Digital
Future Society

Governing algorithms: perils and powers of AI in the public sector

A programme of

GOBIERNO
DE ESPAÑA

VICEPRESIDENCIA
SEGUNDA DEL GOBIERNO
MINISTERIO
DE ASUNTOS ECONÓMICOS
Y TRANSFORMACIÓN DIGITAL

SECRETARÍA DE ESTADO
DE DIGITALIZACIÓN
E INTELIGENCIA ARTIFICIAL

red.es

MOBILE
WORLD CAPITAL™
BARCELONA

About Digital Future Society

Digital Future Society is a non-profit transnational initiative that engages policymakers, civic society organisations, academic experts and entrepreneurs from around the world to explore, experiment and explain how technologies can be designed, used and governed in ways that create the conditions for a more inclusive and equitable society.

Our aim is to help policymakers identify, understand and prioritise key challenges and opportunities now and in the next ten years in the areas of public innovation, digital trust and equitable growth.

Visit digitalfuturesociety.com to learn more

A programme of

GOBIERNO DE ESPAÑA
VICEPRESIDENCIA SEGUNDA DEL GOBIERNO
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

red.es

MOBILE WORLD CAPITAL™ BARCELONA

Published
May 2021

Disclaimer
The information and views set out in this report do not necessarily reflect the official opinion of Mobile World Capital Foundation. The Foundation does not guarantee the accuracy of the data included in this report. Neither the Foundation nor any person acting on the Foundation's behalf may be held responsible for the use which may be made of the information contained herein.

Contents

1. We need to talk about AI

2020, a year with more questions than answers

Artificial Intelligence (AI) was a hot topic in 2020, thanks, in part, to the availability of increasingly well-developed products offering mature and useful AI-powered services. In 2020, AI-based systems carried out more routine tasks than ever before, from planning step-by-step travel directions to translating text between different languages. Furthermore, 2020 will be remembered as the year when AI came to the forefront of many high-impact government decisions. An alarming example of this is the mainstream public sector use of AI systems, such Automatic Decision-Making Systems (ADMS) to support the provision of social benefit entitlements, often with a lack of quality data and poor algorithm accuracy.

There is also a fear building up around artificial intelligence with, among other uses of AI, the proliferation of facial recognition systems (FRS) in public spaces including by the police, causing unease. Unnecessary surveillance and human rights limitations and breaches, especially in non-democratic regimes, are now seemingly in the cold hands of machines, with those same machines "providing governments with unprecedented capabilities to monitor their citizens and shape their choices but also by giving them new capacity to disrupt elections, elevate false information, and delegitimise democratic discourse across borders".[1]

Accordingly, 2020 was also marked by ethical discussions around the use of more advanced AI systems to support the management of administrative tasks including, but not limited to, facial recognition systems, algorithmic predictions about and even control of citizen behaviours. The use of AI-enabled tools by the police and military was also the subject of discussion, as was machine-based discrimination bias.

The Covid-19 outbreak has only exacerbated the threats AI systems pose further. Governments have had to quickly reorient human resources, create contact-tracing apps and adopt new, fully digital ways to carry out administrative work and deliver public services. In this context, the risks include, even without the intention of wrongdoing, the mishandling or infringement of data protection rules on the use of non-anonymised records to develop machine learning tools for early detection of specific, real or expected, behaviours. In fact, this is what had occurred in most public sector applications of AI, even before Covid-19 piled more pressure onto government administrations.

AI is often seen as a silver bullet, but the complexities below the surface represent risks that need to be taken seriously. There is a clear need for policymakers to better grasp the challenges and risks that AI implementation brings, especially to the public sector, in order to implement solutions that can be truly beneficial for all.

[1] Feldstein 2019

What to expect from this whitepaper

AI could well have been nominated Person of the Year 2020 by Time magazine due to huge media attention, in-depth scientific scrutiny and hot policy and regulatory debates that swirled around the great opportunities and enormous risks it poses. However, in 2021 and beyond, we should not stop talking about AI.

The goal of this whitepaper is to contribute towards an inclusive development of AI and help restore and strengthen trust between policymakers and the public. This calls for a greater effort to understand AI's effects more clearly and develop explainable and accountable algorithms. Furthermore, there is a need for strong evaluation frameworks that can assess not only the efficiency but also the performance and socio-economic impact of AI.

In the words of Stephen Hawking, "Success in creating effective AI, could be the biggest event in the history of our civilization. Or the worst. We just don't know."[2]

This whitepaper contains five AI use case studies that have raised concern due to the considerable public backlash that emerged following their adoption. Each fuelled strong debate among politicians, academics, practitioners and citizens. These examples all come from European countries with other international examples also included throughout the whitepaper.

Today our attention is focused primarily on what is properly known as "narrow AI" (or weak AI), which is AI designed to perform a narrow task (eg only facial recognition, internet search or the analysis of specific datasets). However, we are only at the beginning of the AI age![3] The fast pace of technological development raises the question of what will happen if many researchers succeed in the long-term goal of creating what is defined as "general AI" (or strong AI) with AI systems becoming better than humans at all cognitive functions.

What makes Europe different

This whitepaper examines mostly European cases because the European Union (EU), seeking to limit the risks associated with AI, took the position of developing a responsible AI that has an ethical purpose and technical robustness. These are two critical components for fostering trust and facilitating uptake. Building on the 2018 communication AI for Europe and inspired by the Ethics Guidelines of the High-Level Expert Group on AI, the European approach seeks to promote a "human-centric AI", while at the same time supporting technological and industrial capacity and adoption across the economy and public sector.[4, 5, 6]

As emphasised in the Strategy for Europe's Digital Future, which was adopted in 2020, the EU expects AI to significantly improve the lives of citizens and bring major benefits to society through better healthcare, sustainable farming, safer transport, and by making

[2] Kharpal 2017

[3] Oxford Insights 2020

[4] European Commission 2018a

[5] European Commission 2019

[6] European Commission 2020a

industry more competitive and public services more efficient.[7] In this respect, the EU's AI whitepaper describes an approach aimed at creating both an "ecosystem of excellence" and an "ecosystem of trust", making AI systems "ethical by design", and also proposes a risk-based approach to the regulatory regime.[8]

According to the EU Commission, it is important to ensure that regulation is proportionate. It envisages a tiered approach with high risk AI systems subject to mandatory certification before gaining access to the market. A high risk AI classification depends on what is at stake, considering whether both the sector and the intended use involve significant risks. The proposed AI Regulatory Requirements, confirmed in April 2021, will elaborate this further and foster discussion at the international level.

The Commission's ambition is to set out and inspire a common approach for nurturing a distinctive form of AI that is ethically robust and protects the rights of individuals and society. The hope is that the AI Regulatory Requirements will follow a similar path to the General Data Protection Regulation (GDPR), which, although opposed by many during preparation, inspired similar approaches worldwide.[9]

[7] European Commission 2020a

[8] European Commission 2020b

[9] European Parliament 2016

2. AI in government: the rise of the known unknown

Traditionally, AI refers to machines or agents that are capable of observing their environment, learning and then taking intelligent actions or proposing decisions, based on the knowledge and experience gained as a result.[10] Typical applications include machine or deep learning software; robotic process automation (RPA) such as those present in voice assistants; image or speech recognition and text translation; and automated decision-making systems (ADMS). It is also possible to embed AI in hardware devices such as advanced robots, autonomous systems and internet of things (IoT) systems and devices.

The use of AI-enabled systems and tools to support decision-making, implementation and interaction already spans the work of most public administrations worldwide, as it has a clear potential to reduce the cost of core government functions, including enforcing regulatory mandates and adjudicating benefits and privileges.[11] However, many use cases also include other critically important governance tasks such as regulatory analysis, rule-making, internal personnel management, citizen engagement and service delivery.

In most cases, AI systems serve to enhance government performance through automatic analysis of huge volumes of data. They are assumed to provide more comprehensive and accurate insights than human-driven analyses. Nevertheless, this is not necessarily the case, as results from computerised data analytics depend on the quality of the available data and the accuracy of the algorithms employed. But in addition to the issues and challenges we do know about, the “known knowns”, and not to mention the many “unknown unknowns” that we do not know about, the inherent characteristics of AI and the learning properties they display emphasise the existence of many “known unknowns”. That is to say the challenges and problems we do know about but do not know how to solve. This means it is urgent to address AI’s current limitations as well as the negative consequences and side-effects the inappropriate use of AI systems can have on citizens.

