-private partnerships. For example, in the care sector, companies typically offer algorithmic models to spot correlations between individual risk factors. The implementation of such models could fundamentally change the logic of social and healthcare systems. The cases discussed here call for an overdue societal debate around ADM.

## A catalog of ADM cases

### / Repairing automated content moderation

Utopia Analytics, a text analytics company specializing in automating content moderation, has been moderating online conversations on one of Finland's largest online forums (Suomi24) since 2017. The company's flagship service is called Utopia AI Moderator and, according to their website, it promises to deliver "the finest machine learning services on the planet" (Utopia Analytics, 2020).

Researchers at the University of Helsinki followed the implementation of the Utopia AI Moderator tool as part of the content moderation process of Suomi24. The Utopia AI Moderator is a service that is tailor-made in collaboration with the customer, rather than a stand-alone product. As with all of its customers, Utopia provides Suomi24 with a machine learning model that has been built using the data from their historical man-made content removal decisions. The model can be adjusted over time, for instance, it is possible to alter the machine's sensitivity to the messages that are posted.

According to Suomi24, approximately 31,000 new threads with 315,000 comments are posted on the forum monthly. Suomi24 automatically sends moderation requests to Utopia, and their AI model either accepts or rejects the content. If there is a borderline case, or something completely new compared to the existing data, the model may request a human moderator to view the post. The AI model is kept up-to-date by using these new samples generated by human moderators.

Ultimately, the automated moderation process is a combination of Utopia's tool, Suomi24's own technical systems and human work.

Soon after the implementation of the Utopia system, users of the online forum discovered that messages from conversations started disappearing randomly. Some writers on the forum started to craft conspiracy theories about attacks on free speech. Others paid attention to the fact that similar messages were disappearing, for instance, those containing links. Writers concluded that the removals were machine-generated and referred to a dysfunctional and weird AI that was sabotaging their discussions.

Instead, it is probable that the removals were generated by bugs, or inconsistencies in the technical system of the online forum. The important point here is that what looked like a problem generated by the AI model was, in fact, an everyday information technology failure.

After realizing what was happening on the forum, the Suomi24 team openly discussed false removal decisions with the users of their site, telling them about their moderation practices and the division of labor between humans and the machine (Suomi24, 2019). The company representatives admitted that they had returned incorrectly removed messages to discussions. In the course of explaining their moderation practices, they also expressed a view on the tasks and the limitations of AI. As they mentioned in a blog post: “as with any AI, it is not (at least for now) perfect, but it is still an indispensable assistance in a service of this scale.” The Utopia AI Moderator model continues to remove a large portion of offensive posts, as well as posts that do not comply with the forum’s standards, while at the same time the machine is also continually learning to remove content more accurately. However, the AI model never decides what is a proper or improper conversation, it only mirrors decisions human moderators have already made. Over time, it will become better at detecting things that have occurred before, but there will always be new decisions to make which need human judgement.

The collaboration between Utopia Analytics and Suomi24 reminds us of the way in which services develop gradually over time, how things might break down, and how essential it is to do repair work during the implementation process. Following recent upgrades, the technical system

of Suomi24 has improved and the moderation process is expected to work much more smoothly in the future. Yet, machine decisions are only as good as the data that the machine uses. Humans – in this case, content moderators – continue to have a central role in procuring the data used for training the machine.

## **/ Increasing the efficiency of permit processes at the Finnish Immigration Service**

Migri (Finnish Immigration Service) is a government agency responsible for making decisions in matters related to immigration, Finnish citizenship, asylum and refugee status. In 2019, Migri made a total of 104,000 decisions concerning these matters.

The long length of time it takes to process permits causes anguish and uncertainty for the people subject to Migri’s decisions. This is the case particularly for those seeking asylum and refugee status. Other problems that repeatedly come up in public discussions include the residency permits of employees hired from abroad by Finnish firms (Keskuskauppakamari 2019), as well as international students who miss the start of their studies due to delays in permit processing (Helsingin Sanomat 2019c). The agency is currently exploring automation in order to speed up permit processes and make its operations more cost-effective (Migri 2019a). According to an IT consultancy spokesperson, the ongoing “smart digital agency” project aims, to provide better services, improve internal processes and reduce costs (Accenture 2016).

Given the agency’s lack of resources to handle an increasing amount of decisions (Migri 2018a), economic efficiency means reducing the need for human participation at different stages of the permit process. Against this background, automation is an attractive way to decrease human service interactions, as well as to focus the available human resources on those parts of the permit process where they are necessary.

Migri’s aim is to reduce the need to contact officials in person so that, by the end of 2020, 90% of all interactions take place via self-service channels. A practical example of this

**THE COLLABORATION BETWEEN UTOPIA ANALYTICS AND SUOMI24 REMINDS US OF THE WAY IN WHICH SERVICES DEVELOP GRADUALLY OVER TIME, HOW THINGS MIGHT BREAK DOWN, AND HOW ESSENTIAL IT IS TO DO REPAIR WORK DURING THE IMPLEMENTATION PROCESS.**

is the use of chatbots to automatically deliver basic service tasks (Migri 2018b). According to Migri, chatbots have eased the burden on their overwhelmed telephone customer service staff. Before the chatbots were launched in May 2018, the majority of calls did not get through at all: only around 20% of incoming calls were responded to. After chatbot automation, the share of responded calls has increased to around 80%. This was possible because many of the calls involved routine issues. Chatbot automation also redirects service-seekers to other government agencies: for example, starting a business in Finland involves Migri, but also other agencies such as the Tax Administration or the Patent and Registration Office (Migri 2018c).

The other aspect of automation concerns the actual decision-making, which Migri considers the key method of speeding up permit processes (Migri 2019b). Piloted automated processes include citizenship applications and seasonal work certificates. Here, as well as in customer service, the aim is to automate those parts of the process that do not require human input. Significantly, Migri makes a distinction between partial and full decision automation: full automation refers to a process where an information system executes all phases of the decision process. According to Migri, full decision automation would only be applied to straightforward decisions that result in a positive outcome with a claim accepted or a permit granted. Negative decisions, or decisions that require case-by-case consideration or a discussion with the applicant, will not be fully automated. Full automation is, therefore, not suitable for processing asylum applications (Valtioneuvosto 2019). However, even asylum decision processes could be partially automated, so that those phases of the process that do not require human judgement could be carried out automatically. Even with these restrictions, it is notable that the potential for automation concerns a large share of all Migri's decisions. In 2019, about 71,000 decisions (more than two thirds of all decisions) concerned residence permits, and about 87% of these decisions were positive. Almost 12,000 citizenship decisions were made with a similar share of positive decisions (Migri 2018).

In its public statements, Migri has pointed to the organizational changes necessary for making the automation of permit processes possible, including rearrangement of the authority's own operations (Migri 2019b). Even more pressing changes include amendments to legislation relating to personal data in the field of immigration administration. A proposal at the Parliament of Finland is aimed at renewing the regulations on Migri's personal data processing to

"meet the requirements related to digitalization", and allow automated decisions to be made when certain preconditions are met (Ministry of the Interior 2019). The proposal underwent evaluation by the Constitutional Law Committee at the Parliament of Finland. In its statement (Eduskunta 2019), the committee expressed a critical view towards the proposal. In the committee's view, the problems with the proposal concern the fulfillment of good governance requirements, and the responsibilities of officials for the legality of their actions. In addition, instead of approaching decision automation with authority-specific legislation, the committee would favor more general legislation on the topic, which should be amended with sector-specific rulings where necessary.

Migri's plans to automate permit processes, then, take place in the context of internal and external pressures on the agency's permit processes: increased number of applications to process, desire to focus resources on issues that require human judgement, demands for speedy application processing, and a legislation that currently delimits the employment of ADM. Simultaneously, proposals for amending the legislation seem to be running into constitutional conflicts, forestalling or at least delaying, the authority's automation plans.

## **/ Automatic identification of individual risk factors in social care and healthcare**

In October 2019, the Japanese multinational IT service provider, Fujitsu, announced that it was developing an AI solution for South Karelia's social care and healthcare district (known by its Finnish acronym Eksote) (Fujitsu 2019). The project employs machine learning methods with the aim of helping Eksote identify factors underlying social exclusion of young adults, as well as predicting associated risks. With the predictive model, social and healthcare professionals will be provided an overview of risk factors. According to Fujitsu's press release, the model identifies some 90% of young adults susceptible to social exclusion. In practical terms, the model that is being used is derived from pseudonymized data taken from the use of Eksote's services by young adults, and it uses this data to predict social exclusion outcomes defined by Eksote's professionals. According to Eksote, the legislation on the secondary and combined use of healthcare data makes it possible to use only non-identifiable, pseudonymized data. This means that Fujitsu's model cannot be used to identify individual young adults considered to be at risk of social exclusion; rather, the

model produces a list of risk factors on a general level. The next step in the project is to examine whether it is possible, under the current legislation, to set up a consent-based system: a client's consent would be asked for before using the predictive model on their individual data when they, for example, have an appointment with a social care or healthcare professional (Etelä-Saimaa 2019).

The project sits well in a continuum of Eksote's projects involving predictive models. In June 2018, Eksote announced that it had developed, in collaboration with the Finnish IT firm, Avaintec, an AI model to predict problems experienced by children and the youth (Eksote 2018a). The aim was to identify problems early on, so that an intervention could be made and support provided to families sooner rather than later. Much like in the 2019 project on young adults and social exclusion, this model was based on explicitly defined undesired "endpoints": low grade averages, disciplinary interventions at school, high class non-attendance figures, being taken into custody, acceptance into psychiatric care, and substance abuse (Eksote 2018b). The model made use of data gathered from the IT systems of maternity clinics, kindergartens, schools, healthcare, and mental healthcare providers, as well as social services – e.g., it combined data from different administrative branches, namely social care, healthcare, and education. The outcome was the identification of a total of 1340 risk factors (Yle 2018), ranging from bad teeth in children, to parents missing maternity clinic appointments, to the child's siblings bullying others at school. These examples also give an idea of the kinds of data that were employed when making predictions. Ideally, the model would make it possible to continuously keep track of risk factors of individual children.

From a technical point of view, Eksote's information systems could be set up to signal an alarm when a given threshold is exceeded. Indeed, this was the ultimate desire expressed by the project's developers, but the practical implementation of such feedback functionalities ran into conflict with existing legislation. While it is permissible to combine non-identifiable, individual-level data from different administrative branches for the purpose of research (Oppimaa n.d.), it would not be legal to do the same thing for identifiable individuals and on a continuous basis. The project, then, was considered a pilot or a trial, the modelling phase being the practically implemented part, and further implementation is pending legislative changes. However, when an AI-branded model remains on the level of general lists of risk factors and their associated probabilities, it provides limited help beyond traditional statistical research, let alone

the intuition of experienced social work professionals. It is probably not surprising that, for example, when a child's hygiene is neglected, their siblings have problems at school, or appointments are repeatedly missed, that there can be, but on the other hand, may not be, other problems later on in the child's development.

Importantly, Eksote's 2018 and 2019 AI projects operate with similar principles: they define negative outcomes for a given group of people, combine data from different administrative sources, and train a machine learning model with the aim of providing a list of risk factors with probabilities associated with negative outcomes. This kind of approach can be replicated across welfare contexts. Eksote also uses a similar model for the purpose of identifying risk factors in elderly people related to their decreased ability to function. It seems obvious to conclude that if it were possible to combine individual-level data from different administrative branches, models like these, seeking correlations for risk factors, could be implemented quickly into social and healthcare systems.

## Policy, oversight and debate

### / AuroraAI for life-event management

The national AI program aims to turn Finland into a leading country in the application of AI. The program's 2017 report on the objectives and recommended measures outlined how AI can facilitate the public sector's development towards becoming a more efficient and effective service provider (Ministry of Economic Affairs and Employment 2017). Indeed, one of the key actions recommended in the report was to build the world's best public services. As a practical step, the report describes a citizen's AI assistant called Aurora.

Since these recommendations, the idea of the citizen's AI assistant has expanded into an AuroraAI program led by the Ministry of Finance (Ministry of Finance 2019a). The above-mentioned 2017 report still described Aurora as something fairly easy to grasp, e.g., the citizens' AI assistant, not unlike commercial digital assistants such as Siri, tirelessly helping with personalized services. Over the past two years, AuroraAI has become much broader in its aims and ultimately



vaguer: it is “a concept for human-centric and ethical society in the age of artificial intelligence”, as the program’s development and implementation plan defines it (Ministry of Finance 2019b). In the “AuroraAI partner network”, it is envisioned that a group of service providers, including public sector organizations, private firms as well as NGOs, will jointly provide the services offered to citizens. The idea is to automatically identify “life events”, e.g., circumstances in people’s lives that give rise to specific service needs. AuroraAI is described by its promoters as “a nanny”, or “a good guardian” that identifies and predicts life events, and helps citizens to utilize public and private services by suggesting and offering the ones needed in these particular circumstances. Use case examples include a student moving to a new city, retirement from work, loss of employment, or changes in family relations. The actual outcomes of the program for citizens remain conceptual, but what is in the plans resembles a recommender system for public services (Ministry of Finance 2019c). In

addition to offering services in a timely and frictionless manner, the program is simultaneously based on a currently dominant economic rationale: timeliness, personalization, targeting, and automated service provision are expected to increase efficiency and remove wastefulness.

From a legal and regulatory perspective, the AuroraAI initiative is far from a straightforward exercise.

One of the foundational principles underlying the “life event” model is that data collected in various parts of the public sector will be brought together and used to proactively develop services for citizens. However, as the final report of the national AI program (Ministry of Economic Affairs and Employment 2019) points out, the existing legislation that covers the using, sharing, and combining of personal data collected by different branches of public administration “is very challenging” when it comes to novel uses of the same data. The report, therefore, identifies agile development of regulations as a necessary step to make

ONE OF THE  
FOUNDATIONAL  
PRINCIPLES UNDERLYING  
THE “LIFE EVENT”  
MODEL IS THAT  
DATA COLLECTED IN  
VARIOUS PARTS OF  
THE PUBLIC SECTOR  
WILL BE BROUGHT  
TOGETHER AND USED  
TO PROACTIVELY  
DEVELOP SERVICES  
FOR CITIZENS.

new services based on life events possible, as well as to facilitate the use of AI to serve the public interest. For this purpose, the report proposes “sandbox” environments, or restricted trial grounds for formulating new regulations and novel technologies. The idea is that consenting citizens would voluntarily participate in trials and pilot projects operating in the “sandbox”, and allow the secondary use of their personal data, and the development and testing of new AI services, without the need to amend existing regulations. Even if the sandbox model is realized, it remains to be seen what response it will receive from citizens; whether or not citizens are willing to “let the digital twin empower them”, as AuroraAI’s promotional material suggests.

In addition to the regulatory constraints related to combining public-sector data, the desirability of a public services recommender system should also be questioned. In the Finnish context – with its high general trust in the state and its services – some of AuroraAI’s aims, including timely and effective provision of public services, are unlikely to be broadly contested as such. However, the way that the project is currently imagined gives rise to a number of questions that merit careful consideration. For example, how would it be determined which services are either offered, or not, to citizens? The recommendations based on predictive models could go against accepted norms or notions of appropriateness. For example, offering divorce-related public services as soon as the statistical criteria for such a “family relations life event” are met, would likely be frowned upon. The services offered to citizens would need to be carefully curated, but on what legal or moral grounds? Another issue concerns the influence that recommendations could have on citizens’ actions. The project’s promoters say that AuroraAI empowers individuals by giving them control over the decisions they make. However, it is well established that “nudging” – by means of the design of choice architecture and appropriately timed suggestions (or lack thereof) – has an effect on the choices made. AuroraAI would, therefore, act as a decision guidance system, affecting, for example, which social benefits to apply for, or which health and well-being services to make use of. Effective recommendations would increase or decrease service use, and would, therefore, have financial consequences for the public sector. What, then, is optimized on the level of the public services system: is service usage maximized, are costs minimized, or is citizen well-being improved? What set of criteria are these decisions based on and who chooses them? So far, these broader questions about the automated offering of

public services have received hardly any attention in public discussion.

## **/ Automated benefit processes at the Social Insurance Institution of Finland (Kela)**

The Social Insurance Institution of Finland, known as Kela, is responsible for settling some 15.5 billion euros of benefits annually under national social security programs. Kela views AI, machine learning, and software robotics as integral parts of its future ICT systems. Kela's ongoing developments in the automation field include customer service chatbots, detection and prevention of misunderstandings or fraud, and data analytics. Legislation on benefits is complex, spanning hundreds of pieces of separate regulations set in place over 30 years, and legislative compliance of automated procedures is a major concern for Kela. One particular issue, identified in the 2019 Automating Society report, was the need to fulfil requirements concerning how benefit decisions are communicated to citizens, so that the reasoning and results of automation are translated into an understandable decision.

The issue of communicating automatic decisions has since sparked an investigation into Kela's ADM processes. Already in 2018, Finland's Chancellor of Justice – an official who supervises how authorities comply with the law, and who advances the legal protection of citizens – received a complaint concerning the communication of Kela's unemployment benefit decisions. When investigating the complaint, the Chancellor paid attention to the fact that Kela had implemented ADM procedures to settle the benefits. There were tens of thousands of such automated decisions, and no one person to contact who could provide additional information on the decisions (Oikeuskanslerinvirasto 2019a). As a result, in October 2019, the Chancellor started a further investigation into the use of automation in Kela's benefit processes.

The Chancellor's information request to Kela (Oikeuskanslerinvirasto 2019b) provides an overview of compliance and legal protection issues relevant to public sector ADM projects. The information request focused on the requirements of good governance, the legal accountability of officials for their actions, and how these requirements and responsibilities relate to automated decision processes. The Chancellor noted that while Kela publicly states that it has a right to make decisions automatically, it does not provide information on which decisions it automates, and

whether this refers to partial or full automation, or whether it should be considered decision-making supported by automation. Based on these considerations, the Chancellor requested more detailed information on Kela's ADM processes, including what benefit decisions it automates and what the legal basis of such decision automation is. Kela's views on ensuring good governance, as well as the legal protection of the citizens concerned, were also requested. Kela was asked to provide details on how it plans to further develop decision automation, for example, what types of ADM can be employed in the future. Furthermore, Kela was asked to take a stand on ADM when it comes to the accountability of officials; in particular, how responsibilities are distributed among Kela's leadership, ADM system developer, and how individual officials handle the decisions. Finally, Kela was requested to provide details on the fulfilment of citizens' rights to information on how decisions are made, how decision-making algorithms work, and the content of programmed decision-making rules.

The information request that the Chancellor sent to Kela also points out a potential distinction between partial and full automation of decisions. This suggests that whether officials employ automation as a tool or an aide for decision-making, or whether decisions are made completely automatically, may have ramifications on the conduct of good governance and the distribution of officials' responsibilities for the legality of their actions. In a broader context, ADM is sometimes framed as an incremental development: a continuation of older, technologically simpler means of automated information processing. In Kela's case, "traditional" means, such as batch processes and traditional software code, have been employed to automate parts of benefit processes over previous decades. However, a distinction between completely and partially automated processes suggests that, after a certain point, increasing the level of decision automation is not simply a shift from less to more automation, but it brings about a qualitative shift to a different kind of process that has new implications for compliance with legal requirements.

## **/ The Deputy Parliamentary Ombudsperson considers automated tax assessments unlawful**

In the previous Automating Society report, we discussed two inquiries by the Deputy Parliamentary Ombudsperson of Finland, Maija Sakslin, resulting from complaints about mistakes the Tax Administration made by using automated

**ACCORDING  
TO THE TAX  
ADMINISTRATION,  
MORE THAN 80%  
OF ALL TAX  
DECISIONS ARE  
FULLY  
AUTOMATED, AND  
IF AUTOMATION  
WERE BANNED,  
MORE THAN  
2000 NEW  
EMPLOYEES  
WOULD NEED TO  
BE HIRED.**

tax assessments. While the mistakes had already been corrected, the complaints resulted in the Deputy Ombudsperson's further investigation into automated tax assessments, paying special attention to how automated processes fulfil the requirements of good governance and accountability of officials, as well as to the legal protection of taxpayers. She stated that the taxpayer's legal right to receive accurate service and justification of taxation decisions are currently unclear, and requested clarification on these issues from the Tax Administration (Oikeusasiamies 2019).

In November 2019, the Deputy Ombudsperson gave a decision on the matter which contains clarification on the scale of automated tax assessment. According to the Tax Administration, hundreds of thousands of issues are handled annually without human involvement, so that the information system automatically carries out the whole process. The automated procedures include sending reminders, but also making decisions on additional taxes, tax collection and monitoring of payments made. According to the Tax Administration, more than 80% of all tax decisions are fully automated, and if automation were banned, more than 2000 new employees would need to be hired (Helsingin Sanomat 2019d).

The Deputy Ombudsperson points out several issues with the lawfulness of the Tax Administration's automated procedures. The first issue is the legal basis of automation. The Deputy Ombudsperson notes that automatic tax assessments were based on more general legislation not directly related to decision automation, and states that their legal basis does not, therefore, fulfil the standards set by constitutional law. Employing automated procedures would require an unambiguous legislative basis that defines, among other things, which issues are to be processed automatically. In addition, the legislation lacked a definition of an algorithm, and one would be necessary in order for them to be made available for public scrutiny. The second issue is the accountability of public servants for automated decisions. The Tax Administration handles official's legal responsibilities for their actions by naming process owners for all automated processes. This process owner, then, was considered responsible for decisions made under that particular process. In the Deputy Ombudsperson's view, such an arrangement means that the accountability of officials has been defined in an indirect manner, which means that personal accountability of individual officials remains unclear under both constitutional and criminal law. The third critical issue concerns the requirements of good governance. According to the Deputy Ombudsperson, the fulfilment of the requirements of good governance would imply that taxpayers are informed when they are subject to automated procedures. Citizens dealing with the Tax Administration, then, have the right to know the basis of automated processes, as well as when their issue has been resolved automatically.

Due to these considerations, the Deputy Ombudsperson states that the legal basis of the Tax Administration's automated procedures is unclear, and the procedures are, therefore, unlawful. The decision does not mean that the

use of automation should be immediately ceased; rather, the Deputy Ombudsperson stresses the urgency of investigating the legislative needs introduced by ADM procedures. Furthermore, she deplores the broad employment of automated procedures in the Tax Administration without a proper investigation into the legislative needs.

## **/ Digital Minds hiring service discontinued**

The most controversial Finnish case in the AlgorithmWatch report (2019) featured a company called Digital Minds, founded by two Finnish psychologists, which aimed to develop “third-generation” assessment technology for employee recruitment. After the report was published, Matthias Spielkamp and Nicolas Kayser-Bril wrote a follow-up story about Digital Minds for Politico (Spielkamp and Kayser-Bril 2019) underlining the need for critical reflection:

A Finnish company has rolled out a new product that lets potential employers scan the private emails of job applicants to determine whether or not they would be a good fit for the organization. The company, Digital Minds, portrays its offering as something innocuous. If applicants give their consent, what’s the harm? The truth is: We don’t know the answer to that question. And that’s what makes new, potentially intrusive and discriminatory technologies like this one so scary.”

For Juho Toivola, one of the two founders of Digital Minds, the piece in Politico was a “German take on the matter,” reflecting a need to have a precedent in the fight against algorithmic powers. According to Toivola, the critique of their service fails to understand the problem they are trying to solve; the goal is to make the employee assessment more convenient and efficient by utilizing existing data, rather than collecting all data separately with each recruitment. The personality analysis is done through IBM’s Watson, which is a commonly used personality test across the globe. The analysis focuses, for instance, on how active people are online and how quickly they react to posts or emails. Five different personality characteristics are analyzed, including agreeableness, conscientiousness, emotional range, and how extrovert or outward-looking a person is.

In order to get an assessment of the possible harm of the Digital Minds service and its legality in relation to GDPR, Nicolas Kayser-Bril reported Digital Minds to the Finnish data protection Ombudsperson. In May 2019, Finland’s national public service broadcasting company’s reporter, Virpi

Hukkanen, wrote a piece about the ongoing process with the Ombudsperson, suggesting that Digital Minds is offering a service that is legally problematic (YLE 2019). The Finnish Data Protection Ombudsperson, Reijo Aarnio, said to YLE that he suspected the personality assessment deduced by means of email correspondence violates the Labor Privacy Act that states that the information must be collected from the jobseekers themselves. Aarnio also questioned the validity of the jobseeker’s consent when asked for permission to analyze emails. If the candidate is in a vulnerable position and needs a job, he or she might not be in a position to decline access to their data. Moreover, as with letters, emails are covered by the confidentiality of correspondence act.

Remarkably, however, the news piece revealed that less than ten jobseekers had agreed to participate in Digital Mind’s analysis of their social media posts, and only one had agreed to participate in the analysis of email correspondence. No such analysis had been performed as part of a recruitment process. The fact that the company had very few actual clients gave a new twist to the debate concerning Digital Minds. As Toivola emphasized, he was talking about ‘a proof of concept’ rather than an actual service. It turned out that either intentionally or unintentionally Toivola had been collapsing contexts: founders of Digital Minds had experience of dozens of clients and they had been involved in thousands of assessments of job applicants. But these assessments had mostly been done by means of conventional analysis methods within the recruitment context. After the news piece, Toivola actively participated in social media discussions about their service, bringing to the fore their aims and intentions. Shortly after that, however, the service was discontinued. According to Toivola, Digital Minds is rethinking how to position their future offerings in an ethically more robust manner.

One of the lessons learned from Digital Minds is the importance of clarifying how exactly ADM becomes a part of making decisions that have real-life consequences. Similarly to other companies that offer hiring services, Digital Minds waived their responsibility in terms of real-life consequences by suggesting that they do not make actual hiring decisions, but merely automate the process of assessment, based on which employers then make decisions. Ultimately, it is the organizational environment and the divisions of labor that socio-technically arrange ADM systems that determine whether uses of automation are just and ethically robust. Machines fail to care about real-life consequences and ADM system implementers should take this into account.

## Key takeaways

The Finnish cases underline the importance of thoroughly investigating cases when assessing the use and the potential harm of ADM. When the aim of companies is to sell their services, they can alter the details and exaggerate numbers in a way that supports their marketing claims. The investigations should be prepared for the fact that the details about ADM systems – which are often laborious to uncover – contain the most important information. This means that in seeking more robust knowledge about the fundamental issues at the heart of the debate on AI, algorithms, and ethics, company representatives need to be part of the dialogue. That dialogue can only be nurtured in a culture that values openness: ideally, developers of ADM systems need to feel that they are not being attacked, but that they are participants in an ongoing societal debate. When assessing a case like Digital Minds, for instance, we are not merely talking about one single start-up company, but a larger trend in automation processes making important decisions about people's lives.

The collaboration between Utopia Analytics and Suomi24 reminds us of the everyday reality of socio-technical systems. The way many technical systems are built, with layers of software written by groups of coders who do not necessarily communicate with each other, makes them vulnerable to various kinds of breakages that might be impossible to predict beforehand. This stresses the fact that the implementation of AI tools is dependent on their integration into existing technical systems. With insufficient quality controls in place it might be impossible to know beforehand how the AI tools will behave.

In terms of public sector cases, the relationship between legislation and plans to implement fully automated decision-making procedures – i.e., procedures that do not involve humans at all – emerges as one key issue. The Tax Administration and Kela have already broadly implemented ADM, and the lawfulness of their procedures is now under scrutiny by supervisory authorities. The Deputy Parliamentary Ombudsperson's investigation and decision on the use of ADM in the Tax Administration points out that the legal basis of fully automated procedures is insufficient. In her decision, the Deputy Ombudsperson encourages relevant authorities to investigate the legislative needs arising from automation.

THE FINNISH  
CASES UNDERLINE  
THE IMPORTANCE  
OF THOROUGHLY  
INVESTIGATING CASES  
WHEN ASSESSING THE  
USE AND THE POTENTIAL  
HARM OF  
ADM.

This interpretation of the law is important for the broader context of ADM in the Finnish public administration, suggesting that technical procedures that are already widely employed are at odds with existing legislation. Indeed, the use of ADM in Kela is undergoing a similar investigation by the Chancellor of Justice.

Driven by efficiency-related pressures among other things, also other Finnish authorities are either employing or planning to employ automated procedures in their decision-making, the latter is exemplified by the Migri case.

When it comes to the legal status of ADM procedures, Migri's case is different from the other cases mentioned in this report. The automation plans are still pending new legislation that suffered a setback in the Parliament's Constitutional Law Committee. This highlights the need for more general, rather than case-by-case, legislation on the ADM procedures used by the public administration. The Ministry of Justice has launched a project to develop general legislation to ensure that the requirements of legality, good governance, legal protection of citizens, and accountability are met when the public sector implements ADM procedures (Oikeusministeriö 2020). A preliminary assessment of legislative requirements was published in February 2020. The way in which Finnish public officials are accountable for their decisions under criminal and constitutional law seems to give the public sector ADM discussion in Finland a distinct flavor.

The other public sector cases point to different ways in which existing legislation holds back the public sector's automation intentions. From the perspective of Eksote, and similar care providers, automated detection of risk factors is well-intentioned, reasonable, and cost-effective. These kinds of projects, however, tend to individualize the treatment of risk, with possible punitive outcomes when potential problems are brought to the awareness of officials. In a similar way, legislation is at odds with AuroraAI's plans to access data from different sectors of public administration. While the purpose of AuroraAI is not to intervene when risks are identified, but rather to recommend when needs are identified, the similarities with Eksote's risk monitoring system are evident. The purpose of the legislation that holds back automation plans is not to prevent data use, but rather to set boundaries for it. With their surveillance logic, monitoring systems that track individuals' actions might undermine societal cohesion and generate a culture of distrust that should be prevented, for good reasons.



# References:

- Accenture (2016): Accenture tukee Migriä Digitaalinen virasto -hankkeessa, <https://www.accenture.com/fi-en/company-news-release-maahanmuuttovirastoa-digitaalisatiahankkeessa>
- Eduskunta (2019): Valiokunnan lausunto PeVL72019 vp – HE 18/2019 vp, [https://www.eduskunta.fi/FI/vaski/Lausunto/Sivut/PeVL\\_7+2019.aspx](https://www.eduskunta.fi/FI/vaski/Lausunto/Sivut/PeVL_7+2019.aspx)
- Eksote (2018a): Lasten ja nuorten ongelmia ennakoita Suomessa laajimmalla keinoälymallilla Lappeenrannassa, <http://www.eksote.fi/eksote/ajankohtaista/2018/Sivut/Lasten-ja-nuorten-ongelmia-ennakoita-Suomen-laajimmalla-keinoälymallilla-Lappeenrannassa.aspx>
- Eksote (2018b): Digitalisaatio tulevaisuuden oppilashuollossa? Case Eksote/Lappeenrannan kaupunki, <https://vip-verkosto.fi/wp-content/uploads/2018/10/Digitalisaatio-tulevaisuuden-oppilashuollossa-Case-Eksote-24.9..pdf>
- Etelä-Saimaa (2019): Eksote yrittää pitää nuoret aikuiset elämässä kiinni keinoälyn avulla – Taistelu syrjäytymistä vastaan jatkuu asiakkaiden kanssa toteutettavalla pilottihankkeella, <https://esaimaa.fi/uutiset/laheella/79450bc3-1641-4a0f-bcea-11df63988bdc>
- Fujitsu (2019): Fujitsu ja Eksote selvittävät eteläkarjalaisten nuorten syrjäytymisriskiä tekoälyratkaisulla, <https://www.fujitsu.com/fi/about/resources/news/press-releases/2019/eksote-tekoaly.html>
- Helsingin Sanomat (2019c): Yli 800 ulkomaisen opiskelijan oleskeluluvat juutuivat käsittelyyn, osa korkeakouluista joutuu palauttamaan kymmenientuhansien eurojen arvosta lukuvuosimaksuja, <https://www.hs.fi/talous/art-2000006306736.html>
- Helsingin Sanomat (2019d): Verottajan robotti on karhunnut ihmisiltä liikaa rahaa, ja apulaisoikeusasiamiehen mukaan automaattinen päätöksenteko rikkoo myös perustuslakia, <https://www.hs.fi/kotimaa/art-2000006321422.html>
- Keskuskaupakamari (2019): Työperäisellä maahanmuutolla korjausliike väestöennusteeseen, <https://kauppakamari.fi/2019/09/30/tyoperaisella-maahanmuutolla-korjausliike-vaestoennusteeseen/>
- Migri (2018a): Annual Report 2017, <https://migri.fi/documents/5202425/6772175/2017+Annual+Report>
- Migri (2018b): Kamu AI Chatbot, <https://www.slideshare.net/DigitalHelsinki/tekoly-ja-me-osa-2-kamuaichatbotmigrinesity2018>
- Migri (2018c): Have you met Kamu, PatRek and VeroBot? Chatbots join forces to offer advice to foreign entrepreneurs in Finland, [https://migri.fi/en/article/-/asset\\_publisher/tunnetko-kamun-patrekin-ja-verobotin-chatbotit-neuvovat-yhdessa-ulkomaalaista-yrittajaa](https://migri.fi/en/article/-/asset_publisher/tunnetko-kamun-patrekin-ja-verobotin-chatbotit-neuvovat-yhdessa-ulkomaalaista-yrittajaa)
- Migri (2019a): Finnish Immigration Service Strategy 2021, [https://migri.fi/documents/5202425/9320472/strategia2021\\_en/b7e1fb27-95d5-4039-b922-0024ad4e58fa/strategia2021\\_en.pdf](https://migri.fi/documents/5202425/9320472/strategia2021_en/b7e1fb27-95d5-4039-b922-0024ad4e58fa/strategia2021_en.pdf)
- Migri (2019b): Finnish Immigration Service to pursue a considerable reduction of processing times, [https://migri.fi/artikkeli/-/asset\\_publisher/maahanmuuttovirasto-tavoittelee-kasittelyaikojen-voimakasta-lyhentamista?\\_101\\_INSTANCE\\_FVTI5G2Z6RYg\\_languageId=en\\_US](https://migri.fi/artikkeli/-/asset_publisher/maahanmuuttovirasto-tavoittelee-kasittelyaikojen-voimakasta-lyhentamista?_101_INSTANCE_FVTI5G2Z6RYg_languageId=en_US)
- Ministry of Economic Affairs and Employment (2017): Finland's age of artificial intelligence, <https://www.tekoalyaika.fi/en/reports/finlands-age-of-artificial-intelligence/>
- Ministry of Economic Affairs and Employment (2019): Finland leading the way into the age of artificial intelligence, <https://www.tekoalyaika.fi/en/reports/finland-leading-the-way-into-the-age-of-artificial-intelligence/>
- Ministry of Finance (2019a): Implementation of the national AuroraAI programme, <https://vm.fi/en/auroraai-en>
- Ministry of Finance (2019b): AuroraAI – Towards a human-centric society. <https://vm.fi/documents/10623/13292513/AuroraAI+development+and+implementation+plan+2019%E2%80%932023.pdf>
- Ministry of Finance (2019c): #AuroraAI – Let your digital twin empower you, [https://www.youtube.com/watch?v=A2\\_hlJrEiWY](https://www.youtube.com/watch?v=A2_hlJrEiWY)
- Ministry of the Interior (2019): Renewal of personal data legislation in the field of immigration administration progressing, [https://valtioneuvosto.fi/artikkeli/-/asset\\_publisher/1410869/maahanmuuttohallinnon-henkilotietolainsaadannon-uudistus-etenee?\\_101\\_INSTANCE\\_YZfcyWxQB2Me\\_languageId=en\\_US](https://valtioneuvosto.fi/artikkeli/-/asset_publisher/1410869/maahanmuuttohallinnon-henkilotietolainsaadannon-uudistus-etenee?_101_INSTANCE_YZfcyWxQB2Me_languageId=en_US)

Oikeusministeriö  
(2020): Automaattista päätöksentekoa koskevan hallinnon yleislainsäädännön valmistelu. OM021:00/2020. <https://oikeusministerio.fi/hanke?tunnus=OM021:00/2020>

Oikeuskanslerinvirasto (2019a): Työmarkkinatukipäätöksessä ilmoitettavat tiedot lisätietojen antajasta, [https://www.okv.fi/media/filer\\_public/f4/06/f4063426-fe67-48e0-ae6a-1ad7eb2c8feb/okv\\_868\\_1\\_2018.pdf](https://www.okv.fi/media/filer_public/f4/06/f4063426-fe67-48e0-ae6a-1ad7eb2c8feb/okv_868_1_2018.pdf)

Oikeuskanslerinvirasto (2019b): Selvityspyyntö OKV/21/50/2019, [https://www.okv.fi/media/filer\\_public/9d/5e/9d5e3a9e-9af5-425a-8c04-8da88de0c058/okv\\_21\\_50\\_2019.pdf](https://www.okv.fi/media/filer_public/9d/5e/9d5e3a9e-9af5-425a-8c04-8da88de0c058/okv_21_50_2019.pdf)

Oikeusasiamies (2019): Verohallinnon automatisoitu päätöksentekomenettely ei täytä perustuslain vaatimuksia, <https://www.oikeusasiamies.fi/r/fi/ratkaisut/-/eoar/3379/2018>

Oppimaa n.d.: Ennustetyökalun kehitystyö – Lasten ja nuorten ongelmia ennakoidaan Suomen laajimmalla keinoälymallilla Lappeenrannassa, <https://www.oppimaa.fi/oppitunti/ennustetyokalun-kehitystyoy/>

Spielkamp, Matthias and Kayser-Brill, Nicolas (2019): Resist the robot takeover, <https://www.politico.eu/article/resist-robot-takeover-artificial-intelligence-digital-minds-email-tool/>

Suomi24 (2019): Moderointi puhuttaa – tällaista se on Suomi24:ssä, [https://keskustelu.suomi24.fi/t/16034629/moderointi-puhuttaa---tallaista-se-on-suomi24ssa-\(osa-1\)](https://keskustelu.suomi24.fi/t/16034629/moderointi-puhuttaa---tallaista-se-on-suomi24ssa-(osa-1))

Utopia Analytics (2020): Utopia Analytics website, <https://utopiaanalytics.com/>

Yle (2018): Reiät lapsen hampaissa voivat kieliä huostaanoton riskistä – tutkimus löysi yli 1000 merkkiä, jotka ennustavat lapsen syrjäytymistä, <https://yle.fi/uutiset/3-10313460>

Yle (2019): Psykologien rekrytyritys skannaa jopa työnhakijan sähköpostin ja tekee siitä analyysin – Tietosuojavaltuutettu haluaa selvityksen palvelusta, <https://yle.fi/uutiset/3-10798074>





BELGIUM  
DENMARK  
ESTONIA  
FINLAND

**FRANCE**

**STORY**

PAGE 98

**RESEARCH**

PAGE 103

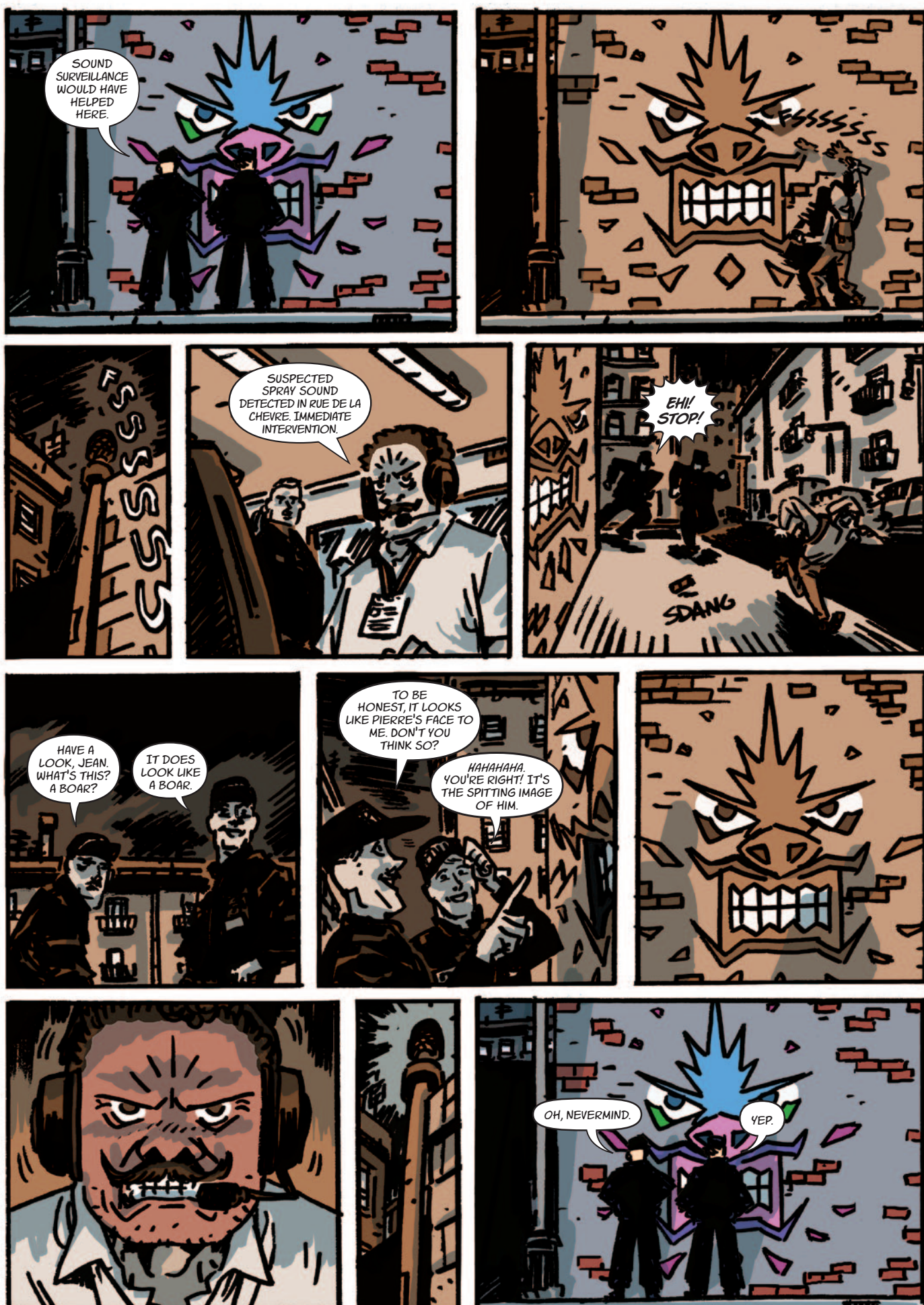
GERMANY  
GREECE  
ITALY  
NETHERLANDS

POLAND  
PORTUGAL  
SLOVENIA

SPAIN  
SWEDEN  
SWITZERLAND

UNITED KINGDOM





Find out more on p. 104 in the research chapter under "Automating surveillance in large cities".



# // Between care and control: 200 years of health data in France

**The French “Health Data Hub” will soon offer health data on all French citizens to AI startups that request it. It is the latest step in a project that began 200 years ago, aimed at centralizing health data. It is a project that has oscillated between care and control, but which has veered mostly towards control.**

By Nicolas Kayser-Bril

Coming back from the 1876 World Hygiene Congress, French doctor Olivier du Mesnil was in awe of Brussels' health information system. There, he wrote, doctors immediately sent a copy of death certificates to the central hygiene bureau. Every week, bureau personnel produced a map of the city where each death was marked with a pin, with different colors for each contagious disease, and shown to the mayor. This technology, Mr. du Mesnil said, allowed even "the person least familiar with scientific research to exert non-stop control over the health status of the population."<sup>1</sup>

Mr. du Mesnil used this Belgian example to underline how much France had to catch up. Like its neighbor to the north, France was a rapidly industrializing country in the 19th century. Railroads sped up transport and factories brought together large numbers of workers packed in unhealthy, precarious dwellings.



Source gallica.bnf.fr/ Bibliothèque de France.

**Data collection, 1930**

## / Controlling cholera

The main beneficiaries of this combination were bacteria that could jump from host to host rapidly over long distances. Two cholera epidemics, in 1832 and 1854, claimed over 100,000 lives each.<sup>2</sup> The bacteria responsible for the illness was not identified until 1884, but governments across Europe understood the need to collect data to follow the spread of cholera and impose quarantines in time to stop its progression. After the first cholera epidemic, the French government began gathering information throughout the country on what was thought to be the cause of the sickness: bad air, rotten food, and even "ignorance".<sup>3</sup>

GOVERNMENTS ACROSS EUROPE UNDERSTOOD THE NEED TO COLLECT DATA TO FOLLOW THE SPREAD OF CHOLERA AND IMPOSE QUARANTINES IN TIME TO STOP ITS PROGRESSION.

Despite its lethality, cholera was not the deadliest disease of the 19th century. Dysentery and tuberculosis were far more dangerous but were limited to the poorest in society.<sup>4</sup> That the government only took action regarding cholera – which killed people from all social classes – was taken by many as proof that the health of the poor was only a concern when it impacted the rich. Distrust of government measures to fight cholera, including data gathering, ran high.<sup>5</sup>

## / Health police

Until well into the 20th century, health was a matter of public order rather than well-being. Until the First World War, health was the purview of the ministry of the interior. To monitor and predict the spread of tuberculosis in large cities, authorities built a "health record" for each building, modeled on the criminal record of each individual. Workers would need to present both to obtain a job. (The health record was discontinued around 1908 after hospital personnel pointed out that building-level data was not of much use.)<sup>6</sup>

A change of perspective occurred in the early part of the 20th century; first, warfare required to conscript not just part of an age cohort, but all of it. The general health level of the population acquired a new military importance. Second, eugenics, a pseudo-scientific craft that claimed to improve a population by rooting out its unhealthy members, increased in popularity. Health and hygiene became political goals in themselves and gained their own ministry in 1920.<sup>7</sup>

## / Knowledge monopoly

Health statistics, once concerned only with epidemics and controlling the poor, started to record well-being. The League of Nations, a newly created and optimistic interna-

Endnotes on page 102

**THEY  
COMPLAINED  
THAT SUCH DATA  
COLLECTION  
WOULD  
ENDANGER  
DOCTOR-PATIENT  
CONFIDENTIALITY,  
BUT THEIR MAIN  
CONCERN MAY  
WELL HAVE BEEN  
THE LOSS OF  
STATUS.**

tional organization, led the movement and commissioned health surveys across countries.

Not all French doctors were enthusiastic about the change.<sup>8</sup> They complained that such data collection would endanger doctor-patient confidentiality, but their main concern may well have been the loss of status. At the time, doctors were the ultimate repository of medical knowledge. Passing on information to the state was seen as a devolution of power. Because doctors were almost entirely financed by their patients, they had little incentive to cooperate in systems they disliked.

In any case, the collection of health data for the well-being of the population was only limited to a fraction of French

taxpayers. In the colonies, health was still seen as a production factor, to be optimized only insofar as it made plantations and mines more profitable. Until 1946, all French colonies together employed, at most, four statisticians, for whom health data was probably not a priority.<sup>9</sup>

## **/ Knitting needles and data processing**

Some in the medical sciences saw an opportunity in structuring data. In 1929, Louis Bazy, a surgeon consulting for the Paris-Orléans railway company, had the idea to use his employer's "statistics machines" to aggregate data on the health of the company's workforce. He designed a system whereby each employee's illness was coded in a punch card, along with her or his personal data. The statistics machine could process 400 cards a minute and, with the help of the tabulating machine, provide information on the spread of a disease or correlate between variables. The scope of applications for medicine and research was endless, he wrote.<sup>10</sup>

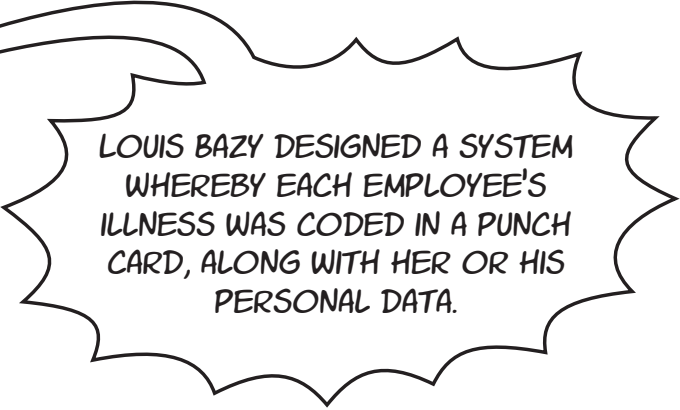
Not every doctor had access to statistics machines, so a professional magazine for progressive doctors explained how to process punch cards with knitting needles.<sup>11</sup> Despite such efforts, there is no evidence that these early attempts at computerized medicine gained many followers.

## **Certainly not GDPR-compliant: Mr. Bazy's structuration of health data on punch cards**

### **/ The drive towards centralization**

During the Second World War, the French government made big strides to implement eugenics. A side effect of this policy was the creation of a National hygiene institute (*Institut national d'hygiène*, INH). From 1942, it conducted large-scale data collection to track the effects of the government's crackdown on alcoholism and venereal disease. It also built a central repository of information on 35,000 cancer patients.<sup>12</sup>

After the war, INH expanded and kept monitoring the nation's health (it became the French National Institute of Health and Medical Research, Inserm, in 1964). On the other hand, the post-war government offered social insurance to all its citizens. With it came a social security number which is given at birth and which remains immutable until death. Having a unique identifier for every citizen revived the old dream of governance through numbers, where decisions could be taken purely on the basis of data.



LOUIS BAZY DESIGNED A SYSTEM WHEREBY EACH EMPLOYEE'S ILLNESS WAS CODED IN A PUNCH CARD, ALONG WITH HER OR HIS PERSONAL DATA.

In France, as in other countries of the Western bloc, central planning was considered a necessity. The government felt it had to collect comprehensive data on morbidity (that is, on the illnesses affecting the population). A first attempt, in 1945, to force hospital doctors to fill out forms after each procedure, to be sent to a central authority, failed. Another attempt was made in 1958 and again in 1972. As in the 1930s, doctors did not comply with their new obligations. They criticized the methodology, complained about the added workload, and failed to see any benefits for them.<sup>13</sup>

## / Digitization

This changed in the 1980s when an attempt at centralizing morbidity data started in 1982. By the beginning of the next decade, all hospitals were feeding data to a central authority.

This success – for the state – might have to do with the economic environment of that decade. The slow economic growth provided an incentive for the government to cut costs in healthcare. Despite initial reluctance from doctors, several health ministers pushed it through and made the new system mandatory in 1991.<sup>14</sup>

The data-gathering effort was, first and foremost, a cost-control mechanism. Knowing how many procedures each hospital carried out, and how much money each hospital received, the health ministry could rank their performance, in financial terms at least. Hospitals that overspent were called to order until, in the 2000s, the government introduced a pay-per-procedure system. A heart attack of level 1 is worth €1,205.57, increased to €1,346.85 if the patient dies within two days.<sup>15</sup> Each procedure that doctors perform is coded, following a strict classification, and hospitals are paid by the social security accordingly.

To navigate the list of over 6,000 procedures, hospitals hire external consultants to “optimize” their coding practices. As [AlgorithmWatch reported in May 2019](#), code optimization is nothing less than “generalized cheating” to maximize revenue, according to a health sociologist at the University of Lille.

## / Quality concerns

Because France has mandatory national health insurance with a single-payer, the morbidity data can be linked to medication usage as soon as a drug is reimbursed by social security. For about 99% of the population, the French national health insurer has comprehensive information on hospital procedures and drug intake since the early 1990s.<sup>16</sup>

This unique data set allowed researchers to find hidden correlations. This is how Benfluorex (sold under the name Mediator) was linked to heart valve disease, leading to the withdrawal of the drug in 2009.

However, all information on hospital procedures is related to accounting and not medical procedures. The optimization of procedure encoding does a great disservice to data quality, but no one knows exactly how bad the situation is, as very few studies have been conducted. One 2011 study showed that, for one specific procedure, the false positive rate was close to 80%, while false negatives reached 35%.<sup>17</sup>

## / FOMO

Despite this abysmal performance, in 2019, the French government pushed to build an even bigger database, called the “Health Data Hub”. Cédric Villani, a mathematician who spearheaded the Artificial Intelligence strategy of president Emmanuel Macron, wrote in a parliamentary report that the real risk of AI in health would be “not to welcome it”.<sup>18</sup> The Hub aims at providing any health-related data to AI projects that request it.

Since 2004, the government has pushed for all French residents to open an electronic health record (EHR). After a slow start, the centralized EHR will be made opt-out in 2021, and should, in time, be versed in the Health Data Hub.<sup>19</sup>

The French data protection authority criticized the project because of its broad aims. Data from the Hub can be used for any “public interest” goal, potentially opening the door to any commercial application. Critics also pointed out that personal data in the Hub is pseudonymized but not aggregated so that it can be easily de-anonymized.<sup>20</sup>

## / Toxic relationships

A doctor, who wished to be identified only as Gilles, started a “data strike” when the Health Data Hub was officially launched in December 2019. He and others called on colleagues to stop filling out the forms that feed the Hub. Since the 1980s, he said, France moved from “a healthcare that cures to a healthcare that counts,” pointing to the cost management systems. He saw no benefits in the new database, saying that research does not need it. “Software only robs time that should be spent on caring for patients,” he added.

Even if the success of the strike cannot be quantified, Gilles’ anger is widely shared. In January 2020, over 1000 doctors resigned their administrative duties, citing the pay-per-procedure system as one of the largest problems.<sup>21</sup>

It was also revealed that the designer of the Health Data Hub quit his job to work for a private sector firm specialized in selling health data. He saw no conflict of interest.<sup>22</sup>

## / Health data shrug

The main breakthrough of the Health Data Hub is that, for the first time, a French government decided to use an English name for an official project.<sup>23</sup> The rationale that led to its creation is a direct continuation of 200 years of efforts by the French government to gather data on its citizens, to make its population more legible, and more governable.

No one knows what the Health Data Hub will bring, but history offers some insights. The information system that Brussels set up in the 1870s, which Mr. du Mesnil so admired, might have worked. The city was spared any large epidemic until the Spanish flu of 1918. But then again, so were all large cities in France. On the other hand, life expectancy in Brussels, relative to the Belgian countryside and other large cities, *decreased* between 1885 and 1910.<sup>24</sup>

It could be that health data and actual health do not always go hand in hand.

## Endnotes

- 1 Du Mesnil, O. L'Exposition et le Congrès d'Hygiène et de Sauvetage de Bruxelles en 1876. *Annales d'Hygiène Publique*, 1877, p. 20.
- 2 Bourdelais, P., Raulot, J. Y. & Demonet, M. La marche du choléra en France: 1832 et 1854. *Annales. Histoire, Sciences Sociales*, 33(1), 1978, p. 125-142.
- 3 Drouet, S. Santé et environnement en France: deux siècles de «symbiose» juridique (1802-2002). *Revue juridique de l'Environnement*, 28(3), 2003, p. 324
- 4 On tuberculosis, see Bello, S., Signoli, M. & Dutour, O. Analyse de l'évolution de la mortalité par tuberculose du XVIIIe au XXe siècle dans quatre grandes villes françaises. *Médecine et Maladies Infectieuses*, 30(5), 2000, 275-283.
- 5 For a British study of the phenomenon, see Morris, R. J. *Cholera 1832: the social response to an epidemic*. Taylor & Francis, 1976. To the best of my knowledge, no such study has been carried out in France, but it is highly unlikely that the processes there were different.
- 6 Fijalkow, Y. Mesurer l'hygiène urbaine en épargnant les propriétaires: Le casier sanitaire des maisons de Paris. *Les Annales de la Recherche Urbaine*, 53(1), 1991, p. 73-78.
- 7 On the administrative organization of health authorities, see Drouet, S., *op. cit.*
- 8 For an interesting discussion among doctors, see Le Secret Médical et la Collaboration des Médecins-Praticiens à l'Hygiène Sociale, *Le Mouvement sanitaire : organe officiel de l'Association des médecins hygiénistes français*, 1927, p. 626-630.
- 9 Le Service Colonial des Statistiques. *Bulletin de la Statistique Générale de France*, octobre-novembre 1945, 377-379.
- 10 Bazy, L. L'emploi des machines pour l'obtention des statistiques médicales. *La Presse Médicale*, 18 janvier 1933, p. 105-106.
- 11 R.H.H, L'emploi des machines pour l'obtention des statistiques médicales [Review]. *Le Mouvement Sanitaire*, 1933, p. 121-122.
- 12 Picard, J. F. Aux origines de l'Inserm: André Chevallier et l'Institut national d'hygiène. *Sciences sociales et santé*, 21(1), 2003, 5-26.
- 13 Chaperon, J., Villard, J. M. & Riou, F. L'information médicale enjeu de la gestion hospitalière. *Politiques et management public*, 6(2), 1988, p. 35-46.
- 14 On the first implementations of PMSI, see Mossé, P. La rationalisation des pratiques médicales, entre efficacité et effectivité. *Sciences sociales et santé*, 16(4), 1988, 35-60.
- 15 Arrêté du 6 mars 2019 fixant pour l'année 2019 les éléments tarifaires mentionnés aux I et IV de l'article L. 162-22-10 du code de la sécurité sociale.
- 16 Bezin, J., Duong, M., Lassalle, R., Droz, C., Pariente, A., Blin, P. & Moore, N. The national healthcare system claims databases in France, SNIIRAM and EGB: Powerful tools for pharmacoepidemiology. *Pharmacoepidemiology and Drug Safety*, 26(8), 2017, p. 954-962.
- 17 Chantry, A. A., Deneux-Tharoux, C., Bal, G., Zeitlin, J., Quantin, C. & Bouvier-Colle, M. H. Le programme de médicalisation du système d'information (PMSI)-processus de production des données, validité et sources d'erreurs dans le domaine de la morbidité maternelle sévère. *Revue d'épidémiologie et de santé publique*, 60(3), 2012, p. 177-188.
- 18 Rapport n° 401 (2018-2019) de MM. Gérard Longuet, sénateur et Cédric Villani, député, fait au nom de l'Office parlementaire d'évaluation des choix scientifiques et technologiques, déposé le 21 mars 2019.
- 19 Cuggia, M., Polton, D., Wainrib, G. & Combes, S. *Health Data Hub, Mission de Préfiguration*. 2019. p.65.
- 20 Hourdeaux, J. «Health Data Hub»: le méga fichier qui veut rentabiliser nos données de santé. *Mediapart*, 2019.
- 21 Hôpital : «Nous avons pris la décision inédite et difficile de démissionner». *Libération*, 2020.
- 22 Foucart, S. & Horel, S. Données de santé : conflit d'intérêts au cœur de la nouvelle plate-forme. *Le Monde*, 2019.
- 23 This claim is debatable. “FranceConnect”, in 2015, used two French words put in a strange order, and “Make our Planet Great again”, in 2017, was only a call for proposals. “Choose France”, in early 2019, was the slogan of another program.
- 24 Eggerickx, T. & Debuissou, M. La surmortalité urbaine: le cas de la Wallonie et de Bruxelles à la fin du XIXe siècle (1889-1892). *Annales de démographie historique*, Jan. 1990, p. 23-41.



# Research

By Nicolas Kayser-Bril

## Contextualization

Many people within French institutions and large corporations see themselves as front-runners in artificial intelligence. However, only a few of these same people would argue that they are the best in the world, and yet almost all of them complain that they are not. This paradox seems to imply that they see their natural position as being in the leading pack.

This state of mind makes for some almost farcical situations, like the time when the government called its AI strategy “AI for humanity” (without, of course, asking the non-French sector of humanity for their opinion). Despite this touch of arrogance, the French make several arguments for their lofty claims. Firstly, France will likely be the first European country to have a centralized database of biometric identity of all its citizens, allowing public services to offer (or require) identification via face recognition. Secondly, almost all ministries and large cities have completed their digitization drives, and many now use their data troves to implement algorithms. Thirdly, the French tax authorities are the most active participants in the European automated information exchange system. Over the course of 2018, they sent and received information on close to 3 million taxpayers, more than any other country. Furthermore, France leads the way when it comes to legislation to keep algorithms in check, even if the law is enforced patchily, if at all.

Private-sector companies such as Thales or Idemia (formerly Morpho-Safran) are world leaders in biometrics. If you’ve ever crossed an international border, chances are you were submitted to automated tools from Idemia, which operates the Schengen area’s Visa Information System and the TSA PreCheck program in the United States. These companies actively lobby public officials at all levels to support their work. This often takes the form of surveillance projects carried out under the umbrella of “smart city” initiatives. However, this sometimes leads to disproportionate measures; for example, the schools where hundreds of children are required to use biometric identification to access canteens.

In a country where measuring racial discrimination is a criminal offense punishable by five years’ in prison<sup>1</sup>, few studies exist on algorithmic bias. Despite this limitation, several watchdog organizations, both within the administration and in civil society, regularly document the many instances where automated decision-making goes against the freedom of French citizens. However, the growing list of occasions when their opinion, or findings, are disregarded by the authorities casts doubt on the sustainability of the French model, at least when it comes to automated and democratic decision-making.

## A catalog of ADM cases

### / ALICEM

Although not an automated decision-making system in itself, Alicem is a government program that should allow any part of the French administration system to offer identification via face recognition. Planned for early 2020, it has been put on hold after a public outcry in late 2019.

Citizens register their face’s biometric characteristics using a smartphone app. They can then use the app to go through administrative procedures that currently require them to physically visit a government agency.

Alicem could work as a centralized database of its citizens’ face biometrics (the government is adamant that no biometric data is stored). Even though officially, there are no plans to use the database further, it potentially opens the way for a wide array of applications, not least blanket face recognition by video surveillance cameras in the country’s streets.

<sup>1</sup> Article 226-19 of the French criminal code forbids the collection of racial statistics. A 2007 ruling by the Constitutional Court confirmed that racial statistics were strictly forbidden.

In early September 2019, a security expert revealed that the developers of the app posted some of the code to Stack Overflow (a forum for computer developers), published private project videos publicly on YouTube, and also made the staging servers of the project freely available online (Alderson, 2019). These are all signs of extremely sloppy security procedures for any project, let alone an official biometric database. As was the case with Aadhaar, India's notoriously leaky biometric database, it is almost certain that the data held by France's Alicem will eventually land in the hands of criminals (BBC, 2018).

After Bloomberg reported about the scheme in October 2019 (Fouquet, 2019), the public outcry led to the government delaying the launch of the app. Now, it is expected to launch in 2020, but no official launch date has been made public.

## / Automating surveillance in large cities

In Saint-Étienne (pop. 175,000), the city planned to deploy microphones in order to automatically detect suspicious sounds. The project was to be implemented in a poor neighborhood, in coordination with CCTV cameras and an autonomous drone equipped with a camera. The plan was to register all "suspicious" sounds, including things like gunshots but also electric drills, sprays, and whistles (Tesquet, 2019 ; La Quadrature, 2019).

Another plan, this time in the two largest southern cities of Nice and Marseille, proposed that some high schools should introduce face recognition at the building's entrance. Under the plan, students would pass a face check prior to entering their high school. The program was to be funded and implemented by Cisco, a US company.

However, both projects are on hold after the French data protection authority considered them illegal. The Saint-Étienne microphones would infringe on the privacy of citi-

zens, and the use of face recognition at the entrance to high schools was deemed disproportionate, they said (Hourdeaux, 2019a).

## / The heart score

In early 2018, and following similar measures for patients with liver and kidney problems, French hospitals introduced a "heart score" for patients in need of a heart transplant. Previously, patients waiting for a heart transplant were divided into two categories – "emergency" and "super emergency" – depending on the severity of their condition. However, a review of past practices revealed that one in four patients classed as "super emergency" for a heart transplant was not actually at high-risk. Whereas, a third of those who were high-risk were not in the "super emergency" category.

The current system computes a score each time a new heart is available for transplant. The algorithm that computes the score is transparent and fairly understandable (younger people, for instance, are given more points. And, where there is lower compatibility between the donor and the potential recipient, this results in fewer points), and the score is available for doctors to access online. Based on the score, doctors decide who will benefit from the fresh organ (Guide du Score Cœur, 2018). The court of auditors praised the new system in a 2019 report. However, they cautioned that patient data was often erroneous because hospital personnel had not updated their digital files. The report stated that bogus data concerned up to one in four patients at one hospital. Such flaws endangered the objectivity and acceptance of the system, the auditors wrote (Cour des Comptes, 2019).

## / Machine-learning to detect tax fraud

The finance ministry introduced a legal amendment in the 2020 budget law that will let them scrape data from social networks, websites for classified advertising, or auction platforms to detect tax-fraud. They could, for instance, compare a person's lifestyle on Instagram with their tax returns. Machine-learning is expected to find the patterns of wrongdoing from a training data set. However, the French data protection authority strongly criticized the plan (Rees, 2019a).

The draft law will let the program run for the next 3 years. Several MPs appealed to the Constitutional Court, in the hope of striking down the measure (Rees, 2019b). But, in

THE SAINT-ÉTIENNE MICROPHONES WOULD INFRINGE ON THE PRIVACY OF CITIZENS, AND THE USE OF FACE RECOGNITION AT THE ENTRANCE TO HIGH SCHOOLS WAS DEEMED DISPROPORTIONATE.

a final decision, the court found the scheme to be legal (La Voix du Nord, 2019).

## / Automation at the employment agency

The national employment agency has digitized every procedure in an attempt to speed up the treatment of cases. While the move proved beneficial for simple situations, it created bottlenecks for complex cases. The scanned documents, for instance, are treated by contractors that sometimes classify them wrongly, leading to erroneous decisions. In the end, caseworkers complained that their workload increased, as they must redo by hand many of the automated processes (Guedj, 2019).

## / "Health Data Hub"

In early December 2019, the government created a legal structure to host the "Health Data Hub" (HDH). HDH is a platform that will gather all the data produced by the public health system, and it will make this data available for start-ups and companies that have a project "in the public interest." The project

follows the AI strategy that President Macron defined in 2018, where health was one of four pillars. Public hospitals are unhappy about the move as it takes the data they have independently maintained over the years away from them. The

THEY COULD, FOR INSTANCE, COMPARE A PERSON'S LIFESTYLE ON INSTAGRAM WITH THEIR TAX RETURNS.

fact that the platform is partly-hosted on Microsoft Azure, a cloud computing solution, compounds the fears that sensitive data might be shared with foreign, third-parties. All the information on the "Health Data Hub" is said to be anonymized, but because it might be shared in a non-aggregated format, re-identification could be possible (Hourdeaux, 2019b).

## / Hate speech on social networks

The "Avia" law, largely copied from the German *NetzDG*, was adopted by the French parliament in early 2020. This law would have forced social networks to remove any piece of content identified as hate speech within 24 hours. Given the required speed, such take-downs would have been made automatically (Rees, 2020).

The law also extended the censorship powers of the police. Under current legislation, police officers can censor any website that "praises terrorism" after giving a 24-hour notification to the website's owner. The new law would have reduced the notification period to one hour. French police interpret "terrorism" very loosely, using the charge against left-wing groups, environmentalists, and many Arab or Muslim communities regardless of their intentions.

However, these key points of the law were struck down by the constitutional court in June. The definition of hate speech was extremely vague, including "praise of terrorism" and "publication of pornographic content that minors could see" (Untersinger, 2020).

# Policy, oversight and debate

## / A handful of the administration's algorithms are now open, many more remain closed

The 2016 "Digital Republic" Act states that any citizen can ask to see the rules that define an algorithm used to make a decision. However, since the law became effective in 2017, only a handful of algorithms have been made public, but the administration remains confident that transparency will happen, eventually (Acteurs Publics, 2019).

Change might be coming, as the law states that, starting 1 July 2020, any decision taken by the administration on the basis of a closed algorithm would be considered void.

A researcher analyzed the freedom of information requests containing the word "algorithm" in the archive of the freedom of information authority (*Commission d'Accès aux Documents Administratifs*, CADA). Out of the 25 requests he analyzed (between 2014 and 2018), his main finding was that the French were not aware of their right to be given the rules governing public-sector algorithms (Cellard, 2019).

## / The much-hated automated radars are back

Gilets jaunes, the "yellow vest" protesters, have destroyed over 3,000 of France's 4,500 automated speed cameras, ac-

cording to the government. The radars were a symbol of government overreach for protesting car-owners.

The speed cameras automatically assess the speed of all vehicles, and newer versions of the radars can check for other illegal behavior, such as a driver using a phone while driving. The much-hated devices are coming back as the government plans to deploy 6,000 new, improved units by the end of 2020 (L'Express, 2019).

## Two watchdogs down, one up

Algotransparency and La Data en Clair, two watchdogs and news organizations that featured in the previous *Automating Society* report, ceased activities in 2018. No reason has been given for the discontinuation of either organization.

La Quadrature du Net, together with 26 other civil society organizations, launched [Technopolice](#) in August 2019, a watchdog for “smart city” initiatives. The group calls for “methodical and continuous resistance” against what it sees as an implementation of the surveillance state.

## / Bob Emploi

Bob Emploi – an online service that claimed it could reduce unemployment by automatically matching job seekers with job offers – failed to gain traction in 2019. Approximately 50,000 accounts were created during the year (there are over three million unemployed people in France). While the project is still under development, the claims of algorithm-driven matching have disappeared from the official website. It is now about providing data to “accompany” job seekers in their job search.

## / Parcoursup – Selection of university students

A students’ union sued the government to obtain the source code behind an algorithm that sorts university applicants. A first tribunal granted them their request; however, that decision was later overruled by the Supreme Court for administrative matters (Conseil d’Etat). Although the administration must, by law, provide the details of any algorithm used to make a decision that affects a citizen’s life (as per the 2016 Digital Republic Act mentioned above), lawmakers carved an exemption for university selection. The judges did not assess the legality of this exemption. Instead, they based their decision solely on it (Berne, 2019).

# Key takeaways

France continues to enthusiastically experiment with automated decision-making at both the national and the local level. Watchdog organizations, such as the data protection authority (Commission nationale de l’informatique et des libertés, CNIL) and La Quadrature, are active and visible in public debates, but their power remains limited, especially given heightened political tensions. After CNIL declared the face recognition program in the Provence-Alpes-Côte d’Azur region illegal, local strongman Renaud Muselier [took to Twitter](#) to declare CNIL a “dusty” organization, which ranked the security of students “below its ideology”.

The government followed through on parts of its AI strategy. Of the four pillars put forward in 2018, transportation, security, and health received support in the form of funding or fast-tracking legislation. We could not find any measures related to ADM and the fourth pillar, the environment.

Despite a highly volatile political situation, issues relating to ADM and data collection do still enter into both public and political debate. Stories on Parcoursup were headline news in early 2019. In the fall, the program of mass data collection by the tax authorities, dubbed “Big Brother Bercy” named after the location of the finance ministry, provoked an outcry loud enough for some members of parliament to propose amendments to the project.

Similarly, a project that would have forced all citizens to use face recognition on a mobile app to access e-government services, called Alicem, was put on hold after newspapers and other detractors denounced the project as disproportionate. Although several civil society organizations have been active on the issue, it took the publication of the story in the US media outlet, Bloomberg, for the uproar to begin in earnest (Saviana, 2019).

The lack of political stability may prevent French proposals and experience from feeding the debate across Europe. Mr. Villani, an MP for the presidential party, for instance, wrote the country’s AI strategy in 2018. However, after he declared his candidacy for the Paris mayoral race against his party’s candidate, he is now a political adversary to President Macron.

# References:

- Saviana, Alexandra (2019). Lancement de la reconnaissance faciale en France: mais qu'allons-nous faire dans cette galère ? in: Marianne [online], <https://archive.is/tW4p7>
- Alderson, Eliot (2019). Various tweets, <http://archive.is/dxafE> <http://archive.is/cuveX>
- Algorithmes: les administrations jouent-elles la transparence ? (2019) in: Acteurs Publics. [online] <https://www.acteurspublics.fr/webtv/emissions/numerique-public/algorithmes-les-administrations-jouent-elles-la-transparence>
- BBC (2018) Aadhaar: 'Leak' in world's biggest database worries Indians, in: BBC [online], <http://archive.is/aWjPG>
- Berne, Xavier (2019): Le Conseil d'État s'oppose à la communication des « algorithmes locaux » de Parcoursup, in: Next INpact, [online] <https://web.archive.org/web/20190613160612/https://www.nextinpact.com/news/107971-le-conseil-detat-soppose-a-communication-algorithmes-locaux-parcoursup.htm>
- Cellard, Loup (2019). Les demandes citoyennes de transparence au sujet des algorithmes publics [research note]. [http://www.loupcellard.com/wp-content/uploads/2019/07/cellard\\_note\\_algo\\_public.pdf](http://www.loupcellard.com/wp-content/uploads/2019/07/cellard_note_algo_public.pdf)
- Fouquet, Hélène (2019). France Set to Roll Out Nationwide Facial Recognition ID Program, in: Bloomberg [online], <https://archive.is/ZyZOf>
- Guedj, Léa (2019): Pôle Emploi dématérialisé: le casse-tête des travailleurs précaires, in: France Inter, [online] <https://web.archive.org/web/20191103130035/https://www.franceinter.fr/pole-emploi-dematerialise-le-casse-tete-des-travailleurs-precaires>
- Guide du Score Cœur (2018), Agence de la biomédecine, [online] [https://www.agence-biomedecine.fr/IMG/pdf/guide\\_score\\_coeur\\_v2.pdf](https://www.agence-biomedecine.fr/IMG/pdf/guide_score_coeur_v2.pdf)
- Hourdeaux, Jérôme (2019): La Cnil juge illégale la reconnaissance faciale à l'entrée des lycées, in: Médiapart, [online] <https://web.archive.org/web/20191030190441/https://www.mediapart.fr/journal/france/281019/la-cnil-juge-illegale-la-reconnaissance-faciale-l-entree-des-lycees>
- Hourdeaux, Jérôme (2019): «Health Data Hub»: le méga fichier qui veut rentabiliser nos données de santé, in: Médiapart, [online] <https://web.archive.org/save/https://www.mediapart.fr/journal/france/221119/health-data-hub-le-mega-fichier-qui-veut-rentabiliser-nos-donnees-de-sante?onglet=full>
- La Quadrature du Net (2019): Mouchards et drones à Saint-Etienne : le maire veut étouffer le débat, in: Technopolice, [online] <https://archive.is/A7zfc>
- La Voix Du Nord (2019): Budget 2020: le Conseil constitutionnel censure très partiellement la collecte de données [online] <https://archive.is/xWsUB>
- Rees, Marc (2019):« Collecte de masse » : la CNIL critique le mégafichier de Bercy, in: Next INpact, [online] <https://web.archive.org/web/20191119134909/https://www.nextinpact.com/news/108248-collecte-masse-cnil-critique-megafichier-bercy.htm>
- Rees, Marc (2019): La controverse #BigBrotherBercy se poursuit à l'Assemblée nationale, in: NextINpact, [online] <https://archive.is/dEaaO>
- Rees, Marc (2020): Cyberhaine: mais que prévoit la proposition de loi Avia ?, in: NextINpact, [online] <https://archive.is/MRMNz>
- Sécurité sociale 2019 (2019), Cour des Comptes, [online] <https://www.ccomptes.fr/system/files/2019-10/RALFSS-2019-08-politique-des-greffes.pdf>
- Tesquet, Olivier (2019): Des micros dans la rue : la CNIL tire les oreilles (intelligentes) de Saint-Etienne, in: Télérama, [online] <https://archive.is/aiKtF>
- Untersinger, Martin (2020): La disposition phare de la loi Avia contre la haine en ligne censurée par le Conseil constitutionnel, in: Le Monde, [online] <https://archive.is/GvHIS>
- "Vitesse, téléphone, ceinture: qu'est-ce que le nouveau radar "tourelle" ? " in: L'Express [online] [https://web.archive.org/web/20190826051258/https://www.lexpress.fr/actualite/societe/vitesse-telephone-ceinture-qu-est-ce-que-le-nouveau-radar-tourelle\\_2070713.html](https://web.archive.org/web/20190826051258/https://www.lexpress.fr/actualite/societe/vitesse-telephone-ceinture-qu-est-ce-que-le-nouveau-radar-tourelle_2070713.html)





BELGIUM  
DENMARK  
ESTONIA  
FINLAND  
FRANCE  
**GERMANY**  
**STORY**

PAGE 110

**RESEARCH**

PAGE 114

GREECE  
ITALY  
NETHERLANDS  
POLAND  
PORTUGAL  
SLOVENIA  
SPAIN  
SWEDEN  
SWITZERLAND  
UNITED KINGDOM



Find out more on page p. 115 in the research chapter under "Suicide Prevention in Jail".



