Telecommunications Policy 44 (2020) 101976

Contents lists available at ScienceDirect

Telecommunications Policy

journal homepage: http://www.elsevier.com/locate/telpol

AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings

Maciej Kuziemski [a], Gianluca Misuraca [b,*]

[a] Berkman Klein Center for Internet and Society, Harvard University, Cambridge, MA, USA
[b] European Commission, Joint Research Centre, Digital Economy Unit, Seville, Spain

A R T I C L E I N F O

Keywords:
Artificial intelligence
Public sector innovation
Automated decision making
Algorithmic accountability

A B S T R A C T

The rush to understand new socio-economic contexts created by the wide adoption of AI is justified by its far-ranging consequences, spanning almost every walk of life. Yet, the public sector's predicament is a tragic double bind: its obligations to protect citizens from potential algorithmic harms are at odds with the temptation to increase its own efficiency - or in other words - to govern algorithms, while governing *by* algorithms. Whether such dual role is even possible, has been a matter of debate, the challenge stemming from algorithms' intrinsic properties, that make them distinct from other digital solutions, long embraced by the governments, create externalities that rule-based programming lacks. As the pressures to deploy automated decision making systems in the public sector become prevalent, this paper aims to examine how the use of AI in the public sector in relation to existing data governance regimes and national regulatory practices can be *intensifying* existing power asymmetries. To this end, investigating the legal and policy instruments associated with the use of AI for strenghtening the immigration process control system in Canada; "optimising" the employment services" in Poland, and personalising the digital service experience in Finland, the paper advocates for the need of a common framework to evaluate the potential impact of the use of AI in the public sector. In this regard, it discusses the specific effects of automated decision support systems on public services and the growing expectations for governments to play a more prevalent role in the digital society and to ensure that the potential of technology is harnessed, while negative effects are controlled and possibly avoided. This is of particular importance in light of the current COVID-19 emergency crisis where AI and the underpinning regulatory framework of data ecosystems, have become crucial policy issues as more and more innovations are based on large scale data collections from digital devices, and the real-time accessibility of information and services, contact and relationships between institutions and citizens could strengthen – or undermine - trust in governance systems and democracy.

1. Introduction

Artificial Intelligence is hardly a novel discipline, however the current amount of policy attention it gathers is a recently modern phenomenon. Already back in the 1960s, calls on Robert F. Kennedy to hold a conference on robotics and labour, and later to start a Federal Automation Commission were both dismissed (Calo, 2017). Although the questions of fairness and ethics of computer systems

* Corresponding author.
E-mail addresses: mkuziemski@cyber.harvard.edu (M. Kuziemski), gianluca.misuraca@ec.europa.eu (G. Misuraca).

https://doi.org/10.1016/j.telpol.2020.101976
Received 10 August 2019; Received in revised form 6 April 2020; Accepted 12 April 2020
Available online 17 April 2020
0308-5961/

have been already asked in the 1990s (Friedman and Nissenbaum, 1994), it is the report by US President's National Science and Technology Council Committee on Technology (The White House, 2016), that has captured the attention of as diverse fora as national governments, international organisations (eg. UNDP, OECD, EU) industrial groups (eg. IEEE, WEF, AAAI, Partnership on AI), and academia. Although scholars fail to agree on the definitional explanation of Artificial Intelligence for over half a century (Russel & Norvig, 2013), the working definition assumed in this document is in line with that proposed by the European Commission: *"systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals"* (European Commission, 2018a). Other than that, the terms *artificial intelligence, machine learning* and *automated decision-making* are used interchangeably.

Artificial intelligence - a deeply technical family of cognitive technologies, that includes i.a. computer vision, machine learning, natural language processing and robotics, is currently experiencing one of its peak hypes as measured by the volume of academic publications, intensity of investment, or regulatory interest. The rush to understand new socio-economic contexts created by the wide adoption of AI is justified by its far-ranging consequences, spanning almost every walk of life - from labour markets (Frey & Osborne, 2013), through human rights protection (Eubanks, 2018) to healthcare (Jiang et al., 2017). Yet, the public sector's predicament is a tragic double bind: its obligations to protect citizens from potential algorithmic harms are at odds with the temptation to increase own efficiency - or in other words - to govern algorithms, while governing *by* algorithms.

Whether such dual role is even possible, has been a matter of debate (Lodge & Mennicken, 2017). The challenge stems from algorithms' intrinsic properties, that make them distinct from other digital solutions, long embraced by the governments: vast computing power - exceeding human cognitive capabilities; 'learning' - autonomous knowledge creation happening without proper supervision; profiling - categorizing traits and behaviours; and nudging - incentivizing compliance - these all create externalities that rule-based programming lacks.

As the pressures to deploy automated decision making systems in the public sector intensify, it is important to examine how machine learning and bureaucracy have both "become generalisable modes of rational ordering based on abstraction and deriving authority from claims to neutrality and objectivity" (McQuillan, 2019). It is with that in mind, that we shall consider our research question: *how the use of AI in the public sector can be intensifying existing power asymmetries and governance practices?* The time to ask such question could not be better: scholars have been pointing to the limitations of focusing solely on *algorithmic bias* (Dave, 2019), calling for the politicization of the discourse on AI (McQuillan, 2018), and proposing more granular frameworks of *algorithmic accountability* (Pasquale, 2019).

Introduction of digital technologies in general, and in the public sector in particular is often portrayed as beneficial to the end users. Yet, are the processes happening under the banners of 'democratization', 'convenience' and 'choice' serving its advertised purposes? Or are these disguised attempts to strengthen the grip of control over the citizens? In other words - is AI facilitating the power shift between the public sector and citizens or merely intensifying existing distribution? Is the use of AI in the processes of governance changing the way power is exercised? Whatever the conclusions, these issues are not neutral to the general public.

2. Methodology

This paper is a part of a wider research project,[1] and focuses on the overview of existing AI-related legal and policy instruments and matching case studies of the use of AI in the public sector in three selected democratic countries.