In principle, there is potential for AI to improve lives by processing huge amounts of data, supporting civil servants in decision-making processes, and providing tailored applications and personalised services.[12] Nevertheless, AI can also increase institutional isomorphism and crystallise dysfunctional systems and structures of power. A layer of AI or machine learning over dysfunctional systems or biased datasets will only worsen pre-existing problems. In addition to this, the public sector is exposed to more in-depth public scrutiny due to the role and functions of the government and the risk of intensifying power asymmetries between policymakers and among citizens.[13]

As some of the examples presented later show, digitalisation processes often touch areas that deal with citizens in very vulnerable situations, which reinforces the need to understand the risks that AI deployment brings to the public sector. In addition, there are other important

[10] Craglia et al. 2018

[11] Engstrom et al. 2020

[12] Algorithm Watch and Bertelsmann Stiftung 2020

[13] Kuziemski and Misuraca 2020

threats inherent to the properties of AI, such as the consequences a machine denying an entitlement through an AI-enabled system, the lack of digital skills of civil servants or how these systems really operate and what the implications are for users.

Previous Digital Future Society work highlights some of the main challenges associated with introducing AI systems into the public sector. These include the "discrimination by default" and inherent bias that the lack of quality of datasets on the lives of vulnerable groups and disadvantaged individuals can generate, the stubborn opacity surrounding the ever-increasing use of solutions in support of what has been labelled the "digital welfare state", and the profound impact these systems may have on the relationship between democratic systems and "algorithmic governance" due to the surveillance power that these technologies can offer public sector institutions.[14, 15, 16, 17]

Discrimination by default

AI offers governments multiple opportunities but it also raises many challenges. For instance, while it can help streamline administrative operations and processes, it could also prove inaccurate and disrupt interoperability between government departments. Artificial intelligence can enable better knowledge gathering and help generate insights by applying advanced predictive analytics, but it also tends to be invasive and can often further engrain social and institutional biases.

Controversial examples include cases of predictive policing, which involve law enforcement agencies using AI technologies to make decisions about pre-trial release and sentencing, or to identify areas where crimes are more likely to occur.[18, 19] An example of this, the Correctional Management Offender Profiling for Alternative Sanctions (COMPAS) in the United States, offers what is likely the most notorious case of AI prejudice.[20] Similar use of AI to predict the likelihood of a criminal reoffending has been widely deployed in various jurisdictions across the US since 2010. A 2016 study from ProPublica reported that "the system predicts that black defendants pose a higher risk of recidivism than they do, and the reverse for white defendants".[21] Even though a later study showed that ProPublica made an important data processing error, which in part affected positive and negative predictive values, the nonprofit organisation asserted "this had little impact on some of the other key statistical measures, which are less susceptible to changes in the relative share of recidivists, such as the false positive and false negative rates, and the overall accuracy".[22]

AI systems aiming to identify hot spots areas for crime have also encountered the same problems. These systems influence police officers on patrol in identified areas, making them more likely to stop or arrest people because of expectations raised by the system's analysis and prediction, rather than the actual circumstances on the ground.[23] Increasing evidence

[14] Digital Future Society 2020a

[15] Alston 2019

[16] Algorithm Watch and Bertelsmann Stiftung 2020

[17] Digital Future Society 2020b

[18] Big Brother Watch 2020

[19] Dencik et al. 2019

[20] Douglas Heaven 2020

[21] COMPAS is an assistive software and support tool used to predict recidivism risk: the risk that a criminal defendant will re-offend.

[22] Barenstein 2019

[23] Babuta and Oswald 2019

suggests, in fact, that as it is biased police data training the machine-learning models, human prejudices are reinforced and consolidated into the AI systems.[24]

Navigating through false positives and negatives

Prediction algorithms are subject to error. In the context of facial recognition technologies, for instance, there are two possible outcomes: a false positive, in which the algorithm draws a positive match between two facial images, when in fact there is no match, and a false negative, in which the algorithm concludes that there is no match, when in fact there is one.[25] A case that has raised significant concern is the use of a fugitive facial recognition system (FRS) by the City of Buenos Aires. Following an April 2019 resolution, the Ministerio de Justicia y Seguridad de la Ciudad Autónoma de Buenos Aires (Municipal Ministry of Justice and Security of Buenos Aires) used a live FRS to identify children accused of committing crimes.[26]

Human Rights Watch (HRW) criticised the system, calling on the city and national government to stop using it to identify suspects, particularly minors, pointing out that the system regularly misidentifies minors. The group argued that these misidentifications could unjustly limit the educational and job opportunities available to children wrongly accused of theft and other crimes. Furthermore, children accused of having committed a crime had their personal information published online, which is against international law.[27]

The global debate surrounding FRS is an important one as this invasive and potentially harmful use of mass surveillance tools is being increasingly implemented across Latin America. The governments in Brazil and Uruguay, for example, are pushing for a legal framework to manage the use of facial recognition systems.

Black-boxing effects of the digital welfare state

Using AI technologies to help organisations detect anomalies within big datasets is also controversial. For example, these technologies use data to automatically detect fraudulent behaviour relating to government service provisions such as subsidies, social welfare or tax (as we will see later) or to identify children and families considered vulnerable and at risk of abuse. A highly discussed case is the Early Help Profiling System (EHPS) deployed by the Hackney Council in London.[28]

[24] Richardson, et al. 2019

[25] European Union Agency for Fundamental Rights 2019

[26] Bronstein 2020

[27] UN Human Rights Office of the High Commissioner 2020

[28] Dencik et al. 2018

The system was supposed to help councils save around 1 million GBP per year by facilitating early, targeted interventions but ended up being heavily criticised due to the nature of the data it collected and the opaque risk assessment it employed. Citizen concerns also related to the fact it seemed the system was only put in place to cope with the UK Government's austerity measures, as it was advertised that it would maximise payments from the Troubled Families programme.[29]

Hackney Council finally halted the project stating it did not realise expected benefits.[30] To a great extent, this is illustrative of the effects that focusing on efficiency and cost-effectiveness has had on the digitalisation of the welfare state in most countries. Little thought goes into the design of the AI-based systems, how to deal with the lack of transparency, or the biased data used to train the algorithms.[31, 32]

AI systems employed to support social assistance applications or calculate healthcare benefits have also shown similar signs of social biases, racial or ethnic discrimination.[33] The problem lies in the fact that it is hard to discern where any bias might come from because often the algorithms are proprietary and so closed off from scrutiny. This brings in an additional challenge linked to the limited capacity of public sector organisations, and the civil servants' ability to deal with such complex systems. Often the humans working with these systems end up relying on the decisions suggested by the machine, without properly being able to question or fully understand the rationale behind them.

In practice, the known unknowns that are emerging as fundamental issues for policymakers to address show that there is an urgent need to ensure government systems and decision-making processes are human-centric and accountable, guarantee transparency and quality of public service management and delivery, and, ultimately, generate well-being for all.

[29] GOV.UK, n.d

[30] Dencik, et al. 2019

[31] Douglas Heaven 2020

[32] Digital Future Society 2019

[33] Eubanks 2018

3. Governance of, with and by AI

It is clear then that policymakers face a difficult dilemma: the obligation to protect citizens from potential algorithmic harms is at odds with the temptation to increase efficiency and enhance the quality of digital services.[34] The challenge they face is two-fold: to govern AI, algorithms and related automated processes, and govern with and by AI, using algorithms and computerised methods and systems to enhance and improve public services.

Governance of AI

Like with any technological innovation, introducing AI into the public sector is not a straightforward process. It must not override existing governance mechanisms and institutions. There are the traditional technological, legal and regulatory barriers to address as well as ethical and social concerns. Furthermore, other factors such as long-term investments, skills and capacities, perceived value, and the sustainability and difficulties faced in the development of basic digital government operations and services, also relate to AI. This means the type of governance "of AI" adopted is critical and not so easy to determine upfront.

Merging of enormous amounts of data with powerful machine learning algorithms is what currently drives the development of AI. Therefore, it is impossible to talk about governance of AI without first looking at existing data regulatory regimes and practices. It would be logical to establish AI governance as an extension of data protection and competition regulation. Unfortunately, however, the current attitude towards AI is driven by the narrative of exceptionalism, AI is perceived with AI as a new phenomenon that lies outside existing policies and laws.

This means governments must first develop a better understanding of the governance mechanisms and regulatory implications that are changing the way that public and private sector organisations operate, as well as the impact they have on citizens' rights. Only then will they be in a position to explore the innovative uses of technologies governments feel they need. The SyRi and Gladsaxe cases, presented in section four of this whitepaper, illustrate this point further.