# Can AI mitigate the **climate** crisis? Not really.

**Several institutions have claimed that AI will contribute to solving the climate crisis, but the evidence is scant. On the contrary, AI has a track record of helping emit more greenhouse gases.**

By Nicolas Kayser-Bril

"Artificial Intelligence (AI) will change our lives by ... contributing to climate change mitigation and adaptation." So [began](#) the White Paper on AI published by the European Commission on 19 February 2020. On 23 October last year, the German Data Ethics Commission [wrote the same thing](#), stating that "AI holds enormous potential for ... implementing effective climate protection measures."

Inspired by these bold statements, I set out to find examples of how AI contributed to solving the climate crisis. Germany, where AlgorithmWatch is located, was the perfect testing ground.

Germany is a land of contrasts. On the one hand, it is home to the most popular Green party in Europe. On the other, its political class supports Europe's largest carmakers, even though some of them pleaded guilty to conspiring to cheat emissions tests and actively harmed the health of millions of Germans. The country has more wind farms than any other in Europe, but it is the continent's largest emitter of CO<sub>2</sub> (its per capita emissions are largely above the European average, too).

If AI can tip the balance one way or another, it should show in Germany first.

## / Smart washing machines

I started by reading through the 500-page [report](#) of the German Advisory Council on Global Change. It stressed that any strategy to mitigate the climate crisis needed a rethink of the moral values on which the market system is based, but it was much less clear on the specifics. New technology, the authors wrote, could be a "low-carbon enabler" as well as a "power drainer". Indeed, data centers alone consumed about 1% of the global electricity output and produced 0.3% of all CO<sub>2</sub> emissions in 2014, a share that could be higher today. Automated home appliances could reduce energy consumption, but the production of devices in itself leaves a large CO<sub>2</sub> footprint.

In the hope of finding more concrete examples of how AI could contribute to climate preservation, I met with two academics close to the authors of the report. After our one-hour conversation in their shiny offices in the center of Berlin, my hopes were in tatters. One of the very few examples they came up with was – half-jokingly – the smart washing machine. An AI-powered washing machine could be automatically programmed to run at night, they said. This would spread electricity consumption more evenly

over the day, allowing for more efficient production systems.

I'm no scientist, but my experience of washing machines has taught me that they already have a timer and that the reason we don't run them at night is that it's very noisy and will anger neighbors.



"ARTIFICIAL INTELLIGENCE (AI)  
WILL CHANGE OUR LIVES BY ...  
CONTRIBUTING TO CLIMATE CHANGE  
MITIGATION AND ADAPTATION. "  
EUROPEAN COMMISSION

## / Muddy waters

If academia couldn't help, maybe civil society would. Germany is home to the largest Friday For Future group, a worldwide movement of pupils and students who demand the implementation of the Paris Agreement on climate. I asked Annika Rittmann, one of their spokespersons, if she knew of an instance where AI led to a net decrease in greenhouse gases emissions.

"We base our work on scientific results," she wrote back. "In this area, they are still very muddy, so that we can't say much on the topic."

I wasn't getting closer to an answer, but it seemed clear that the European Commission and the Data Ethics Committee might have been overenthusiastic in their statements.

## / Wind turbines

Another civil society organization, GermanWatch (no relation to AlgorithmWatch), published a [report](#) in 2019 on the chances and risks of AI for the *Energiewende*. This is a German concept that literally means "energy turnabout" and concretely means that German consumers pay one third more than the EU average for their electricity in exchange for promises of making energy generation sustainable.

I finally found concrete examples. BirdVision, for instance, is the subsidiary of Bürgerwindpark Hohenlohe, a company operating nine wind farms in southern Germany. It operates



computer vision software to detect birds and automatically reduces the turbine speed if they come too close. Eight turbines are equipped with the system, and the company says that, although the product is still in development, potential users have expressed interest in seeing the system as soon as it is brought to market. This, GermanWatch wrote, could lead to more acceptance of wind energy and encourage the construction of more wind turbines.

Energy efficiency is the area where AI is most promising, according to the report. Smart appliances could be programmed to run only when electricity from renewable sources was abundant, for instance. On the distribution side, “smart grids” could better balance production and consumption, reducing demand at peak times.

### **/ “AI is not an end”**

By now, I knew that report writers could get carried away. I asked a dozen companies active in renewable energy how they used AI.

Next Kraftwerke, which distributes electricity, uses machine learning to forecast the production levels of solar farms and the consumption volumes of its clients. The information is used by the commercial department and by personnel in charge of grid stability. A software developer at the company said that other projects were less successful (they tried to model the energy market, for instance). AI is “not an end in itself,” he said, but “a tool in a wider toolbox.”

Others, such as Naturstrom AG, a production and distribution company with over 250,000 customers, or VSB, a turn-key supplier that has installed over 650 wind turbines and close to 60 solar farms, told me they didn’t use AI at all.

### **/ The efficiency problem**

Mitigating the climate crisis requires reducing the number of greenhouse gases in the atmosphere, either by capturing them, a very expensive and immature technology, or by preventing their release.

While it seems clear that AI can boost energy efficiency, whether or not energy efficiency does anything to mitigate the climate crisis remains an open question. Take light bulbs, for instance. LED technology allows light to be produced without heat so that a 15-watt LED bulb is equivalent to a 120-watt tungsten bulb, an eightfold efficiency gain. However, it’s unclear whether this translates to lower ener-

gy consumption, as cheap LEDs open the way for brighter, interactive advertising screens, among other innovations. They require more than 2,000 watts each, adding up to about six times the electricity consumption of a household each year.

***“AI IS NOT AN  
END IN ITSELF,  
BUT A TOOL  
IN A WIDER  
TOOLBOX.”***

***A SOFTWARE DEVELOPER AT  
NEXT KRAFTWERKE***

The basic tenet of a market economy is that lower consumption for some translates into lower energy costs and higher consumption for others. This is why the German Advisory Council on Global Change warned that a change in rules was needed before discussing any technical solutions.

### **/ Fire detectors**

In the last two years, over 2,000 hectares of German forest went up in flames. A low number compared to southern Europe (43,000 hectares burnt in Portugal alone in 2018), but very high for Germany. Only Sweden and Latvia have seen larger increases compared to the 2008-2017 average.



Considering the fact that each hectare of burnt land releases between 15 and 35 tons of CO<sub>2</sub>, German wildfires were responsible for over 40,000 tons of CO<sub>2</sub> emissions in 2019. Less than one-hundredth of one-hundredth of a percent of all the greenhouse gases emitted in the country that year, but still a consequent amount. And one where the rebound effect does not apply: Preventing a single wildfire is an actual contribution to climate mitigation, not a potential one.

AI can contribute to the fight against wildfires in several ways, from predicting where they'll flare up to detecting fire activity using satellite imagery. The technology is so promising that *Fast Company*, a US magazine, ran an article two years ago entitled "[Future wildfires will be fought with algorithms](#)."

Again, I was hopeful that I could find an example of AI having a net positive effect on climate. I emailed the emergency services and forest management authorities of the German regions most affected by wildfires (Brandenburg, Saxony, and Saxony-Anhalt). I was eager to know how they used or planned to use AI. Alas! They don't. Not that they're technology-shy. They do use sensors and software to detect fires, such as automated smoke detectors, satellite imagery, drones, and infrared cameras, just nothing that involves Artificial Intelligence. (The Brandenburg region does not rule out using AI, but a 2015 [doctoral thesis](#) showed that the system it operates, FireWatch, outperformed all machine learning models).

## / AI for oil

I was about to look into another area where AI might be used to mitigate the climate crisis when Greenpeace ruined my quest. On 19 May, they published [Oil in the Cloud](#), a report on how AI helped energy companies extract more oil and gas, a net contribution to the climate catastrophe.

Even though the report focused on the United States, it revealed that Amazon, Microsoft, and, to a lesser extent, Google, encouraged the fossil fuel industry to use their services. AI could, and did, help them find more hydrocarbons, faster.

In one case, Microsoft noted that its technologies could support production growth by "as much as 50,000 oil-equivalent barrels per day." Once burned in a combustion engine or power plant, that's over 20,000 tons of CO<sub>2</sub> released into the atmosphere, half the amount released by German wildfires in 2019, *every day*, for just *one* project.

## / Getting to the source

Any efficiency gain that AI offers to operators of renewable energy also applies to those who burn hydrocarbons. Because fossil fuel companies are much bigger and more powerful, it seems fair to say that AI contributes to more CO<sub>2</sub> emissions, not less. As long as AI is not demonstrably used in projects that lead to net decreases in CO<sub>2</sub> concentrations, such as carbon capture and storage, stating that it contributes to mitigating the climate crisis is, at best, wishful thinking.

Finally, I sought out the source the European Commission relied on to come to the conclusion that "AI would change our lives by contributing to climate change mitigation and adaptation." After all, they might have access to documents and experts that I might have missed.

A source with knowledge of the matter sent me three links. The first one was a listicle titled "[8 ways AI can help save the planet](#)," written by a PwC consultant, containing zero references and published by the World Economic Forum, a business group famous for its yearly gathering in Davos. The second one was [a blog post](#) based entirely on a report by the same World Economic Forum. The third was [a peer-reviewed article](#) in *Nature*, which stated that AI had the potential to mitigate the climate crisis, but very clearly mentioned that it also had the potential to make it worse.

The lead researchers of the German Data Ethics Commission did not answer my request to see their sources.

"WE WOULD BE HAPPY IF  
CLIMATE MITIGATION USED LESS  
ARTIFICIAL INTELLIGENCE, AND  
MORE HUMAN REASON."  
KARSTEN SMID

I'll leave the last word to Karsten Smid, who's in charge of climate-related issues at Greenpeace Germany. In an email to AlgorithmWatch, he wrote: "We would be happy if climate mitigation used less Artificial Intelligence, and more human reason."

# Research

By Louisa Well

## Contextualization

Issues concerning ADM systems keep coming up in public discussions, but they do not usually make the headlines. ADM is commonly referred to under the label “Artificial Intelligence” (AI) and the lines between digitization, automation, and machine learning tend to be blurry. Discussion around ADM was sparked when the federal police piloted a face recognition system at Südkreuz railway station in Berlin. The program was criticized by civil society actors for its low-quality output and the potential for fundamental rights infringements. Many ADM systems are rolled out in Germany and most of them do not get as much attention as the Südkreuz case.

Risk-prediction applications are on the rise. Within the police force they are used to analyze the threat levels of Islamists under observation or to predict burglary hot spots. The Federal Foreign Office uses software to identify where the next international crisis is likely to occur. Schools and offices use an automated analysis tool to identify risks of violent behavior, be it from a school shooter, a terrorist, or a violent partner. Automated text, speech, and image recognition are deployed to prevent suicides in jail, analyze migrants’ dialects or to identify child pornography. Automation can also be found in the welfare system where it is mainly used to speed up administrative processes.

These cases show that ADM systems are beginning to permeate many different sectors of society, a trend that will intensify over the coming years. At a governmental level in Germany, the course is set for AI applications to be developed and rolled out. The government’s AI Strategy focuses on funding research and supporting

THE FEDERAL FOREIGN OFFICE USES SOFTWARE TO IDENTIFY WHERE THE NEXT INTERNATIONAL CRISIS IS LIKELY TO OCCUR.

small and medium sized enterprises (SME) and start-ups to make Germany a strong player in the European AI landscape.

## A catalog of ADM cases

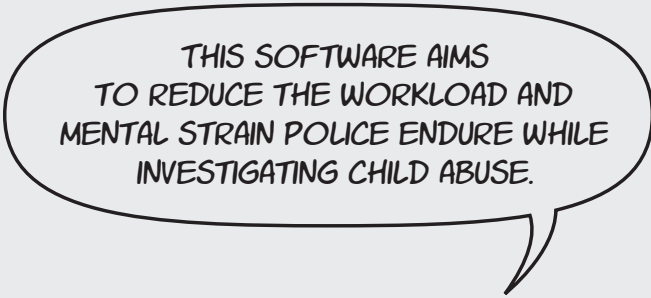
### / Predictive Policing

#### / Federal Level

Since 2017, the Bundeskriminalamt (Federal Crime Agency, BKA) has used the risk-assessment tool [RADAR-ITE](#) (Bundeskriminalamt, 2017) to sort “militant Salafists” into three threat levels (high, conspicuous, and moderate). The system was developed in cooperation with the Department of Forensic Psychology at the University of Konstanz. In order to assess a person already known to the authorities, the caseworker fills in a standardized questionnaire about the “observable behavior” of the subject, drawing on data the police previously gathered on the person and everything the police is legally authorized to access. Once the results of the corresponding threat levels are provided, the caseworker (or the respective department) decides what action to take.

#### / Federal State Level

At the federal state level, several police forces have been running trials with [predictive policing software](#) (Heitmüller, 2019), some of which are currently in use. In Bavaria, the tool [PRECOBS](#) (Institut für musterbasierte Prognosetechnik, 2018) calculates where burglaries are most likely to occur, however the same software was discontinued in Baden Württemberg in 2019 due to [data quality issues](#) (Mayer, 2019). In the field of burglary prevention, the state of Hesse deploys the [KLB-operativ](#) forecast tool (Polizei Hessen, 2017), and Berlin uses [KrimPro](#) (Dinger, 2019). Both of these tools analyze data to identify where a potential break-in is most likely to occur. Based on IBM products, North Rhine-



THIS SOFTWARE AIMS  
TO REDUCE THE WORKLOAD AND  
MENTAL STRAIN POLICE ENDURE WHILE  
INVESTIGATING CHILD ABUSE.

Westphalia developed [SKALA](#) (Polizei Nordrhein-Westfalen, 2020), while Lower Saxony uses [PreMAP](#) (Niedersächsisches Ministerium für Inneres und Sport, 2018); these two tools make predictions about burglary hot spots, which the police incorporate into action plans.

## / Identifying Child Pornography

An [AI tool to identify pornographic images of children was developed in a collaboration between](#) (Richter, 2019) Microsoft and the Ministry of Justice of the State of North Rhine-Westphalia, the Zentral- und Ansprechstelle Cybercrime (Contact Office for Cyber Crime, ZAC NRW) based in the department of public prosecution in Cologne. In order to comply with strict regulations related to data and child protection, software is used to blur images in police investigations on the police servers in Germany before they are uploaded to Microsoft servers in a non-identifiable way. In the cloud, algorithms analyze the images for pornographic content, identify the faces of victims and abusers and compare them to existing profiles stored in a database. These results are returned to the police for further analysis. This software aims to reduce the workload and mental strain police endure while investigating child abuse.

## / Suicide Prevention in Jail

In 2019, the Ministry of Justice of the State of North Rhine-Westphalia launched a program aimed at [preventing suicide in jail](#) (Schmalen, 2019). The ministry employed the Chemnitz-based firm [FusionSystems GmbH](#) (Systems, 2019) to build a video surveillance system that can detect suspicious objects such as a knotted rope or a fire lighter inside a cell and alert officers on duty. The system is supposed to be used on inmates who are categorized at a medium to high suicide risk level and it is meant to substitute the current in-person checks carried out at 15-minute intervals, [which have been criticized](#) (Schmalen, 2019) because they potentially increase the emotional strain on inmates.

## / Crises Management in Foreign Policy

The Federal Foreign Office uses the data analytics tool [PREVIEW- Prediction, Visualization, Early Warning](#) (Auswärtiges

Amt, 2019) to identify evolving international crises. The tool analyzes publicly available data related to current political, economic, and societal trends and conflicts in order to identify developing crises. According to the Federal Foreign Office, AI is used to process the data, which is then used to produce infographics and maps to help provide insights into the state of a particular conflict. Furthermore, trend analyses illustrate how political and societal developments may evolve. *PREVIEW* is deployed by the Federal Foreign Office's *Department S*, which oversees international stabilization measures and crises engagement. The output also supports the civil servants in determining which steps to take next.

## / Identity Check of Migrants

The Bundesamt für Migration und Flüchtlinge (Federal Office for Migration and Refugees, BAMF) has been using automated text and speech recognition systems to [identify refugees](#) (Thüer, Köver and Fanta, 2018) since 2017. Agency employees can ask asylum seekers to give them access to their cell phone, tablet, or laptop to verify if they are telling the truth about where they come from. The agency has the ability to obtain all the data contained on the devices and run software on it. The software presents the employee with a limited overview of the content, which also includes language analysis of the text retrieved. According to the BAMF, both the software and the hardware was provided by the [firm Atos SE](#) (Biselli, 2017), however [VICE Magazine found evidence](#) (Biselli, 2018b) that the mobile forensic technology firm MSAB was also involved. Another tool deployed by the BAMF [aims to identify disguised dialects in speech](#) (Biselli, 2018a). When an asylum seeker does not have a valid proof of ID, a two-minute voice recording of the person describing a picture in their mother tongue is analyzed by software, which then calculates a percentage of how close the speech comes to a certain dialect.

## / Social Services/ Welfare Administration

Since 2012, the Behörde für Arbeit, Soziales, Familie und Integration (Agency for Labor, Social Affairs, Family and Integration) in Hamburg has been using a piece of [software called JUS-IT](#) (Behörde für Arbeit Soziales Familie und Integration, 2018) for the *General Social Service*, the *Child Care Service* and the *Economic Help for Youths Programme*. It is used to administer cases and automate payments and is equipped with interfaces that connect it to police reports and health insurance funds. The system is based on *Cúram*

– a modular off-the-shelf IBM product that can be tailored for specific needs – which has been criticized for [functioning inaccurately in Canada](#) (Human Rights Watch, 2018). In 2018, an [expert commission on child protection](#) (BÜRGER-SCHAFT DER FREIEN UND HANSESTADT HAMBURG, 2018) found that JUS-IT lengthens administrative processes, leaving less time for much needed family visits. As a result, the commission recommends major revisions or a [complete shut-down](#) (Lasarzik, 2019) of the software.

The Bundesagentur für Arbeit (Federal Labor Agency) uses an IT-system called ALLEGRO to [administer unemployment benefit](#). The agency workers input an applicants' data and the system calculates the corresponding benefit levels. The system can connect to health insurance and pension funds and cooperate with customs and the [Central Register of Foreign Nationals \(Ausländerzentralregister\)](#) (Deutscher Bundestag, 2018). In 2012, ALLEGRO replaced the administration's previous software, called A2LL, which was an [error prone system developed by T-Systems](#). The new software was [developed in-house at the Federal Labor Agency](#) (Borchers, 2008).

## / Risk-Scoring of Violence

Similar to the tool used by the Federal Crime Agency described above, researchers from the Institute of Psychology and Threat-Management at the University of Darmstadt have developed a tool called [Screener Islamismus \(Screener Islamism\)](#) (DyRiAS, 2019c). This tool is used to calculate the level of risk a person has of causing violence due to Islamic extremism. The *Screener Islamismus* tool is based on the [DyRiAS](#) IT-system (short for Dynamic Risk Assessment Systems) (DyRiAS, 2019a). The philosophy behind the tool is that violence is not linked to character traits but that it occurs at the end of a process where perpetrator, victim, and situational influences interact. Users answer several questions about the behavior of the person they want to check. The system organizes the answers into an overview and [adds guidance](#) (DyRiAS, 2019c) on how to handle the situation. It remains to be seen, which institution will implement the system as it does not seem to have been built for the police, but other institutions such as schools and offices, etc.

Other [DyRiAS applications are already in use](#). For example (DyRiAS, 2019b), protection centers for women in the cities of Singen and Weimar use *DyRiAS-Intimpartner* (DyRiAS intimate partner) to determine the threat levels of abusive male partners. Another example can be found at the

Landesschulamt Sachsen-Anhalt (Federal State Education Agency). It deploys *DyRiAS-Schule* (DyRiAS school) to analyze the threat levels of students who have the potential to go on a school shooting rampage. In another case, Swiss employers screen employees with *DyRiAS-Arbeitsplatz* (DyRiAS workplace) to look at their potential for violent behavior or if they might become a stalker. After a member of staff fills in a questionnaire about the person they suspect of potentially acting violently, the tools “automatically create a report” that provides the user with a threat score.

# Policy, oversight and debate

## / Policy – Government and Parliament

### / German AI Strategy

The [German AI Strategy](#) (Die Bundesregierung, 2019), published in November 2018, has three key aims:

- 1) To make Germany, and Europe, leaders in the development and deployment of AI and to keep Germany competitive at an international level.
- 2) To ensure the responsible development and deployment of AI for the common good.
- 3) To embed AI ethically, legally, culturally, and institutionally into society, shaped by politics and based on a broad dialogue across society.

The strategy mandates 3 billion euros to fund AI projects and, following a [written query from the Green Party](#) (Deutscher Bundestag, 2019), the Federal Government disclosed how the funds will be distributed.

The lion's share goes to the Federal Ministry of Education and Research (170 million euros). This will consolidate existing German AI research institutions, fund additional professorships, attract international experts to Germany, improve academic teaching and, allow for investment in junior scientists. A slightly smaller portion goes to the Federal Ministry of Economic Affairs and Energy (147 million euros), followed by the Federal Ministry of Labor and Social Affairs (74 million euros) while the rest is divided between seven further ministries and the Federal Chancellery.

As part of the AI Strategy, the German government decided

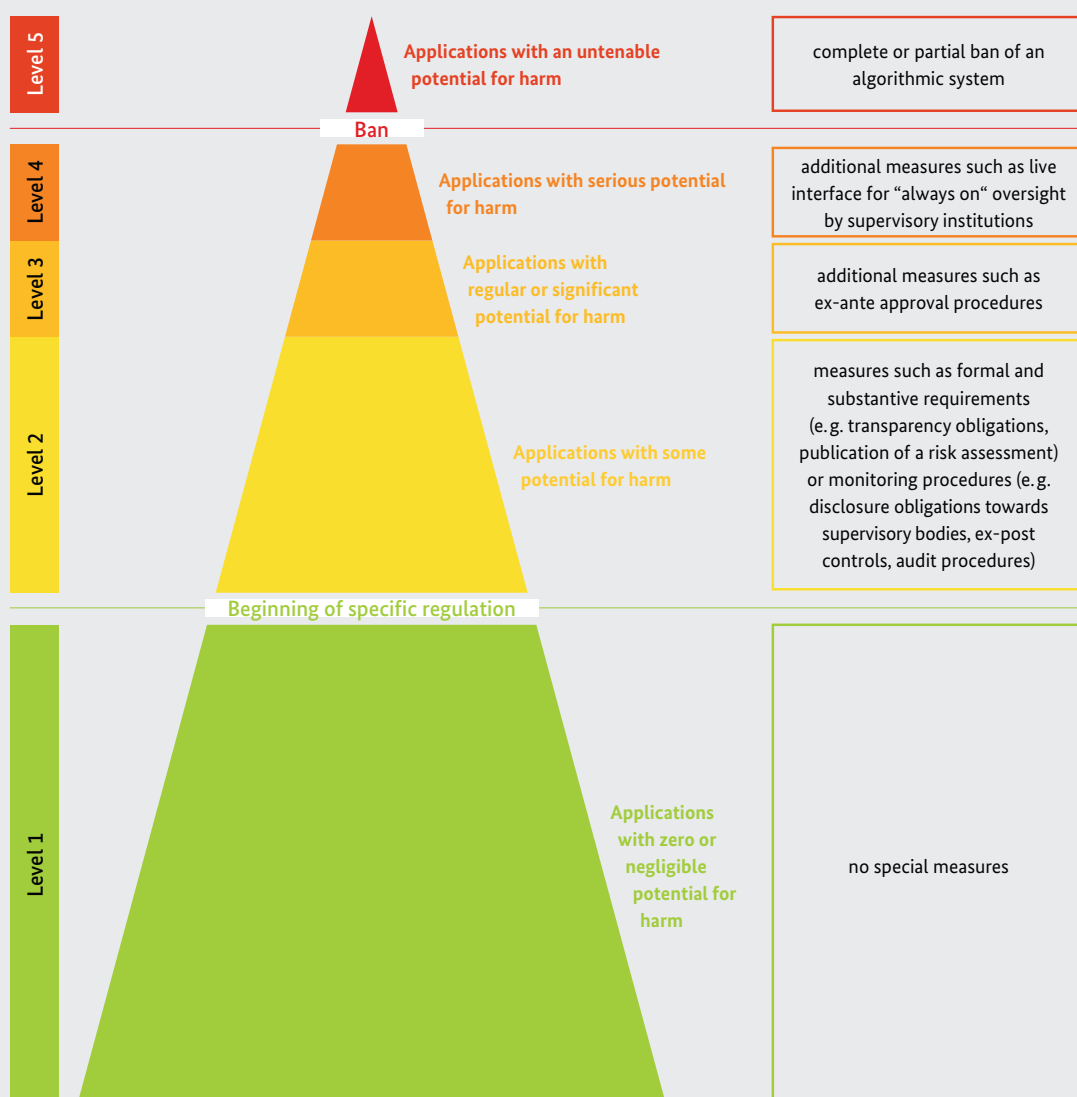
to create an [‘AI Observatory’](#) at the Federal Ministry of Labor and Social Affairs. This will carry out technological assessments of the distribution and effects of AI regarding the labor market and society in general. The aim is to foster European interinstitutional cooperation, develop a shared understanding of AI and design a framework of guidelines and principles for the deployment of AI in the labor sector.

The [AI Strategy will be updated](#), and many projects concerned with issues around research, innovation, infrastructure, administration, data sharing models, international standards, and civic uses are currently in the making.

## / German Data Strategy

In November 2019, the government published the [key points](#) it would like to include in the forthcoming data strategy. The strategy’s goals are to make data more accessible, support responsible data use, improve data competencies in society, and make the state a pioneer in data culture. As data plays an integral part in ADM-systems, the strategy will have a high impact on which applications can be rolled out in the future. The development process of the data strategy was accompanied by [expert hearings](#) with participants from civil society, the economy and science, and the public took part online.

Criticality pyramid and risk-adapted regulatory system for the use of algorithmic systems





## / Data Ethics Commission

In 2018, the German government appointed the [Data Ethics Commission](#) (Datenethikkommission, 2019) to discuss the ethical implications of Big Data and AI. A year later, the Commission presented a report with several recommendations concerning algorithmic systems. A general requirement they called for is that algorithms should have a human-centric design. This includes; consistency with societal core values, sustainability, quality and efficacy, robustness and security, minimal bias and discrimination, transparency, explainability and traceability, and clear accountability structures. Interestingly, they also introduced the concept of “system criticality”, especially with respect to transparency and control. With this approach, they demand to include the likeliness and scope of potential damage caused by algorithmic systems. In addition, the commission recommended that algorithms should be audited according to a “criticality pyramid”. The criticality of an algorithmic system is assessed using a scale where 1 = not critical, no action needed, to 5 = high potential of damage, partial or complete ban of the system. Between these two extremes, they propose heightened transparency obligations, ex-ante approval mechanisms and continuous supervision by oversight bodies.

## / Enquete Commission on Artificial Intelligence

At the end of 2019, one year after the commission was installed in parliament, the working groups published interim results from expert hearings and discussions carried out during the previous year. The [economic working group](#) (Enquete-Kommission Künstliche Intelligenz, 2019c) stated that AI holds the potential to both heighten productivity and improve sustainability as well as to reinforce social injustices and restrict participation in the labor market and society. The [working group on issues around government and the state](#) (Enquete-Kommission Künstliche Intelligenz, 2019b) stressed the duty of care binding the state when rolling out AI systems, and it proposed mechanisms that support the ability of citizens to decide and to trace back decisions. Bearing in mind the fact that health data is very sensitive and needs to be highly protected, the [working group on health](#) (Enquete-Kommission Künstliche Intelligenz, 2019a) sees AI as one potential driver for improved diagnoses, therapies, and care for the sick and elderly.

## / German French Working Group on Disruptive Innovations and Artificial Intelligence

To face global competition in the area of technology and innovation, especially in AI, a [working group](#) (Deutsch-Französische Parlamentarische Versammlung, 2019) on the topic was established between the French and the German parliaments in 2019. The group aims to coordinate the work of the two governments regarding AI issues and to foster a European frame for enhanced innovation capabilities guided by European values.

In 2017/18, a pilot project on face recognition was run by the police at the Südkreuz train station in Berlin. It was [criticized](#) by civil rights and data protection activists and also provided a [high number of false positives and negatives](#). When Horst Seehofer, Minister of the Interior, introduced a bill that would authorize the federal police to roll out face recognition software nationwide, his plans were met with criticism from many parties. Early on in 2020, [Minister Seehofer withdrew the bill](#).

## / Civil Society and Academia

### / Civil Society

Civil society is active in the field of digitalization in general and around ADM processes in particular. An important stream in public discussions focuses on AI and sustainability. Highlights in this debate were the [Bits & Bäume Conference](#) (Bits and Trees) (Bits und Bäume, 2019) in 2018, which aimed to connect both the ecology and technology communities to discuss the interaction between digitization issues and the environment. In 2019, and after analyzing the complex interactions between the two topics, the [German Advisory Council on Global Change](#) (German Advisory Council on Global Change (WBGU), 2019) presented a thorough report called [Towards Our Common Digital Future](#). The report pointed out a gap in the Sustainable Development Goals (SDGs) concerning ADM-processes, and called for traceability, legally enforceable rights, and a regulatory discussion around liability rights concerning algorithmic systems. The Bertelsmann Foundation [Ethics of Algorithms](#) (Bertelsmann Stiftung, 2019b) project developed the [Algo. Rules](#) (Bertelsmann Stiftung, 2019a). These consist of nine guidelines on how to design an algorithmic system. In addition, they also published working papers on a broad range of issues as well as conducting studies on the societal [state](#)

of knowledge and acceptance for algorithms in Germany and Europe (Grzymek, 2019). The [Stiftung Neue Verantwortung](#) think tank (Stiftung Neue Verantwortung, 2019) ran a two-year project on algorithms and the common good, focusing on the importance of a strong civil society, health issues, recruitment, predictive policing, corporate digital responsibility, and the charity sector. The investigative platform [Netzpolitik.org](#) (Netzpolitik.org, 2019) and the watchdog-organization [AlgorithmWatch](#) (AlgorithmWatch, 2020a), continue to report on [new ADM cases](#) (AlgorithmWatch, 2020b). AlgorithmWatch also carries out research projects on the topics of platform governance, human resource management and credit scoring, as well as mapping the state of ADM in Germany and Europe.

## / Academia

Academic discourse on ADM occurs at different levels: some of it focuses on developing ADM systems, while other parts focus on research into ethical implications. An example of the research into ADM in action can be found in the health sector. The Federal Ministry for Economic Affairs and Energy funds two medical projects (under the frame of the [Smarte Datenwirtschaft Wettbewerb](#) Smart Data Economy Competition) (Hans Böckler Stiftung, 2018) in which AI is supposed to be used on patient data. In 2019, the [Smart Medical Doctor – From Data to Decision began](#) (Bundesministerium für Wirtschaft und Energie, 2019). This is a project aimed at developing a medical data platform where AI anonymizes and prepares data from hospital cases to make them available for use in a diagnosis support tool for doctors called Ada DX. Ada Health GmbH, Helios Kliniken GmbH, the University Hospital Charité, and the Beuth University of Applied Sciences are all involved in the project. A second project, called [Telemed5000](#) (Telemed5000, 2019), aims to develop a remote patient management system which allows doctors to manage large numbers of cardiologic patients. This is a [German-Austrian collaboration](#) and it aims to combine Internet of Things (IoT) systems with deep learning technology to monitor heart patients.

There are also several research projects underway that are looking into the societal implications of algorithms in Germany. A few examples worth mentioning are the [Weizenbaum Institut – Research for a Networked Society](#) (Weizenbaum Institut, 2019); the [Algorithm Accountability Lab](#) (TU Aachen, 2019) and the [Socioinformatics Degrees](#) at the Technical University Kaiserslautern; and the newly founded [Politics Department at the Technical University Munich](#) (TUM, 2019).

## Key takeaways

The narratives around ADM run between the poles of technological solutionism, where AI and ADM are praised to be the fix to all our societal problems and technophobic rejections of everything that is new. In many cases, the truth lies somewhere between and it is important to learn more about the intricacies of ADM systems to judge good from bad. Maybe the cure for cancer will be found with the help of machine learning, and ADM can potentially help to manage the task of reducing carbon emissions. At the same time, a ubiquitous datafication of many aspects of life and their incorporation into ADM systems can pose dangers to fundamental rights and must be scrutinized for injustices. One thing is certain; things are going to change. The shifts that ADM systems bring to the economy, science, labor, health care, finance, social services, and many other sectors will need to be accompanied by a society that is ready to critique flaws and, at the same time, embrace changes where they are useful and wanted. For too long, discussions about ADM linked back to autonomous driving or the image of robots replacing humans. It is timely and necessary for society to learn more about the challenges and opportunities that come with ADM systems. At a governmental level, the Enquete Commission on AI is a good start for policy makers to get a better perspective on the topic. Civil society actors such as AlgorithmWatch continuously labor to shed light on the workings of hidden algorithms. What is needed now, is a broad societal discussion about how ADM can make life better and where we see red lines that should not be crossed.

# References:

AlgorithmWatch (2020a)

AlgorithmWatch. Available at: <https://algorithmwatch.org/en/> (Accessed: 17 January 2020).

AlgorithmWatch (2020b) Stories.

Available at: <https://algorithmwatch.org/en/stories-en/> (Accessed: 17 January 2020).

Auswärtiges Amt (2019) PREVIEW.

Available at: [https://www.auswaertiges-amt.de/de/aussenpolitik/themen/krisenpraevention/8\\_Krisenfrueherkennung](https://www.auswaertiges-amt.de/de/aussenpolitik/themen/krisenpraevention/8_Krisenfrueherkennung) (Accessed: 30 November 2010).

Behörde für Arbeit Soziales Familie

und Integration (2018) JUS-IT Hamburg. Available at: <https://www.hamburg.de/jus-it/> (Accessed: 17 January 2020).

Bertelsmann Stiftung (2019a) Algo.

Rules. Available at: <https://algorules.org/en/home/> (Accessed: 17 January 2020).

Bertelsmann Stiftung (2019b) Ethics

of Algorithms. Available at: <https://www.bertelsmann-stiftung.de/en/our-projects/ethics-of-algorithms/project-description/> (Accessed: 17 January 2020).

Bertelsmann Stiftung (2019c):

Predictive policing mit Algorithmen vor die Lage kommen Available at: <https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/predictive-policing-mit-algorithmen-vor-die-lage-kommen/>

Biselli, A. (2017) Syrien oder Ägypten?

Software zur Dialektanalyse ist fehleranfällig und intransparent, Netzpolitik.org. Available at: <https://netzpolitik.org/2017/syrien-oder-aegypten-software-zur-dialektanalyse-ist-fehleranfaellig-und-intransparent/> (Accessed: 17 January 2020).

Biselli, A. (2018a) Die IT-Tools des

BAMF: Fehler vorprogrammiert, Netzpolitik.org. Available at: <https://netzpolitik.org/2018/die-it-tools-des-bamf-fehler-vorprogrammiert/> (Accessed: 17 January 2020).

Biselli, A. (2018b) 'Handys von

Asylbewerbern zu analysieren, kostet viel mehr als geplant', Vice Magazine, April. Available at: <https://www.vice.com/de/article/kzxy8n/handys-von-asylbewerbern-zu-analysieren-kostet-viel-mehr-als-bisher-geplant>.

Biselli, A. (2019) Asylbehörde

sucht mit Künstlicher Intelligenz nach auffälligen Geflüchteten <https://netzpolitik.org/2019/asylbehoerde-sucht-mit-kuenstlicher-intelligenz-nach-auffaelligen-gefluechteten/#vorschaltbanner>

Bits und Bäume (2019) Bits und

Bäume. Available at: <https://bits-und-baeume.org/en> (Accessed: 17 January 2020).

Borchers, D. (2008) 'Hartz-IV-Software:

Allegro soll A2LL ablösen', Heise, 10 March. Available at: <https://www.heise.de/newsticker/meldung/Hartz-IV-Software-Allegro-soll-A2LL-abloesen-189165.html>.

Bundeskriminalamt (2017)

Presseinformation: Neues Instrument zur Risikobewertung von potentiellen Gewaltstraftätern. Available at: [https://www.bka.de/DE/Presse/Listenseite\\_Pressemitteilungen/2017/Presse2017/170202\\_Radar.html](https://www.bka.de/DE/Presse/Listenseite_Pressemitteilungen/2017/Presse2017/170202_Radar.html) (Accessed: 17 January 2020).

Bundesministerium für Wirtschaft und

Energie (2019) Telemed5000. Available at: [https://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/Smarte-Datenwirtschaft-Projekte/smarte-datenwirtschaft\\_telemed5000.html](https://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/Smarte-Datenwirtschaft-Projekte/smarte-datenwirtschaft_telemed5000.html) (Accessed: 17 January 2020).

BÜRGERSCHAFT DER FREIEN UND

HANSESTADT HAMBURG (2018) Bericht der Enquete-Kommission „Kinderschutz und Kinderrechte weiter stärken: Available at: [https://www.buergerschaft-hh.de/parldok/dokument/65251/bericht\\_der\\_enquete\\_kommission\\_kinderschutz\\_und\\_kinderrechte\\_weiter\\_staerken\\_ueberpruefung\\_weiterentwicklung\\_umsetzung\\_und\\_einhaltung\\_gesetzlicher\\_gru.pdf](https://www.buergerschaft-hh.de/parldok/dokument/65251/bericht_der_enquete_kommission_kinderschutz_und_kinderrechte_weiter_staerken_ueberpruefung_weiterentwicklung_umsetzung_und_einhaltung_gesetzlicher_gru.pdf).

Datenethikkommission (2019) Opinion of the Data Ethics Commission.

Available at: [https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten\\_DEK\\_EN.pdf?\\_\\_blob=publicationFile&v=1](https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN.pdf?__blob=publicationFile&v=1) (Accessed: 17 January 2020).

Deutsch-Französische

Parlamentarische Versammlung (2019) Deutsch-Französische Offensive für disruptive Innovationen und Künstliche Intelligenz. Available at: [https://www.bundestag.de/resource/blob/659556/6050efd45a80ac7a7214e2c49d03427a/20190924\\_beschluss\\_arbeitsgruppe\\_riki-data.pdf](https://www.bundestag.de/resource/blob/659556/6050efd45a80ac7a7214e2c49d03427a/20190924_beschluss_arbeitsgruppe_riki-data.pdf).

Deutscher Bundestag (2018) 'Drucksache19/5014'. Available at: <https://dip21.bundestag.de/dip21/btd/19/050/1905014.pdf>.

Deutscher Bundestag (2019) 'Drucksache19/10101'. Available at: <https://dip21.bundestag.de/dip21/btd/19/101/1910101.pdf>.

Die Bundesregierung (2019) KI-Strategie. Available at: <https://www.ki-strategie-deutschland.de/home.html> (Accessed: 17 January 2020).

Dinger, A. (2019) 'Vorhersage-Software: So genau ist das Programm der Polizei', Berliner Morgenpost, 11 September. Available at: <https://www.morgenpost.de/berlin/article216410297/Vorhersage-Software-So-genau-ist-das-Programm-der-Polizei.html>.

DyRiAS (2019a) DyRiAS. Available at: <https://www.dyrias.com/en/> (Accessed: 17 January 2020).

DyRiAS (2019b) Referenzen. Available at: <https://www.dyrias.com/de/referenzen.html> (Accessed: 17 January 2020).

DyRiAS (2019c) 'Screener Islamismus'. Available at: [https://www.dyrias.com/images/Screener\\_Islamismus/DyRiAS\\_Broschuere\\_Screener\\_Radikal.pdf](https://www.dyrias.com/images/Screener_Islamismus/DyRiAS_Broschuere_Screener_Radikal.pdf).

Enquete-Kommission Künstliche Intelligenz (2019a) Projektgruppe „KI und Gesundheit“. Available at: <https://www.bundestag.de/resource/blob/672950/fed938366dcf1b3f79c2ff177e0f86f5/PG-3-Projektgruppenbericht-data.pdf>.

Enquete-Kommission Künstliche Intelligenz (2019b) Projektgruppe „KI und Staat“. Available at: <https://www.bundestag.de/resource/blob/672932/8dafccf73bce9d9560223d1683d82cb9/PG-2-Projektgruppenbericht-data.pdf>.

Enquete-Kommission Künstliche Intelligenz (2019c) Projektgruppe „KI und Wirtschaft“. Available at: <https://www.bundestag.de/resource/blob/672894/e82fdd5cb437de0aaf-0da3d7033f84a3/PG-1-Projektgruppenbericht-data.pdf>.

German Advisory Council on Global Change (WBGU) (2019) Towards our Common Digital Future. Available at: <https://www.wbgu.de/en/> (Accessed: 17 January 2020).

Grzymek, V. (2019) Was Europa über Algorithmen weiß und denkt. Available at: <https://algorithmenethik.de/2019/02/06/was-europa-ueber-algorithmen-weiss-und-denkt/> (Accessed: 17 January 2019).

Hans Böckler Stiftung (2018) Smarte Datenwirtschaft. Available at: [https://www.boeckler.de/pdf/p\\_mbf\\_praxis\\_2018\\_010.pdf](https://www.boeckler.de/pdf/p_mbf_praxis_2018_010.pdf) [https://www.digitale-technologien.de/DT/Redaktion/DE/Kurzmeldungen/Aktuelles/2018/smartedatenwirtschaft\\_20181217.html](https://www.digitale-technologien.de/DT/Redaktion/DE/Kurzmeldungen/Aktuelles/2018/smartedatenwirtschaft_20181217.html) (Accessed: 20 November 2019).

Heitmüller, U. (2019) 'Missing Link: Predictive Policing – die Kunst, Verbrechen vorherzusagen', Heise, 19 May. Available at: <https://www.heise.de/newsticker/meldung/Missing-Link-Predictive-Policing-die-Kunst-Verbrechen-vorherzusagen-4425204.html>.

Human Rights Watch (2018) May 2019 Submission to the UN Special Rapporteur on Extreme Poverty & Human Rights Regarding His Thematic Report on Digital Technology, Social Protection & Human Rights. Available at: <https://www.ohchr.org/Documents/Issues/Poverty/DigitalTechnology/HumanRightsWatch.pdf>.

Institut für musterbasierte Prognosetechnik (2018) PRECOBS in Bayern. Available at: [https://www.ifmpt.de/project\\_bayern.html](https://www.ifmpt.de/project_bayern.html) (Accessed: 17 January 2020).

Lasarzik, A. (2019) "Empathie ist wichtiger als strenge Kontrolle", Die Zeit, 16 January. Available at: <https://www.zeit.de/hamburg/2019-01/kinderschutz-christian-prasser-jugendamt-hamburg-enquete-kommission-yagmur>.

Mayer, N. (2019) 'Strobl entscheidet sich gegen Precobs', Stuttgarter Nachrichten, 3 September. Available at: <https://www.stuttgarter-nachrichten.de/inhalt.aus-fuer-die-einbruchvorhersage-software-strobl-entscheidet-sich-gegen-precobs.19a18735-9c8f-4f1a-bf1b-80b6a3ad0142.html>.

Netzpolitik.org (2019) Netzpolitik.org. Available at: <https://netzpolitik.org/> (Accessed: 17 January 2010).

Netzpolitik.org(2020) Predictive Policing <https://netzpolitik.org/?s=predictive+policing>

Niedersächsisches Ministerium für Inneres und Sport (2018) Zweite Pilotphase startet: Einsatz der Prognosesoftware PreMAP wird zum 1. November weiter intensiviert. Available at: [https://www.mi.niedersachsen.de/aktuelles/presse\\_informationen/zweite-pilotphase-startet-einsatz-der-prognosesoftware-premap-wird-zum-1-november-weiter-intensiviert-170622.html](https://www.mi.niedersachsen.de/aktuelles/presse_informationen/zweite-pilotphase-startet-einsatz-der-prognosesoftware-premap-wird-zum-1-november-weiter-intensiviert-170622.html) (Accessed: 17 January 2020).

Polizei Hessen (2017) Prognose-Software gegen Wohnungseinbruchdiebstahl. Available at: <https://www.polizei.hessen.de/icc/internetzentral/nav/f7b/broker.jsp?uCon=50460710-abca-6f51-a3ac-106409c082ee&uTem=bff71055-bb1d-50f1-2860-72700266cb59&uMen=f7b602efe369-b951-df5a-47e60ef798e7> (Accessed: 17 January 2020).

Polizei Nordrhein-Westfalen (2020) Projekt SKALA (Predictive Policing in NRW) – Ergebnisse. Available at: <https://polizei.nrw/artikel/projekt-skala-predictive-policing-in-nrw-ergebnisse> (Accessed: 17 January 2020).

Richter, I. (2019) Automatische Bilderkennung hilft im Einsatz gegen Kinderpornografie, Microsoft. Available at: <https://news.microsoft.com/de-de/ki-im-einsatz-gegen-kinderpornografie/> (Accessed: 17 January 2020).

Schmalen, L. (2019) 'https://www.nw.de/nachrichten/zwischen\_weser\_und\_rhein/22594348\_Wie-Technik-hilft-Suizide-im-Gefaengnis-zu-vermeiden.html', Neue Westfälische, 11 October. Available at: [https://www.nw.de/nachrichten/zwischen\\_weser\\_und\\_rhein/22594348\\_Wie-Technik-hilft-Suizide-im-Gefaengnis-zu-vermeiden.html](https://www.nw.de/nachrichten/zwischen_weser_und_rhein/22594348_Wie-Technik-hilft-Suizide-im-Gefaengnis-zu-vermeiden.html).

Spiegel (2019): KI soll Kinderpornografie schneller erkennen Available at: <https://www.spiegel.de/netzwelt/netzpolitik/nrw-kuenstliche-intelligenz-soll-kinderpornografie-schneller-erkennen-a-1280480.html>

Stappenbeck, Jens (2017): Frühwarnung weiter denken: Errichtung einer digitalen Vorhersageplattform <https://peacelab.blog/2017/03/fruehwarnung-weiter-denken-errichtung-einer-digitalen-vorhersageplattform>

Stiftung Neue Verantwortung (2019) Stiftung Neue Verantwortung. Available at: <https://www.stiftung-nv.de/en> (Accessed: 17 January 2020).

Systems, F. (2019) Fusion Systems. Available at: <https://www.fusionsystems.de/home.html> (Accessed: 17 January 2020).

Telemed5000 (2019) Telemed5000. Available at: <https://www.telemed5000.de/> (Accessed: 17 January 2020).

Thüer, L., Köver, C. and Fanta, A. (2018) Asylverfahren: Handy-Durchsuchung bringt keine Vorteile, Netzpolitik.org. Available at: <https://netzpolitik.org/2018/asylverfahren-handy-durchsuchung-bringt-keine-vorteile/> (Accessed: 17 January 2020).

TU Aachen (2019) Algorithmic Accountability. Available at: <http://aalab.informatik.uni-kl.de/en/projekte/> (Accessed: 17 January 2020).

TUM (2019) A Unique Profile – Bringing Politics & Technology Together. Available at: <https://www.hfp.tum.de/en/home/> (Accessed: 17 January 2020).

Weizenbaum Institut (2019) Research for the Networked Society. Available at: <https://www.weizenbaum-institut.de/index.php?id=1&L=5> (Accessed: 17 January 2020).







BELGIUM  
DENMARK  
ESTONIA  
FINLAND  
FRANCE  
GERMANY

**GREECE**  
**STORY**

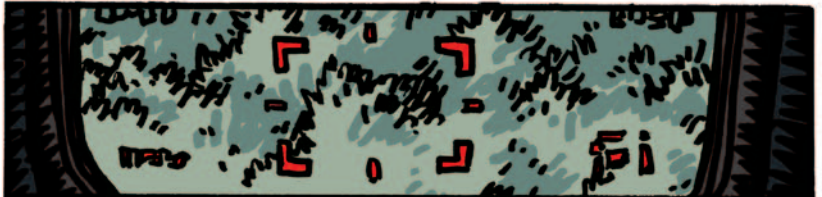
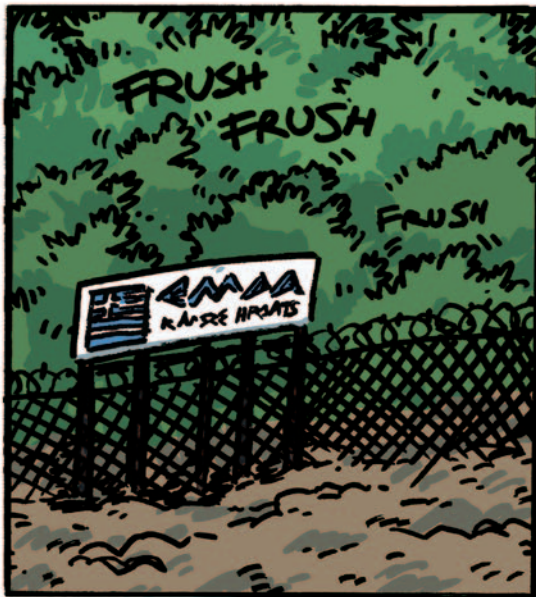
PAGE 126

**RESEARCH**

PAGE 131

ITALY  
NETHERLANDS  
POLAND  
PORTUGAL  
SLOVENIA  
SPAIN  
SWEDEN

SWITZERLAND  
UNITED KINGDOM



Find out more on p. 131 in the research chapter under "Patrolling and surveillance at the borders".

# Broken Horizon:

**In Greece,  
research into  
automation  
fails to find  
applications**

**Greek researchers have led hundreds of research projects involving automation, but very few found their way into real-life uses. Opinions differ as to where the innovation pipeline is clogged.**

By Nikolas Leontopoulos



“Roborder” is a European research project. If its name sounds fancy or dystopian, wait for the tagline: “Autonomous swarms of heterogeneous robots for border surveillance”.

Roborder is endowed by Horizon 2020. The 80 billion euro fund is “the biggest EU Research and Innovation programme ever” according to the European Commission, which oversees the program. Horizon 2020 is divided into thematic “challenges”. Roborder belongs to the “Secure Societies” challenge, which aspires to protect Europeans against a series of threats: Crime and terrorism, natural and man-made disasters, border security, and cyber-security. (The European Commission makes little secret that “Secure Societies” is about preventing asylum seekers from reaching Europe, according to research by Mark Akkerman, who studies border militarization.)

Almost all Horizon 2020 consortia consist of a coalition of research centers, large and small companies, and public bodies, which the European Commission calls “end users”. Roborder’s consortium is made up of 26 partners from 12 countries; its coordinator is CERTH, a Greek research center in Thessaloniki. This should come as no surprise.

## / Greek champions

Greece is only the 9th-largest EU country by population, but Greek researchers rank first in winning Horizon 2020 grants in the field of security, according to data from Horizon 2020’s dashboard. “This is a huge breakthrough – and evidence that the security research ecosystem in Greece is blossoming,” said Dimitris Kyriazanos, a leading researcher for Greece’s top research organization, NCSR Demokritos, in Athens. Mr. Kyriazanos has been Greece’s delegate to the “programme committee” for Secure Societies, an advisory body to the European Commission, since 2019.

Greek organizations were part of several Horizon 2020 projects that developed automated decision-making systems. Thessaloniki’s CERTH is part of Anita, a program to spot criminals active in online marketplaces, and of Tensor, which aims at “retrieving and analyzing heterogeneous online content for terrorist activity recognition.” Kemea received half a million euros to develop Foldout, a project

where machine learning shall be put to use to detect asylum-seekers when they enter the EU overland. Spirit, where the Greek police is an end-user, “will take a novel approach in the development of a new scalable privacy preserving intelligence analysis for resolving identities system prototype (sic).” (The Greek police did not answer requests for clarification.) These are just a few examples, there are many more.

However, despite this flurry of research projects, where Greek research institutions or end-users play a major role, very little evidence of automation can be found in Greece. This raises the question: if Greece is such a prolific producer of AI research, why does it have so little to show for it?

## / Not more than meets the AI

George Vardoulas, a researcher at the American College of Greece, who has extensive research experience both in academia and in the industry, deplores the resistance to embrace automation: “Often automation is not compatible with the end-users’ wishes and stumbles on human prejudice. In Greece, we tend to politicize everything, labeling things as right-wing or left-wing, looking [for] who to blame for the tech, who is behind it. In the past ten years, irrationalism has grown globally, not just in Greece. More and more people see science and engineering as a tool of control and oppression.”

We challenged various stakeholders of the Greek research ecosystem, from research centers to end-users, to provide examples of Horizon 2020 research projects that were transformed into products or have been applied in real life. Almost all our interlocutors had little or nothing to say.

**GREECE IS ONLY THE 9TH-LARGEST EU COUNTRY BY POPULATION, BUT GREEK RESEARCHERS RANK FIRST IN WINNING HORIZON 2020 GRANTS IN THE FIELD OF SECURITY.**

Only one official from the Ministry of Maritime Affairs, one of Horizon 2020’s main end-users, pointed to one emblematic ‘success-story’ involving AI: “Optinet”, which was funded by the European Commission’s regional development fund. Optinet is “an innovative tool to optimize the network of coastal navigation,” according to Athina Foka, the head of the EU Structural Funds service at the ministry. But Optinet started in 2018 and is scheduled to be delivered in 2021. Currently, it does not even have a website, and it is unclear what parts of the project involve automated decision-making.



Mr. Vardoulas, the researcher at the American College of Greece, argues that “major breakthroughs are rare in research by definition. In most cases, EU-funded research does not directly lead to products; but it makes procedures better, it changes the way of thinking, it enhances cross-border collaborations. The closest you can get to a tangible effect is the impact research has on the rationale of designing the products in the next five to ten years. And the only way to achieve this is in the ‘casual’ environment of research.”

## / No blue skies in Greece

But another Greek researcher, who asked to be identified only as “Artemis” and who has worked on several Horizon 2020 projects, strongly disagrees: “This is not ‘blue skies’ research [research projects with no implementation goals]; this is supposed to be applied science. Projects are funded in order to lead to products or platforms, to serve real-life needs.”

Mr. Artemis is vindicated by Horizon 2020’s official motto: “Taking great ideas from the lab to the market”. Even the European Commission acknowledged this shortcoming. In the “interim evaluation” of the Horizon 2020 programme, among the “lessons learned” was the need for more “impact” and “breakthrough innovation”.

If Horizon 2020 does not lead to impact or breakthrough innovation, then where is the added value?

Mr. Vardoulas lays out its implicit rationale: “Since direct funding of companies is prohibited in the EU, the alternative was to find an indirect way to provide funding for the companies so that they remain competitive on a global scale. The US and China support their companies, so the EU should do the same. This is a very competitive process and funds are very limited. But it can provide both an initial boost for start-ups and an incentive for innovation for larger companies.”

## / Greece’s research on life support

The benefits are even more critical for a country such as Greece. Since 2010, hundreds of thousands of highly educated graduates have emigrated. Austerity, imposed by foreign creditors including the EU Commission, exhausted the last drops of public funding to research. Mr. Vardoulas points out that during the crisis years, one of the only imports of capital was via the research programs. They still

fulfill an existential need for university graduates active in AI research, which is extremely difficult in a country with very limited industry.”

Mr. Kyriazanos from Demokritos agrees: “Currently, many highly-skilled executives in Greek companies of the security sector originate from the research ecosystem. This would not have been possible without EU research programmes and the contribution of research institutions.”

However, research programs are not without flaws, says Mr. Vardoulas. “The research projects’ mantra is innovation. But calls are not always compatible with reality, with what is needed and what is feasible. In some cases, this along with the fierce competition leads the applicants to promise far-fetched things. Proposals that sound cool may get promoted at the expense of others that may be [a] better fit for the reality imposed by budget and time constraints.”

## / Short life spans

Mr. Kyriazanos says: “At the end of each project, there will be some tangible results that could be placed on track for commercialisation. However, if there are no national grants, support, or calls to assist the researcher, they will move on to ensure funding for the next project. If there are no favorable conditions for the ‘next logical step’ in innovation, results either remain on a low maturity prototype or worse: they are shelved. This means lots of lost opportunities.”

In other words, Greece might only lack a way to take automated decision-making from the lab to the streets. But not all agree.

In 2018, a wildfire at Mati on the outskirts of Athens cost 102 lives. It was the second-deadliest forest fire in the 21st century after the 2009 bushfires in Australia. Mr. Kyriazanos says: “It makes you wonder if things could have turned out differently. In 2014 we ran a big drill in forest firefighting in the framework of AF3” (Advance Forest Fire Fighting), a research project financed by the European Commission under Horizon 2020’s predecessor, called FP7. AF3 included several automated systems to predict the evolution of fires and evacuate communities. “And this has been hanging over me, Mr. Kyriazanos added. It was never used – actually, even if we offered it for free to the Fire Service, they still would not use it.”

In the aftermath of the wildfires, it was reported in the Greek press that several projects and platforms that pro-

vided an automated response to such incidents and that had been developed at a cost of dozens of millions had been cast aside.

## **/ “Researchers should get out of their labs”**

One of the main end-users of Horizon 2020 projects is the Greek Coast Guard. Ms. Foka, of the Ministry of Maritime Affairs, oversees the European research projects. She is critical of the work done by research centers, and not just in Greece. “Researchers should get out of their labs and see the real needs. If the deliverable is not tailored to the needs of the end-user, then nobody’s going to use it.”

But Mr. Artemis, who has dealt with the Coast Guard within Horizon 2020 projects, tells a completely different story. “By design, the end-user is tasked to define specifications for the project. The Coast Guard rarely does so, and in the end, researchers themselves have to define the specifications.”

Almost two years after the end of a Horizon 2020 project in the field of security, Mr. Artemis received a call from a Coast Guard officer. The officer asked the researcher on which workstation the software had been installed. Mr. Artemis realized that for almost two years the so-called ‘deliverable’ had been standing idle in one of the ministry’s computers. The ministry’s own staff probably did not know where it had been installed, and certainly did not use it.

## **/ Erasmus for grown-ups?**

Mr. Artemis says: “The sad reality is that behind a significant proportion of Horizon 2020 projects on AI, there is nothing but hot air. The acronyms are cool, the wording too but what’s really cool is the lifestyle! European research institutions and companies are in for the money. As for individuals, both in the research institutions and the end-users, it’s about a lifestyle of free travel, food, and socializing around Europe. It’s an ecosystem addicted to EU funding; without it, it would die.”

**CURRENTLY, MANY  
HIGHLY-SKILLED  
EXECUTIVES IN GREEK  
COMPANIES OF THE  
SECURITY SECTOR  
ORIGINATE FROM  
THE RESEARCH  
ECOSYSTEM. THIS  
WOULD NOT HAVE  
BEEN POSSIBLE  
WITHOUT EU RESEARCH  
PROGRAMMES AND  
THE CONTRIBUTION  
OF RESEARCH  
INSTITUTIONS.**

Mr. Vardoulas rejects this approach as an over-simplification, but he asserts that, for some European companies, funding depending on EU research grants becomes an end in itself, as if this were the product: “It’s a Europe-wide problem. There are many start-ups across the continent that exist only to attract EU funds. In reality, they have no marketable product. What is their activity? European research funding.”

## **/ KEMEA, a rookie that tops the charts**

Historically, research in Greece has mostly been done by publicly-funded research centers such as NCSR Demokritos in Athens, the Centre for Research and Technology Hellas (CERTH) in Thessaloniki, and the Foundation for Research and Technology-Hellas (FORTH) in Crete. But in the last few years, there’s a new entry on the scene: KEMEA, the Centre for Security Studies.

Despite being a rookie, KEMEA has grown to become the biggest recipient in total funding for the entire Horizon 2020 Secure Societies program. For the years 2018-2020, it has received more than 24 million euros when the second European recipient lags behind with 11 million euros. (KEMEA’s share is so big that it still ranks first even when earlier years are taken into account.)

KEMEA is a hybrid. From an administration point of view, it is fully controlled by the Ministry of Citizen Protection (the equivalent of the Ministry of the Interior, also responsible for the police) which appoints its seven-member board of directors and its executive director. From a legal point of view, KEMEA is a private structure, which frees it from some public scrutiny. Although it answers directly to the government, KEMEA is a member of the European Organisation for Security, the European lobby representing corporations such as Airbus, Thales, and G4S.

KEMEA has participated in more than 100 European research programs. To what avail?

We tried to approach KEMEA for months in order to understand the impact of the many projects it contributed to in

the field of security and automated decision-making. Despite communications via phone, email, and a preliminary meeting with its executive director, KEMEA declined to answer questions.

We also addressed questions to three of the most important end-users in Greece: the Ministry of Defense, the police, and the coast guard. Only the latter responded – and not in writing but over a casual phone conversation.



***RESEARCHERS SHOULD GET OUT  
OF THEIR LABS AND SEE THE REAL  
NEEDS. IF THE DELIVERABLE IS NOT  
TAILORED TO THE NEEDS OF THE  
END-USER, THEN NOBODY'S GOING  
TO USE IT.***

# Research

By Eleftherios Chelioudakis

## Contextualization

The discussions in Greece about automation have become intense both at the institutional and the public level. According to the Digital Government factsheet for Greece ([European Commission, 2019](#)), the Greek government – in cooperation with the Ministry of Digital Policy, Telecommunications and Information – aims to drastically transform the Hellenic State. This involves simplifying procedures and implementing big changes, especially in the field of new technology.

Moreover, with its National Digital Strategy ([Hellenic Ministry of Digital Policy, Telecommunications & Media, 2016](#)), Greece aims to kick-start its economic development, create new jobs, enhance the efficiency of public administration, eliminate geographical and social exclusion, and strengthen the inclusion of all citizens in the economic and social life of the country. This strategy focuses on seven priorities, with an emphasis on the digitalization of public governance and the strengthening of security.

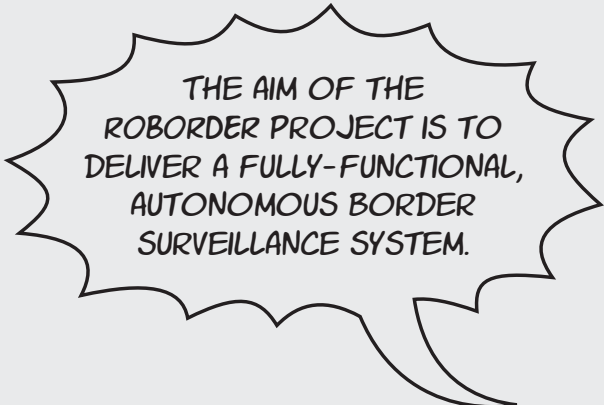
The Greek government has not yet adopted a national AI strategy. However, ADM systems are thriving at the national level, especially on research projects, and there are a considerable amount of use-cases. Academic institutions and scientific research centers in Greece focus their activities on the development and deployment of ADM tools based on national grants, funds from the European Union (such as the Horizon 2020 programs), and collaboration with private industry. Finally, the Greek state is struggling to keep up with the technological developments in ADM and reflect on the challenges that arise by participating in common initiatives at the EU and Council of Europe level, and by establishing scientific committees for the examination of the role of algorithms and AI.

## A catalog of ADM cases

### / Patrolling and surveillance at the borders

The Center for Research and Technology Hellas (ITI/CERTH), the National and Kapodistrian University of Athens (UoA), and the Hellenic Ministry of Defense participate in the research project consortium ROBORDER (“autonomous swarm of heterogeneous Robots for BORDER surveillance”). The aim of the ROBORDER project is to deliver a fully-functional, autonomous border surveillance system. The system consists of unmanned mobile robots including aerial, water surface, underwater and ground vehicles, capable of functioning both on a standalone basis and in swarms. The project started in May 2017 and will conclude in February 2021. Twelve other countries are also participating in this consortium: Portugal, United Kingdom, Romania, Germany, Italy, Switzerland, Bulgaria, Spain, Hungary, Belgium, Finland, and Estonia ([ROBORDER, 2017](#)).

Furthermore, a considerable number of public and private stakeholders in Greece are taking part in a research project called TRESSPASS (“robust Risk basEd Screening and alert System for PASSengers and luggage”). The Greek National Center for Scientific Research “Demokritos” is coordinating the project, while six other Greek actors are part of the project’s consortium. These include the Piraeus Port Authority, the Greek Ministry of Maritime Affairs and Insular Policy, the Hellenic Customs Administration, and the Center for Security Studies (KE-MEA), which is a research center within the Hellenic Ministry of Citizen Protection. TRESSPASS promotes the idea of



THE AIM OF THE ROBORDER PROJECT IS TO DELIVER A FULLY-FUNCTIONAL, AUTONOMOUS BORDER SURVEILLANCE SYSTEM.

“risk-based” security checks and offers an analytic framework for modelling risk as well as a systematic approach of quantifying risk, based on a set of indicators. Once an estimate of these indicators is made using data collected from background information, sensors and applications, a level of risk is calculated for each traveler, based on the risk management approach introduced in the TRESSPASS project ([TRESSPASS, 2018](#)).

In the field of border management, another research project that has attracted attention over the past year is the iBorderCtrl project (“Intelligent Portable Control System”). According to the project’s website, the aim of this project is to enable faster and more thorough border control for third country nationals crossing the land borders of EU Member States. iBorderCtrl was implemented on the Hungarian, Greek, and Latvian borders as a pilot project, but it is not an authorized law enforcement system. KEMEA and the Train-OSE S.A., a Greek railway company, participated in piloting the project in Greece. The project came to an end in August 2019 ([iBorderCtrl, 2016](#)). In November 2018, the Greek civil society organization Homo Digitalis filed a petition to the Greek Parliament regarding the pilot implementation of the iBorderCtrl project at Greek borders (Homo Digitalis, 2018). Finally, based on a successful access to information request it was revealed that during the pilot project in Greece the role of travelers was primarily “played” both by the Border Guards and Hellenic Police Officers. Furthermore, staff from KEMEA and other consortium members voluntarily played the role of travelers. Thus, it appears that no real travelers participated in the Greek pilots ([iBorderCtrl, 2019](#)).

Finally, KEMEA is also involved in another research project – related to the use of ADM for border management – known as FOLDOUT (“Through-foliage detection, including in the outermost regions of the EU”). The aim of FOLDOUT is to build a system which combines various sensors and technologies and fuses them into an intelligent detection platform for border surveillance. According to FOLDOUT’s website, the project will improve the speed of the work of border guards by combining events from various sensors to give a situation threat assessment followed by suggested reaction scenarios. According to the project’s website, a two-year pilot project in Bulgaria and demonstrations of the project in Greece, Finland, Lithuania, and French Guiana are all planned. However, there are no details on the exact dates of such demonstrations. Nevertheless, the goals are described as follows: detection of irregular border crossings (illegal migrants + vehicles) in forest terrain border surveillance for the Bulgarian and Greek demonstrations, and

detection of illegal transport and entry of goods (trafficking) in temperate broadleaf forest and mixed terrain border surveillance for the demonstrations in Finland, Lithuania, and French Guiana ([FOLDOUT, 2018](#)).

## / Automated Car Insurance

In November 2017, two companies – Interamerican S.A. and OSEVEN – informed the Hellenic Data Protection Authority (HDPa) about their car insurance program called “Smart Drive”. The price of the insurance for the “Smart Drive” program is determined by the driver’s behavior based on an application installed in the driver’s smartphone. Following two years of wide consultations between the HDPa and the companies, the HDPa announced they would postpone the adoption of a decision on this case until the European Data Protection Board has finalized its guidelines on processing personal data in the context of connected vehicles and mobility related applications ([Hellenic Data Protection Authority, 2020a](#)). The related decision that the HDPa will hand down in the coming months is awaited with much interest.

## / Autonomous driving

In 2015, the city of Trikala was in the spotlight because of a self-driving, 10-passenger electric bus. The autonomous vehicle was part of a European Union-funded project called CityMobil2 (CM2). The maximum speed of the bus was only 20 kilometers per hour, and it could not change lanes, however it stopped when any obstacles, such as parked cars, got in its way ([Rainey, 2018](#)). Now, the city is looking into autonomous driving a little further. AVINT (“Autonomous vehicles integrated within the urban context”) is a project set up to study urban transport in Trikala. It will integrate a line of automated buses – three buses in total, carrying 10 passengers each – into the city’s transportation network. This automated bus line will connect Trikala city center with the university campus. The project started in January 2019 and will be completed in December 2021. According to the project’s website, Greece is a pioneer in autonomous driving, as it has had a legal framework in place since 2014. The framework allows for the operation of automated buses without a driver under strict conditions. The partners in the project are the e-Trikala company, owned by the municipality of Trikala, the Institute of Communications and Computer Systems (ICCS) of the School of Electrical and Computer Engineering at the National Technical University of Athens, and the private company Space Hellas ([AVINT, 2019](#)).



Lastly, the University of Patras, ITI/CERTH, and the Foundation for Research and Technology Hellas (FORTH) took part in the ADAS&ME project ("Adaptive ADAS to support incapacitated drivers & Mitigate Effectively risks through tailor made HMI under automation").

The aim of the research project is to develop robust detection/prediction algorithms to monitor the physical and emotional state of the driver/rider to mitigate accidents related to driver fatigue, sleepiness, stress, inattention or high emotions. The project focused on seven use-cases, covering a wide variety of critical driving scenarios, using different types of vehicles (cars, trucks, buses, and motorcycles). The results of the project were demonstrated in December 2019 and the project came to an end in February 2020 ([ADAS&ME, 2016](#)).

THE ANPR CAN RECOGNIZE  
15,000 LICENSE PLATES PER HOUR,  
AND THE CAMERAS CAN DETECT  
LICENSE PLATES AT SPEEDS ABOVE  
170 KM PER HOUR.

Finally, the pharmaceutical company Pfizer has chosen the city of Thessaloniki for the creation of one of its six digital laboratories, aimed at capitalizing on technologies such as AI and big data analytics for the development of new treatments. The investment was announced in September 2019 by the Greek Prime Minister, Mr. Kyriakos Mitsotakis, at the 84th edition of the Thessaloniki International Fair (TIF), the largest exhibition event in the Balkans ([Konti, 2019](#)).

## / Healthcare and Medicine

NCSR Demokritos is the coordinator of the research project IASIS ("Integration and analysis of heterogeneous big data for precision medicine and suggested treatments for different types of patients"). In addition, the private Greek company Athens Technology Center also participates in this project. The goal of IASIS is to integrate data from various sources, including genomics, electronic health records and bibliography, and apply advanced analytical methods to discover useful patterns. Data resources for two different disease categories, namely dementia, and lung cancer, will be explored during the project's implementation phase. It is not known whether and how the consortium aims to integrate such a system into clinical decision making ([IASIS, 2017](#)).

Furthermore, Aristotle University of Thessaloniki (AUTH) coordinated the research project i-PROGNOSIS ("Intelligent Parkinson early detectiOn Guiding NOvel Supportive Interventions"), while three other Greek stakeholders, including ITI/CERTH, the Greek telecommunications company COSMOTE, and the Microsoft Innovation Center – Greece, participated in it. Among the aims of the project was a plan to develop and employ advanced big data analytics and machine learning techniques, in a distributed and privacy-aware fashion, to create a Parkinson's Disease Behavioral Model and construct reliable early symptoms detection alarms. The project concluded in January 2020 ([i-PROGNOSIS, 2016](#)).

## / Automated processing of traffic offenses and Automatic Number Plate Recognition cameras

Automated processing of traffic offenses was introduced in Greece back in 2001 based on Presidential Decree 287/2001. However, in recent years, because of the financial crisis, the purchase of photo radars has been limited. As a result, the number of available traffic enforcement cameras on national highways is low. According to a 2017 interview with the then Deputy Minister for Citizen Protection, Mr. Nikolaos Toskas, traffic enforcement cameras were only present on a single section of the entire national highway network. However, 14 traffic enforcement cameras have been installed on the Attica toll road, which is a privately owned motorway system ([Kathimerini, 2017](#)). It is also worth noting that the number of normal traffic cameras on national highways is high, while police officers use manually operated tools to detect drivers who exceed the speed limit ([Lialos, 2017](#)).

In addition, Automatic Number Plate Recognition (ANPR) cameras were openly introduced in 2019. According to a Greek news report, around 50 Hellenic police vehicles are equipped with such cameras and they are capable of detecting the license plates of stolen vehicles or vehicles used in criminal activities. The ANPR can recognize 15,000 license plates per hour, and the cameras can detect license plates at speeds above 170 km per hour. In the near future, they will also be able to detect uninsured vehicles ([To Vima, 2019](#)).