The research underpinning the paper is structured in three stages: i) the analysis of legal and data governance implications of the use of AI in government; ii) the investigation of the complementarities in data and AI governance processes; and iii) the assessment of which AI governance methods best support trust and therefore strengthen government legitimacy.

This paper in particular serves as an input to a landscaping exercise of AI governance and regulatory frameworks in the EU and its comparison with countries that are considered vanguard in the field. It relies on the overview of existing legal and policy instruments that affect three selected OECD countries (Canada, Finland and Poland) as well as matching case studies of the use of AI in the public sector in those countries. These two elements are complemented by the forward looking analysis of the goals, drivers, barriers and risks for the use of AI in the public sector.

By deciding on the focus of the analysis, authors have attempted to select democratic countries that represent diverse socio-economic models of development and represent mid-to high-ranking position in the Government AI Readiness Index[2] (Oxford Insights, 2019). Similarly, case studies have been completed according to the principle of an informed-oriented selection (Flyvbjerg, 2006), allowing to derive critical insights about emerging paradigms and practices of the use of AI in the public sector, while providing an overview of diverse tech applications across government branches. In particular, the selection of three democratic countries have been dictated by the necessity to have a solid basis for a comparison (such as a commitment to the common regime).

The paper draws from a range of qualitative research methods, including literature and regulatory review, semi-structured interviews and case studies. In particular, to inform the case studies part of this paper, the authors have conducted a number of background interviews with public servants, delivery leads, government contractors, and academics. The findings have been clustered under the framework of *goals, drivers, barriers* and *risks* to account for underlying objectives and context of the use of AI in the public sector.

[1] The paper is results of a study on AI and Data governance conducted as part of the AI Watch, a joint initiative of DG CONNECT and the European Commission's Joint Research Centre.

[2] Poland - 27th, Canada - 6th, Finland - 4th.

3. Public sector and automated decision-making

3.1. Goals

Within the framework of the current narrative of the great potential of AI to transform our societies and economic systems, the potential benefits of this "new" set of technologies are indeed massive. But risks must also be governed while democratic values and human rights respected. For this reason, the EU in particular, aims to develop trusted AI based on ethical and societal values building on the European Charter of Fundamental Rights.

In this context, the public sector plays a vital role in the development and uptake of AI. However, most of the debate tends to place government either in the role of "regulatory actor" or at best "facilitator", setting out the framework conditions for private actors and citizens to use AI in an ethical manner. This leaves the alternative role of the public sector as "first buyer" and direct beneficiary of AI take-up and implementation rather obscure, if not neglected. In other words, the current policy discourse focuses on the governance "by" AI, far less on the governance "with" AI.

Indeed, under the first respect, it is to be stressed the direction taken by the EU Member States with the signature, already in 2018, of the Declaration on Cooperation on AI, containing the commitment of joining forces and engaging in a common policy approach, to leverage on the achievements and investments in AI of the European research and business community, while at the same time dealing with related social, economic, ethical and legal issues appropriately. This adds to the intense policy design work at national level, which has originated so far. These efforts document a firm intention of European governments to be the main actors in regulating the use of AI in society and stimulate its development by e.g. a more clear discipline of access to valuable data sets.

However, and mirroring a trend that has fastened its pace in the last 3–5 years in the private sector worldwide, the adoption of AI within public administration processes and internal operations has the potential to provide enormous benefits in terms of improved efficiency and effectiveness of policy making and service delivery to business and citizens, ultimately enhancing their level of satisfaction and trust in the quality of governance and public service. Nevertheless, the role of government as "user" of AI technologies has received far less attention than the "regulator" role in the strategies adopted so far.

Indeed, when used in a responsible way, the combination of new, large data sources with advanced machine learning algorithms could radically improve the operating methods of the public sector, paving the way to pro-active public service delivery models and relieving resource constrained organisations from mundane and repetitive tasks. However, the continuous collection and analysis of data combined with the use of AI-enabled systems by governments have also raised significant concerns about the power relations between the State and the citizens, while at the same time the opportunities for citizens to contest recommendations and results of the AI systems used for public services are also rising, with citizens finding unacceptable to use algorithms to make decisions with real-life consequences for humans, especially when it comes to: violations of privacy; lack of fairness; removal of the human element from important decisions; as well as the inability to capture the nuance and complexity of human nature (Pew Research Center, 2018).

3.2. Drivers

One of the seminal questions of our times is: what are the metrics of our collective success? With the growing global disenchantment with the current state of digital economy (also defined surveillance capitalism by Zuboff, 2015), the only constant is the intrinsic need to innovate. Whenever there is a problem – and regardless of what the problem is – the answer is to innovate this way or another: "it seems as if all governmental functions must cater to a discourse of innovation in order to appear economically defensible, politically legitimate and suited to this historical moment" (Pfotenhauer & Jasanoff, 2017).

The pursuit of novelty is a powerful driver that forces the governments into adopting technological solutions more often than perhaps necessary. Global thought-leadership circuit consisting of few key international fora, and a relentless nature of parliamentary politics of round the clock campaigning, forces political leaders to constantly overpromise and compare their country's performance to those of the immediate socio-economic surroundings. Hence, the narratives of competition and "arms race" is born and mirrored in the policy choices of the decision-makers. The introduction of 'business-like' practices that New Public Management brought into the government over 30 years ago, has translated into the wide adoption of market solutions, such as standards, incentives and benchmarks. These, have only been amplified by the turn towards evidence-based policy-making. However, with the rise of predictive analytics and automated decision making, the very nature of knowledge base is changing towards overwhelming quantification. Carving through insurmountable heaps of data, algorithms derive insights and find correlations that are not apparent to human cognition. Although regularly escaping scrutiny, their findings often pass as 'objective' and 'neutral'. It is by no means the bureaucrats fault to be tempted to embrace risk-based assessments. Assigning numeric value to any activity and hiding behind the machine produced 'evidence' shields imperfect humans from the accusations of bias, misdemeanor and inefficiency. At the same time, the drive towards rivalry and international benchmarking increases data collection. More fine grained data about citizens may mean increased oversight, and is in itself a powerful tool of global governance (Johns, 2017). Paraphrasing the old adage that "what gets measured, gets done", one could say that "what gets data mined, gets done" - data collection and analysis is indeed a powerful tool of meaning-making, ordering knowledge about the world and focusing our collective attention.