Governance with AI

Another important, yet often overlooked, aspect to explore and assess is the effective use and value AI can offer governments when they are redesigning internal administrative processes to enhance the quality and impact of public services.[35]

[34] Kuziemski and Misuraca 2020

[35] As the literature review by Desouza et al. (2020) highlights, the focus of research on AI take-up lies — almost exclusively — in the development and applicability of AI in the private sector. Only a very small portion (59 out of 1438) of articles published between 2000 and 2019 discuss AI for and in the public sector.

Governing "with AI" means humans should still remain in the classical situation of using and controlling a technology that reinforces our capacity, through a process that requires human supervision. Crucially, however, this requires a better understanding of the potential benefits and risks associated with the use of AI in the public sector. These include safeguarding human rights and deploying AI ethically, especially in sensitive policy areas and domains of public interest that have direct and stringent implications on the trust-relationship that exists between governments and citizens.

In Poland, for example, both the public and internal civil servants criticised an algorithmic profiling system introduced as part of a reform of the *Powiatowe Urzędy Pracy*, *PUP* (Public Employment Services). The system divided unemployed citizens into three different categories with each category establishing an indicated level of support and resource burden. It drew criticism as it was very opaque with citizens unaware of the score it gave them or how this score had been determined.

Furthermore, the idea was for the profiling system to serve solely as an advisory tool, with a human operator ultimately deciding the appropriateness of each categorisation. However, in practice, internal staff questioned less than one percent of the algorithm's decisions due to a lack of time, fear of management repercussions and the presumed objectivity of the AI system.

In the end, the system was judged unconstitutional and dismantled by the government following formal complaints about the discriminations it caused. This case clearly shows that while humans-in-the-loop could offer a solution, they must be empowered to question AI decisions – especially when the systems have been introduced to help save costs and improve efficiency. Section four of the whitepaper explores another similar case from Austria.

Governance by AI

The true power – and related perils – of AI use in the public sector lies, however, in governance "by AI", which implies that human decisionmakers should surrender to the "superhuman capacities" of AI.

Although applications of this type of AI system are still in the early stages, particularly in government, we are already witnessing the rapid development of intelligent/autonomous systems that do not simply execute predefined instructions or tasks. More sophisticated AI applications would not rely on human intervention, and could learn and adapt on their own. They can be used as a collaborative tool to identify problems, find new solutions and execute them faster in innovative ways. However, if used maliciously, they can cause harm and shift the cognitive capacities of human beings, which in turn, would have a profound impact on the world we live in, in personal and social spaces.

This development further exacerbates the tensions that exist because of the unequal relationships that exist between data subjects and data analysers (be they "augmented" humans with computer-assisted capacities, or solely a machine). This is called the "algocracy" risk.[36]

[36] Danaher 2016

AI can bring about better outcomes for everyone, but before embarking on a potentially radical transformation of the way policies and services are designed and delivered, the possible risks and unintended consequences and side effects must be considered. Not least of all, challenges relating to accountability and trust, but also liability. Who will be held responsible when an AI system causes damage through accidents or mistakes?

As the cases in the following section illustrate, the fundamental question about how governments design and manage AI systems (or AI systems can manage governments?) and the role of private sector providers that often control the data and automated decision-making system processes, needs to be addressed. In this respect, since empirical studies on the use of algorithmic models in policymaking have so far been scarce, limiting the academic understanding of their use and effects, a dedicated effort is required, at the policy and research level, to ensure that the use of AI in the public sector receives more attention.[37] AI is conceived as an important driver of change for governance systems, as it can enable a paradigmatic shift in the power relations between stakeholders. However, this change is often driven by techno-deterministic approaches. As Evgeny Morozov warns, "rather than fixing social support structures or the true causes of crises, solutionists deploy technology to avoid politics, and explore more and more ways to nudge our behaviour to cope with the problems".[38]

The legal and ethical implications of AI use (be it with or by AI) are of key importance to ensure the legitimacy and trustworthiness of governments and the delivery of fair and inclusive public services. At the same time, the public sector plays a central role in defining the regulatory mechanisms and technical solutions for further development of AI based systems across society.[39]

[37] Kolkman 2020

[38] Morozov 2020

[39] Misuraca and van Noordt 2020

4. Learning from European examples of AI in government

In many countries, governments are experimenting with AI to improve policymaking and service delivery. This is already having impacts on various aspects of the public sector, and these are often taken for granted as being positive. However, there are many examples of misuse, and the negative consequences and harms the use of AI in government can cause, including a number of widely publicised cases that have garnered huge public interest and subsequent policy debates.

As anticipated, many challenges underpin the effective use of AI in the public sector, undermining its mainstream implementation. But indeed, whereas it may be little more than a minor nuisance if your text predictor suggests one word when you mean another, it becomes all the more important for an AI system to do what you mean it to, if it aims to support decisions about your health or social care benefits for instance.

From this perspective, the following European cases illustrate some of the main risks associated with governmental use of AI in crucial areas of public service and policymaking. These case studies focus on Europe because, as mentioned earlier, the EU's position is to develop a responsible AI that has an ethical purpose and technical robustness.

From fraud detection to government resignation

Systeem Risico Indicatie (SyRi) is an AI system used by various Dutch municipalities and the national government "to prevent and combat fraud in the fields of social security and income-related schemes, tax and social insurance contributions, and labour laws".[40] It made the headlines after The District Court of The Hague judged, in early 2020, that it was non-compliant with Article 8 of the European Charter of Human Rights (ECHR) which stipulates that every citizen has the right to protection of their private life, with the benefits of new technologies needing to be weighed against it. SyRi links and analyses data from various public agencies and generates a risk report to assist in tackling the misuse of funds and detecting fraud. Although based on a legal basis with clear information on which data SyRi could capture, store, or share between different departments, the use of the system was highly controversial as it targeted and scrutinised mostly poor and vulnerable citizens as supposedly more likely to commit fraud.

[40] Rechtbank Den Haag 2020

A coalition of various civil society organisations and a large labour union complained on the grounds that the system was unfair and unjust as it did not screen all citizens equally and was only used in disadvantaged neighbourhoods: "if you only search in certain places, you will only find something in those places".[41] This highlighted the fact that unintentional links could be made on the basis of bias, such as a lower socio-economic status or an immigration background, especially considering that the data modelling methods were not open to scrutiny.

The problems the SyRi case showcased were further amplified by the recent Dutch Tax Authorities scandal. The secretive **Fraude Signalering Voorziening (FSV)** system supported incorrect risk analyses that led to people being incorrectly labelled as fraudulent. After several complaints, investigations showed that the system was using restricted data to detect signals of possible fraud, including entries registered in the FSV that did not have distinctions for meaning and severity, which caused the inclusion of incomplete, incorrect and outdated information.[42, 43]

This patchwork in practices did not only fail to comply with GDPR, it also created unclear and incorrect civil servant working practices regarding FSV data.[44] The resulting malpractices led to a parliamentary committee of inquiry report concluding that "fundamental principles of the rule of law have been violated" in reclaiming childcare support payments from parents identified as fraudsters for minor errors, such as missing signatures on paperwork.[45] Families, often from minority groups and immigrant backgrounds, were forced to pay back tens of thousands of euros with no means of redress, plunging many into financial and personal hardship. FSV is argued to have been at the heart of many of these incorrect fraud classifications.

Despite Government officials apologising for the scandal and earmarking 500 million EUR to compensate affected parents in March 2020, the Rutte Government resigned in early January 2021 to avoid losing a confidence vote in a parliamentary debate.[46]

Tracing "ghetto" models for children at risk

In Denmark, some local authorities ran an experiment attempting to trace young children who were vulnerable due to social circumstances. The **Gladsaxe model**, named after the suburban Copenhagen municipality that initiated the project, utilised a machine learning model that combined external information with data from different sources related to unemployment, healthcare and social conditions to analyse over 200 risk indicators. The model used a points-based system, with parameters such as mental illness (3,000 points), unemployment (500 points), and missing a doctor's (1,000 points) or dentist's appointment (300 points). Divorce was also included in the risk estimation, which was then rolled out to all families with children, to support identifying socially vulnerable situations. The model gave or deducted points from

[41] Blauw 2020

[42] Vijlbrief and van Huffelen 2020a

[43] Vijlbrief and van Huffelen 2020b

[44] KPMG 2020

[45] NL Times 2020

[46] BBC News 2021

families depending on the data found in the system. Children identified as at risk of abuse could then be targeted for an early intervention, possibly resulting in forced removals.[47]

The project received significant public backlash with complaints from civil society organisations and academics. Critics complained that the shift to algorithmic administration weakens government accountability, allows governments to consolidate their power and inevitably leads to increasingly draconian measures policing individual behaviour. In practice, the AI system was considered to pose a threat to liberal democracy, drawing comparisons to the Social Credit System used by the Chinese government.[48]