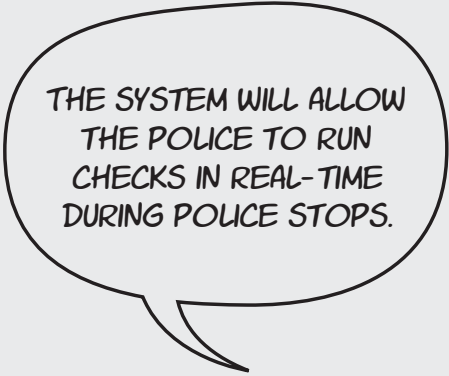
## / Predictive Policing

Greece participates in several predictive policing projects. Firstly, in the spring of 2019, Intracom Telecom, a global

telecommunication systems and solutions vendor, signed a four million euro contract with the Hellenic Police and the Hellenic Ministry of Citizen Protection for the development of “Smart Policing” software. The contract states that Intracom Telecom will develop an integrated information system which will include the core building blocks for face recognition, automated fingerprint identification, document authentication, and complex information searches in existing databases. The system will allow the police to run checks in real-time during police stops ([Intracom Telecom, 2019](#) and [Hellenic Police, 2019](#)). Based on the implementation timetable of the project, the vendor will deliver the system to the Hellenic Police within 20 months of signing the contract ([Hellenic Police 2018](#)).

Secondly, the Hellenic Police and four other law enforcement authorities (i.e. the West Midlands Police (UK), the Police and Crime Commissioner for Thames Valley (UK), the Serbian Ministry of Interior (RS), and the Police Academy in Szczytno (PL)) are taking part in a project called SPIRIT (“Scalable privacy preserving intelligence analysis for resolving identities”). The information available on the project’s website is very limited, even though the project has been up and running since August 2018. Nevertheless, it is evident from the website that the project aims to use tools such as face extraction and matching, to correlate information from social media data, and to continuously initiate complex associative searches across all sources relevant to criminal investigation. A successful freedom of information request revealed that trial runs will be organized, including in Greece, involving end users and practitioners, such as law enforcement agencies, and other stakeholders. The Hellenic Police (Border Unit) is involved in trials that run between January and August 2020, while reports of trial runs and tests are expected in the form of deliverables in April 2020 and January 2021 (Ask the EU, 2020). Two of the Greek stakeholders in SPIRIT’s consortium are the private companies Singular Logic and Nydor System Technologies. The project is due to end in July 2021 ([SPIRIT, 2018](#)).

Thirdly, the Center for Security Studies (KEMEA) is a member of the COPKIT project, which is developing data-driven policing technologies to support law enforcement agencies (LEAs) and help prevent, investigate and mitigate crime and terrorism. Europol and nine LEAs from eight different countries are members of this project as well. The first private demo of the COPKIT tools took place in Athens in November 2019 at the premises of KEMEA. A second private demo was held in Madrid in May 2020 and the third will take place in Brussels in November 2020, fol-



THE SYSTEM WILL ALLOW  
THE POLICE TO RUN  
CHECKS IN REAL-TIME  
DURING POLICE STOPS.

lowed by a public demonstration at the project’s final conference ([COPKIT, 2018](#)).

Furthermore, ITI/CERTH is a participant in the ANITA research project (“Advanced tools for fighting oNline Illegal TrAfficking”), which started in May 2018 and is set to conclude in April 2021. The aim of the project is to design and develop a knowledge-based user-centered investigation system for analyzing heterogeneous (text, audio, video, image) online and offline content for fighting the trafficking of drugs, counterfeit medicines, and firearms. According to the project’s website ([ANITA 2018](#)), ANITA will combine the following benefits:

- a. data source analysis of crypto-currency networks, transactions, and blockchain technologies
- b. advanced big data analytics tools for the automatic analysis of vast amounts of multimodal content from identified sources
- c. methodologies for capturing, modelling and inferring knowledge in human understandable forms (e.g. expressive ontologies), also extracting relevant and new knowledge from neural networks and formally storing it in the form of ontologies.
- d. domain-related and user-oriented intelligence applications, which will enable users to identify patterns for correlations among illegal trafficking events, entities and activities, and will support decision-making processes for countermeasures.
- e. lastly, the project’s goal is to develop a cognitive user modelling framework that will capture, analyze, interpret, mimic, and integrate key human cognitive and information processing functions to represent the incorporation of human perception/cognition principles in the system processing pipelines (i.e. integrating the investigators/officers/practitioners ‘in-the-loop’ of the overall analysis process).

Finally, ITI/CERTH also participates in a project called CONNEXIONS (“InterCONnected NEXt-Generation Immersive IoT Platform of Crime and Terrorism DetectiON, PredictiON, InvestigatiON, and PreventiON Services”), which started in 2018 and will end in August 2021. The aim of the project is to enhance operational and (near) real-time situational awareness of the consortium’s law enforcement authorities, through automated identification, interpretation, fusion and correlation of multiple heterogeneous big data sources. Such multimodal data includes surface, deep, and dark web, as well as social media content in seven languages (English, French, German, Portuguese, Romanian, Spanish, and Arabic). Other sources include data acquired by IoT devices and digital evidence ([CONNEXIONS, 2018](#)).

## Policy, oversight and public debate

### / Policy

Establishment of the Legislative Committee for the transposition of the new Copyright Directive and its controversial.

### / Article 17

In September 2019, Ms. Lina Mendoni, Greek Minister of Culture and Sports, issued her decision to establish a legislative committee entrusted with the transposition of Directive 2019/790 on copyright and related rights. There is great interest in the transposition of the controversial Article 17, which relates to the use of content moderation software in the Greek legal framework. The Committee is scheduled to deliver its work by the end of June 2020 ([Hellenic Ministry of Culture and Sports, 2019](#)).

### / Greece at the Ad Hoc Committee on AI of the Council of Europe

In September 2019, the Committee of Ministers of the Council of Europe set up an Ad Hoc Committee on AI (CAHAI). The Greek delegation at the CAHAI is composed of representatives from the Ministry of Foreign Affairs and the Ministry of Digital Governance. The Greek digital rights organisation Homo Digitalis is an observer of the CAHAI since June 2020, as well. The Committee aims to examine the feasibility and potential elements of a legal framework for the development, design and application of AI, based on the

Council of Europe’s standards on human rights, democracy and the rule of law ([Council of Europe, 2019a](#)).

### / Greece signs the Protocol to amend Convention 108 of the Council of Europe

In September 2019, ambassador Mr. Panayiotis Beglitis, Permanent Representative of Greece at the Council of Europe, signed the protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data ([Council of Europe, 2019b](#)). The new Article 9 of the Modernized Convention 108 states that every individual shall have a right not to be subject to a decision significantly affecting him or her based solely on the automated processing of data without having his or her views taken into consideration ([Council of Europe, 2018a](#)). According to the protocol’s explanatory report, the data subject should have the chance to argue the irrelevance of the profile to be applied to his or her particular situation, or other factors that could influence the result of the automated decision. For instance, when the application of algorithmic reasoning ends up limiting an individual’s right to social benefits ([Council of Europe, 2018b](#)).

### / Greece finally enforces the GDPR and Law Enforcement Directive, but raises criticism

In August 2019, the Greek Parliament adopted the Law 4624/2019, which enforces the provisions of the GDPR and implements the Directive 2016/680 (the Law Enforcement Directive – LED) into the national legal framework. Article 52 of L.4624/2019 transposes Article 11 of the LED (Automated individual decision-making). In September 2019, the Greek digital rights organization Homo Digitalis and the Greek consumer rights organization EKPIZO challenged the clarity of L.4624/2019 and requested the Hellenic Data Protection Authority to issue an Opinion on its conformity with the provisions of the GDPR and the LED ([Homo Digitalis & EKPIZO, 2019](#)). Also, in October 2019, Homo Digitalis filed a complaint (CHAP(2019)03059) to the European Commission against a breach of the provisions of the EU data protection law (GDPR and LED) by Greece and L.4624/2019. According to Homo Digitalis, while Article 52, para.1 of the L.4624/2019 states that a decision based solely on automated processing, including profiling, which produces an adverse legal effect concerning the data subject or significantly affects him or her is only allowed when it is authorized by law. It does not state that such a law shall provide appropriate safeguards for the

rights and freedoms of the data subject, not even the right to obtain human intervention on the part of the controller. This omission is very serious and could jeopardize the rights of the data subjects in the age of technology-led predictive policing ([Homo Digitalis, 2019a](#)). The Hellenic DPA issued its Opinion on L. 4624/2019 in January 2020, which heavily criticized the lack of conformity of some of its provisions with the GDPR and the LED. Concerning Article 52, the Hellenic DPA stated that Article 11 LED has been poorly transposed because the L. 4624/2019 does not provide the appropriate safeguards for the rights and freedoms of the data subject, at least the right to obtain human intervention on the part of the controller ([Hellenic Data Protection Authority, 2020b, p.23](#)).

### **Scientific Committee to examine the impact of AI in the Greek Judicial System**

In February 2019, Mr. Michael Kalogirou, the then Minister of Justice, Transparency, and Human Rights, issued a Ministerial Decision to establish a scientific committee entrusted with the task of examining the impact of AI on the Greek judicial system. The scientific committee is composed of members of the Greek Courts and Supreme Courts, academics, members of Greek law bars, and civil servants. The committee will submit proposals, inter alia, to the ministry, for the modernization of the existing legal framework taking into account technological developments ([Hellenic Ministry of Justice, Transparency, and Human Rights, 2019](#)).

### **/ The potential impact of AI on human rights in the Constitution?**

In February 2019, during the Greek Parliament's Plenary Session on the revision of the Greek Constitution, Mr. Yianis Maniatis acting as a rapporteur for the Movement for Change (KINAL) political party, proposed an amendment to Article 25 of the Greek Constitution regarding the potential impact of AI on the rule of law, and respect for human rights and fundamental freedoms. The suggestion was not voted through ([Maniatis, 2019](#)).

### **/ Greece signs a motion for the examination of the role of algorithms and AI in the criminal justice system**

In September 2018, Ms. Ioanneta Kavvadia, the then representative of Greece at the Parliamentary Assembly (PACE) of the Council of Europe, signed a joint motion with more

than twenty other representatives. The petition requested the PACE to develop a recommendation examining the role of algorithms and AI in the criminal justice system bearing in mind the Council of Europe standards on human rights and the rule of law ([Cilevičs et al, 2018](#)).

### **/ Greece signs a declaration of cooperation on AI with EU Member States**

In April 2018, 25 EU Member States signed a declaration of cooperation on AI. With this declaration, the Member States agreed to work together on different challenges raised by AI, from ensuring Europe's competitiveness in the research and deployment of AI, to dealing with social, economic, ethical, and legal questions. In May 2018, Mr. Costas Fotakis, the then Alternate Minister for Education, Research, and Religious Affairs, signed the declaration on behalf of Greece ([Hellenic Republic, 2018](#)).

### **/ Civil Society and Academia**

#### **/ Hellenic Data Protection Authority (HDP)**

The Hellenic Data Protection Authority (HDP) is a constitutionally consolidated independent authority. The HDP posted guidance on its website stating that citizen rights should not to be subject to a decision based solely on automated processing, including profiling under Article 22 GDPR. By using simple language, the HDP has attempted to clarify the GDPR's provisions of Article 22 and explain its connection with Articles 13, 14, 17, 18, and 21 ([Hellenic Data Protection Authority, 2019](#)).

#### **/ National Commission for Human Rights in the Hellenic Republic (GNCHR)**

The [Greek National Commission for Human Rights \(GNCHR\)](#) is an independent advisory body to the state and specializes in human rights issues. It was established by Law 2667/1998, in accordance with the UN Paris Principles, and inaugurated on 10 January 2000.

#### **/ Hellenic Artificial Intelligence Society (EETN)**

The [Hellenic Artificial Intelligence Society \(EETN\)](#) started in 1996 and it is a non-profit scientific organization devoted to organizing and promoting AI research in Greece and abroad.

### / The Center for Security Studies (KEMEA)

The [Center for Security Studies – KEMEA](#) is a scientific, consulting, and research agency supervised by the Minister of Citizen Protection involved in many of the research projects described above.

### / The Foundation for Research and Technology – Hellas (FORTH)

The [Foundation for Research and Technology – Hellas \(FORTH\)](#) was founded in 1983 and is one of the largest research centers in Greece and it has a reputation for being a top-level international research institute.

### / The National Center of Scientific Research “Demokritos” (NCSR Demokritos)

Founded in 1961, the [National Center of Scientific Research Demokritos](#) is the largest multidisciplinary research center in Greece. Together with Ernst & Young Greece, it founded the Hellenic Artificial Intelligence Center of Excellence ([NCSR Demokritos & EY, 2019](#)).

### / The Center for Research and Technology Hellas (ITI/CERTH)

The [Center for Research and Technology Hellas \(ITI/CERTH\)](#) was founded in 2000 and it is listed among the TOP 20 EU research institutions with the highest participation in competitive research grants.

### / The Research and Innovation Center in Information, Communication and Knowledge Technologies “ATHENA”

The [ATHENA Research and Innovation Center](#) was founded in 2003. It is comprised of the Institute for Language and Speech Processing (ILSP), the Information Management Systems Institute (IMSI) and the Industrial Systems Institute (ISI).

### / The Institute of Communications and Computer Systems (ICCS)

The [Institute of Communication and Computer Systems \(ICCS\)](#) was founded in 1992 and belongs to the School of Electrical and Computer Engineering (ECE) at the National Technical University of Athens (NTUA).

### / The Hellenic Scientific Council for Information Society (HSCIS)

The [Hellenic Scientific Council for Information Society \(HSCIS\)](#) was established in 2001 and its main purpose is to contribute to the development of the Information Society within a privacy preserving environment.

### / The Hellenic Association of Data Protection & Privacy (HADPP)

The [Hellenic Association of Data Protection & Privacy \(HADPP\)](#) is a non-profit group of professionals with a passion to exchange, analyze, communicate and educate on the ongoing issues of data protection, privacy and security.

### / The Institute for Privacy Law, Data Protection and Technology (IPL)

The [Institute for Privacy Law, Data Protection and Technology \(IPL\)](#) was established by the European Public Law Organization and aims at advancing and promoting knowledge and policy in the fields of privacy and data protection law, technology, society, and fundamental rights.

### / The Open Technologies Alliance (GFOSS)

The [Open Technologies Alliance \(GFOSS\)](#) is a non-profit organization founded in 2008. Its main goal is to promote openness through the use and the development of open standards and open technologies in education, public administration, and business.

### / Consumers’ Association The Quality of Life (EKPIZO)

[EKPIZO](#) was established in 1988 to protect consumer rights and improving the quality of life. It is a non-governmental, non-profit association with more than 24,000 members. EKPIZO is a member of the European Consumer Organization (BEUC).

### / Center for Consumer protection (KEPKA)

[KEPKA](#) is a non-governmental, non-profit, independent consumer organization founded in 1982. KEPKA is a member of the European Consumer Organization (BEUC).



## / Homo Digitalis

Homo Digitalis is the first Greek civil society organization focused on the protection and promotion of digital rights. Its activities cover a range of topics related to data protection, privacy, freedom of expression and information online, intellectual property, non-discrimination, and net neutrality. It was founded in 2018, and it is a member of European Digital Rights (EDRI).

## Key takeaways

The use of Automatic Number Plate Recognition (ANPR) cameras by the Hellenic Police is a practice that requires special attention, since it could pose important challenges for the protection of the rights of individuals in Greece. Since the cameras are deployed in police vehicles they could, in theory, monitor nearby pedestrians or scan details of vehicles that are not stolen. This raises important questions regarding the rights to privacy and data protection. The national law should provide thorough details and safeguards on the collection of this data, the purpose of the processing activities, and the retention period, as well as details of access to this information by competent authorities.

In addition, the four million euro contract given to Intra-com Telecom to work with the Hellenic Police to develop an information system related to face recognition, automated fingerprint identification, document authentication, and complex information searches in existing da-

tabases, raises important human rights' concerns. The police intend to use the software during routine stops to verify the identity of individuals. However, a national legal framework providing for the legal basis of such activities and for the appropriate safeguards for the rights and freedoms of the data subjects involved is lacking. Also, one could argue that such a measure does not meet the criterion of strict necessity that allows for the processing of biometric information provided for in Article 10 of the Law Enforcement Directive.

Moreover, it is clear that there are many Greek, or Greek stakeholder-led, research projects with a focus on the development and deployment of ADM systems currently taking place. The fields of predictive policing and border management appear to be two of the most developed. However, the participation of state actors in these research projects, including national research centers such as KEMEA, the Hellenic Police, and various ministries, including the Hellenic Ministry of Defense, is particularly noticeable. It is important to mention that through these projects valuable technical expertise is acquired by the partners involved, which at a later stage could potentially be applied in real-life case scenarios.

At the same time, projects that focus on the use of ADM in the fields of healthcare, transportation, and medicine are in line with the National Digital Strategy of Greece and the aim is for these projects to assist the country in kick-starting its economic development. Such projects could create the infrastructure needed to boost the economy and make Greek stakeholders in these fields more competitive.

***THE USE OF AUTOMATIC NUMBER PLATE RECOGNITION (ANPR) CAMERAS COULD POSE IMPORTANT CHALLENGES FOR THE PROTECTION OF THE RIGHTS OF INDIVIDUALS IN GREECE.***

# References:

- ADAS&ME (2016), <https://www.adasandme.com/about-adasme/use-cases/>
- Ask the EU, [https://www.asktheeu.org/en/request/7495/response/25180/attach/2/REA%20reply%20confirmatory%20signed.pdf?cookie\\_passthrough=1](https://www.asktheeu.org/en/request/7495/response/25180/attach/2/REA%20reply%20confirmatory%20signed.pdf?cookie_passthrough=1)
- ANITA (2018), <https://www.anita-project.eu/index.html>
- ATHENA Research and Innovation Center, <https://www.athena-innovation.gr/en/home>
- AVINT (2019), <https://trimis.ec.europa.eu/project/autonomous-vehicles-integrated-within-urban-context#tab-partners>
- Center for Research and Technology Hellas (ITI/CERTH), <https://www.certh.gr/5B4D1A98.en.aspx>
- Center for Security Studies – KEMEA, <http://www.kemea.gr/en/kemea/about-kemea>
- Cilevičs et al (2018), <http://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=25062&lang=en>
- CONNEXIONS (2018), <https://www.connexions-project.eu/>
- COPKIT, 2018, <https://copkit.eu/>
- Council of Europe (2018a), [https://search.coe.int/cm/Pages/result\\_details.aspx?ObjectId=09000016807c65bf](https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016807c65bf)
- Council of Europe (2018b), <https://rm.coe.int/cets-223-explanatory-report-to-the-protocol-amending-the-convention-fo/16808ac91a>
- Council of Europe (2019a), <https://www.coe.int/en/web/artificial-intelligence/cahai>
- Council of Europe (2019b), <https://www.coe.int/en/web/data-protection/-/greece-32nd-country-to-sign-convention-108->
- EKPIZO, [https://www.ekpizo.gr/en/about\\_us](https://www.ekpizo.gr/en/about_us)
- European Commission (2019), <https://joinup.ec.europa.eu/collection/nif-national-interoperability-framework-observatory/digital-government-factsheets-2019>
- FOLDOUT (2018), <https://foldout.eu/publications/>
- Foundation for Research and Technology – Hellas (FORTH), [https://www.forth.gr/index\\_main.php?c=1&l=e](https://www.forth.gr/index_main.php?c=1&l=e)
- General Secretariat for Civil Protection (2020), <https://www.civilprotection.gr/el/koini-syntentyxi-typoy-gia-ton-periorismo-kykloforias-politon-se-oli-tin-epikrateia-logo-koronoioy>
- Greek National Commission for Human Rights (GNCHR), <http://www.nchr.gr/index.php/en>
- Hellenic Artificial Intelligence Society (EETN), <https://www.eetn.gr/>
- Hellenic Association of Data Protection & Privacy (HADPP), <https://www.dataprotection.gr/>
- Hellenic Data Protection Authority (2019), [https://www.dpa.gr/portal/page?\\_pageid=33,211383&\\_dad=portal&\\_schema=PORTAL](https://www.dpa.gr/portal/page?_pageid=33,211383&_dad=portal&_schema=PORTAL)
- Hellenic Data Protection Authority (2020a), <http://www.dpa.gr/APDPXPortlets/htdocs/documentDisplay.jsp?docid=44,198,200,125,117,79,104,104>
- Hellenic Data Protection Authority (2020b), <https://www.dpa.gr/APDPXPortlets/htdocs/documentSDisplay.jsp?docid=182,151,200,123,234,153,149,126>
- Hellenic Ministry of Culture and Sports (2019), [https://www.opi.gr/images/library/nomothesia/ethniki/ypoyrgikes\\_apofaseis/12521\\_2019.pdf](https://www.opi.gr/images/library/nomothesia/ethniki/ypoyrgikes_apofaseis/12521_2019.pdf)
- Hellenic Ministry of Digital Policy, Telecommunications & Media (2016), [https://www.eydamth.gr/images/site\\_2017/ArticleID\\_283/GR\\_Digital\\_Strategy\\_2016\\_2021.pdf](https://www.eydamth.gr/images/site_2017/ArticleID_283/GR_Digital_Strategy_2016_2021.pdf)
- Hellenic Ministry of Justice, Transparency, and Human Rights (2019), <http://www.et.gr/idoes-nph/search/pdfViewerForm.html?args=5C7QrtC22wFqnM3eAbJzrXdtvSoClrL8m5HxDW-de055MXD0LzQTLf7MGgcO23N88knBzLCmTXKaO6fpVZ6Lx9hLslJUqeiQcGF2jxEfw7ZsehGiT34iyOTcRTmhFV6byKTou5KWlv4>
- Hellenic Police (2018), <http://www.astynomia.gr/images/stories/2018/pr-okirikseis18/26052018texnikesprodiagrafes.pdf>
- Hellenic Police (2019), <http://www.astynomia.gr/images/stories/2019/pr-okirikseis19/14122019anakoinosismartpolicing.pdf>

Hellenic Republic (2018), [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=52938](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=52938)

Hellenic Scientific Council for Information Society (HSCIS), <http://www.hscis.eu/>

Homo Digitalis, <https://www.homodigitalis.gr/>

Homo Digitalis (2018), <https://www.homodigitalis.gr/en/posts/3044>

Homo Digitalis (2019a), <https://www.homodigitalis.gr/posts/4423>

Homo Digitalis (2019b), [https://www.asktheeu.org/en/request/access\\_to\\_documents\\_request\\_rela\\_2#incoming-24832](https://www.asktheeu.org/en/request/access_to_documents_request_rela_2#incoming-24832)

Homo Digitalis (2019c), [https://www.asktheeu.org/en/request/information\\_about\\_four\\_4\\_h2020\\_p#incoming-24777](https://www.asktheeu.org/en/request/information_about_four_4_h2020_p#incoming-24777)

Homo Digitalis & EKPIZO (2019), <https://www.homodigitalis.gr/en/posts/4217>

IASIS (2017), <http://project-iasis.eu/>

iBorderCtrl (2016), <https://www.iborderctrl.eu/Greek-Pilot>

iBorderCtrl (2019), p.15, <https://www.asktheeu.org/en/request/7488/response/24777/attach/3/D6%204%20700626%20Eval%20report%20final%20prototype%20pilot%20deploy%20BestPractices%20redacted.pdf>

Institute of Communication and Computer Systems (ICCS), <https://www.iccs.gr/en/>

Institute for Privacy Law, Data Protection and Technology (IPL), <https://www1.eplo.int/ipl>

Intracom Telecom (2019), [http://www.intracom-telecom.com/en/news/press/press2019/2019\\_07\\_02.htm](http://www.intracom-telecom.com/en/news/press/press2019/2019_07_02.htm)

i-PROGNOSIS (2016), <http://www.i-prognosis.eu/>

Kathimerini (2017), <https://www.kathimerini.gr/898573/article/epikairothta/ellada/epivevaiwnei-o-toskas-to-reportaz-ths-k---den-yparxoy-n-kondylia-gia-rantar-elegxoy-taxythtas>

KEPKA, <http://www.kepka.org/>

Konti (2019), <http://www.ekathimerini.com/244384/article/ekathimerini/business/pfizer-selects-thessaloniki-for-a-new-digital-hub>

Lialos (2017), <https://www.kathimerini.gr/898380/article/epikairothta/ellada/xiliades-kameres-se-komvika-shmeia-aytokinhthodromwn>

Maniatis (2019), <http://www.maniatisy.gr/park-blog/11464-arthro-g-maniati-stin-efimerida-to-vima-to-syntagma-i-klimatiki-allagi-i-texniti-noimosyni-03-02-2019>

National Centre of Scientific Research Demokritos, <http://www.demokritos.gr/>

NCSR Demokritos & EY (2019), <https://www.iit.demokritos.gr/newsevents/opening-ceremony-for-the-collaboration-between-ey-and-ncsr-demokritos-at-snfcc/>

Open Technologies Alliance (GFOSS), <https://ellak.gr/Rainey>, 2018, <https://www.theguardian.com/cities/2018/sep/04/trikala-greece-first-smart-city-dont-need-to-know-a-politician-to-get-something-done>

ROBORDER (2017), <https://roborder.eu/>

SPIRIT (2018), <https://www.spirit-tools.com/index.php>

To Vima (2019), <https://www.tovima.gr/2019/08/06/society/i-astynomia-skanarei-aytomata-tis-pinakides-ton-oximaton/>

TRESSPASS (2018), <https://www.tresspass.eu/The-project>





BELGIUM  
DENMARK  
ESTONIA  
FINLAND  
FRANCE  
GERMANY  
GREECE

**ITALY**  
**STORY**  
PAGE 144  
**RESEARCH**  
PAGE 148

NETHERLANDS  
POLAND  
PORTUGAL  
SLOVENIA  
SPAIN  
SWEDEN  
SWITZERLAND  
UNITED KINGDOM





Find out more on p. 150 in the research chapter under "Is my complaint algorithmically sound? Predictive justice in Genoa (and beyond)".

# In Italy, an appetite for **face** **// recognition** in football stadiums

**Right before the pandemic, the government and top sports authorities were planning a massive deployment of face recognition and sound surveillance technologies in all Italian football stadiums. The reason? To help fight racism.**

By Fabio Chiusi



At the beginning of 2020, just as the whole world was grappling with increasing [evidence](#) of the discriminatory and racist outcomes of face recognition technologies, Italy mulled its widespread adoption in football stadiums as an essential tool in the fight against racism.

The issue had been at the forefront of the country's mainstream media coverage over the previous months, most notably because of the repeated racist chants and slurs made by supporters of [Verona](#), [Lazio](#), and other teams against Brescia superstar Mario Balotelli, who is black.

In February, right before the COVID-19 pandemic took Italy by storm, the Minister for Youth Policies and Sports, Vincenzo Spadafora, had had enough, and he thought that technology should come to the rescue. He, therefore, [announced](#) that the Italian government was about to experiment with, and deploy, "new technologies to support the work of police forces" in football stadiums all over the country. And not just to prevent individuals who are banned from sports competitions from entering Italian stadiums: he wanted them to spot racist supporters, in real-time.

Spadafora did not specify which "new technologies" he was talking about. But others did. Luigi De Siervo, CEO of Lega Serie A, Italy's top football league, [spoke](#) about "the silent work" that institutions were doing. "We need to get all of the people who are ruining this wonderful sport, one by one", he said. "With face

WE NEED TO GET ALL OF THE PEOPLE WHO ARE RUINING THIS WONDERFUL SPORT, ONE BY ONE. WITH FACE RECOGNITION, THIS IS NOW POSSIBLE.

recognition, this is now possible." According to [La Repubblica](#), De Siervo even went to Israel to study its latest applications.

The head of the Italian Football Federation (Federazione Italiana Giuoco Calcio, or FIGC), Gabriele Gravina, also

[talked](#) about the same idea: "using technology will lead to high-definition face recognition of individuals who will be guilty of committing such crimes," he had said in January. To Gravina, the project was meant to see the light in a relatively short time span: "Experimentation will start rapidly, and then we will sensitize all interested parties, both clubs and leagues, towards using this technology."

## **/ The police want to listen to your conversations in the football stands**

But it's not just face recognition: the plan also included adopting sound surveillance systems, so-called "sonic radars". Already, in October 2019, Gravina had [hypothesized](#) the adoption of the "passive radars" currently in use by anti-terrorism units, to "spot the source of a noise", including private conversations between supporters in the stands – a feature that Gravina himself described as problematic in terms of privacy. However, this supposedly meant that such systems would also be able to pick up racist conversations. Offenders could then be identified through face recognition.

ONLY THOSE CLUBS IMPLEMENTING THE ADEQUATE TECHNOLOGIES WOULD BE SHIELDED FROM LIABILITY FOR ACTS COMMITTED BY SUPPORTERS.

Deploying such technologies would also bear legal consequences, according to Gravina. In fact, Federcalcio was contextually [planning](#) to adopt a new regime of "objective responsibility" for football clubs: only those implementing the adequate technologies would be shielded from liability for acts committed by supporters. This way, according to Federcalcio, they would also be free from extortion by hardcore, "ultra" fans, who often use racism as a threat to attain the goals of their blackmailing strategies. For example, in September 2019, when Juventus FC [stopped](#) giving in to alleged extortion – demanding cheap season tickets, free drinks at stadium bars, and even invitations to club parties – ultras started singing racist chants as a reprisal, knowing all too well that it would have meant sanctions for the club. As a result, 12 ultras were arrested.

While various media reports [argued](#) that some staffers within the Interior Ministry had serious doubts over the effectiveness of such technologies in real-life scenarios, [others](#) uncritically praised them as "the next frontier of anti-racism". De Siervo even [compared](#) himself to former UK

Prime Minister, Margaret Thatcher: “We will achieve in two years what Thatcher achieved in ten”, he argued.

But then, the pandemic struck, and everything changed.

## / The government vaguely confirms its plans, others remain silent

With the COVID-19 outbreak raging throughout northern Italy, the government brought Serie A to a halt on March 10. Stadiums would remain closed to the public until June, crucially delaying the planned deployments.

And yet, in an email exchange in May 2020, a spokesperson for the Ministry of Youth and Sports “surely” – but vaguely – confirmed that “we are trying to implement a series of tools to prevent racism in stadiums and that we’re working on that together with the FIGC.” No updates were deemed necessary in a further written exchange in September, and no further answers have been provided to the precise questions asked by AlgorithmWatch concerning the plan.

Neither Lega Calcio Serie A nor Federcalcio replied to our questions also. And yet, that hasn’t stopped clubs from experimenting with face recognition in stadiums. Thanks to media reports, we are aware of pilot deployments both in [Udine](#) and [Naples](#), where 190 face recognition cameras were activated in September 2019. A month later, video surveillance at the Neapolitan San Paolo stadium proved crucial in [identifying](#) some 32 supporters who were then each issued with a fine of 500 euros – 166 euros if paid within five days of notice – for the violation of the stadium’s security regulations.

## / Yet another job for S.A.R.I.

But it’s the Udine pilot that introduces the most relevant development. As detailed in the [first edition of our ‘Automating Society’ report](#), an “automated system for image recognition” (“Sistema Automatico di Riconoscimento Immagini”, or “S.A.R.I.”) was already in use by the Italian police to apprehend criminals, both to match the face image of a suspect with those included in databases maintained by law enforcement (“ENTERPRISE” function) and to perform

live face recognition from real-time video feeds (“REAL TIME” function).

Conceived in [2014](#), and deployed in several different settings (from the port of Bari to a casino in Las Vegas), this system was also trialed in June 2019 at the Stadio Friuli, Udine, before the final match of the Under-21 European Cup. The objective was to pilot the monitoring function at the gates to keep supporters out who had previously received a restraining order that bans them from entry.

EVEN IN THE  
ABSENCE OF ANY  
PUBLICLY AVAILABLE  
EVIDENCE, DE SIERVO  
HINTED AT A FUTURE  
ADOPTION OF S.A.R.I.  
FOR ALL ITALIAN  
FOOTBALL STADIUMS.

AlgorithmWatch repeatedly asked both Udinese FC and Reco 3.26, S.A.R.I.’s Puglia-based creators, about the results of that trial deployment, but the football club delegated all answers to the developers, and the developers never replied to any questions.

And yet, even in the absence of any publicly available evidence, De Siervo hinted at a future adoption of S.A.R.I. for all Italian football stadiums, [explicitly stating](#) that the system should realize this technological revamp of football stadium security would be “the same as that in use by police forces”.

The lack of any public discussion around this topic did not prevent S.A.R.I. from being [portrayed](#) by mainstream media as a default security measure in both Inter and Milan FC’s plans for a new stadium to replace San Siro in the Lombardy capital. Here, it would not only check for banned individuals, but also for “potential terrorists”, according to *Corriere della Sera Milano* – its main selling point being that the software could allegedly analyze “a face at a 60 meters distance”.

But, again, there’s more. According to the same article, the plan that both Inter and Milan football clubs have shared with the municipality’s office includes “geolocalisation and sound sensors”, together with “software that is able to recognise abnormal behaviour, such as “loitering” or “the sudden presence of an object in a certain place.”

## / Former president of Italian DPA: “caution” should be applied

Racism. Terrorism. Loitering. Face recognition in stadiums takes on many guises – and even more deployments. Face recognition cameras are, for example, also [installed](#) at Ata-

lanta's newly named "Gewiss Stadium" in Bergamo. The re-styled "Curva Pisani" features seven turnstiles and 40 video surveillance cameras, some of them equipped with face recognition technology. In 2016, the Italian Data Protection Authority authorized the use of the technology in Rome's Stadio Olimpico.

But things have changed since then. In recent years, evidence of inaccurate and discriminatory outcomes of face recognition technologies has increased. And that evidence should matter, argues the former president of the Italian DPA, Antonello Soro, in a written interview with Algorithm-Watch conducted when he was still in office.

In the interview, Soro acknowledged that "the absolutely peculiar context of [a football] stadium is one... which ... extensive use of biometric technologies... [is]... more evident." As a result, safeguards are all the more important. And, since 2016, the "peculiar dangers" of biometric surveillance – "especially when instrumental to an algorithmic decision-making process" – have become apparent, Soro calls for "further caution" before deploying such systems. In particular, they have to be made consistent with the law and need a privacy impact assessment. This is what the DPA has been talking about in his dialogue with the FIGC too, Soro wrote.

However, the "real time" function of S.A.R.I., the one that would be relevant for Spadafora's technological plans against racism, has only been subject to "preliminary dialogues" with the Interior Ministry during his time in office, wrote Soro. As a result, the proposal "has not been translated into a detailed project" – mainly because of the lack of an "adequate normative framework", one that "details checks and balances" of the treatment of biometric data, especially for such a "significant one, both in terms of size and typology of performed operations."

No further updates have been provided under the current presidency of the DPA, which started in July 2020.

## **/ The COVID-19 pandemic as a way to repurpose face recognition**

In the absence of clear transparency rules, pilots and deployments of face recognition technology began to multiply,

hidden from public view. But then, the COVID-19 pandemic struck, halting competitions and emptying public spaces.

A foundational issue then came to the fore: when stadiums reopen, supporters will need to wear a mask to be allowed to enter the premises. But that spells doom for the accuracy of face recognition technology, as documented in a [recent](#) National Institute of Standards and Technology (NIST) [study](#). What to do? Could face recognition be upgraded in such a way that it can still recognize people, even though they are wearing masks?

While many companies [claim](#) to have solved the issue, the available evidence suggests otherwise. This is possibly the reason why current pilots (and PR material from vendors) mostly focus on face recognition technology, without, rather ambitiously, claiming to recognize an individual who is wearing a mask. The solutions that claim to do this mostly work in conjunction with thermal scanning technology and other tools used to scan public health during the pandemic.

One example of a [pilot project is in Turin](#). The "Feel Safe" system – an "automated anti-virus gate" – was trialed at the match between Turin and Hellas Verona. The gate is designed to safely speed up the entrance of fans and is produced by the Milan-based company, Worldwide Exhibition System (WES). No information is available regarding what kind of biometric data the system takes into account, not even in corporate or PR material. A WES presentation only says that the FeelSafe system can "interface with a company's entry management system, thanks to a system of biometric controls."

Italy is, of course, not alone in experimenting with face recognition technology in stadiums. While [Denmark](#), [Belgium](#), and the [Netherlands](#) deployed the system for security checks – with varying results and legal [challenges](#) – it is also being [thought about](#) as a tool "to give fans personalized experiences" at the Dodger's Stadium in Los Angeles.

And yet, no country – apart from Italy – has tried to apply face and sound recognition systems to identify the authors of racist slurs.

Whether this idea will actually be implemented remains to be seen. We'll be watching.

IN RECENT YEARS, EVIDENCE OF INACCURATE AND DISCRIMINATORY OUTCOMES OF FACE RECOGNITION TECHNOLOGIES HAS INCREASED. AND THAT EVIDENCE SHOULD MATTER.



# Research

By Fabio Chiusi

## Contextualization

### / A louder conversation, but mostly around “AI”

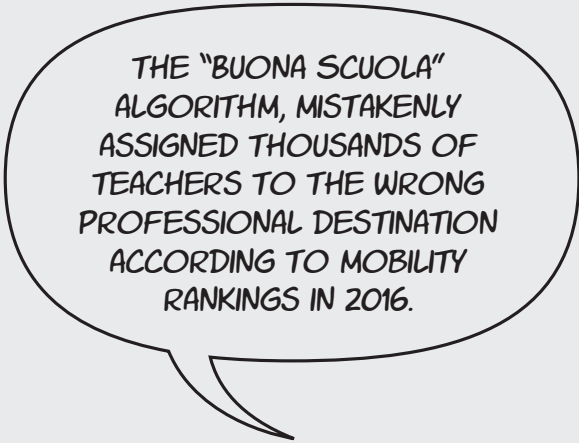
Even though it is still unrecognized as such, automated decision-making systems are increasingly at the forefront of policy and public debates in Italy. Specific legislative measures are absent, and — as noted in the first edition of this report — there is no formal definition of ADM for both government and private use. And yet, two sentences pronounced in 2019 now prescribe clearer limitations around the use of ADM systems in education. This particularly refers to the “Buona Scuola” algorithm, that mistakenly assigned thousands of teachers to the wrong professional destination according to mobility rankings in 2016, and which has since been discontinued.

ADM systems are at the heart of the debate surrounding the use of data to fight the COVID-19 outbreak, through a mobile phone app which has received consistent media coverage, and also related to Big Data predictive analytics when checking for fiscal consistency with the ISA Fiscal Reliability Index implemented by the current government. Predictive policing has also made its mark in Italy and findings unearthed in this report clearly show how solutions like XLAW are widely deployed, tested, and implemented. Predictive justice and health projects are also being explored, mostly at the local level.

However, for most Italians, the debate is about “artificial intelligence” and not ADM. In fact, the social and political implications of “AI” still constitute the bulk of the mainstream rhetoric around the challenges and opportunities of automation, strongly hingeing on moral panics and “clickbait” mainstream media articles. At the same time, the public dimension of the tech-policy debate around algorithmic fairness, discrimination, and manipulation often proceeds almost entirely devoid of the mounting evidence coming from both academic literature and investigative journalism.

### / Face recognition is suddenly everywhere

This might help explain why the increasing deployment of invasive ADM solutions — such as face recognition technologies — is still not properly appreciated by the population at large in Italy, despite the fact that an October 2019 Comparitech report (Forbes Italia 2019) bashed Italy’s “extensive use of biometrics, including face recognition in airports” adding that it was “causing concern among citizens”. And while face recognition has become wildly popular through smartphone applications like FaceApp, experiments with its uses have been taking place in public venues, and institutions, such as stadiums, hospitals, and schools throughout the country and in the absence of any meaningful democratic debate, more experiments are already being planned.



THE “BUONA SCUOLA”  
ALGORITHM, MISTAKENLY  
ASSIGNED THOUSANDS OF  
TEACHERS TO THE WRONG  
PROFESSIONAL DESTINATION  
ACCORDING TO MOBILITY  
RANKINGS IN 2016.

As a result, no significant activism campaign has grown in response, thus allowing these experiments to go on mostly unchallenged or, when challenged by journalists and pundits, to be met with a disconcerting lack of transparency. This is most apparent in the case of the — ever-evolving — SARI face recognition system used by law enforcement agencies. The system’s mysterious functioning has yet to be made fully public.

This lack of detailed scrutiny has also led to an increasing tendency to frame algorithms in a “dystopian” fashion, and (unjustly) identify them as the culprit of many of the current

drawbacks within Italian society. Algorithms are regularly blamed for polarization, lack of trust in institutions and the media, “fake news” — whether automated or not — and for the crisis in democracy. Furthermore, rising racism, hate speech, far-right nationalism, and intolerance have been substantially attributed to the personalization and micro-targeting algorithms of digital platforms. Again, this has mostly happened without solid, peer-reviewed evidence or reporting on the actual persuasive effects of social media campaigns.

## **/ What Italians think about when they think about ADM**

But what do Italians actually think about when they think about ADM? The best available proxy is to look into how and what Italians think about “AI” in decision-making contexts. In 2019, two polls tried to understand several aspects of the role of artificial intelligence in the daily lives of Italians, and some of the findings are relevant to the ADM debate. The AIDP-Lablaw survey, conducted by BVA DOXA and published in November 2019 (BVA DOXA 2019), argues that Italians are generally hopeful that AI will unearth solutions and discoveries that would otherwise be inconceivable (94%), and they widely believe that AI is beneficial to well-being and quality of life (87%). Also, 89% think robots will never fully replace human intervention — this broadly refers to the job market, but it might also be relevant in terms of ADM.

However, the Italians who were surveyed still fear that AI will end up having a commanding role in their lives (50% think there is a risk of machine predominance on mankind), and a large majority (70%) believe that “many jobs” will be lost as a result of the increase in automation. A crucial finding was the startling consensus around regulating AI: 92% agreed that laws are necessary.

Another survey, also published in November 2019, tried to address a narrower issue: Are Italians concerned about “killer robots” — part of what we may label “ADM in war”? According to a YouGov poll within the international “Campaign to Stop Killer Robots” (Rete Disarmo 2019), managed in Italy by Rete Disarmo, 75% of Italians are in favor of an international ban on fully autonomous systems equipped to kill human beings, and they go as far as proposing Italy as a leader in the design of international norms around the issue. A previous IRIAD Archivio Disarmo poll (March 2019) showed similar results. Support for the ban is also consistent across the whole ideological and party spectrum.

Also, relevant to help understand attitudes towards ADM in Italy is a monster.it poll concerning automation in the hiring process. Published in December 2018, the poll concluded (Sole 24 Ore 2018) that almost half of the sample (45%) firmly believe that “robots will never be able to assess an applicant better than a human”. However, almost a quarter of respondents think that algorithms might be — and will be — helpful in assisting the evaluation of candidates. Six percent agreed that “algorithms will be so reliable and precise that they’ll be able to substitute humans”.

## **A catalog of ADM cases**

### **/ Your boarding pass is your face: face recognition at airports**

Face recognition technologies are being deployed, or are about to be deployed, at airports in Milan, Linate, Malpensa, and at Rome Fiumicino — three of the biggest hubs for both national and international flights in the country. The idea behind the adoption of the technology is to allow travelers to experience a “frictionless journey” by replacing identification and boarding passes with the traveler’s own face. Airport authorities claim that this will make checks quicker, while at the same time the technology will increase security.

On February 12, 2020, a pilot deployment of face recognition technology started at Linate airport, and will last until the end of the year (Corriere Milano, 2019; Corriere della Sera, 2020). Called “Face Boarding”, the project initially involves Alitalia flights to Roma Fiumicino only, and just for frequent and “Comfort” (i.e. business) class flyers.

A trial deployment of face recognition is also expected to begin at Malpensa airport by the end of 2020 (Mobilita 2019).

At Rome Fiumicino, the trial started in November 2019, and initially involved just one company, KLM, and one destination, Amsterdam (TgCom24, 2019). However, airport managers claim that it will soon expand and that it will ultimately apply to all flights and all destinations in and out of the international airport.

Developers of the system, deployed through a collaboration between ADR (Aeroporti di Roma) and electronic iden-

tity solutions provider Vision-Box, claim that it only acquires biometric information of the traveler temporarily, and that all data is erased “within an hour of departure”: “Nothing will remain on ADR’s servers”, they write on the airport’s website.

The biometric-based technological overhaul of both the Milan hubs, Linate and Malpensa, will cost 21 million euros, wrote all-news broadcaster TgCom24, and will include “six kiosks for enrollment (subscription to the program that links biometric data to a passport or ID card), 25 electronic boarding gates, 10 pre-security gates and 7 face security spots”.

Face recognition-powered “Automated Border Control Gates” (or “ABC Gates”) were first installed at Naples airport in 2016, and manage the transit of some 2,000 passengers a day (Repubblica Napoli 2016). “The new technology helped providing Naples Airport passengers with a better experience”, says Alessandro Fidato, the facility’s Infrastructure & Operations Director, “making transit and controls quicker and at the same time assuring that the stricter standards in security are observed at Italian borders”. The IT solution is provided by Sita, described in the media as a “world leader” in airport security technology. The company claims to already serve some thirty governments worldwide.

More deployments are to be expected, especially in the aftermath of the COVID-19 pandemic, according to media reports (Il Messaggero, 2020).

No independent auditing or finding concerning any of these face recognition systems is publicly available, and none has been adequately debated among the general population.

## **/ These automated systems predict your health**

Experiments with automated decision-making systems for predictive health are not new in Italy. And while projects such as RiskER and Abbiamo i numeri giusti are still ongoing, another interesting predictive health project has been added to the Italian landscape throughout 2019. In the city of Vimercate, in the Lombardy region of northern Italy, the local branch of the public health authority (Azienda Socio Sanitaria Territoriale, ASST) has adopted open source cloud solutions (developed by Al maviva) which use algorithms to

predict the beginning of chronic pathologies and post-surgery complications (Giornale Monza 2019).

Building on a decade spent digitizing medical records, Vimercate hospital aims to more efficiently manage patients, including by personalizing treatments according to machine learning-powered analytics, thus reducing costs and optimizing logistics, while at the same time improving its “precision medicine” solutions.

“This is the first structured usage of AI within a hospital in Italy”, says Head of AI Solutions at Al maviva, Antonio Cerqua, who also revealed that the objective is to provide his company’s IoT platform, Giotto, “to a national network of hospitals”. According to Cerqua, “many have already shown their interest” (Il Giorno 2019).

**NO INDEPENDENT  
AUDITING OR FINDING  
CONCERNING ANY OF THESE  
FACE RECOGNITION SYSTEMS IS  
PUBLICLY AVAILABLE, AND NONE  
HAS BEEN ADEQUATELY DEBATED  
AMONG THE GENERAL  
POPULATION.**

Representatives from the consulting giant Deloitte also visited the Vimercate hospital specifically because of this project, wrote Giornale di Monza.

“AI must not replace physicians”, says Head of Informatics at ASST Vimercate, Giovanni Delgrossi, “but support them, a tool that helps them making better decisions” (Sole 24 Ore 2019b).

## **/ Is my complaint algorithmically sound? Predictive justice in Genoa (and beyond)**

In 2019, the first experiment with automated decision-making within the justice system was developed at LIDER Lab Scuola Superiore Sant’Anna in collaboration with EMbeDS, KDD Lab and the Tribunal of Genoa (Gonews 2019).

Called “Predictive Jurisprudence”, it allows researchers to access and analyze — through machine learning techniques — the corpus of rulings pronounced by the judiciary in the Liguria capital. The purpose of this is to extract meaningful information for further processing, starting with the identification of common trends in jurisprudence focused on a specific subject matter (e.g., past judgments on a certain typology of crime or around similar court cases).

This would provide a benchmark against which each human judge might assess the case before him or her, and

easily check for consistency with previous rulings in analogous situations. Ideally, it might even produce predictions regarding a practitioner's future behavior in similar cases.

Genoa Tribunal president, Enrico Ravera, argues that the aim of the project is not a mere "extrapolation of statistics from case studies", but rather to "apply artificial intelligence techniques to jurisprudence". More precisely, "Predictive Jurisprudence" is described by the research team, coordinated by Prof. Giovanni Comandé, as "a multilayer project unfolding into five interconnected but autonomous levels".

In a written reply to an email request by AlgorithmWatch, Comandé and colleague Prof. Denise Amram said: "In the start-up phase, the project aims to analyze decisions with the corresponding files of trial courts according to the criteria and methodologies developed in the Observatory on personal injury, applicable to areas of litigation other than non-pecuniary damages (level 1). The same materials are used also through techniques of Machine Learning to develop both tools for annotation and automatic extraction of information from legal texts (level 2) and algorithms for analysis and prediction (so-called Artificial Intelligence level 3)."

In particular, the researchers argue, "the architecture of the database designed to host the data acquired by the courts will be designed and trained for developing algorithms to automatically identify trends with reference to the criteria known to the interpreter, as well as to highlight new trends on the basis of possible bias/tendencies found by the algorithm."

Especially notable in terms of automated decision-making is the fact that "the algorithm aims to recreate and mimic the legal reasoning behind the solution(s) adopted in the judgments by making predictable subsequent decisions on the same subject." Automation is, therefore, applied to the summarization and explanation of ideas: "These tools should also help to explain the reasoning underlying each decision, while the development of suitable tools to explain the criteria defined by the developed AI (level 4) will be tested."

Finally, in the last phase, "efforts and results in the different levels of research and development will be traced back to the attempt to structure the analysis of the legal argument at such a level of abstraction and systematicity as to contribute to the simplification of all tasks (level 5)."

Selected case-studies currently involve a first pilot on "alimony in case of divorce". Here, "queries are pre-determined by law, but their judicial interpretations continuously evolve", write the researchers. "At this regard, current reform bills are proposing to introduce new criteria, whose efficacy could be discussed in light of our analysis."

"The second pilot and the third one," wrote Comandé and Amram, "develop within the Observatory on Personal Injury Damage studies. The algorithm may contribute to the identification of criteria for awarding non-pecuniary losses compensation beyond the current interpretations and attempts to standardize these highly subjective head of damages. Within this core-analysis, the algorithm could be better-trained to explain non-pecuniary losses in case of burn out, whose boundaries are still discussed both from clinical and legal perspectives."

According to the researchers, "several Tribunals are joining the project. Among them the Tribunal of Pisa and Bologna."

In 2018, the Appeals Court and Tribunal in Brescia also started their own experiment with "predictive justice" through a pilot project that aims to extract predictions about the length of a lawsuit, and the principles that will most likely be adopted in evaluating it and goes as far as providing an estimated probability of having a complaint approved. (BresciaOggi 2018)

This is part of a two-year long collaboration between the justice system and the University of Brescia which consists of the following steps: identification of the subject matter to analyze, creation of a database for each topic, designing of the "work groups" among university resources that will operationally interface with the tribunal, extraction of predictions by researchers and, finally, publication and dissemination of findings (AgendaDigitale 2019) — which, at the time of writing, are not yet available.

## Policy, oversight and debate

### / Learning from the Buona Scuola algorithm debacle

The most important developments in the field of automated decision-making over the last year concern how the Italian

Courts responded to the claims of teachers affected by the structural mistakes made by La Buona Scuola algorithm. The algorithm was used to sort 210,000 mobility requests from teachers in 2016. The evaluation and enforcement of the mobility request procedure were delegated entirely to a faulty automated decision-making system (Repubblica, 2019).

The algorithm has since been discontinued, and yet appeal cases against the decisions it made are still in the thousands (10,000 teachers were affected, according to La Repubblica), and, as noted in the first edition of the Automating Society report, these complaints are strongly grounded in facts — from bad programming choices, to deliberate obfuscation strategies to prevent meaningful auditing of the source code of La Buona Scuola algorithm itself.

As a result, teachers were unjustifiably moved from Puglia to Milan, instead of within their own region, or displaced from their natural location in Padua. The system even automatically forced two teachers with autistic children to move from the southern region of Calabria to Prato, in the northern region of Tuscany.

Over the course of 2019, two specific rulings determined precisely how and why these automated mistakes resulted in illegal discrimination and choices over the professional and personal lives of thousands of individuals employed in the public education system. The first case was from the Consiglio di Stato (State Council 2019), and the second was from Section 3rd-Bis of the Tribunale Amministrativo (Administrative Tribunal, TAR) of Lazio (n. 10964, September 2019) (TAR Lazio 2019).

And while both institutions recognize the technical flaws that affected the performance of the ranking algorithm, they are also clear that the actual flaw is even more fundamental, as it involves the very rationale behind its deployment — according to the TAR, even contradicting the Italian Constitution and the European Convention on Human Rights.

In fact, even though the State Council reminds us that the automation of decision-making processes within the administration brings about “unquestionable benefits”, it also prescribes that these benefits can only be realized when such processes do not involve any “discretionary judgment” on the part of the system. In particular, its “technical rules” (the programming and functioning) and decisions must always ultimately be the result of human judgment. Algorithms might even work by themselves, but only when highly standardized, repetitive procedures are involved.

Similarly, the TAR of Lazio concluded that “informatics procedures, were they to attain their best degree of precision or even perfection, cannot supplant, by entirely replacing it, the cognitive, acquisitive and judgmental activity that only an investigation entrusted to a physical person can attain”. Crucial, again, is the fact that the “whole” procedure was delegated to an “impersonal algorithm”, when instead automation may only serve an “instrumental”, “merely auxiliary” role within the administrative process — clearly specifying that algorithms may “never” be put in a “dominant” position when compared to human beings.

This has clear implications for the transparency and explainability of the motives behind the public administrations’ decisions, as both rulings remark. However, the interpretative frameworks differ.

The State Council does not rule out the possibility of a transparent, fully automated decision. It only bases its favorable response to claimants in the fact that the particular algorithm deployed for La Buona Scuola was so compromised and chaotic as to be “unknowable”—neither to the affected teachers nor to the Court itself. However, the ruling immediately adds that when full automation is involved, an even stronger notion of “transparency” should apply. A notion that “implies the full knowability of any rules expressed in languages other than the judicial”.

The TAR of Lazio, instead, defines the very possibility of such full automation as “deleterious Orwellian perspective”, suggesting that both transparency and explainability would be fundamentally compromised as a result.

**APPEAL CASES  
AGAINST THE  
DECISIONS BUONA  
SCUOLA MADE  
ARE STILL IN THE  
THOUSANDS.**



## / Algorithmic fiscal reliability rankings in practice

First introduced by the Gentiloni government in 2017, the ISA Fiscal Reliability Index finally came into use in 2019. Developed by SOSE (Soluzioni per il sistema economico), a subsidiary of both the Ministry of Finance (MEF) and Banca d'Italia, the system consists of an algorithmic scoring of fiscal trustworthiness. The system assigns precise benefits to higher rankings (above 8 out of a scale of 10), while at the same time prescribing stricter controls on those below 6. This is based on the mere automated presumption of anomalies in consumption, properties, investments, and other financial behavior. Its rationale is to give fiscal subjects a way to self-assess their position and preemptively fill any gaps that might alert authorities and lead to further scrutiny.

And again, detractors pointed out that the system is essentially flawed. *Corriere della Sera's* Dataroom interviewed “several accountants” in “many different regions of Italy”, and even though the methodology does not give statistical attribution and/or scientific weight to her findings, investigative journalist Milena Gabanelli concluded that errors might concern as many as 50% of taxpayers. As a result of the introduction of ISA “we found”, Gabanelli wrote, “that 40 to 50% of taxpayers shifted from a “congruous and coherent” fiscal position in income tax return for 2018 to an “insufficient” one in 2019, and vice versa” (*Corriere della Sera* 2019).

Errors allegedly resulting from the fact that taxpayers are not allowed to share which factors actually contribute to their annual incomes, thus facilitated misrepresentations and misunderstandings. It only takes a pregnancy, some vacant real estate, or a contingent surge in legal expenses to alert the system and potentially find yourself automatically flagged for “anomalies”, claims Gabanelli, as the algorithm is “badly programmed” and even less flexible and affordable than the widely criticized “studi di settore”, based on an “inductive” method of fiscal scrutiny. Official data are still lacking.

Also, claims of a “Fiscal Big Brother” being algorithmically implemented, noted in the previous edition of this report, still represent the main framing around automated decision-making about fiscal matters in public debate. This is especially the case regarding a new, “anonymized” tool for predictive, Big Data analysis of taxpayers’ behavior which was also introduced with the 2019 Budget Law (Sole 24 Ore 2019c).

Called “Evasometro Anonimizzato”, the tool will include “specific algorithms”, wrote financial newspaper *Il Sole 24 Ore*, and will be able to cross-check the different databases held by the Italian fiscal authorities for inconsistencies in consumption patterns or in any other financial operation. Anomalies recognized by the “digital eyes” of the system will alert fiscal authorities, who can then summon flagged individuals for further scrutiny.

This algorithmic-driven monitoring has to be coupled with that of “open sources”, such as “news articles, websites and social media”, by the Italian fiscal authority, Agenzia delle Entrate. The monitoring system has been active since 2016, but was largely unknown to the public until January 2020. It was devised to “acquire all elements that could be useful in knowing the taxpayer”, thereby helping the institution check the consistency of each individual’s fiscal behavior (*Corriere della Sera* 2019b).

Proof of how this system works revealed in several Court rulings — for example, from the Appeals Court of Brescia and Ancona — in which judges explicitly mentioned damning “documentation extrapolated from Facebook”. One example is that of an ex-husband who was forced to pay living expenses to his former wife because of Facebook posts that depicted a lifestyle patently incompatible with the 11,000 euros income he declared in his tax return for the previous year.

## / Predictive policing: the new normal?

Predictive, automated policing is rapidly widening its scope in Italy. While the 2019 Automating Society report showed how the KeyCrime software had been deployed in the city of Milan, experiments of implementation of another algorithm-based solution, called XLAW, have recently spread to several other important municipalities.

Like KeyCrime, the XLAW software was also developed by a law enforcement official and it has already been deployed in the cities of Naples, Modena, Prato, Salerno, Livorno, Trieste, Trento, and Venice (*Business Insider* 2019), where it led to the arrest of a 55 year-old man accused of theft (*Polizia di Stato* 2018). Law information website Altalex reports that the system’s accuracy is 87-93% in Naples, 92-93% in Venice, and 94% in Prato, but, at the moment, no independent auditing or fact-checking is publicly available for each of these trials (Altalex 2018).

XLAW has a strong predictive component to it. In the case of Venice, for example, State Police boasted that “84% of felonies (“fatti-reato”) that have been either attempted or committed had been foreseen by the system”. As a result, XLAW, just like KeyCrime, is regularly portrayed as a success story.

And yet, the only official response concerning the actual results produced by KeyCrime in Milan over the years tells a different story. In reply to a FOIA request by VICE journalist Riccardo Coluccini, the Dipartimento per le politiche del personale dell'amministrazione civile e per le risorse strumentali e finanziarie of the Interior Ministry stated that it does not possess any “data or study concerning the software’s impact on tackling crimes”, not even around the “variables” considered by the system in making its automated decisions. (Vice Italy 2018)

In addition, full disclosure has been opposed by the administration as the software is patented and proprietary, and as such the data enjoys “special protection” in terms of intellectual property rights. The Direzione Centrale Anticrimine of the Interior Ministry only uses it on a gratuitous loan, Vice reported.

The creator of XLAW, Elia Lombardo, describes the software as a probabilistic, machine learning-powered solution for trends and pattern discovery in crimes. The rationale behind its workings is the “hunting reserves” model of crime-spotting, that assumes “predatory” crimes (e.g. burglary, robbery) to be both “recurrent” and “residing”. This implies that one can deduce — or more accurately, induce — how criminal behavior will unfold, before it has happened, by intelligent analysis of the criminal history of a location over time. Lombardo claims to have done this by carefully analyzing 20 years of data and also by calling on his experience in the field. As a result, thanks to XLAW, law enforcement officials can be alerted and deployed on the scene before a crime has even happened. At the same time, XLAW provides police officers with details as precise as “genre, height, citizenship, distinguishing features, and biometrics” of a potential suspect.

Lombardo states that this approach was validated by the “independent” analysis of two universities in Naples — the Federico II and Parthenope — before it was implemented.

## / Strategic documents on AI, and their relevance for ADM

Even in the face of a proliferation of examples of automated decision-making systems in every Italian citizen’s public life, policy responses have been notably absent.

The hypothesized Parliamentary Committee on AI Ethics has been shelved, with all three of the current governing parties proposing one focused on disinformation — still labeled “fake news” even in the face of wide opposition from academia — instead (Valigia Blu 2019).

A trans-disciplinary center on AI, proposed in the 2018 Digital Transformation Team’s White Paper, has not been realized either.

However, from January to June 2019, thirty experts convened by the Italian Ministry of Economic Development (Ministero dello Sviluppo Economico, MISE) outlined a document that provides a foundation for the government when drafting the Italian Strategy on AI (“Strategia nazionale per l’intelligenza artificiale”).

The resulting document was open for public consultation between 19 August 2019 to September 13 2019 (MISE 2019), and its final version was published on July 2, 2020 (MISE 2020).

The strategy claims to be fundamentally rooted in the principles of anthropocentrism, trustworthiness, and sustainability of AI, with a strong accent on the necessity of an ethical use of AI and on the need to ensure technical trustworthiness starting from the very design of such systems. The ambition is no less than to spur a new “Renaissance” based on artificial intelligence, called “RenAIssance” in the final document.

To steer the country in that direction, the report outlines 82 proposals, some of which are relevant to understand how automated decision-making is framed in Italy. Even though not explicitly mentioned, ADM systems might benefit from the “national mapping” of experiments in regulating, testing, and researching AI (“regulatory sandboxes”) that, according to the Strategy, would be held by a newly hypothesized inter-ministerial and multi-stakeholder “cabina di regia” dedicated to tackling the multifaceted regulatory, educational, infrastructural, and industrial issues raised by the implementation of AI.

Echoing EU law, the document also asserts that explainability of algorithmic decisions should be ensured through procedural transparency, “every time a decision that strongly impacts on personal life is taken by an AI system”. As per the EU Commission’s White Paper on AI, transparency should be proportional to the risk embodied by the AI system under consideration.

Creators and vendors of such systems should also be obliged to inform users whenever they deal with a machine, instead of a human, reads the Strategy, while appropriate sanctions and remedies should be put in place for automated decisions that entail discriminatory/illegal effects. Consistent with the GDPR, this means that users must be able to effectively oppose any automated decisions, in a “simple and immediate” way.

Even though still under development, “Trustworthy AI Impact Assessment” is also explicitly mentioned as a tool to better understand actual risks implied by the use of automated systems. Benefits are strongly highlighted, though, as adoption of “AI” solutions should be promoted, according to the document, for every “societal challenge”, and in particular to achieve a “better quality of life”, for example through deployments in the health, education, and “digital humanities” sectors.

Artificial intelligence is also prominently featured in the “2025 Strategy for Technological Innovation and Digitization” by the newly created Ministry for Technological Innovation and Digitization within the current, second Conte administration. In It, AI is envisioned as “in the Nation’s service”, with an implicit reference to automated decision-making processes, and especially useful in enabling “efficiency” in repetitive administrative procedures that require a low degree of discretion.

At the same time, an ethical, trustworthy AI could help with respect regarding the constitutional principle of a fair and timely trial, according to the document. In fact, the judiciary system is explicitly mentioned as the first sector of public life in which the 2025 Strategy envisions experimentation. A newly formed “Alliance for Sustainable Artificial Intelligence”, a joint committee made of both private entities and public institutions, will, therefore, have the task of devising an “Ethical and Judicial Statute for AI”.

“Beyond fixing a minimal set of guidelines”, the Alliance — through the so-called “AI Ethical LAB-EL” — will “establish a set of basic requirements for the acceptance of AI solutions

for both the public and private sector”. According to the 2025 Strategy, respect of these guidelines could become, in time, equivalent to an official “certification” that a certain AI solution is both ethically and judicially viable.

The document, however, contains no mention of the National Strategy on AI, and provides no coordination among the two official plans for Italy’s AI-powered future.

Lastly, in April 2019, +Europa MP Alessandro Fusacchia launched an “Intergruppo parlamentare” on AI (CorCom 2019).

## Key takeaways

Italy needs to be more aware of how quickly automated decision-making systems are spreading in many sectors of public administration and public life. Face recognition technologies and other invasive ADM systems are being increasingly adopted without any meaningful democratic debate and/or dialogue with civil society and academia, thus potentially setting the stage for further problematic deployments.

The Buona Scuola algorithm failure, now enshrined in multiple rulings, should stand as a warning: automation can lead to very painful consequences when rushed and ill-conceived, both for specific categories (in this case, the thousands of teachers unjustly displaced by the faulty automated system) and the general population (the taxpayers who are ultimately paying the price of their legitimate appeals).

A general principle seems to be emerging. If it is to be truly beneficial and trustworthy, and bring about what Paola Pisano from the Innovation and Digitization Ministry, defines as a “new Humanism of AI and a digital Renaissance”, machine intelligence must work with human intelligence, augmenting it rather than replacing it. Whether this means that full automation must be ruled out for public uses of ADM, or that we only need a better definition of what tasks and decisions should be fully automated is still a contentious issue. And it is an issue that will probably shape the Italian policy response to ADM-related issues — when there is one.

# References:

Aeroporti di Roma, Il controllo biometrico, <http://www.adr.it/controllo-biometrico#>

AgendaDigitale (2019), Giustizia predittiva: così a Brescia l'intelligenza artificiale aiuta a prevedere i tempi delle cause, <https://www.agendadigitale.eu/cultura-digitale/giustizia-predittiva-cosi-a-brescia-lintelligenza-artificiale-aiuta-a-prevedere-tempi-e-esiti-delle-cause/>

Altalex (2018), X-LAW: la polizia predittiva è realtà, <https://www.altalex.com/documents/news/2018/11/28/x-law-la-polizia-predittiva>

BVA DOXA (2019), Lavoro, la sfida dei robot e dell'IA, <https://www.bva-doxa.com/lavoro-la-sfida-dei-robot-e-dellia/>

BresciaOggi (2018), Tempi dei processi, un aiuto dalla "giustizia predittiva", <https://www.bresciaoggi.it/territori/citt%C3%A0/tempi-dei-processi-un-aiuto-dalla-giustizia-predittiva-1.6415532>

Business Insider (2019), Parla l'ispettore di polizia che ha creato un algoritmo che prevede scippi, rapine, furti. E funziona, da anni, <https://it.businessinsider.com/perla-il-dirigente-di-polizia-ha-creato-un-algoritmo-che-prevede-scippi-rapine-furti-e-funziona-da-anni/>

Corriere Milano (2019), Linate, riapre l'aeroporto ma il restyling durerà un anno e mezzo, [https://milano.corriere.it/notizie/cronaca/19\\_ottobre\\_23/liniate-riapre-l-aeroporto-ma-restyling-durera-anno-mezzo-5c30c9ee-f55c-11e9-ad25-1b4f9a2d3eb8.shtml](https://milano.corriere.it/notizie/cronaca/19_ottobre_23/liniate-riapre-l-aeroporto-ma-restyling-durera-anno-mezzo-5c30c9ee-f55c-11e9-ad25-1b4f9a2d3eb8.shtml)

Corriere della Sera (2019), Tasse, ecco come un algoritmo difettoso ti fa pagare più di quanto devi, <https://www.corriere.it/dataroom-milena-gabanelli/tasse-isa-ecco-come-algoritmo-difettoso-ti-fa-pagare-piu-quanto-devi/77bbc760-03dd-11ea-a09d-144c1806035c-va.shtml>

Corriere della Sera (2019b), Evasione, da social network e big data la nuova caccia ai "furbetti", [https://www.corriere.it/economia/tasse/20\\_gennaio\\_12/evasione-social-network-big-data-nuova-caccia-furbetti-016742ae-3521-11ea-b1ac-094115aa5a02.shtml](https://www.corriere.it/economia/tasse/20_gennaio_12/evasione-social-network-big-data-nuova-caccia-furbetti-016742ae-3521-11ea-b1ac-094115aa5a02.shtml)

CorCom (2019), Nasce l'interruppe parlamentare sull'intelligenza artificiale, <https://www.corrierecomunicazioni.it/digital-economy/nasce-lintergruppo-parlamentare-sullintelligenza-artificiale/>

Forbes Italia (2019), Riconoscimento facciale: l'Italia è ultima in Ue per la protezione della privacy, <https://forbes.it/2019/10/21/riconoscimento-facciale-litalia-ultima-in-europa-per-protezione-della-privacy/>

Giornale Monza (2019), Intelligenza artificiale in corsia: l'esempio dell'ASST di Vimercate, <https://giornaledimonza.it/attualita/intelligenza-artificiale-in-corsia-lesempio-dellast-di-vimercate/>

Il Giorno (2019), Vimercate, svolta hi-tech in ospedale: se ti ammalai lo prevede l'algoritmo, <https://www.ilgiorno.it/monza-brianza/cronaca/ospedale-vimercate-algoritmo-1.4593259>

Gonews (2019), Giustizia e tecnologia: l'intelligenza artificiale potrà prevedere le sentenze?, <https://www.gonews.it/2019/09/17/giustizia-tecnologia-lintelligenza-artificiale-potra-prevedere-le-sentenze/>

Il Messaggero (2020), Coronavirus, il futuro degli aeroporti: intelligenza artificiale e plexiglass, [https://www.ilmessaggero.it/roma/senzarete/aeroporto\\_coronavirus\\_fiumicino-5176984.html](https://www.ilmessaggero.it/roma/senzarete/aeroporto_coronavirus_fiumicino-5176984.html)

Ministero per l'Innovazione e la Digitalizzazione (2019), 2025 Strategia per l'innovazione tecnologica e la digitalizzazione del Paese, [https://innovazione.gov.it/assets/docs/MID\\_Book\\_2025.pdf](https://innovazione.gov.it/assets/docs/MID_Book_2025.pdf)

MISE (2019), Strategia nazionale per l'intelligenza artificiale, <https://www.mise.gov.it/images/stories/documenti/Strategia-Nazionale-Intelligenza-Artificiale-Bozza-Consultazione.pdf>

MISE (2020), Proposte per una strategia nazionale per l'intelligenza artificiale, [https://www.mise.gov.it/images/stories/documenti/Proposte\\_per\\_una\\_Strategia\\_italiana\\_AI.pdf](https://www.mise.gov.it/images/stories/documenti/Proposte_per_una_Strategia_italiana_AI.pdf)

Mobilita (2019), Aeroporto Fiumicino. Parte il riconoscimento facciale che permetterà di viaggiare senza documenti, <https://mobilita.org/2019/11/17/aeroporto-fiumicino-parte-il-riconoscimento-facciale-che-permetterà-di-viaggiare-senza-documenti/>

Polizia di Stato (2018), Venezia: i reati vengono previsti da un algoritmo, <https://www.poliziadistato.it/articolo/225bdd72498c36f619718140>

Repubblica (2019), Scuola, trasferimenti di 10 mila docenti lontano da casa. Il Tar: “L’algoritmo impazzito fu contro la Costituzione”, [https://www.repubblica.it/cronaca/2019/09/17/news/scuola\\_trasferimenti\\_di\\_10mila\\_docenti\\_lontano\\_da\\_casa\\_il\\_tar\\_l\\_algoritmo\\_impazzito\\_fu\\_contro\\_la\\_costituzione\\_-236215790/](https://www.repubblica.it/cronaca/2019/09/17/news/scuola_trasferimenti_di_10mila_docenti_lontano_da_casa_il_tar_l_algoritmo_impazzito_fu_contro_la_costituzione_-236215790/)

Repubblica Napoli (2016), All’aeroporto di Napoli arrivano i varchi automatici: controlli in 20 secondi [https://napoli.repubblica.it/cronaca/2016/01/19/news/all\\_aeroporto\\_di\\_napoli\\_arrivano\\_i\\_gates\\_automatici\\_controlli\\_in\\_20\\_secondi-131604084/](https://napoli.repubblica.it/cronaca/2016/01/19/news/all_aeroporto_di_napoli_arrivano_i_gates_automatici_controlli_in_20_secondi-131604084/)

Rete Disarmo (2019), Il 73% dei cittadini di 10 paesi europei vuole la messa al bando dei “killer robots”, <https://www.disarmo.org/rete/a/47017.html>

State Council (2019), ruling n. 2270, April 8, 2019, [https://www.avvocatirandogurrieri.it/images/easyblog\\_articles/9469/CdS-sentenza-n.-2270-dell8-aprile-2019.pdf](https://www.avvocatirandogurrieri.it/images/easyblog_articles/9469/CdS-sentenza-n.-2270-dell8-aprile-2019.pdf)

Sole 24 Ore (2018), Algoritmi, no grazie! I candidati vogliono un recruiter in carne e ossa, <https://www.ilsole24ore.com/art/algoritmi-no-grazie-candidati-vogliono-recruiter-carne-e-ossa-AE5zR8jG>

Sole 24 Ore (2019b), Se l’algoritmo prevede l’evoluzione delle malattie e la disponibilità dei letti, <https://www.ilsole24ore.com/art/se-l-algoritmo-prevede-l-evoluzione-malattie-e-disponibilita-letti-ACeQkjQ>

Sole 24 Ore (2019c), Dagli scontrini ai conti bancari, il Fisco lancia la lotta all’evasione con i dati anonimi, <https://www.ilsole24ore.com/art/dagli-scontrini-conti-bancari-fisco-lancia-lotta-all-evasione-i-dati-anonimi-AC5wGau>

TAR Lazio (2019), ruling n. 10964, section 3-bis, September 13, 2019, [http://images.go.wolterskluwer.com/Web/WoltersKluwer/%7B15d38a3f-3bd0-4c31-a437-b325dfa65f35%7D\\_tar-lazio-roma-sentenza-13-settembre-2019-numero-10964.pdf](http://images.go.wolterskluwer.com/Web/WoltersKluwer/%7B15d38a3f-3bd0-4c31-a437-b325dfa65f35%7D_tar-lazio-roma-sentenza-13-settembre-2019-numero-10964.pdf)

TgCom24 (2019), Volare si cambia: con riconoscimento facciale e intelligenza artificiale tornano i liquidi in valigia, [https://www.tgcom24.mediaset.it/tgtech/volare-si-cambia-con-riconoscimento-facciale-e-intelligenza-artificiale-tornano-i-liquidi-in-valigia\\_11228327-201902a.shtml](https://www.tgcom24.mediaset.it/tgtech/volare-si-cambia-con-riconoscimento-facciale-e-intelligenza-artificiale-tornano-i-liquidi-in-valigia_11228327-201902a.shtml)

Valigia Blu (2019), Perché diciamo no a una Commissione parlamentare d’inchiesta sulle “fake news”, <https://www.valigiablu.it/commissione-inchiesta-fake-news/>

Vice Italy (2018), La polizia predittiva è diventata realtà in Italia e non ce ne siamo accorti, <https://www.vice.com/it/article/pa5apm/polizia-predittiva-italia-lombardi-xlaw-prevedere-crimini-algoritmi>





BELGIUM  
DENMARK  
ESTONIA  
FINLAND  
FRANCE  
GERMANY  
GREECE  
ITALY

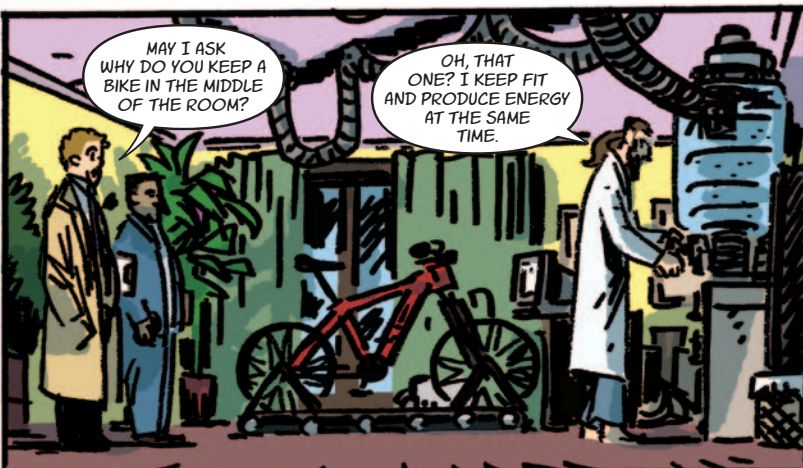
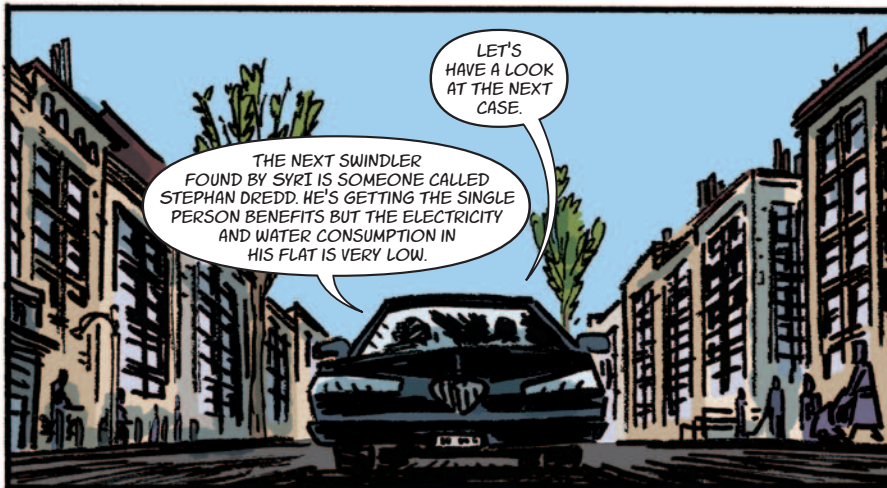
# **NETHERLANDS STORY**

PAGE 160

## **RESEARCH**

PAGE 164

POLAND  
PORTUGAL  
SLOVENIA  
SPAIN  
SWEDEN  
SWITZERLAND  
UNITED KINGDOM



Find out more in the story on the next page.