3.3. Barriers

Successful deployment of automated decision making in the public sector is subject to numerous barriers - some of them general, some of them context specific. First of all, although governments have followed various digitalisation pathways, many are held hostage by the sunk costs of the legacy IT systems. Negative experiences of developing and procuring new technologies in the past affects risk aversion levels and appetite for experimentation.

AI governance is a multi-level game characterized by the systemic resistance to steering, due to the sheer volume of actors, velocity of change and the perceived inevitability of the very technology at stake. High-level national AI strategies seem to have little to do with the lived experience of the bureaucrats dealing with citizens perplexed by unfavorable machine powered verdicts. A recent paper (Veale & Brass, 2019) has proposed a useful three level categorization of barriers to the development of public sector machine learning:

i) Macro level requires creation of new cross-cutting individual rights and obligations. These need to be supported by upskilling of bureaucrats who now need to be able to fully assess the intended and unintended consequences of AI against public values, such as accuracy, fairness, transparency and equity
ii) Meso level requires the development of more dynamic ways to measure, monitor and evaluate the inputs, process information, outputs, outcomes and impacts of public programmes using machine learning, which poses a challenge to established measures of public sector performance, quality and risk assessment
iii) Micro level requires the emergence of new tensions between the legitimacy of algorithmic decisions used in frontline service delivery, the discretion of street-level bureaucrats when employing, assessing or overriding automated decisions, and the rights of the data subjects when these processes are used to inform the allocation of public goods and services.

Agenda setting bottlenecks are further perpetuated by misaligned incentives, goals and measures: public sector's duties towards the citizens are at odds with those of the profit maximizing private sector. Agency problems are not uncommon - politicians driving the agenda have different goals and reward structures than ordinary bureaucrats. Finally, general AI principles and best practices can provide little guidance when contrasted with high-stake context-specific applications at the frontier of law enforcement, healthcare or service delivery. Much has been written about how workplace culture can hamper innovation in the public sector (Arundel, Bloch, & Ferguson, 2019). Skills shortage, lack of technology literacy and inability to meaningfully audit commissioned technologies are all potential hindering factors. Introduction of the comprehensive General Dave Protection Regulation (GDPR) in May 2018 as an overarching data governance mechanism (European Commission, 2016) could on its own terms be seen as a potential barrier, given how much emphasis it places on individual vs collective rights, and how inconsistent its enforcement has been across constituencies (European Commission, 2019).

3.4. Risks

Relying on automated methods follows an all too familiar pattern (Dzindolet, 2003) - stakeholders who initially consider decision aids trustworthy, after observing it makes errors happen to distrust even its reliable applications - too early adoption of faulty applications puts the trust in the system at risk. Similarly, public sector's reliance on voluntary best practices and self-regulation fares well, as long as no misdemeanor is found on the side of data processors - as exemplified by the public outrage and calls for regulation of Internet platforms that have continuously ignored its self-imposed standards. Introducing new resource-intensive processes inside the public sector - especially if they require reskilling and a lot of taxpayer dollars - enters the logic of path-dependency - it is much harder to abandon a flagship and politically salient project that has promised to "revolutionize" a given sector.

In a recent study, a number of scholars pointed to the abstraction traps specific to machine learning - or how algorithms fail to properly account for or understand the interactions between technical systems and social worlds (Selbst, Boyd, Friedler, Venkatasubramanian, & Vertesi, 2019):

- the **framing trap** - failure to model the entire system over which a social criterion, such as fairness, will be enforced;
- the **portability trap** - failure to understand how repurposing algorithmic solutions designed for one social context may be misleading, inaccurate, or otherwise do harm when applied to a different context;
- the **formalism trap** - failure to account for the full meaning of social concepts such as fairness, which can be procedural, contextual, and contestable, and cannot be resolved through mathematical formalisms;
- the **ripple effect trap** - failure to understand how the insertion of technology into an existing social system changes the behaviours and embedded values of the pre-existing system;
- The solutionism trap - failure to recognize the possibility that the best solution to a problem may not involve technology.

4. Case studies: strengthening border control, enhancing services delivery and improving user experience

As anticipated in the methodology section to investigate the various issues illustrated above in terms of goals, drivers, barriers and risks of AI in the public sector, case studies have been selected to guarantee a thematic and geographic diversity of examples of the public sector's use of automated decision-making systems in mid-to high-ranking countries within the *Government AI Readiness Index*. This section provides a summary of the context of case studies and the key elements derived from the analysis of specific systems in the countries.

4.1. Canada: immigration process control system

4.1.1. Canada's AI policy environment

Despite being a research institute, The Canadian Institute For Advanced Research (CIFAR) has played a leading role in driving domestic Canadian AI policy efforts. Most significantly, it has been in charge of developing the *Pan-Canadian AI Strategy*, launched in March 2017 and supported with CAD$100M government funding.[3]

Canada's ambition to shape global AI policy is realized also through international initiatives. The governments of Canada and France announced in July 2018 that they would work together to establish a new International Panel to guide the responsible adoption of AI systems that are human-centred and grounded in human rights, inclusion, diversity, innovation and economic growth (OECD, 2019).