At first, the Danish government downplayed the criticism, emphasising the opportunity the Gladsaxe model offered for identifying children at risk earlier, and planned to roll it out nationwide. This was part of a larger "ghetto-plan" to fight "parallel societies", initiated in 2010.[49] The government's plan included using changing sets of criteria to help publish annual "ghetto lists", defining areas deemed to present a concentration of social problems.In these areas, special legal provisions would apply concerning crime prevention, integration, data protection, welfare and the allocation of public housing. For example, a 2018 initiative made it a legal obligation for children living in specific neighbourhoods to attend at least 25 hours of mandatory day-care from a week the age of twelve months. The same initiative also allowed for a doubling of criminal penalties in "ghetto areas".[50]

However, upon the unveiling of the scheme the Gladsaxe model used to evaluate children's well-being and development, it emerged that individual evaluations were prepared and stored without the knowledge of parents and in breach of existing legislation. In September 2018, the minister responsible mentioned a planned legal act, but by December of the same year the proposal for scaling up of the Gladsaxe model had been put on hold, despite some politicians still vouching for the system to be reinstated – although in adjusted form – in the future.[51]

Algorithmic profiling: the new glass ceiling

Similar to the Polish employment case discussed in section three, the **Austrian ArbeitsMarktService (AMS),** known as the AMS algorithm, is another example of public employment services (PES) using algorithmic profiling models to predict a jobseeker's probability of finding work, in a bid to cut costs and improve efficiency. AMS automates the profiling of job seekers to make its counselling process more efficient and to improve the effectiveness of active labour market programs. Based on statistics from previous years, the system calculates the future chances of job seekers on the labour market using the computer-generated "re-integration chance" indicator (IC value). In practice, the algorithmic system looks for connections between successful employment and job seeker characteristics, including age, ethnicity, gender, education, care obligations and health impairments as well as

[47] Thapa 2019

[48] Mchangama and Hin-Yan 2018

[49] The government's official use of such a historically loaded term as 'ghetto' has led to the Danish ghetto lists being widely discussed both within Denmark and beyond its borders. Source: Bendixen 2018.

[50] Seemann 2020

[51] Algorithm Watch and Bertelsmann Stiftung 2020

past employment, contacts with the AMS and the state of the labour market in the job seekers' place of residence. It then classifies job seekers into three groups based on their forecasted IC value: those with high chances to find a job within six months, those with a one-year prospect, and those likely to be employed within two years. Subsequently, different levels of assistance and resources for further education become available to the diverse categories of job seekers with the aim of investing primarily in those for whom the support measures are most likely to lead to reintegration into the labour market.[52]

The algorithm was strongly criticised by civil society organisations, journalists and academics and even the independent Volksanwaltschaft (ombudsman) raised concerns about its application. The criticisms stemmed from the perceived discriminatory elements within the algorithm, with specific regard to women or people aged over 50. Although it was partly made public (though only 2 out of 96 model variations were available), the algorithm was further criticised on other relevant points, including lack of transparency, bias in the system, and for diminishing caseworkers' ability to make independent decisions.[53, 54]

In fact, although the AMS system was only intended to provide staff with an additional function in the care of jobseekers, a recent study shows that it had far-reaching consequences for the entire organisation. These consequences included an increase in the efficiency of the counselling process, but only when associated with a predominantly routine adoption of the AI system, and an improvement in "training effectiveness" by concentrating funding on the middle of the three groups. On the other hand, it was confirmed that "in the development of the system, hardly any procedures were used to avoid bias in the system, and it does not offer any indications in its application to prevent possible structural inequalities in treatment", in particular with regard to gender equality.[55]

As this and the earlier Polish example show, these statistical methods are used to segment jobseekers into groups in a bid to improve identification of those at-risk of becoming long-term unemployed. But at the same time, they also induce discrimination. Predictive systems reflect institutional and systemic biases, and since they are based on past hiring decisions and evaluations, they can both reveal and reproduce patterns of inequity, penalising disadvantaged and minority groups, including women.[56]

Computer says no: nudging social service paths

Social protection marks another important area where governments are experimenting with AI. Among the examples emerging from many countries across the world, the **Trelleborg**

[52] Allhutter, et al. 2020

[53] Wimmer 2018

[54] Allhutter, et al. 2020

[55] Institute of Technology Assessment of the Austrian Academy of Sciences 2020

[56] Digital Future Society 2020a

Case deserves particular attention.[57, 58] In 2016 the Swedish municipality of Trelleborg began using a specific automated decision-making system based on robotic process automation (RPA) to manage welfare applications such as home care and sickness and unemployment benefits. RPA is an application governed by expert rule-based systems aimed at automating routine administrative tasks such as the calculation of home care fees and benefits, with an RPA case handler then executing the results. In practice, however, the software is usually based on different rules that lead to a yes or no decision that the case handler typically follows.[59]

This required the structuring and engineering of internal data and data about the applicant as well as the analysis and redesign of administrative processes. It shows how AI, implemented alongside a digital transformation process, can improve public administration operations.[60] The municipality argues, in fact, that they have considerably reduced the number of people receiving social benefits incorrectly and that future development would have allowed the program to learn how to perform more complex tasks, therefore, widening the scope of process automation within the public sector.[61]

However, despite the apparent success of the programme, which led the National Innovation Agency Vinnova, and the Swedish Association of Local Authorities and Regions, partnering with Trelleborg in a bid to replicate it in other municipalities, the system has faced resistance. From the outset, many social workers feared losing their jobs, understandably so as the number of caseworkers dropped from 11 to 3 and were uneasy about handing sensitive social tasks over to computers. Other Swedish municipalities aiming to follow the Trelleborg example also met opposition, with some staff members resigning. Case reports also mentioned the strong need for making the automation process trustworthy. While trying to increase efficiency, up to 15% of the system's decisions (up to 500,000 cases) were incorrect, leading to a shutdown of the system and many protests concerning the risk of excluding vulnerable citizens as RPA makes it more challenging to assess individual needs.[62]

In practice, as other cases of using AI systems to automate social welfare benefits decisions also show, the existence of both computer and paper-based documentation processes can lead to duplication and inefficiencies.[63] Moreover, a lack of trust in the use of AI obliges staff to double check all processes, which actually increases service time and reduces effectiveness.[64]

To grade or not to grade: the A-level disgrace

In 2020 the Covid-19 pandemic outbreak had a major impact on education systems worldwide. Given the critical situation, the UK government decided not to hold exams for students aged 16–18. As an alternative, the UK's exam regulator developed the **Ofqual grading algorithm system.** The aim was to find an objective way to standardise the final grades of all students, as Ofqual had found that an assessment based only on teachers' evaluations

[57] Misuraca and van Noordt 2020

[58] Engstrom, et al. 2020

[59] Algorithm Watch and Bertelsmann Stiftung 2020

[60] Codagnone et al. 2020

[61] UIPath, n.d.

[62] Wills 2019

[63] Ranerup and Zinner Henriksen 2019

[64] Wihlborg et al. 2016

would be unfair due to differences in schools.[65] The AI system, therefore, combined both previous grades as well as the teacher assessment to avoid inflation and maintain a proper distribution.[66]

On 13 August 2020 thousands of students in the UK received their A-level exam grades. Almost 40% of them received grades lower than they had anticipated based on teacher assessments, with 3% down two grades.[67] This sparked public outcry and legal action. The decision to optimise the algorithm to maintain standards and avoid grade inflation, instead led to other unexpected consequences. In particular, the algorithm's consistent downgrading of the results of those who attended state schools and upgrading of the results of pupils attending privately-funded independent schools drew heavy criticism. Effectively, due to the algorithm's behaviour around small cohort sizes, it was disadvantaging pupils from lower socio-economic backgrounds.[68]

In practice, bright and promising students from underperforming schools had much higher chances of having their grades lowered, reducing their chances of getting into their preferred university programmes.[69] In Scotland, for instance, the higher pass rate for students coming from the most disadvantaged groups was reduced by 15.2% compared with only 6.9% for those from wealthier backgrounds.[70]

Faced with wide criticism, the government announced that the results would be changed to the original teacher estimates. Furthermore, to prepare for the following year's exams, the government also announced a public consultation to seek views on the proposal that it should be the teachers' assessments of the standard at which a student performs throughout the year that should determine their grades. The UK exam debacle clearly illustrates the very real concerns that exist about when or how to ensure legitimacy when using AI to make decisions that will highly impact the life opportunities available to citizens.