# How Dutch activists got an invasive **fraud detection** algorithm banned

**The Dutch government has been using SyRI, a secret algorithm, to detect possible social welfare fraud. Civil rights activists took the matter to court and managed to get public organizations to think about less repressive alternatives.**

By Koen Vervloesem



In its fight against fraud, the Dutch government has been cross-referencing personal data about citizens in various databases since 2014. This system, called SyRI (for “system risk indication”), seeks out “unlikely citizen profiles” that warrant further investigation. Despite major objections from the Dutch Data Protection Authority and the Council of State, SyRI has been implemented without any transparent process for citizens to see what happens to their data.

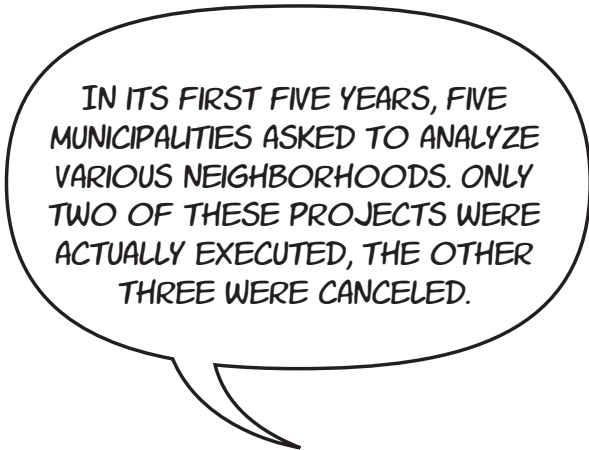
The idea is this: if some government agency suspects fraud with benefits, allowances, or taxes in a specific neighborhood, it can make use of SyRI. Municipalities, the Employee Insurance Agency (UWV), the social security bank, inspectors of the Ministry of Social Affairs, and the employment and the tax authorities all have access to the system. SyRI decides which citizens in the neighborhood require further investigation.

SyRI has not been a success for the government. In its first five years, five municipalities asked to analyze various neighborhoods. Only two of these projects were actually executed, the other three were canceled. According to research in 2019 by the Dutch newspaper *De Volkskrant*, none of these algorithmic investigations have been able to detect new cases of fraud.

## / False positives

Government agencies that want to use SyRI must follow a detailed procedure. Two agencies should cooperate and ask the Ministry of Social Affairs and Employment (SZW) to conduct an analysis. Before a SyRI project starts, SZW publishes an advisory in the online version of the official gazette. “The municipality has no obligation to inform citizens of a neighborhood that they are being analyzed,” said Ronald Huissen, from [Platform Bescherming Burgerrechten](#) (Platform for Civil Rights Protection). “And if they are informed, it is by a city bulletin that is not necessarily read by them, and in very vague terms, without the details of what data SyRI uses and how.”

The agency that asked for the analysis cannot just penalize citizens who are flagged for an unlikely combination of data: it has to investigate, for every flagged citizen, whether an actual case of fraud took place. Moreover, flagged citizens are first examined at the Ministry of Social Affairs and Employment for false positives. Data on citizens who are deemed false positives is not handed over to the agency that asked for the analysis.



IN ITS FIRST FIVE YEARS, FIVE MUNICIPALITIES ASKED TO ANALYZE VARIOUS NEIGHBORHOODS. ONLY TWO OF THESE PROJECTS WERE ACTUALLY EXECUTED, THE OTHER THREE WERE CANCELED.

## / No transparency

But even with these checks in place, the lack of transparency is still a big issue. Residents of whole neighborhoods were put under a magnifying glass without them even knowing which privacy-sensitive data SyRI had about them. Each ‘risk indication’ is logged into a register that citizens can look into if they ask. But citizens are not automatically warned if they are flagged as a fraud risk by SyRI, and they cannot access the reasons why they have been flagged.

At the beginning of 2018, Platform Bescherming Burgerrechten, together with a couple of other Dutch civil rights organizations, filed a case against the Dutch state to stop the use of SyRI. At the same time, they wanted to spark a public debate about SyRI via a media campaign [Bij Voorbaat Verdacht](#) (Suspected from the outset).

According to [the official resolution](#) that is the legal basis for SyRI, the system can cross-reference data about work, fines, penalties, taxes, properties, housing, education, retirement, debts, benefits, allowances, subsidies, permits, exemptions, and more. These are described so broadly that, in 2014, the Council of State concluded in its negative opinion on SyRI that there is “hardly any personal data that cannot be processed”.

## / Black box

SyRI pseudonymizes the data sources it uses with a ‘black box’ method. That is, for each data source that is linked, all citizen names are replaced by a unique identifier for each individual. The identifier makes it possible to link data about the citizen from these various data sources. After the analysis, the result is a list of identifiers that represent possibly fraudulent beneficiaries. These identifiers are then translated back to their real names.

In the case of Platform Bescherming Burgerrechten versus the Dutch State, the latter gave some examples of “discrepancies” that could lead to a risk indication. One of these discrepancies is a low usage of running water. This could be a sign that someone who receives benefits is living together with someone else at another address and thus does not have the right to the higher benefit for singles. However, there are many other possible causes for low water usage, such as using rainwater, a frugal life, or even a broken water meter.

## / A secret sauce

It is still unclear what is happening in this ‘black box’, and the Dutch government blocked all attempts from concerned parties to shed light on this. In 2017, the Ministry of Social Affairs decided that the risk models it used should be kept secret. In 2018, the political party D66 wrote a motion (but did not file it) to publish SyRI’s algorithms or to conduct a technical audit if publishing the algorithms is not possible.

Tamara van Ark, State Secretary for Social Affairs and Employment, strongly advised against filing the motion (so that it was never put to a vote), and she warned that potential offenders could adapt their behavior if the state disclosed SyRI’s risk models. But many of the factors in the risk models are already known or expected, or have already been used before SyRI to detect fraud, such as low water usage. It is hard to imagine that someone who commits fraud to get a higher benefit will leave the faucet open to increase their water usage.

## / Primarily used in low-income neighborhoods

There’s another problem with SyRI: according to freedom of information requests by Platform Bescherming Burgerrechten, it turns out SyRI has been primarily used in low-income neighborhoods. This exacerbates biases and discrimination: if the government only uses SyRI’s risk analysis in neighborhoods that are already deemed high-risk, it is no wonder that it will find more high-risk citizens there.

Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights, expressed his concerns about SyRI in [a letter](#) to the Dutch court on 26 September 2019: “Whole neighborhoods are deemed suspect and are made subject to special scrutiny, which is the digital equivalent of fraud inspectors

knocking on every door in a certain area and looking at every person’s records in an attempt to identify cases of fraud, while no such scrutiny is applied to those living in better-off areas.”

## / A repressive position

Mr. Alston does not question that welfare fraud exists and that it should be punished, but he warns that SyRI’s focus seems to be wrong: “If the focus on fraud seen to be committed by the poor is highly disproportionate to equivalent efforts targeted at other income groups, there is an element of victimization and discrimination that should not be sanctioned by the law.”

Maranke Wieringa, a PhD candidate at Utrecht University researching algorithmic accountability in Dutch municipalities, sees another problem: “One goal of municipalities using SyRI for specific neighborhoods is to improve their living standards. However, SyRI is not designed for that purpose. If you take a national instrument that is designed for fraud detection and then apply it with a social purpose of improving living standards and social cohesion in a neighborhood, you can question whether you should depart from the same repressive position for both goals.”

## / SyRI is not necessary

On 29 November 2019, SyRI won the Big Brother Award from the Dutch digital rights organization, Bits of Freedom. This prize is awarded to the biggest privacy intrusion of the year. When director-general Carsten Herstel accepted the award in the name of the Ministry of Social Affairs and Employment, he told the audience “I find it logical that the government gets alerted when someone gets a rental allowance and owns the house at the same time.”

“WHOLE NEIGHBORHOODS ARE  
DEEMED SUSPECT AND ARE MADE  
SUBJECT TO SPECIAL SCRUTINY, WHILE  
NO SUCH SCRUTINY IS APPLIED TO  
THOSE LIVING IN BETTER-OFF AREAS.”

PHILIP ALSTON



According to Mr. Huissen, from Platform Bescherming Burgerrechten, the government does not need this kind of mass surveillance to prevent fraud: “The government already has information about who owns which house, so it could check this before granting the person a rental allowance. For all big fraud scandals in social security we have seen in the past decades it became clear afterwards that they could have been prevented with simple checks beforehand. That happens far too little. It is tempting to look for solutions in secret algorithms analyzing big data sets, but often the solution is far simpler.”

Ms. Wieringa agrees that this is a better way. “SyRI has been introduced from the point of view of a repressive welfare state: it does not trust the citizens. But that is just one stance of many possible ones. For instance, the government could check, perhaps even while using fewer data sources, who has the *right* to an allowance.”

## / No fair balance

On 5 February 2020, the Dutch court of The Hague **ordered** the immediate halt of SyRI because it violates article 8 of the European Convention on Human Rights (ECHR), which protects the right to respect for private and family life. Article 8 requires that any legislation has a “fair balance” between social interests and any violation of the private life of citizens.

SyRI’s goal, or “social interest”, is to prevent and fight fraud. The Dutch state claimed that the SyRI legislation offered sufficient guarantees to do this while protecting the privacy of citizens, but the court disagreed. The legislation is insufficiently transparent and verifiable, and there are not enough safeguards against privacy intrusions, the judges wrote.

According to Ms. Wieringa, the court’s decision makes it clear that the biggest problem with SyRI is not that it is used to battle fraud (this is a legitimate aim, the court says), but it is the way it does it: “The system is deemed too opaque by the judges. If the government wants to ‘fix’ this problem, it will have to add more transparency to SyRI. A ‘SyRI 2.0’ will likely be SyRI but less opaque. Previous experiences point to that course of action. The “Waterproof” system, a forerunner of SyRI, was deemed illegal in 2007 on privacy grounds. Later, the government simply passed a law to circumvent the problem, thus creating SyRI. Another ‘lazy fix’, this time geared towards increasing transparency, would be a logical step for SyRI.”

## / A new way of dealing with algorithms

On the other hand, two public organizations in the Netherlands, the UWV and the tax authority, have reacted to the court’s decision by reassessing their own algorithmic systems for fraud detection. “This is a sign that the court’s decision is urging everyone to find a new way of dealing with algorithms in the public sector,” Ms. Wieringa added.

Tijmen Wisman, chairman of Platform Bescherming Burgerrechten, hopes that the government will do more. “Just adapting SyRI to be more transparent will still result in information asymmetry. Citizens do not want to give information anymore to their government if the latter can use this information in all possible ways against them.”

According to Mr. Wisman, public organizations need to change the way they manage data: “Data must no longer be allowed to roam freely, but must reside in an authentic data source. Each organization should keep logs for every time that their data is consulted. Citizens should be able to easily access these logs. This way it becomes clear to citizens what their data is used for, and they can challenge this use. This requirement of transparency also follows from the court’s ruling in the SyRI case, as well as the GDPR.”

In the Netherlands, welfare fraud is estimated at 150 million euros a year. Together, the municipalities, the UWV, and the social security bank have detected fraudulent claims totaling more than 744 million euros between 2013 and 2018. This compares to an estimated 22 billion euros lost to tax fraud, each year.

*Editor’s note: Since this article was written, the Dutch government introduced the Data Processing by Partnerships Act (Wet Gegevensverwerking door Samenwerkingsverbanden), on 24 April 2020, which was immediately dubbed “Super SyRI”. Several political parties and civil society organizations expressed dismay at the new project, which, they claim, does not address the grave faults of the initial SyRI.*

# Research

By Ronan Ó Fathaigh & Naomi Appelman

## Contextualization

The use of automated decision-making (ADM) in the Netherlands came to international attention in late 2019 and early 2020. This followed the intervention by one of the UN's top human rights experts in a high-profile court case in The Hague concerning the controversial SyRI (System Risk Indication) system. Professor Philip Alston, the UN Special Rapporteur on extreme poverty and human rights, described the case as the first legal challenge he was aware of that "fundamentally and comprehensively" contested the use of an ADM system in the welfare state on human rights grounds. Particularly noteworthy in the Court's ruling that declared SyRI to be in violation of Article 8 of the European Convention of Human Rights (Rechtbank Den Haag, 2020) [\[here\]](#) is that governments have a "special responsibility" for safeguarding human rights when implementing new technologies such as these automated profiling systems (*Idem*, par. 6.84). The case led to a major public and political debate on ADM systems, and the use of algorithms in the Netherlands, and by the government more generally.

In contrast, the previous edition of Automating Society described how ADM and artificial intelligence (AI) was "predominantly discussed as part of the larger Dutch strategy on digitization," and that no work, specifically focused on ADM and AI, had been done in the Netherlands on a national agenda. Yet, in the space of a year, a lot has changed. As we will discuss in this chapter, the Dutch government has now published a national Strategic Action Plan for Artificial Intelligence, with actions including major government investment in research on the legal aspects of decision-making algorithms, and the transparency, explainability, and supervision of algorithms. Indeed, in summer 2019, the main coalition government party, *Volkspartij voor Vrijheid en Democratie* (VVD), published a report proposing a new supervisory authority for the use of algorithms (Middendorp, 2019a) [\[here\]](#). In a similar vein, in autumn 2019, another coalition government party, *Democraten 66* (D66), also published proposals on the establishment of an oversight body for ADM systems. The proposals suggested prohibiting the

use of certain ADM systems (like SyRI) and automated face recognition systems until proper legislation and supervision is first put in place (D66, 2019) [\[here\]](#).

In this edition of Automating Society, we aim to highlight new applications of ADM systems in the Netherlands. This chapter provides updates on both the political and public debates, and the regulatory and self-regulatory measures concerning ADM and AI. We also look at how civil society and the media in the Netherlands continue to play a crucial role in the current public debate on ADM systems, and the government's use of algorithms that affect individuals. For example, in summer 2019, the Dutch public broadcaster, *Nederlandse Omroep Stichting* (NOS), revealed the widespread use of predictive algorithms by government agencies, based on confidential inventories on the use of such algorithms obtained by the broadcaster (Schellevis & de Jong, 2019) [\[here\]](#). This report led the Chairman of the Dutch Data Protection Authority (*Autoriteit Persoonsgegevens*) (AP) to state publicly that the Dutch government must be more transparent about its use of predictive algorithms (NOS, 2019) [\[here\]](#). Indeed, in November 2019, the AP announced that two of its high-risk focus areas for supervisory work during the 2020-2023 period would be digital government, and the use of algorithms and AI (*Autoriteit Persoonsgegevens*, 2019) [\[here\]](#). The AP is concerned about the increased use of algorithms and AI by more and more government authorities and private companies, which carries risks of harmful effects - while the irresponsible use of algorithms can lead to incorrect decisions, exclusion, and discrimination. The AP also highlighted an ongoing trend within local government and law enforcement authorities giving them access to large amounts of data on individuals. The AP added that it is important that government authorities handle such data responsibility.

The chapter also shows how it is not just the data protection authority that is increasing its oversight of the use of ADM systems, but that other regulatory authorities are too. A prominent example of this is the Dutch Central Bank (*De Nederlandsche Bank*) (DNB), which supervises financial institutions in the Netherlands. In 2019, the DNB recognized the

risks associated with the use of AI by financial institutions in decision-making processes (De Nederlandsche Bank, 2019) [\[here\]](#). As a result, the DNB published general principles for the use of AI in the financial sector, in order to ensure that financial institutions use AI in a responsible manner. The DNB will be examining the issue of responsible use of AI in its supervision of financial institutions in 2020.

## A catalog of ADM cases

### / Alternative Dispute Resolution

Over the past few years in the Netherlands, the use of online and automated dispute resolution systems as an alternative to costly and time-consuming court cases has increased. Organizations, such as the E-Court and Stichting Digitrage (Digitrage Foundation), are the most well-known examples. Legally, these dispute resolution mechanisms are a kind of automated arbitration, where both parties contractually agree to use one specific automated method to resolve a dispute. At the E-Court, the entire process is conducted online. The parties upload the required documents and an ADM system creates a verdict. A (human) arbiter then reviews that verdict and signs it. Especially in sectors with a high turnover of small court cases, these fast and cheap alternatives are attractive. For instance, in 2018 most Dutch health insurers, and many debt collection agencies, included these automated dispute resolution companies in their policy agreements (Kuijpers, Muntz & Staal, 2018) [\[here\]](#).

However, as the use of these automated online arbitration systems increased, a large societal backlash ensued. The E-Court, for example, was criticized for the opacity of its underlying system that produces the verdicts. This was due to the fact that the verdicts were not published, and because people were inadequately informed that they had the choice of going to a state court (Kuijpers, Muntz & Staal, 2018) [\[here\]](#). There was extensive negative reporting on these practices in the media, and a report by the Landelijke Organisatie Sociaal Raadslieden (National Association of Social Lawyers) heavily criticized the E-Court (Sociaal Werk Nederland, 2018) [\[here\]](#). These criticisms and concerns grew to such an extent that the lower courts decided to temporarily stop confirming the arbitration decisions of the E-Court, effectively shutting down their operation. This

resulted in an extended legal battle (e-Court, 2018) [\[here\]](#). In February 2020, Dutch state and e-Court were still engaged in this dispute, now arguing over whether a settlement has been reached (Driessen, 2020) [\[here\]](#). Despite these concerns, the Minister of Legal Protection recently stated that he views digital innovation and online arbitration, such as the E-Court, as positive developments, as long as they operate within the bounds of the law (Minister voor Rechtsbescherming, 2019a) [\[here\]](#).

### / Municipal Governments

ADM systems are used extensively at a municipal level in the Netherlands. Many municipal governments either already use ADM systems to aid them with executing their policies or are conducting pilot programs on how to productively use ADM. However, exact numbers are not known, and most municipal governments do not clearly explain how their systems operate, and how they are employed. This makes for a very fragmented, and quite vague landscape. Nevertheless, several clear cases can be found. For example, the municipal government of Amersfoort in the province of Utrecht is using an ADM system to predict which children are most at risk of dropping out of school in order to effectively allocate resources to help these children. Another example is a co-operation between several municipal governments

**AS THE USE OF  
AUTOMATED  
ONLINE  
ARBITRATION  
SYSTEMS  
INCREASED, A  
LARGE SOCIETAL  
BACKLASH  
ENSUED.**

to use an ADM system to create risk profiles for cases of welfare fraud (Schellevis & de Jong, 2019) [\[here\]](#). Another municipality, 's-Hertogenbosch, is looking into the ways in which ADM systems can help reintegrate people back into the labor market. The ADM system would predict the rate of success of a possible reintegration track. Based on the score the system comes up with, the municipal government would then decide whether to invest in such a track for an individual (Wesselink et al, 2018) [\[here\]](#). The four largest municipalities, Amsterdam, Rotterdam, The Hague, and Utrecht, use so-called 'scan vehicles' – cars equipped with smart cameras – to automatically detect parking violations, which has increased the number of fines considerably (Nieuwold 2018) [\[here\]](#). The city of Amsterdam is developing additional software to make the scan vehicles capable of recognizing litter polluting the city (AT5 2019) [\[here\]](#). The police are also cooperating with municipal governments in local projects involving ADM systems. For example, in the municipality of Roermond, the police are using "smart cameras" to detect shoplifters (Andringa, 2018) [\[here\]](#).

## Health Care

A striking example of ADM is the use of the growth tracking software, called Growth Watch, used in the Dutch health care system. The software tracks the development of children and automatically flags discrepancies that indicate either disease or child abuse (RTV Utrecht, 2019) [\[here\]](#). The software is said to contain several anomalies that make it unreliable for children under a certain age, or with a specific health history. In one instance the system flagged a case of probable child abuse, which resulted in an infant being unjustly separated from its parents (RTV Utrecht, 2018) [\[here\]](#) and (van Gemert et al., 2018) [\[here\]](#). This case prompted doctors to advise against using the software without a proper disclaimer or major adaptations (van Gemert, 2019) [\[here\]](#).

DUTCH LAW  
ENFORCEMENT  
IS EXPERIMENTING  
EXTENSIVELY WITH ADM  
SYSTEMS.

## Journalistic Reporting

Automated news reporting is also on the rise in the Netherlands. The previous Automating Society report discussed several news agencies that had implemented automated recommender-systems that semi-automatically decide which articles are shown to each visitor or subscriber. These trials have continued and have been deemed successful. In the past year, ADM in journalistic reporting has been taken

a step further, and now there are several examples of automatically generated news articles. For example, during the provincial elections of March 2019 the national broadcaster NOS published automatically generated articles with updates on the election results. The system works with standard texts and automatically inserts the latest results (Waarlo, 2019) [\[here\]](#) and (Duin, 2019) [\[here\]](#). A large private broadcaster, RTL, also recently implemented an automatic news generator. The automatic news generating software is called ADAM, short for "Automatische Data Artikel Machine" (automatic data article machine), and is mainly used to create local news, such as traffic updates. ADAM will also be used to generate articles on trends where much of the information is publicly available, such as national trends regarding schools or hospitals. The initiative was partially funded by the Google Digital News Initiative (Wokke, 2019) [\[here\]](#) and (Bunskoek, 2019) [\[here\]](#).

## Finance and Trading

One of the largest Dutch banks, the ING, has created several financial services that are based on ADM systems. For instance, the ING Smart Working Capital Assistant uses ADM to predict events "that have a positive or negative impact on the working capital of the company in question". Such assessments are sent directly to the client's smartphone, who can then use the information accordingly (Visser, 2018) [\[here\]](#). Another example is the ING Katana Lens that the bank created in cooperation with a pension fund. This web-based application is aimed at helping investors by giving them predictions of price trends for specific bonds. The ADM system works by collecting and analyzing past and current price data, and predicting the future price trends based on past patterns (Banken.nl, 2018) [\[here\]](#) and (Mpozika, 2019) [\[here\]](#). The ING also has a similar product, Katana, but this is aimed at the sales-side of bond trading and not investors (ING, 2019) [\[here\]](#).

## Law Enforcement Initiatives

Dutch law enforcement is experimenting extensively with ADM systems. The police are conducting several local pilots, such as the project in the municipality of Roermond, discussed earlier, and several other nationwide projects. ADM is also used to fine people for using a mobile device whilst driving. The Dutch national police issued a state-

ment that, from October 2019 onwards, the police will use advanced cameras to automatically detect drivers using phones while operating a vehicle, as this is illegal in the Netherlands. These so-called “smart” cameras will automatically flag when a driver is using a mobile device. The camera will then automatically take a picture of that person and register the license plate. This information is subsequently sent to a police officer for review. When the police officer determines that the information is correct, a fine of 240 euros is automatically sent to that driver’s home address. (RTL nieuws, 2019) [[here](#)] and (Politie, 2019b) [[here](#)].

The police are also using ADM to try and solve “cold cases” – or unsolved crimes. The national police conducted an experiment using artificial intelligence to analyze several cold cases, and to help prioritize which cases should be reopened. This experiment has been deemed successful and the police will continue using ADM to analyze cold cases. (Politie, 2018) [[here](#)] and (Tweakers Partners, 2019) [[here](#)]. Furthermore, the national police are actively collaborating with academia to develop ways in which artificial intelligence can help law enforcement. In 2019, the University of Utrecht established the National Police Artificial Intelligence Lab, where seven PhD candidates will, in close collaboration with the police, research how artificial intelligence can be used in a law enforcement context. This lab builds on the work already done in the Police Data Science Lab that the University of Amsterdam created in 2018 (Politie, 2019a) [[here](#)]. Another example of the Dutch police using ADM is in their online reporting system for Internet fraud. ADM is used to adjust the reporting questionnaire, give people advice, and automatically alert the police when there is a high likelihood that a crime is committed (Blik op nieuws, 2019) [[here](#)].

Of specific importance is the use of the predictive policing tool ‘Crime Anticipation System’ (CAS, *Criminaliteits Anticipatie Systeem*). Based on past police reports and conviction records, CAS predicts which places have the highest risk of specific types of violent crimes, with the aim to effectively spread police presence throughout a city to prevent these crimes from happening (Minister van Veiligheid en Justitie, 2015) [[here](#)]. On a bi-weekly basis, the system produces a 125 by 125-meter color coded grid indicating the highest risk areas where police presence is subsequently increased. Different versions of CAS have been in use since 2014 (Politie, 2017) [[here](#)].

Finally, the Dutch police’s Catch face recognition system is of particular importance. An investigation published by VICE in July 2019 uncovered the existence of over 2.2 million

images in the Catch system of a total of 1.3 million individuals who may be suspected of committing a serious criminal offense (van Gaal, 2019) [[here](#)]. In November 2019, the police confirmed that its face recognition system now has access to a complete database of people suspected of serious crimes. Current figures are unknown, but in 2017, 93 suspects were identified using the system (NU.nl, 2019) [[here](#)]. Furthermore, the national police introduced a smartphone app last year that allows police officers to send images to Catch (NU.nl, 2018) [[here](#)].

## Policy, oversight and debate

Having described the ADM systems currently in use in the Netherlands, this section examines how Dutch society is debating these developments. We will look at the debate from the perspective of the government, civil society and academia.

### / Government and Parliament

It was noted in the last Automating Society report that little work had been done on a national agenda concerning ADM or artificial intelligence by the Dutch government. However, over the last year, there have been major developments in terms of new national agendas concerning the use of ADM systems and artificial intelligence.

### / Strategic Action Plan for Artificial Intelligence

One of the most important developments at a policy level has been the Dutch government’s new *Strategisch Actieplan voor Artificiële Intelligentie* (Strategic Action Plan for Artificial Intelligence) (SAPAI), which was launched in October 2019, and contains a number of initiatives in relation to ADM and algorithms (De Rijksoverheid, 2019) [[here](#)] (see also the letter to parliament from the Minister for Legal Protection, 2019b) [[here](#)]. The SAPAI is built upon three tracks:

Track 1 involves “capitalising on societal and economic opportunities” presented by AI, and how the Dutch government will make optimal use of AI in the performance of public tasks. The government also intends to engage in intensive public-private partnerships to realize the benefits of AI.



Track 2 is “creating the right conditions”. This includes ensuring that the Netherlands has access to more usable data for AI applications to realize better AI developments.

Importantly, track 3 of the SAPAI is “strengthening the foundations”. On this track, the government will ensure that public values and human rights are protected. In this regard, the SAPAI’s actions will include research into the legal aspects of decision-making algorithms, research into the risks of face recognition technology, and European certification of AI applications in the administration of justice. In addition, there will be action on the effective supervision of algorithms, including research into the transparency/explainability of algorithms, and crucially, the supervision of algorithms. The Dutch government will also establish a transparency lab for government organizations, and stimulate the participation of Dutch companies and public organizations in the pilot phase of the ethical guidelines for AI from the High-Level Expert Group of the European Commission.

The SAPAI also tasked the *Nederlandse Organisatie voor Wetenschappelijk Onderzoek* (Dutch Research Council) (NWO) with funding research on explainable, socially conscious, and responsible AI. In November 2019, the NWO

launched the first national AI research agenda, (Artificial Intelligence Research Agenda for the Netherlands) (NWO, 2019) [[here](#)].

## / Dutch Digitalization Strategy 2.0

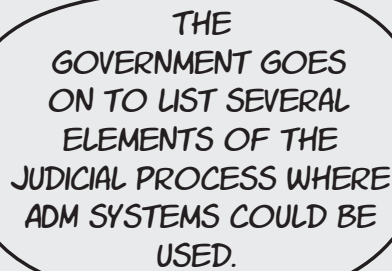
The previous edition of Automating Society noted that the first *Nederlandse Digitaliseringsstrategie: Nederland Digitaal* (Dutch Digitalization Strategy) was launched in 2018. In 2019, the Dutch government built upon this by implementing an updated *Nederlandse Digitaliseringsstrategie 2.0* (Dutch Digitization Strategy 2.0) (De Rijksoverheid, 2019b; 2019c) [[here](#) and [here](#)]. This strategy sets out the government’s priorities in relation to the impact of digitization on society, the economy, and government. These include applying AI solutions to resolve social issues and stimulate economic growth and developing digital inclusion, digital government, digital connectivity, and digital resilience. Notably, the strategy recognizes that AI can put pressure on key basic rights and public values (such as human dignity, autonomy, and the prohibition of discrimination). As a result, the importance of inclusiveness and transparency and the explicability of algorithms will be key issues. Furthermore, the potentials and risks of algorithm-based decision-making will also be assessed in the context of safeguarding public values; and the government will engage in case studies in the area of social security, law, self-driving cars, and content moderation. In March 2019, the Ministry of the Interior and Kingdom Relations also presented the NL DIGI-TAAL: Data Agenda Government to the House of Representatives (Ministerie van Binnenlandse Zaken, 2019a) [[here](#)].

In addition, in February 2019, the first Conferentie Nederland Digitaal (Digital Netherlands Conference) was held, bringing together government ministers, industry, researchers, and civil society, in order to better cooperate on digitization (Nederland Digitaal, 2019) [[here](#)]. Notably, Nederland ICT, an organization representing industry, presented its AI ethical code of conduct to the State Secretary for Economic Affairs.

Finally, in the previous Automating Society report, the government was supposed to react to a report on algorithms and fundamental rights during the summer of 2018. In March 2019, the government sent its response to parliament. In its reaction, the government acknowledges that new technologies can have a large impact on fundamental rights. The government mainly focused on the need for procedural safeguards and fairness. Central to its policy is neutrality, non-discrimination, transparency, and data pro-

**THE STRATEGY  
RECOGNIZES  
THAT AI CAN  
PUT PRESSURE  
ON KEY BASIC  
RIGHTS AND  
PUBLIC VALUES.**

tection. The government is looking into additional legal safeguards for reducing bias in its big data initiatives and it is creating guidelines to be used for the ex-ante review. Additionally, the government stressed the importance of robust data protection laws and effective enforcement (Minister van Binnenlandse zaken, 2019b) [[here](#)].



THE  
GOVERNMENT GOES  
ON TO LIST SEVERAL  
ELEMENTS OF THE  
JUDICIAL PROCESS WHERE  
ADM SYSTEMS COULD BE  
USED.

## / Supervision of ADM System and Algorithms

Members of the current Dutch government have also been quite active in terms of the debate concerning ADM systems and algorithms. As mentioned in the introduction above, in May 2019, the main coalition government party, VVD, published a report proposing a new supervisory authority for the use of algorithms (Middendorp, 2019a) [[here](#)]. The following month, Jan Middendorp, a government MP, sent an initiative to the House of Representatives, proposing supervision of the use of algorithms by the government (Middendorp, 2019b) [[here](#)].

Furthermore, in November 2019, another coalition government party, D66, also published a report with proposals to limit the use of ADM systems by government and technology companies (D66, 2019) [[here](#)]. The proposals include:

- creating a proper legal framework for the use of ADMs systems, where linking and analyzing individuals' data would only be permissible where it is required for a compelling purpose, and with sufficient guarantees for protecting individuals' human rights.
- establish a central overview of the number of databases, the degree of linking, and the use of ADM systems.
- establish an algorithm authority, and a reporting obligation for algorithm use.
- prohibit face recognition systems and certain uses of algorithms until there is proper legislation and oversight.

## / Algorithms and the Administration of Justice

In the previous Automating Society report, it was noted that the Minister for Legal Protection promised to send a letter to parliament in the fall 2018 on the possible meaning

of ADM and AI for the judicial branch.

At the time of publication, this letter had not been received. However, in December 2018, the Minister sent a letter to parliament about the application and use of algorithms and artificial intelligence in the administration of justice. The overall aim of the letter was to give an overview of current experiments in the field of ADM

and the administration of justice and to identify benchmarks for judicial use of ADM, based on fundamental rights. These benchmarks are accessibility, internal and external transparency, speed, independence, and impartiality of the judge and the fairness of the trial. The government goes on to list several elements of the judicial process where ADM systems could be used and also identifies possible pitfalls. The overall conclusion is to tread carefully, to conduct experiments apart from actual cases, and to properly assess each experiment to see what the added value of the use of technology is (Ministerie van Justitie en Veiligheid, 2018) [[here](#)]. Finally, in March 2020, the Minister for Legal Protection sent a letter to parliament in response to questions on algorithmic analysis of court judgments (Minister voor Rechtsbescherming, 2020) [[here](#)]. The Minister set out the benefits of algorithmic analysis of judgments, including allowing more insight into judicial reasoning and pointed to a scheme at the District Court of East Brabant on AI analysis of court documents and cases. The Minister confirmed that only 2-3% of court rulings are published online in the Netherlands, but it is planned to increase this to 5% over the next three years; and that there is currently no specific legal framework for applying AI to court rulings in the Netherlands.

## / Algorithms and Law Enforcement

The extensive use of ADM systems by the Dutch police, as described in the previous section, became a topic of debate in the Dutch Parliament after the Minister of Justice and Safety outlined the Government's policy on the use of AI by the police (Kamerstukken 2019) [[here](#)]. In response, the Parliamentary Committee for Justice and Safety drafted more than 80 questions. The Minister answered in February 2020 emphasizing, firstly, that the Dutch police currently is not using AI systems on a large scale with only the online Internet fraud reporting system as a concrete example. However, the Minister did confirm that algorithmic systems, as outlined above, are used throughout the Dutch police, and outlined several fields where algorithmic or AI systems

are being developed. Examples of these, are speech to text software, image/face recognition, and natural text processing, all to shorten the time needed for desk research. Further, the letter elaborated on the ethical, legal, and privacy safeguards in place for the use of AI. Crucially, the Minister clearly stated that it is the government's conviction that the future development and use of AI-systems for the police is deemed necessary (Kamerstukken 2020) [[here](#)].

## / Regulatory and self-regulatory measures

### / Automatic Number Plate Recognition

On 1 January 2019, a law that had been enacted in 2017 on Automatic Number Plate Recognition (ANPR) came into effect [[here](#)]. The civil society organization, Privacy First, is currently preparing legal proceedings, arguing that the law violates international and European privacy and data protection laws (Privacy First, 2019) [[here](#)].

### Data Processing Partnership Act

In the previous edition of Automating Society, it was reported that a public consultation was completed on a proposal for the Wet gegevensverwerking door samenwerkingsverbanden (Data Processing Partnership Act). The act aims to provide a basis for public-private cooperation, and make collaboration between them easier in relation to the processing of data, specifically used for surveillance or investigation purposes (e.g., to prevent crimes or detect welfare fraud). In June 2019, the Minister for Justice and Security sent a letter to parliament on the progress of the proposed law [[here](#)]. The Ministerraad (Council of Ministers) have agreed on the bill, and it has been sent to the Raad van State (Council of State) for its advice.

### / Data Protection Authority focus on AI and algorithms

As mentioned in the Introduction, the Autoriteit Persoonsgegevens (Dutch Data Protection Authority) (AP) announced in November 2019 that two of its high-risk focus areas for supervisory work during the 2020-2023 period would be digital government, and the use of algorithms and AI (Autoriteit Persoonsgegevens, 2019a) [[here](#)]. This is because the AP is concerned about the increased use of algorithms and AI by more government authorities and private

companies, which carries the risk of harmful effects, while the irresponsible use of algorithms can lead to incorrect decisions, exclusion, and discrimination. The AP also highlighted the continued trend of local government and law enforcement authorities now have access to large amounts of data on individuals, and that government authorities must handle such data responsibly.

In addition, the AP also announced in 2019, that it would be investigating the developments of Smart Cities by municipalities in the Netherlands, to ensure the privacy of residents and visitors is protected. The AP noted that municipalities are increasingly using data-driven and automated decisions and that the digital tracking of people in (semi) public places is an invasion of privacy that is only permitted in exceptional cases (Autoriteit Persoonsgegevens, 2019b) [[here](#)].

Finally, in April 2020, in the context of the coronavirus crisis, the AP released a statement on possible government use of location data from telecommunication companies to track the spread of the virus. The AP stated that, under Dutch privacy law and the Telecommunication Act (Telecommunicatiewet), telecom companies “may not simply share customer data with the government”. The AP concluded that statutory regulation would be needed, which would be required to be proportionate and contain sufficient safeguards (Autoriteit Persoonsgegevens, 2020) [[here](#)].

### / Dutch Central Bank recommendations on AI and ADM

In July 2019, the Dutch Central Bank (de Nederlandsche Bank, DNB) published a discussion document titled “General principles for the use of Artificial Intelligence in the financial sector”. This document was created as a response to the growing use of automated systems within the financial sector and is meant to give guidance to financial service providers planning to use ADM systems. The DNB explicitly recognizes the risks involved in using ADM in a financial context. The bank summarized their recommendations in the form of an acronym: “SAFEST”. This stands for soundness, accountability, fairness, ethics, skills, and transparency. The guidelines are meant to start a sector-wide conversation on the use of ADM. The DNB called on relevant stakeholders to send input and will report on the results at the end of 2020 (De Nederlandsche Bank, 2019) [[here](#)].

## / Civil Society and Academia

### / Civil Society

The previous Automating Society report listed several civil society organizations doing important work on ADM, including Bits of Freedom [\[here\]](#), De Kafkabrigade (The Kafka Brigade) [\[here\]](#), Platform Bescherming Burgerrechten (Platform for the Protection of Civil Rights) [\[here\]](#), and Privacy First [\[here\]](#). In the past year, these organizations have continued their efforts. For example, Bits of Freedom is actively campaigning against face recognition software and it recently published a report arguing for a ban on face recognition in public spaces. For that report, they conducted an experiment where they showed how even one publicly-accessible camera in the inner city of Amsterdam made it possible to identify the people walking past (Hooyman, 2019) [\[here\]](#). Furthermore, the Platform for the Protection of Civil Rights has been one of the driving forces behind, and the main claimant in, the SyRI system court case. This organization has been campaigning against the system for years (Platform Bescherming Burgerrechten, 2019) [\[here\]](#).

### / Public-Private Partnerships

Consistent with the Dutch government's policy of promoting public-private partnership, there are now several prominent examples of such partnerships. Two such examples are the launch of the Strategic Action Plan for AI by the State Secretary for Economic Affairs in 2019, and the Dutch AI Coalition, which, was also launched the same year (Nederlandse AI Coalitie, 2019) [\[here\]](#). It is a public-private partnership comprised of over 65 parties from industry, government, education, and research institutions (Martens, 2019; Wageningen University & Research, 2019) [\[here\]](#) and [\[here\]](#). In addition, the Kickstart AI Initiative is also relevant. This brings together a range of large Dutch companies Ahold Delaize, ING, KLM, Philips, and the NS, to fund AI research and education (Kickstart AI, 2019) [\[here\]](#).

### / Scientific Research

There is also a great deal of scientific research on ADM systems and AI taking place in Dutch universities. As mentioned previously, the Dutch Research Council (NWO) launched the first national AI research agenda in 2019. In addition, one of the most significant research initiatives has been the establishment of the national Innovation Center for Artificial Intelligence (ICAI). This is a national initiative focused on joint technology development between academia, industry and

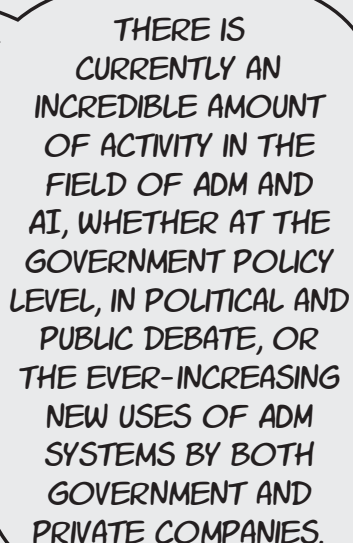
government in the area of artificial intelligence. The ICAI involves a number of research institutions, including Delft University of Technology, Radboud University, University of Amsterdam, Utrecht University, and Vrije Universiteit Amsterdam (Innovation Center for Artificial Intelligence, 2019) [\[here\]](#). There are a number of labs established under this collaboration, such as the National Police Lab AI. As mentioned above, at Utrecht University, the National Police Lab AI works in collaboration with the police, to research how artificial intelligence can be used in a law enforcement context.

Furthermore, at the University of Amsterdam, there is the Digital Transformation of Decision-Making research initiative. The purpose of this initiative is to examine how automated decision-making systems are replacing human decision makers in a range of areas, from justice, to media, commerce, health, and labor. In particular, researchers are looking at how the digitization of decision-making affects democratic values and the exercise of fundamental rights and freedoms (UvA, 2019a) [\[here\]](#). In relation to this, there is the AI & the Administration of Justice research initiative which is also at the University of Amsterdam. This initiative is engaged in normative research into the fundamental rights, societal values, and ethics of automated decision-making, specifically oriented towards the role and responsibilities of judges, public prosecutors, and lawyers (UvA, 2019b) [\[here\]](#). And finally, there is the Human(e) AI research priority area of the University of Amsterdam which stimulates new research on the societal consequences of the rapid development of artificial intelligence and ADM in a wide variety of societal areas (UvA, 2019c) [\[here\]](#).

## Key takeaways

This chapter reveals several important new findings about ADM and AI in the Netherlands during the past year. Perhaps the most important point is that there is currently an incredible amount of activity in the field of ADM and AI, whether at the government policy level, in political and public debate, or the ever-increasing new uses of ADM systems by both government and private companies. Secondly, it is striking that the use of ADM systems and AI by law enforcement has seen major development in the Netherlands. Thirdly, it also seems that a lot of the use of ADM systems by the government is in the realm of local government and government agencies that have discretion in how to allocate recourse, such as welfare, unemployment, tax

exemptions, etc. However, in terms of the challenges that ADM systems present, the case of the Netherlands perhaps highlights major issues concerning the lack of transparency with regards to ADM systems used by government. This has been one of the key criticisms of ADM systems, such as the SyRI system for welfare-fraud detection. Indeed, in the SyRI judgment, one of the biggest faults was the lack of transparency, with the Court even noting the “deliberate choice” by the government, of not providing “verifiable information” on the nature of SyRI. Finally, sector-specific regulators are also stepping-up, such as the data protection authorities and financial institution regulators, which are bringing a new level of guidance and scrutiny to both public and private uses of ADM systems.



THERE IS CURRENTLY AN INCREDIBLE AMOUNT OF ACTIVITY IN THE FIELD OF ADM AND AI, WHETHER AT THE GOVERNMENT POLICY LEVEL, IN POLITICAL AND PUBLIC DEBATE, OR THE EVER-INCREASING NEW USES OF ADM SYSTEMS BY BOTH GOVERNMENT AND PRIVATE COMPANIES.

## References:

Andringa, Remco (2018) Politie wil zakkenrollers en plofkraken vangen met data, *NOS Nieuws*, <https://nos.nl/artikel/2250767-politie-wil-zakkenrollers-en-plofkrakers-vangen-met-data.html>

AT5 (2019) 'Dijksma wil scanauto gebruiken in strijd tegen afvalbergen' <https://www.at5.nl/artikelen/196944/dijksma-wil-scanauto-gebruiken-in-strijd-tegen-afvalbergen>

Autoriteit Persoonsgegevens (2019a) AP legt focus in toezicht op datahandel, digitale overheid en AI, <https://autoriteitpersoonsgegevens.nl/nl/nieuws/ap-legt-focus-toezicht-op-datahandel-digitale-overheid-en-ai>

Autoriteit Persoonsgegevens (2019b) Waarborg privacy in de ontwikkeling van Smart Cities, <https://autoriteitpersoonsgegevens.nl/nl/nieuws/waarborg-privacy-de-ontwikkeling-van-smart-cities>

Autoriteit Persoonsgegevens (2020) Gebruik telecomdata tegen corona kan alléén met wet, <https://autoriteitpersoonsgegevens.nl/nl/nieuws/gebruik-telecomdata-tegen-corona-alleen-met-wet>

Banken.nl (2018) ING verbetert trading met kunstmatige intelligentie, <https://www.banken.nl/nieuws/21220/ing-verbetert-trading-met-kunstmatige-intelligentie>

Bunskoek, Jasper (2019) RTS Nieuws introduceert redactierobot ADAM, *rtl nieuws*, <https://www.rtlnieuws.nl/nieuws/artikel/4914671/adam-robot-nieuws-journalistiek-innovatie-rtl-nieuws-google-dni>



Blink op nieuws (2019) Sneller advies bij aangifte, <https://www.blikopnieuws.nl/digitaal/276509/sneller-advies-bij-aangifte-internetoplichting-door-kunstmatige-intelligentie-ai>

Driessen, Camil (2020), 'Robotrechter en staat ruziën over schikking' *het NRC*, <https://www.nrc.nl/nieuws/2020/02/26/robotrechter-en-staat-ruzien-over-schikking-a3991842>

Duin, Roelf Jan (2019), Een robot kan dit (nog) niet schrijven, *Het Parool*, <https://www.parool.nl/kunst-media/een-robot-kan-dit-nog-niet-schrijven-bc838dbf/>

D66 (2019) D66 pleit voor digitale revolutie, <https://d66.nl/digitale-revolutie/>.

e-Court persberichten (2018) e-Court sleept rechterlijke macht voor de rechter, <http://www.e-court.nl/persberichten/>

van Gaal (2019) Gezichtsherkenning op de Nederlandse straten: moeten we dat willen?, *Motherboard*, <https://www.vice.com/nl/article/8xyzdz/gezichtsherkenning-op-de-nederlandse-straten-moeten-we-dat-willen>

van Gemert e.a. (2018) Pediatric Condition Falsification Misdiagnosed by Misjudged Weight Growth from the Curve of Measured Weights. *Am J Case Rep.* 2018, <https://www.ncbi.nlm.nih.gov/pubmed/29946058>

van Gemert e.a. (2019) Limitations of Weight Velocity Analysis by Commercial Computer Program Growth Analyser Viewer Edition, *Annals of Biomedical Engineering* 2019/47, <https://link.springer.com/article/10.1007/s10439-018-02118-8>

Hooyman, Paula (2019) Het ware gezicht van gezichtsherkenningstechnologie, <https://www.bitsoffreedom.nl/wp-content/uploads/2019/11/het-ware-gezicht-van-gezichtsherkenningstechnologie.pdf>

ING (2019) Katana named innovator 2019, <https://www.ing.com/Newsroom/News/Katana-named-Innovator-2019.htm>

Kamerstukken (2019), 2019-2020, 26 643 en 32 761, nr. 652, <https://zoek.officielebekendmakingen.nl/kst-26643-652.html>

Kamerstukken (2020), 2019-2020, 26 643 en 32 761, nr. 669, <https://zoek.officielebekendmakingen.nl/kst-26643-669.html>

Kuijpers, Karlijn, Thomas Muntz and Tim Staal (2018), Vonnis te koop, *De Groene Amsterdammer*, <https://www.groene.nl/artikel/vonnis-te-koop>

Landelijke Organisatie Sociaal Raadslieden LOSR/Sociaal Werk Nederland, (2018) Rechtspraak op bestelling?! Stop commerciële rechtspraak, [https://schuldingo.nl/fileadmin/Publicaties/Rechtspraak\\_op\\_bestelling\\_Stop\\_commerciele\\_rechtspraak.pdf](https://schuldingo.nl/fileadmin/Publicaties/Rechtspraak_op_bestelling_Stop_commerciele_rechtspraak.pdf)

Middendorp, Jan (2019a) Initiatiefnota Menselijke grip op algoritmen, <https://www.vvd.nl/content/uploads/2018/11/35212-2-Initiatiefnota-van-het-VVD-lid-Middendorp-Menselijke-grip-op-algoritmen-PDF.pdf>

Middendorp, Jan (2019b) Initiatiefnota: Menselijke grip op algoritmen, <https://www.tweedekamer.nl/kamerstukken/detail?id=2019D22106>

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2019a) Aanbiedingsbrief bij NL Digitaal: Data Agenda Overheid, <https://www.rijksoverheid.nl/documenten/kamerstukken/2019/03/15/aanbiedingsbrief-bij-nl-digitaal-data-agenda-overheid>

Minister van Binnenlandse Zaken en Koninkrijksrelaties (2019b) Kamerbrief met reactie op UU-rapport Algoritmes en grondrechten, <https://www.rijksoverheid.nl/documenten/kamerstukken/2019/03/28/kamerbrief-met-reactie-op-uu-rapport-algoritmes-en-grondrechten>

Minister van veiligheid en Justitie (2015) Antwoord op vragen van het Lid Marcouch, Tweede Kamer, 2014-2015, 1321, <https://zoek.officielebekendmakingen.nl/ah-tk-20142015-1321.html>

Ministerie van Justitie en Veiligheid (2018) Kamerbrief over toepassing artificieel intelligentie en algoritmen in de rechtspleging, <https://www.rijksoverheid.nl/documenten/kamerstukken/2018/12/19/ek-artificieel-intelligentie-en-algoritmen-in-de-rechtspleging>

Minister voor Rechtsbescherming (2019a), Antwoord op vragen van het lid Jasper van Dijk over het bericht 'deurwaarders die onterecht kosten in rekening brengen', <https://www.tweedekamer.nl/kamerstukken/kamervragen/detail?id=2019D47735&did=2019D47735>

Minister voor Rechtsbescherming (2019b) Brief van de Minister voor Rechtsbescherming aan de Voorzitter van de Tweede Kamer der Staten-Generaal, <https://zoek.officielebekendmakingen.nl/kst-26643-641.html>

Minister voor Rechtsbescherming (2020), Antwoord op vragen van het lid Groothuizen over algoritmische analyse van vonnissen, <https://www.tweedekamer.nl/kamerstukken/kamervragen/detail?id=2020Z03232&did=2020D11984>

Mpozika, Nicole (2019) Disentangling the Bond Universe with Katana Lens, *Medium*, <https://medium.com/wbaa/disentangling-the-bond-universe-with-katana-lens-4bd271fe97c2>

Nederlandse AI Coalitie (2019) <https://nlaic.com/>

Nederland Digitaal (2019) Opbrengsten Conferentie Nederland Digitaal 2019, <https://www.nederlanddigitaal.nl/actueel/nieuws/2019/03/21/opbrengsten-conferentie-nederland-digitaal-2019>

De Nederlandsche Bank (2019) DNBulletin: DNB komt met richtlijnen voor gebruik kunstmatige intelligentie, <https://www.dnb.nl/nieuws/nieuwsoverzicht-en-archief/DNBulletin2019/dnb385020.jsp>

Niewold, Michaël (2018), 'Inzet scanauto's: meer opbrengst, minder agressie', RTL Nieuws, <https://www.rtlnieuws.nl/economie/artikel/3788081/inzet-scanautos-meer-opbrengst-minder-agressie>

NOS (2019) Privacywaakhond: overheid moet transparanter zijn over algoritmes, NOS Nieuws, <https://nos.nl/artikel/2286880-privacywaakhond-overheid-moet-transparanter-zijn-over-algoritmes.html>

NPPL (2019a) Beregenen op maat met Raindancer, <https://www.proeftuinprecisielandbouw.nl/beregenen-op-maat-met-raindancer/>

NPPL (2019b) Sensor-gestuurde gewasbescherming in fruit, <https://www.proeftuinprecisielandbouw.nl/techniek/sensor-gestuurde-gewasbescherming-in-fruit/>

NU.nl (2018) Gezichtsherkenning politie Identificeerde 93 verdachten In 2017, <https://www.nu.nl/internet/5141948/gezichtsherkenning-politie-identificeerde-93-verdachten-in-2017.html>

NU.nl (2019) Gezichtendatabase van politie bevat foto's van 1,3 miljoen mensen, <https://www.nu.nl/tech/5967571/gezichtendatabase-van-politie-bevat-fotos-van-13-miljoen-mensen.html>

NWO (2019) Artificial Intelligence Research Agenda for the Netherlands, <https://www.nwo.nl/binaries/content/documents/nwo/algemeen/documentation/application/nwo/beleid/open-science/ai-research-agenda-for-the-netherlands-airea-nl/AI+Research+Agenda+for+the+Netherlands+%28AIR+EA-NL%29>

Platform Bescherming Burgerrechten (2019) Uitspraak in SyRI-rechtszaak 5 februari: lees de pleitnota van onze coalitie, <https://platformburgerrechten.nl/2019/10/29/uitspraak-in-syri-rechtszaak-5-februari-lees-de-pleitnota-van-onze-coalitie/>

Politie (2017) Criminaliteits Anticipatie Systeem verder uitgerold bij Nationale Politie <https://www.politie.nl/nieuws/2017/mei/15/05-cas.html>

Politie (2018) Nieuwe technologie in oude politiezaken, <https://www.politie.nl/nieuws/2018/mei/23/00-nieuwe-technologie-in-oude-politiezaken.html>

Politie (2019a) Kunstmatige intelligentie vergroot onze slagkracht, 16 januari 2019, <https://www.politie.nl/nieuws/2019/januari/16/%E2%80%98kunstmatige-intelligentie-vergroot-onze-slagkracht.html>

Politie (2019b) Inzet slimme camera's tegen afleiding In het verkeer, 30 September 2019, <https://www.politie.nl/nieuws/2019/september/30/00-inzet-slimme-cameras-tegen-afleiding-in-het-verkeer.html>

Privacy First (2019) Submission of Privacy First to the fifth periodic review of the Netherlands by the UN Human Rights Committee, [https://www.privacyfirst.nl/images/stories/PDFs/Privacy\\_First\\_submission\\_addendum\\_HRC\\_2019.pdf](https://www.privacyfirst.nl/images/stories/PDFs/Privacy_First_submission_addendum_HRC_2019.pdf)

Rechtbank Den Haag (2020), 5 February 2020, ECLI:NL:RBDHA:2020:865. Official Dutch language version: <http://deeplink.rechtspraak.nl/uitspraak?id=ECLI:NL:RBDHA:2020:865>. Unofficial English translation: <http://deeplink.rechtspraak.nl/uitspraak?id=ECLI:NL:RBDHA:2020:1878>.

De Rijksoverheid (2019a) Strategisch Actieplan voor Artificiële Intelligentie, <https://www.rijksoverheid.nl/documenten/beleidsnotas/2019/10/08/strategisch-actieplan-voor-artificiele-intelligentie>

De Rijksoverheid (2019b) Nederlandse Digitaliseringsstrategie 2.0, <https://www.rijksoverheid.nl/documenten/rapporten/2019/07/05/nederlandse-digitaliseringsstrategie-2.0>

De Rijksoverheid (2019c) Dutch Digitisation Strategy 2.0, <https://www.nederlanddigitaal.nl/binaries/nederlanddigitaal-nl/documenten/publicaties/2019/11/13/english-version-of-the-dutch-digitalisation-strategy-2.0/Dutch+Digitisation+Strategy+2.0.pdf>

RTL nieuws (2019) Politie zet vanaf vandaag camera's in tegen appende automobilisten, <https://www.rtlnieuws.nl/nieuws/nederland/artikel/4867361/camera-controle-appen-achter-stuur-auto-politie-boete-240-euro>

RVT Utrecht (2018) Wetenschappers kritisch over kinderartsen en Veilig Thuis, <https://www.rtvutrecht.nl/nieuws/1793482/wetenschappers-kritisch-over-kinderartsen-en-veilig-thuis.html>

RTV Utrecht (2019) Wetenschappers: „Software bij kinderartsen geeft mogelijk foute uitkomst“, <https://www.rtvutrecht.nl/nieuws/1970056/wetenschappers-software-bij-kinderartsen-geeft-mogelijk-foute-uitkomst.html>

Schellevis, Joost, and Winny de Jong (2019), Overheid gebruikt op grote schaal voorspellende algoritmes, „risico op discriminatie“, *NOS Nieuws*, 29 May 2019 <https://nos.nl/artikel/2286848-overheid-gebruikt-op-grote-schaal-voorspellende-algoritmes-risico-op-discriminatie.html>

Tweakers Partners (2019) Kunstmatige intelligentie moet politie helpen bij cold cases, 10 April 2019, <https://tweakers.net/plan/2072/kunstmatige-intelligentie-moet-politie-helpen-bij-cold-cases.html>

University of Amsterdam (2019a) The Digital Transformation of Decision-Making, <https://www.uva.nl/en/faculty/amsterdam-law-school/research/research-themes/digital-transformation/digital-transformation-home.html>

University of Amsterdam (2019b) Kick-off meeting AI & the Administration of Justice, <https://www.uva.nl/shared-content/faculteiten/nl/faculteit-der-rechtsgeleerdheid/evenementen/evenementen/2019/09/kick-off-meeting-ai--the-administration-of-justice.html>

University of Amsterdam (2019c) RPA Human(e) AI, <https://humane-ai.nl/>

Visser, Cees (2018) „ING neemt ai-applicatie op in bankomgeving“, *Computable*, <https://www.computable.nl/artikel/nieuws/development/6417555/250449/ing-neemt-ai-applicatie-op-in-bankomgeving.html>

Waarlo, Niels (2019) NOS zet robotjournalist in voor verkiezingsuitslagen, *de Volkskrant*, <https://www.volkskrant.nl/cultuur-media/nos-zet-robotjournalist-in-voor-verkiezingsuitslagen~b0e1159e/>

Wageningen University & Research (2019), Precisielandbouw, <https://www.wur.nl/nl/Dossiers/dossier/dossier-precisielandbouw.htm>

Wesselink, Henk et al. (2018), Datedreven sturing bij gemeenten. Van data tot (gedeelde) informatie voor beter (samen) sturen, *Vereniging Nederlandse Gemeenten, Berenschot*, [https://vng.nl/files/vng/nieuws\\_attachments/2018/datedreven\\_sturing\\_bij\\_gemeenten\\_lr.pdf](https://vng.nl/files/vng/nieuws_attachments/2018/datedreven_sturing_bij_gemeenten_lr.pdf)

Wokke, Arnoud (2019) RTL Nieuws gebruikt software om lokale artikelen te maken, *Tweakers*, <https://tweakers.net/nieuws/159930/rtl-nieuws-gebruikt-software-om-lokale-artikelen-te-maken.html>



BELGIUM  
DENMARK  
ESTONIA  
FINLAND  
FRANCE  
GERMANY  
GREECE  
ITALY  
NETHERLANDS

**POLAND**

**STORY**

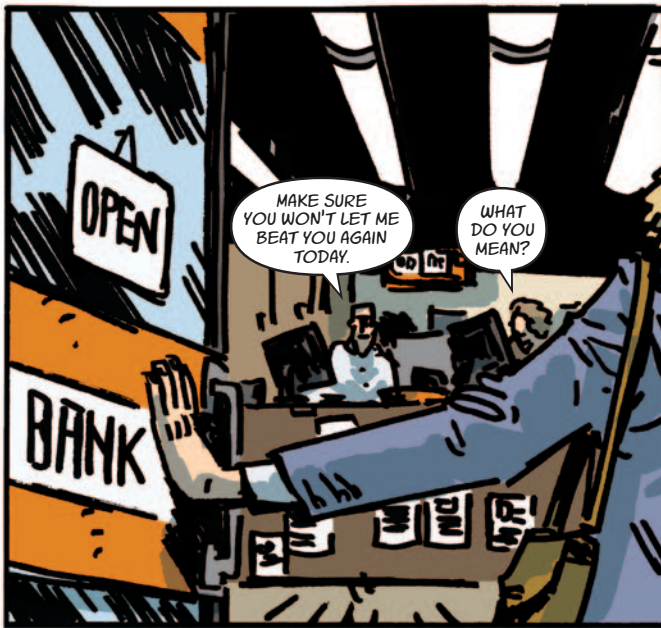
PAGE 178

**RESEARCH**

PAGE 183

PORTUGAL  
SLOVENIA  
SPAIN  
SWEDEN  
SWITZERLAND  
UNITED KINGDOM





Find out more on p. 185 in the research chapter under "Smile-monitoring of employees and consumers by PKO bank".



# Pre-crime at the **tax office**: How Poland automated the fight against VAT fraud.

In their fight against fraud, Polish tax authorities use STIR, an algorithm that sifts through the data of millions of entrepreneurs. The government claims success, but dozens of companies have been hit, some say wrongly.

By Konrad Szczygieł

“We have broken the group of VAT fraudsters”, “We have detected the artificial vegetable oil trading being a carousel fraud”, “National Revenue Administration liquidated the Asian mafia thanks to STIR”.

These are just a few of the headlines from recent months. They showcase the success of the tax authorities in their fight against tax fraud, and all of them feature STIR prominently.

## / It's not about stirring

To begin with, a brief explanation – STIR has nothing to do with stirring. This is an abbreviation (*System Teleinformatyczny Izby Rozliczeniowej*) for a technical system aimed at automatically identifying suspicious transactions and fraudulent entrepreneurs.

STIR was set up in 2017 by the government but it was built by the National Clearing House (*Krajowa Izba Rozliczeniowa*, KIR), a private company owned by private banks as well as the national bank, which holds a third of the shares. KIR is a technical hub that provides Polish banks with the infrastructure they need for inter-bank transactions.

Representatives of the Ministry of Finance who discussed STIR with AlgorithmWatch described STIR as a “warehouse of data”. “It is not a magic box showing final results. There is always a human at the end of it, analyzing what the system found,” said one of them, who wishes to remain anonymous.

Since April 2018, the main user of STIR has been the National Revenue Administration (*Krajowa Administracja Skarbowa*, KAS). This new public authority was set up in 2017 after the Law and Justice (PiS) government made fighting VAT (value-added tax) fraud a priority. KAS was born of the merger of the tax administration and the customs service.

## / VAT carousel

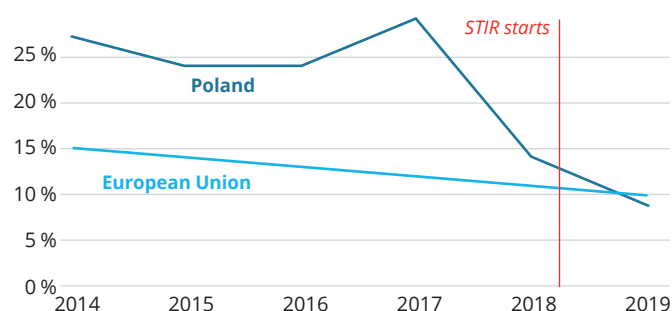
As in all Member States of the European Union, VAT fraud is a large and profitable business. It is commonly measured using the “VAT gap”, which represents the difference between expected VAT revenues and actual VAT revenues (part of the gap is due to other causes, such as bankruptcies).

In 2018, according to CASE, a think-tank, the VAT gap was about 140 billion euros in the EU and 5.8 billion euros in Poland, with a clear downward trend.

The fraud, also known as the VAT carousel, involves at least two companies. Company A sells a product with VAT to company B, but does not transfer the VAT it received to the public treasury as is legally required. Company B, which paid VAT to company A, then asks the public treasury for reimbursement, as it legally should. In real-life, the scheme involves many more intermediary companies and cross-border transactions.

## Closing the gap

Missing VAT as a share of expected VAT.



Data for 2018 is preliminary

Chart: AlgorithmWatch Source: CASE Created with Datawrapper

## / A data warehouse

What information about Polish entrepreneurs does STIR collect? Quite a lot. The list includes all bank accounts opened and maintained by a company, including the daily statements of transactions, identification data of senders and recipients of transactions and their account numbers; the date, amount, currency, title, and description of transactions, and the initial and final balances of statements. Since July 2019, banks must also provide STIR with the IP addresses from which account holders log into their accounts.

“Determining the IP addresses which were used to carry out untypical transactions and the dates and times of these transactions will allow [us] not only to identify the persons, but also to check the places from which the instructions to carry out these transactions were made, the links between these places and the persons using these IP addresses,” the Ministry of Finance explained in a memorandum.

## / Over 5 million people under surveillance

Data [released](#) by KAS, the National Revenue Administration, shows that STIR collected data on over 11 million transactions in 2019, related to close to four million entities. Five and a half million individuals were connected to these entities, according to the KAS report.

STIR can be accessed by analysts working in a special unit of KAS. Every day, reports from STIR land on their desks, which include information on transactions that were automatically labeled as suspicious as well as “entities classified as high-risk groups using the financial sector for tax evasion”, according to KAS documents.

Analyses based on STIR data are then submitted for approval to the head of the National Revenue Administration, who can decide to freeze bank accounts that are deemed suspicious. Freezing an account can be done without informing the bank or the account owner, and can last 72 hours. The freeze can then be extended up to three months. Since March 2019, the decision to freeze an account can also be made by regional heads of tax offices.

STIR does not seem to have links to judicial authorities. It is less a tool to find past fraud and bring wrongdoers to justice than a predictive tool aimed at preventing fraud.

## / Ministerial satisfaction

How effective is STIR? A glance at the statistics shows that the system is developing slowly.

In 2018, thanks to the algorithm, 41 accounts belonging to 23 entities [were frozen](#), containing about 10.2 million Polish złoty (2.3 million euros). KAS reported that the operation prevented the theft of 132 million złoty (30 million euros) in VAT fraud, slightly above 0,5% of the estimated VAT gap.

The number [increased](#) five-fold in 2019. 537 accounts belonging to 113 entities were frozen, holding over 67 million złoty (15 million euros) in assets. The estimated savings for the state coffers amount to 584 million złoty (133 million euros).

According to the latest data we obtained, data from STIR has already lead to the freezing of 383 accounts in 2020.

SINCE 2018, 52 COMPLAINTS  
HAVE BEEN FILED BY  
ENTREPRENEURS WHOSE BANK  
ACCOUNTS HAD BEEN FROZEN  
AFTER A DECISION BY KAS  
BASED ON STIR DATA.

The Ministry of Finance is satisfied with the algorithm, according to a [report](#) by the newspaper, *Dziennik Gazeta Prawna*. “In December 2019, the Ministry hosted a delegation of the Chinese treasury to learn about the functioning of the latest solutions in the implementation of the split payment mechanism and the STIR system,” Mariusz Szulc, the journalist, wrote. A system similar to STIR, the Central Electronic System of Payment information ([CESOP](#)), is expected to take effect throughout the EU in 2024. It will collect information on the recipients and senders of cross-border transfers.

## / Seeking redress

Entrepreneurs whose accounts have been blocked by KAS, based on the result of STIR analysis, are often helpless in the fight against the algorithm. They seek help in the courts.

Until March 2019, when regional tax offices received the right to freeze bank accounts themselves, the Voivodeship Administrative Court in Warsaw was the only place where entrepreneurs could file complaints against KAS. Judge Ewa Marcinkowska of the Court Information Department told AlgorithmWatch that, since 2018, 52 complaints have been filed by entrepreneurs whose bank accounts had been frozen after a decision by KAS based on STIR data.

Ms. Marcinkowska added that in 41 cases the court dismissed the complaints, and thus ruled against the entrepreneurs.

## / No investigation needed

In a judgment of the Warsaw Administrative Court of 20 September 2018, judges wrote that the tax office does not have to conduct evidentiary proceedings when freezing an account.

In January 2019, the court made itself even clearer, stating that “a rational entrepreneur should reckon with the fact that

the authority may apply procedures in force against him, in particular those consisting in blocking [their bank] account”.

## / Complaining to the ombudsperson

What can a long-term account lockout mean for an entrepreneur? This can be seen in the content of complaints received by the office of the Polish Commissioner for Human Rights. “In one of the complaints, the complainant raised general allegations about the nature of the provisions introduced, claiming that this was an abuse of power. In his opinion, the blocking of the entrepreneur’s funds may cause the company to become insolvent and consequently bankrupt,” said Łukasz Starzewski, from the press office of the Commissioner for Human Rights. The entrepreneur pointed out that the most painful thing in the case of account blockage was the inability to pay out salaries to employees. (Current legislation allows for lifting an account freeze to pay salaries, but the procedure is long and cumbersome.)

*Rzeczpospolita*, a daily newspaper, [wrote](#) about this case: “The blockade ‘by surprise’ prevented the entrepreneur from defending his rights, leading to the deregistration of his company from the register of active VAT payers. The Commissioner for Human Rights stated that such actions by the tax authorities may violate civil rights and freedoms and the rule of law”.

## / False positives

How does KAS explain their approach? Their philosophy is simple: better safe than sorry. In simple terms, the idea is to ensure that the money belonging to the state treasury will not disappear from accounts of suspected companies, for example.

But there is a silver lining for entrepreneurs. In June 2020, another court, in Poznań, [changed](#) its stance, compared to the Warsaw court. The Poznań judges ruled that if the decision to block an account for 72 hours is insufficiently justified, it would also be illegal to extend it for three months.

## / The secret algorithm

What kind of algorithm does STIR use? What does it look at? It is a closely guarded secret. One official from the Ministry of Finance I talked to, when asked about the details of the algorithm, quickly cut the conversation off, saying that the cook does not reveal the secrets of his recipes.

According to the KAS press office, “the tasks of KIR include running STIR and processing data in this system, determining algorithms and indicators of risk”.

I then asked the National Clearing House for more details. Agnieszka Sobczak-Malinowska, from the KIR press office, referred me to the provisions of the law that allowed the deployment of STIR. “Due to the confidentiality of information about STIR, also protected by law, it is not possible to provide data in response to your questions in a wider scope than it results from the above-mentioned documents,” she said.

The [law](#) in question, passed on 24 November 2017, does not answer our questions. It refers in general terms to the use of the algorithm and that KAS, by means of STIR, is tasked to establish risk indicators.


The 2017 law states that STIR’s algorithm is to be developed by KIR, taking into account “best practices of the banking sector and the cooperative savings and credit unions in terms of preventing their activities from being used for fiscal evasion.” According to the law, the algorithm, when selecting suspicious transactions, uses criteria such as whether an entity made unusual, sudden movements on its bank account and whether it transfers money to “countries where there is a high risk of fraud”.

## / Stirring discontent

Since the beginning of work on STIR, the lack of transparency related to the algorithm has been a matter of debate. Objections were raised by NGOs, including Panoptykon, which advocates for the citizens’ right to privacy.

During the development of the 2017 STIR law, Panoptykon experts [frowned at](#), among other things, the collection of IP addresses that account holders use to log into their accounts. “In what situations should this IP address be disclosed? Does it concern only entrepreneurs or regular people? And what about people who are permanently (e.g. by phone) logged into banking: will the IP address used at the time of logging in or the one from the moment of executing the transaction be transferred?” the organization asked.

These questions were not addressed but the change in regulations passed anyway. The collection of IP addresses has been effective since July 2019. No public consultation on this matter took place.