Apart from implementing Canada's AI Strategy, CIFAR plays a significant role in setting the agenda for the future directions of policymaking in this area. For instance, its recent series of AI Futures Policy Labs (CIFAR, 2019) organized with the Brookfield Institute for Innovation and Entrepreneurship brought together 125 policymakers to explore the future frontiers of regulation, delivering recommendations across topics such as: reskilling, antitrust regulation, consumer protection, data protection, public awareness and incentivizing responsible innovation.

At the same time, the government has issued a *Directive on Automated Decision Making* in February 2019 (Government of Canada, 2019a). Its objective is to secure that automated decision systems are deployed in a possibly risk-free manner and lead to more efficient, accurate, consistent, and interpretable decisions. The Directive imposes a number of obligations on the public sector's delivery leads, including:

i) Algorithmic Impact Assessments (Government of Canada, 2019b), in the form of a questionnaire asking around 60 questions related to the business process, data and system designed decisions. The results demonstrate (on a scale from 1 to 4) a potential impact level towards the rights, health or well-being of individuals or communities; the economic interests of individuals, entities, or communities; and the ongoing sustainability of an ecosystem;
ii) Transparency standards, including the provision of notice before the introduction of the new system; a meaningful explanation of the system's decisions; ensuring the access to the system's software components; and releasing custom source code owned by the government;
iii) Assuring quality of the system in the form of: testing for biases before rolling out the system; monitoring the outcomes the system produces; ensuring the accuracy and relevance of the data used; providing review of the system; establishing internal oversight capabilities and literacy; conducting risk assessments and maintaining contingency plans; ensuring human intervention is possible;
iv) Providing recourse options to challenge administrative decisions;
v) Publishing regular reports on the system's efficiency.

One of the recent governmental guidelines (Government of Canada, 2018) proposes such AI applications to help deliver public service through:

i) Smarter search; natural language processing can advance public sector's user interaction. Over time the algorithms will learn patterns to better understand what citizens want when accessing government's services;
ii) Chatbots; it is envisioned that they could help filter routine questions away from human service agents so that they may focus on helping users through complex or distressing cases, or cases where a user is uncomfortable relaying their circumstances to a machine. They may also assist with public consultations on policies or programmes, by being able to ask follow-up questions and react to user feedback in a much more nimble fashion than a survey;
iii) Automated decision support; envisioned to increase quality of service by cutting wait times. For instance - AI can be applied to electronic forms to help ensure that data entered meets institutional standard of quality.

4.1.2. Automating immigration process

4.1.2.1. Drivers. As a part of an effort to make the immigration process more efficient, Canada has begun using automated decision making in sorting and filtering of people's applications as early as 2014. The urgency of the process is amplified by the fact that Canada is projecting to admit up to 340,000 new permanent residents annually until 2020 (Government of Canada, 2017).

Canada's immigration system is federally regulated by the Ministry of Immigration, Refugees and Citizenship Canada (IRCC). All initial immigration decisions are made by either an administrative tribunal such as the Immigration and Refugee Board or individual

[3] CIFAR's goal is two-fold: i) to boost Canada's scientific excellence by investing in three major centres (Edmonton, Montreal and Toronto) and increasing the number of outstanding AI researchers; ii) to develop global thought leadership on socio-economic implications of AI. Regional governments of Quebec and Montreal have provided additional CAD$120M funding for the research programmes. Three newly established institutes - Alberta Machine Intelligence Institute (Amii) in Edmonton, Mila in Montreal and the Vector Institute in Toronto are aggressively poaching top AI talent (so called Canada CIFAR AI Chairs), through a designated CAD$66M fund over the next five years. Up till June 2019, 46 researchers have been hired this way.

immigration officers employed by Immigration, Refugees and Citizenship Canada, or by the enforcement arm of the immigration system, the Canadian Border Services Agency (CBSA). These decisions are then reviewable either by an appeals body such as the Refugee Appeal Division and/or by the Federal Court of Canada and the Federal Court of Appeal, before moving up to the Supreme Court of Canada (University of Toronto, 2018).

4.1.2.2. Goals. IRCC has been developing a predictive tool to automate the activities currently conducted by immigration officials. Such system would recommend whether the immigration application merits acceptance, and spot potential red flags. As of June 2018, the system has been in a pilot mode within the Express Entry application stream. Since the actual pilot's scope has been a subject of a number of pending freedom of information requests, the following discussion outlines how automated decision-making could potentially affect the immigration system.

4.1.2.3. Risks. Even before setting foot on the Canadian soil, the system could determine factors such as i.a. whether one's application is complete, how likely it is that it is fraudulent, and if one's child is biologically or legally hers (University of Toronto, 2018). This raises several questions - what data would be harvested to determine these factors? How to appeal or redress the automated decision systems' verdict?

Upon arrival you can file for a number of temporary and permanent applications: a hypothetical automated decision system may shortlist you for a second screening or deny entrance entirely. How the data containing these decisions will be stored and shared between agencies? Will it impact the provision of other public services?

4.1.2.4. Barriers. The controversy is not limited to the system's opacity. Migrants and refugees are by definition in a rather precarious situation, and the decisions that they are subject to are complex, highly discretionary, and not easily reduced to a binary option. Experiences using predictive analytics in similar high stake contexts, such as policing, suggest that technological solutions are subject to the very same biases and errors as human decision-making, yet amplifying systemic injustices and automating inequalities, while providing suboptimal appeal routes and transparency standards.

Evidence suggests that the government seeks to expand its use of automated decision making in the administrative processes. In April 2018 a Request for Information has been published by the authorities, seeking industry inputs on a prospective tender commissioning "AI/ML powered solutions leveraging data-driven insights […] to assist and inform: legal research and development of legal advice/legal risk assessments; prediction of outcomes in litigation; and trend analysis in litigation" (Government of Canada, 2019c).