[65] BBC News 2020a

[66] Taylor 2020

[67] Education Technology 2020

[68] Lee 2020

[69] The Conversation 2020

[70] BBC News 2020b

5. Lessons learned: turning away from dystopian futures

It is clear that for all their advanced capabilities and somewhat mythic reputation, AI systems face real-world issues when it comes to being smart, safe, and efficient tools to support government decision-making and the provision of public services.[71]

Naturally, AI alone cannot be held responsible for the bias and mistakes associated with the scandals outlined in previous sections. Nevertheless, the risks produced by heavily relying on machines serve to highlight, for example, as Philip Alston notes the systemic failure of some governments to protect vulnerable families from overzealous tax inspectors "generating mistakes on all levels that have led to great injustice for thousands of families and criminalizing innocent people".[72] The case studies demonstrate the imbalance between the state's economic interests to combat fraud, and the social interest of privacy, as confirmed by the Dutch Court in the **SyRi case.**

The great hope that AI is a benign technology, inherently more transparent, accountable, and fair than human decision-making has also been challenged. Not least by the lack of transparency and safeguards to guarantee individual rights that emerged in the **Gladsaxe case**, for example. This is of particular relevance when it comes to discussing if, and to what extent, certain situations justify the collection and combination of personal data, be it for ensuring child welfare or fighting a pandemic. Here AI use will resonate with the security and safety principles embedded in societies as well as the values that underpin them.

As a matter of fact, "models are opinions embedded in mathematics", as the data scientist Cathy O'Neil has written. "Despite their reputation for impartiality, they reflect human goals and ideology."[73] Models are useful because they let us strip out extraneous information and focus only on what is most critical to the outcomes we are trying to achieve. But they are also abstractions. Choices about what goes into them reflect the priorities of their creators. This is evident in the **AMS algorithm** which represents the transformation towards an "enabling state",[74] with a shift to activation regimes, turning rights-based access to welfare into consumer-oriented services.[75]

The inherent political nature of AI can also be found in the **Trelleborg case.** The AI system deployed there strongly improved one specific government process but could not ensure organisational interoperability, or gain the trust of the public or even internal staff, with expressed concerns including the risk of excluding vulnerable citizens, and "losing the control"

[71] Clasen 2021

[72] UN Human Rights, Office of the High Commissioner 2020

[73] O'Neil 2016

[74] Deeming and Smyth 2015

[75] Penz et al. 2017

through the automisation of all processes.[76] This shows how important it is to understand both the challenges related to the collection and analysis of data and the potential dangers derived from the design of proactive public services. The challenge is heightened even further due to the possible stigma effect attached to a person being classified a future problem at an early stage, as seen in the **Gladsaxe** case. And talking about the future, the scandalous results put forward by the **Ofqual grading algorithm** illustrate the risks of giving AI systems control of crucial decision effecting citizens' lives.

But does this mean algorithms will never be able to 'make the grades' or 'take decisions'?

[76] Codagnone et al. 2020

6. What to watch out for

Clearly, deploying AI in the public sector offers huge potential for improving the lives of citizens. Unfortunately, and as this whitepaper has shown, it is not a simple matter. To the contrary, unless AI is deployed sufficiently well, it will not only simply replicate existing human biases and limitations, but as applications become more sophisticated, these will increasingly go unperceived with the potential to cause serious societal harms.

Let's take the example of **facial recognition systems (FRS),** used by millions of people on a daily basis to log into their smartphones, organise their photos or secure their devices. As well as consumer applications, FRS has many other beneficial uses, such as assisting, for instance, blind and low-vision communities or helping law enforcement agencies find missing children and victims of human trafficking.

However, despite the promise of these supposed benefits, in the last two years a notable resistance towards such biometric technologies has emerged due to the risks to privacy, data protection and human rights their indiscriminate use poses.[77] Several cases involving the unlawful deployment of FRS have come to the attention of digital rights organisations and the general public all over the world. For instance, many cities have moved to ban police from deploying the technology over fears that it paves the way for potential privacy violations and mass surveillance.[78, 79]

In some cases, the piloting of facial recognition technology to identify potential criminals in public places has also been forbidden, such as at the Zaventem airport in Brussels.[80] Criticisms and negative advice have also been issued regarding requests to experiment on the use FRS in schools in France and Sweden. There is also a debate on the deployment of Body Worn Cameras (BWC) for policing in several countries.[81]

Similar applications were also tested in the UK and France. For instance, the **London Metropolitan Police** used two facial recognition cameras in King's Cross Station, one of London's most crowded places. The experiment lasted months, and the authorities had no concerns about transparency or thoughts about offering information mechanisms to passers-by whose data they had collected.[82]

On the contrary, in light of the Covid-19 security measures, the **Paris metro authority** tested using AI to detect whether travellers were wearing face masks by analysing closed circuit television cameras (CCTV) feeds. The initiative had been part of the city's efforts to help prevent the spread of the virus, but it sparked a warning from the data protection authority after a three-month test at the Chatelet-Les-Halles station in the heart of Paris. The station normally sees about 33 million passengers a year.[83]

[77] Moraes et al. 2020

[78] Gershgorn 2020

[79] Roussi 2020

[80] Misuraca and van Noordt 2020

[81] Misuraca et al. 2020

[82] Togawa and Deeks 2018

[83] Vincent 2020

The *Commission Nationale de l'Informatique et des Libertés, CNIL* (National Commission for Informatics and Freedoms) argued that this type of technology carries a risk that the identity of person analysed could be reconstructed and that the measures would also qualify under GDPR because the cameras will be collecting personal data without consent.[84]

The message emerging from this analysis is clear. As outlined in the February issue of the MIT Technology Review, eloquently titled This is how we lost control of our faces!, this technology has not simply eroded our privacy but has "fuelled an increasingly powerful tool of surveillance. The latest generation of deep-learning-based facial recognition has [also] completely disrupted our norms of consent."[85]

Reporting results from a recent study, complete with an analysis of the largest FRS survey ever conducted, including over 100 face datasets compiled from 1976 to 2019 and containing 145 million images of about 17 million subjects, offer some interesting insights. They suggest that the way advanced recognition technologies deeply impact on individual "intimacy" will have implications for how different facets of society respect privacy, as well as how this has evolved over the past 30 years.[86]

This gives an idea into how the parameters defining the use of FRS will be shaped over the next 30 years and beyond. As the authors underline in the conclusions: "FRS pose complex ethical and technical challenges. Neglecting to unpack this complexity, to measure it, analyse it and then articulate it to others, is a disservice to those who are most impacted by its careless deployment."[87]

But AI is not just about data, many more factors contribute to AI-enabled innovation. In addition to ensuring the availability of high-quality data for developing and adopting AI, it is also crucial to make sure its deployment aligns with the public sector's organisational scope and values, as well as the specific requirements the AI must meet.[88] For this, different policy options are being proposed, considering, for example, approaches based on ethics-by-design, ex ante conformity assessment or standard convergence, and the development of innovative public procurement.[89, 90]

Also, public trust is essential to ensure these systems are legitimate and effective, particularly when it comes to the public sector. The rapidly growing literature in the field, which shows the unique challenges the use of AI in government presents, confirms the importance of public trust, as does the attention numerous institutions, including the European Commission's AI Watch and the OECD AI Policy Observatory, pay it.[91, 92, 93, 94]

Considering that the development of AI is driven by the "combination of enormous amounts of data with powerful computation and sophisticated mathematical models", positive regulation, as Gruson and colleagues describe, should carefully consider and seek to address the risks that inaccuracy and lack of transparency pose.[95] There is a need, therefore, to ensure safeguards in the form of soft law, oversight, international standards and regulatory sandboxes for trialling.

[84] Fouquet 2020

[85] Hao 2021

[86] Raji and Fried 2021

[87] Ibid.

[88] Misuraca and Viscusi 2020

[89] World Economic Forum 2020

[90] UK Government 2021

[91] Desouza et al. 2020

[92] Sun and Medaglia 2019

[93] European Commission, n.d.

[94] Berryhill et al. 2019

[95] Gruson et al. 2019

There are moves advocating specific regulation approaches in both the USA and the EU, with each taking a different path. While the existence of such avenues in China and in other non-democratic countries is unclear, the number of so-called like-minded countries is growing, with the group sharing the need to find a common approach to develop responsible, human-centric AI.[96]

In this vein, an interesting example to watch out for is the **AI experiment in Espoo, Finland** that aimed to develop an evidence-based segmentation of social and health risks. The objective was to predict the future service paths for individuals, potentially allowing new forms of proactive care and prevention. Initiated, as part of the Six City Strategy for testing "Future societies" in Finland, the experiment used more than 37 million social and health related contact data points from approximately 520,000 residents.[97] The system integrated these data points with the childhood education data of all citizens from between 2002 and 2016, as well as data from private health care services and national statistics relating to basic social protection.