*THE ALGORITHM, WHEN  
SELECTING SUSPICIOUS  
TRANSACTIONS, USES CRITERIA  
SUCH AS WHETHER AN ENTITY  
MADE UNUSUAL, SUDDEN  
MOVEMENTS ON ITS BANK  
ACCOUNT.*



# Research

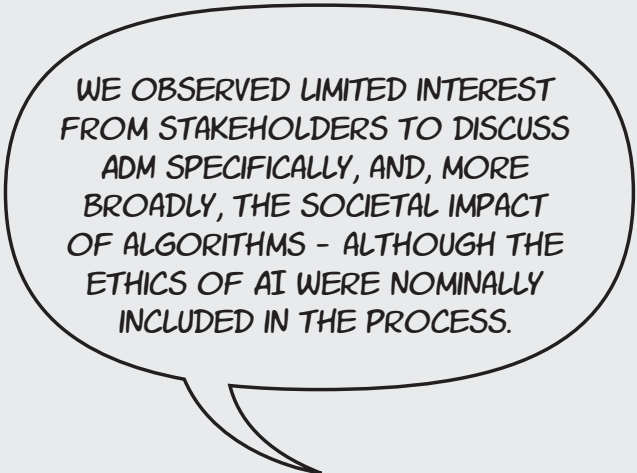
By Natalia Mileszyk & Alek Tarkowski

## Contextualization

In 2018, the Polish government initiated a public debate on artificial intelligence technologies by preparing a draft AI strategy through a participatory process that involved a broad range of experts and stakeholders. Throughout the debate, we observed limited interest from stakeholders to discuss ADM specifically, and, more broadly, the societal impact of algorithms – although the ethics of AI were nominally included in the process. The debate among stakeholders, initiated by the government, has focused on defining Poland's strategy for supplying AI solutions. Key topics concern: support for Polish entrepreneurs, provision of necessary digital talent, and the capacity of Polish research institutions. Some attention has been paid to regulatory issues, in particular those related to the availability of data for AI solutions.

At the same time, for the purposes of this report, we looked at institutions employing ADM systems but we found a very limited number of cases. One prominent example stood out: an automated system for assigning court cases to judges. This is a case that received public attention by virtue of being part of a broader, controversial (political) reform of the judicial system. Yet even in this case, the lack of transparency around the project made it difficult to establish the characteristics of the system and its real impact. There were a few other cases, but they received little attention from the media or wider society and caused limited controversy. One example – an important ADM development case study – is the system used to profile unemployed people. However, it did not garner any significant public attention.

At the end of 2018, the government unveiled its draft AI strategy, which was scheduled for finalization the following year. 2019 was to be the year in which the Polish government would define the final shape of its AI strategy. In Poland, the public administration is seen as a key actor that can help shape the national approach to AI. This is due to the high level of public funding available to core parts of the IT sector. Additionally, academic researchers – whose



WE OBSERVED LIMITED INTEREST FROM STAKEHOLDERS TO DISCUSS ADM SPECIFICALLY, AND, MORE BROADLY, THE SOCIETAL IMPACT OF ALGORITHMS – ALTHOUGH THE ETHICS OF AI WERE NOMINALLY INCLUDED IN THE PROCESS.

work is predominantly funded from public sources – are also seen as an important part of the Polish AI ecosystem.

The third key group of actors – startups and other IT companies – have not been very active in these public conversations, despite the fact that other stakeholders assign them key roles. These private companies are largely software houses that provide AI-based solutions directly to their clients, largely foreign ones. As a result, they have little interest in participating in national debates about AI and ADM.

Throughout 2019, it became clear that it would not be possible to finalize the national AI strategy quickly. A final draft of the strategy was made available for public consultation in late August. However, the whole process had yet to be finalized by the end of 2019. During 2019, no significant ADM-related initiatives were announced in the public sector and a very limited number of AI-based solutions were introduced by commercial actors. The ADM systems that directly impact Polish citizens can mostly be found in the financial sector and on various online platforms – yet these receive very limited public attention or oversight, mostly due to their global scope. Finally, 2019 has been the year in which ADM as a public policy issue was introduced into the public debate, largely due to the activity of several non-profit organizations.

The broad issue of “artificial intelligence” remains a popular topic in the media – for example, the development of “deep fake” technologies has received widespread attention in the mainstream media. However, media coverage rarely concerns specific ADM cases. Mainstream media outlets generally offer limited coverage of “tech” topics and focus on subjects that are seen by media titles as controversial or possibly interesting to the general public. For this reason, there is very limited coverage of ongoing developments that concern the development of AI or ADM solutions in Poland. In this context, it should be noted that in early 2019 the National Information Processing Institute (OPI) launched a [dedicated service covering topics related to AI technologies](#).

However, the most controversial ADM-related debate focused on “content upload filters” used by Internet platforms, which was a highly contentious issue in the European copyright reform debate. Poland was one of the countries, in which [this debate was the most heated](#). In late March 2019, the ruling party (Prawo i Sprawiedliwość) declared that protecting users’ freedom online through an appropriate implementation of the new Directive would be one of its electoral promises in the upcoming elections to the European Parliament. In practical terms, this declaration meant that Poland will strive not to implement the rules that enforce the controversial content upload filters. The decision was an attempt to build support among right-wing voters, seen by the party as being largely opposed to the new content filtering provisions. In May 2019, [Poland challenged the content filtering provisions at the European Court of Justice](#). While the term ADM was never used during this policy debate, this challenge should be seen as an important case where the societal impact of ADM systems came up for public debate.

## A catalog of ADM cases

There were no major developments regarding the use of ADM in 2019. Having said that, significant change can be seen when it comes to policy debate and judicial decisions although not when it comes to implementation. The reasons for such a state of play are multifaceted. Firstly, Polish administration, which is perceived as a key actor when it comes to the development of AI and ADM solutions, does not treat the development of ADM solutions as a priority. Up until now, the development of an e-administration infra-

structure – which is the task of the Ministry of Digital Affairs – has focused on systems that do not employ ADM mechanisms. Secondly, Poland is still a country where the supply of ADM solutions remains higher than demand. Whereas, Polish IT companies provide ADM solutions across the world, there is no need to implement them locally – whether it be in the public or the private sector. Generally, [the demand for AI services in Poland is limited](#) and this creates a natural tendency to seek out clients abroad. In Poland, 73% of companies receive at least some of their revenue from clients abroad and 33% receive most of their revenue from foreign clients.

According to the [Map of the Polish AI 2019](#) report, published by Digital Poland Foundation, the commercial ADM solutions that are most frequently supplied by Polish AI companies are analytics, big data, and business intelligence, together with sales, marketing, and advertising (one of the examples, Mr. Wolf, is presented below). In addition, companies often provide services in areas such as; financial services and insurance, the Internet of things, and industry 4.0. This is probably due to the relative competitiveness and openness to innovation in the Polish financial and industrial sectors. However, it has to be underlined that AI solutions used by companies are usually related to automation, not necessarily to taking any decisions. The more general overview of such companies is presented in the report [Iloraz Sztucznej Inteligencji 2](#) (only available in Polish). ADM solutions are employed to a limited extent by the Polish business sector, however growth in this regard can still be seen. As mentioned above, Polish companies develop AI solutions largely for foreign markets, and thus have limited influence upon the Polish ecosystem.

### / Detection of bank accounts being used for illegal activity

STIR – System Teleinformatyczny Izby Rozliczeniowej (Clearance Chamber ICT System) is a new tool that gathers monetary information from banks, Cooperative Savings, and Credit Unions. The aim of the system is to examine financial activity and to discover potential illegal activity. The system is operated by Krajowa Izba Rozliczeniowa (State Clearance Chamber – a key entity of the Polish payment system infrastructure, which renders complex clearing services and provides solutions for the banking and payment sectors). The ADM tool in this system provides suggestions to tax authorities by assigning risk indicators to accounts. For example, if there is suspicion that an offense has been committed related to a certain account, then (at the request of

the tax authorities) the bank can block the account for 72 hours (which can be extended).

Legal regulations regarding STIR have been in force since January 13, 2018, when amendments introduced to the Act of August 29, 1997 – on [Tax Ordinance – came into force](#).

The algorithm used by STIR is not publicly available, nor transparent. Moreover, the law introducing STIR states that disclosing or using algorithms or risk indicators, without being entitled to do so, is an offense. A person who discloses algorithms can be imprisoned for up to five years (if the act was unintentional, then a fine can be given).

- 1) The risk indicators, which are a result of an assessment to see if money has potentially been used for illegal activities, is determined by the algorithms developed by the Clearance Chamber. This risk indicator [takes into account the best practices of the banking sector in the field of counteracting financial and tax offenses](#), and it is a key factor when deciding whether or not to block an account. These criteria might be as follows (according to Tax Ordinance): economics (based on the assessment of overall economic activities of an entity, especially taking into consideration any transactions unjustified by the nature of the business),
- 2) geography (consisting of transactions with entities from countries in which there is a high risk of tax fraud),
- 3) subject-specific nature of certain entities (conducting high-risk business activities when it comes to the possibility of tax extortion),
- 4) behavior (any unusual behavior of the entity) and
- 5) connections (the existence of links between the entity and entities that are at risk of participating in activities related to tax fraud).

The Clearance Chamber can assign the task of maintaining or modifying the STIR system to an external contractor. Such delegation is made based on a civil law contract, but the public supervision of the process whereby an entrepreneur is selected is not sufficient. This is due to the limited application of public procurement rules (the State Clearance Chamber is a stock company). This might lead to insufficient transparency and oversight of this ADM system. In reply to a freedom of information request by the [ePaństwo Foundation](#) for more information about this particular entrepreneur, and the contract itself, the Ministry of Finance refused to disclose the information on the basis that the State Clearance Chamber is a private entity and that its operations cannot be revealed under FOI laws.

In December 2018, the Administrative Court in Warsaw issued a [precedent judgment in the STIR case](#). The court stated that in the case of extending the account lock by three months, there is no need for an evidence proceeding to be conducted. In such a case, it is enough for a public officer to analyze the flows of money and to assume that there is a risk of using a certain bank account for a tax scam.

## **/ Smile-monitoring of employees and consumers by PKO bank**

A rare example of media coverage of ADM occurred in September 2019, when the *Gazeta Wyborcza* daily [wrote about a test pilot of Quantum CX](#), an ADM system that has at its heart a machine vision algorithm which is able to distinguish the smiles on people's faces. As part of the pilot, the system was installed in several branches of the PKO BP bank. The system tracks smiles of both employees and clients. The former can exchange smiles for prizes, while on the basis of the number of smiles of the latter, the company pays funds to a charity.

The creators of the system [present](#) it as a modern, AI-based approach to employee motivation and corporate charity programs. Participation in the pilot is voluntary and em-

**A MACHINE  
VISION  
ALGORITHM  
WHICH IS ABLE  
TO DISTINGUISH  
THE SMILES  
ON PEOPLE'S  
FACES.**

employees are able to independently regulate their own privacy settings in a specially designed application. Using the highest-level privacy settings means that the manager cannot check the results of a particular employee, but only the sum of smiles in a given bank branch.

After the initial article, the system was criticized by lawyers working on digital rights and labor law issues. The issue was addressed by the Polish Ombudsperson (Rzecznik Praw Obywatelskich), who [questioned the legality of the system on the basis of both labor law and GDPR compliance](#). The Ombudsperson has sent a formal request to the Chief Labor Inspector to investigate the case.

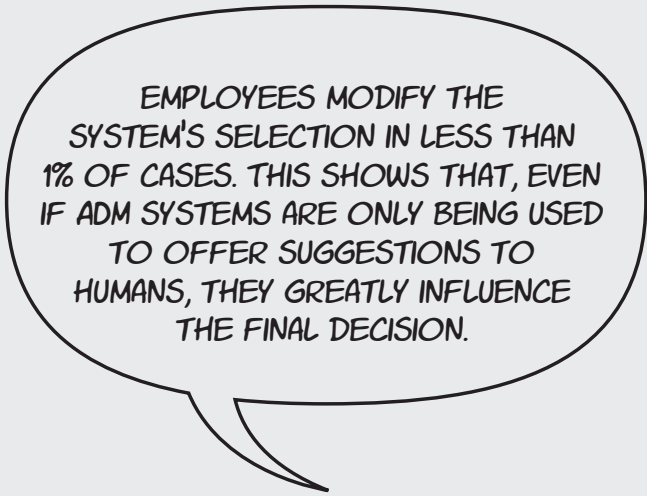
At the same time, the system has received a positive response from potential clients. Another pilot has been initiated by the office of the city of Sopot. In addition, several companies have expressed an interest in the system as well.

The case of Quantum CX clearly demonstrates the varied responses to ADM systems, and we will have to wait and see how the case will further develop in 2020 and whether any further evaluation of this system will be conducted.

## The ADM system for profiling the unemployed is scrapped

As presented in last year's report, in May 2014, the Ministry of Labor and Social Policy introduced a [simple ADM system that profiles unemployed people](#), and assigns them three categories that determine the type of assistance they can obtain from local labor office. Panoptykon Foundation, and other NGOs critical of the system, have been arguing that the questionnaire used to evaluate the situation of unemployed people, and the system that makes decisions based on it, is discriminatory, lacks transparency, and infringes data protection rights. Once the system makes a decision based on the data, the labor office employee can change the profile selection before approving the decision and ending the process. Yet according to official data, employees modify the system's selection in less than 1% of cases. This shows that, even if ADM systems are only being used to offer suggestions to humans, they greatly influence the final decision.

The issue was finally solved in 2018, when [the Constitutional Tribunal decided](#) that the system needs to be better secured in a legislative act – although its main objection was rather formal: the scope of data used by the profiling



EMPLOYEES MODIFY THE SYSTEM'S SELECTION IN LESS THAN 1% OF CASES. THIS SHOWS THAT, EVEN IF ADM SYSTEMS ARE ONLY BEING USED TO OFFER SUGGESTIONS TO HUMANS, THEY GREATLY INFLUENCE THE FINAL DECISION.

tool should have been set out in the legal act adopted by parliament and not decided by the government. As a result, the government decided to [end its experiment with profiling the unemployed](#), and the system was finally scrapped in December 2019.

The case of this ADM system shows that assessing the impact of even a simple ADM system is a resource and time-intensive effort. The case also proves that – while the public administration lacks the necessary procedures, institutions, and skilled staff – the role of non-profit watchdogs who monitor ADM development is crucial.

## / Mr. Wolf

Mr. Wolf is a commercial tool that helps to automate text communication in customer service by automatic classification of notifications, interpretation of their content, and by providing automated answers to repetitive questions. The system makes automated decisions related to consumer service communication at various levels, starting from classification to content decisions. After being sent to the system, notifications are categorized by algorithms. Mr. Wolf's task is to identify what kind of information is contained in the notification: is it a bug report, parcel order or a password reset? It then assigns the case to the appropriate category. In the event that the question cannot be categorized, it redirects the task to a human and then the process of answering begins. After receiving the information, the software provides answers or transfers the matter to the employee. What is important is that the cases that are resolved by employees teach the system how to solve more difficult cases – which make ADM systems more effective and failure-proof. Mr. Wolf is an example of an ADM tool that makes repetitive actions less burdensome and time-consuming. At the same time, from the way the software

has been programmed, it accepts the fact that not all cases can be solved automatically and that there is a need for humans in the process.

# Policy, oversight and debate

## / Policy

### / Artificial Intelligence Development Policy for 2019-2027 and other governmental actions

The Polish government [has been working on the issue of Artificial Intelligence](#) since June 2018, when Jarosław Gowin, Deputy Prime Minister and Minister of Science and Higher Education, declared that [Poland will create its own Artificial Intelligence strategy](#). As of July 2020, the final version of the national AI strategy still hasn't been published.

The work on a national AI strategy, and programs emerging from this strategy, have been relatively slow. Efforts related to AI development were initiated by two different Ministries: The Ministry of Entrepreneurship and Technology, and the Ministry of Digital Affairs. These two agencies have coordinated their efforts to a small extent. The former is focused on industrial policy related to the "Industry 4.0" model, while the latter is developing educational programs aimed at increasing the supply of skilled digital talent in the IT sector.

On February 26th, 2019, the Minister of Digitization, Minister of Entrepreneurship and Technology, Minister of Science and Higher Education, and the Minister of Investment and Development all signed [a memorandum on the development of Artificial Intelligence in Poland](#) which aimed to provide a strategic and cross-sectoral framework for AI development. The Ministries agreed to create an AI ecosystem, to coordinate activities aimed at providing interdisciplinary education and research on AI, to provide citizens with skills related to data science, and to cooperate in order to create a roadmap to get there.

In late August 2019, the Ministry of Digital Affairs published the much awaited final draft of the [Artificial Intelligence Development Policy for 2019-2027](#) for [public consultation](#). At the time of writing this report neither the results of the con-

sultation nor the final version of the policy document have been made public.

The key goal of this document is to provide Poland with an important place in the global economy, which is already being shaped – and will be further shaped in the near future – by AI technologies and applications. In the opinion of its authors, the use of AI by Polish companies is necessary to build a strong internal market for AI solutions. As written in the document, the mission of Poland's strategic policy in the field of AI is to support AI science, research and development for the growth of innovation, and productivity of the knowledge-based economy, as well as to support citizens in the process of transformation in the work environment, and the improvement of competences while taking into account the protection of human dignity, and ensuring conditions for fair competition. So far, AI development in Poland has led to the creation of a supply strategy in which Poland is a country providing AI-based solutions that are then used in other parts of the world. Creating a demand strategy is necessary if AI technology is to benefit Polish society. Such ambition is visible in the draft policy.

Even if the term ADM is not mentioned in the draft of the document, the policy refers in many places to decisions taken by AI, which constitute part of general ADM mechanisms. The authors of the policy acknowledge that using AI in ADM is inevitable, but that these systems shall be implemented with respect to certain standards, such as transparency (including how algorithms work). Moreover, ADM development shall be in accordance with international standards, legal framework, and ethical considerations. The documents also put legal challenges related to ADM in the spotlight, such as the automatic conclusion of civil contracts. There are also certain proposals to help regulate ADM, such as a system of norms, certification or data management standards – although a single perfect solution has not been identified.

The authors of the document summarize the framework by concluding that, in order to strengthen AI development in Poland, coordination is necessary in the following areas: construction of a 5G network; development of digital competences at all levels of education; support for the possibility of creating inter-university networks to create consortia dealing with specific research problems; promote the latest technological solutions among entrepreneurs to raise the level of development and establish a dialogue with other EU economies; support and promote joint entrepreneurial solutions (e.g., joint R&D, exchange of experiences, cluster



solutions); support cooperation between academic centers and business entities, support projects in the field of cyber security; support the creation of an API standard for access to industrial data and open interoperability standards.

## / Oversight

### / The right to explain credit decisions

Up until recently, in Poland and other European Union countries, only entrepreneurs had the right to ask a bank to explain a decision to decline credit. Poland is the first EU country to extend this right to individual customers and to clarify what information banks must provide to their customers – including cases where ADM was used to calculate a credit decision. At the end of April 2019, [an Act of 21 February 2019 amending other acts](#) (including banking law and the Consumer Credit Act) in connection with ensuring the application of the General Data Protection Regulation came into force.

According to the new law, everyone has the right to clarify their credit assessment if:

- they are applying for any kind of loan. It does not matter if the decision was negative or positive or whether the application is for a mortgage or just for the ability to shop in installments.
- they are applying for a consumer loan at a loan institution (companies such as Provident, Vivus, and Stork operate in the Polish market), and the final decision was negative.

If a credit or loan decision was made automatically (without human intervention), citizens have the additional right to have the basis for that decision explained to them. This permission is independent of whether the automatic decision was positive or negative. In this case, the customer also has the right to express their own position. If someone does not agree with the ADM decision, their application can be reviewed, and the resulting decision will be made by a human (e.g., a bank analyst). This procedure applies both at banks and at lending institutions.

The clarification can be requested as soon as the bank, or loan institution, issues a credit decision. The additional right to explain the grounds for an automatic decision can also be exercised as soon as the result of the decision is known. In practice, this occurs immediately after submitting the ap-

UP UNTIL  
RECENTLY, IN  
POLAND AND OTHER  
EUROPEAN UNION  
COUNTRIES, ONLY  
ENTREPRENEURS  
HAD THE RIGHT  
TO ASK A BANK TO  
EXPLAIN A DECISION  
TO DECLINE CREDIT.  
POLAND IS THE  
FIRST EU COUNTRY  
TO EXTEND THIS  
RIGHT TO INDIVIDUAL  
CUSTOMERS.

plication – automatic decisions are made as soon as all the required data has been entered into a computer.

The Polish non-governmental organization, [Panoptikon Foundation](#) (which focuses on privacy) played a key role in getting this right extended to customers.

## / Civil society and academia

While the general public and policy debate in Poland focuses on AI and its commercial potential, civil society actors are trying to shift the narrative more towards ADM-related issues. According to NGOs (Centrum Cyfrowe Foundation, Panoptikon Foundation, Klub Jagielloński, and others), discussion about ADM systems – as opposed to discussion about AI – allows for a better understanding of the impact of new technologies on the individual and society. What is worth underlining is the fact that this shift is not purely dominated by NGOs, but also allows the space for dialogue with various other actors, especially administration and business.

### / AlgoPoland. Automated Decision Making in service of the society

The [AlgoPoland Automated Decision Making in service of the society](#) report is the result of cooperation between the [Centrum Cyfrowe Foundation](#) (Polish digital think-and-do tank, of which the authors of this chapter are members) and

Klub Jagielloński. The key assumption of the report's authors is that Poland has so far implemented fewer ADM systems than other European states and developed economies. To encourage sustained development of ADM systems, Poland can take advantage of the benefits of its delayed economic growth. While implementing these systems, Poland is able to learn from the mistakes made by other countries which have pioneered the way, but have often made faulty attempts at implementation. A rational strategy for the implementation of ADM systems – one based on clear principles aimed at serving the public interest – can be of great benefit to the Polish economy, society, and public administration.

The AlgoPoland report consists of three parts: presentation of various applications of ADM systems around the world,

and specifically in Europe and Poland, in the following key areas of social life: predictive policing, job market, social welfare, and healthcare, education, access to information and culture, finance, the judiciary system, and finally, the challenges and opportunities related to ADM and recommendations.

The recommendations from the report are as follows:

- To build socio-economic trust regarding ADM systems, it is vital to proceed within the quadruple helix model which combines academia, industry, public administration, and civil society.
- ADM systems must be regulated. Due to their considerable impact on individuals and societies, we

***WHILE THE GENERAL PUBLIC  
AND POLICY DEBATE IN  
POLAND FOCUSES ON AI AND  
ITS COMMERCIAL POTENTIAL,  
CIVIL SOCIETY ACTORS  
ARE TRYING TO SHIFT THE  
NARRATIVE MORE TOWARDS  
ADM-RELATED ISSUES.***

need to cooperate to develop regulatory measures which will ensure that ADM systems serve the joint interests of the economy and society.

- ADM systems are only as good as the data which they process. High quality and, most importantly, the availability of datasets are necessary for avoiding biased and corrupt results of using ADM systems.
- ADM systems must be based on a transparent and explainable mechanism.
- The key to sustainable development of ADM is public awareness: understanding of the mechanism and results of the application of such systems.

## / ALGOVrithms. State of play report

In May 2019, another digital NGO – ePaństwo Foundation, in cooperation with partners from the Czech Republic, Georgia, Slovakia, Hungary, and Serbia, published a report called [ALGOVrithms. State of play report](#). The report identifies examples of algorithms used in the public administration systems of the previously mentioned countries. The report describes how they work, verifies whether they were introduced through legal regulations, checks how open and transparent they are, and sees if there is any system in place that secures citizens' rights against potential abuse as a result of algorithms. While working on the report, researchers mainly used three methods to collect the required information: desk research, freedom of information requests, and interviews with recognized experts and decision-makers.

The authors did not outline the existing overall state policy regarding the implementation of alGOVrithms in any of the countries participating in the research. Moreover, they did not find any examples of the existence of a legal framework comprehensively describing the rights and obligations of the states and citizens regarding ADM. The legal documents that do exist refer to some aspects of examples of alGOVrithms, such as systems to allocate judges to specific court cases. The authors proved that algorithms used in ADM software are not subject to transparency and that access to the algorithms or the source code is not possible. In addition, the authors did not find any cases where a single institution oversees, or even possesses comprehensive knowledge of, which ADM systems exist in each country. Apart from the example in the Serbian system, where judges are allocated to cases, and where the donor (EU) has audited the system, no external and independent audits are in place to monitor the accuracy and fairness of the operation of algorithms. The report shows that there is a long way to go in the region

before a sustainable and human rights-proof policy framework for ADM development is created.

## / Digital Sustainability Forum

The [Digital Sustainability Forum](#) (DSF) is a joint initiative between the Centrum Cyfrowe Foundation, Digital Poland Foundation, and Microsoft. The goal of the initiative is to develop recommendations for decision-makers who are responsible for building strategies, policies, and regulating emerging technology. These recommendations are the result of cooperation between representatives of science, business, non-governmental organizations, and public administration. The cooperation of all participants of this innovative ecosystem has resulted in the development of pragmatic recommendations which take into account the reality of a constantly changing world. It is also a tangible example of a project using a quadruple helix model which combines academia, industry, public administration, and civil society.

Up until now, the Digital Sustainability Forum was the place to discuss two issues: AI and cybersecurity. The AI discussion addressed the questions of how to build digitally sustainable policies in the AI sector, how to ensure that the opportunities provided by the development of AI technology overcome social risks and challenges, and how to use Poland's digital transformation to achieve its sustainable development goals. The issue of safeguards that need to be implemented when it comes to ADM systems that use AI was also discussed. The scope of the forum shows that the issue of ADM (and other systems) need to be addressed from various perspectives. Such an approach creates room to help answer the challenges of ADM from both a societal and a business perspective – the inclusion of diverse voices will be crucial in this process.

## Key takeaways

In 2019, the public debate on AI, which was initiated when the government began work on the national AI strategy, continued with a significant number of events and reports covering the issue. AI is now a prominent topic of discussion among stakeholders interested in the development and regulation of digital technologies in Poland.

Within this broad context, the specific concept of ADM has mostly been promoted by several non-profit organizations, but this has received limited recognition from the government. The overall debate remains largely focused on AI and its economic potential, whereas the public administration has been paying a degree of attention to issues such as AI ethics and data governance.

In 2019, we did not observe any significant new deployments of ADM systems in Poland. In particular, it was noted that the public sector did not deploy any such solutions throughout the year. Several ADM systems, which we have identified in this report, have very limited scope and have not sparked any significant debate on ADM systems in Poland.

Poland has a strong IT sector and many companies are working on services based on AI solutions. These companies are commonly seen as the strongest asset of the Polish AI ecosystem, but they usually develop bespoke systems for foreign and not domestic clients, and for this reason they have a very limited impact on the Polish AI ecosystem and surrounding public debate.

## References:

Borowiecki Łukasz and Mieczkowski Piotr (2019): Map of the Polish AI, [online] <https://www.digitalpoland.org/assets/publications/mapa-polskiego-ai/map-of-the-polish-ai-2019-edition-i-report.pdf> [20.12.2019]

Calini Leonardo, Marczuk Piotr, Mieczkowski Piotr, Paszcza Bartosz (2019): Iloraz Sztucznej Inteligencji, [online] <https://www.digitalpoland.org/assets/publications/iloraz-sztucznej-inteligencji/iloraz-sztucznej-inteligencji-edycja-2-2019.pdf> [19.12.2019]

Ilić Vujo, Izdebski Krzysztof, Kerekeš Daniel, Merényi Miklos, Škop Michal, Turashvili Teona (2019): alGOVrithms state of play, [online] <https://epf.org.pl/wp-content/uploads/sites/2/2019/05/alGOVrithms-State-of-Play-Report.pdf> [17.12.2019]

Mileszyk Natalia, Paszcza Bartosz and Tarkowski Alek (2019): AlgoPolska. Zautomatyzowane podejmowanie decyzji w służbie społeczeństwu, [online] <https://centrumcyfrowe.pl/wp-content/uploads/sites/16/2019/12/AlgoPolska-raport.pdf> [20.12.2019]

Niklas Jędrzej, Sztandar-Sztanderska Karolina, Szymielewicz Katarzyna (2015): Profiling the unemployed in Poland: social and political implications of algorithmic decision making, [online] [https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon\\_profiling\\_report\\_final.pdf](https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf) [17.12.2019]

Niklas Jędrzej (2019): Poland: Government to scrap controversial unemployment scoring system, [online] <https://algorithmwatch.org/en/story/poland-government-to-scrap-controversial-unemployment-scoring-system/> [18.12.2019]



BELGIUM  
DENMARK  
ESTONIA  
FINLAND  
FRANCE  
GERMANY  
GREECE  
ITALY  
NETHERLANDS  
POLAND

**PORTUGAL**  
**STORY**

PAGE 194

**RESEARCH**

PAGE 197

SLOVENIA  
SPAIN  
SWEDEN  
SWITZERLAND  
UNITED KINGDOM





Find out more on p. 198 in the research chapter under "Casinos".

# **Automated verification of prescriptions helped crackdown on medical fraud**

**Portugal's national health service introduced a centralized, automated system to verify medical prescriptions in 2016. One year later, it flagged 20 million euros of fraud.**

By Paula Simoes

Which doctors give more medical prescriptions? Which ones prescribe more expensive medicines? How many doctors have more than 100 packages dispensed from the same pharmacy, on the same day? How many packages are dispensed from each pharmacy when the expiry date is about to end?

These are some of the questions answered by the Control and Monitoring Center from the Portuguese Public National Health Service, *Centro de Controlo e Monitorização do Sistema Nacional de Saúde* (CCM-SNS), using data from several databases, to detect and fight prescription fraud. The “paperless prescriptions” program and the electronic medical prescription system started in 2016 and were made mandatory for the public sector, although exceptions were allowed in some cases.

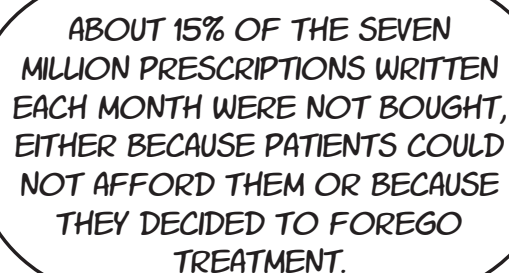
Doctors fill in the prescription, using software certified by the Shared Services of the Ministry of Health, *Serviços Partilhados do Ministério da Saúde* (SPMS), sign it digitally, either with their Portuguese Citizen Card or with their Order of Doctors’ card, and send it to the patient’s mobile phone by text message or email, or print it out. Doctors can also use a mobile app to issue prescriptions, even remotely, by signing them with their Digital Mobile Key. This is a means of authentication created by the Portuguese government that connects a mobile phone number to a Portuguese citizen identification number. Patients can also use their Citizen Card to buy prescribed medicines from a pharmacy.

When the system was announced, back in 2015, the national newspaper, [Público, reported](#) that about 15% of the seven million prescriptions written each month were not bought, either because patients could not afford them or because they decided to forego treatment. The new system lets doctors know if the prescriptions were bought or not. Another novelty is that patients can buy the prescribed medicines from different pharmacies, in case one pharmacy does not stock all of them.

## / Controlling fraud

One of the major motivations, as well as optimizing resources and reducing costs, for using the electronic prescription system was to control the use and detect fraud.

Fraud was reduced by 80% during the first year of the new system, according to the National Health Service (SNS). In 2019, 97.28% of the prescriptions in the public health sector were electronic (the share of electronic prescriptions is



ABOUT 15% OF THE SEVEN MILLION PRESCRIPTIONS WRITTEN EACH MONTH WERE NOT BOUGHT, EITHER BECAUSE PATIENTS COULD NOT AFFORD THEM OR BECAUSE THEY DECIDED TO FOREGO TREATMENT.

much lower in the private sector, where the practice is not mandatory). The system allows the SNS control center to quickly monitor and detect irregularities by automatically analyzing patterns in the prescription and dispensation of medicines.

The Control and Monitoring Center uses data from invoices issued by pharmacies and other providers of services, such as medical exams. This data set is complemented by information from electronic or paper prescriptions from the databases of the Ministry of Health, such as the National Prescription Database (*Base de Dados Nacional de Prescrições*, BDNP); the Vignettes and Prescription Requisition Portal; and the National Medicine Database. The control is made at the prescription phase, where the system verifies that the doctor is registered and can make the prescription and that the patient exists and can benefit from it.

## / Phony prescriptions

In 2011, RTP, the public sector broadcaster, [reported](#) that the police were investigating a case of fraud that happened in 2009 and 2010, before the new system was in place. Prescriptions had been written in the name of deceased doctors, using falsified signatures, or to deceased patients. Several doctors were using the same professional license number and one doctor made 32,000 prescriptions in only one year – one every 3 minutes. The targets were expensive medicines with a high percentage of reimbursement by the SNS. With a drug costing 100 euros and a reimbursement rate of 90%, a pharmacy can get 90 euros from the SNS by presenting a prescription written by a doctor, without having sold the medicine. The news report mentioned that medicines represented 40% of all public expenditure fraud.



Another type of fraud can happen when patients from the public sector are referred to the private sector. In January 2014, SIC, a private TV channel, [reported](#) that the private health group Sanfil invoiced ADSE, a public health sub-system, for medicines and medical exams that had not been provided to the referred patients.

RIGOR,  
TRANSPARENCY AND  
PERMANENT MONITORING  
ARE ESSENTIAL  
INSTRUMENTS IN THE  
FIGHT AGAINST FRAUD AND  
CORRUPTION.

The Invoice Control Center (CCM) started operations in 2010. Two years later, the Information Exploration Unit was created to analyze data gathered by the CCM and to detect anomalous and potentially fraudulent situations, to develop a fraud risk analysis model, and to forward suspicious cases to the police.

The fraud control system also uses the Ministry of Health's National Registry of Health Users, which contains information about the benefits patients are entitled to (the reimbursement rate varies depending on employment status, with different schemes for pensioners and public-sector employees). This allows for the immediate verification of each transaction to the different health providers. The National Commission for Data Protection authorized the practice.

## / Back in public hands

The computerization of the system of fraud detection for complementary means of diagnosis and therapy dates back to the 1980s. The system was extended to medicine prescriptions in 2003, under the aegis of the Central Administration of the SNS, with data coming from regional health administrations.

In 2007, the Council of Ministers earmarked over 30 million euros over four years to create and implement the Invoice Control Center (*Centro de Conferência de Faturas*, CCF), now called Control and Monitoring Center. The project was carried out by a private company, PT Comunicações, from the design to the implementation and operation phases.

Until then, the system to control the invoices involved around 500 workers spread across a variety of offices around the country, leading to high costs for the National Health Service and difficulties in obtaining results in a timely manner. In 2005, there were 23 million prescriptions for complementary means of diagnosis and therapy and 55 million prescriptions for medicine.

At the same time, the Ministry of Health tried to increase the use of electronic prescriptions – from an estimated 40% at the time. The promise of timely and easy fraud detection was a big motivation for the push towards digitization. However, it could only be done if the entire technological ecosystem was updated to deal with it.

Over the first seven years, the central unit detected irregularities amounting to hundreds of millions of euros and was instrumental in starting many prosecutions for fraud in the health system. According to the Ministry of Health, in the first semester of 2017, the Invoice Control Center processed prescriptions worth 356.2 million euros and flagged 21.1 million euros to the State prosecutor.

Although the Invoice Control Center was always under the Central Administration of the SNS, the management, maintenance, and operation of the center were provided by a private operator until June 2018. At that time, the Council of Ministers transferred responsibilities for this unit to the Shared Services of the Ministry of Health (SPMS), putting it entirely under the control of the public sector.

The government justified the transition by considering the activity of the center as strategic to the efficiency of the SNS and to the management of public expenditure. Data protection considerations also played a role.

In December 2019, the adjunct Secretary of State for Health, Jamila Madeira, made an official visit to the premises of the Invoice Control Center. She said the center was a priority for the government because “rigor, transparency and permanent monitoring are essential instruments in the fight against fraud and corruption”. Each month, the center checks 8 million documents, 70% of them digital.

# Research

By Eduardo Santos

## Contextualization

ADM in Portugal is still at an early stage. A lot of work has already been done, and there is a lot of knowledge in the field. However, in general, citizens do not yet feel the presence of ADM in their daily lives. This is either; because it is not present at all or, when it is present, it is not obvious and/or it plays a limited role.

In recent years, new companies have emerged that mostly aim to work with and sell to international markets. These companies use decent AI tools provided by local universities and build upon scientific knowledge from the field and the human resources sector. Some of these companies, although they are not well known in Portugal, have achieved significant commercial success abroad. This has led to increased attention from the national media who are interested in the activities of these companies and the potential of AI for the country.

Unlike other countries, where the use of ADM has led to public criticism and debate, in Portugal, the absence of ADM processes – or the lack of awareness of them – has so far meant that there has been little public debate on the issue. The media often writes about AI from a critical angle, pointing out the dangers and challenges. However, the coverage mostly focuses on events in other countries which can seem distant to the reality of life for Portuguese people.

This might change soon, as many projects are starting to develop, or are already being implemented, in the country. At the same time, the government has taken the lead by presenting the first National Strategy for Artificial Intelligence and by promoting programs to support AI initiatives in Public Administration. The government is also encouraging Portuguese companies to develop in the field of AI and ADM and to take advantage of opportunities abroad. The prime minister, António Costa, recently traveled to India where he visited a Portuguese company working on a pilot project for the installation of biometric face identification technology at New Delhi International Airport. In addition,

Costa has encouraged Portuguese companies to take advantage of the economic opportunities offered by India, especially for engineering and tech companies in this sector, as India is currently building one hundred new airports.

## A catalog of ADM cases

### / Banking

Banks and financial service companies are promoting services that they claim give their clients an immediate decision on their personal loan requests. For example, BPI's "immediate credit" solution states that clients can simulate the desired loan conditions and, upon submitting their data, "know immediately if the funding is approved". However, the small print indicates that there is a human decision before the credit is granted.

Younited Credit Portugal also promises an immediate first answer with their online system.

Puzzle, from Banco BNI, advertises a one-minute response to online loan requests. The small print states that it will take 48 hours if the client uses a manual process instead, although it is not clear what that means.

In early 2019, Jornal de Negócios reported that Caixa Geral de Depósitos, the public bank founded in 1876 and one of the biggest banks in the country, expected "automated credit decision (services) to be coming soon"

### / Police Forces

In January 2020, for the first time, the Data Protection Authority (DPA) has given an opinion on the use of AI in video surveillance systems. It refused to give its approval in two projects that would implement such systems in public spaces in the cities of Leiria and Portimão, as requested



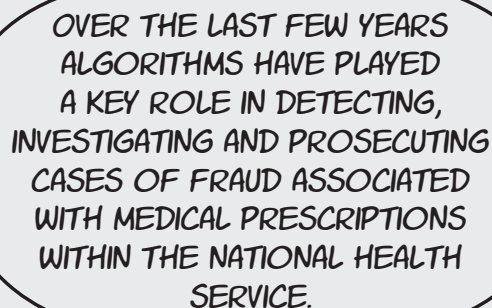
by the police. This system would allow the use of technology “to search for people by physical characteristics, gender, clothing or hair colours”. However, the DPA pointed out the lack of justification for the use of this technology. This conclusion refers not only according to proportionality criteria and taking into account the amount and type of information that can be collected, but also to the opacity of the process of pattern analysis. The DPA strongly criticized the projects, as they involved “large-scale systematic monitoring and tracking of people and their habits and behavior, as well as identifying people from data relating to physical characteristics, being undeniable the high risk to the fundamental rights to data protection and respect for privacy”. The DPA also considered that there was a risk of “not being able to understand whether the results presented by the system, which would serve as (the) basis for the police to make decisions about the targeted citizens, are discriminatory”, pointing out that this would be “inadmissible” under the Portuguese Constitution. (Sources: TSF, [Público](#)).

In January 2018, [Diário de Notícias reported](#) on the use of face recognition by the Laboratory of the Forensic Science Police to monitor foreign citizens in airports. However, it remains unclear the exact context in which such a system is used. The involvement of a forensic laboratory from the police might indicate that such technology is not used as a large-scale standard solution for the control of foreign citizens, but rather as an alternative resource, available to the police, for investigation purposes.

## / Casinos

Under the Portuguese gambling law ([DL n.º 422/89, de 02 de Dezembro](#)), specific people can be banned from casinos. This includes the possibility of [self-exclusion](#), to help people who are struggling with gambling addiction. Casinos cannot allow persons on the non-authorized list to enter their premises. Traditionally, this law is enforced by the security staff of the casinos using CCTV and personal contact for recognition purposes. In this context, the rise of face recognition was seen as an opportunity to improve the efficiency of law enforcement.

In 2016, Turismo de Portugal (the national tourism board) [contracted IBM](#) for 110 licenses and the central software needed to activate and run the face recognition modules of the CCTV systems installed in casinos. The contract cost 337,833.63 euros (plus VAT) and was comprised of 401 CCTV cameras which were set up in all 11 casinos in Portu-



OVER THE LAST FEW YEARS  
ALGORITHMS HAVE PLAYED  
A KEY ROLE IN DETECTING,  
INVESTIGATING AND PROSECUTING  
CASES OF FRAUD ASSOCIATED  
WITH MEDICAL PRESCRIPTIONS  
WITHIN THE NATIONAL HEALTH  
SERVICE.

gal. According to the publicly available contract, the implementation period should have taken 90 days.

In August 2018, the [Jornal de Notícias newspaper reported](#) (paywall) that casinos were installing face recognition systems intended to stop, for example, self-excluded gambling addicts from entering the premises.

However, in May 2019, the [Público newspaper reported](#) (paywall) that the system was still not up and running. The article cites a representative from a casino who confirmed that the system did not produce the expected results, “for technical reasons”, but he thought that such a goal may still be achievable. A representative from another casino confirmed the claim, saying that the system “does not work very well, (but it) is in the implementation phase”. [Público](#) further reported that the system was installed back in 2017, but it has not been operational since then. The technical difficulties are related to the poor lighting at the entrances to the casinos, and to the fact that the cameras need to zoom in on the customer’s face. The failure of the system meant that the casinos had to hire more staff to do this work. Turismo de Portugal did not reply to questions concerning the renewal of licenses and their plans to continue the contract with IBM. [Público](#) also reported, back in 2017, that it was said that the CCTV cameras would only work with IBM software. This would explain why the contract was signed directly with IBM instead of going through a normal public tender process.

## / Prescription fraud detection in the National Health Service

Over the last few years algorithms have played a key role in detecting, investigating and prosecuting cases of fraud associated with medical prescriptions within the national

health service (Serviço Nacional de Saúde). [Sábado reported](#) how a team of three or four people processed between six to seven million prescriptions a month, written by 40,000 medical doctors and picked up at the 3000 pharmacies throughout Portugal. The team uses algorithms to detect patterns and to flag cases of higher fraud probability. For example, the system checks to see if there is a large concentration of prescriptions from a single doctor filled in at the same pharmacy, if a doctor prescribes above-average quantities, which doctors give out the most prescriptions, or who gives the most value, etc.

This information is then forwarded to the criminal investigation police. Between 2012 and 2016, the reports forwarded for police investigation involved 432 doctors, and more than 122 service providers and prescriptions worth a total of 1,098 million euros. The police then investigated which of these cases might have illicit origins, which resulted in several police operations and subsequent criminal prosecutions with significant impact in the press and with the public.

## **/ Pensioners prove they are alive via face or voice recognition**

Every year, approximately 11,000 retired Portuguese citizens who live abroad must demonstrate “proof of life” in order to continue receiving their pension. Now, a pilot project launched by Caixa Geral de Aposentações, a welfare institution for civil servants, aims to simplify the process by allowing pensioners who live abroad to demonstrate the required proof of life via face or voice recognition. The pilot project started in Macau (a former colony that remained under Portuguese administration until 1999), where hundreds of people are already using the system.

A representative of a pensioner’s association in Macau [told Antena 1](#) that, while he found the system easy to

use, he also heard criticism from other pensioners who had trouble with it. The problems were mainly related to the fact that these are senior citizens who are not familiar with computers and, in some cases, even mobile phones.

## **/ Uber announces real time face recognition to identify drivers**

In December 2019, Uber announced that it had started using its new driver identity verification and validation system in Portugal. The company said that “the new feature will require drivers and delivery partners to take a picture in ‘selfie’ mode before going online. The picture is then immediately cross-checked with the profile photo of the driver or delivery partner and so we make sure you are the right person behind the wheel or making the delivery”, [Expresso reported](#). The Real Time Identity Verification system was already being tested in the United Kingdom, after the passenger transport platform faced problems in London, with the local taxi regulator refusing to extend its license due to safety concerns.

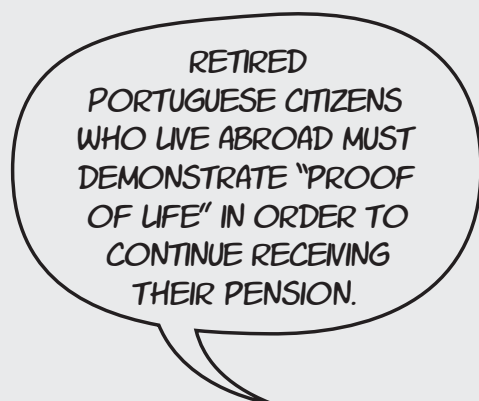
## **/ Airports**

### **/ Portuguese Immigration and Border Service**

In 2019, the Portuguese Immigration and Border Service spent 4.2M euros + VAT to update Project RAPID (Automatic Identification of Passengers Holding Traveling Documents), a system that started in 2007 and that automatically controls passengers who cross Schengen’s external borders. “This system combines the operations of reading and checking electronic passports with an innovating feature for assessing biometric data which operates an automatic gate opening device. This device checks on a first phase the genuineness of electronic passports and validates all data stored in the chip and, on a second phase, appraises the passenger’s identification by establishing a comparison between the photo stored in the chip and the information of the passenger in loco, automatically opening the border gate when the features of both images are coincident”. The recently purchased devices will be installed in the airports of Lisbon, Porto, Faro and Madeira (source).

### **/ BOM – Biometrics on the Move**

During October 2019, the Humberto Delgado International Airport in Lisbon tested a voluntary trial called Biometrics



on the Move. This is a Frontex (the European Border and Coast Guard Agency) project together with the Portuguese Immigration and Border Service and the Lisbon Airport Authority (ANA). It uses face recognition and touchless scanning of fingerprints enabling passengers to cross the border “almost seamlessly”, “without even taking out their passport or other documents”, according to Frontex.

## **/ “Seamless flow” system in Lisbon’s Airport**

Vinci Airports, a company that develops and operates airport platforms, [chose](#) the Humberto Delgado International Airport in Lisbon to pilot a center of excellence on flow management, within the network of airports the company operates.

The center of excellence in Lisbon is currently developing innovative projects related to a barrier-free system, which will allow a seamless flow of passengers from the entrance of the terminal until boarding the flight, without having to stop or show documentation. In order to use this system, the passenger will have to submit passport data and allow a photograph to be taken. Then, biometric reading devices (fingerprint or face recognition) scattered throughout all the necessary stages will read this data as the passenger is moving, and allow him or her to pass without stopping.

The company intends to start implementing the new system in Terminal 2, which is used mostly by low-cost airlines, but there’s no date set for the implementation.

## **/ Justice: Algorithm for assigning judges to a case under suspicion**

“Operação Marquês” is the nickname of a highly controversial judicial process where the ex-prime minister of Portugal, José Sócrates, was detained (preventive detention) while facing charges of corruption, qualified tax fraud, document forgery, and money laundering.

Following an indictment at the investigation phase, the defense requested the opening of an optional phase (“Instrução”). This phase takes place before the trial phase and allows the investigating judge to assess the consistency of the evidence gathered by the public prosecutor. The judge can then determine whether the evidence is sufficiently robust to lead to the possible conviction of the defendant. For this specific case, this optional phase would fall under the jurisdiction of a specialized court that has only two judges,

one being the judge that oversaw the investigation phase and made decisions that the ex-PM and his lawyers heavily contested in public. They also argued why the same judge who oversaw the investigation should end up being the one to lead the next phase of the process.

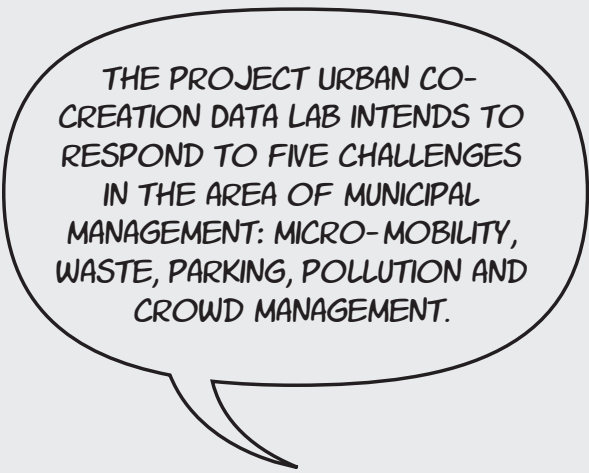
The selection of the judge in charge of the next phase became a contentious topic. Public opinion on which judge should be selected was divided, and often influenced by personal views on the ex-PM. Some argued that the same judge who was thought to be tougher should continue. Whereas, others preferred to have the alternative judge, who, in their view, was fairer. In the end, it was left up to “the algorithm” to randomly “choose” which judge it would be.

A TV cameraman and a photojournalist [were brought into the room](#) as the random algorithm decided who would be the judge to take the case. This led to public criticism from the first judge, effectively putting the algorithm under suspicion. According to him, the algorithm was not purely random as it took into account several factors, such as the number of opening cases each judge had been involved in, and, as such, it could be manipulated, to a degree. This led to the opening of a disciplinary process for the judge, later filed. In a press release, the independent judicial entity responsible for the management and disciplinary action of judges [said](#) that “the electronic distribution of cases is always random, not balancing on a daily basis or in any other time period, that may be known in advance, the cases distributed to each judge”. However, a fact-checking agency [has confirmed](#) from several sources that the system is not entirely neutral and that it is influenced by external factors, including the number of cases assigned to each judge.

None of the criticism has ever focused on the fact that [true randomness is actually very difficult to generate](#) in a computer system or the fact that the algorithm is not public. In fact, very little is known about the algorithm, as the previous contradictions show. For example, it is not known if it was ever audited. The judicial process is still ongoing, and it is still not known if the ex-prime minister, José Sócrates, will face a trial.

## **/ Human resources**

In 2017, Expresso [reported](#) how human resource companies in Portugal were taking the first steps in automation. At the time, Manpower confirmed that they were not us-



THE PROJECT URBAN CO-CREATION DATA LAB INTENDS TO RESPOND TO FIVE CHALLENGES IN THE AREA OF MUNICIPAL MANAGEMENT: MICRO-MOBILITY, WASTE, PARKING, POLLUTION AND CROWD MANAGEMENT.

ing it in the country yet, but that they were preparing for such a possibility. Randstad confirmed that they were using assessment and applicant's validation tools. Expresso says that despite the clear and fast disruption of the human resources sector in the EUA, change has been slow-paced in Portugal, "both because of legal issues (such as data protection) and because the human component continues to be highly valued in this market".

## / Urban Co-creation Data Lab

With one million euros in funding from the European Union (under the Connecting Europe Facility), the project [Urban Co-creation Data Lab](#) intends to respond to five challenges in the area of municipal management: micro-mobility, waste, parking, pollution, and crowd management. The project started in Lisbon, but it will also be tested in two other European cities which have yet to be selected. According to the project's leader, the main objective is "bringing together the best European brains in urban analytics to collaboratively create models of artificial intelligence to support decision-making in municipal planning and management". ([source](#))

## / Irene – Your citizen's card assistant

In 2019, at the Web Summit, the Institute of Records and Notaries (IRN) presented Irene – a virtual assistant whose mission is to help users on matters regarding the Citizen's Card. At the event, IRN's vice president [described](#) the capabilities of the service: "In addition to answering a set of questions, Irene engages in a conversation in which she

asks a set of questions to the citizen, in order to understand their particular situation and thus be able to create an optimal experience in the renewal of the Citizen's Card". The system is currently being tested and is available on [IRN's website](#).

## / Security video analytics

Founded in 2017, Heptasense offers a video analytics service for surveillance cameras to automate and optimize security management. The company [claims](#) to be "against facial recognition" and believes that "video surveillance systems have always taken this path of espionage, of invasion of privacy". Their idea is to detect suspicious behavior without face recognition turning conventional surveillance cameras into intelligent systems. In their own words: "the system is able to analyze a person's behavior and thus anticipate its possibly illegal or suspicious actions, helping the security to prevent dangerous incidents", adding that "it is the first AI software that learns and understands human behavior patterns, mimicking the way the human brain works and identifying potential threats to safety and security".

It is not clear if this is the system used in the previously mentioned projects in Leiria and Portimão related to the use of AI in video surveillance systems, although there are similarities in the description of the features. However, the criticism put forward by the Data Protection Authority also applies, namely the issues concerning fundamental human rights and the opacity of the system. (please see "Police Forces" section).

# Policy, oversight and public debate

## / Government

### / AI Portugal 2030 – The National Strategy for Artificial Intelligence

In 2019, the government presented the [National Strategy for Artificial Intelligence](#). The document was created by the Ministry of Science, Technology and Higher Education and its main objective is to make Portugal part of the European network of centers of excellence in artificial intelligence, in the short to medium term.

The Action Plan includes seven points of action:

- 1) Inclusion and Education – widespread dissemination of knowledge in AI,
- 2) qualification, and specialization,
- 3) thematic areas for research and innovation in European and international networks,
- 4) public administration and its modernization,
- 5) specific areas of specialization in Portugal with international impact,
- 6) new developments and support areas in European and international networks, and
- 7) new challenges of society brought by AI: Ethics and Safety.

### / Program in Data Science and AI in Public Administration

In 2018, the government [launched a program](#) to support new R&D projects in a partnership between the Public Administration and scientific institutions: “The projects should be oriented to deepen the processing of public data and stimulate the production of new knowledge relevant to citizens, based on the use of advanced techniques of artificial intelligence and data science”.

In 2018, the following projects were the winners: Use of Artificial Intelligence to Enhance Teledermatological Screening; Data Science for Emergency Medical Service Optimization; Identification and Prediction of Hospital Emergency Demand; Ground Recognition System. A new edition of the program was launched in 2019.

### / Oversight

While there is no specific law or oversight entity for ADM processes, ADM might be relevant for the activity of sector regulators.

### / The National Competition Authority warns: algorithms might be breaking the law

In July 2019, the Competition Authority [alerted](#) companies that sell goods or services online to the possible implications of the algorithms they use. In its [report](#), the authority included a survey of 38 companies with an online presence in Portugal. 37% of the companies surveyed admitted using algorithms to monitor the prices of competitors. Of those, 78.6% said they adjusted prices according to the results,

and 7.9% of those made those adjustments automatically, and without human intervention.

The regulator considers that “the monitoring of competitors’ strategies and the adjustment of the price or other strategic variables based on the information collected have always been part of the normal functioning of the markets”. However “their frequency and extent intensify in the digital economy”, and so “the frequency of their use can facilitate collusion strategies”. It adds that ultimately companies “are responsible for the algorithms they use”.

## / Civil Society and academia

### / APPIA

Founded in 1984, the main purpose of the [Associação Portuguesa Para a Inteligência Artificial](#) (Portuguese Association for Artificial Intelligence, APPIA) is “promoting research, teaching and dissemination of artificial intelligence as a scientific field”. With over [300 senior researchers as members](#), APPIA has long been an active voice in the field, especially within academia. APPIA organizes multiple events around AI, the main ones being a yearly AI [conference](#) and an advanced course on AI.

### / Calouste Gulbenkian Foundation

Calouste Gulbenkian Foundation is a Portuguese institution dedicated to the promotion of the arts, philanthropy, science, and education. As a major reference in the promotion of such areas, especially through funding, the foundation plays a vital role in the areas of activity it covers. Furthermore, Gulbenkian [offers scholarships](#) intended to attract students from Portuguese higher education institutions to research in Artificial Intelligence.

### / Associação D3 – Defesa dos Direitos Digitais (D3)

D3 is a digital rights organization, and a member of the European Digital Rights network. The organization has voiced [its public position against](#) the use of artificial intelligence systems by police forces for public surveillance purposes.



## Key takeaways

Portugal seems to be in a very good position to develop its ADM capabilities, despite being far from a world leader in the industry. As a result of extensive academic work, together with a strong, well prepared technological sector and competitive wages, there is the potential for unique growth opportunities for the small and medium-sized companies that have come to dominate the Portuguese business sector in recent years. This is increasingly felt by various actors, especially those in both the business and political arenas. The creation of the National Strategy for Artificial Intelligence by the government is one such example. However, there is a feeling that, while the economic opportunities of AI are immense, there is a limited time-frame, which cannot be missed. The main economic barriers are the usual difficulties of the Portuguese economic fabric relating to access to capital and the lack of scale of most enterprises.

On the other hand, public criticism and debate related to the dangers and opportunities created by ADM processes remain limited. The fact that in Portugal, many of the controversial solutions adopted in other European countries are not present has meant that citizens have not yet felt the need to bring these issues up in public discussion, despite the fact that both the public and the press follow developments abroad very closely. Regardless of this fact, the research shows that new initiatives are emerging very rapidly in Portugal, within the AI and ADM sectors. However, most of these initiatives are still in the very early stages. Therefore, even when ADM mechanisms are in place, citizens may not feel their presence as they might have a very limited application for small specific groups within the larger population.

## References:

- AI Portugal 2030 (2020): AI Portugal <https://www.portugal.gov.pt/download-ficheiros/ficheiro.aspx?v=236848b1-fcb6-4c65-9773-292d1c5b9ad1>
- Antena 1 (2019): Prova de vida inovadora simplifica obrigação dos pensionistas [https://www.rtp.pt/noticias/pais/prova-de-vida-inovadora-simplifica-obrigacao-dos-pensionistas\\_a1191491](https://www.rtp.pt/noticias/pais/prova-de-vida-inovadora-simplifica-obrigacao-dos-pensionistas_a1191491)
- Baptista,Sofia Correia (2020): Há um laboratório de dados para ajudar a resolver os problemas das cidades <https://www.publico.pt/2020/01/03/local/noticia/ha-laboratorio-dados-ajudar-resolver-problemas-cidades-1892557>
- Barbado, Rute (2017): Algoritmos ajudam a encontrar o profissional certo num milhão <https://expresso.pt/economia/2017-08-12-Algoritmos-ajudam-a-encontrar--o-profissional-certo-num-milhao>
- BPI (2020): Crédito Imediato BPI <https://www.bancobpi.pt/particulares/credito/credito-pessoal/credito-imediato>
- Base (2016): Contrato e Aquisição de Licenças <http://www.base.gov.pt/Base/pt/Pesquisa/Contrato?a=2804641>
- Figueria, Alexandra (2018): Câmaras em casinos para travar viciados <https://www.jn.pt/nacional/camaras-em-casinos-para-travar-viciados-9736330.html>

Fundação Calouste Gulbenkian (2020): Novos Talentos em Inteligência Artificial <https://gulbenkian.pt/grant/novos-talentos-em-inteligencia-artificial/#>

Guerra, Ana Rita(2019): Tecnologia portuguesa leva inteligência a câmaras de videovigilância <https://www.dinheirovivo.pt/vodafone-business-conference/tecnologia-portuguesa-leva-inteligencia-a-camaras-de-videovigilancia/>

Guerra, An Rita (2019): Portugal “tem competências” em inteligência artificial e pode “dar o salto” <https://www.dinheirovivo.pt/vodafone-business-conference/portugal-tem-competencias-em-inteligencia-artificial-e-pode-dar-o-salto/>

EPIA Conference on Artificial Intelligence (2019): <https://epia2019.utad.pt/>

Instituto dos registos e do notário (2020): <https://irn.justica.gov.pt/>

Lopes, Bruno Faria (2019): Os caca-fraudes na saúde <https://www.sabado.pt/portugal/detalhe/os-caca-fraudes-na-saude>

Lusa (2019): Costa despede-se da Índia com visita a empresa portuguesa com tecnologia modelo <https://expresso.pt/politica/2019-12-20-Costa-despede-se-da-India-com-visita-a-empresa-portuguesa-com-tecnologia-modelo>

Lusa (2019): Aeroporto de Lisboa escolhido para ser centro de excelência na área de gestão de fluxos [https://www.rtp.pt/noticias/economia/aeroporto-de-lisboa-escolhido-para-ser-centro-de-excelencia-na-area-de-gestao-de-fluxos\\_n1183809](https://www.rtp.pt/noticias/economia/aeroporto-de-lisboa-escolhido-para-ser-centro-de-excelencia-na-area-de-gestao-de-fluxos_n1183809)

Lusa (2019): Sorteio da Operação Marquês. Processo disciplinar contra juiz Carlos Alexandre arquivado <https://observador.pt/2019/11/19/sorteio-da-operacao-marques-processo-disciplinar-contra-juiz-carlos-alexandre-arquivado/>

Mateus, Cátia (2019): Uber cria sistema para verificar da identidade dos motoristas em Portugal <https://expresso.pt/economia/2019-12-17-Uber-cria-sistema-para-verificar-da-identidade-dos-motoristas-em-Portugal>

Moreira, António Vasconcelos (2019): Algoritmos já concedem crédito <https://jornaleconomico.sapo.pt/noticias/algoritmos-ja-concedem-credito-414670>

Pereira, Micael(2018): Sorteio do juiz da Operação Marquês decorre esta sexta e vai ser filmado pela SIC <https://expresso.pt/sociedade/2018-09-27-Sorteio-do-juiz-da-Operacao-Marques-decorre-esta-sexta-e-vai-ser-filmado-pela-SIC>

Pequinino, Karla (2019): Autoridade da Concorrência avisa que algoritmos de preços podem infringir a lei <https://www.publico.pt/2019/07/02/tecnologia/noticia/autoridade-concorrenca-alerta-empresas-portuguesas-uso-abusivo-algoritmos-compras-online-1878456>

Polígrafo (2018): É verdade que o sorteio do Processo Marquês pode ter sido manipulado? <https://poligrafo.sapo.pt/fact-check/e-verdade-que-o-sorteio-do-processo-marques-pode-ser-manipulado>

Procuradoria Geral Distrital de Lisboa (2020): Reformula a Lei do Jogo [http://www.pgdlisboa.pt/leis/lei\\_mostra\\_articulado.php?artigo\\_id=242A0038&nid=242&tabela=leis&pagina=1&ficha=1&so\\_miolo=&nversao=%23artigo](http://www.pgdlisboa.pt/leis/lei_mostra_articulado.php?artigo_id=242A0038&nid=242&tabela=leis&pagina=1&ficha=1&so_miolo=&nversao=%23artigo)

Secretaria-Geral do Ministério da Justiça (2019): Tem dúvidas sobre o Cartão de Cidadão? A assistente Irene pode ajudar <https://justica.gov.pt/Noticias/Tem-duvidas-sobre-o-Cartao-de-Cidadao-A-assistente-Irene-pode-ajudar>

Younited credit (2020): Como funciona o empréstimo online <https://pt.younited-credit.com/como-funciona-o-emprestimo-on-line>

Trigueirao, Sónia (2019): Casinos: Turismo de Portugal gastou 338 mil euros mas reconhecimento facial não funciona <https://www.publico.pt/2019/05/06/economia/noticia/turismo-gastou-338-mil-euro-reconhecimento-facial-nao-funciona-1871551>

Turismo de Portugal (2020): Autoexclusão e proibição <https://www.srij.turismodeportugal.pt/pt/jogo-responsavel/autoexclusao-e-proibicao/>

Urban Data-Lab (2020): 5 Challenges <http://www.urbandatalab.pt/>

Wikipedia (2020): Random number generation [https://en.wikipedia.org/wiki/Random\\_number\\_generation](https://en.wikipedia.org/wiki/Random_number_generation)



BELGIUM  
DENMARK  
ESTONIA  
FINLAND  
FRANCE  
GERMANY  
GREECE  
ITALY  
NETHERLANDS  
POLAND  
PORTUGAL  
**SLOVENIA**  
**STORY**  
PAGE 208  
**RESEARCH**  
PAGE 214  
SPAIN  
SWEDEN  
SWITZERLAND  
UNITED KINGDOM





Find out more on p. 214 in the research chapter under "E-social services".



# Slovenian police acquire automated tools first, // legalizes them later

**The Slovenian police legalized its use of face recognition five years after it started to use it. Despite formal safeguards, no institution can restrain the Interior ministry.**

By Lenart J. Kučič

When Slovenian journalists or activists ask officials whether the police, the secret service, and the army are using any technological tools for mass surveillance, they are often reluctant to admit even the existence of such devices. However, when they meet their colleagues at international security conferences or receive questions from foreign journalists, they want to make a positive impression. “Of course, we are testing (or using) the latest new gadgets to fight terrorism, organized crime or illegal migration,” they say. “We are a modern and technologically advanced security force, after all!”

This may sound anecdotal but it happens to be true.

When Slovenian civil society activists and privacy experts read the 2019 AlgorithmWatch [report on face recognition](#) in the EU, they were surprised to find Slovenia listed among the countries where the police use face recognition.

“Slovenian police personnel have been using face recognition since 2014. The software has been developed internally,” the report claimed, quoting the Slovenian Police as the source.

It turned out Slovenian journalists and privacy activists might have been following the wrong sources.

How do you detect evil aliens before they can invade human society? You need to follow the “hot sheets”, said special agent “K”, played by American actor Tommy Lee Jones in the hit sci-fi movie “Men in Black”.

What he meant by “hot sheets” were the supermarket tabloid magazines reporting on alien abductions and other strange and bizarre stories. You may also get lucky by reading the *New York Times*, “K” explained to the rookie agent “J” (Will Smith). But tabloids are a much better source of information if you want to control alien activity on Earth and protect humanity.

But the “hot sheets” do not only reveal potential aliens. They can also alert privacy advocates to unusual police activity.

“Aggressive pensioners recognized on Facebook,” [claimed](#) the headline in *Slovenske novice*, the biggest Slovenian print tabloid. A woman from Ljubljana, the Slovenian capital, left her car on a bicycle track. She left it with the blinkers on, and she ran to a nearby restaurant. It only took her a minute to come back, the tabloid wrote, but her parked car had already annoyed an elderly couple.

The elderly man smashed one of the side mirrors with his fist and then tried to hit the woman as well. She took a photograph of the couple with her phone and reported the attack to the police. She also published the photographs on Facebook where some users said they recognized the violent couple.

The tabloid article mentioned that the police had been using a new piece of face recognition software called Face Trace to find suspects by using open source investigation methods (such as searching social media and other online sources).

The tabloid article was originally published in December 2015 – four years ahead of the AlgorithmWatch report on face recognition. The CEO of the company that developed the Face Trace software even linked to the piece from his social media account. After thorough research of the Slovenian media archives for this feature, I found an even [earlier mention](#) of Face Trace in one of the major Slovenian daily newspapers, *Delo*, where a journalist talked to the police operator who was responsible for face recognition in 2014.

But no-one in Slovenian civil society took notice at the time.

## **/ “I saw my friend in the police database”**

The articles in *Slovenske novice* and *Delo* about face recognition are not isolated examples.

In January 2017, a woman contacted a journalist writing for the national television website MMC. She was stopped by traffic police. She was not carrying her ID, so the police officer brought her to his car to confirm her identity on the police computer. She gave her personal information (name, address, date of birth) and photographs of her appeared on the screen. She expected to see her official photographs –



**SLOVENIAN POLICE  
HAVE BEEN USING FACE  
RECOGNITION SOFTWARE  
SINCE 2014.**

from her ID, passport, and driving license – but she also noticed photographs that she had published on her personal Facebook profile.

Why would the police store her Facebook photographs in their database? she wondered. Was that even legal?


The police declined to confirm or deny that they gathered pictures from social media profiles in their database. Their spokesperson then suggested to the journalist that the police officer may have also checked publicly available photographs on the Internet during the procedure, which could have given a wrong impression to the driver (that her Facebook photographs were stored in the police database). Instead, they claimed that everybody could access the information on social media and provided a lengthy legal argument on why citizens must always carry a valid ID to avoid such police procedures and paying a fine.

A similar example dates back even further – before Face Trace was officially introduced. A citizen told us that they filed a complaint to the Information Commissioner in September 2013. The complaint included a passage where a young woman reported an attempted sexual assault to the police. When she browsed through the police database to find and recognize the alleged attacker, she also saw a photo of her friend.

“He was never accused or convicted of any wrongdoings,” the complaint states. He did join some peaceful public protests, though. Is it possible that the police take photographs of protesters and store them in a database of photographed individuals, or in a database of “extremists”? she wondered. And on what legal basis could the police do such a thing?

The Information Commissioner replied to the complaint in February 2014. The police denied that a database of “extremist” protesters existed, the Commissioner wrote. But they explained that the database of photographed individuals could also contain a photograph of one’s friend or relative if this person has ever been suspected of a criminal act. The police can also record or photograph public events and keep “relevant pieces” of the material. At that time, the police database of photographed individuals contained around 30,000 people, according to the Commissioner.

From the examples above, it is possible to conclude that Slovenian police have been using face recognition software since 2014. Also, the police confirmed to me that they use face recognition systems to compare, for example, police



THE POLICE ARE OFTEN ALLOWED TO INTERPRET THE LEGISLATION AS IT BEST SUITS THEIR NEEDS - WITHOUT MUCH OPPOSITION FROM THE PUBLIC, POLITICIANS, OR PRIVACY EXPERTS.

sketches to pictures in their database of photographed individuals. This is important because the database could contain photographs from personal social profiles and mass public events.

We are using conditionals because we do not know for certain what photographs the police actually keep in their database. During our investigation, the official answers provided by the police remained ambiguous and the Information Commissioner has not yet decided to investigate the matter and evaluate the use and content of police databases.

## / Hacking the law

In addition, the police are often allowed to interpret the legislation as it best suits their needs – without much opposition from the public, politicians, or privacy experts. “The police always follow the same pattern. First, their representatives dramatically demonstrate all the new threats to public life: international terrorism and organized crime, migrant crisis, and online threats – from pedophilia to cyberattacks. Next, they express their frustrations and concerns because they cannot efficiently fight the bad guys who have all the latest tools and no legal restrictions on how to use them. So, they wait for every opportunity to amend the existing legislation and expand their powers,” explained a public official who took part in several negotiation rounds with the police and who requested anonymity.

The police always present a very ambitious wish list, our source said. Some wishes are clearly not realistic and would never pass the legislative process: weakening the encryption on communication apps or using drones for mass surveillance. But they can be used in negotiations with the Information Commissioner and other privacy advocates.

“Fine, let us not touch the encryption,” say the police. But we still need to amend the outdated law on duties and powers of the police and include, for example, biometrics. After biometrics is mentioned in the legislation the police can interpret this as a legal basis for using face recognition.

“They are hacking the legislation this way. This is like installing a piece of malware to the operating system and then using the exploits to legalize whatever you are doing,” concluded the anonymous source.

Such hacking of the legislative process by the police is not merely hypothetical. When the Slovenian police bought its first IMSI catcher (a tool to snoop on cell phones) in 2004, there was no legal basis for its use, the Slovenian online medium *Slo-Tech* revealed. Nevertheless, the police secretly used the machine in more than 300 cases, from 2006 to 2012, and then bought another IMSI catcher in 2012.

When lawyers, journalists, and activists first challenged the allegedly illegal use of the IMSI catchers the police spokesperson denied the existence of such devices. However, they had to change their communications strategy after a senior police official said in an interview on public radio that they were using IMSI catchers “to find missing persons”. Instead of denying their purchase and use, the police tried – with a lot of help from the government – to retroactively legalize IMSI catchers.

The police approach to face recognition was very similar.

The Slovenian Information Commissioner confirmed that the police informed them about the use of the face recognition software, Face Trace, back in 2014. The police claimed that the Commissioner agreed to their suggested uses of the product. But the Commissioner’s office actually issued several critical comments regarding the use of biometric methods and face recognition between 2015 and 2019. The Commissioner’s office also appealed to the Slovenian Human Rights Ombudsman to file a formal complaint to the constitutional court when the Slovenian Ministry for Internal Affairs introduced a new law on the duties and powers of the police in 2017.

The new amendments allowed the police to use biometric and face recognition tools, and thus provided a legal framework for the use of the Face Trace software. Again, the police tried to retroactively legalize its use and the government ignored all the warnings from the Information Commissioner. How was this possible? There was no public pressure, argued the Information Commissioner. Nobody seemed to care about their warnings against potential police abuse of biometry and face recognition. As a consequence, the police are now allowed to automatically combine face recognition with other biometric data like fingerprints and DNA profiles.

***THE POLICE TRIED TO  
RETROACTIVELY LEGALIZE ITS  
USE AND THE GOVERNMENT  
IGNORED ALL THE WARNINGS  
FROM THE INFORMATION  
COMMISSIONER.***

## / No public pressure?

The police, nonetheless, did face some opposition. In the 2019 Automating Society report, we learned that the Slovenian Human Rights Ombudsman, and the Information Commissioner, filed a formal complaint to the constitutional court in 2017. They claimed that the new law on the duties and powers of the police legalized some excessive and inadmissible measures for gathering personal data without sufficient protection of citizens who have not been accused or suspected of any wrongdoing. In addition, they warned against the use of drones and automated systems for recognizing license plates, as well as the unregulated use of passenger name records at Slovenian airports. But for some reason, biometric systems (such as face recognition) were not included in the complaint.

How so? The Information Commissioner explained that they suggested to the Slovenian Human Rights Ombudsman that biometrics should also be included in the complaint. But that did not happen. The representatives of the Ombudsman explained (to us) that there had been no public debate or controversy surrounding the use of biometrics by the police at the time. They had a legal option to file another complaint to the constitutional court and address the use of face recognition but they did not take any action. Why? Because they did not want to over-use this exceptional legal measure, according to the Ombudsman.

This argument clearly illustrates some of the problematic aspects regarding the implementation and use of Artificial Intelligence (AI), algorithms, and automated decision-making systems (ADM) in Slovenia. During our research for the 2019 Automating Society report, very few civil society organizations or individuals addressed the social impacts of automation on society. In addition, conferences about AI are usually sponsored and organized by commercial companies and industry groups who want to promote their solutions. Furthermore, many of the specialized journalists who cover such issues have quit the media in recent years. As a result, there is no rigorous coverage on the topic and privacy experts or activists have lost access to the general public.

Civil society should not be blamed for this state of affairs. "Slovenian non-governmental organizations (NGOs) are small and understaffed. As far as I know, no NGO is monitoring the police and the use of their powers systematically," said Katarina Bervar Sternad from the [Legal Centre for the Protection of Human Rights and Environment](#). "We

cannot afford to dedicate one staffer to this particular field. This is unfortunate because we receive many questions regarding potential abuses of police powers."