In September 2018, the University of Toronto's International Human Rights Programme and the Citizen Lab at the Munk School of Global Affairs and Public Policy have published a report (University of Toronto, 2018) scrutinizing the use of AI in the immigration system. Final recommendations to the federal government include: publishing a report on how automated decision systems are currently used within Canada's immigration and refugee system; freezing all efforts to procure, develop, or adopt any new automated decision system technology until existing systems are fully in compliance with a government-wide safety standard; and adopting a binding, government-wide standard for the use of automated decision systems. They also request: establishing an independent, arms-length body with the power to engage in all aspects of oversight and review of all use of automated decision systems by the federal government; creating a methodology to determine which public services are appropriate for the experimental use of automated decision systems and which are not; making complete source code for all federal government automated decision systems public and open source by default, and launching a federal Task Force to better understand the impacts of automated decision system technologies on human rights and the public interest.

The report on the use of automated decision-making in the Canadian immigration system, has caused a significant backlash and forced the government to reinvent its relation with the automated tools. Thanks to the pressure from academia and advocacies, some of the most controversial practices have been put to halt.

As a result, a guideline for the *Responsible use of AI* has been created. Its guiding principles include the government's commitment to (Government of Canada, 2018): understanding and measuring the impact of using AI by developing and sharing tools and approaches; being transparent about how and when using AI, starting with a clear user need and public benefit analysis; providing meaningful explanations about AI decision making, while also offering opportunities to review results and challenge these decisions. This implies being as open as possible by sharing source code, training data, and other relevant information, all while protecting personal information, system integration, and national security and defence,a s well as providing sufficient training so that government employees developing and using AI solutions have the responsible design, function, and implementation skills needed to make AI-based public services better.

4.2. Poland: employment services "optimisation"

4.2.1. Poland's AI policy environment

Polish political leadership displays serious commitment to the industrial focus, and the WEF-originated narrative of the "4th Industrial Revolution". In the words of Prime Minister Mateusz Morawiecki: "only active state involvement can secure truly sustainable development" whereas "the liberal model adopted in Poland in times of economic transition is no longer able to cope with the new economic conditions brought in by the Fourth Industrial Revolution".[4] Such premise has become a cornerstone of the government's flagship Responsible Development Strategy - an overarching master-plan setting country's industrial goals all the way until 2030 (European Commission Digital Transition Monitor, 2018).

Since its inception, the plan has introduced a number of new institutions that shall support the introduction of automation in the public and private sector. For instance, The Łukasiewicz Research Network[5] has been started by the Ministry of Science and Higher Education to solidify cooperation between some 37 state-owned institutes and help achieve economies of scale. At the same time, the Ministry of Entrepreneurship and Technology has initiated The Future Industry Platform Foundation[6] tasked with supporting the digital transformation of the industry and promoting automation among public and private sector alike.

In 2018, one of the ministerial research institutes has been elevated and transformed into the newly established Polish Economic Institute[7] that aims to build a knowledge base and specific expertise for the public sector. In 2019, the government has introduced a GovTech Poland[8] programme to incentivize modernization of the public sector with the help of crowdsourced start-up solutions. These developments have to be assessed against the backdrop of wider market conditions that affect Poland - the fourth largest IT graduates pool in the EU (McKinsey, 2016), and the fourth lowest Digital Economy and Society Index score in the EU (European Commission, 2018b).

In June 2018, Poland's Deputy Prime Minister Jarosław Gowin announced that the country will create its own AI Strategy. It has been drafted by a fairly heterogeneous group of over 130 stakeholders working pro bono over the course of few months under the guidance of the Ministry of Digital Affairs. The final document presented in November 2018 concludes the proceedings of four working groups: i) data-based economy; ii) financing and development; iii) education; and iv) law and ethics. Its key recommendations include: increasing the volume of targeted public sector procurement; investing €1.9B in AI development by 2023; positioning Poland as a leading AI vendor; creating a center of digital excellence; and establishing an ethical oversight mechanisms for public sector-led AI projects.

Interestingly, however, while the document extensively addresses the ethical and legal issues of automated decision-making in its last section, it does not propose any practical steps either towards safeguarding against misuse of AI, or towards its implementation within the public sector. In the words of one of the participants of the working group that prepared the Strategy: "*regulation is seen only as a means for providing more effective public* support *for research, prototyping and implementation of AI in the economy*".

4.2.2. Employment service "optimisation"

4.2.2.1. Drivers. Within this context, as early as in 2012, the Polish Ministry of Labour and Social Policy (MLSP) started working on the reform of 340 labour offices (PUP - Powiatowe Urzędy Pracy), charged with analyzing and supporting the development of the labour market.

The reform has been prompted by the global financial crisis and the subsequent domestic economic slowdown.

Its purpose was to minimize the structural problems of unemployment, such as low professional and territorial mobility and high risk of female post-maternity professional inactivity (Panoptykon Foundation, 2015).

4.2.2.2. Goals. The urgency of the reform has been underlined by the general perception of PUP's being inefficient, understaffed and unfit to address the challenges posed by the modern labour market. With that in mind - and without significant public spending increases - MLSP has scoped solutions that would ensure more efficient budget allocation. In this light, resorting to profiling using the automated system has come across as ticking all the boxes - both as a modern, cost efficient and individualized method of service delivery.

The process of profiling divides unemployed into three categories taking into consideration a number of characteristics. Assignment to a given category determines what types of programmes a subject is eligible for (i.a. job placement, vocational training, apprenticeship, activation allowance).

The system is based on data collected during an initial interview (i.a. age, gender, disability and duration of unemployment), and a subsequent computer based test that scores the unemployed across 24 different dimensions. The algorithmic process is opaque - the subject is aware neither of her score, nor of how certain features determine the final categorization. Remarkably, the rules guiding the algorithm have only been disclosed by MLSP after the lengthy legal battle started by a prominent non-profit.