Although considered successful, the system has been put on hold for the time being. This is to allow discussions about the ethical concerns related to the role of the public sector in the development of such systems and the need to ensure citizen trust, as well as how to combine the various datasets while safeguarding privacy and security.

At the same time, and with an opposite viewpoint, it will be also interesting to follow the possible successor of SyRi, which aims to fight subversive crime that critics are already jokingly calling **SuperSyRi.**

This confirms it is not enough to be attentive to AI's technological aspects, including data quality and accuracy and algorithm transparency, but there is also a need to build trust in this disruptive technology. To this end, ethical and secure-by-design algorithms are crucial, but there is also a need for a broader engagement of civil society on the values to be embedded in AI and the directions future developments should take.[98]

[96] Feijóo et al. 2020

[97] Engels et al. 2019

[98] Ada Lovelace Institute 2020

7. Recommendations

If deployed wisely, AI holds the promise to address some of the world's most intractable challenges. But the likely destabilising effects AI can have on many aspects of economic and social life frustrate the significance of the positive impacts it can make.[99]

The multiple dilemmas faced by policymakers require further investigation due to the unforeseen implications and side-effects they may have. Below are seven recommendations to consider in this regard:

Beware of techno-solutionism

First of all, avoid thinking of AI as some sort of super-agent able to do more or less everything. Relying on automated methods follows an all too familiar pattern: stakeholders initially consider decision-making aids trustworthy then, after observing errors, distrust even the most reliable applications. Adopting faulty applications too early puts trust in the system at risk. Similarly, reliance on voluntary best practices and self-regulation can only do so much, with success depending on good faith from actors such as Facebook and other data processors.[100] This requires also taking into account the perceptions citizens have of data sharing, which may vary due to diverse cultural and administrative backgrounds and guaranteeing the possibility to include local content to ensure multiple perspectives are considered.

Be suspicious of ethical shortcuts

At the same time, be aware of the fact that AI-based technologies may, if superficially handled, infringe upon the principles of privacy and data protection to the extent that the collective security or quality of public service gains they offer cannot be justified. It is, therefore, important to maintain the link between the consideration of ethical risks and potential harms to social cohesion and the advantages in terms of efficiency or productivity that AI adoption offers a government body or agency. Carefully considering the barriers that could prevent public sector exploitation is essential, including looking at unintentional or unexpected effects as well as potential benefits, and comparing ex ante and ex post impacts. The focus should be on legal, technical and organisational aspects, but also on citizen acceptance.

[99] European Commission 2018b

[100] Kuziemski and Misuraca 2020

Look for concrete evidence

The actions of many governments worldwide clearly demonstrate the growing interest in exploring and experimenting with using AI to redesign public sector internal processes, enhance policymaking mechanisms and improve public service delivery. However, as there is still no straightforward evidence matching the positive impact expectations placed upon AI, the imbalance between potential and effective adoption of AI solutions must be underlined.[101]

Also, to duly address the ethical and political risks of using AI in the public sector, regulatory convergence towards a common approach to AI adoption is paramount. This should include re-using and sharing public service AI-based systems and solutions and engaging relevant stakeholders from academia, the private sector and civil society in the design of AI systems, as well as testing alternative solutions and assessing ex ante both conformity requirements and impacts.

Adopt a public value perspective

Adopting a public value perspective, focusing on the effective implementation of AI in both public administration and service delivery will address the complex challenges associated with the use of AI in government. In fact, it is vital to consider that, with AI, we are dealing with "boundary objects", a concept used in sociology to describe phenomena that "have different meanings in different social worlds but which structure is common enough to more than one world to make them recognisable means of translation".[102] In practice, the reasons for introducing AI and the perception of results achieved is different for diverse groups of stakeholders. Whereas for some, performance and accuracy are the most important feature to address, for others the traceability, transparency, and redressability options are fundamental. The same also applies to individual definitions of the "quality" of services, as related to the data or citizen satisfaction, for example.

Be ready to handle disruption

While experimenting with a variety of AI technologies in diverse policy domains, it is important to take into account the concept of "re-framing public sector innovation" which refers to "the need to consider both tangible changes in procedures, functions and institutions, as well as a 'cognitive restructuring' that concerns values, culture and shared understandings to articulate a reinforced set of values for the public sector ethos".[103] This meta-framing is required when

[101] Misuraca and van Noordt 2020

[102] Star and Griesemer 1989

[103] Misuraca et al. 2020

coping with complex and possibly disruptive and open-ended social dynamics, such as AI, to better evaluate the effects of direct and indirect consequences of action on institutions, citizens and society at large.[104] Ultimately, this will also imply the need to rethink how services are designed and delivered, the way data is shared and managed, and the manner algorithmic decision-making is implemented.

Look for stakeholder alliances

Acknowledging and appreciating the different opinions and levels of understanding that exist about AI among key groups in society is central for the success of complex endeavours, such as the adoption of AI in the public sector. This implies the need to carry out interdisciplinary analyses and undertake multi-stakeholder communication and interaction, in parallel to public sector transformation. In this context, it may be relevant to consider the potential effects AI could have on the achievement of the Sustainable Development Goals set out in the UN Agenda 2030. This would ensure the AI is not only trustworthy but also human-centric, harnessing its power to increase wellbeing for all.[105, 106]

Design new models of governance

Governance is a relevant concept for AI in three regards. Firstly, the use of AI opens up the potential for the public sector to achieve unprecedented gains and, secondly, it also opens up the capacity to nudge citizens towards behaving in one way or another, under the condition of ensuring an appropriate balance between personal privacy and human rights. This requires a commitment to governance of AI, guaranteeing that AI generates public value and is beneficial to all, and is not just seen as a goal in itself. Finally, it is necessary to learn how to govern the use of AI in the public sector to progressively link it to the wider impact it can have on various policy domains. Despite the limited number of successful implementations, it is crucial to identify and share use cases in order to learn from, replicate, scale and institutionalise AI into mainstream services.[107] Only in this way will we overcome the impasse of "ever-piloting" and "neverinstalling" what really works, while at the same time banishing forever the actual threats that put the stability of our societies at risk.

[104] Rossel 2010

[105] Vinuesa et al. 2020

[106] Feijóo et al. 2020

[107] Misuraca and van Noordt 2020

References

Ada Lovelace Institute. (2020). Examining the Black Box: Tools for assessing algorithmic systems. [online] Available at: https://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/

AI Council. (2021). AI Roadmap. Office for Artificial Intelligence, Department for Business, Energy & Industrial Strategy, and Department for Digital, Culture, Media & Sport. Gov.uk. [online] Available at: https://www.gov.uk/government/publications/ai-roadmap

Algorithm Watch and Bertelsmann Stiftung. (2020) Automating Society Report 2020. [online] Available at: https://automatingsociety.algorithmwatch.org

Allhutter, D., Cech, F., Fischer, F., Grill, G. and Mager, A. (2020). Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective. Frontiers in Big Data. [online] Available at: https://www.frontiersin.org/article/10.3389/fdata.2020.00005

Alston, P. (2019). Digital technology, social protection and human rights: Report. United Nations. [online] Available at: https://www.ohchr.org/EN/Issues/Poverty/Pages/DigitalTechnology.aspx

Alston, P. (2020) Landmark ruling by Dutch court stops government attempts to spy on the poor. UN Human Rights Office of the High Commissioner. [online] Available at: https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?LangID=E&NewsID=25522

Babuta, A. and Oswald, M. (2019). Data Analytics and Algorithmic Bias in Policing, Briefing paper, Royal United Services Institute for Defence and Security Studies. UK government's Centre for Data Ethics and Innovation. [PDF] Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/831750/RUSI_Report_-_Algorithms_and_Bias_in_Policing.pdf

Barenstein, M. (2019). ProPublica's COMPAS Data Revisited. Cornell University. [online] Available at: https://arxiv.org/abs/1906.04711v3

BBC News. (2020a). A-levels: Why are students so unhappy about this year's results? [online] Available at: https://www.bbc.co.uk/newsround/53803651

BBC News. (2020b). Scotland's results day: Thousands of pupils have exam grades lowered. [online] Available at: https://www.bbc.com/news/uk-scotland-53636296

BBC News. (2021) Dutch Rutte government resigns over child welfare fraud scandal. [online] Available at: https://www.bbc.com/news/world-europe-55674146

Bendixen, M. (2018) Denmark's 'anti-ghetto' laws are a betrayal of our tolerant values. The Guardian. [online] Available at: https://www.theguardian.com/commentisfree/2018/jul/10/denmark-ghetto-laws-niqab-circumcision-islamophobic