Domen Savič, from the NGO Državljan D, makes a similar point. "The NGO sector is mostly funded by the government or international EU grants," Savič said. If there are no tenders available for the information society, then NGOs cannot afford to cover this field. Furthermore, the media and politicians rarely understand how technology impacts human rights. "When AlgorithmWatch revealed that Slovenian police [were] using face recognition software to profile... citizens and some Slovenian media investigated the story, there was not a single voice of protest raised in the parliament."

To make things worse, there is almost no collaboration between politics, the media, and civil society, warned Savič. "It takes an extraordinary amount of effort, time, and pressure to actually change a policy or a police practice. But most of the issues die out pretty quickly even if there is someone who tries to get answers from the people involved."

Consequently, the police face almost no opposition when handing in their "wish-list" to the government.

Another worrying example is a recent attempt by the Ministry of Justice to use the GDPR – to which the necessary amendments to Slovenian legislation must comply – to exclude the police from civilian control. The Information Commissioner warned that the ministry wanted to redefine its role under the new (amended) law on personal data protection. According to the ministry's proposal, the Information Commissioner would no longer be able to oversee the police.

## / Authoritarian tendencies

Such safety mechanisms are critical and monitoring of the police should not be left only to NGOs and activists. Mr. Savič, Ms. Bervar Sternad, and the anonymous official all agreed that a common European regulatory approach would be necessary because national NGOs cannot effectively monitor how the police use the technology. EU Member States, on the other hand, can use the sense of emergency to adopt new pieces of legislation and bypass parliamentary debate. The current health crisis is the most recent example of such an attempt.

Soon after the COVID-19 pandemic started earlier this year, the Slovenian government began using the crisis as an excuse to significantly expand police powers.



The method they used was similar to the previous attempts at legislation hacking. In March 2020, the Slovenian government [proposed](#) a first draft of the “Act on the intervention measures to mitigate the consequences of the communicable disease SARS-CoV-2 (COVID-19) for citizens and the economy”. They submitted the text to the National Assembly for consideration and adoption under an emergency procedure. However, the draft also included two articles that dramatically increased the powers of the police.

Article 103 suggested that the police could use various methods to ensure that the citizens obey the quarantine and the Communicable Diseases Act. They could – among other measures – use face recognition when stopping and identifying individuals, enter their houses or apartments, limit their movement as well as collect and process protected personal information such as medical data provided by the National Institute of Public Health. Article 104 went even further. It suggested that the police could trace the location of the mobile phones of individuals without a court warrant.

All the suggested measures were introduced under an emergency procedure – without any consultations or public debates. The Information Commissioner thus [described](#) the anti-COVID-19 measures as an attempt to potentially “establish a police state”. They considered the new police powers to be too broad and potentially unconstitutional and undemocratic. The Human Rights Ombudsman [wrote](#) that it was hard to believe that such measures were really necessary and proportional (neither institutions were consulted during the process). A critical commentary was also [published](#) by the members of the Institute of Criminology who wrote that mass surveillance is not compatible with European legal culture.

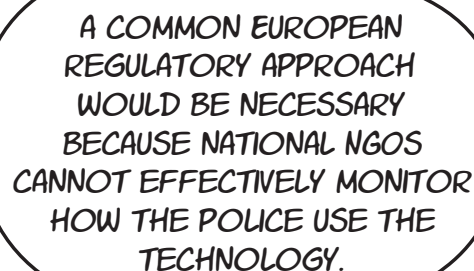
Article 104 was eventually removed from the amended act because of strong criticisms from the public and the opposition political parties. However, article 103 on the powers of the police remained in the “[COVID Act](#)” that was adopted in April 2020. Furthermore, the government insisted that contact tracing applications were necessary to help health officials stop the pandemic. They also suggested that citizens would have to install such an application when traveling across the country. Data from the application would be collected and used by the National Institute of Public Health, but the police would also be allowed to access the database.

## / Public protests

When the first groups of citizens started protesting against the government in April 2020, Interior minister Aleš Hojs [demanded on Twitter](#) that the police use their new powers to identify and prosecute the protesters, such as collecting and analyzing all available images from traditional and social media. The government claimed that the protesters were violating laws on public health, and that any public gathering was illegal.

According to what we know, the police can collect and analyze photographs from social media. The police also routinely record public events, protests, and other mass gatherings. Furthermore, the new legislation could allow the police to access other kinds of personal information (contact tracing) to curb future anti-government protests.

It takes many small steps to build a – partly automated – surveillance infrastructure. The Slovenian police have been consistently taking such steps and the “COVID Act” has most likely legalized most of the items on their “wish-list”, from face recognition to contact tracing.



A COMMON EUROPEAN  
REGULATORY APPROACH  
WOULD BE NECESSARY  
BECAUSE NATIONAL NGOS  
CANNOT EFFECTIVELY MONITOR  
HOW THE POLICE USE THE  
TECHNOLOGY.

# Research

By Lenart J. Kučič

## Contextualization

In biology and medicine, viruses are considered to be agents that provide a stress test for the host organism. They can show how robust the immune system is, and where its weaknesses are. The new coronavirus is no exception. Not only has the virus attacked its human hosts, it has also shown that many European governments could not resist the temptation to use the pandemic to increase the powers of the police. And that their immune system – civil society, experts or activists – has often been too weak to prevent their attempts.

Why such an analogy?

### / Hacking the legislation

There are no policies, strategies, or guidelines on how to implement new technologies for collecting and processing data in Slovenia. The systems for automatic decision-making (ADM) have, therefore, been implemented in many branches of government, such as social services and tax authorities, without much (if any) public debate. Furthermore, whenever the police have attempted to “hack the legislation” – to use any kind of crisis to increase its powers – it has faced only weak opposition from individual privacy experts and activists.

The global pandemic of “the new coronavirus” COVID-19 is the most recent example. During the first wave of the pandemic, the government introduced the Act on intervention measures to mitigate the consequences of COVID-19. But this huge legislative package also included some articles that have significantly increased the powers of the police, which was allowed to collect and process some “special kinds” of personal information, e. g., medical, location, and biometric data (photographic and video images for face recognition).

Furthermore, the government adopted another package of anti-COVID-19 measures in July, which provided a legal

basis for introducing the mobile application for contact tracing before the application was introduced and tested. According to the new law, the app is obligatory for citizens who are tested positive for the coronavirus or who are in quarantine.

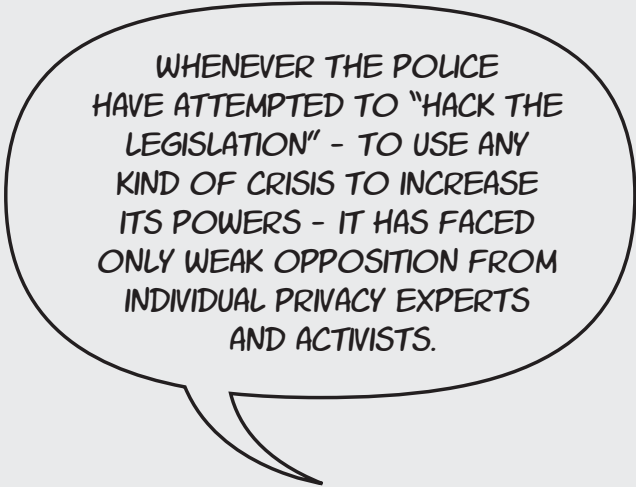
The recent pandemic has shown that the power of the state has increased significantly during the crisis and that the “immune system” was too weak to prevent the new increase of police powers. The European Union should, thus, prepare a common framework that is also legally binding for Member States. Slovenian NGOs are too small and do not have enough expert knowledge to be the principle protectors against state and private misuse of ADM. And Slovenia is probably not the only such example.

## A catalog of ADM cases

### / E-social services

In 2010, the Slovenian government introduced the e-Sociala (e-social services) program to optimize social transfers, such as social and unemployment benefits, child benefits, subsidies, etc. that make up the welfare system. The system consists of many connected databases (Ministry of Public Administration, 2009) and was designed to optimize the work process for social workers, evaluate their work, and prevent errors when processing applications for social welfare. We learned from the Ministry of Labor, Family, Social Affairs and Equal Opportunities that the system has been upgraded since its introduction and that it now includes AI, ADM, and machine learning capabilities.

The ministry spokesperson provided several examples of ADM. For example, citizens can use ADM to learn their social rights and see a prediction of benefit payments so that they can make an informed evaluation and calculation before applying for social benefits. Furthermore, the



WHENEVER THE POLICE  
HAVE ATTEMPTED TO "HACK THE  
LEGISLATION" - TO USE ANY  
KIND OF CRISIS TO INCREASE  
ITS POWERS - IT HAS FACED  
ONLY WEAK OPPOSITION FROM  
INDIVIDUAL PRIVACY EXPERTS  
AND ACTIVISTS.

application processes for several social rights/benefits is fully automated, e.g., to research previous decisions related to social benefits and to see new (informative) ones for particular applicants. E-sociala can also detect people from the same family and access their financial information, thereby simplifying the process for social workers.

The goal is to further automate the process so that the system can create a full "social rights profile" for every applicant. The system will then be able to see what benefits a particular citizen is eligible for, for how long, and for what amount of money. For legal reasons, social workers still need to personally evaluate, print, sign, and send every application. At the moment, the process still requires some manual (human) intervention which is recorded and analyzed in order to improve the system and prevent errors.

The ministry admits that some human input will always be necessary because of changing and ever more complex life scenarios, and the quality of the input data. However, they hope to automate their system to such a level that social workers will only need to intervene in very specific situations.

## / Face recognition at Jože Pučnik International Airport

In May 2019, the former Slovenian airline company, Adria Airways, announced ([STA, 2019](#)) that they had successfully tested a pilot project on biometric boarding together with Polish airline operator LOT.

The two airline companies were a testing ground for the US IT company, Amadeus, which is developing and implementing a large centralized database of biometric data for the airline industry. The biometric boarding system helped 175 passengers to board in "a record short time". Boarding took only two seconds per passenger instead of five to seven seconds, according to the press release.

The passengers were asked to install the Amadeus smart-phone app, take a selfie, photograph their passport (including their portrait photograph), and upload the data to a remote server. The passenger photographs were then captured at the time of boarding, and the system automatically verified each individual. Following a positive match, a confirmation message was sent to the boarding system, and the passenger was allowed to board. All the biometric data used during the test was deleted within 48 hours of the test, the Adria representatives said.

During the research for this report, we learned that Amadeus has signed a partnership agreement with Ljubljana airport, which is operated by the German company, Fraport. Amadeus helped Fraport to upgrade its information systems, and biometric boarding was also a part of this update. The system should have been fully functional by the end of 2019. However, there was no additional information available at the time of writing.

## / Smart Cities

An increasing number of Slovenian cities, and local municipalities, have joined the "smart city" initiative. These cities include Maribor, Kranj, Ljubljana, and Novo Mesto, among others.

However, there are many problems with these "smart city" projects. The state provided no strategy or guidelines on how to implement "smart" solutions (like ADM) and it has no oversight on what is happening at the local level. Furthermore, some local municipalities seem to have little or no knowledge of how their systems are designed and operated, who hosts and analyzes the data, and whether or not there are any ADM systems, etc.

Furthermore, "smart city" initiatives are still relatively unsophisticated. For example, the most ambitious "smart city" plans were announced in the city of Novo Mesto in 2018 ([Rajšek, 2018](#)). The city authorities announced that – in collaboration with the German IT company, SAP, and the Slovenian national telecommunications operator, Telekom

Slovenije – Novo Mesto would become the “first reference smart city” in Slovenia. Among the “smart” services were the real-time measurement of air quality, availability of public parking spaces, traffic analysis, intelligent street lighting systems, and the measurement of public opinion (i.e. the public’s satisfaction with public services). The city is also testing IoT (Internet of Things) solutions to gather and analyze data more effectively.

However, according to a spokesperson for Novo Mesto, this “first reference smart city” is not using any ADM or AI systems as the city officials are still trying to find practical use of the data.

**SOME LOCAL MUNICIPALITIES SEEM TO HAVE LITTLE OR NO KNOWLEDGE OF HOW THEIR SYSTEMS ARE DESIGNED AND OPERATED, WHO HOSTS AND ANALYZES THE DATA, AND WHETHER OR NOT THERE ANY ADM SYSTEMS.**

## / Financing Hate Speech

In 2018, Slovenian NGO Državljan D started a [campaign](#) ([Državljan D, 2018](#)) to document online ads from big Slovenian advertisers – including state and state-owned companies – that appeared on various media websites, and regularly featured problematic, hateful, and false articles on migrants, ethnic, and religious minorities (Muslims, LGBTQ members, etc.). Državljan D argued that advertisers were financing hate speech and called on companies to stop advertising on such websites.

At the beginning of the campaign, the advertisers tried to blame “advertising algorithms” for the problematic placement of their ads and they denied any responsibility. However, what looked like an anecdotal example of ADM turned bad, soon became a political (and politicized) story. In November 2018, Slovenian Prime Minister Marjan Šarec joined the debate when he published a public statement calling for advertisers to reconsider their marketing strategies when their clients promote hate speech (e.g., racism and homophobia).

However, state and state-owned companies continue to advertise on websites that promote hate speech. Among such companies are the national incumbent telecommunications operator Telekom Slovenije, Slovenian Railways, and some other companies where the state has some influence over the advertising budgets. Meanwhile, financing hate speech through (state) advertising has become a regional and international problem. The Slovenian investigative

medium Pod črto [published](#) a series of articles about Slovenian media ([Kučič, 2019](#)) and learned that political actors have built a parallel media system that also includes a network consisting of 17 regional Internet portals.

Since 2017, the portals have published more than 20,000 articles. Most of the content is neutral, but the anonymous authors also publish a lot of articles about migrants and other political (or politicized) topics. This content is then promoted on social media platforms to target individuals on their profiles. The

financing for the portals likely comes from

Hungarian media businessmen close to the Hungarian political party Fidesz and Prime Minister, Viktor Orban, who is expanding his party media empire to neighboring states to support his political allies according to many cross-border investigations by journalists ((see, for example, [Cvetkovska, Zöldi, and Deli, 2020](#)).

Similar attempts have also been documented in other countries in the region, including Macedonia, Hungary, and Poland, among others. State advertising and the money from state-owned companies have been used – sometimes illegally – to promote political agendas of the ruling political parties and influence public opinion. Playing with algorithms on social media platforms has become an essential part of the game. And this “game” also includes targeted attacks on the political opposition, journalists, NGO’s, and private individuals, according to media reports.

Since we included the story in our first report, the focus of this example has shifted from ADM to computational propaganda. But it should be seen as a warning, because some recent examples have shown (e.g., The Cambridge Analytica case) that the use of big data, machine learning, and targeted advertising (which is often automated) has a strong potential to distort public opinion. Especially when the government or a political party can finance such an operation with state and public money.

## / Tax evasion and tax fraud

The Ministry of Finance’s financial administration has been using machine learning to detect tax evasion schemes,

and tax fraud, and to find errors in tax reports. During the research for this report, we also learned that the ministry ranks “risky” citizens who it thinks are more likely to become tax evaders.

This year, we learned about some additional information about this system. In 2017, the financial administration started using a fraud management software system, produced by SAP. The administration also formed working groups to find the most likely scenarios for tax evasion and tax fraud. The groups were asked to define the indicators of risk that may help detect illegal activities. They listed 158 risk factors and used them to select around 17,500 individuals and companies for tax inspection. The inspectors found irregularities in more than 75% of all the selected cases.

The system proved to be a useful tool for tax inspectors, according to the spokesperson. The tax authority would not give us specific examples on how they are using their prediction models because they would not want help future tax-evaders to “game their system”.

The spokesperson explained that every new tax evasion indicator (alert) is instantly shared with all users of the system (tax inspectors). The system can also reduce the risk of tax inspectors deliberately avoiding more demanding or “untouchable” cases. However, a spokesperson for the financial administration explained that fraud detection cannot be fully automated. The quality of data and the rapid evolution of tax-evasion systems being the biggest obstacles for a tax ADM.

***THE MINISTRY OF FINANCE'S  
FINANCIAL ADMINISTRATION  
HAS BEEN USING MACHINE  
LEARNING TO DETECT TAX  
EVASION SCHEMES, AND  
TAX FRAUD, AND TO FIND  
ERRORS IN TAX REPORTS.***

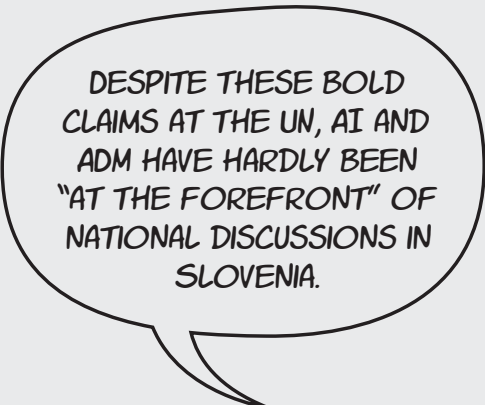


## / Early warning for educational problems

Slovenia was one of the first countries in the world to develop and test an intelligent system to detect early learning problems in primary schools (METIS).. The system was designed to monitor pupils' grades and absences, to search for specific learning patterns, and help teachers find "problematic" pupils. Unfortunately, METIS was also an example of how not to implement an ADM process into the educational system.

Firstly, the budget was too small. The Ministry of Education, Science and Sport allocated only 70,000 euros for the project. Thus, the researchers had to take many shortcuts. They apparently used "interpolated" data, together with the "real data", to build a detection model. The educational experts were also critical of the methodology and they claimed that grades and absences by themselves were not sufficient indicators to reliably detect learning problems. "You do not need AI to see that a pupil has got worse grades in the second semester. You can do that with Excel," said a speaker at the Grounded festival.

The ministry argued that they never endorsed METIS as an official tool for detecting learning and behavioral problems. However, the public debate also addressed another related issue: that a lot of school data is collected and owned by a private (commercial) contractor. And that the contractor is probably the only institution in Slovenia capable of building an efficient ADM system for educational purposes. Or to provide personalized profiles of pupils for future employers and universities that could predict their performance as workers and students.



DESPITE THESE BOLD CLAIMS AT THE UN, AI AND ADM HAVE HARDLY BEEN "AT THE FOREFRONT" OF NATIONAL DISCUSSIONS IN SLOVENIA.

# Policy, oversight and debate

## / Government

In the first edition of this report, we learned that Slovenia had not adopted any strategic document at the national level regarding the use of AI, algorithms, or ADM systems.

That has not changed. However, according to a spokesperson, the Ministry of Public Administration has initiated preliminary activities that will eventually lead to a national AI strategy. The ministry formed a cross-sector working group that was expected to present a first draft of the strategy by the end of February 2020 (the draft was still not published at the time of publication). The ministry also invited representatives from industry groups, research institutions, and NGOs to join the working group. In addition, the ministry states that it wants to establish an efficient model of support for research, development, and implementation of AI solutions by 2025.

In October, the Slovenian mission to the UN, the Council of Europe, and UNESCO hosted an event to discuss the challenges of AI at the UN headquarters in New York. The debate focused on the legal and ethical issues of existing and future uses of AI. Slovenia's Ambassador to the UN, Darja Bavdaž Kuret, said in conclusion that Slovenia would remain at the forefront of discussions on AI, [according to a report by the Slovenian Press Agency STA \(STA, 2019\).](#)

## / Civil Society, Arts, and Academia

Despite these bold claims at the UN, AI and ADM have hardly been "at the forefront" of national discussions in Slovenia. Only a few events have attracted public and media attention.

## / International AI research center

In November 2019, UNESCO's general conference decided that the first UNESCO-sponsored international center for AI will be located in Ljubljana. The Department of Intelligent Systems, at the Jožef Stefan Institute (JSI), will transform into a center that focuses on governance and policies surrounding AI. Furthermore, the International Research Centre for Artificial Intelligence (IRCAI) was also established in May 2020, said the spokesperson for the JSI.

The new facility will “aim to provide an open and transparent environment for AI research and debates on AI, providing expert support to stakeholders around the globe in drafting guidelines and action plans for AI,” [according to](#) the Slovenian Press Agency ([STA, 2019](#)). The center will advise governments, organizations, legal personnel, and the public on systemic and strategic solutions related to the introduction of AI in various fields. It will also run consultations with the wider public about the impact of AI.

The new center will retain its existing research staff and funding structure. It currently has an annual budget of [around €500,000](#) from a combination of national and EU funds ([Banovic, 2019](#)).

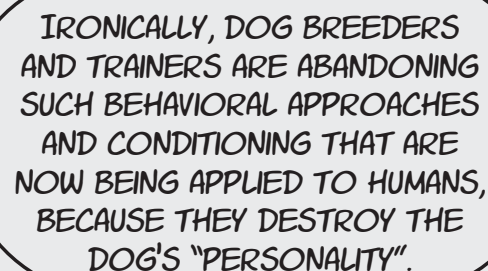
## / Grounded Festival

Automation and Power was the central theme of the [Grounded festival](#) in November 2019 ([Grounded, 2019](#)). The festival addresses current social issues by combining theoretical debates, social activism and progressive electronic club culture.

Speakers at the festival presented some examples of ADM that were mentioned in the first edition of this report. Among these examples were a system for detecting children with potential learning and educational problems and another for predicting potential tax evaders. In many cases, it was the first time that discussions, Q&A sessions, and presentations on these topics had happened in public. It was also the first time that state representatives – from the Ministry of Education, the financial administration, and other departments – had to answer some hard questions about the current and future uses of AI and ADM.

Such opportunities are important. The speakers and organizers agreed that Slovenian civil society rarely has any insight into the operation of such systems, let alone any influence on the desirability of their introduction. They believe that the political, social, economic, health, justice, and education systems can no longer be discussed without a simultaneous debate on the automation that underpins their development. Instead, automation (as a process) is shying away from democratic scrutiny due to the complexity and specialization of knowledge, high financial input, and related interests.

In addition, Slovenia is in the process of implementing a 5G infrastructure, under pressure from the EU, and without any public participation, following the



IRONICALLY, DOG BREEDERS AND TRAINERS ARE ABANDONING SUCH BEHAVIORAL APPROACHES AND CONDITIONING THAT ARE NOW BEING APPLIED TO HUMANS, BECAUSE THEY DESTROY THE DOG'S "PERSONALITY".

mandate of ever-faster and more data-rich connectivity, warned the Grounded festival organizers.

## / Brute Force Art Project

What can AI learn from dogs? This is a question of practice-oriented, artistic research aimed at developing an algorithm for artificial neural networks co-programmed by dogs.

[Brute Force](#) is one of the first art projects in Slovenia to look at aspects of AI that challenge our social and political reality ([Ars Electronica, 2019](#)).

The artist, Maja Smrekar, who started Brute Force explained (to us) that the algorithms for automating societal processes often follow behavioral doctrine once used for training dogs. Technology companies rely heavily on behavioral experiments by the late psychologist B. F. Skinner. They use social constructivism and social conditioning to influence human behavior – to either buy a product, vote for a particular political candidate, or change our mood. Ironically, dog breeders and trainers are abandoning such behavioral approaches and conditioning that are now being applied to humans, because they destroy the dog's “personality”.

Smrekar uses the analogy with dog training to unveil this behaviorist regime and its possible social impact. Can algorithms predict a dog's behavior, and how can we use such information? To learn from it, and to train them better and influence their behavior? Can algorithmic prediction be used to control and enslave humans as feared by some contemporary thinkers like Yuval Noah Harari? Such futures are possible, Maja Smrekar believes. But resistance to such conditions also requires technological understanding, which might lead to turning the system's logic against itself and to the emancipation of humans.

## Key takeaways


The absence of a national strategy and public debate on AI and ADM are the two defining factors that influence the use and implementation of IT solutions for automating Slovenian society.

There is a global trend towards automation, and Slovenia is no exception. Slovenian companies, ministries, public institutions, and local authorities are buying and testing all kinds of solutions that use AI and ADM, but nobody knows the full extent of such testing and implementation.

In addition, there are many inconsistencies. On the one hand, the politicians and state officials emphasize the role of AI and ADM for the future of society and the economy, but they are unwilling to finance research and development in this field. Furthermore, they are vocal about

humane uses of AI, social equality, and respect for human rights, but they ignore all the warnings against the increasing economic inequality and the potentially unconstitutional powers of the police to collect personal information and monitor Slovenian citizens secretly or even illegally.

Consequently, the state plays a retroactive instead of a proactive role. The government, and regulators, react when something goes wrong or catches public attention, but they fail to start solving problems that could most likely be both predicted and prevented. In addition, too much pressure is put on NGOs and privacy activists to protect the citizens from the police surveillance infrastructure because they do not have enough staff, specialized knowledge, and funding for such a task. Nor can they rely on European institutions to provide a legal framework and political pressure to the Member States that have shown repressive tendencies during the pandemic.



***THE POLITICIANS AND STATE OFFICIALS EMPHASIZE THE ROLE OF AI AND ADM FOR THE FUTURE OF SOCIETY AND THE ECONOMY, BUT THEY ARE UNWILLING TO FINANCE RESEARCH AND DEVELOPMENT IN THIS FIELD.***

# References:

Ars Electronica (2019), !brute\_force: [https://ars.electronica.art/outofthebox/en/brute\\_force/](https://ars.electronica.art/outofthebox/en/brute_force/)

Banovic, Rebecca (2019), Pioneers of Artificial Intelligence: Slovenia to Set Up International AI Research Centre with UNESCO: <https://www.forbes.com/sites/rebeccabanovic/2019/04/07/pioneers-of-artificial-intelligence-slovenia-to-set-up-international-ai-research-centre-with-unesco/#4fae1c0caea> [7.4.2019]

Cvetkovska, Saška; Zöldi Bianka, and DeliĆ, Anuška (2020): Oiling Orbán's Propaganda Machine: <https://www.ostro.si/en/stories/oiling-orbans-propaganda-machine> [28. 2. 2020]

Državljan D (2018), Sovraštvo: <http://www.sovrastvo.si>

Grounded (2019), Automation and Power: <https://www.grounded.si/festival2019/>

Kučić, Lenart Jurij (2019), Mediji SDS (3. del): madžarski prevzem desnega medijskega stebra: <https://podcrto.si/mediji-sds-3-del-madzarski-prevzem-desnega-medijskega-stebra/> [11.11.2019]

Slovenska tiskovna agencija (2019), AI challenges discussed at UN event co-hosted by Slovenia: <http://znanost.sta.si/2692287/ai-challenges-discussed-at-un-event-co-hosted-by-slovenia> [29.10.2019]

Slovenska tiskovna agencija (2019), Ljubljana seat of UNESCO-sponsored global AI research center: <http://znanost.sta.si/2701513/ljubljana-seat-of-unesco-sponsored-global-ai-research-centre> [25.11.2019]

Ministry of Public Administration (2009), E-sociala: jasen koncept in napredna tehnološka platforma – predpogoj za uspeh: [http://dsi2010.dsi-konferenca.si/upload/predstavitve/informatika%20v%20javni%20upravi/kricej\\_e\\_sociala%20jasen%20koncept%20in%20napredna%20tehnoloska%20platforma.pdf](http://dsi2010.dsi-konferenca.si/upload/predstavitve/informatika%20v%20javni%20upravi/kricej_e_sociala%20jasen%20koncept%20in%20napredna%20tehnoloska%20platforma.pdf) [10.11.2009]

Slovenska tiskovna agencija (2018), Ljubljansko letališče v pilotnem projektu za preprostejšo prijavo na let: <https://krog.sta.si/2635573/ljubljansko-letalisce-v-pilotnem-projektu-za-preprostejso-prijavo-na-let> [18.5.2019]

Rajšek, Bojan (2018), Novo mesto bo postalo pametno: <https://www.delo.si/lokalno/dolenjska/novo-mesto-bo-postalo-pametno-88495.html> [4.9.2018]



BELGIUM  
DENMARK  
ESTONIA  
FINLAND  
FRANCE  
GERMANY  
GREECE  
ITALY  
NETHERLANDS  
POLAND  
PORTUGAL  
SLOVENIA

**SPAIN**  
**STORY**

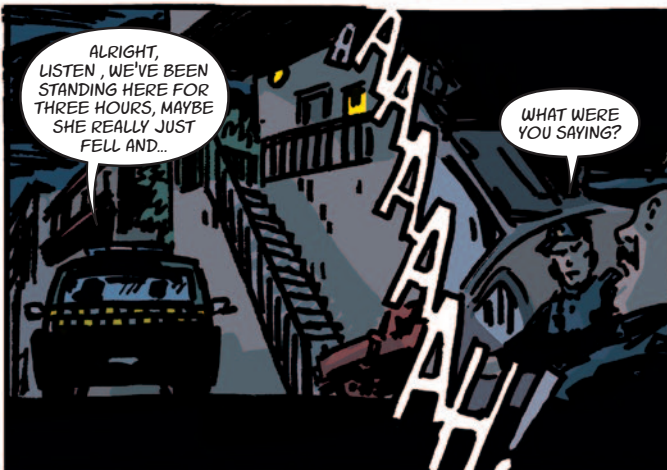
PAGE 224

**RESEARCH**

PAGE 229

SWEDEN  
SWITZERLAND  
UNITED KINGDOM





Find out more in the story on the next page.

# In Spain, the VioGén algorithm attempts to forecast gender violence

**As part of a program to curb feminicides, Spain built VioGén, an algorithm that assesses the risk faced by victims of gender violence. It remains a work in progress.**

By Michele Catanzaro

In the early morning of 24 February 2018, Itziar P., a psychologist living in the Spanish city of Castellón, went to a police station to report threats from her husband, Ricardo C..

In audio recordings she made with her mobile, the husband could be heard saying: “We will all end up dead and me in jail,” or “I will take away from you what you love most.”

According to Itziar P., Ricardo C. had also broken in pieces the buggy of their smaller child (Martina, two-years-old) and slapped the older one (Nerea, six-years-old) when both children were in his custody.

The police officer asked Itziar P. a set of questions and fed the answers into VioGén, a piece of software that helps the Spanish police estimate the risk of recidivism in gender violence. The officer issued a report in which that risk was deemed as low.

## / Critical failure

In the following days, both Itziar P. and Ricardo C. were called to declare in court. She asked that he be forbidden to visit their children, but the judge denied the request, based on the low-risk estimation made by the police, among other reasons.

Seven months later, on 25 September 2018, Nerea and Martina were sleeping at Ricardo C.’s place. In the early morning, he killed them “with cruelty” and threw himself out of a window.

Itziar P.’s story was shocking. Why was the case deemed low-risk? VioGén had failed in its role to support police personnel in assessing the risk of new assaults, and, therefore, assigning the right level of protection. Since the software was first deployed, in 2007, several “low-risk” cases have ended in the homicide of women or children.

## / Better than nothing

The VioGén program is by far the most complex of its sort in the world. It has reasonable performance indexes. Nobody believes that things would be better without it – except the far-right, who make spurious claims that it helps women report innocent men.

But critics point out some flaws. Police officers are seldom educated in gender-based violence, while others blindly rely on the outcome of the software. Moreover, the pro-

gram may systematically underestimate risk. Some victims’ organizations believe that the possibility of a low-risk score is nonsense. They say that reporting to the police is a high-risk situation in itself because abusers perceive it as a challenge.

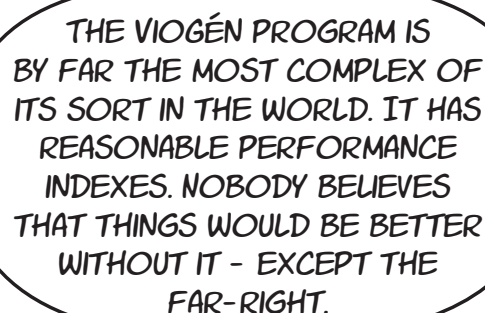
As of January 2020, 600,000 cases had gone through VioGén. About 61,000 of them were considered active, meaning they were being followed-up by the police (the system is designed to periodically check on women until they are deemed safe).

## / Reporting an assault

When a woman goes to report an assault from an intimate partner, she triggers a process that takes at least a couple of hours. First, the police officer goes through an online form with her. The officer ticks each of the items of the VPR form (from the Spanish initials of “Police Risk Assessment”) as “present” or “non-present”. There are 39 items in the latest published version of the form (the VPR<sub>4.0</sub>). Officers can also rely on police databases, witnesses, and material proofs.

Questions explore the severity of previous assaults (for example, whether weapons were ever used); the features of the aggressor (jealous, bully, sexual abuser, unemployed, drug addict, etc.); the vulnerability of the victim (pregnant, foreign, economically dependent, etc.); and aggravating factors (like assaults by other men).

Answers are thrown automatically into a mathematical formula that computes a score, measuring the risk of whether the aggressor will repeat violent actions. This quantitative approach is different from the one used in DASH, the British equivalent of VioGén. The latter is a paper check-list that helps agents to get an idea of the situation.



THE VIOGÉN PROGRAM IS BY FAR THE MOST COMPLEX OF ITS SORT IN THE WORLD. IT HAS REASONABLE PERFORMANCE INDEXES. NOBODY BELIEVES THAT THINGS WOULD BE BETTER WITHOUT IT - EXCEPT THE FAR-RIGHT.



## / Keeping the score

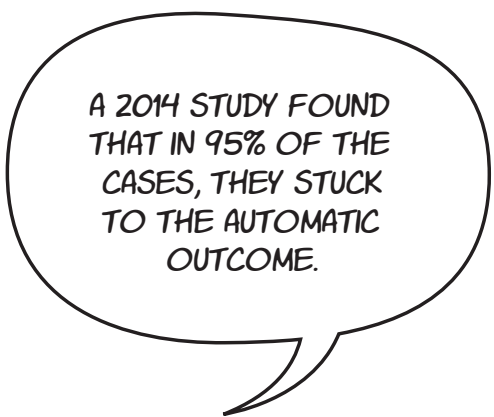
In theory, Spanish agents can increase the score manually if they believe there is a higher risk. But a 2014 study found that in 95% of the cases, they stuck to the automatic outcome.

The formula used in VioGén is a “simple algorithm”, according to Juan José López Ossorio, a psychologist who has been in charge of VioGén from its early stages, in a written statement to AlgorithmWatch. The algorithm gives more weight to items that empirical studies have shown to be more related to recidivism, Mr. López Ossorio wrote. He declined to disclose the exact formula.

Once a case's score is established, the officer decides on a package of protection measures associated with that level of risk. For the lowest scores, officers will discreetly check on the woman from time to time. For the highest, the police will give the victim an alarm button, track the aggressor's movements, or guard her house. Officers also send the forms and risk scores to the prosecutors and judges that will see the woman's case.

After the first report, the police meet again with the woman to fill in a second form, in order to assess whether the situation has worsened or improved. This follow-up happens periodically and depends on the risk level. Police only stop following-up if judicial measures are not pursued, and the risk level falls below medium.

VioGén is one of the outcomes of a pioneering law on gender-based violence that Spain approved in 2004, ten years



A 2014 STUDY FOUND  
THAT IN 95% OF THE  
CASES, THEY STUCK  
TO THE AUTOMATIC  
OUTCOME.

before the Council of Europe adopted a common framework on the subject, the Istanbul Convention. Nowadays, the software is used by the main Spanish police forces (Policía Nacional and Guardia Civil) and by hundreds of local police forces (except in Catalonia and the Basque Country that have independent police bodies).

## / The best available system

VioGén is the best device available to protect women's lives, according to Ángeles Carmona, president of the Domestic and Gender-Based Violence Observatory of the Spanish General Council of the Judiciary (CGPJ).

She recalled a case she saw in a court in Seville, of an aggressor who had a high-risk of recidivism, according to VioGén. A control wristband was attached to the man. One day, the police saw that the signal of the wristband was moving fast towards the victim's home. They broke into the home just in time to prevent him from suffocating her with a pillow.

**COMPARED WITH WHAT IS AROUND  
AND WITHIN THE EXISTING  
LIMITATIONS, VIOGÉN IS AMONG  
THE BEST THINGS AVAILABLE.**

It's impossible to know how many lives have been saved thanks to VioGén, according to Antonio Pueyo, a professor of psychology at the University of Barcelona who has advised VioGén from the beginning.

However, a 2017 study by Mr. López Ossorio and his team tried to measure how good the protocol was. They found that VioGén's [Area Under the Curve](#) (AUC), a widely-used measure of performance for predictive models, stood between 0.658 and 0.8. An AUC of 0.5 is as good as the toss of a coin, and an AUC of 1 means the model never fails. Cancer screening tests are considered good when their AUC is between 0.7 and 0.9. In other words, VioGén works.

"Compared with what is around and within the existing limitations, VioGén is among the best things available," says Juanjo Medina, a professor of quantitative criminology at the University of Manchester, who has compared instruments that assess the risk of intimate partner violence.

Spain is the only place where victims can be followed-up across different regions. Close to 30,000 police officers and other agents across the country had access to VioGén in 2018.

However, the cases that have slipped through the cracks of VioGén have raised concerns. The latest one happened in February 2020, when a 36-year-old woman, and mother of two, had her throat cut by her former partner, who then threw her body in a container in the town of Moraira. The two had been registered in the VioGén system after the police reported him for attacking her, but the case had become inactive after a judge cleared him.

## / False negatives

In 2014, the newspaper El Mundo [published](#) a leaked document from the General Council of the Judiciary that showed that 14 out of 15 women killed that year, having reported their aggressor before, had low or non-specific risk (the classification used for any person reporting a threat to the police).

Some critics say that low-risk should not even be an option. Reporting is a maximum risk moment for a woman, according to Carme Vidal Estruel, spokesperson of Tamaia, an association that helps victims in Barcelona. She says that the situation is akin to divorcing or becoming pregnant, both moments in which the aggressor realizes that he is losing grip on the victim.

Another widespread criticism is that few officers among those who should validate the computer's outcome receive enough training in gender issues. Some VioGén items are embarrassing, like those related to sexual violence, humiliation, or intimate messages on mobile phones.

Officers should ask circular questions (instead of blunt, direct questions) and avoid transmitting the feeling that the object of investigation is the woman, according to Chelo Álvarez, president of Alanna, an association of former victims in Valencia. Ms. Carmona of the General Council of the Judiciary recalls a woman who reported her husband for robbing her car keys. She was so scared that she could not say anything else. The day after, the man killed her.

Few officers are aware of these nuances. In 2017, there was a total of 654 officers in the whole of Spain attached to the Women-Children Teams (EMUME) of the Guardia Civil. That is much less than one for every police station.

## / Ignored requirements

This situation is very different from what the 2004 law that created VioGén required. According to the law, cases should be dealt with by an interdisciplinary team, including psychologists, social workers, and forensic doctors.

This team should go into psychological aspects that the VioGén form does not cover. Moreover, the team should carry out a forensic assessment of the aggressor. Critics point out that the current system evaluates how dangerous a person is, without ever talking to him. Several teams were created after the law was passed in 2004, but the process was cut back, following the 2008 financial crisis.

Mr. Pueyo, the psychology professor, acknowledges some of the criticism but believes that VioGén should be judged on its ability to predict new assaults, not homicides, because these events are very rare. The probability that a woman will be killed, after reporting, is about one in ten thousand, according to Mr. López Ossorio.

However, the Istanbul Convention seeks to reduce the risk of death, and not only of women but also of children. Overlooking the risk to children is another criticism VioGén faces.

The convention came into force in Spain in 2014, but VioGén forms were not changed accordingly until Itziar P.'s case in 2018, according to her lawyer.



## / VioGén 5.0

A new protocol was put in place in March 2019, the fifth big change VioGén has gone through since its first deployment in 2007. Now, the program identifies cases “of special relevance”, in which the danger is high, and includes cases “with minors at risk”.

This is done through the “dual evaluation procedure” of the new VPR form (VPR<sub>5.0</sub>-H), Mr. López-Ossorio explained. Two calculations are carried out in parallel: one related to recidivism and a new one related to lethal assault.

Depending on the outcome of the latter (called the “H-scale”), the risk score increases automatically. Moreover, the case can then be signaled to the prosecutors and judges as being “of special relevance”.

Mr. López-Ossorio declined to disclose how the H-scale was built, but he wrote that it was based on a four-year study his group carried out to find which factors specifically relate to cases that end up in homicides.

The new protocol seems to have triggered a major shift in the risk scores of VioGén. Passing from VPR<sub>4.0</sub> to VPR<sub>5.0</sub>-H, the number of extreme risk cases rose, and those of high risk almost doubled, according to Mr. López Ossorio.

As the president of Valencia’s former victim’s association, Ms. Álvarez puts it: “Things are improving, but they should go faster because we are being killed.”

*THE NEW  
PROTOCOL  
SEEMS  
TO HAVE  
TRIGGERED A  
MAJOR SHIFT  
IN THE RISK  
SCORES OF  
VIOGÉN.*

# Research

By Jose Miguel Calatayud

## Contextualization

Public administration bodies at the local, regional, and national levels in Spain have been using ADM for years now, and while entirely autonomous ADM systems seem to be rare – according to publicly available information – protocols in which ADM makes part of the whole decision process seem to be fairly common.

“Automated administrative actions” are regulated by law, and before implementing any action, public bodies are supposed to establish which competent authority will define its function and, if necessary, take responsibility for auditing the source code. Also, the transparency law mandates public bodies to be proactively transparent and to generally grant citizens access to any information held by the public administration.

However, in practice, public bodies proactively release very little information on the ADM systems they use, and they are also reluctant to do so when requested by citizens or organizations. This could all change depending on the outcome of an ongoing court case, in which a non-profit foundation is asking the government to publish the source code of an ADM system in the same way law texts need to be public (we discuss this case below in this chapter).

In recent years, universities have been producing academic research on the use of ADM by the public administration, which some watchdog organizations are also trying to monitor. But such ADM systems rarely feature in the mainstream media and political discourse, and there seems to be a general lack of awareness among the public of the use of ADM by public bodies and its consequences for public and individual life.

Spanish companies, big international corporations, Spanish start-ups, and universities have all been developing and providing the public administration with ADM systems.

As reported in last year's chapter and as updated in this one, ADM is most common in Spain in the surveillance and predictive policing fields, the health sector, and the analysis of social media content; and it's also been increasing its presence in the delivery of financial and social aid, and the automatization of different administrative procedures.

The fact that, at the time of writing [11 May 2020], Spain hasn't had a stable central government for more than four years (since late 2015), and that the public administration is quite decentralized (with regions and also towns and cities having a great deal of autonomy to implement their practices), means that there can be normative differences between regions, and that towns and cities have had an almost free rein to implement ADM processes at the local level, especially within the frame of the so-called Smart City platforms.

## A catalog of ADM cases

### / Distribution of financial aid

In November 2017, the office of the Secretary of State for Energy released some software, known as BOSCO, to companies providing electricity. The aim of BOSCO was to determine whether people were entitled to financial aid to help them with their electricity bills. The reasoning behind the ADM was twofold. Firstly, that it would make the process much easier for aid applicants (although, this didn't seem to be the case judging by the large number of complaints the system received) and secondly, that it would also make the process easier and more efficient for the public utility companies.

After receiving many reports that the software was not functioning properly, Civio, a Madrid-based non-profit investigative newsroom and citizen lobby, discovered that BOSCO was systematically denying aid to eligible applicants. Civio

asked the government for the BOSCO source code to identify why those errors were happening. The request passed through three different ministries before ending up at the [Committee for Transparency and Good Governance](#), which refused to share the code saying that it would violate copyright regulations (even though the software had been developed by the public administration itself).

In July 2019, [Civio filed an administrative complaint](#) arguing that the source code of any ADM system used by the public administration should be made public by default, in the same way that legal texts are made public (Civio, 2019). The case – which at the time of writing was ongoing – could end up at the Spanish Supreme Court, and could set a legal precedent.

## / Risk assessment in cases of domestic violence

Probably the one ADM system that has received the most public scrutiny in Spain is [the VioGén protocol](#), which includes an algorithm that evaluates the risk that victims of domestic violence are going to be attacked again by their partners or ex-partners.

VioGén was launched in 2007 following [a 2004 law on gender-based violence](#) that called for an integrated system to protect women. Since then, whenever a woman makes a complaint about domestic violence a police officer must give her a set of questions from a standardized form. An algorithm uses the answers to assess the risk that the women will be attacked again. These range from: no risk observed, to low, medium, high or extreme risk. If, later on, an officer in charge of the case thinks a new assessment is needed, VioGén includes a second set of questions and a different form, which can be used to follow up on the case and which the algorithm uses to produce an updated assessment of the level of risk. The idea was that the VioGén protocol would help police officers all over Spain produce consistent and standardized evaluations of the risks associated with domestic violence, and that all the cases that are denounced would benefit from a more structured response by the authorities, including follow-up assessments when needed.

After the protocol has been followed, the officer can override the algorithm's response and decide to give the case a higher level of risk. Each of the lev-

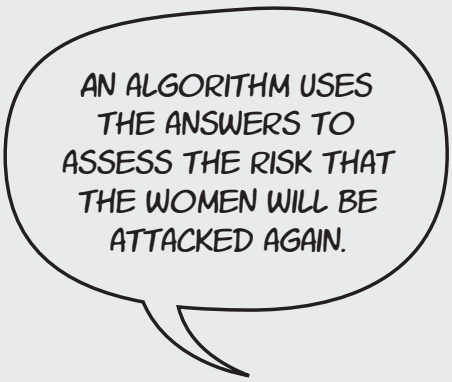
els implies different compulsory protection measures, and only the “extreme” level of risk asks for permanent protection. However, according to official documents about VioGén, the police may decide to grant additional protection measures apart from those established by the protocol.

The set of questions and the algorithms were developed in collaboration between police and academic experts. They have been updated five times since their first implementation, as officers and experts on domestic violence learned from its application.

There is a fair amount of public information available about the whole VioGén protocol, and [a book, published in September 2018 by the Ministry of the Interior](#), provides a candid (by the usual standards of available information on ADM systems) and very informative account of the protocol's history, design, and implementation (Spanish Ministry of the Interior, 2018). However, the book stops short of revealing the algorithms' code or inner workings to calibrate and measure the weight given to the different factors and their interrelations.

The sensitivity of the subject of domestic violence and the fact that a fair amount of information about the whole protocol (if not about the algorithm itself) is available have contributed to making VioGén [quite visible in the mainstream media \(Precedo, 2016\)](#). That's especially the case when official figures about victims of domestic violence are released, and they include references to cases not deemed high or extreme risk by VioGén and [which nevertheless ended up in the aggressor killing the woman \(Álvarez, 2014\)](#).

However, a careful revision of [the available figures](#) shows that these are published inconsistently, in a way that aggregates together the initial and the follow-up risk assessments of different cases (with some reports describing the number of assessments as the total number of individual cases), and that on some occasions several years had gone by between an initial assessment of low risk and the aggressor finally attacking again and this time killing the woman after having been convicted and served his prison sentence. All of which makes it hard to extract coherent conclusions about any possible correlations between the cases calculated to be of low risk and those in which the aggressor ended up killing the woman.



AN ALGORITHM USES THE ANSWERS TO ASSESS THE RISK THAT THE WOMEN WILL BE ATTACKED AGAIN.

When discussing VioGén, most of the experts on sexual and gender-based violence quoted in the press complain that although [police officers follow the protocol, they often haven't been properly trained in how to deal with cases of domestic abuse](#) (Precedo, 2016). The book which was published by the Ministry of the Interior quotes a 2014 study stating that officers who followed the protocol didn't change the level of risk given by the algorithm in 95% of the cases. This would seem to support the argument that the code should be made public so that it can be properly audited.

## / Detecting fraud in public procurement

In October 2018, the regional Valencian parliament passed a law [that included the use of ADM to detect possible cases of fraud in public procurement](#) (Europa Press, 2018). The parliamentary majority in government at the time was made up of a coalition of progressive parties that came to power in 2015. This coalition came about after 20 years of rule by the Conservative People's Party, which in the end became marred in a series of corruption scandals regarding, among other things, public procurement.

The ADM system, known as Saler, [was funded by the Valencian government](#) and the human rights organization FIBGAR, and developed by the [Polytechnic University of Valencia](#).

Saler cross-checks data from various Valencian public administration databases and [the official gazette of the Commercial Registry](#). The purpose is to raise red flags when it finds suspicious behavior (for example, a high number of small contracts given to the same company). Saler then automatically sends the information to the corresponding authority: the Anti-fraud Agency, the Public Prosecutor's Office or the Court of Audits.

The idea, yet to be proven, is that Saler will become more effective in raising relevant red flags as it's fed more databases and as it learns by being used by public servants. Its creators would like to export it to other regional administrations in Spain, and to the national central administration too, but the fact that each region uses different databases and indexes makes this quite a complicated process for the time being.

The public launch of this ADM by the Valencian authorities got [a fair amount of press coverage](#) by usual standards, probably because it dealt with corruption and that its original acronym was Satan (Cid, 2018).

## / Video surveillance and (quasi-) face recognition

In November 2019, [Ifema](#), a state-owned consortium, [started installing face recognition surveillance cameras](#) in the big congress facilities it runs in Madrid (Peinado, 2019). Ifema specifically stated in [the public procurement information from March 2019](#) that it was looking for "technical improvements in face recognition licenses".

It should be noted that these cameras weren't still in use by the time of the [UN Climate Change Conference COP25](#), hosted at Ifema between 2 and 13 December 2019, according to a representative of the consortium [quoted in the press](#) (Peinado, 2019).

In late 2018, the city of Marbella, a tourist hub on the southern coast of Spain, [started using a video surveillance system with, reportedly, the highest definition in Spain](#), and developed by the American firm Avigilon (Pérez Colomé, 2019). Only the border between the Spanish enclave of Ceuta and Morocco and some football stadiums use similar cameras, according to press reports. Officially, the system doesn't have face recognition features, but the software used by these cameras does use "appearance search" and "face analytics". Apparently this allows the system to search people by how they look by identifying a series of unique facial features, clothes, age, gender, and hair color, [according to Avigilon's PR](#).

A company representative [quoted in the press](#) said their face recognition software wasn't being used by public authorities in Spain, but he added that the company had indeed installed face recognition software in private surveillance systems in Spain (Pérez Colomé, 2019).

In both the Ifema and Marbella cases – in which the systems are justified as they are said to increase security while also increasing the efficiency of surveillance – it's not clear how exactly the recognition software works and what kinds of checks and other measures may be in place to protect people's biometric information and other personal data.

## / Predictive policing

In 2016, the local police of Rivas-Vaciamadrid, a town of 86,000 people in the Madrid region, [ran a pilot of the Pred-Crime software](#), developed by Spanish company EuroCop (Europa Press, 2015). [Pred-Crime](#) analyzes historical data to predict where and when it's more likely that different types

of common misdemeanors and offenses, like traffic violations and robberies, will be committed.

Reportedly, the plan was to fully implement this software during 2016, but after testing it for nine months the municipality decided not to go on using it. “It needs to keep on being developed to become properly efficient”, a representative of the municipality is quoted as saying [in a May 2019 press report \(García, 2019\)](#).

EuroCop says on its website [that it has dozens of municipalities from all over Spain among its customers](#), but it doesn't specify whether those local authorities are using its predictive software or any of the other non-predictive tools the company also markets.

In another case of predictive policing, the Spanish police have reportedly been using software that analyzes the available data about the victim of a killing and the context of the crime, and then produces the likely profile of the killer. Between 2018 and 2019, agents of the office of the Secretary of State for Security have collaborated with the police by using such software in at least five investiga-

tions, [according to a press report \(Pérez Colomé, 2019b\)](#). No more details are known about this case.

According to the same press article, in mid-2019, the Environmental Ministry signed an agreement with the Autonomous University of Barcelona to develop software that can predict the most likely profile of pyromaniacs.

## / Automated analysis of social media content

In January 2018, the regional Government of Catalonia started a pilot project to try and measure the impact among the public of its STEMcat's initiatives, a plan aimed at promoting scientific and technological vocations among young people.

For a month and a half, the Catalan government used [Citi-beats](#). This text-analysis software used machine-learning algorithms to collect and analyze around 12,000 tweets that spoke about the STEM disciplines in Catalonia. One of the insights the authorities said that they had gained was that women were more responsive to messages about natural sciences than about technology, [according to the Catalan government's PR](#). The authorities then used that and other insights to “optimize their strategy and propose new initiatives” to make young people interested in the STEM disciplines. The project was part of [SmartCAT](#), the Catalan government's strategy to become a smart region (as noted in this chapter, regional governments in Spain have a high degree of autonomy to develop their policies). The SmartCAT director said the software had allowed them “to evaluate in a more objective way the impact of the (government's) initiatives” to make people interested in science and technology.

Citibeats, [developed by Social Coin](#), a Barcelona-based start-up, was also used in December 2017 by the Barcelona municipality to gather people's attitudes towards public transport and mobility in the city by analyzing around 30,000 comments by more than 15,000 people. [In a case study of this project](#), Citibeats spoke of “citizens as sensors”

In the two cases described above, and while the authorities praise the software's ability to gather and analyze thousands of online comments (something it would take much more time and money to do using traditional survey methods), it's not clear how representative those samples are and how valid the conclusions might be; and there doesn't seem to be any information on how those analyses then influenced public policy.

**IT NEEDS  
TO KEEP  
ON BEING  
DEVELOPED  
TO BECOME  
PROPERLY  
EFFICIENT.**



Since March 2019, [the regional government of Navarra has also been using the Citibeats software](#) to detect online hate speech by analyzing text published on Facebook, Twitter and Instagram.

As with other cases dealing with personal data, it's not clear how the software works and what oversight mechanisms the public authority may have in place when using it.

## **/ Automated evaluation of offers for public procurement through e-tendering**

[The current law governing public procurement in Spain](#), which was passed in November 2017 to adhere to EU regulations, allows public authorities to run digital bidding processes using electronic devices that can sort the different offers "through automated evaluation methods".

This kind of digital bidding can only be used when the " requisite specifications of the contract to be awarded can be established in a precise way" and the "public services that are its object are not of an intellectual type, as in engineering, consultancy and architecture". Contracts to do with food quality can't be awarded through digital bidding either.

According to the law, before using this method a public authority needs to make public which "electronic device" it is going to use in the bidding process.

Before running the digital bidding itself, the authority needs to carry out a complete evaluation of all the offers and then send a simultaneous electronic invitation to all the eligible bidders. When such an invitation is sent, the authority has to include "the mathematical formula" that will be used in the automated classification of the eligible offers.

## **/ Automated assistance in tax filing**

Since July 2017, the Tax Authority in Spain has been using the IBM Watson software [to provide automated assistance regarding a particular aspect of VAT filing](#) which mostly affects big companies. In public communications about it, both [IBM and the Tax Authority highlighted](#) that the software can work 24/7 and, as such, was freeing public servants from having to deal with a large number of emails from people trying to do their VAT filing. According to their PR, between July 2017 and February 2018, the number of emails to civil servants about VAT issues decreased from

900 to 165 per week. And, reportedly, the automated assistant went from receiving around 200 questions per week, when it was launched in July 2017 [to around 2,000 by November 2017 \(Computing, 2018\)](#).

## **/ Smart Cities: from smart waste to smart tourism**

As cities aim to start using big data and automated processes to become "smart", a series of ADM systems have been embedded at the municipal level in Spain. These systems usually support decision-making rather than working entirely autonomously.

[One such example is the Smart Waste platform](#), which collects data from sensors installed in bins and trucks, and also from social media, surveys, the census, and satellite information. The combined data helps local authorities decide what services will be needed when and where.

Smart Waste was developed by [The Circular Lab](#) (from the innovation center of [Ecoembes](#)) which is a non-profit organization charged with collecting plastic packaging, cans, cartons, paper, and cardboard packaging for recycling. [Minisat](#), a division of [Indra](#), a Spanish multinational transport, defense, and security technology consultancy company also helped develop the platform.

During 2018, the Logroño municipality and the La Rioja and Cantabria regional governments first ran the platform as a pilot and today it is available to local and regional authorities all over Spain.

As is the norm in almost every case in Spain, there does not seem to be any available information on how the software works, what kind of output it produces, and what decisions or changes the different authorities have adopted due to its use.

In another example of the involvement of local governments with ADM systems, the Barcelona municipality and the [Sagrada Familia basilica](#), one of the city's tourist attractions, partnered with the private company Bismart to develop a mobile application, [called Smart Destination](#). This application analyzes publicly available data about people (including social media posts and the number of people queuing to enter a tourist attraction), vehicles and traffic, the weather, bar and hotel occupancy, etc., and also the preferences given by the user to automatically generate personalized tourist plans and routes.

# AS IS THE NORM IN ALMOST EVERY CASE IN SPAIN, THERE DOES NOT SEEM TO BE ANY AVAILABLE INFORMATION ON HOW THE SOFTWARE WORKS.

Once again, it is not known exactly how the system works and how it comes up with the output it gives to the public.

## Policy, oversight and debate

### / Government

#### / The delayed Spanish National AI Plan and the National Strategy for AI

In November 2017, the Ministry of Industry and Trade presented a [Group of Experts on AI and Big Data](#), formed by individuals from academia and the private sector (representing entities like [BBVA](#) – the second biggest bank in Spain – , [Telefónica](#) – one of the largest phone operators and network providers in the world – , and [the Vodafone Institute](#) – which the British company describes as its European think-tank) and tasked it with writing a White paper on AI to be published in July 2018. That paper was expected to inform the elaboration of a code of ethics about the use of data in public administration. It was to be the first step towards a Spanish National AI Plan, but nothing seems to be publicly known about it as of the time of writing (Full disclaimer: when it was created, this group of experts included among its members Lorena Jaume-Palasi, co-founder of Algorithm Watch who, at the time, was still involved with the organization.)

As a prelude to that plan, in March 2019, the Ministry of Science, Innovation and Universities published a [Spanish AI National Strategy](#). It was released shortly before the April 2019 general elections, and it was criticized as vague and rushed [by the specialized press \(Merino, 2019\)](#). Indeed, the strategy reads like a mere overview of the possibilities and dangers of AI for the state and the public administration.

The document welcomes the use of AI systems by the public administration, which the strategy assumes will increase efficiency (for instance by advancing towards interoperability between public bodies and to generate “automated administrative procedures”) and save the state huge amounts of money (“billions of euros” alone in the health sector, the document says, referencing [a study from June 2017 by the global consultancy company PricewaterhouseCoopers](#)) (PwC, 2017). The strategy also says that AI-based threats to security require AI-based solutions, and it advocates for transparency in “algorithms and models” and wants to see an “honest use of the technology”, without describing any precise plans or norms to explain how it will advance in that direction.

On 18 November 2019, [in the closing act of the II International Congress on AI](#), the acting Science Minister, Pedro Duque, [said that he expected the Spanish AI National Plan to be ready in the following weeks](#) (Fiter, 2019). However, at the time of writing there has not been any other news about it.

In general, Spain hasn’t had a stable government since 2015 – in fact, on 10 November 2019 [Spain held its fourth](#)

[general election in four years](#) (El País, 2019). In addition, at the time of writing, a minority government established on 13 January has been struggling to contain one of the worst COVID-19 outbreaks in the world, and has probably contributed to those delays and the lack of a more comprehensive and decisive political leadership in the development of strategies and plans concerning the use of AI systems by the public administration.

## / Oversight

### / The law regulating public administration

In Spain, [the use of ADM processes by the public administration falls under the law 40/2015](#). This text defines “automated administrative action” as “any act or action entirely performed through electronic means by a public administration body as part of an administrative procedure and with no direct intervention by any employee”. And the text also states that before any automated administrative action is taken it must be established which of the competent authorities are “to define the specifications, programming, maintenance, supervision and quality control and, if applicable, the audit of the information system and its source code”. It must also be known which body will be responsible in case of any legal challenge to the automated action.

### / Transparency and access to information

Besides that, [the 2013 law regulating citizens' access to public information](#) mandates that public bodies have to be proactively transparent and generally grant citizens access to any content and documents held by the public administration.

Yet, in practice, public bodies rarely publish detailed information on the ADM systems they use. When citizens or organizations request information about them, the public administration is usually reluctant to grant it, which in effect means their use of ADM processes happens in a context of opacity. The ongoing court case launched by non-profit Civio, commented on earlier in this report, illustrates that point.

Since 2014, the central public administration has been making [Open Government Plans](#), which outline its commitments to advance towards a transparent government and administration. The current one, [the III Open Government Plan 2017-19](#), doesn't include any actions explicitly related

to ADM systems. [The IV Open Government Plan 2019-21](#) is, at the time of writing, being developed by an interministerial working group.

## / Data protection law

When it comes to people's data protection, [Spanish law 3/2018](#) (which was passed after the [GDPR](#) was approved) gives citizens the right to consent to the collection and processing of their data by the authorities. In principle, citizens then also have the right to request and access information about which body has their data and what it intends to do with it. The text explicitly mentions that the law applies to “any completely or partially automated processing of personal data”. It also says that when personal data is used “to create profiles” then the people affected “must be informed of their right to oppose automated individual decisions that have legal effect over them or that significantly affect them in a similar way”.

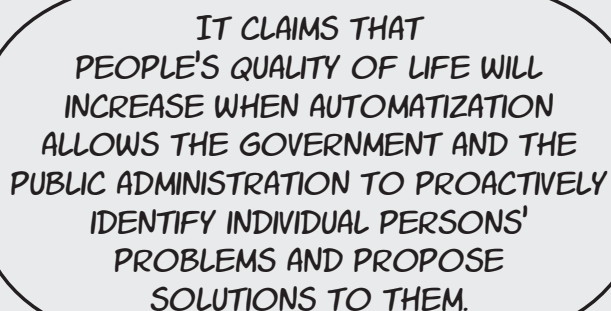
## / Towards a Digital Rights Charter for Spain

On 16 June 2020, the Spanish government announced it had kicked off the process to develop a Digital Rights Charter for Spain by naming a group of experts who will make suggestions. The aim of the Charter is to add new specific ‘digital rights’ to those included already in Spanish law. This new Charter should include rights relating to “protecting vulnerable groups, new labour conditions, and the impact of new technologies like AI”, among other issues, [according to the government's press release](#). The announcement said at a later stage the process would be open to the public too, and in the end the government will write up the Charter taking into consideration the expert's and the public's inputs. At the time of writing, there was no information on the calendar of the whole process or any other details about the scope and content of the Charter.

## / Political decision-making and oversight

In terms of political decision-making, [the office of the Secretary of State for Digital Advancement](#), which depends on the [Ministry of Economy and Business](#), is charged with coordinating the different strategies, plans, and actions for the “connectivity and digital transformation of Spain”, and includes a section focusing on AI.

Only one of the 11 specific plans promoted by that office, [the one devoted to “smart territories”, which was](#)



IT CLAIMS THAT PEOPLE'S QUALITY OF LIFE WILL INCREASE WHEN AUTOMATIZATION ALLOWS THE GOVERNMENT AND THE PUBLIC ADMINISTRATION TO PROACTIVELY IDENTIFY INDIVIDUAL PERSONS' PROBLEMS AND PROPOSE SOLUTIONS TO THEM.

published in December 2017, mentions the use of ADM systems. It claims that people's quality of life will increase when automatization allows the government and the public administration to proactively identify individual persons' problems and propose solutions to them. Regarding data collection, the plan states that "the data collection and processing mechanisms must be defined, (aiming to) limit the amount and type of data to collect, control their uses, and facilitating access to them and to the logic of the algorithms on which future decisions will be based".

The body charged with "executing and deploying the plans of the Digital Agenda for Spain" is [Red.es](#), a public corporate entity that depends on the office of the Secretary of State for Digital Advancement. Red.es includes a [National Observatory for Telecommunications and the Information Society](#), which, as of the time of writing [11 May 2020], doesn't seem to have produced work explicitly on ADM systems.

Spain is a decentralized state, with each of the 17 regions having a high degree of autonomy over their public administrations, which can be different from one another. In the Spanish central public administration, the top governance body regarding the use of IT is the [Commission of IT Strategy](#), in which all ministries are represented and which takes on an intermediary role between the executive power, embodied by the Council of Ministers, and the public administration itself. [The 2014 legal text that regulates this Commission](#) states that digital transformation should mean "not only the automatization of services [offered by the public administration] but their complete redesign", but it doesn't go into any other details regarding ADM.

## / Civil Society and Academia

### / Academic research and debate

In recent years, there has been a lively debate in academia regarding the use of ADM systems in the public administration (as well as in the private sector), and there are several academic researchers specifically looking into ADM in the public administration. However, few of these debates seem to make it into the mainstream public discourse and to the general public.

In April 2019, the [Network of Specialists in IT Law](#) noted the "striking lack of algorithmic transparency and the absence of the appropriate awareness by the public administration of the necessity to approve a specific legal frame. We only see some concerns over compliance on data protection, which is perceived as a limitation". [The document](#) reads like a fair summary of the situation in Spain, and calls on the authorities to "launch new mechanisms to guarantee [people's] fundamental rights by design and by default". The conclusions also point to the topic of smart cities as one area in need of urgent attention, as municipalities are freely experimenting with ADM systems without a proper debate or regulations (Cotino et al., 2019).

Some other notable academic research groups working in this field are [Idertec at the University of Murcia](#), and the [Research Group on Web Science and Social Computing at the Universitat Pompeu Fabra in Barcelona](#), which has a research area devoted to algorithmic fairness and transparency.

### / Non-profit watchdogs

In recent years, several private organizations have been doing research and advocacy that includes ADM. In Barcelona, there is the [Eticas Foundation](#), a non-profit associated with the private consultancy company, Eticas Consulting, which researches issues that include algorithmic discrimination, law and practice, migration and biometrics, and policing and security. Its founder and director, Gemma Galdón, has published several op-eds on these questions and is often quoted in the mainstream press. But, according to its website, the foundation itself doesn't seem to have been very active recently.

In September 2019, a new association called [OdiselA](#) launched in Madrid, describing itself as an independent "observatory of the social and ethical impact of AI". Odis-

eIA was founded by 10 people from academia, the public administration, and the private sector, some of whom are well-known in expert circles regarding ADM in Spain.

## **/ ADM in the press**

When it comes to the general public and mainstream political and media discourse, there seems to be a general lack of awareness about the use of ADM systems in public administration, the opportunities, challenges, and risks posed by such systems, and the implications for public and individual life. This is something about which all experts consulted have remarked upon and complained about.

While the specialized press has had a more consistent critical approach regarding ADM systems in the public administration, until recently, most of the mainstream press

coverage of ADM seemed to be based on the PR materials released by the companies developing the software and by the public bodies using it. However, and maybe driven by the bad image that the word “algorithm” has been getting in connection with scandals at Facebook and Google, more recently the mainstream press has been taking a more critical approach. Having said that, the majority of articles and items in the general press still seem to be about individual ADM systems that gain visibility for a particular reason (like VioGén, discussed above, which deals with the sensitive issue of domestic violence) rather than about the use of ADM by the public administration, its normative frame and its implications for public and individual life as a subject on its own. Often, when the media talk about the lack of transparency and accountability regarding ADM systems in the public administration, it is on an individual case-by-case basis rather than as a structural issue.

***THERE SEEMS TO BE A GENERAL  
LACK OF AWARENESS ABOUT THE  
USE OF ADM SYSTEMS IN PUBLIC  
ADMINISTRATION, THE OPPORTUNITIES,  
CHALLENGES, AND RISKS POSED BY  
SUCH SYSTEMS.***



## Key takeaways

Even though they don't explicitly mention ADM systems as such, the present laws governing public administration in Spain (which include a definition for "automated administrative actions") and citizens' access to public information do offer a good opportunity to call the authorities to account regarding the use of ADM in public administration.

Such current normative frames will be put to the test by the ongoing court case in which a non-profit foundation is asking the government to release the source code of an ADM system in the same way that legal texts are made public.

There are also good practices, like the proactive way in which the authorities have been publishing information concerning the VioGén protocol, which could serve as an example for other authorities to follow. Even though more

transparency – including ways to audit the functioning of the ADM systems themselves – is welcome, a greater explanatory effort by the authorities is also needed.

The number of active academic researchers looking into the field of ADM in public administration (and also in the private sector) and the increasingly comprehensive coverage of the issue by the mainstream press offer another opportunity for the general public to become more aware of how ADM systems are used to deliver public services, and the consequences for public and individual life of such systems.

On the other hand, the normalization of the use of ADM by the public administration without a proper public debate or putting clear oversight mechanisms in place, especially as low-key ADM systems spread at the local level among municipalities, risks creating an opaque administrative environment in which citizens may end up lacking the means to hold the authorities accountable about such systems.



# References:

Álvarez, Rafael J. (2014): Las asesinadas que denunciaron se valoraron como 'riesgo bajo o nulo', [online] <https://www.elmundo.es/espana/2014/12/09/54861553ca4741734b8b457e.html> [11.01.2020]

Cid, Guillermo (2018): Ingenieros valencianos crean 'Satan', un 'software' para cazar corruptos: así funciona, [online] [https://www.elconfidencial.com/tecnologia/2018-10-22/algorithmo-anticorrupcion-valencia-satan\\_1632428/](https://www.elconfidencial.com/tecnologia/2018-10-22/algorithmo-anticorrupcion-valencia-satan_1632428/) [11.01.2020]

Civio (2019): "Being ruled through secret source code or algorithms should never be allowed in a democratic country under the rule of law", [online] <https://civio.es/novedades/2019/07/12/being-ruled-through-secret-source-code-or-algorithms-should-never-be-allowed-in-a-social-and-democratic-state-under-the-rule-of-law> [11.01.2020]

Computing (2018): IBM Watson ficha por la Agencia Tributaria para la gestión del IVA, [online] <https://www.computing.es/analytics/noticias/1103993046201/ibm-watson-ficha-agencia-tributaria-gestion-del-iva.1.html> [11.01.2020]

Cotino, Lorenzo et al. (2019): Conclusiones del I Seminario Internacional Derecho Administrativo e Inteligencia Artificial, [online as a PDF file] [https://www.dropbox.com/s/5px5jkkvauiz06vu/CONCLUSIONES\\_DAIaVfinal.pdf](https://www.dropbox.com/s/5px5jkkvauiz06vu/CONCLUSIONES_DAIaVfinal.pdf) [11.01.2020]

El País (2019): Socialists win repeat Spanish election, Vox becomes third-biggest force in Congress, [online] [https://elpais.com/elpais/2019/11/10/inenglish/1573407794\\_574125.html](https://elpais.com/elpais/2019/11/10/inenglish/1573407794_574125.html) [11.01.2020]

Europa Press (2015): La Policía Local usará un 'software' que facilita la prevención de delitos, [online] <https://www.elmundo.es/tecnologia/2015/12/10/566991ee268e3ee63c8b4634.html> [11.01.2020]

Europa Press (2018): Les Corts aprueban la creación de 'Satan', un sistema de alertas contra la corrupción, [online] <https://www.europapress.es/comunitat-valenciana/noticia-les-corts-aprueban-creacion-satan-sistema-alertas-luchar-contra-corrupcion-20181017150930.html> [11.01.2020]

Fiter, Miguel (2019): El Gobierno reformará las universidades para hacerlas más flexibles, [online] <https://www.elindependiente.com/politica/2019/11/18/el-gobierno-reformara-las-universidades-para-hacerlas-mas-flexibles/> [11.01.2020]

García, Ter (2019): Minority Report en las policías europeas: llegan los sistemas de predicción de delitos, [online] <https://www.elsaltodiario.com/tecnologia/minority-report-policias-europa-sistemas-algoritmos-prediccion-delitos> [11.01.2020]

Merino, Marcos (2019): La 'Estrategia para la Inteligencia Artificial en I+D' del Gobierno, poco más que un teaser de la Futura Estrategia Nacional de IA, [online] <https://www.xataka.com/inteligencia-artificial/estrategia-para-ia-i-d-i-presentada-gobierno-poco-que-teaser-futura-estrategia-nacional-ia> [11.01.2020]

Peinado, Fernando (2019): Las cámaras que leen la cara se extienden por Madrid, [online] [https://elpais.com/ccaa/2019/11/26/madrid/1574801864\\_377093.html](https://elpais.com/ccaa/2019/11/26/madrid/1574801864_377093.html) [11.01.2020]

Pérez Colomé, Jordi (2019): Marbella, el mayor laboratorio de videovigilancia de España, [online] [https://elpais.com/tecnologia/2019/11/21/actualidad/1574348695\\_231540.html](https://elpais.com/tecnologia/2019/11/21/actualidad/1574348695_231540.html) [11.01.2020]

Pérez Colomé, Jordi (2019b): Así trabaja el equipo de Interior que predice el perfil de los asesinos, [online] [https://elpais.com/tecnologia/2019/06/26/actualidad/1561500733\\_085160.html](https://elpais.com/tecnologia/2019/06/26/actualidad/1561500733_085160.html) [11.01.2020]

Precedo, José (2016): Las primeras medidas de protección a una mujer maltratada las decide un algoritmo informático, [online] [https://www.eldiario.es/sociedad/primeras-proteccion-maltratada-algoritmo-informatico\\_0\\_561344191.html](https://www.eldiario.es/sociedad/primeras-proteccion-maltratada-algoritmo-informatico_0_561344191.html) [11.01.2020]

PwC (2017): Sherlock in Health. How artificial intelligence may improve quality and efficiency, whilst reducing healthcare costs in Europe, [online as a PDF file] <https://www.pwc.de/de/gesundheitswesen-und-pharmastudie-sherlock-in-health.pdf> [11.01.2020]

Spanish Ministry of the Interior (2018): La valoración policial del riesgo de violencia contra la mujer en España, [online as a PDF file] <http://www.interior.gob.es/documents/642012/8791743/Libro+Violencia+de+G%C3%A9nero/19523de8-df2b-45f8-80c0-59e3614a9bef> [11.01.2020]

BELGIUM  
DENMARK  
ESTONIA  
FINLAND  
FRANCE  
GERMANY  
GREECE  
ITALY  
NETHERLANDS  
POLAND  
PORTUGAL  
SLOVENIA  
SPAIN

**SWEDEN**

**STORY**

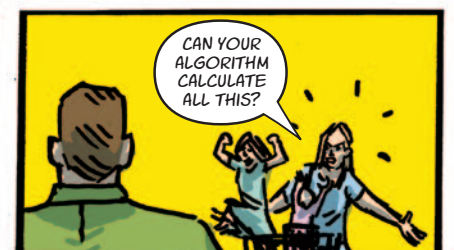
PAGE 242

**RESEARCH**

PAGE 246

SWITZERLAND  
UNITED KINGDOM





Find out more in the story on the next page.