Assignment to one of the three profile groups indicates the needed level of support and resource-heaviness. The first profile includes people with a high level of education, who are active and have enough professional qualifications to find a job relatively quickly (ePanstwo Foundation, 2019). The support they can count on could take the form of job training vouchers or seed funding to start a business. The second profile consists of subjects having more trouble re-entering the job market (i.a. due to the lack of skills) but showing great promise, which is why PUP's direct majority of resources towards this group. The final category makes up around 30% of the subjects who face serious obstacles preventing them from seeking employment (i.a. chronic diseases, disability, addiction) - due

[4] https://www.weforum.org/agenda/2017/01/middle-income-economies-fourth-industrial-revolution.
[5] https://lukasiewicz.gov.pl.
[6] https://przemyslprzyszlosci.gov.pl.
[7] http://pie.net.pl/en/o-nas.
[8] https://www.govtech.gov.pl/en/about-govtech-polska.

to resource constraints, almost none support is offered to this group. Importantly, in this case categorization translates into life-changing, binary decisions: state support or lack thereof.

4.2.2.3. Risks. The idea behind the profiling mechanism was to serve solely as an advisory tool, while retaining a human in the loop who would have a final say over the fairness of the categorization. Surprisingly - as one study has found - less than 1 in 100 decisions made by the algorithm have been questioned by the responsible clerks. Unless for the outstanding precision of the algorithm (which is to be considered unlikely), the reasons for not challenging automated decisions include lack of time to ponder its details; fear of repercussions from the supervisors; and a belief in the objectivity of the process - all in all rendering what was supposed to be an advisory mechanism the ultimate "automated" decision-maker.

4.2.2.4. Barriers. Unsurprisingly, the system has received a significant backlash, both internally and from the wider ecosystem. Many of the unemployed have complained through administrative courts, claiming the categorization to be unjust. The Panoptykon Foundation, that has analyzed the system in depth, has found its regulatory basis to be unclear: "[these] legal acts enigmatically determine what a profile is and how the procedure for its determination looks like." (Panoptykon Foundation, 2015). The Supreme Audit Office has carried a thorough control of PUPs, only to conclude the ineffectiveness of the profiling system and its potential to lead to discrimination. Finally, the Human Rights Commissioner has filed a formal complaint to the Constitutional Tribunal over a procedural issue, and the latter has ruled the profiling to be unconstitutional.[9] As of June 14th 2019, the profiling tool has been officially disbanded.[10]

4.3. Finland: addressing citizens' needs through "AuroraAI"

4.3.1. Finnish AI policy environment

Contrary to other countries, Finland has centred its AI policy efforts not only around increasing the competitiveness of the industry and making its public services more efficient, but also underlining the importance of the citizens wellbeing. The publication *Finland's Age of Artificial Intelligence* by the Ministry of Economic Affairs and Employment in 2017 (Finland Government, 2017) provides an overview of the country's ambition and understanding of its competitive edge.

Among its strengths, Finland identifies seamless collaboration between academia and the private sector, significant and long standing investments in the R&D, highly educated and tech-savvy population, as well as a convenient environment for piloting new products and services, thanks to a limited and harmonised market, as well as the culture of experimentation within the public sector. On top of that, Finland has a uniquely well functioning data governance infrastructure, exemplified by MyData[11] - a world class model for personal data management that offers individuals access to and control over the data collected about them.

At the same time, Finland is quite self aware regarding its weaknesses. Historic focus on the domestic market that is relatively small, lackluster prospects for the economies of scale, and weak international linkages are among the most important barriers for the country's sustainable growth. Over the course of two years in the making, Finland has formulated a vision that can be summarized through eleven commitments[12]:

Enhance business competitiveness through the use of AI
Effectively utilise data in all sectors
Ensure that AI can be adopted more quickly and easily
Ensure top-level expertise and attract top experts
Make bold decisions and investments
Build the world's best public services
Establish new models for collaboration
Make Finland a forerunner in the age of artificial intelligence
Prepare for artificial intelligence to change the nature of work
Steer AI development into a trust-based, human-centric direction
Prepare for security challenges

Adapted from www.tekloyaika.fi.

The country has also made great strides to offer free and accessible digital skills trainings. In June 2018, *The Elements of AI*[13] course has been launched, aiming to attract as much as 1% of Finnish population with a hassle-free introduction to machine learning that requires no prior experience. A year forward, almost 170,000 citizens have completed the course. Framing AI-related challenges in lay terms, and promoting the course to general audience underlines Finland's seriousness about inclusiveness and accessibility: senior citizens are encouraged to participate in the training and then share newly acquired knowledge and skills through the community centres and adult learning facilities. As Finland is grappling with the aging population, its public sector is placing bets on AI to increase

9 http://trybunal.gov.pl/postepowanie-i-orzeczenia/wokanda/art/10105-zarzadzanie-pomoca-kierowana-do-osob-bezrobotnych.
10 https://www.prawo.pl/kadry/bezrobotni-nie-beda-profilowani-utrudnialo-to-ich-aktywizacje,394701.html.
11 https://www.lvm.fi/-/finland-to-lead-the-way-in-mydata-980446.
12 https://www.tekoalyaika.fi/en/2019/06/11-key-actions-to-make-finland-a-leader-in-artificial-intelligence.
13 https://www.elementsofai.com.

the efficiency of the healthcare system, and become a global leader of innovative wellbeing solutions.

In March 2019, the Government's Analysis, Assessment and Research Centre has published a policy brief on *Finnish AI Competences* (Finland Governemnt, 2019a), comparing how the country scores across the board. For the purpose of analysis, AI has been divided into ten subfields.[14] Finland's strongest publishing record happens to be in *Platforms and services; Ecosystems*; *Robotics and machine autonomy;* and *Sensing and situation awareness*.

The Finnish Center for Artificial Intelligence (FCAI), initiated by Aalto University, University of Helsinki, and VTT Technical Research Centre of Finland has become the main hub for basic and applied research, animating the nation's ecosystem and facilitating knowledge transfer. In this perspective, Finland's Presidency of the Council of the European Union has made its priority to discuss digitalisation and data economy, through the lens of sustainability, well-being and rule of law (Finland Governemnt, 2019c).