Berryhill, J., Kok Heang, K., Clogher, R. and McBride, K. (2019). Hello, World Artificial intelligence and its use in the public sector. OECD Working Papers on Public Governance No. 36, November 2019. [online] Available at: https://www.oecd.org/governance/innovative-government/working-paper-hello-world-artificial-intelligence-and-its-use-in-the-public-sector.htm

Big Brother Watch. (2020). Big Brother Watch briefing on Algorithmic Decision-Making in the Criminal Justice System. [PDF] Available at: https://bigbrotherwatch.org.uk/wp-content/uploads/2020/02/Big-Brother-Watch-Briefing-on-Algorithmic-Decision-Making-in-the-Criminal-Justice-System-February-2020.pdf

Blauw, S. (2020). An algorithm was taken to court – and it lost. The Correspondent. [online] Available at: https://thecorrespondent.com/276/an-algorithm-was-taken-to-court-and-it-lost-which-is-great-news-for-the-welfare-state/36504050352-a3002ff7

Bronstein, H. (2020). Rights group criticizes Buenos Aires for using face recognition tech on kids. Reuters. [online] Available at: https://www.reuters.com/article/ctech-us-argentina-rights-idCAKBN26U23Z-OCATC

Clasen, S. (2021). When the government uses AI: Algorithms, differences, and trade-offs. ASU. W. P. Carey News. [online] Available at: https://news.wpcarey.asu.edu/20210119-when-government-uses-ai-algorithms-differences-and-trade-offs

Codagnone, C., Liva, G., Barcevičius, E., Misuraca, G., Klimavičiūtė, L., Benedetti, M., Vanini, I., Vecchi, G., Ryen Gloinson, E., Stewart, K., Hoorens, S. and Gunashekar, S. (2020). Assessing the impacts of digital government transformation in the EU. Publications Office of the EU. [online] Available at: https://op.europa.eu/en/publication-detail/-/publication/7e715248-aac0-11ea-bb7a-01aa75ed71a1/language-en

Craglia, M., Annoni, A., Benczur, P., Bertoldi, P., Delipetrev, P., De Prato, G., Feijoo, C., Fernandez Macias, E., Gomez, E., Iglesias, M., Junklewitz, H, López Cobo, M., Martens, B., Nascimento, S., Nativi, S., Polvora, A., Sanchez, I., Tolan, S., Tuomi, I. and Vesnic Alujevic, L. (2018). Artificial Intelligence - A European Perspective. Publications Office, Luxembourg. [online] Available at: https://www.researchgate.net/publication/329449889_Artificial_Intelligence_A_European_Perspective

Danaher, J., (2016). The Threat of Algocracy: Reality, Resistance and Accommodation. Philosophy & Technology. [online] Available at: https://www.scinapse.io/papers/2242985385

Deeming, C. and Smyth, P. (2015). Social Investment after Neoliberalism: Policy Paradigms and Political Platforms. [online] Available at: https://www.cambridge.org/core/journals/journal-of-social-policy/article/social-investment-after-neoliberalism-policy-paradigms-and-political-platforms/C8E670BB1F0E2185F0EDDFB4B8C5AB8E

Dencik, L., Hintz, A., Redden, J. and Warne, H. (2018). Data Scores as Governance: Investigating uses of citizen scoring in public services. Open Society Foundations. [PDF] Available at: https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf

Dencik, L. Redden, J. Hintz, A. and Warne, H. (2019). The 'golden view': data-driven governance in the scoring society. Internet Policy Review. [online] Available at: https://doi.org/10.14763/2019.2.1413

Desouza, K., Dawson, G. and Chenok, D. (2020). Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector. Business Horizons, 63(2), 205–213. [online] Available at: https://doi.org/10.1016/j.bushor.2019.11.004

Digital Future Society. (2020a). Exploring Gender-Responsive Designs in Digital Welfare. [online] Available at: https://digitalfuturesociety.com/report/exploring-gender-responsive-designs-in-digital-welfare/

Digital Future Society. (2020b). Towards Gender Equality in Digital Welfare. [online] Available at: https://digitalfuturesociety.com/report/hacia-la-igualdad-de-genero-en-el-estado-de-bienestar-digital/

Douglas Heaven, W. (2020). Predictive policing algorithms are racist. They need to be dismantled. MIT Technology Review. [online] Available at: https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/

Education Technology. (2020). 36% of A-levels in England downgraded by Ofqual algorithm. [online] Available at: https://edtechnology.co.uk/he-and-fe/36-of-a-level-grades-in-england-downgraded-by-ofqual-algorithm

Engels, F., Wentland, A. and Pfotenhauer, S.M. (2019). Testing future societies? Developing a framework for test beds and living labs as instruments of innovation governance. Research Policy. [online] Available at: https://www.sciencedirect.com/science/article/pii/S0048733319301465

Engstrom, D. F., Ho, D. E., Sharkey, C. M. and Cuéllar, M. F. (2020). Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies. NYU School of Law, Public Law Research Paper No. 20-54. SSRN Electronic Journal. [online] Available at: https://doi.org/10.2139/ssrn.3551505

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. New York: St Martin's Press.

European Commission. (n.d.). AI for the public sector. [online] Available at: https://knowledge4policy.ec.europa.eu/ai-watch/topic/ai-public-sector_en

European Commission. (2018a). Artificial Intelligence for Europe. [online] Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:237:FIN

European Commission. (2018b). The Age of Artificial Intelligence Towards a European Strategy for Human-Centric Machines. European Political Strategy Centre. [online] Available at: https://ec.europa.eu/jrc/communities/sites/jrccties/files/epsc_strategicnote_ai.pdf

European Commission. (2020a). Shaping Europe's digital future. [online] Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0067

European Commission. (2020b). White Paper on Artificial Intelligence: a European approach to excellence and trust. [online] Available at: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

European Parliament. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union. [online] Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj

European Union Agency for Fundamental Rights. (2019). Facial recognition technology: fundamental rights considerations in the context of law enforcement. [online] Available at: https://op.europa.eu/en/publication-detail/-/publication/0de97f99-10db-11ea-8c1f-01aa75ed71a1/language-en

Feijóo, C., Kwon, Y., Bauer, J., M., Bohlin, E., Howell, B., Jain, R., Potgieter, P., Vu, K., Whalley, J. and Xia, J. (2020). Harnessing artificial intelligence to increase wellbeing for all: The case for a new technology diplomacy. Telecommunications Policy. [online] Available at: https://doi.org/10.1016/j.telpol.2020.101988

Feldstein, S. (2019). The Global Expansion of AI Surveillance. Carnegie Endowment for International Peace. [online] Available at: https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847

Fouquet, H. (2020). Paris Tests Face-Mask Recognition Software on Metro Riders. Bloomberg. [online] Available at: https://www.bloombergquint.com/politics/paris-tests-face-mask-recognition-software-on-metro-riders

Gershgorn, D. (2020). Live Facial Recognition Is Spreading Around the World. OneZero. Medium. [online] Available at: https://onezero.medium.com/live-facial-recognition-is-spreading-around-the-world-13f128c671dc

GOV.UK. (n.d.). Troubled Families Programme. [online] Available at: https://troubledfamilies.blog.gov.uk/

Gruson, D., Helleputte, T., Rousseau, P. and Gruson, D. (2019). Data science, artificial intelligence, and machine learning: Opportunities for laboratory medicine and the value of positive regulation. Clinical Biochemistry. [online] Available at: https://pubmed.ncbi.nlm.nih.gov/31022391/

Hao, K. (2021). This is how we lost control of our faces. MIT Technology Review. [online] Available at: https://www.technologyreview.com/2021/02/05/1017388/ai-deep-learning-facial-recognition-data-history/

High-Level Expert Group on Artificial Intelligence. (2019). Ethics guidelines for trustworthy AI. European Commission. [online] Available at: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-en

Institute of Technology Assessment of the Austrian Academy of Sciences. (2020). An Algorithm for the unemployed? Socio-technical analysis of the so-called "AMS Algorithm" of the Austrian Public Employment Service (AMS). [online] Available at: https://www.oeaw.ac.at/en/ita/projects/finished-projects/2020/ams-algorithm

Kharpal, A. (2017). Stephen Hawking says A.I. could be 'worst event in the history of our civilization'. CNBC. [online] Available at: https://www.cnbc.com/2017/11/06/stephen-hawking-ai-could-be-worst-event-in-civilization.html

Kolkman, D. (2020). The usefulness of algorithmic models in policy making. Government Information Quarterly, 37(3), 101488. [online] Available at: https://doi.org/10.1016/j.giq.2020.101488