# Central authorities slow to react as Sweden's cities embrace automation of **welfare management**

**Trelleborg is Sweden's front-runner in automating welfare distribution. An analysis of the system's source code brought little transparency – but revealed that the personal data of hundreds was wrongly made public.**

By Katarina Lind & Leo Wallentin



Trelleborg is a city of 40,000 in Sweden's far south. Three years ago, it became the first municipality to introduce fully automated decision-making in its social services. They named their robot *Ernst* and introduced it as a digital co-worker.

Sweden's social services are governed at the local authority level. The 1992 Local Government Act gave decision-making powers to municipal committees, but this right can be further delegated to an employee. With the exception of Trelleborg and their lawyers, all other instances forbid the delegation of decision-making to an automated system, and, therefore, made automated decision-making illegal.

The same does not apply to state agencies. In 2018, automated decisions were allowed after a change in the Administrative Procedure Act (*Förvaltningslagen*), the law that regulates governmental agencies. Welfare payments, such as parental benefits and dental care subsidies, are now allocated without any human intervention.

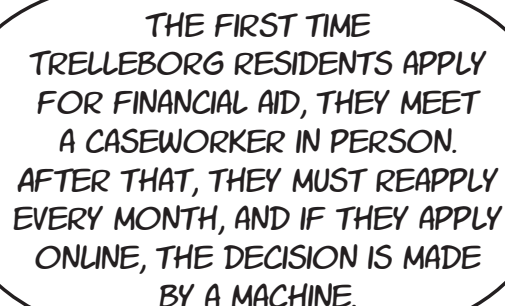
## / Full automation

Trelleborg uses a process called robotic automation, or RPA, to handle applications for financial aid. The software is based on different rules that lead to a yes or no decision.

The first time Trelleborg residents apply for financial aid, they meet a caseworker in person. After that, they must reapply every month, and if they apply online, the decision is made by a machine. They fill in details on their income and expenses, which the RPA compares with the previous month. It also pulls information such as tax and income statements and student loans from a database that gathers personal data from seven agencies for control purposes. A decision is then made based on these data points.

Should the applicant's situation significantly change from one month to the next, the software stops and forwards the application to a human caseworker. Around one in three reapplications are currently handled by the software. The rest is treated by caseworkers, because of circumstances the software cannot handle.

Because every beneficiary meets a caseworker the first time they apply – and new circumstances are checked by a human being – an individual assessment is always made, Ms. Schlyter, of Trelleborg municipality, said.



THE FIRST TIME  
TRELLEBORG RESIDENTS APPLY  
FOR FINANCIAL AID, THEY MEET  
A CASEWORKER IN PERSON.  
AFTER THAT, THEY MUST REAPPLY  
EVERY MONTH, AND IF THEY APPLY  
ONLINE, THE DECISION IS MADE  
BY A MACHINE.

## / Saving time

The main reason for deploying RPA was to save time and re-allocate resources to meet beneficiaries in person instead of handling documents, according to Ms. Schlyter. It also shortens the time beneficiaries are notified of a decision, as decisions that previously could have taken two days can now be reached in less than a minute.

The introduction of the RPA and the relocation of staff also led to lower payments for the municipality, Ms. Schlyter said.

In recent years, many towns started using online applications for welfare distribution, a first step towards automating the process. A report by the Board of Health and Welfare (*Socialstyrelsen*), a national authority, showed that the number of municipalities that introduced online applications for welfare more than tripled over the last three years, from 9% in 2017 to 29% in 2019.

Another report, published in November 2019 by the Swedish Association of Local Authorities and Regions (SKR), showed that the trend continued upwards, with 36% of municipalities saying that they used online application systems.

However, few municipalities use automated processes. The SKR survey found that 8% of the municipalities used some form of automation and only one (Trelleborg) used it for decision-making. Things may change rapidly, as 40% of the municipalities said they were planning to automate administrative work over the next few years.

## / Redefining social work

Most of the automated tasks, such as handling invoices, are uncontroversial. These programs are not especially “smart”: they are quite simple rule-based algorithms. But the introduction of automated decision-making into the welfare system sparked a discussion about the profession of social work and what social assistance should be.

“Financial aid is society’s safety net, and has to be assessed individually by a professional social worker. When you replace these professionals with software, many social workers feel it is a threat to their profession,” said Lupita Svensson, a researcher at Lund University’s School of Social Work.

Ms. Svensson recently wrote a report about automating the welfare sector (*Technology is the easy part*, published in November 2019). She said that, over the last 20 years, decisions about financial aid moved away from individual assessments and towards more general, rule-based decisions.

**THE MUNICIPAL  
SECTOR HAS A  
NAIVE VIEW OF  
TECHNOLOGI-  
CAL ADVANCES.  
THEY THINK A  
“ROBOT” WILL  
BE IMPARTIAL  
AND OBJECTIVE.**

“Initially, the legal text about financial aid gave social workers a great deal of room to manoeuvre, since the law was saying that you couldn’t generalise. When this law is converted to code, it becomes clear that social work has changed. By converting law to software, the nature of financial aid changes, as you can’t maintain the same individual assessments as before.”

Ms. Svensson is also concerned by the idea that an algorithm could be impartial.

“The municipal sector has a naive view of technological advances. They think a “robot” will be impartial and objective. But how were these robots constructed? When I asked municipalities about this, they told me they followed the social workers’ processes. This means there’s a risk of copying in the norms, ideas and values that are already present in the system. There’s very little critical discussion of this.”

## / Mass resignation

When Kungsbacka, a town of 20,000 inhabitants, 300 kilometers north of Trelleborg, introduced the “Trelleborg model”, as it became known, in 2018, 12 of 16 social workers left their work in protest. Some of them have returned to their jobs, but the majority left for good.

Inger Grahn, a local representative of the Union for Professionals in Kungsbacka, said that the protest was about two things. Firstly, the “Trelleborg model”, or at least its automated component, might not be legal. (As of early 2020, Kungsbacka has not implemented full automation.)

Secondly, implementing the Trelleborg model requires a major reorganization of municipal services. It shifts responsibility for financial aid from the department of social services to the department of work.

Kungsbacka’s caseworkers said that this model might prevent them from getting the whole picture of a beneficiary. By focusing on getting beneficiaries directly into work, social issues such as children’s welfare could be missed.

Technology cannot solve everything, Ms. Grahn said. “As far as we know, there aren’t yet any algorithms that take individual cases into account sufficiently to follow the law. Not when it comes to children with special needs, or any other kind of individual case.”

## / Looking for transparency

One central concern with automated decision-making is transparency. How can automated decisions and the underlying algorithms be explained in a way everyone understands? And are algorithms official records that can be communicated to the public?

Simon Vinge, chief economist at the Union for Professionals (*Akademikerförbundet SSR*), has sought answers to these questions for over a year. In June 2018, he asked Trelleborg how the algorithm made decisions and how their system worked, but he did not receive satisfactory answers. After he sent a complaint to the Swedish Parliamentary Ombudsman (JO) in September 2019, he received some screenshots and a flow chart. Mr. Vinge and the Union for Professionals argued that the information is not sufficient to really understand how the program works, and he asked for 'meaningful information', in the sense of Article 15 of the GDPR, about how a decision is made.

"When it comes to automated decision-making, no one knows what they have to share, or when you've received enough information to understand how an automated decision was made. I still don't know which parameters lead to a declined application, or what is being fed into the formula," Mr. Vinge said.

Trelleborg said that they had given all the information they have been asked for. The JO will decide on the case in the coming months.

"AS FAR AS WE KNOW, THERE AREN'T YET ANY ALGORITHMS THAT TAKE INDIVIDUAL CASES INTO ACCOUNT SUFFICIENTLY TO FOLLOW THE LAW. NOT WHEN IT COMES TO CHILDREN WITH SPECIAL NEEDS, OR ANY OTHER KIND OF INDIVIDUAL CASE."

INGER GRAHN

"If it's difficult to explain how a simple rule-based algorithm works, how can we hope to explain more complex systems like machine learning?" Mr. Vinge said.

## / Analyzing the code

Last fall, Freddi Ramel, a journalist, requested the source code of the software being used in Trelleborg under Sweden's Freedom of Information Act. When Trelleborg said it was not an official document, Mr. Ramel appealed to the administrative court. Trelleborg argued that the code was a trade secret, but the court decided otherwise. The source code is an official document, judges said, and the code was then communicated to Mr. Ramel.

The code that Trelleborg finally shared is made of 136,000 lines of rules, spread out across 127 XML files. Some of the files seem to contain older, unused rulesets. Without access to the data used by the software, it is impossible to understand the rules with any certainty. The code interacts with other pieces of software, making the deciphering effort all the more difficult. But it is possible to (quite painstakingly) start outlining a general decision tree.

Without a clear explanation from the municipality, the system remains a black box. Having the code does not change anything, Mr. Vinge of the SSR union wrote in an email.

## / Personal data leak

The analysis of the code yielded more than just some of the rules guiding the RPA. It contained the names and social security numbers of approximately 250 people, seemingly citizens who previously had contact with the municipality related to welfare. This data seems to have been in the code since 2017 and is now visible for anyone who files an FOI request to see the code, as well as the subcontractors working on it.

Trelleborg municipality is currently investigating why the personal data ended up in the code, and why the code was not screened before it was made public.

Even though Trelleborg introduced their "robot" three years ago, the government has only just begun looking into this issue. In January 2020, Stockholm ordered an investigation into the use of automated decision-making by municipalities and regions. It will be published in March 2021.

# Research

By Anne Kaun & Gabriela Taranu

"WE ARE SUPPOSED TO BE ROBOTS, WE ARE SUPPOSED TO BE IDIOTS"

FROM AN INTERVIEW WITH AN EMPLOYEE  
AT THE SWEDISH PUBLIC  
EMPLOYMENT SERVICES

## Contextualization

As the previous *Automating Society* report stressed, the Swedish government is strongly invested in placing the country at the forefront of the development of artificial intelligence (AI) and automated decision-making (ADM).

The previous report addressed the AI initiatives that have been in place since 2017, including government reports on self-driving cars and an AI commission. These initiatives continue to function, but not all of them have been extended.

In 2018, a new public agency for digitalization (DIGG – Agency for Digital Government) was established to coordinate general efforts in digitalization and especially to develop AI tools in support of the public administration. One of the major private investments in the development of AI – by Wallenberg Autonomous Systems and Software Programme (WASP) – was extended with special research application calls and a graduate school focusing on AI from the perspective of social sciences and the humanities. The program is led by Virginia Dignum at the Department for Computing Science at Umeå University.

Besides these initiatives in research and development, there have been developments in terms of regulation and

oversight of automated decision-making in the public sector. Among other things, the National Audit Office began large-scale oversight of automated decision-making within public agencies at the national level. Furthermore, we have seen several litigation cases related to automated decisions for social benefit applications in the municipality of Trelleborg (the so-called Trelleborg model) as well as the first GDPR ruling in Sweden concerning a face recognition system in a gymnasium.

## A catalog of ADM cases

To understand how the public sector is using ADM, we conducted as systematic mapping exercise. In November 2019, we sent freedom of information requests to 321 government agencies and supporting units to find out if they:

- a. Worked with ADM
- b. And, if they did, in which areas they use ADM.

By the end of November, we had received 213 responses. 84 agencies did not respond within the timeframe and 24 could not be reached. Out of the 213 agencies who replied, only 15 said that they relied on ADM (often in the form of robotic process automation (RPA) or automated process flows).

The replies showed that there is some confusion as to what the term ADM encompasses. We received several questions concerning the definition of ADM, as well as longer elaborations of how ADM is understood at particular institutions. Several respondents only referred to RPA as a form of ADM, while others also included algorithms that are part of external infrastructure that they rely on for their work (e.g. social media tools that are used for external communication).

The agencies that are at the forefront of ADM are the *Pensionsmyndigheten* (Swedish Pensions Agency), *Försäkrings-*

*skassan* (Swedish Social Insurance Agency) and *Skatteverket* (the Tax Agency). Another important agency working with proactive open data initiatives is the *Lantmäteriet* (Mapping, Cadastral and Land Registration Authority) that has, among other things, automated decisions related to real estate mortgages.

One agency that has automated many processes that directly affect applicants is *Försäkringskassan* (Swedish Social Insurance Agency). This agency, together with the Swedish Association of Local Authorities and Regions, owns and develops the SSTBK database. SSTBK is a comprehensive platform that combines databases from different public agencies to allow for decisions on social benefits (agencies included are Swedish Public Employment Services, Swedish Board of Student Finance, Swedish Federation of Unemployment Insurance Funds, Swedish Social Insurance Agency, Swedish Tax Agency).

84 AGENCIES  
DID NOT RESPOND  
WITHIN THE  
TIMEFRAME AND  
24 COULD NOT BE  
REACHED. OUT OF  
THE 213 AGENCIES  
WHO REPLIED, ONLY  
15 SAID THAT THEY  
RELIED ON ADM.

Currently, the Public Employment Services are working with automation in three areas:

- self-registration
- activity reports
- financial checks of companies applying for support when employing people who have been unemployed for a long time (*företagsstöd*)

At the moment, the first contact unemployed people have with the Public Employment Services is exclusively digital and self-served. Job seekers have to self-register online and are then booked for a telephone or video call within a month to evaluate their prospects of returning to the work force (*arbetsmarknadspolitisk bedömning*).

If clients are eligible for unemployment benefits and enter an educational or training program, they have to report on their job search and program activities to the Public Employment

Services on a monthly basis. The registration, evaluation, and decision on whether to provide continued support are automated. Finally, ADM is also used to perform financial checks on companies seeking financial support to help them employ the long-term unemployed. The eligibility of a company to receive this support is automatically rated: green = eligible for support, yellow = needs an additional manual check, and red = not eligible). According to an interviewee we spoke with, there are limited possibilities for caseworkers to challenge the automated decision in this specific area.

## / Public Employment Services

In the following section, we will focus more closely on the Public Employment Services for two reasons. Firstly, the agency is actively working and experimenting with ADM and machine learning. Secondly, the organization is currently undergoing a major restructuring including substantial budget cuts that have contributed to the accelerated development of technological solutions for automation.

Budget cuts of around 2.9 billion SEK in 2019<sup>1</sup> led to 1700 employees being laid off and the closure of offices, especially those outside the big cities where many vulnerable communities reside. The agency aims to direct most of its clients to remote services and only provide assistance at local offices for a minority of cases. Prior to the budget cuts, the agency had plans to increasingly automate decision processes. The cuts have led to an increased need to implement automation to handle the workload. However, this has to be done with decreasing resources at the same time as the agency prepares for an anticipated recession and a subsequent rise in the number of unemployed people.

## / Adult Education and Economic Support – Nacka Municipality

Nacka municipality is part of Stockholm county, and it has developed and implemented a so-called robot employee called Yasmine. The robot is part of an RPA service platform that was provided by the Norwegian company Basefarm. So far, the administration of Nacka has automated three processes. Two of these are in the so-called Establishment Unit (*Etablering*), and the other is in the Work and Business Unit. The Establishment Unit is responsible for economic support, unaccompanied children and minors (refugees). The Work and Business Unit administers adult education,

<sup>1</sup> The new budget suggests there will be 1 billion SEK of additional funding for 2020, but lay-offs of 1,700 will not be revoked.



societal orientation, and labor market initiatives (Nacka Kommun 2018).

The current automation project was modelled on a similar automation project in Trelleborg municipality (see previous report), where decisions on applications for social benefits have been fully automated. Trelleborg is the first municipality to work with fully automated decision-making on social benefit applications.

Nacka has developed a centralized approach that allows all units to start automation projects. Different units and departments are supported in developing a detailed project plan and workflow description for implementing ADM (Hertin & Lindberg, 2019). Units interested in automating can apply and implement a pilot project and receive support from the central administration unit to do so.

This centralized approach is supposed to increase the degree of automation of the whole administration of the municipality while using a common approach across different units and work processes. The processes that are currently automated are related to adult education (15 rules developed for RPA) and decisions on economic support (200 rules developed for RPA). Basefarm provided the plat-

form, and the tool, which was designed by a company part-owned by Google Alphabet (CapitalG investment) is called UiPath (UiPath 2020).

## / Nectarine – AI in elderly care

*Nectarine Health* (a company previously called *Noomi* and *Aifloo*) has developed a wristband for the elderly. It is fitted with multiple sensors that monitor eating habits, movement, sleep patterns, and can detect if the wearer trips or falls (Nanalyze 2018)(Svensknarings 2018). Based on data from the wristband, caregivers are automatically alerted in case of an emergency. In 2016, *Nectarine Health* started collaborating with Aleris, a major private healthcare provider in Sweden, when the wristbands first came into use.

Apart from registering movements, the wristbands also monitor other health-related data mainly to organize elderly care more efficiently. This example is included here to illustrate that Sweden is uniquely placed in terms of collecting and sharing patient data in different public and private registries that are increasingly harvested by international and transnational private companies (Gustafsson & Röstlund, 2019).

## / Swedish microchipping movement

The Swedish microchipping movement, which took off in 2015, has been in the international news again (Läsker 2019)(Schwartz 2019). It is estimated that around 5000 people in Sweden currently have RFID technology-based microchip implants that allow them to automatically open doors and use photocopiers.

On several occasions, microchipping has become a popular feature at parties. For example, one of these microchipping events was arranged during Almedalen, a high-profile political week held during the summer of 2017, by the Swedish state-owned company Statens Järnvägar (SJ) (Petersén, 2019).

## / School Placement Algorithm

In spring 2020, a report by the state (*statens offentliga utredning*) into school segregation was made public (Statens Offentliga Utredning 2020). Among other things, the report suggests implementing a more complex algorithm for the placement of school children. At the moment, school selection is based solely on the relative proximity of the school.

**APART FROM  
REGISTERING  
MOVEMENTS,  
THE WRISTBANDS  
ALSO MONITOR  
OTHER HEALTH-  
RELATED DATA.**

In Gothenburg, one case illustrates the failure of the automated system as the algorithm only considered the straight line distance to a school, and not the geography, or the actual time it takes to get to the school building by foot or car (Dagens Nyheter 2020). After school placement decisions are made in the spring, reports often emerge of parents who are unhappy with the automated decisions. In 2018, parents in Uppsala, a city north of Stockholm, developed a placement algorithm that they believed would make the assignment fairer. However, the city board in charge never considered the algorithm (Heimer 2018).

REPORTS OFTEN  
EMERGE OF PARENTS  
WHO ARE UNHAPPY WITH  
THE AUTOMATED SCHOOL  
PLACEMENT DECISIONS.

and can solve different tasks with a certain degree of independence. AI, according to their definition, is based on algorithms made up of mathematical formulas or rules that define how AI should act. The report focuses, in particular, on machine learning as a specific form of AI, based on algorithms, that are self-learning, self-adjusting, and self-improving over time, and that can derive rules, and detect patterns in large datasets.

The report focused on two municipalities in each of the five Nordic countries (Denmark, Finland, Iceland, Norway, and Sweden) and mapped their work on AI and ADM. One of the main takeaways from the report is that, while neither AI nor ADM are used very broadly in any of the Nordic countries, local municipalities are running several different pilot studies. These studies tend to focus on chatbots, which are increasingly common at the municipal level.

The report emphasizes that the use of AI and ADM might have a large impact on trust in public institutions if it is not implemented in a sustainable and sensible manner. The Nordic countries have a long history of strong social trust, and trust in public institutions, and the authors of the report provide three general suggestions. Firstly, municipalities in the Nordic countries should continuously exchange experience with AI implementation projects. Secondly, there should be shared ethical guidelines across the Nordic countries, and thirdly, they should come together to invest in research and development projects to concentrate resources and spend funding efficiently.

## / Swedish Police Piloting Face Recognition Software

As part of a broadly discussed special operation to fight organized crime (operation Rime Ice), the Swedish police implemented several face recognition pilot projects including searching through visual material such as pictures from crime scenes and CCTV footage to identify reoccurring people. The police did not inform the Swedish Data Protection Authority about the pilot testing. Last year the authority approved that the police could continue matching face shots with an offender register (*signalementregister*). However, their application for testing face recognition at Skavsta airport was dismissed in 2019. In 2020, the government submitted a revised application to implement a pilot project from January 2021 (Munther 2020). At the moment, the Swedish Data Inspection Authority is reviewing the police's use of the American face recognition app Clearview.

# Policy, oversight and debate

## / Report by the Nordic Council of Ministers

In 2019, the Nordic Council of Ministers published a report about the use of artificial intelligence at the municipal level across the Nordic countries (Andreasson & Stende, 2019). The authors focused explicitly on AI that they define as autonomous systems that are apparently intelligent. The report does not explicitly define ADM, but refers to a number of tasks that are or can potentially be automated with the help of AI. These systems are based on data analysis

## / Report by the Union for Professionals

In 2019, the Union for Professionals (Akademikerförbundet SSR) published a report written by Lupita Svensson, a senior lecturer at the School of Social Work at Lund University, on ADM within the Swedish social service system (Svensson, 2019). Svensson evaluated and mapped ADM across all municipalities with the help of RPA. Her report showed that only 16 out of a total of 290 municipalities have implemented RPA in their administration of social benefits. However, only one – Trelleborg municipality – has implemented fully automated decisions, while the other 15 municipalities use RPA for decision support. Besides mapping the extent to which Swedish municipalities use automated processes in social services, Svensson also conducted interviews with politicians, project leaders, and case workers

across municipalities. She came to the conclusion that the technological shift that the Swedish social services are currently going through can be compared to computerization in the 1980s. However, there are important differences. The policy for social services during the 1980s emphasized the importance of trust, whereas current policies, and the technological solutions they utilize, emphasize control. At the same time, caseworkers argue that they have more time to actually do social work, although they need to renegotiate what their basic tasks encompass, besides a strong focus on administration and documentation. This new wave of “deep” digitalization that goes beyond earlier e-governance initiatives is still ongoing and is characterized by contradictions. In the broader context, the need for support including social benefits is increasing, while resources are systematically shrinking. Automation and digitalization are often motivated with efficiency and cost saving arguments however, the necessary infrastructure is often cost-intensive and needs continuous maintenance and these problems are often downplayed in the discussion. Svensson’s findings point to a fundamental shift in how social services are organized and delivered to citizens.

## **/ Report by DIGG – Agency for Digital Government**

In Spring 2020, the Swedish Agency for Digital Government (founded in 2018) published a report for the government entitled *To Support the Public Administration in Using AI* (DIGG, 2020). The report maps the pre-conditions for the use of AI in public administration, the main challenges, and develops a set of recommendations to increase the use of AI in public administration. The starting point is that the costs and resource requirements for both the public and welfare sector will increase considerably over the next decade. The report states that AI offers huge economic potential that, until now, has not been fully exploited. AI is connected with the possibility to increase, and make more equal, access to health services, improve education, and implement more just and efficient decision-making in public administration. The report points out two particular challenges. Firstly, questions of data security and secondly, ethical considerations. In order to support the implementation of AI in public administration, the report makes the following suggestions:

- Create a center for expertise in AI
- Establish a platform for coordination, co-development, and innovation
- Develop an AI guide for public administration

THE DATA PROTECTION  
AUTHORITY ISSUED THE FIRST  
RECORDED GDPR FINE IN SWEDEN -  
FOR UNLAWFUL USE OF FACE  
RECOGNITION SOFTWARE.

- Establish the juridical foundation to allow for test-bed activities
- Develop professional and role-based education for AI
- Develop a national data strategy for public administration

## **/ First GDPR fine in Sweden for the use of face recognition at a public school**

In 2018, a public high school in Skellefteå, in Västerbotten County, used face recognition to register student attendance (Lindström 2019). One class took part in a three-week pilot of a system implemented with Tieto (IT Pedagogen 2019), a Nordic IT software and service company with headquarters in Finland. However, the Data Protection Authority (Datainspektionen, 2019) issued a fine of 20,000 euros (200,000 SEK) to Skellefteå municipality – which is the first recorded GDPR fine in Sweden – for unlawful use of face recognition software.

The Data Protection Authority ruled that GDPR had been violated in several ways (Heinje 2019). Firstly, the use of face recognition technology was considered excessively intrusive and, in that sense, it violated the principle of data minimization (article 5(1)(c)). Secondly, face recognition falls under ‘special categories of data’ (article 9) that require specific consent by the data subject. Even though the students gave their informed consent, their position as students (and minors) meant that the decision to participate was not a real choice. Thirdly, the school did not document a Data Protection Impact Assessment (DPIA) and hence violated article 35.

The relatively modest fine was imposed because of the brief duration of the pilot scheme, and the fact that only 22 students were involved. Skellefteå has challenged the

decision by arguing that participation in the scheme was voluntary and that students and their parents who opted into the pilot gave their full consent to the gathering of data (Lindström 2019).

## **/ Trelleborg municipality reported to Parliamentary Ombudsperson**

Our first report for Automating Society highlighted the automation project of Trelleborg municipality that fully automated decisions on social benefit applications and has worked on this particular ADM project since 2017. In 2019, Simon Vinge of the Union for Professionals (Akademikerförbundet SSR) reported the municipality of Trelleborg to the Parliamentary Ombudsperson (Justitieombudsmannen JO). His complaint was that the authorities failed to respond to a Freedom of Information (FOI) request related to the use of RPA to make automated decisions. As of May 2020, they have not heard back from the Parliamentary Ombudsmen.

Similarly, the journalist Freddi Ramel started a discussion about transparency related to the ADM system used in Trelleborg. After several failed attempts to get access to the source code from Trelleborg municipality, and after reaching out to the Danish company responsible for the coding, he took legal action. In contrast to Simon Vinge, Ramel submitted an appeal to the Administrative Court of Appeal, by arguing that the source code of the software falls under the Swedish *Offentlighetsprincipen* (the principle of public access to official records). Following his appeal, the court ruled that the source code had to be made publicly accessible and fully included in the principle of public access (Dagens Samhälle 2020).

After this ruling was made public, another journalist at the financial newspaper, *Dagens Industri*, submitted an FOI request to view the source code. Trelleborg municipality forwarded the request to the Danish software company, who then shared the source code with the journalists. However, the code was not checked beforehand, and sensitive data,

***FOLLOWING HIS APPEAL,  
THE COURT RULED THAT  
THE SOURCE CODE HAD  
TO BE MADE PUBLICLY  
ACCESSIBLE AND FULLY  
INCLUDED IN THE PRINCIPLE  
OF PUBLIC ACCESS.***

including personal identification numbers (similar to social security numbers), names, and information about home care for the elderly, was also shared with the journalists (Dagens Samhälle 2020).

## / Mapping and review by the National Audit Office

In 2019, the National Audit Office (*Riksrevisionen*) conducted an initial mapping of ADM currently in use by public agencies. Based on the results, it decided to undertake a larger review of ADM used by the public sector, and that has a direct impact on citizens. Their report will be published in October 2020<sup>2</sup>.

## Key takeaways

There is an ongoing and vivid debate about ADM, which is often framed as AI in the public sector, and there have been numerous events, panel discussions, reports, etc. At the same time, there are not too many instances of fully automated decision-making. Having said that, public institutions continue to test the water and navigate grey legal areas. The dominant forms of ADM are those made up of RPA and based on solutions provided by companies like Blue Prism and UiPath. Automation and pilot projects are often implemented in collaboration with external consulting firms such as PricewaterhouseCoopers or Boston Consulting Group. The Swedish Association of Local Authorities and Regions coordinated the introduction of ADM<sup>3</sup>. Regarding automation projects, private companies are increasingly becoming part of, and are responsible for, critical social infrastructure.

## References:

Andreasson, Ulf & Stende, Truls (2019). Nordiske Kommuners Arbeid Med Kunstig Intelligens. Nordisk Ministerråd. Retrieved from: <http://norden.diva-portal.org/smash/get/diva2:1365285/FULLTEXT01.pdf>.

Dagens Nyheter (2020): Nytt dataprogram placerar skolbarn i Göteborg: "Helt koko för Ellen!" <https://www.dn.se/nyheter/sverige/nytt-dataprogram-skapar-kaos-i-goteborgs-antagning-till-grundskolan-helt-ko-ko-for-ellen/>

Dagens Samhälle (2020): Känsliga uppgifter spreds via kod till bistandsrobot <https://www.dagenssamhalle.se/nyhet/kansliga-uppgifter-spreds-kod-till-bistandsrobot-31911>

Datainspektionen (2019). Beslut. <https://www.datainspektionen.se/globalassets/dokument/beslut/beslut-ansiktsigenkanning-for-narvarokontroll-av-elever-dnr-di-2019-2221.pdf>

DIGG (2020). Att främja den offentliga förvaltningens förmåga att använda AI. Retrieved from: <https://www.digg.se/globalassets/slutrapport---framja-den-offentliga-forvaltningens-formaga-att-anvanda-ai.pdf>

Försäkringskassan (2015): SSBTEK Allmän beskrivning [https://www.forsakringskassan.se/wps/wcm/connect/283d29b0-563d-4e0f-8bf8-d7bf90594fa7/allman-beskrivning\\_ssbtek\\_revb.pdf?MOD=AJPERES&CVID](https://www.forsakringskassan.se/wps/wcm/connect/283d29b0-563d-4e0f-8bf8-d7bf90594fa7/allman-beskrivning_ssbtek_revb.pdf?MOD=AJPERES&CVID)

Gustafsson, Anne & Röstlund, Lisa (2019). Konsulterna. Stockholm: Mondial.

2 <https://www.riksrevisionen.se/nu-granskas/inledda-granskningar/automatiserat-beslutsfattande-i-statsforvaltningen.html>

3 <https://www.computerweekly.com/news/252471846/Swedish-authorities-introduce-robots-to-help-social-workers>



Granskning, P. (2019): Automatiserat beslutsfattande i statsförvaltningen <https://www.riksrevisionen.se/nu-granskas/pagaende-granskningar/automatiserat-beslutsfattande-i-statsforvaltningen.html>

Heimer, Johan (2018): Föräldrarnas algoritm löser skolvalsproblem <https://unt.se/nyheter/uppsala/foraldrarnas-algoritm-loser-skolvalsproblem-4929283.aspx>

Heinje, Thomas (2019): Skolans ovanliga test: Elevernas närvaro registreras med kamerateknik <https://www.svt.se/nyheter/lokalt/vasterbotten/skolans-ovanliga-test-registrerar-elevernas-narvaro-med-kamera>

Hertin, Emma & Lindberg, Kristine (2019). Införande av digitala medarbetare för att hjälpa eller ersätta mänskliga kollegor. MA thesis Linköping University. Retrieved from: <http://www.diva-portal.org/smash/get/diva2:1326430/FULLTEXT02>

IT Pedagogen (2019): Tieto och Skellefteå kommun testar ansiktsgenkänning i klassrummet <https://it-pedagogen.se/tieto-och-skelleftea-kommun-testar-ansiktsgenkanning-i-klassrummet/>

Läcker, Kristina (2019): The Ethical Minefield of Microchipping <https://www.spiegel.de/international/business/the-human-microchipping-trend-sweeping-sweden-a-1292924.html>

Lindström, Karin (2019): Skellefteå överklagar GDPR-bot för ansiktsgenkänning <https://computersweden.idg.se/2.2683/1.723024/skelleftea-overklagar-gdpr-bot>

Munther, Victor (2020): Regeringen vill testa ansiktsgenkänning på svensk flygplats <https://www.svt.se/nyheter/inrikes/regeringen-vill-testa-ansiktsgenkanning-pa-svensk-flygplats>

Nacka Kommun (2018): Roboten Yasemin snabbar på ansökningar till vuxenutbildning <https://www.nacka.se/forskola-skola/vuxenutbildning/nyheter-vuxenutbildning/2018/11/roboten-yasemin-snabbar-pa-ansokningar-till-vuxenutbildning/>

Nanalze (2018): Top AI startups in Sweden <https://www.nanalyze.com/2018/12/top-ai-startups-sweden/>

Petersén, Moa (2019). The Swedish Microchipping Phenomenon. Bingley: Emerald Publishing.

Schwartz, Oskar (2019): The rise of microchipping: are we ready for technology to get under the skin? <https://www.theguardian.com/technology/2019/nov/08/the-rise-of-microchipping-are-we-ready-for-technology-to-get-under-the-skin>

Statens Offentlig Utredning (2020): En Mer Likvardig Skola <https://www.regeringen.se/498b68/contentassets/fcf0e59defe04870a39239f5bda331f4/en-mer-likvardig-skola--minskad-skolsegregation-och-forbattrad-resurstilldelning-sou-202028>

Svensson, Lupita (2019). "Tekniken är den enkla biten" Om att implementera digital automatisering i handläggning av försörkningsstöd. <https://akademssr.se/sites/default/files/files/LupitaSvensson.pdf>

Svensktnarings (2018): Ny teknik i välfärden Vår tids välfärdskliv [https://www.svensktnaringsliv.se/migration\\_catalog/Rapporter\\_och\\_opinionsmaterial/Rapporter/ny-teknik-i-valfarden\\_714707.html/BINARY/Ny%20teknik%20i%20v%C3%A4lf%C3%A4rden.pdf](https://www.svensktnaringsliv.se/migration_catalog/Rapporter_och_opinionsmaterial/Rapporter/ny-teknik-i-valfarden_714707.html/BINARY/Ny%20teknik%20i%20v%C3%A4lf%C3%A4rden.pdf)

Svt Nyheter (2020): Polisen testade ny teknik för ansiktsgenkänning – utan att informera <https://www.svt.se/nyheter/inrikes/polisen-testade-ny-teknik-utan-att-informera>

UiPath (2020): Our investors [https://www.uipath.com/company/investors?utm\\_source=GoogleSearch&utm\\_medium=cpc&utm\\_term=uipath-e&utm\\_content=391571747460&utm\\_campaign=AR19FWV](https://www.uipath.com/company/investors?utm_source=GoogleSearch&utm_medium=cpc&utm_term=uipath-e&utm_content=391571747460&utm_campaign=AR19FWV)

DENMARK  
ESTONIA  
FINLAND  
FRANCE  
GERMANY  
GREECE  
ITALY  
NETHERLANDS  
POLAND  
PORTUGAL  
SLOVENIA  
SPAIN  
SWEDEN

**SWITZERLAND**

**STORY**

PAGE 256

**RESEARCH**

PAGE 259

UNITED KINGDOM





Find out more on p. 259 in the research chapter under "Cancer diagnoses and treatments".

# Swiss police automated crime predictions but have little to show for it

**A review of 3 automated systems in use by the Swiss police and judiciary reveals serious issues. Real-world effects are impossible to assess due to a lack of transparency.**

By Nicolas Kayser-Bril



## / Predicting burglaries

Precobs has been used in Switzerland since 2013. The tool is sold by a German company that makes no mystery of its lineage with “Minority Report”, a science-fiction story where “precogs” predict some crimes before they occur. (The plot revolves around the frequent failures of the precogs and the subsequent cover-up by the police).

The system tries to predict burglaries from past data, based on the assumption that burglars often operate in small areas. If a cluster of burglaries is detected in a neighborhood, the police should patrol that neighborhood more often to put an end to it, the theory goes.

Three cantons use Precobs: Zürich, Aargau, and Basel-Land, totaling almost a third of the Swiss population. Burglaries have fallen dramatically since the mid-2010s. The Aargau police even [complained](#) in April 2020 that there were now too few burglaries for Precobs to use.

But burglaries fell in every Swiss canton, and the three that use Precobs are nowhere near the best performers. Between 2012-2014 (when burglaries were at their peak) and between 2017-2019 (when Precobs was in use in the three cantons), the number of burglaries decreased in all cantons, not just in the three that used the software. The decrease in Zürich and Aargau was less than the national average of -44%, making it unlikely that Precobs had much of an actual effect on burglaries.

A 2019 [report](#) by the University of Hamburg, could not find any evidence of the efficacy of predictive policing solutions, including Precobs. No public documents detail how much Swiss authorities have spent on the system, but Munich paid 100,000 euros to install Precobs (operating costs not included).

## / Predicting violence against women

Six cantons (Glarus, Luzern, Schaffhausen, Solothurn, Thurgau, and Zürich) use the Dyrias-Intimpartner system to predict the likelihood that a person will assault their intimate

partner. Dyrias stands for “dynamic system for the analysis of risk” and is also built and sold by a German company.

According to a 2018 [report](#) by Swiss public-service broadcaster SRF, Dyrias requires police officers to answer 39 “yes” or “no” questions about a suspect. The tool then outputs a score on a scale from one to five, from harmless to dangerous. While the total number of persons tested by the tool is unknown, [a tally by SRF](#) showed that 3,000 individuals were labeled “dangerous” in 2018 (but the label might not be derived from using Dyrias).

The vendor of Dyrias claims that the software correctly identifies eight out of ten potentially dangerous individuals. However, another study looked at the false positives, individuals labeled dangerous who were in fact harmless, and found that six out of ten people flagged by the software should have been labeled harmless. In other words, Dyrias boasts good results only because it takes no risks and assigns the “dangerous” label liberally. (The company behind Dyrias disputes the results).

Even if the performance of the system was improved, its effects would still be impossible to assess. Justyna Gospodinov, the co-director of BIF-Frauenberatung, an organization that supports victims of domestic violence, told Algorithm-Watch that, while cooperation with the police was improving and that the systematic assessment of risk was a good thing, she could not say anything about Dyrias. “When we take in a new case, we do not know whether the software was used or not,” she said.

## / Predicting recidivism

Since 2018, all justice authorities in German-speaking cantons use ROS (an acronym for “Risikoorientierter Sanktionenvollzug” or risk-oriented execution of prison sentences). The tool labels prisoners ‘A’ when they have no risk of recidivism, ‘B’ when they could commit a new offense, or ‘C’ when they could commit a violent crime. Prisoners can be tested several times, but subsequent tests will only allow them to move from category A to B or C and not the other way around.

PRECABS HAS BEEN USED IN SWITZERLAND SINCE 2013. THE TOOL IS SOLD BY A GERMAN COMPANY THAT MAKES NO MYSTERY OF ITS LINEAGE WITH “MINORITY REPORT”, A SCIENCE-FICTION STORY WHERE “PRECOGS” PREDICT SOME CRIMES BEFORE THEY OCCUR.



A [report by SRF](#) revealed that only a quarter of the prisoners in category C committed further crimes upon being released (a false-positive rate of 75%) and that only one in five of those who committed further crimes were in category C (a false-negative rate of 80%), based on a [2013 study](#) by the University of Zürich. A new version of the tool was released in 2017 but has yet to be reviewed.

The French and Italian-speaking cantons are working on an alternative to ROS, which should be deployed in 2022. While it keeps the same three categories, their tool will only work in conjunction with prisoner interviews that will be rated.

## / Mission: Impossible

Social scientists are sometimes very successful when predicting general outcomes. In 2010, the Swiss statistics office predicted that the resident population of Switzerland would reach 8.5 million by 2020 (the actual 2020 population is 8.6 million). But no scientist would try to predict the date a given individual will die: Life is simply too complicated.

In this regard, demography is no different from criminology. Despite claims to the contrary by commercial vendors, predicting individual behavior is likely to be impossible. In 2017, a group of scientists tried to settle the issue. They asked 160 teams of researchers to predict school performance, the likelihood of being evicted from home, and four other outcomes for thousands of teenagers, all based on precise data collected since birth. Thousands of data points were available for each child. The results, [published in April 2020](#), are humbling. Not only could not a single team predict an outcome with any accuracy, but the ones who used artificial intelligence performed no better than teams who used only a few variables with basic statistical models.

Moritz Büchi, a senior researcher at the University of Zürich, is the only Swiss scholar who took part in this experiment. In an email to AlgorithmWatch, he wrote that while crime was not part of the outcomes under scrutiny, the insights gained from the experiment probably apply to predictions of criminality. This does not mean that predictions should not be attempted, Mr. Büchi wrote. But turning simulations into ready-to-use tools gives them a “cloak of objectivity” which can discourage critical thinking, with potentially devastating consequences for the people whose future is predicted.

Precobs, which does not attempt to predict the behavior of specific individuals, does not fall into the same category, he

TURNING SIMULATIONS INTO  
READY-TO-USE TOOLS GIVES  
THEM A “CLOAK OF OBJECTIVITY”  
WHICH CAN DISCOURAGE CRITICAL  
THINKING, WITH POTENTIALLY  
DEVASTATING CONSEQUENCES  
FOR THE PEOPLE WHOSE  
FUTURE IS PREDICTED.

added. More policing could have a deterrent effect on criminals. However, the detection of hotspots relies on historical data. This might lead to the over-policing of communities where crime was reported in the past, in a self-reinforcing feedback loop.

## / Chilling effects

Despite their patchy track record and evidence of the near-impossibility to predict individual outcomes, Swiss law enforcement authorities keep using tools that claim to do just that. Their popularity is due in part to their opacity. Very little public information exists on Precobs, Dyrias, and ROS. The people impacted, who are overwhelmingly poor, rarely have the financial resources needed to question automated systems, as their lawyers usually focus on verifying the basic facts alleged by the prosecution.

Timo Grossenbacher, the journalist who investigated ROS and Dyrias for SRF in 2018, told AlgorithmWatch that finding people affected by these systems was “almost impossible”. Not for lack of cases: ROS alone is used on thousands of inmates each year. Instead, their opacity prevents watchdogs from shedding light on algorithmic policing.

Without more transparency, these systems could have a “chilling effect” on Swiss society, according to Mr. Büchi of the University of Zürich. “These systems could deter people from exercising their rights and could lead them to modify their behavior,” he wrote. “This is a form of anticipatory obedience. Being aware of the possibility of getting (unjustly) caught by these algorithms, people may tend to increase conformity with perceived societal norms. Self-expression and alternative lifestyles could be suppressed.”

# Research

By Nadja Braun Binder and Catherine Egli

## Contextualization

Switzerland is a distinctly federalist country with a pronounced division of powers. Therefore, technical innovations in the public sector are often first developed in the cantons.

One example of this is the introduction of an electronic identity (eID). At the federal level, the legislative process required to introduce the eID has not yet been completed, whereas an officially confirmed electronic identity is already in operation in one canton. In 2017, and as part of Switzerland's cantonal eGovernment strategy, Schaffhausen canton became the first to introduce a digital identity for residents. Using this eID, citizens can apply online for a fishing permit, calculate tax liabilities of a real estate profit or a capital statement, or request a tax declaration extension, among other things.

Also, an adult account can be set up at the Child and Adult Protection Authority, and doctors can apply for credit for patients who are hospitalized outside their district. Another example, that began as part of a pilot project in the same canton in September 2019, enables residents to order extracts (via smartphone) from the debt collection register. These services are in a constant state of expansion (Schaffhauser 2020). Although the eID itself is not an ADM process, it is an essential prerequisite for access to digital government services, and therefore, could also facilitate access to automated procedures, e.g., in the tax field. The fact that a single canton has progressed further down this path than the Federal Government is typical for Switzerland.

Direct democracy is another defining element of the Swiss state. For example, the legislative process for a national eID has not yet been completed because a referendum is going to be held on the corresponding parliamentary bill (eID – Referendum 2020). Those who have asked for the referendum do not fundamentally oppose an

official eID, but they want to prevent private companies from issuing the eID and managing sensitive private data.

Another element that must be taken into account is the good economic situation in Switzerland. This allows great progress to be made in individual areas, such as automated decisions used in medicine, and in many areas of research. Although there is no central AI or ADM strategy in Switzerland, due to the distinct federal structure and the departmental division of responsibilities at the federal level, sectoral research is conducted at a globally competitive level.

## A catalog of ADM cases

### / Cancer diagnoses and treatments

At the moment, Switzerland is investigating the use of automated decisions in medicine, which is why ADM has been developed further in the healthcare sector than in other domains. Today, more than 200 different types of cancer are known and almost 120 drugs are available to treat them. Every year, numerous cancer diagnoses are made, and, as each tumor has its own particular profile with gene mutations that help the tumor grow, this creates problems for oncologists. However, once they have made the diagnosis and determined the possible gene mutation, they have to study the ever-growing medical literature in order to select the most effective treatment.

THIS USE OF ADM  
TO ANALYZE MEDICAL  
IMAGES IS NOW STANDARD  
PRACTICE AT THE  
UNIVERSITY HOSPITAL  
OF ZURICH.

This is why the Geneva University Hospitals are the first hospitals in Europe to use the IBM Watson Health tool, Watson for Genomics®, to better find therapeutic options and suggest treatment for cancer patients. The doctors

still examine the gene mutations and describe where and how many of them occur, but Watson for Genomics® can use this information to search a database of about three million publications. The program then creates a report classifying genetic alterations in the patient's tumor and providing associated relevant therapies and clinical trials.

Until now, cancer doctors have had to do this work themselves – with the risk that they might overlook a possible treatment method. Now the computer program can take over the research, but oncologists still have to carefully check the literature list that the program produces, and then they can decide on a treatment method. As a result, Watson for Genomics® saves a great deal of time in analysis and provides important additional information. In Geneva, the report from this ADM-tool is used during the preparation of the Tumor Board, where physicians take note of the treatments proposed by Watson for Genomics® and discuss them in plenary to jointly develop a treatment strategy for each patient (Schwerzmann/Arroyo 2019).

ADM is also in use at the University Hospital of Zurich, as it is particularly suitable for repetitive tasks, predominantly in radiology, and pathology, and, therefore, it is used to calculate breast density. During mammography, a computer algorithm automatically analyzes X-ray images and classifies the breast tissue into A, B, C or D (an internationally recognized grid for risk analysis). By analyzing risk based on breast density, the algorithm greatly assists physicians in assessing a patient's breast cancer risk, since breast density is one of the most important risk factors in breast cancer. This use of ADM to analyze medical images is now standard practice at the University Hospital of Zurich. Furthermore, research into advanced algorithms for the interpretation of ultrasound images is ongoing (Lindner 2019).

Having said this, more than one-third of breast cancers are overlooked in mammography screening examinations, which is why research is being carried out on how ADM can support the interpretation of ultrasound (US) images. The interpretation of US breast images contrasts sharply with standard digital mammography – which is largely observer-dependent and requires well-trained and experienced radiologists. For this reason, a spin-off of the University Hospital of Zurich has researched how ADM may support and standardize US imaging. In doing so, the human decision process was simulated according to the breast imaging, reporting, and data system. This technique is highly accurate, and, in the future, this algorithm may be used to mimic human decision-making, and become the standard for the de-

tection, highlighting, and classification of US breast lesions (Ciritisi a.o. 2019 p. 5458–5468).

## **/ Chatbot at the Social Insurance Institution**

To simplify, and support administrative communication, certain cantons also use so-called chatbots. In particular, a chatbot that was tested in 2018 at the “Sozialversicherungsanstalt des Kantons St. Gallen” (Social Insurance Institution in the canton of St. Gallen, SVA St. Gallen). The SVA St. Gallen is a center of excellence for all kinds of social insurance, including a premium reduction for health insurance. Health insurance is compulsory in Switzerland and covers every resident in the event of illness, maternity, and accidents, and offers everyone the same range of benefits. It is funded by the contributions (premiums) of citizens. The premiums vary according to the insurer, and depend on an individual's place of residence, type of insurance needed, and it is not related to income level. However, thanks to subsidies from the cantons (premium reduction), citizens on a low income, children, and young adults in full-time education or training, often pay reduced premiums. The cantons decide who is entitled to a reduction (FOPH 2020).

Towards the end of each year, the SVA St. Gallen receives approximately 80,000 applications for premium reductions. To reduce the workload associated with this concentrated flood of requests, they tested a chatbot via Facebook Messenger. The object of this pilot project was to offer customers an alternative method of communication. The first digital administrative assistant was designed to provide customers with automatic answers to the most important questions regarding premium reductions. For example: what are premium reductions and how can a claim be made? Can I claim a premium reduction? Are there any special cases, and how should I proceed? How is premium reduction calculated and paid out? Also, if it was indicated, the chatbot could refer customers to other services offered by the SVA St. Gallen, including the premium reduction calculator and the interactive registration form. While the chatbot does not make the final decision to grant premium reductions, it can still reduce the number of requests as it can inform unauthorized citizens of the likely failure of their request. It also performs a significant role in disseminating information (Ringeisen/Bertolosi-Lehr/Demaj 2018 S.51-65).

Due to the positive feedback from this first test run, the chatbot was integrated into the SVA St. Gallen's website in

2019 and there is a plan to gradually expand the chatbot to include other insurance products covered by the SVA St. Gallen. It is possible that the chatbot will also be used for services related to contributions to Old-Age and Survivor's Insurance, Disability Insurance, and Income Compensation Insurance (IPV-Chatbot 2020).

## **/ Penal System**

The Swiss Execution of Penal Sentences and Justice is based on a system of levels. According to this system, inmates are generally granted increasing amounts of freedom as the duration of their imprisonment continues. This makes it a collaborative process between executive authorities, penal institutions, therapy providers, and probation services. Of course, the risk of escape and recidivism are decisive factors when it comes to granting these greater freedoms.

In recent years, and in response to convicted felons committing several tragic acts of violence and sex offenses, the ROS (Risk-Oriented Sanctioning) was introduced. The primary objective of ROS is to prevent recidivism by harmonizing the execution of sentences and measures across the various levels of enforcement with a consistent focus on recidivism prevention and reintegration. ROS divided the

work with offenders into four stages: triage, assessment, planning, and progress. During triage, cases are classified according to their need for recidivism risk assessment. Based on this classification, a differentiated individual case analysis is carried out during the assessment stage. During the planning stage, these results are developed into an individual execution plan for the sanction of the corresponding offender, which is continuously reviewed during the progress stage (ROSNET 2020).

Triage plays a decisive role at the beginning of this process – both for the offender and in terms of ADM, as it is performed by an ADM-tool called the Fall-Screening-Tool (Case Screening Tool, FaST). FaST automatically divides all cases into classes A, B, and C. Class A signifies that there is no need for assessment, class B equates to a general risk of further delinquency, and class C corresponds to the risk of violent or sexual delinquency.

This classification is determined by using criminal records and is based on general statistical risk factors including age, violent offenses committed before the age of 18, youth attorney entries, number of previous convictions, offense category, sentences, polymorphic delinquency, offense-free time after release, and domestic violence. If risk factors

***ROS DIVIDED THE WORK  
WITH OFFENDERS INTO  
FOUR STAGES: TRIAGE,  
ASSESSMENT, PLANNING, AND  
PROGRESS.***

are met that, according to scientific findings, have a specific connection with violent or sexual offenses, then a C classification is made. If risk factors are met that have a specific connection with general delinquency, then a B classification is applied. If no risk factors, or hardly any, are found, then an A classification is made. Therefore, the classification consists of items (risk factors) in a closed answer format, each of which has a different weighting (points). If a risk factor is given, the relevant points are included in the total value. For the overall result, the weighted and confirmed items are added to a value, leading to the A, B or C classification that, in turn, serves as a basis to decide if further assessment is necessary (stage 2).

This classification is carried out fully automatically by the ADM-application. However, it is important to note that this is not a risk analysis, but a way of screening out the cases with increased assessment needs (Treuhardt/Kröger 2018 p. 24-32).

Nevertheless, the triage classification has an effect on how those responsible at a particular institution make decisions and which assessments are to be made. This also determines the offender's so-called 'problem profile' regarding how the execution of sentences and measures are planned (stage 3). In particular, this planning defines any possible facilitation of enforcement, such as open enforcement, day release employment, or external accommodation. Furthermore, no ADM applications are apparent in the other stages of ROS. FaST is, therefore, only used during the triage stage.

## / Predictive Policing

In some cantons, in particular in Basel-Landschaft, Aargau, and Zurich, the police use software to help prevent criminal offenses. They rely on the commercial software package "PRECOBS" (Pre-Crime Observation System), which is solely used for the prognosis of domestic burglaries. This relatively common crime has been well researched scientifically, and police authorities usually have a solid database regarding the spatial and temporal distribution of burglaries as well as crime characteristics. Furthermore, these offenses indicate a profes-

sional perpetrator and thus show an above-average probability of subsequent offenses. In addition, corresponding prognosis models can be created using relatively few data points. PRECOBS is, therefore, based on the assumption that burglars strike several times within a short period if they have already been successful in a certain location.

The software is used to search for certain patterns in the police reports on burglaries, such as how the perpetrators proceed and when and where they strike. Subsequently, PRECOBS creates a forecast for areas where there is an increased risk of burglary in the next 72 hours. Thereupon, the police send targeted patrols to the area. PRECOBS thus generates forecasts on the basis of primarily entered decisions and it does not use machine learning methods. Although there are plans to extend PRECOBS in the future to include other offenses (such as car theft or pickpocketing) and consequently create new functionalities, it should be noted that the use of predictive policing in Switzerland is currently limited to a relatively small and clearly defined area of preventive police work (Blur 2017, Leese 2018 p. 57-72).

## / Customs clearance

At the federal level, ADM is expected to be particularly present at the Federal Customs Administration (FCA), since this department is already highly automated. Accordingly, the assessment of customs declarations is largely determined electronically. The assessment procedure can be divided into four steps: summary examination procedure, acceptance of a customs declaration, verification, and inspection, followed by an assessment decision.

THE DETERMINATION  
OF THE CUSTOMS  
DUTY WILL BE DECIDED  
FULLY AUTOMATICALLY.

IN CONTRAST,  
HUMAN CONTACT IS  
INTENDED TO BE  
CONCENTRATED  
ONLY ON CHECKING  
SUSPICIOUS GOODS  
AND PEOPLE.

The summary examination procedure represents a plausibility check and is carried out directly by the system used in the case of electronic customs declarations. Once the electronic plausibility check has been completed, the data processing system automatically adds the acceptance date and time to the electronic customs declaration, meaning the customs declaration has been accepted. Up to this point, the procedure runs without any human intervention by the authorities.

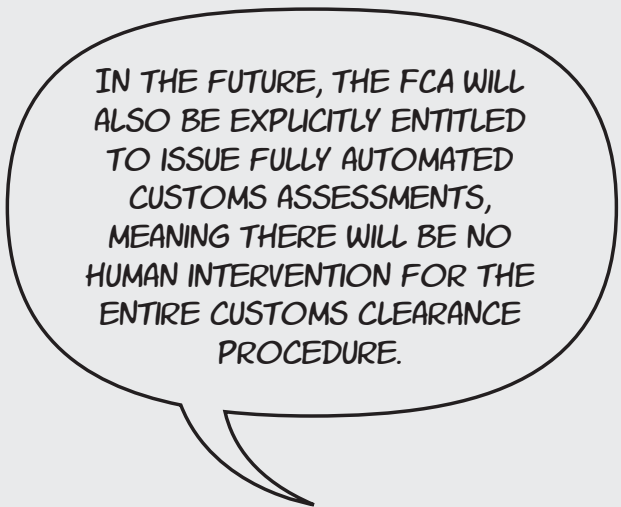
However, the customs office may subsequently carry out a full or random inspection and verification of the declared



goods. To this end, the computerized system carries out a selection based on a risk analysis. The final stage of the procedure is when the assessment decision is issued. It is not known whether or not this assessment decision can already be issued without any human intervention. However, the DaziT program will clarify this uncertainty.

The DaziT program is a federal measure to digitize all customs processes by 2026 to simplify and accelerate border crossings. Border authorities' customer relations relating to the movement of goods and people will be fundamentally redesigned. Customers who behave correctly should be able to complete their formalities digitally, independent of time and place. While the exact implementation of the DaziT program is still at the planning stage, the revision of the Customs Act that correlates to DaziT is included in the revision of the Federal Act on Data Protection.

This is explained in more detail below, and should serve to clarify the previously mentioned uncertainty regarding the automated customs assessment procedure: In the future, the FCA will also be explicitly entitled to issue fully automated customs assessments, meaning there will be no human intervention for the entire customs clearance procedure. Thus, the determination of the customs duty will be decided fully automatically. In contrast, human contact is intended to be concentrated only on checking suspicious goods and people (EZV 2020).



IN THE FUTURE, THE FCA WILL ALSO BE EXPLICITLY ENTITLED TO ISSUE FULLY AUTOMATED CUSTOMS ASSESSMENTS, MEANING THERE WILL BE NO HUMAN INTERVENTION FOR THE ENTIRE CUSTOMS CLEARANCE PROCEDURE.

tabase. This technology can be used for various purposes, but at the moment in Switzerland it is only used to a limited extent. At the federal level, the system for automatic vehicle detection and traffic monitoring is only used as a tactical tool depending on the situation and risk assessments, as well as economic considerations, and only at state borders (parlament.ch 2020). The half-canton of Basel-Landschaft has enacted a legal basis for the automatic recording of vehicle registration plates and alignment with corresponding databases (EJPD 2019).

## / Accident and Military Insurance

Throughout the revision of the Data Protection Act (explained in more detail below), it was decided that the accident and military insurance companies will be entitled to automatically process personal data. It is not clear what automated activities the insurance companies intend to use in the future. However, they may, for example, use algorithms to evaluate policyholder's medical reports. Through this fully automated system, premiums could be calculated, and benefit claim decisions made and coordinated with other social benefits. It is planned that these bodies will be authorized to issue automated decisions.

## / Automatic Vehicle Recognition

In recent years, both politicians and the public have become concerned with the use of automatic systems, such as cameras that capture vehicle license plates, read them using optical character recognition, and compare them with a da-

## / Allocation of primary school pupils

Another algorithm that has been developed, but is not yet in use, is designed to allocate primary school pupils. International studies indicate that social and ethnic segregation between urban schools is increasing. This is problematic, as the social and ethnic composition of schools has a demonstrable effect on the performance of pupils, regardless of their background. In no other OECD country are these so-called 'compositional effects' as pronounced as in Switzerland. The different composition of schools is mainly due to segregation between residential quarters and the corresponding school catchment areas. Therefore, the Centre for Democracy Aarau proposed not only to mix pupils according to their social and linguistic origin but also when defining catchment areas, so that the highest possible level of mixing between schools can be achieved. To optimize this process, a novel, detailed algorithm was developed that could be used in the future to support school allocation and classroom planning. The algorithm was trained to

reconstruct the school catchment areas and to survey the social composition at individual schools using the census data of first to third graders in the canton of Zurich. Traffic load data, the network of pavements and footpaths, underpasses and overpasses, were also taken into account. This data could be used to calculate where pupils need to be placed to mix the classes more. At the same time, the capacity of school buildings will not be exceeded and the length of time spent traveling to school will remain reasonable (ZDA 2019).

## Policy, oversight and public debate

### / Switzerland's federal structure as the prevailing circumstance

When reporting on policy in Switzerland, the prevailing federal structure must be emphasized. It is a situation that has already been reflected upon in the previously mentioned ADM examples. Switzerland is a federal state, consisting of 26 highly autonomous Member States (cantons), which in turn grant their municipalities extensive leeway. As a result, the political and public debate on ADM depends greatly on the corresponding government, which cannot be described exhaustively in this report. Furthermore, this fragmentation on policy, regulation, and in research, introduces the risk of working in parallel on overlapping issues, which is also why the confederation strives for advanced coordination as stated below. However, the Federal Government has full responsibility over certain relevant legal fields and political governance, which legally binds all the governments in Switzerland, and thus impacts the entire population. Hence, set out below are those parts of the current federal political debate.

#### Government

At the moment, the role of ADM in society, generally referred to as AI, is predominantly treated as part of a larger discussion on digitization. The Federal Government does not have a specific strategy concerning AI or ADM, but in recent years it launched a "Digital Switzerland Strategy", where all aspects regarding AI will be integrated. More generally, the national legal framework concerning digitization will be adjusted simultaneously through the revision of the Federal Act on Data Protection (FADP).

THE CENTRE FOR DEMOCRACY AARAU PROPOSED NOT ONLY TO MIX PUPILS ACCORDING TO THEIR SOCIAL AND LINGUISTIC ORIGIN BUT ALSO WHEN DEFINING CATCHMENT AREAS, SO THAT THE HIGHEST POSSIBLE LEVEL OF MIXING BETWEEN SCHOOLS CAN BE ACHIEVED.

### / Digital Switzerland

In 2018, and against a background of increasing digitization beyond government services, the confederation launched the "Digital Switzerland Strategy". One focus of this is on current developments in AI (BAKOM 2020). Responsible for the strategy, especially its coordination and implementation, is the "Interdepartmental Digital Switzerland Coordination Group" (Digital Switzerland ICG) with its management unit "Digital Switzerland Business Office" (Digital Switzerland 2020).

As part of the Digital Switzerland Strategy, the Federal Council set up a working group on the subject of AI and commissioned it to submit a report to the Federal Council on the challenges associated with AI. The report was acknowledged by the Federal Council in December 2019 (SBFI 2020). Alongside discussing the most central challenges of AI – those being traceability and systematic errors in data or algorithms – the report details a concrete need for action. It is recognized that all challenges, including this need for action, depend strongly on the subject area in question, which is why the report examined 17 subject areas in greater depth, such as AI in the fields of healthcare, administration, and justice (SBFI 2020 b).

In principle, the challenges posed by AI in Switzerland have, according to the report, already been largely recognized and addressed in various policy areas. Nevertheless, the interdepartmental report identifies a certain need for action which is why the Federal Council has decided on four measures: In the area of international law and on the use of AI in public opinion-forming and decision-making, addition-

al reports will be commissioned for in-depth clarification. Further on, ways of improving coordination, relating to the use of AI in the Federal Administration, will be examined. In particular, the creation of a competence network, with a special focus on technical aspects of the application of AI in the federal administration, will be examined. Finally, AI-relevant policy will be taken into account as an essential component of the “Digital Switzerland” strategy. In this context, the Federal Council has decided that interdepartmental work should continue and that strategic guidelines for the confederation should be developed by spring 2020 (SBFI 2020 c).

In addition, at its meeting on 13 May 2020, the Federal Council decided to create a national Data Science Competence Center. The Federal Statistical Office (FSO) will establish this interdisciplinary center on 1 January 2021. The new center will support the federal administration in implementing projects in the field of data science. To this end, the transfer of knowledge within the Federal Administration as well as the exchange with scientific circles, research institutes, and the bodies responsible for practical application will be promoted. In particular, the center of excellence will contribute to the production of transparent information while taking data protection into account. The reasoning behind the new center is backed up by a statement from the Federal Council, which said that data science is becoming increasingly important, not least in public administration. According to the Federal Council data science includes “intelligent” calculations (algorithms) so that certain complex tasks can be automated (Bundesrat 2020).

## **/ Oversight**

Since the Federal Data Protection Act is no longer up to date, due to rapid technological developments, the Federal Council intends to adapt the FADP to these changed technological and social conditions, and in particular, improve the transparency of data processing and strengthen the self-determination of data subjects over their data. At the same time, this total revision should allow Switzerland to ratify the revised Council of Europe Data Protection Convention ETS 108 and to adopt the Directive (EU) 680/2016 on data protection in the area of criminal prosecution, which it is obliged to do due to the Schengen Agreement. In addition, the revision should bring Swiss data protection legislation as a whole closer to the requirements of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of people regarding the processing of personal data and on the free movement of such

data, and repealing Directive 95/46/EC (GDPR). The revision is currently being discussed in parliament (EJPD 2020).

While the total revision of the existing federal law on data protection will evidently revise the entire Act with all its various aspects, one new provision is of particular interest in regard to ADM. In the case of so-called “automated individual decisions”, there should be an obligation to inform the data subject if the decision has legal consequences or significant effects. The data subject may also request that the decision is reviewed by a human or that he or she should be informed of the logic on which the decision is based. Thereby, a differentiated regulation is provided for decisions by federal bodies. Correspondingly, even though federal bodies must also mark the automated individual decision as such, the possibility that the data subject can request a review by a human may be restricted by other federal laws. In contrast to the EU’s GDPR, there is neither a prohibition of an automated decision, nor is there a claim not to be subject to such a decision. (SBFI 2019)

## **/ Civil Society, Academia and other Organizations**

In addition, a number of fora in Switzerland research, discuss, and work on digital transformation and its opportunities, challenges, needs, and ethics. Most of them address this on a general level, however some address ADM or AI specifically.

## **/ Research Institutes**

Switzerland has a number of well-known and long-established research centers regarding AI technology. These include the Swiss AI Lab IDSIA in Lugano (SUPSI 2020) and the IDIAP Research Institute in Martigny (Idiap 2020) as well as the research centers of the Swiss Federal Institute of Technology in Lausanne (EPFL 2020), and Zurich (EPFL 2020). In addition, private initiatives such as the Swiss Group of Artificial Intelligence and Cognitive Science (SGAICO) complement the academic initiatives, by bringing together researchers and users, promoting knowledge transfer, confidence building, and interdisciplinarity (SGAICO 2020).

## **/ Government Research Funding**

The confederation also addresses the topic of AI via research funding. For example, the Federal Government invests in two national research programs via the Swiss

# *MACHINE LEARNING METHODS ARE NOT USED IN STATE ACTIVITIES IN THE NARROWER SENSE, AS FAR AS CAN BE SEEN.*

National Science Foundation (SNSF) (SNF 2020). Firstly, the National Research Programme 77 “Digital Transformation” (NRP 77) (NRP77 2020) and, secondly, the National Research Programme 75 “Big Data” (NRP 75) (NRP 75 2020). The former examines the interdependencies and concrete effects of digital transformation in Switzerland, and focuses on education and learning, ethics, trustworthiness, governance, the economy, and the labor market (NFP 77 2020). The latter aims to provide the scientific basis for an effective and appropriate use of large amounts of data. Accordingly, the research projects examine questions of the social impact of information technology and address concrete applications (SNF 2020).

Another institute working in this area is the Foundation for Technology Assessment (TA-Swiss). TA-Swiss is a center of excellence of the Swiss Academies of Arts and Sciences, whose mandate is laid down in the Federal Law on Research. It is an advisory body, financed by the public sector, and it has commissioned various studies on AI. The most pertinent of these is a study published on 15 April 2020 on the use of AI in different areas (consumption, work, education, research, media, public administration, and justice). According to the study, a separate law on the use of AI is not considered to be effective. Nevertheless, citizens, consumers, and employees in their dealings with the state, companies or their employer should be informed as transparently as possible about the use of AI. When public institutions or companies use AI they should do so according to clear rules, in an understandable and transparent manner (Christen, M. et al. 2020).

## **/ Digital Society Initiative**

The Digital Society Initiative was launched in 2016. It is a center of excellence at the University of Zurich for critical reflection on all aspects of the digital society. Its goal is to reflect on and help shape the digitization of society, democracy, science, communication, and the economy. In addition, it aims to critically reflect and shape the current change in thinking brought about by digitization in a future-oriented manner and to position the University of Zurich nationally and internationally as a center of excellence for the critical reflection of all aspects of digital society (UZH 2020).

## **/ Digitale Gesellschaft**

The Digitale Gesellschaft (Digital Society) is a non-profit society and broad-based association for citizen and consumer protection in the digital age. Since 2011, it has been working as a civil society organization for a sustainable, democratic and free public sphere and it aims to defend fundamental rights in a digitally networked world. (Digitale Gesellschaft 2020)

## **/ Other organizations**

Several other Swiss organizations are also worth a mention. These organizations concentrate on digitization in general, especially in an economic context, e.g., Swiss Telecommunication Association (asut 2020), digitalswitzerland (Castle 2020), Swiss Data Alliance, and Swiss Fintech Innovations.

## Key takeaways

ADM is used in various parts of the public sector in Switzerland, but these tend not to be in a centralized or comprehensive manner. Only a few cantons use ADM in police work, for example, and the systems used vary. The advantage of such an approach is that cantons or the federal government can benefit from the experience of other cantons. The drawback is that efficiency losses may occur.

There are selective legal foundations, but no uniform ADM law or e-government law or anything similar. Neither is there a specific AI or ADM strategy, but recently attention has been paid to greater coordination, both between different departments at the federal level and between the Federal Government and the cantons. Machine learning methods are not used in state activities in the narrower sense, e.g., in police work or the criminal justice system, as far as can be seen.

Also, at that same level, ADM is used or discussed selectively, but not in a comprehensive manner. In the wider public sector, ADM is used more often and more widely. A good example is a deployment in the Swiss health system, where the Geneva University Hospital became the first hospital in Europe to use ADM to suggest treatments for cancer patients.

## References:

Asut (o. J.): in: asut.ch, [online] <https://asut.ch/asut/de/page/index.xhtml> [30.01.2020]

Bundesamt für Kommunikation BAKOM (o. J.): Digitale Schweiz, in: admin.ch, [online] <https://www.bakom.admin.ch/bakom/de/home/digital-und-internet/strategie-digitale-schweiz.html> [30.01.2020].

Der Bundesrat (o.J.): Der Bundesrat schafft ein Kompetenzzentrum für Datenwissenschaft, In: admin.ch, [online] <https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-79101.html> [15.05.2020].)

Christen, M. et al. (2020): Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz, in: TA-Swiss, [online] <https://www.ta-swiss.ch/themen-projekte-publikationen/informationsgesellschaft/kuenstliche-intelligenz/> [15.05.2020].

Ciritsis, Alexander / Cristina Rossi / Matthias Eberhard / Magda Marcon / Anton S. Becker / Andreas Boss (2019): Automatic classification of ultrasound breast lesions using a deep convolutional neural network mimicking human decision-making, in: European Radiology, Jg. 29, Nr. 10, S. 5458–5468, doi: 10.1007/s00330-019-06118-7.

Digitale Gesellschaft (o. J.): Über uns, in: Digitale Gesellschaft, [online] <https://www.digitale-gesellschaft.ch/uber-uns/> [30.01.2020].



digitalswitzerland (Castle, Danièle Digitalswitzerland (2019): Digitalswitzerland – Making Switzerland a Leading Digital Innovation Hub, in: digitalswitzerland, [online] <https://digitalswitzerland.com> [30.01.2020])

Digital Switzerland (2020): (Ofcom, Federal Office Of Communications (o. J.): Digital Switzerland Business Office, in: admin.ch, [online] <https://www.bakom.admin.ch/bakom/en/homepage/ofcom/organisation/organisation-chart/information-society-business-office.html> [30.01.2020].)

EPFL (o. J.): in: epfl, [online] <https://www.epfl.ch/en/> [30.01.2020]

EJPD (o. J.): Stärkung des Datenschutzes, in: admin.ch, [online] <https://www.bj.admin.ch/bj/de/home/staat/gesetzgebung/datenschutzstaerkung.html> [30.01.2020c].

E-ID Referendum (o. J.): in: e-id-referendum.ch/, [online] <https://www.e-id-referendum.ch> [31.1.2020].

EJPD (o. J.): Stärkung des Datenschutzes, in: admin.ch, [online] <https://www.bj.admin.ch/bj/de/home/staat/gesetzgebung/datenschutzstaerkung.html> [30.01.2020c].

EJPD Eidgenössisches Justiz- und Polizeidepartement (2019): Änderung der Geschwindigkeitsmessmittel-Verordnung (SR 941.261) Automatische Erkennung von Kontrollschildern, in: admin.ch, [online] [https://www.admin.ch/ch/d/gg/pc/documents/3059/Erl\\_Bericht\\_de](https://www.admin.ch/ch/d/gg/pc/documents/3059/Erl_Bericht_de).

EZV (2020): EZV, Eidgenössische Zollverwaltung (o. J.): Transformationsprogramm DaziT, in: admin.ch, [online] <https://www.ezv.admin.ch/ezv/de/home/themen/projekte/dazit.html> [30.01.2020].

ETH Zurich – Homepage (o. J.): in: ETH Zurich – Homepage | ETH Zurich, [online] <https://ethz.ch/en.html> [30.01.2020].

Federal office of public health FOPH (2020): (Health insurance: The Essentials in Brief (o. J.): in: admin.ch, [online] <https://www.bag.admin.ch/bag/en/home/versicherungen/krankenversicherung/krankenversicherung-das-wichtigste-in-kuerze.html> [13.02.2020].)

Geschäft Ansehen (o. J.): in: parlament.ch, [online] <https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaefte?AffairId=20143747> [30.01.2020].

Heinhold, Florian (2019): Hoffnung für Patienten?: Künstliche Intelligenz in der Medizin, in: br.ch, [online] <https://www.br.de/br-fernsehen/sendungen/gesundheit/kuenstliche-intelligenz-ki-medizin-102.html> [30.01.2020].

Idiap Research Institute (o. J.): in: Idiap Research Institute, Artificial Intelligence for Society, [online] <https://www.idiap.ch/en> [30.01.2020]

Der IPV-Chatbot – SVA St.Gallen (o. J.): in: svasg.ch, [online] <https://www.svasg.ch/news/meldungen/ipv-chatbot.php> [30.01.2020].

Leese, Matthias (2018): Predictive Policing in der Schweiz: Chancen, Herausforderungen Risiken, in: Bulletin zur Schweizerischen Sicherheitspolitik, Jg. 2018, S. 57–72.

Lindner, Martin (2019): KI in der Medizin: Hilfe bei einfachen und repetitiven Aufgaben, in: Neue Zürcher Zeitung, [online] <https://www.nzz.ch/wissenschaft/ki-in-der-medizin-hilfe-bei-einfachen-und-repetitiven-aufgaben-id.1497525?reduced=true> [30.01.2020]

Medinside (o. J.): in: Medinside, [online] <https://www.medinside.ch/de/post/in-genf-schlaegt-der-computer-die-krebsbehandlung-vor> [14.02.2020].

NRP 75 Big Data (o. J.): in: SNF, [online] <http://www.snf.ch/en/researchinFocus/nrp/nfp-75/Pages/default.aspx> [30.01.2020].

NFP [Nr.] (o. J.): in: nfp77.ch, [online] <http://www.nfp77.ch/en/Pages/Home.aspx> [30.01.2020]

NRP 75 Big Data (o. J.): in: SNF, [online] <http://www.snf.ch/en/researchinFocus/nrp/nfp-75/Pages/default.aspx> [30.01.2020].

NFP [Nr.] (o. J.): in: nfp77.ch, [online] <http://www.nfp77.ch/en/Pages/Home.aspx> [30.01.2020]

Ringeisen, Peter / Andrea Bertolosi-Lehr / Labinot Demaj (2018): Automatisierung und Digitalisierung in der öffentlichen Verwaltung: digitale Verwaltungsassistenten als neue Schnittstelle zwischen Bevölkerung und Gemeinwesen, in: Yearbook of Swiss Administrative Sciences, Jg. 9, Nr. 1, S. 51–65, doi: 10.5334/ssas.123.