4.3.2. AuroraAI

4.3.2.1. Drivers. In this context, Finland's Artificial Intelligence endeavours have been run in a different manner than other national strategies. Beyond publishing a comprehensive strategy, the government has attempted to change its own *modus operandi* with the use of AI services. One of the programmes that have been piloted is AuroraAI, "*an operations model based on people's needs, where artificial intelligence helps citizens and companies to utilise services in a timely and ethically sustainable manner*" (Finland Government, 2019b).

4.3.2.2. Goals. Run by the Ministry of Finance, and cutting across several departments, the programme attempts to organize public service provision in a more individualized way, with the help of reinforced learning. The pilot ran for five months, and was followed by a three month consultation process that concluded in April 2019. Rather than a single service, AuroraAI is an attempt to revolutionize current public management practices - from focused on efficiency and production outputs to placing customers need at the centre of the interaction.

This 'human-centricity' - as the documents call it - envisages an iterative development of public services with significant inputs from citizens and businesses, and through the lens of 'life-events', that is focal points/situations that require increased interaction between the citizen and the state. AuroraAI is thought of as a platform or a service network, where the public operator sets specific technological- and process-requirements, as well as ethical boundaries, allowing anyone to develop their value proposition within the platform. The rationale of the project stems from the growing sustainability gap of the public finances, and the deterioration of the dependency rate, and a hope that new, personalized service chains will cater better to the changing realities of the XXI century.

One of the experimental applications that has been run during the pilot focused on the "moving to a place of study" life-event. Two cities involved (Tampere and Turku) carried out well-being surveys among its student populations, and - based on its results - clustered the students into the groups in need of different support. The study has discovered that factors such as reliability of the public transportation, and quality of the natural environment plays an important role in students' well-being.

Data retrieved from the study helped create a new cross-cutting public offering - a bot that would supply information about the topics that students found important. However, the uptake of the new service has not been particularly high, leading the authorities to believe that only proactive marketing could change customers' behaviour. At the same time, it has been concluded that in order to pursue such projects further, municipalities need to develop user experience awareness and data analytics skills.

4.3.2.3. Barriers. After the conclusion of the pilot, it has been decided to roll out the €100M programme on a wider scale in 2019-23. During the consultation and planning phase, relevant stakeholders have pointed out to programme's ambition and the extent of the cultural change that it seeks, commenting that it may not be realistic to achieve with the current resources and timeframe. The programme has been praised for being developed incrementally and through experiments.

4.3.2.4. Risks. Programme's opaqueness makes it hard to comprehend for civil servants and to communicate to wider general public. Its success relies on the introduction of a common legal framework, and elaborate data protection impact assessment, to ensure sensitive data of Finnish citizens is properly managed.

Existing arrangements known from the MyData environment have been referenced as a best practice of purpose-limited personal data collection by the public administration.

During the next phase of the programme's rollout it is planned to run several other pilots of the "life-event" services, i.a. by preventing the marginalisation in school settings; and increasing the safety and convenience of active children. These pilots will need extra scrutiny, as the field of education is particularly prone to data spillages and mismanagement.

5. Discussion and way forward

The analysis of case studies has provided an overview of the regulatory approaches towards AI across three jurisdictions, along with three public sector use cases of automated decision systems. It has suggested that even seemingly trivial application of AI by the public sector can be an instrument of exercising control over the citizens: in the case of Canada, delineating the boundaries of political community by making value judgments on who can and who cannot enter; in the case of Poland classifying citizens' as a 'good

[14] See https://www.tietokayttoon.fi/julkaisu?pubid=29903

investment' or not; and in the case of Finland – as benevolent as it sounds – overhauling service provision into a perfect tool to centrally amass sensitive data about citizens.

Power has proven to be a central consideration for the use cases of AI in the public sector – by embracing automated methods, one gains control over the physical space, vital resources and information. Both Canadian and Polish cases have underlined the strong role civil society and academia can play in scrutinizing automated decision making systems – both at the stage of goal-setting, procurement and implementation. At the same time, it is becoming evident that the role of the state in AI policymaking is not to be downplayed – even if it takes a form of governing through adopting technological solutions at the center of its operations, and not writing laws.

That being said, current AI policy debate is heavily skewed towards voluntary standards and self-governance, somehow disregarding power-related considerations. In particular, recent meta analyses of AI ethical guidelines have pointed out (Jobin et al., 2019) to its reluctant approach towards enforcement mechanisms; lack of clarity to which norms should be prioritized; and a significant gap between agenda setting and its implementation. To the contrary, some have been hopeful about the potential of professional norms (Gasser & Schmitt, 2019) and soft, horizontal regulation (Veale & Brass, 2019), while others urged to move beyond GDPR's art. 22 as the basis of automated decision making governance (Jobin et al., 2019).

It is now key to discern and amplify voices that try to envision alternative ways to organize our digital world, either through rethinking how do we create public value (Mazzucato, 2016; Misuraca, Geppert, & Codagnone, 2017) or through reimagining our relation with the majoritarian powers (Zittrain, 2018), while remembering that the process is inherently political and needs to happen within the boundaries of democratic scrutiny (Kuziemski, Frahm, & Schioelin, 2019).

Finally, it is important to ask ourselves: "For what are we optimising?" (Bavitz & Hessekiel, 2018). No proper guidance for the public sector use of automated decision systems can fail to imagine the states of the world it envisions, and the values that it wants to support.