KPMG. (2020). Rapportage verwerking van risicosignalen voor toezicht. KPMG Advisory N.V. [PDF] Available at: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2020/07/10/kpmg-rapport-fsv-onderzoek-belastingdienst/kpmg-rapport-fsv-onderzoek-belastingdienst.pdf

Kuziemski, M. and Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. Telecommunications Policy. [online] Available at: https://doi.org/10.1016/j.telpol.2020.101976

Lee, G. (2020). Did England exam system favour private schools? Channel 4 News. [online] Available at: https://www.channel4.com/news/factcheck/factcheck-did-england-exam-system-favour-private-schools

Mchangama, J. and Hin-Yan, L. (2018). The Welfare State Is Committing Suicide by Artificial Intelligence. Foreign Policy. [online] Available at: https://foreignpolicy.com/2018/12/25/the-welfare-state-is-committing-suicide-by-artificial-intelligence/

Misuraca, G., Barcevičius, E. and Codagnone, C. (2020). Exploring Digital Government Transformation in the EU – Understanding public sector innovation in a data-driven society. European Commission. [online] Available at: https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/exploring-digital-government-transformation-eu-understanding-public-sector-innovation-data

Misuraca, G. and van Noordt, C. (2020). AI Watch - Artificial Intelligence in public services: Overview of the use and impact of AI in public services in the EU. European Commission. [online] Available at: https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/ai-watch-artificial-intelligence-public-services

Misuraca, G. and Viscusi, G. (2020). AI-Enabled Innovation in the Public Sector: A Framework for Digital Governance and Resilience. International Conference on Electronic Government. EGOV 2020. [online] Available at: https://link.springer.com/chapter/10.1007%2F978-3-030-57599-1_9

Moraes, T. G., Almeida, E.C. and de Pereira, J.R.L. (2020). Smile, you are being identified! Risks and measures for the use of facial recognition in (semi-)public spaces. AI Ethics. [online] Available at: https://doi.org/10.1007/s43681-020-00014-3

Morozov, E. (2020). The tech 'solutions' for coronavirus take the surveillance state to the next level. The Guardian. [online] Available at: https://www.theguardian.com/commentisfree/2020/apr/15/tech-coronavirus-surveilance-state-digital-disrupt

NL Times. (2020). Parents faced 'unprecedented injustice' for years in childcare subsidy scandal. [online] Available at: https://nltimes.nl/2020/12/17/parents-faced-unprecedented-injustice-years-childcare-subsidy-scandal

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group, USA.

Oxford Insights. (2020). AI Readiness Index 2020. [online] Available at: https://www.oxfordinsights.com/government-ai-readiness-index-2020

Penz, O., Sauer, B., Gaitsch, M., Hofbauer, J. and Glinsner B. (2017). Post-bureaucratic encounters: Affective labour in public employment services. Critical Social Policy. [online] Available at: https://doi.org/10.1177/0261018316681286

Raji, I., and Fried, G. (2021). About Face: A Survey of Facial Recognition Evaluation. [PDF] Available at: https://arxiv.org/pdf/2102.00813.pdf

Ranerup, A. and Zinner Henriksen, H. (2019). Value positions viewed through the lens of automated decision-making: The case of social services. Government Information Quarterly 36, 101377. [online] Available at: https://doi.org/10.1016/j.giq.2019.05.004

Rechtbank Den Haag. (2020). SyRI legislation in breach of European Convention on Human Rights. de Rechtspraak. [online] Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-legislation-in-breach-of-European-Convention-on-Human-Rights.aspx

Richardson, R., Schultz, J. and Crawford, K. (2019). Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. 94 N.Y.U. L. REV. ONLINE 192. [online] Available at: https://ssrn.com/abstract=3333423

Rossel, P. (2010). Making anticipatory systems more robust. Foresight. [online] Available at: https://doi.org/10.1108/14636681011049893

Roussi, A. (2020). Resisting the rise of facial recognition. Nature. [online] Available at: https://www.nature.com/articles/d41586-020-03188-2

Seemann, A. (2020). The Danish 'ghetto initiatives' and the changing nature of social citizenship, 2004–2018. Critical Social Policy. [online] Available at: https://doi.org/10.1177/0261018320978504

Star, S. L. and Griesemer, J. (1989). Institutional ecology, 'translations' and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology. Social Studies of Science, Vol. 19, p. 387-420.

Sun, T. Q. and Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. Government Information Quarterly 36. [online] Available at: https://doi.org/10.1016/j.giq.2018.09.008

Taylor, R. (2020). Written statement from Chair of Ofqual to the Education Select Committee. GOV.UK. Ofqual. [online] Available at: https://www.gov.uk/government/news/written-statement-from-chair-of-ofqual-to-the-education-select-committee

Thapa, E.P. (2019). Predictive Analytics and AI in Governance: Data-driven government in a free society – Artificial Intelligence, Big Data and Algorithmic Decision-Making in government from a liberal perspective. European Liberal Forum. [PDF] Available at: https://www.liberalforum.eu/wp-content/uploads/2019/11/PUBLICATION_AI-in-e-governance.pdf

The Conversation. (2020). Gavin Williamson, Ofqual and the great A-level blame game. [online] Available at: https://theconversation.com/gavin-williamson-ofqual-and-the-great-a-level-blame-game-144766

Togawa Mercer, S. and Deeks, A. (2018). 'One Nation Under CCTV': The U.K. Tackles Facial Recognition Technology. Lawfare blog. [online] Available at: https://www.lawfareblog.com/one-nation-under-cctv-uk-tackles-facial-recognition-technology

UIPath. (n.d.). RPA in the Public Sector: UiPath Helps Swedish Citizens Regain Self-Sufficiency. [online] Available at: https://www.uipath.com/resources/automation-case-studies/trelleborg-municipality-enterprise-rpa

Vijlbrief, J.A. and van Huffelen, A.C. (2020a). Informatie over de Fraude Signalering Voorziening (FSV) en het gebruik van FSV binnen de Belastingdienst. Tweede Kamer Der Staten-General. [online] Available at: https://www.tweedekamer.nl/kamerstukken/detail?id=2020Z13850&did=2020D29414

Vijlbrief, J.A. and van Huffelen, A.C. (2020b). Kamerbrief Fraude Signalering Voorziening (FSV). Ministrie van Financien [PDF] Available at: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2020/04/28/kamerbrief-fraude-signalering-voorziening-fsv/Kamerbrief+Fraude+Signalering+Voorziening+%28FSV%29.pdf

Vincent, J. (2020). France is using AI to check whether people are wearing masks on public transport. The Verge. [online] Available at: https://www.theverge.com/2020/5/7/21250357/france-masks-public-transport-mandatory-ai-surveillance-camera-software

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Daniela Langhans, S., Tegmark, M. and Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications. [online] Available at: https://doi.org/10.1038/s41467-019-14108-y

Wihlborg, E., Larsson, H. and Hedström, K. (2016). "The Computer Says No!" -- A Case Study on Automated Decision-Making in Public Authorities. 49th Hawaii International Conference on System Sciences (HICSS). [online] Available at: https://ieeexplore.ieee.org/document/7427547

Wills, T. (2019). Sweden: Rogue algorithm stops welfare payments for up to 70,000 unemployed. Algorithm Watch. [online] Available at: https://algorithmwatch.org/en/rogue-algorithm-in-sweden-stops-welfare-payments/

Wimmer, B. (2018). Der AMS-Algorithmus ist ein Paradebeispiel für Diskriminierung. Kurier – futurezone. [online] Available at: https://futurezone.at/netzpolitik/der-ams-algorithmus-ist-ein-paradebeispiel-fuer-diskriminierung/400147421

World Economic Forum. (2020). The Global Risks Report 2020. World Economic Forum. [PDF] Available at: http://www3.weforum.org/docs/WEF_Global_Risk_Report_2020.pdf

Acknowledgements

Lead author

- **Gianluca Misuraca –** Vice President, Inspiring Futures

Supporting author

- **Tanya Álvarez –** Researcher, Digital Future Society Think Tank

Think Tank team

- **Carina Lopes** - Head of the Digital Future Society Think Tank
- **Patrick Devaney** - Editor, Digital Future Society Think Tank
- **Olivia Blanchard** - Researcher, Digital Future Society Think Tank

Citation

Please cite this report as:

- Digital Future Society. (2021). Governing algorithms: perils and powers of AI in the public sector. Barcelona, Spain.

Contact details

To contact the Digital Future Society Think Tank team, please email: thinktank@digitalfuturesociety.com

Digital
Future Society