ROSNET > ROS allgemein (o. J.): in: ROSNET, [online] <https://www.rosnet.ch/de-ch/ros-allgemein> [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (o. J.): Künstliche Intelligenz, in: admin.ch, [online] <https://www.sbf.admin.ch/sbf/de/home/das-sbf/digitalisierung/kuenstliche-intelligenz.html> [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (2019): Herausforderungen der künstlichen Intelligenz – Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat, in: admin.ch, [online] <https://www.sbf.admin.ch/sbf/de/home/das-sbf/digitalisierung/kuenstliche-intelligenz.html> [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (o. J.): Künstliche Intelligenz, in: admin.ch, [online] <https://www.sbf.admin.ch/sbf/de/home/das-sbf/digitalisierung/kuenstliche-intelligenz.html> [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (2019): Herausforderungen der künstlichen Intelligenz – Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat, in: admin.ch, [online] <https://www.sbf.admin.ch/sbf/de/home/das-sbf/digitalisierung/kuenstliche-intelligenz.html> [30.01.2020].

Schaffhauser eID+ – Kanton Schaffhausen (o. J.): in: sh.ch, [online] <https://sh.ch/CMS/Webseite/Kanton-Schaffhausen/Beh-rde/Services/Schaffhauser-eID--2077281-DE.html> [30.01.2020].

SGAICO – Swiss Group for Artificial Intelligence and Cognitive Science (2017): in: SI Hauptseite, [online] <https://swissinformatics.org/de/gruppierungen/fg/sgaico/> [30.01.2020]

SNF, [online] <http://www.snf.ch/en/Pages/default.aspx> [30.01.2020]

Srf/Blur;Hesa (2017): Wie «Precobs» funktioniert – Die wichtigsten Fragen zur «Software gegen Einbrecher», in: Schweizer Radio und Fernsehen (SRF), [online] <https://www.srf.ch/news/schweiz/wie-precobs-funktioniert-die-wichtigsten-fragen-zur-software-gegen-einbrecher>

SUPSI – Dalle Molle Institute for Artificial Intelligence – Homepage (o. J.): in: idsia, [online] <http://www.idisia.ch> [30.01.2020].

Swissdataalliance (o. J.): in: swissdataalliance, [online] <https://www.swissdataalliance.ch> [30.01.2020].

Swiss Fintech Innovations (SFTI) introduces Swiss API information platform (2019): in: Swiss Fintech Innovations – Future of Financial Services, [online] <https://swissfintechinnovations.ch> [30.01.2020]].

Schwerzmann, Jacqueline Amanda Arroyo (2019): Dr. Supercomputer – Mit künstlicher Intelligenz gegen den Krebs, in: Schweizer Radio und Fernsehen (SRF), [online] <https://www.srf.ch/news/schweiz/dr-supercomputer-mit-kuenstlicher-intelligenz-gegen-den-krebs>

Treuthardt, Daniel / Melanie Kröger (2019): Der Risikoorientierte Sanktionenvollzug (ROS) – empirische Überprüfung des Fall-Screening-Tools (FaST), in: Schweizerische Zeitschrift für Kriminologie, Jg. 2019, Nr. 1–2, S. 76–85.; (Treuthardt, Daniel / Melanie Kröger / Mirjam Loewe-Baur (2018): Der Risikoorientierte Sanktionenvollzug (ROS) – aktuelle Entwicklungen, in: Schweizerische Zeitschrift für Kriminologie, Jg. 2018, Nr. 2, S. 24–32.

ZDA (2019): Durchmischung in städtischen Schulen, in: zdaarau.ch, [online] <https://www.zdaarau.ch/dokumente/SB-17-Durchmischung-Schulen-ZDA.pdf> [30.01.2020].

BELGIUM  
DENMARK  
ESTONIA  
FINLAND  
FRANCE  
GERMANY  
GREECE  
ITALY  
NETHERLANDS  
POLAND  
PORTUGAL  
SLOVENIA  
SPAIN  
SWEDEN  
SWITZERLAND



# UNITED KINGDOM STORY

PAGE 272

# RESEARCH

PAGE 277





Find out more on p. 281 in the research chapter under "Challenging face recognition".

# **Brexit:** How EU nationals navigate the automated checks of the "settled status" program

 The British Home Office deployed a mobile app so that EU citizens can apply for permission to stay in the UK. Some are confident about the partly automated process; others live Kafkaesque experiences.

By Hugo Barbieux



Pablo Duro, a 28-year-old Spaniard, is a PhD student and member of staff at Manchester Metropolitan University. He started his research work in October 2019. After applying to stay in the UK, he was notified that his status was “pre-settled”. To obtain “settled status”, which would allow him to remain in the UK for good, he would need to be resident in the country for five years. He said that remaining a European citizen in the UK was not a concern for him even though Brexit preparations were already in full swing. Being an immigrant was no big deal, he said.

Mr. Duro first heard about the settlement procedure and the mobile app it involves during a chat with another Spanish citizen. He visited the Home Office website and followed the steps. He had to complete a form, download a mobile application, scan his passport, and upload a photograph of his face in order to verify his identity. He said the process was very quick. Within five days, he received a notification in his mailbox confirming that he had been granted the pre-settled status.

Mr. Duro said the process was clear and concise. He believed there was no risk in sharing his data via the app as an official institution created it.

## / Failed attempts

The Tilbrook family, in Newcastle, would certainly envy such an experience. The grandmother of the family, an 80-year-old who requested anonymity before being interviewed, found the settlement process distressing.

The elderly Danish woman worked in paid employment in England for three years from 1961 to 1964 before working at home, like many women of her generation, Jane, her daughter, explained. The 80-year-old asked her son-in-law to download the app and fill out the application form for her.

The first attempt failed – her legitimate claim for settled status was refused – and she received a “temporary status”. Her daughter and son-in-law shared the experience on social media and contacted the administration to understand why their elderly relative’s application had failed. They had to upload additional documents, following which Jane Tilbrook’s mother was granted settled status.

Highly dissatisfied with the process, Mr. Tilbrook, 60, said his mother-in-law was tech-savvy for her age – she uses a smartphone – but this operation was “way beyond her”. He

said using near field communication, or NFC, to upload the details of a passport chip is beyond the abilities of most people he knew, let alone an 80-year-old.

He explained that she would have been unable to upload the required documents. When he did it, the first attempt took him two hours. The second attempt was even longer.

Mr. Tilbrook described the process as “very intimidating”. His mother-in-law, reluctantly, had to share private information, and, like many women of her age, most of the required documents were written in her husband’s name. Understandably, she wanted a written document confirming her status, but the app only provides an electronic confirmation.

The Tilbrook’s case is not isolated. Many similar examples can be found on social media. For example, the Twitter account [“EU Settled Status Horror Stories”](#) lists tweets from EU nationals who have complained about the process.

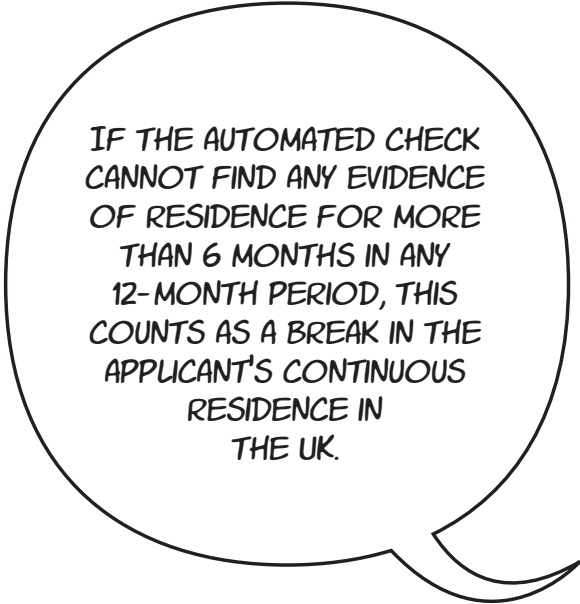
## / Beta tests

By the end of the public test of the app, on 21 January 2019, the system had received over 200,000 applications from nationals of all 27 EU countries. Then, on 30 March 2019, the Home Office officially launched the simple, free system.

In the [“Processing Level Memorandum of Understanding”](#) document, the Home Office justifies the data sharing process and automated checks they use. The document states that the algorithm used was developed to support the EU Settlement Scheme and manage the influx of applications as the UK leaves the EU. According to the document, this approach improves the user experience, while reducing fraud, errors, case processing time, and excessive paperwork.

Applicants who downloaded the [“EU Exit: ID document check”](#) app are required to complete five key steps: firstly, they take an image of the personal details page of either their passport or identity card. Then, they must scan this document, via a smartphone, by using the NFC feature. The smartphone then reads the contents of the electronic chip stored inside modern identity documents. The applicant must then take a video and a photo of their face. Finally, she or he must answer an online questionnaire. Additional documentation may be required later in the process.

However, this is not the end of the journey. Applicants are invited to provide their National Insurance (NI) number to



IF THE AUTOMATED CHECK CANNOT FIND ANY EVIDENCE OF RESIDENCE FOR MORE THAN 6 MONTHS IN ANY 12-MONTH PERIOD, THIS COUNTS AS A BREAK IN THE APPLICANT'S CONTINUOUS RESIDENCE IN THE UK.

verify their residence history in the UK. This is not a mandatory step, but official documentation says that applicants who submit their NI number will not have to provide further evidence. This step uses an algorithm-based verification process.

Automated processing of tax and other benefits records is only carried out when an applicant has provided their NI number. This automated check is done against records held by the tax authority (HMRC) and the Department for Work and Pensions (DWP).

A Home Office document titled [“How the automated check calculates the period of residence”](#) mentions that, if the automated check finds evidence of UK employment or some benefits in a month, then that month counts as a month of continuous residence. If the automated check cannot find any evidence of residence for more than 6 months in any 12-month period, this counts as a break in the applicant's continuous residence in the UK. If this happens, the applicant is told which time periods this applies to.

But not everyone has an NI number. And the online process has caused several problems (such as email providers erroneously marking official messages as spam), some of which were spotted during beta trials and subsequently fixed. At the end of this process, a caseworker looks at the application and makes the final decision.

Alternatively, applicants can send their identity documents by post or visit their local authority in person to have their passport or biometric residence card checked manually.

## / Not so simple

Settled.org is a UK charity dedicated to guide, inform, and assist those vulnerable and hard-to-reach EU citizens who may be at risk of not having the authorization to remain in the UK by the time the EU Settlement Scheme ends. They say the EU Settlement Scheme is not as “simple and straightforward” as the government has stated.

The Home Office uses a surprisingly reassuring tone to explain the app, considering that Mr. Tilbrook said he did need to enter an NI number for his mother-in-law, adding that he was not informed of an alternative to using the app. “The algorithm simply doesn't fit,” he said.

Settled.org strongly recommends EU citizens to begin applying for settled status well before the application period ends on 30 June 2021. Applicants with atypical profiles may find it challenging to obtain the right status, a spokesperson for the charity said.

Lily Beurrier, a French citizen who moved to the UK in 2003, applied for settled status in June 2019, two years before the official end of the process. She said she was fired from her work and was worried that this could impact her chances of obtaining settled status.

She said she is confident using IT software and she thought the process would be quite straightforward. She said that using the app to submit her application was fairly easy, yet sometimes frustrating. For example, it took a couple of attempts to scan her passport before it worked. However, the most frustrating part, she said, was to get the right light in the room so that the app could recognize her passport and accept her photo.

## / “Application denied”

The process took about ten minutes, but then she received her result: “application denied”.

“I never thought this could have happened to me,” she said. “I have been living in the UK for over 15 years with the same amount of years employed by the same employer. It was a real shock.”

In vain, she phoned the government helpline. Then, she complained on social media, which is when the BBC contacted her. Her subsequent interview, and contacts with journalists, allowed things to move forward slightly.

The Home Office asked her to provide further evidence of residence. Her employer wrote a letter confirming her 15 years of continuous employment in the UK.

Interestingly, she said that she was advised to use her computer rather than the app to upload the letter. A couple of weeks later, she was granted settled status.

Ms. Beurrier said this was a very challenging and upsetting process and that she felt “cheated” by the app. For exam-

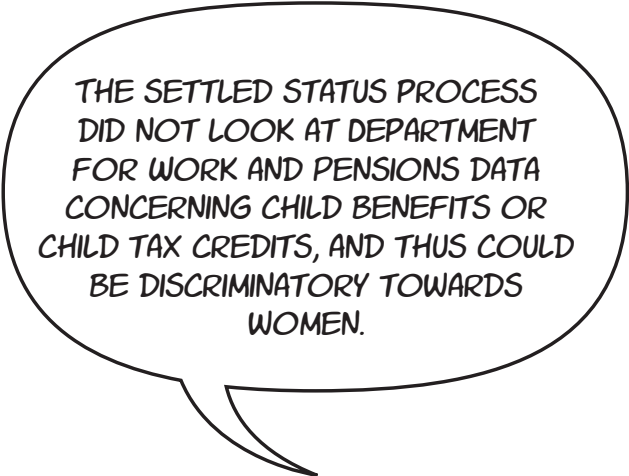
ple, the questionnaire does not consider the social aspect of her application. “How could an app or automated platform decide on my future in the UK? I am married to an English man and both our children were born in the UK.”

### **/ Incompatible devices**

Lady-Gené Waszkewitz’s story also defuses the theory of the age gap between young tech-savvy students who can easily deal with apps, and elderly people who might be confused by smartphones. The 28-year-old German moved to the UK in 2011 as a pupil barrister. In total, she studied for five years in the UK. But she, too, struggled to obtain settled status.

***“I NEVER THOUGHT THIS COULD  
HAVE HAPPENED TO ME. I HAVE  
BEEN LIVING IN THE UK FOR  
OVER 15 YEARS WITH THE SAME  
AMOUNT OF YEARS EMPLOYED  
BY THE SAME EMPLOYER. IT WAS  
A REAL SHOCK.”***

***LILY BEURRIER***



THE SETTLED STATUS PROCESS DID NOT LOOK AT DEPARTMENT FOR WORK AND PENSIONS DATA CONCERNING CHILD BENEFITS OR CHILD TAX CREDITS, AND THUS COULD BE DISCRIMINATORY TOWARDS WOMEN.

She said the app was not compatible with all devices, meaning she needed to take time out of her workday and go to Southwark Council to have her passport scanned to complete this first part of the process.

Then, she had to photograph five years of bank statements, council tax exemption forms, enrolment certificates, tenancy, and employment contracts, and forms to account for five continuous years of living in the UK - the minimum amount of time needed to obtain settled status.

She said the process was not at all straightforward because she, like many others, did not have all these documents to hand and she had to contact the relevant agencies and authorities to obtain them – including her university overseas.

“I did not feel particularly overwhelmed by the process but rather annoyed at having to obtain this information which should be in the government’s hands through my National Insurance number, for example,” she said.

## / Possible discrimination

Lawyers instructed over the summer by The Legal Education Foundation (TLEF) wrote in a public opinion that the settled status process did not look at Department for Work and Pensions (DWP) data concerning child benefits or child tax credits, and thus could be discriminatory towards women. According to their report, 87% of child benefit recipients were female.

Ms. Beurrier confirmed that she was indeed a recipient of child tax credit for her children. It is very likely that the automated checks could not find evidence of residence for more than six months in any 12-month period. “I never thought to mention or refer back to the child tax credit,” she said.

Furthermore, the lawyers pointed out that the process of proving documentary evidence is far from easy and that the Home Office gave very few clues as to what type of document would better support an application. Connected to this idea, Mr. Tilbrook said the process does not take pensions into account. His mother-in-law receives the basic pension. “They know she’s been in the UK,” he said.

Despite these woes, the Home Office published satisfactory figures. Public data released puts the total number of applications received, up to 31 January 2020, at over 3.1 million. 88% of those applications have been processed: 58% were granted settled status, 41% were granted pre-settled status, and seven applications were refused on suitability grounds. However, given that there are no precise figures for the number of EU citizens living in the UK, it is too early to call the scheme a success.

# Research

By Tom Wills and Fabio Chiusi


## Contextualization

Last year, the United Kingdom was notable for its wide range of policy and research initiatives looking at the topic of automated decision-making (ADM) across government and civil society. These fell broadly into two categories: those driven by the hype around Artificial Intelligence (AI), which sought mainly to promote the supposed economic potential of the technology, with ethics as a secondary consideration; and critical initiatives sounding the alarm about the possible adverse impacts of ADM, especially on minority groups.

Institutional responses to ADM generally remain in the initial stages of research and consultation. However, a number of themes are already emerging, both in the identification of risks associated with ADM and in the broad approach to how to moderate the impacts of this technology. In particular, the possibility of discriminatory bias has been raised in relation to policing, public services, and other areas of civic life. The need for explicability and accountability in automated decision-making has been raised in a number of reports. So too has the current lack of frameworks and guidelines for organizations to follow in their evaluation of AI. This is a crucial time for the development of Britain's regulatory approach to ADM, with several official organizations collecting their final evidence before launching more concrete policy recommendations.

Public interest and controversy around several specific ADM implementations has noticeably increased. This is particularly so in the case of face recognition. Attentive journalists have investigated and drawn attention to the roll out of this technology across high-profile public spaces in London, which would otherwise be happening silently and without oversight. The civil society response has been strong enough to halt the deployment of face recognition in the case of the King's Cross estate.

The spectacular fiasco, in August 2020, of the algorithm deployed by the Office of Qualifications and Examinations Re-



PUBLIC INTEREST AND  
CONTROVERSY AROUND SEVERAL  
SPECIFIC ADM IMPLEMENTATIONS  
HAS NOTICEABLY INCREASED.

gulation (Ofqual) to grade A-level<sup>1</sup> and GCSE<sup>2</sup> students also received mainstream attention, [leading](#) the government to backtrack and resort to human judgment, and to protestors both in Parliament Square and outside the Department for Education. Their motto? "Fuck the algorithm".

Brexit is however the all-consuming political question, cutting across all kinds of issues, and touching on ADM in a number of ways. EU citizens living in the UK have been told to register their residency in the country through a smartphone app powered by automated face recognition and automated data matching across government departments. Meanwhile, the debate on how to facilitate the transit of goods and people across the border between Northern Ireland and the Republic of Ireland in the event of leaving the European Single Market has raised the question of whether [automated decision-making could be used](#) (Edgington, 2019) to create a virtual border.

However, there remains little public awareness of automated decision-making as an issue in itself. While there is growing discourse around 'Artificial Intelligence', including criti-

1 "A levels are British educational qualifications which school children take when they are seventeen or eighteen years old. People usually need A levels if they want to go to university in Britain." (<https://www.collinsdictionary.com/dictionary/english/a-level>)

2 "GCSE is the qualification taken by 15 and 16 year olds to mark their graduation from the Key Stage 4 phase of secondary education in England, Northern Ireland and Wales." (<https://www.bbc.co.uk/bitesize/levels/z98jmp3>)



cal coverage in the popular technology press, the framing of AI as an issue does not always correlate with the questions that need to be asked of ADM systems. The lack of a public register of automated decision-making systems is a problem that has been identified by some of the policy responses, which would help facilitate public understanding.

## A catalog of ADM cases

### / Brexit: Settled status applications from EU citizens

As part of preparations for Brexit, the UK government launched a scheme to establish the residency rights of the estimated 3.4 million EU citizens living in the country.

Until Brexit, EU nationals have been free to enter the UK, and work and live in the country indefinitely without any need to apply for a visa or register as a foreigner.

Although at the time of writing, the terms of Britain's exit from the EU are still up in the air, the government's policy has been that EU citizens who enter the UK before Brexit must apply for status as "settled" or "pre-settled" residents in order to have the right to stay. Those with at least five years of residency are supposed to receive "settled" status, while others will be given "pre-settled" status and then required to apply for "settled" status when they reach the five-year threshold.

Given that there was no system to record when EU nationals entered the UK or became resident in the country, the process of establishing that someone came to the UK before Brexit is reliant on a combination of records from various government agencies, which may differ from person to person.

The government implemented the registration system in the form of a smartphone app. No doubt aiming to streamline the process as much as possible and control the costs of this huge bureaucratic exercise, the government has employed a number of elements of automated decision-making.

In October 2019, the Guardian reported on some of the [challenges and problems of the scheme](#) (Gentleman, 2019).

It mentioned how, after collecting the user's details, the app attempted to confirm their history of residency by "making instant checks with HMRC", (Her Majesty's Revenue and Customs, the UK tax office). It described how "the app checks identity through facial matching in the same way travellers are checked in airports at e-passport gates." The newspaper article also noted that "although much of the system is automated, every case is checked by a human before a decision is made."

After a deadline, which may arrive in December 2020 or June 2021, the Guardian pointed out, "those who have failed to apply will gradually be transformed from legal residents into undocumented migrants [...] they will gradually discover that normal life has become impossible."

Swee Leng Harris, a law expert, researched the automated aspects of the settled status app for an article to be published in the forthcoming [Data & Policy](#) (Harris, forthcoming) journal.

Harris wrote that where the system found only a "partial match" with tax or welfare benefit records, the applicant would be granted pre-settled status as a default unless they challenged the decision.

According to government statistics analyzed for the article, although only a small number of appeals had been processed by the time the figures were provided (253 cases as of 31 May 2019), around 91% had been decided in favor of the applicant.

Harris argues this raises the question of how many people were given pre-settled status when they had, in reality, fulfilled the requirements for settled status but did not challenge the decision.

### / "Racist" algorithm for visa applicants suspended before legal challenge

In August 2020, the UK government's Home Office abandoned an ADM system to determine the risk represented by visa applicants. Commonly referred to as "Streaming tool", the algorithm had been subject to legal challenge by the [Joint Council for the Welfare of Immigrants \(JCWI\)](#) and the non-profit organization Foxglove.

[According](#) to them, the algorithm "entrenched racism and bias into the visa system", as it "suffered from 'feedback

loop' problems known to plague many such automated systems – where past bias and discrimination, fed into a computer program, reinforce future bias and discrimination". This meant that the system "was prone to fast-tracking applicants from predominantly white countries and sorting people from other nations into a longer and more onerous review process", as [Motherboard](#) puts it.

As a result, the Home Office said it will redesign the ADM system, this time explicitly "including issues around unconscious bias and the use of nationality, generally."

## **/ Welfare benefit applications**

Fraud in the welfare system is a topic beloved of the country's popular press. The woman who [falsely claimed the man she lived with was not her partner](#) (Mutch, 2019) in order to qualify for benefits worth £30,000 and the Brit who claimed £90,000 from various funds [despite living in Spain](#) (Mahmood, 2019) are just two examples of "benefit cheats" who have made the newspapers recently.

This is part of the reason the government is keen to be seen to detect fraudulent applications as efficiently as possible. To that end, people who need to claim benefits because their income is not enough to pay the costs of housing are increasingly subject to automated decision-making, in an

other example detailed by Swee Leng Harris in a [forthcoming article](#) (ibid.) in Data & Policy.

The system is called Risk Based Verification (RBV), Harris writes, and in some parts of the country uses a propensity model in order to assign claims to a risk group, on which basis they may be subject to more stringent manual checks, such as documentation requirements or home visits by officials.

Harris points out that there are no requirements for local government offices, which administer the benefits, to monitor whether their algorithms are disproportionately impacting people with protected characteristics (such as belonging to a minority ethnic group).

An investigation in the Guardian identifies two companies which it says supply RBV systems to about 140 local councils in the UK. One is a company called TransUnion, a US credit-rating business. The other is Xantura, a British firm that focuses on providing analytics systems to local authorities.

The Guardian reported that [several councils had abandoned RBV systems after reviewing their performance](#) (Marsh, 2019). North Tyneside, for example, found that the TransUnion system did not provide an explanation for flagging welfare applications as 'high-risk', and additional

***THOSE WHO HAVE FAILED TO APPLY  
WILL GRADUALLY BE TRANSFORMED  
FROM LEGAL RESIDENTS INTO  
UNDOCUMENTED MIGRANTS [...] THEY  
WILL GRADUALLY DISCOVER THAT  
NORMAL LIFE HAS BECOME IMPOSSIBLE.***

checks were not revealing any reason to reject the claims. The warnings were simply leading to valid claims being delayed. By contrast, [the review](#) (NTC, 2019) said that access to government tax data, rather than RBV, had helped speed up verification of applications.

North Tyneside Council had implemented RBV on [advice issued by the Department for Work and Pensions](#) (DWP 2011), the central government ministry responsible for overseeing welfare benefits, which had already started using the system to evaluate applications for unemployment benefit, which is administered on a nationwide basis.

The [council's own evaluation of RBV](#) (NTC, 2015) recorded that “In researching the use of RBV, there have been no issues raised with regard to human rights,” suggesting that the implementation of this automated decision-making technology spread under the radar of public scrutiny. It did, however, identify the “possibility that people with certain protected characteristics may be under or over represented in any of the risk groups,” and promised to monitor whether the system ended up disadvantaging particular groups.

The evaluation also noted that the cost of the “software and the manpower” required for RBV would be borne by Cofely GDF Suez, the private-sector outsourcing firm (now known as Engie) which had been given the contract to process welfare applications in North Tyneside. This raises the question of whether the company also stood to gain from the expected cost savings.

## / The grading algorithm fiasco

With schools closed since March due to the COVID-19 outbreak, A level and GCSE students in England, Wales, and Northern Ireland could not take their exam. In response, the government decided to entrust their marks to an algorithm.

The mathematical model was supposed to put grades and rankings estimated by teachers into historical perspective, by [taking into account](#) grades previously achieved by students and the distribution of grades across the country. This way, [argued](#) the government, grades would have been consistent with past results.

The opposite, however, ensued. “When A-level grades were announced in England, Wales and Northern Ireland on 13 August, nearly 40% were lower than teachers’ assess-

ments”, wrote the BBC, reporting on official figures. What the search for historical consistency actually produced was downgraded marks for bright students from historically underperforming schools<sup>3</sup>.

This also produced discriminatory results, largely benefiting private schools at the expense of public education. As the [BBC](#) clearly puts it: “Private schools are usually selective – and better-funded – and in most years will perform well in terms of exam results. An algorithm based on past performance will put students from these schools at an advantage compared with their state-educated equivalents”.

The UK’s Data Protection Authority, ICO, immediately issued a [statement](#) that is highly relevant for the use of ADM in this context. “The GDPR places strict restrictions on organisations making solely automated decisions that have a legal or similarly significant effect on individuals”, it reminded, before adding that “the law also requires the processing to be fair, even where decisions are not automated”.

The case rapidly reached mainstream attention. As a consequence, protests [erupted](#) in Parliament square and outside the Department of Education, while the government – after similar moves in Scotland, Wales, and Northern Ireland – quickly [reversed course](#), scrapping the algorithm and resorting to the teachers’ judgment instead. The chief civil servant at the Department of Education, Jonathan Slater, was also [sacked](#) as a result.

Prime Minister Boris Johnson eventually [blamed](#) the fiasco on the “mutant algorithm”, but many commentators saw this as a [policy, not technology, debacle](#), one that should be blamed on human decision-makers rather than “algorithms”, and that should remind us that they are [never neutral](#) instead. The Ada Lovelace Institute also [warned](#) that “the failure of the A-level algorithm highlights the need for a more transparent, accountable and inclusive process in the deployment of algorithms, to earn back public trust”.

An even more important consequence is that, after the Ofqual algorithm fiasco, councils started to “quietly” scrap “the use of computer algorithms in helping to make decisions on benefit claims and other welfare issues”, the Guardian [revealed](#). Some 20 of them “stopped using an algorithm to flag claims as “high risk” for potential welfare fraud”, the newspaper wrote.

3 A detailed explanation of the algorithm’s composition and functioning is provided by the Guardian, <https://www.theguardian.com/education/2020/aug/21/ofqual-exams-algorithm-why-did-it-fail-make-grade-a-levels>

## / Challenging face recognition

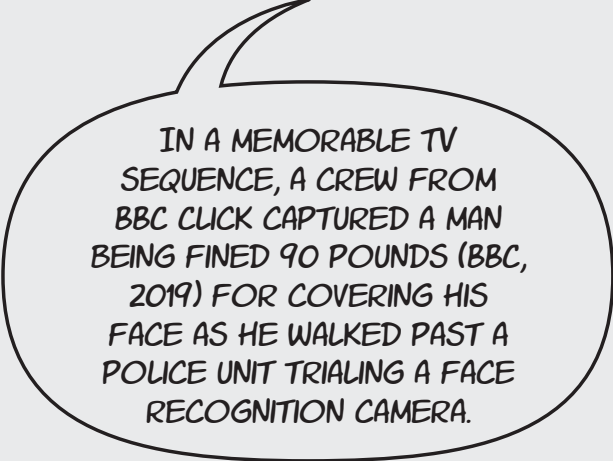
In last year's report, we noted that police forces had begun trialing the use of automatic face recognition systems to automatically decide when people should be apprehended as suspects. This development was challenged by the civil society organization Big Brother Watch, which monitored the use of face recognition at Notting Hill Carnival in London and observed a high number of false positives.

Controversies around face recognition have increased in media profile significantly in the last year. In particular, there was an outcry over the use of face recognition in the King's Cross area of central London.

It is the site of one of London's main railway stations, but the formerly run-down district has in recent years been re-developed and now also features an art school, corporate offices, shops, and restaurants and – crucially – what appear to the casual visitor to be public squares, walkways, and piazzas.

In August, the Financial Times revealed in an [exclusive investigation](#) (Murgia, 2019) that Argent, a property developer, was using the technology to track tens of thousands of people as they went about their business across the 67-acre King's Cross estate, which it owns.

The story was picked up by the [BBC](#) (Kleinman, 2019) and [The Guardian](#) (Sabbagh, 2019) and sparked a public debate. Within a matter of weeks, the pressure forced the company to switch off the system and declare that it had no further plans to use any form of face recognition technology in the future.



IN A MEMORABLE TV SEQUENCE, A CREW FROM BBC CLICK CAPTURED A MAN BEING FINED 90 POUNDS (BBC, 2019) FOR COVERING HIS FACE AS HE WALKED PAST A POLICE UNIT TRIALING A FACE RECOGNITION CAMERA.

In the course of the scandal, it emerged that the Metropolitan Police had [provided a database of images of persons of interest](#) (Murgia, 2019b) for Argent to load into the system.

The Information Commissioner, who is responsible for enforcing data protection laws, has [launched an investigation](#) (Rawlinson, 2019) into the King's Cross system.

Other examples of face recognition systems also made the news. In a memorable TV sequence, a crew from BBC Click [captured a man being fined £90](#) (BBC, 2019) for covering his face as he walked past a police unit trialing a face recognition camera.

London's Metropolitan Police reported their first arrest using face recognition in February. According to newspaper reports (Express 2020), a woman wanted for allegedly assaulting police officers after an altercation at a London shopping mall in January was identified by the system walking down Oxford Street three months later.

The arrest came just days after the head of the force, Cressida Dick, had [publicly made the argument for algorithmic policing at a conference](#) (Dick, 2020).

"If, as seems likely, algorithms can assist in identifying patterns of behaviour by those under authorised surveillance, that would otherwise have been missed, patterns that indicate they are radicalising others or are likely to mount a terrorist attack; if an algorithm can help identify in our criminal systems material a potential serial rapist or killer that we could not have found by human endeavour alone; if a machine can improve our ability to disclose fairly then I think almost all citizens would want us to use it," she was [reported by ITV News as saying](#) (ITV, 2020).

A legal challenge to police face recognition backed by Liberty, the civil rights group, was [rejected](#) (Bowcott, 2019) by the High Court in 2019. The group is [appealing the decision](#) (Fouzder, 2019).

Researchers from the University of Essex, meanwhile, [reviewed police trials of face recognition](#) (Fussey, 2019) in the UK and found that out of 42 matches made, only 8 could be confidently confirmed as accurate by human review.

# Policy, oversight and debate

## / Policy

### / Parliament

The All-Party Parliamentary Group on Artificial Intelligence (APPG AI) is a cross-party group of MPs with a special interest in AI, founded in 2017. It is sponsored by a number of large companies including the 'Big Four' accountancy firms – Deloitte, Ernst & Young, KPMG and PricewaterhouseCoopers – plus Microsoft and Oracle. Its latest program, published in summer 2017, is called "[Embracing the AI Revolution](#)" (APPG AI, 2019). The introduction explains: "We spent the first two years understanding key economic and socio-ethical implications [of AI] and we are now ready to take action." The four pillars of the group's work from 2019-2020 are: Education, Enterprise Adoption of AI, Citizen Participation, and Data Governance.

Its planned outputs include an 'AI 101' course and lesson plan to be used in schools; awards to be given to companies leading in the adoption of AI; and a public survey on how AI should be regulated. Its programme does not specifically address the ethical considerations around ADM, although it refers to the need to "ensure all AI developers and users become ethical AI citizens".

### / Office for Artificial Intelligence

The Office for AI is now up and running. In June 2019, it published guidance in conjunction with the Government Digital Service (GDS) and the Alan Turing Institute (ATI) called "[A guide to using artificial intelligence in the public sector](#)" (GDS, 2019). This guidance is supposed to help civil servants decide whether AI could be useful and how to implement it ethically, fairly, and safely.

### National Data Strategy

In July 2019, the government launched a call for evidence to build its [National Data Strategy](#) (DCMS, 2019). Its stated aim is to "support the UK to build a world-leading data economy". As part of the call for evidence, it highlighted the need to consider "fairness and ethics" in how data is used. One of its priorities is "To ensure that data is used in a way that people can trust". Following the election in December 2019,

the government is expected to publish a draft strategy and run a second phase of evidence-gathering, in the form of a full public consultation on the draft strategy.

## / Oversight

### / Information Commissioner's Office

The Information Commissioner's Office (ICO) is the UK's data protection regulator, funded by the government and directly answerable to parliament. It oversees and enforces the proper use of personal data by the private and public sectors.

The ICO has issued [detailed guidance](#) (ICO, 2018) on how organizations should implement the requirements of the General Data Protection Regulation (GDPR) in regards to automated decision-making and profiling. It has also appointed a researcher working on AI and data protection on a two-year fixed term to research and investigate a [framework for auditing algorithms](#) (ICO, 2019), including public consultation.

### / Centre for Data Ethics & Innovation (CDEI)

The CDEI is a body set up by the UK government to investigate and advise on the use of "data-driven technology". It describes itself as an "independent expert committee". Although the body has some independence from the government department that hosts it, critics have pointed out what they say is the "[industry-friendly](#)" (Orlowski, 2018) make-up of its board.

Nonetheless, the body has published several reports advocating more regulation of algorithms since its establishment in 2018.

In February 2020, CDEI published a report on online targeting – the personalization of content and advertising shown to web users using machine learning algorithms. While the report recognized the benefits of this technology, helping people navigate an overwhelming volume of information and forming a key part of online business models, it said targeting systems too often operate "without sufficient transparency and accountability."

Online targeting translated to "significant social and political power", the report said, that could influence people's



perception of the world, their actions and beliefs, and their ability to express themselves.

The report stopped short of recommending specific restrictions on online targeting. Rather, it advocated “regulatory oversight” of the situation under a new government watchdog. The watchdog would draw up a code of practice requiring platforms to “assess and explain the impacts of their systems”. It would also have legal powers to request “secure access” to platforms’ data in order to audit their practices.

CDEI’s [interim report on bias in ADM](#) (CDEI, 2019) was published in summer 2019. This noted the importance of collecting data on personal characteristics such as gender and ethnicity so that algorithms can be tested for bias. However, it acknowledged “some organisations do not collect diversity information at all, due to nervousness of a perception that this data might be used in a biased way”.

It also identified a lack of knowledge of “the full range of tools and approaches available (current and potential)” to combat bias.

Finally, it discussed the challenges of fixing algorithmic systems once bias has been identified: “Humans are often trusted to make [value judgements and trade-offs between competing values] without having to explicitly state how much weight they have put on different considerations. Algorithms are different. They are programmed to make trade-offs according to unambiguous rules. This presents new challenges.”

A final report was due to follow in March 2020, but was postponed due to the COVID-19 crisis and is expected to be published by the end of 2020.

The Royal United Services Institute (RUSI), a respected independent defense and security think tank, was commissioned by CDEI to produce a [report on algorithmic bias in policing](#) (RUSI, 2019) which was published in September 2019. It confirmed the potential for “discrimination on the grounds of protected characteristics; real or apparent skewing of the decision-making process; and outcomes and processes which are systematically less fair to individuals within a particular group” in the area of policing. It emphasized that careful consideration of the “wider operational, organisational and legal context, as well as the overall decision-making process” informed by ADM was necessary in addition to scrutinizing the systems themselves. And it highlighted “a lack of organisational guidelines or clear pro-

cesses for scrutiny, regulation and enforcement for police use of data analytics.”

Police forces will need to consider how algorithmic bias may affect their decisions to police certain areas more heavily, the report found. It warned of the risk of over-reliance on automation and said discrimination claims could be brought by individuals on the basis of age or gender discrimination.

CDEI is also working with the government’s Race Disparity Unit to investigate the risk of people being [discriminated against according to their ethnicity](#) (Smith, 2019) in decisions made in the criminal justice system.

## / Committee on Standards in Public Life

This government committee was established in 1994 after a series of corruption scandals. Its usual remit is to oversee measures to ensure politicians and civil servants adhere to the Nolan Principles, a seven-point guide to ethical conduct in public office that the committee drew up when it was created: selflessness, integrity, objectivity, accountability, openness, honesty, and leadership.

The committee has announced an inquiry into “technologically-assisted decision making”. In a [blog post](#) (Evans, 2019), the committee chair explained: “We want to understand the implications of AI for the Nolan principles and examine if government policy is up to the task of upholding standards as AI is rolled out across our public services.”

## / Academia and Civil Society

### / Data Justice Lab

The Data Justice Lab is a research lab at Cardiff University’s School of Journalism, Media, and Culture. It seeks to examine the relationship between what it calls ‘datafication’ – the collection and processing of massive amounts of data for decision-making and governance across more and more areas of social life and social justice. Its major research project DATAJUSTICE has published [working papers](#) on [how to evaluate ADM systems](#) (Sánchez-Monedero & Dencik, 2018); the [difference between ‘fairness’ and ‘social justice’](#) (Jansen, 2019) as aims for ADM and AI in policing; [data-driven policing trends](#) (Jansen, 2018) across Europe; and the [datafication of the workplace](#) (Sánchez-Monedero & Dencik, 2019).

**/ AI Council**

This committee of people from the private sector, public sector, and academia met for the first time in September 2019. [Minutes of the meeting](#) (AI Council, 2019) suggest it did not discuss ADM specifically, although in discussing its future work a need for “positive news stories on AI compared with other kinds of messaging” was raised.

## Key takeaways

**/ Narrative around data rights**

The growing discourse around data rights in the UK is a welcome development. However, centering data as the referent object to which rights need to be attached minimizes, at best, crucial questions about how data is used in automated decision-making.

The concept that data is a valuable commodity over which individuals have personal rights of ownership and control may provide some limits on how personal data is circulated without consent.

However, automated decision-making raises the question not only of what data is accessible to whom, but the very serious implications of how and to what end that data is processed. The notion of data rights is of limited use in regulating this aspect.

Several policy initiatives have focused on the potential for discrimination in ADM systems. This is a much more solid basis on which to assess ADM, and the robustness of the tools that emerge from these initiatives will have a big impact on the kind of regulatory environment for ADM in the UK.

**/ Emerging industry lobby**

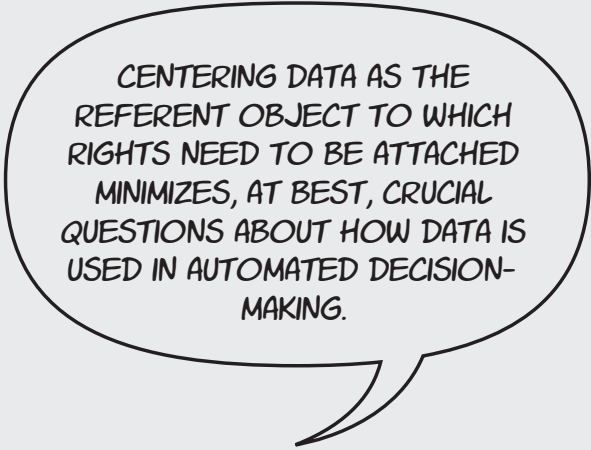
The emergence of a self-conscious artificial intelligence industry, with its attendant political lobbying functions, can be seen clearly in the structure and program of the All-Party Parliamentary Group on AI. There are great hopes for the economic potential of automation, which are encapsulated in the hype around AI (even if the narrow technical definition of the latter term is not strictly adhered to). This presents risks that the automated systems will be promoted and implemented at ever-increasing speed on the basis of possibly over optimistic eco-

nomic arguments, without proper consideration of their social impacts.

Civil society responses can take advantage of pop culture ideas about AI in order to challenge this, although they must take care not to attack the straw man of the AI singularity, and focus on the real and present-day concerns that will affect the lives of those who are subject to automated decision-making in the immediate future.

**/ Brexit and opportunities for good ADM policy**

The conclusion of several consultations in the coming year presents significant opportunities for policymakers to move towards taking action on ADM. Brexit has all but paralyzed policy implementation across all areas, but at the same time it has raised multiple new issues of ADM. While the situation is unpredictable, debate on the nature of Britain’s relationship with the EU and enormous policy decisions in this connection are likely to continue for several years. Civil society actors have an important role to play in highlighting the reality and risks of automated decision-making as the UK moves through this transformative period.



CENTERING DATA AS THE REFERENT OBJECT TO WHICH RIGHTS NEED TO BE ATTACHED MINIMIZES, AT BEST, CRUCIAL QUESTIONS ABOUT HOW DATA IS USED IN AUTOMATED DECISION-MAKING.

# References:

AI Council (2019), Department for Digital Culture, Media, and Sport, [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/836901/AI\\_Council\\_Meeting\\_Minutes\\_9\\_September\\_2019.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/836901/AI_Council_Meeting_Minutes_9_September_2019.pdf)

APPG AI (2019), Big Innovation Centre, [https://www.appg-ai.org/wp-content/uploads/2018/11/appg-ai-programme-2019-2020\\_june18.pdf](https://www.appg-ai.org/wp-content/uploads/2018/11/appg-ai-programme-2019-2020_june18.pdf)

BBC Click (2019), YouTube, <https://www.youtube.com/watch?v=0oJqJkFTdAg>

Bowcott, Owen (2019), The Guardian, <https://www.theguardian.com/technology/2019/sep/04/police-use-of-facial-recognition-is-legal-cardiff-high-court-rules>

CDEI (2019), Centre for Data Ethics and Innovation, <https://www.gov.uk/government/publications/interim-reports-from-the-centre-for-data-ethics-and-innovation/interim-report-review-into-bias-in-algorithmic-decision-making>

DCMS (2019), Department for Digital, Culture, Media & Sport, <https://www.gov.uk/guidance/national-data-strategy>

Dick (2020), YouTube, [https://www.youtube.com/watch?v=8lv7qw6X\\_rs](https://www.youtube.com/watch?v=8lv7qw6X_rs)

DWP (2011), Department for Work and Pensions, [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/633018/s11-2011.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/633018/s11-2011.pdf)

Edgington, Tom (2019), BBC, <https://www.bbc.com/news/uk-47047993>

Evans, Jonathan (2019), Committee on Standards in Public Life, <https://cspl.blog.gov.uk/2019/03/28/why-cspl-are-reviewing-artificial-intelligence-in-the-public-sector/>

Fouzder, Monidipa (2019), Law Gazette, <https://www.lawgazette.co.uk/news/liberty-heads-to-court-of-appeal-over-facial-recognition-technology-ruling/5102005.article>

Fussey, Pete and Murray, Daragh (2019), University of Essex, <https://www.hrbdt.ac.uk/hrbdt-researchers-launch-new-report-on-london-metropolitan-polices-trial-of-live-facial-recognition-technology/>

GDS (2019), Government Digital Service, <https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector>

Gentleman, Amelia (2019), The Guardian, <https://www.theguardian.com/politics/2019/oct/08/ive-been-here-50-years-the-eu-citizens-struggling-for-the-right-to-stay-in-britain>

ICO (2018), Information Commissioner's Office, <https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling-1-1.pdf>

ICO (2019), Information Commissioner's Office, <https://ico.org.uk/about-the-ico/news-and-events/ai-blog-automated-decision-making-the-role-of-meaningful-human-reviews/>

ITV (2020), <https://www.itv.com/news/london/2020-02-24/cressida-dick-facial-recognition-critics-should-face-up-to-victims-of-crime/>

Jansen, Fieke (2018), Cardiff University, <https://datajusticeproject.net/wp-content/uploads/sites/30/2019/05/Report-Data-Driven-Policing-EU.pdf>

Jansen, Fieke (2019), Cardiff University, <https://datajusticeproject.net/2019/03/18/engineering-for-social-justice-instead-of-fairness-a-perspective-from-the-context-of-policing/>

Kleinman, Zoe (2019), BBC, <https://www.bbc.com/news/technology-49320520>

Mahmood, Basit (2019), Metro, <https://metro.co.uk/2019/05/19/benefits-cheat-given-900-years-pay-88000-claimed-living-spain-9598666/>

Marsh, Sarah (2019), The Guardian, <https://www.theguardian.com/society/2019/oct/15/councils-using-algorithms-make-welfare-decisions-benefits>

Murgia, Madhumita (2019), Financial Times, <https://www.ft.com/content/8cbcb3ae-babd-11e9-8a88-aa6628ac896c>

Murgia, Madhumita (2019b), Financial Times, <https://www.ft.com/content/e9e57c2a-cff1-11e9-99a4-b5ded7a7fe3f>

Mutch, Michael (2019), The Mirror, <https://www.mirror.co.uk/news/uk-news/benefit-cheat-who-pocketed-30000-20095768>

NTC (2015), North Tyneside Council, [http://ntc-web-democratic-archive-public.s3-website.eu-west-2.amazonaws.com/Files/CBT/CBT-2015-03-09\\_7\\_I\\_Risk\\_Based\\_Verification.pdf](http://ntc-web-democratic-archive-public.s3-website.eu-west-2.amazonaws.com/Files/CBT/CBT-2015-03-09_7_I_Risk_Based_Verification.pdf)

NTC (2019), North Tyneside Council, <http://my.northtyneside.gov.uk/sites/default/files/meeting/related-documents/5c%20Risk%20Based%20Verification%20Report.pdf>

Orlowski, Andrew (2018), The Register, [https://www.theregister.co.uk/2018/11/20/can\\_you\\_trust\\_an\\_ai\\_data\\_trust\\_not\\_to\\_slurp\\_your\\_data/](https://www.theregister.co.uk/2018/11/20/can_you_trust_an_ai_data_trust_not_to_slurp_your_data/)

Rawlinson, Kevin (2019), The Guardian, <https://www.theguardian.com/technology/2019/aug/15/ico-opens-investigation-into-use-of-facial-recognition-in-kings-cross>

RUSI (2019), Royal United Services Institute, <https://www.gov.uk/government/publications/report-commissioned-by-cdei-calls-for-measures-to-address-bias-in-police-use-of-data-analytics>

Sabbagh, Dan (2019), The Guardian, <https://www.theguardian.com/uk-news/2019/aug/12/regulator-looking-at-use-of-facial-recognition-at-kings-cross-site>

Sánchez-Monedero, Javier & Dencik, Lina (2018), Cardiff University, <https://datajusticeproject.net/wp-content/uploads/sites/30/2018/12/WP-How-to-evaluate-automated-decision-systems.pdf>

Sánchez-Monedero, Javier & Dencik, Lina (2018), Cardiff University, <https://datajusticeproject.net/wp-content/uploads/sites/30/2019/05/Report-The-datafication-of-the-workplace.pdf>

Smith, Beckie (2019), Civil Service World, <https://www.civilserviceworld.com/articles/news/data-ethics-centre-and-race-disparity-unit-investigate-algorithm-bias>



# Team

## / Brigitte Alfter

Author of the **research** chapter on **Denmark**



Brigitte Alfter is a senior German-Danish journalist, director for Arena for Journalism in Europe, and lecturer at the University of Gothenburg. Having practiced journalism at local, national, and European levels, she realized there was a need for cross-border collaborative journalism structures. Since 2008, she has practiced, and assisted colleagues in practicing, in-depth journalism in Europe on European issues. She develops European support and infrastructures for cross-border journalism, combines journalism practice, entrepreneurial activities, teaching/training with academic research.

## / Naomi Appelman

Author of the **research** chapter on the **Netherlands**



Naomi Appelman is a PhD researcher at the Institute for Information Law (IViR), University of Amsterdam. Her PhD looks at platform regulation and automated decision-making through the lens of democratic values and fundamental rights within the broader context of the digital

transition of decision-making research initiatives. She combines legal and philosophical perspectives in her research, which focuses on strengthening the position of people in online speech regulation. Within the IViR, she has worked on research reports on disinformation and political advertisements for the Dutch Ministry of Internal Affairs, and she has worked on an online content moderation report. Naomi also studied philosophy at the University of Amsterdam.

## / Beate Autering

**Graphic designer** and layout artist



Beate Autering is a freelance graphic designer. She graduated in design and runs the beworx studio. She creates designs, graphics, and illustrations and also provides image editing and post-production services.

## / Hugo Barbieux

Author of the journalistic **story** on the **United Kingdom**



Hugo Barbieux studied European journalism in Brussels, where he completed a successful thesis entitled *The Political and Sociological Reasons for the Resurgence of Populism*. Since then, he has studied diverse techniques of data journalism and investigative journalism including the use of computational thinking at Birmingham City University.

## / Rosamunde van Brakel

Author of the **research** chapter on **Belgium**



Rosamunde van Brakel works as a part-time research professor at the LSTS research group of the Vrije Universiteit Brussel (VUB) in Belgium, where she coordinates the Chair in Surveillance Studies and conducts research on democratic governance of surveillance and crime control. In addition, she works as a part-time postdoctoral researcher at the imec-SMIT-VUB research group working on the SPECTRE project on smart cities and DPIAs. Also, she is co-director of the Computers, Privacy and Data Protection annual international conference, co-founder and executive director of the non-profit organization Privacy Salon, and co-director of the Surveillance Studies Network. She studied educational sciences and criminology at the Katholieke Universiteit Leuven and the University of Ottawa. Previously, she conducted research on electronic monitoring of offenders, freedom infringements of transport security technologies, evaluation and certification schemes for security products and Big Data, privacy, and security.



### / Nadja Braun Binder

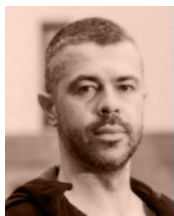
Author of the research chapter on **Switzerland**



Nadja Braun Binder studied law at the University of Berne and received her doctorate there. Her academic career took her to the Research Institute of Public Administration in Speyer in 2011, where she conducted research on the automation of administrative procedures, among other things. In 2017, she was habilitated by the German University of Administrative Sciences, Speyer, and followed a call to the Faculty of Law at the University of Zurich, where she worked as an assistant professor until 2019. Since 2019 Nadja has been Professor of Public Law at the University of Basel. Her research focuses on legal issues related to digitization in government and administration. She is currently conducting a study on the use of artificial intelligence in public administration in the canton of Zurich.

### / Jose Miguel Calatayud

Author of the **research** chapter on **Spain**



Jose Miguel Calatayud is a Spanish freelance journalist. Based in Barcelona between 2016 and 2019, Jose relocated to Berlin in 2020. Currently, he covers Europe and focuses on long-term investigations, and he is especially interested in democracy and human rights. Since 2019, he has also worked with Arena for Journalism in Europe, a non-profit foundation that promotes cross-border collaborative journalism. In 2017, he was awarded an Open Society fellowship to carry out journalistic research into the political engagement of citizens across Europe. Between 2012 and 2014, he was based in Istanbul and covered Turkey and the region for the Spanish daily newspaper, El País. Earlier, from 2009, he was based in East Africa, where he was a correspondent for the Spanish news agency, Efe, and, from 2011, for El País. His work has also appeared in Foreign Policy, Al Jazeera, New Statesman, The Independent, Agence France-Presse, Radio France Internationale, and Deutsche Welle, among others.

### / Michele Catanzaro

Author of the journalistic **story** on **Spain**



Michele Catanzaro is a freelance journalist based in Barcelona, Spain. He has a PhD in physics, has written for Nature, El Periódico de Catalunya, Süddeutsche Zeitung, and other outlets, and he is the co-author of the book “Networks: A Very Short Introduction” (2012) and the documentary “Fast Track Injustice: The Óscar Sánchez Case” (2014), that received the Golden Nymph Award in 2015. His work has also been recognized by other awards: King of Spain International Journalism Prize, BBVA Innovadata, Valors, European Science Writer of the Year 2016, Prismas, and Colombine. He has received grants from the Journalism Fund, Journalism Grants, and Climate Investigation Grants. He has experience in teaching, exhibitions, TV, and events, and he coordinates the PerCientEx project on excellence in science journalism in Spain and Latin America. He was also a Journalist in Residence at the Heidelberg Institute for Theoretical Studies.

### / Eleftherios Chelioudakis

Author of the **research** chapter on **Greece**



Eleftherios Chelioudakis is co-founder of Homo Digitalis, the Greek civil society organization which is a member of the European Digital Rights (EDRI) network. He is a lawyer admitted to practice in Greece, while he also works as a research associate at the Centre for IT & IP Law (CiTiP) at Katholieke Universiteit (KU) Leuven. Eleftherios holds an LLB from National and Kapodistrian University of Athens, an LL.M. in law and technology from Tilburg University Law School, and a master's in digital humanities from the Computer Science School of KU Leuven. Finally, he holds a Fellow of Information Privacy (FIP) designation, as well as CIPP/E and CIPM certifications from the International Association of Privacy Professionals (IAPP).

## / Fabio Chiusi

**Editor**, author of the journalistic **story** and **research** chapter on **Italy** and the **UK** as well as of the **introduction** and the **EU** chapter



Photo: Julia Bornkessel

Fabio Chiusi works at AlgorithmWatch as the co-editor and project manager for the 2020 edition of the Automating Society report. After a decade in tech reporting, he started as a consultant and assistant researcher in data and politics (Tactical Tech) and AI in journalism (Polis LSE). He coordinated the “Persuasori Social” report about the regulation of political campaigns on social media for the PuntZero Project, and he worked as a tech-policy staffer within the Chamber of Deputies of the Italian Parliament during the current legislation. Fabio is a fellow at the Nexa Center for Internet & Society in Turin and an adjunct professor at the University of San Marino, where he teaches journalism and new media and publishing and digital media. He is the author of several essays on technology and society, the latest being “Io non sono qui. Visioni e inquietudini da un futuro presente” (DeA Planeta, 2018), which is currently being translated into Polish and Chinese. He also writes as a tech-policy reporter for the collective blog ValigiaBlu.

## / Samuel Daveti

**Comic artist**



Samuel Daveti is a founding member of the Cultural Association, Double Shot. He is the author of the French language graphic-novel, Akron Le guerrier (Soleil, 2009), and he is the curator of the anthological volume Fascia Protetta (Double Shot, 2009). In 2011, he became a found-

ing member of the self-produced comics collective, Mammaiuto. Samuel also wrote Un Lungo Cammino (Mammaiuto, 2014; Shockdom, 2017), which will become a film for the media company Brandon Box. In 2018, he wrote The Three Dogs, with drawings by Laura Camelli, which won the Micheluzzi Prize at Napoli Comicon 2018 and the Boscara-to award for the best webcomic at the Treviso Comic Book Festival.

## / Catherine Egli

Author of the **research** chapter on **Switzerland**



Catherine Egli recently graduated with a double bilingual master's in law degree from the Universities of Basel and Geneva. Her thesis focused on automated individual decision-making and the need for regulation in the Swiss Administrative Procedure Act. Alongside her studies, she

worked for the chair of Prof. Dr. Nadja Braun Binder by conducting research on legal issues related to automated decision-making. Her favorite topics of research include the division of powers, digitization of public administration, and digital democracy.

## / Ronan Ó Fathaigh

Author of the **research** chapter on the **Netherlands**



Ronan Ó Fathaigh is a Senior Researcher at the Institute for Information Law (IViR), University of Amsterdam. He is a member of the Digital Transformation of Decision-Making research initiative, examining the normative implications of the shift toward automated decision-making and

the effect on democratic values and fundamental rights. Ronan has published his work in a number of international academic journals and co-authored reports for international institutions such as the Council of Europe and the European Parliament. He has also co-authored a report for the Dutch Ministry of Internal Affairs on online disinformation and political advertisements, and he has conducted research on automated decision-making in content moderation systems, as part of a joint research initiative of IViR and the Annenberg Public Policy Center, University of Pennsylvania. Ronan has a PhD in law from Ghent University, and he is a former visiting research fellow at Columbia University. Ronan is a regular contributor to the Strasbourg Observers blog, commenting on current human rights issues.

### / Sarah Fischer

#### Editor



Sarah Fischer is a project manager for the “Ethics of Algorithms” project at the Bertelsmann Stiftung, in which she is primarily responsible for the scientific studies. She has previously worked as a post-doctoral fellow in the graduate program “Trust and Communication in a Digitalized World” at the University of Münster where she focused on the topic of trust in search engines. In the same research training group, she earned her doctorate with a thesis on trust in health services on the Internet. She studied communication science at the Friedrich Schiller University in Jena, and she is the co-author of the papers “Where Machines can err. Sources of error and responsibilities in processes of algorithmic decision making” and “What Germany knows and believes about algorithms”.

### / Leonard Haas

#### Additional editing



Leonard Haas works as a research assistant at AlgorithmWatch. Among other things, he was responsible for the conception, implementation, and maintenance of the AI Ethics Guidelines Global Inventory. He is a master’s student in the field of social sciences at the Humboldt University Berlin and holds two Bachelor’s degrees from the University of Leipzig in Digital Humanities and Political Science. His research focuses on the automation of work and governance. In addition, he is interested in public interest data policy and labor struggles in the tech industry.

### / Graham Holliday

#### Copy editing



Photo: Josh White

Graham Holliday is a freelance editor, author, and journalism trainer. He has worked in a number of roles for the BBC for almost two decades, and he was a correspondent for Reuters in Rwanda. He works as an editor for CNN’s Parts Unknown and Roads & Kingdoms – the international journal of foreign correspondence. The late Anthony Bourdain published Graham’s first two books, which were reviewed in the New York Times, Los Angeles Times, Wall Street Journal, Publisher’s Weekly, Library Journal, and on NPR, among other outlets.

### / Anne Kaun

#### Author of the **research** chapter on **Sweden**



Anne Kaun is an associate professor in media and communication studies at Södertörn University, Stockholm, Sweden. Her research interests include media activism, media practices, critical studies of algorithms, automation, and artificial intelligence. Together with Fredrik Stiernstedt, she is currently pursuing a larger project on the media history and media future of prisons. In addition, she is studying automated decision-making in the public sector and its implications for the welfare state as well as democratic values. As part of this work, she facilitates the Swedish Network on Automated Decision-Making in the Public Sector that gathers scholars from diverse disciplines studying algorithmic automation in Sweden.

## / Nikolas Kayser-Bril

**Editor** and author of the **research** chapter on **France** and author of the journalistic **stories** on **Denmark, France, Germany** and **Switzerland**



Nicolas Kayser-Bril is a data journalist, and he works for AlgorithmWatch as a reporter. He pioneered new forms of journalism in France and Europe and is one of the leading experts on data journalism. He regularly speaks at international conferences, teaches journalism

in French journalism schools, and gives training sessions in newsrooms. A self-taught journalist and developer (and a graduate in Economics), he started by developing small interactive, data-driven applications for Le Monde in Paris in 2009. He then built the data journalism team at OWNI in 2010 before co-founding and managing Journalism++ from 2011 to 2017. Nicolas is also one of the main contributors to the Data Journalism Handbook, the reference book for the popularization of data journalism worldwide.

## / Aleksi Knuutila

Author of the journalistic **story** on **Finland**



Alexsi Knuutila is an anthropologist and a data scientist who studies new forms of political culture, communities, and communication. He is interested in the application of interdisciplinary methods that combine interpretative, qualitative methods to the computational study of large-scale digital data. Knuutila works as a postdoctoral researcher in the Computational Propaganda Project.

## / Lenart J. Kučič

Author of the journalistic **story** and the **research** chapter on **Slovenia**



Lenart J. Kučič is a journalist, lecturer, and podcaster. He is a member of Pod črto, a non-profit and independent media collective specialized in investigative reporting, data journalism, and in-depth stories. He is a co-editor of the online magazine Disenz and a co-founder of the podcasting network, Marsowci. In addition, he is the author and co-author of several books, expert articles, and studies on the subjects of media ownership and new media. His recent work focuses on computational propaganda, the use of algorithms, and the social impact of ICT's.

## / Tuukka Lehtiniemi

Author of the **research** chapter on **Finland**



Tuukka Lehtiniemi is an economic sociologist and a postdoctoral researcher at the Centre for Consumer Society Research at the University of Helsinki. He is broadly interested in how the uses we invent for new technologies are shaped by how we imagine the economy to work.

His current research focuses on the data economy and automated decision-making technologies. He previously worked at Aalto University, and, prior to that, in expert positions in the Finnish public sector. In 2018, he was a fellow at the Alexander von Humboldt Institute for Internet and Society in Berlin. His work has been published in New Media & Society, Big Data & Society, and Surveillance & Society.

### / Nikolas Leontopoulos

Author of the journalistic **story** on **Greece**



Nikolas Leontopoulos is a Greek journalist based in Athens. He is the co-founder of Reporters United, a new center for investigative journalism that incorporates a network of reporters in Greece. For ten years, Nikolas worked for the Athens daily, Eleftherotypia. He collaborated with Reuters on investigations into Greek banking, shipping, and media, and he reported for the “Outlaw Ocean” New York Times series. During the Greek financial crisis, he worked as a fixer for several international media outlets, field-produced documentaries for VRPO, ARTE, and Al Jazeera, and he researched for ARD’s documentary “The Trail of the Troika”. He made two documentaries for VICE Greece: “The European Commission and the Defense Industry Lobby” and “The German Big Brother”. He was a co-founder of Investigate Europe, a consortium of European journalists publishing with media partners across Europe.

### / Katarina Lind

Author of the journalistic **story** on **Sweden**



Katarina Lind is a Swedish data journalist who worked with J++, an international team of data journalists, on the chapter about automating the welfare system in Sweden.

### / Maris Männiste

Author of the **research** chapter on **Estonia**



Maris Männiste is a media and communications PhD student and works as an information systems assistant at the Institute of Social Studies at the University of Tartu. Her research interests include the social consequences of big data and algorithms, perceptions of privacy, and datafication practices. Her recent research explores algorithmic governance ideals, where Estonian experts are considered to be a pioneer community engaged in the analysis and management of data and actively proposing novel data solutions.

### / Natalia Mileszyk

Author of the **research** chapter on **Poland**



Natalia Mileszyk is a lawyer, public policy expert, and head of digital and technology practice at CEC Government Relations, based in Warsaw. Previously, she was a public policy expert at Centrum Cyfrowe Foundation, a leading Polish think-and-do-tank, dealing with the social aspects of technology and the need for human-centered digital policy. Recently, this involved such issues as online platform regulations, AI, automated decision-making, and sharing economy regulations. She graduated from Warsaw University and Central European University in Budapest (LL.M.).

### / Lorenzo Palloni

**Comic artist**



Lorenzo Palloni is a cartoonist, the author of several graphic novels and webcomics, an award-winning writer, and one of the founders of comic artists collective, Mammaiuto. At the moment, he is working on forthcoming books for the French and Italian markets. Lorenzo is also a Scriptwriting and Storytelling teacher at La Scuola Internazionale di Comics di Reggio Emilia (International Comics School of Reggio Emilia).

### / Kristina Penner

Author of the **EU** chapter



Photo: Julia Bornkessel

Kristina Penner is the executive advisor at AlgorithmWatch. Her research interests include ADM in social welfare systems, social scoring, and the societal impacts of ADM, as well as the sustainability of new technologies through a holistic lens. Her analysis of the EU border management system builds on her previous experience in research and counseling on asylum law. Further experience includes projects on the use of media in civil society and conflict-sensitive journalism, as well as stakeholder involvement in peace processes in the Philippines. She holds a master’s degree in international studies/peace and conflict research from Goethe University in Frankfurt.



## / Gerda Kelly Pill

Author of the **journalistic story** on **Estonia**



Gerda Kelly Pill is currently working in the ICT sector as a marketing and communications manager. Her job includes working with making complex topics understandable, communicating to an international audience, and bringing digitization to the construction, real estate, and facility management industries. Gerda is most passionate about information and media literacy, privacy, freedom of speech, and gender equality. She is currently in her second year of master's studies in communications at the University of Tartu. Gerda intends to write her thesis on media literacy and media education in Estonia.

## / Alessio Ravazzani

**Comic artist**



Alessio Ravazzani is an editorial graphic designer, cartoonist, and illustrator who collaborates with the most prestigious comics and graphic novel publishers in Italy. He is an author with the Mammaiuto collective, of which he has been a member since its foundation.

## / Friederike Reinhold

Additional editing of the **introduction** and **policy recommendations**



Photo: Julia Bornkessel

As a senior policy advisor, Friederike Reinhold is responsible for advancing AlgorithmWatch's policy and advocacy efforts. Prior to joining AlgorithmWatch, she worked as a Humanitarian Policy Advisor at the German Federal Foreign Office, with the Norwegian Refugee Council (NRC) in Iran, with Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) in Afghanistan, and at the WZB Berlin Social Science Center.

## / Minna Ruckenstein

Author of the **research** chapter on **Finland**



Minna Ruckenstein is an associate professor at the Consumer Society Research Centre and the Helsinki Centre for Digital Humanities at the University of Helsinki. She directs a research group that explores economic, social, emotional, and imaginary aspects of algorithmic systems and processes of datafication. Recently projects have focused on algorithmic culture and rehumanizing automated decision-making. The disciplines of anthropology, science and technology, economic sociology, and consumer research underpin her work. Minna has been published widely in respected international journals, including *Big Data & Society*, *New Media & Society*, and *Social Science & Medicine*. Prior to her academic work, she was a journalist and an independent consultant, and this professional experience has shaped the way she works, in a participatory manner with stakeholders involved. Her most recent collaborative projects have explored practices of content moderation and data activism.

## / Eduardo Santos

Author of the **research** chapter on **Portugal**



Eduardo Santos is a Lisbon-based lawyer. He graduated from the faculty of law at the University of Lisbon, and he is a co-founder, and the current president, of D3 - Defesa dos Direitos Digitais, the Portuguese Digital Rights Association, part of the European Digital Rights (EDRI) association. Eduardo's work at D3 focuses on tech policy, advocacy, and lobbying related to digital rights issues including, privacy, net neutrality, data retention, video surveillance, copyright, and regulation of platforms both at a local and a European level. He has also contributed to research projects within the areas of law and technology and has been an active civil society representative on international Internet governance fora such as IGF and EuroDIG.

### / Paula Simões

Author of the journalistic **story** on **Portugal**



Paula Simões has a degree in journalism and is currently doing her PhD in digital humanities. She is an active member of ANSOL, the Portuguese National Association for Free Software, and she is also an active member of D3, the Portuguese Digital Rights Association. She was also

president of AEL, an association that promoted F/LOSS, Creative Commons, Public Domain, and Open Access in education and scientific research, and she has been writing about these themes for more than a decade.

### / Matthias Spielkamp

**Editor**



Photo: Julia Bornkessel

Matthias Spielkamp is co-founder and executive director of AlgorithmWatch. He has testified before several committees of the German Bundestag on AI and automation. Matthias serves on the governing board of the German section of Reporters Without Borders and the advisory councils of Stiftung Warentest and the Whistleblower Network.

He was a fellow of ZEIT Stiftung, Stiftung Mercator, and the American Council on Germany. Matthias founded the online magazine mobilisicher.de, reporting on the security of mobile devices, with an audience of more than 170,000 readers monthly. He has written and edited books on digital journalism and Internet governance and was named one of 15 architects building the data-driven future by Silicon Republic. He holds master's degrees in journalism from the University of Colorado in Boulder and in philosophy from the Free University of Berlin.

### / Beate Stangl

**Graphic designer** and layout artist



Beate Stangl is a qualified designer and she works in Berlin on editorial design projects for beworx, Friedrich Ebert Stiftung, Buske Verlag, UNESCO Welterbe Deutschland e.V., Sehstern Agency, iRights Lab, and Landesspracheninstitut Bochum.

### / Konrad Szczygieł

Author of the journalistic **story** on **Poland**



Konrad Szczygieł is a Polish journalist working with Reporters Foundation (Fundacja Reporterów) and VSquare.org. He is a former investigative reporter at OKO.press and Superwizjer TVN. In 2016, and 2018, he was shortlisted for the Grand Press award. Konrad is based in Warsaw.

### / Gabriela Taranu

Author of the **research** chapter on **Sweden**



Gabriela Taranu is a Romanian marketing specialist based in Sweden and interested in social media and digital communication. She studied communication and public relations at the University of Bucharest before completing a master's degree in media, communication and cultural analysis at Södertörn University. Her studies focussed on the Korean music industry and its influence upon different demographics to analyze trends in popular culture and how they connect with consumers through social media and marketing campaigns. After finishing her studies, Gabriela worked in digital marketing and social media at a tech start-up to improve communication channels and create strong relations between the product and its users.

## / Alek Tarkowski

Author of the **research** chapter on **Poland**



Alek Tarkowski is a sociologist, digital rights activist, and strategist. He is co-founder and president of Centrum Cyfrowe Foundation, a think-and-do tank. His work focuses on strategies and public policies through which digital technologies serve openness, collaboration, and engagement. Recently, he has also focused on the social and cultural dimensions of the latest phase of technological change. Alek is a board member of Creative Commons, founder of Creative Commons Poland, and co-founder of Communia, a European advocacy association supporting the digital public domain. He is an alumnus of New Europe Challenger 2016 and Leadership Academy of Poland (class of 2017), a member of the Steering Committee of the Internet Governance Forum Poland, a former member of the Polish Board of Digitization (an advisory body to the Minister of Digitization), and a member of the Board of Strategic Advisors to the Prime Minister of Poland.

## / Marc Thümmeler

Publications **coordinator**



Photo: Julia Bornkessel

Marc Thümmeler is in charge of public relations and outreach at AlgorithmWatch. He has a master's degree in media studies, has worked as a producer and editor in a film company, and managed projects for the Deutsche Kinemathek and the civil society organization Gesicht Zeigen. In addition to his core tasks at AlgorithmWatch, Marc has been involved in the crowdfunding and crowdsourcing campaign OpenSCHUFA, and he coordinated the first issue of the Automating Society report, published in 2019.

## / Koen Vervloesem

Author of the journalistic **stories** on **Belgium** and the **Netherlands**



Koen Vervloesem is a Belgian freelance journalist. He holds master's degrees in engineering (computer science) and in philosophy. He has been writing about artificial intelligence, programming, privacy, computer security, Linux, open-source, and the Internet of Things for more than 20 years.

## / Leo Wallentin

Author of the journalistic **story** on **Sweden**



Photo: Sara Mac Key

Leo Wallentin is a Swedish data journalist and CEO at J++ Stockholm, a consultancy agency doing data-driven research, storytelling, and training for newsrooms. Leo has spent the last decade specializing in data-driven and data-critical journalism.

## / Louisa Well

Author of the **research** chapter on **Germany**



Louisa Well is a policy advisor to Dr. Anna Christmann, MdB, who is Spokeswoman for Innovation and Technology Policy of the parliamentary group Alliance 90/The Greens at the German Bundestag. In this position, Louisa works on policies in AI and digitization. She received a master's degree in science on the subject of science and technology in society from the University of Edinburgh and a Bachelor of Arts in political science and English literature from the University of Heidelberg. For her master's thesis — on the question of whether automation fosters discrimination in the automated distribution of unemployment benefits in Germany — she did a research placement with AlgorithmWatch.

### / Tom Wills

Author of the **research** chapter on the  
**United Kingdom**



Tom Wills is a freelance data journalist and researcher based in Berlin. His specialty is using computational techniques for journalistic research and analysis — a type of data journalism. Previously, he led the data journalism team at The Times of London, using computational techniques

to drive investigations on topics ranging from Westminster politics to the gender pay gap. Using tools such as the Python coding language to design algorithms for the purposes of public interest journalism has given him an insight into the perils of automation. He has also reported on the consequences of data-driven decision-making, including a major investigation into the World-Check bank screening database which resulted in stories published in six countries as part of an international collaboration. Tom graduated from City University, London with a master's degree in investigative journalism in 2012.

# ORGANISATIONS

## / AlgorithmWatch

AlgorithmWatch is a non-profit research and advocacy organization that is committed to watch, unpack and analyze algorithmic / automated decision-making (ADM) systems and their impact on society. While the prudent use of ADM systems can benefit individuals and communities, they come with great risks. In order to protect human autonomy and fundamental rights and maximize the public good, we consider it crucial to hold ADM systems accountable to democratic control. Use of ADM systems that significantly affect individuals' and collective rights must not only be made public in clear and accessible ways, individuals must also be able to understand how decisions are reached and to contest them if necessary. Therefore, we enable citizens to better understand ADM systems and develop ways to achieve democratic governance of these processes – with a mix of technologies, regulation, and suitable oversight institutions. With this, we strive to contribute to a fair and inclusive society and to maximize the benefit of ADM systems for society at large.

<https://algorithmwatch.org/en/>



## / Bertelsmann Stiftung

The Bertelsmann Stiftung works to promote social inclusion for everyone. It is committed to advancing this goal through programmes aimed at improving education, shaping democracy, advancing society, promoting health, vitalizing culture and strengthening economies. Through its activities, the Stiftung aims to encourage citizens to contribute to the common good. Founded in 1977 by Reinhard Mohn, the non-profit foundation holds the majority of shares in the Bertelsmann SE & Co. KGaA. The Bertelsmann Stiftung is a non-partisan, private operating foundation. With its "Ethics of Algorithms" project, the Bertelsmann Stiftung is taking a close look at the consequences of algorithmic decision-making in society with the goal of ensuring that these systems are used to serve society. The aim is to help inform and advance algorithmic systems that facilitate greater social inclusion. This involves committing to what is best for a society rather than what's technically possible – so that machine-informed decisions can best serve humankind.

<https://www.bertelsmann-stiftung.de/en>

## | BertelsmannStiftung