This is of particular relevance nowadays in light of the crisis that the COVID-19 Pandemia generated worldwide. As the COVID-19 outbreak rages across the world, governments have started developing specific applications to trace and track citizens' behaviours in an attempt to mitigate the risks of contagion. This has brought to the fore a new old dilemma, on "*whether key tenets of European democracies, including the protection of the fundamental right to privacy, should be set aside during the pandemic to enable a more effective response*" as for instance Andrea Renda emphasised in a recent blog post.[15]

The danger however is not on the emergency reaction, rather on the post-crisis management of data and digital infrastructures, and how governments will deal with the temptation of keep imposing restrictions on individual rights, such as privacy, and the control of the free movement of people, as well as the digital tracing that may be guaranteed by supposedly benevolent AI-enabled applications and predictive modelling systems, reflecting a sort of "Leviathan governance" (Misuraca, Broster, & Centeno, 2012) to reduce the spread to healthy individuals.

As governments, municipalities and public agencies around the world resort to automation in as diverse sectors as healthcare, law enforcement, and social services - sometimes with suboptimal, or downright unfair results - it is key to consider desired directions of the development of the field, and scrutinize existing algorithmic practices.

What goals should public sector organisations pursue when commissioning automated decision systems? Whose benefits should be prioritized? Despite existing body of work on decision support systems and Human Computer Interaction (HCI), the dilemmas policymakers are currently facing are far from straightforward and binary: *"If a police department turns to a machine learned predictive model to anticipate crime risk in different parts of a city, they face a range of debates. A desired end might be to treat all crime equally. But does that imply police should focus resources on areas of high crime at the expense of those with low crime, to maximise total arrests? Or does it mean that a crime in a low-risk area is just as likely to be intervened in as a crime in a high risk area? Areas conceived of as 'high risk' are rarely distributed at random, coupled instead to communities with different demographic or vulnerability distributions. The means are also unclear. Should models be used to increase preventative measures, such as community policing, or to heighten response capacity after crimes have been reported?"* (Veale, Van Kleek, & Binns, 2018).

Having these conversations - about the ends and means - while acknowledging the trade-offs and communicating with the populations at stake cannot be substituted by the creation of any indicator. Yet, a useful starting point is to think about the desired states of the world such interventions are set to advance.

Here once again the current COVID-19 outbreak comes in our help, as it is often remarked that crisis – like wars - are always dramatic accelerators of change. So as discussed by Geoff Mulgan in a recent blog post,[16] "*Coronavirus could be used to accelerate changes that were long overdue*" as it served as an extreme stress test for governments of all kinds and with specific impacts on digital resilience, institutional governance capacity and welfare systems.

Since the goal of many public sector's AI applications is to improve the productivity and quality of services - that is, not unlike many other technology agnostic efforts of the public sector - it is useful to assess the suitability of existing performance measurement tools, and how these could be further extended in the future, especially if the post-crisis situation will not bring us to "normality" rather force as collectively to change and embrace the complexity of our social systems and the preventative measures required to anticipate new challenges.

To this end, it is clear that a common framework for evaluating the potential impact of the use of AI in the public sector – and the specific effects of automated decision support systems on public services and their governance – is needed.

In this regard, it is generally accepted that the main difference in assessing private sector and public sector innovation, is the

[15] See https://www.ceps.eu/will-privacy-be-one-of-the-victims-of-covid-19.
[16] See https://www.geoffmulgan.com/post/how-not-to-waste-a-crisis-possibilities-for-government-after-covid-19.

possible absence of the market in the latter (Bloch & Bugge, 2013), and as a consequence, outcome monitoring has historically relied mainly on self-reported measures such as interviews and surveys (such as for instance Innobarometer or the European Innovation Scoreboard among others).[17]

The OECD's *Oslo Manual,* developed by the Working Party of National Experts on Science and Technology Indicators (NESTI), and last updated in 2018 (OECD, 2019), serve as a universal guideline for collecting data related to innovation.

It has been pointed out how it serves as a good basis for developing specific, public-sector related tools and indicators, as it relies on National Statistics Office (NSOs) that have extensive data gathering experience; allows for collecting insights in a comparable format between private and public sectors; allows to compare advantages and disadvantages of public and private sector provision of service innovations; and is designed for highly heterogeneous environments, while maintaining flexibility needed for sector-specific questions (Arundel et al., 2019). Among other popular survey, the Australian Public Sector Innovation Indicators (APSII) focuses on five themes: inputs, processes, outputs, outcomes, and environmental conditions that affect innovation (Australian Government, 2017), while the EU's European Public Sector Innovation Scoreboard (*unfortunately discontinued*) has proposed a more robust classification: enablers (human resources and quality of public services), activities (capacities, drivers and barriers), and outputs (innovators, effects on business performance and government procurement) (European Commission, 2013).

Yet another avenue for standard and indicator development is the development of voluntary guidelines that have been particularly popular in the context of AI policy. For instance, the European Commission's High-Level Expert Group on AI, after a year-long deliberative process, published the *Ethics Guidelines for Trustworthy AI* (European Commission's HLGE on AI, 2019), complemented by over 500 comments received through the open consultation. The report has identified seven requirements deemed to be key for the ethical development of AI-related technologies. These include: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental wellbeing; and accountability. Such fairly general guiding poles are complemented by a pilot *Trustworthy AI Assessment List* that includes a set of 140 context- and sector-agnostic questions linked to the seven requirements, and is addressed to the *'developers and deployers of AI systems'* as well as *'AI practitioners'*. This pilot list is set to be reviewed after gathering the feedback on its usability from relevant stakeholders.

Building on the preliminary results of this analysis and case studies, further research is therefore recommended to develop a standardized format to catalogue and promote best practices of the use of AI in the public sector with the aim to create a wide knowledge base and benchmark of the use of Automated Decision Systems in the Public Sector.

In a longer term perspective, and taking a regional approach (e.g., in the EU), it is also worth organizing surveys and public sensing to test citizens' acceptance and approach of automated decision making in specific sector domains.

For this purpose a methodology to determine which public services are appropriate for the experimental use of automated decision systems and which are not should also be created and tested through *ad hoc* pilot experiments.

Results of these research actions would feed insights into the policymaking processes at EU and global level, and mechanisms for enabling peer to peer learning and knowledge sharing between policy makers and public sector delivery leads, with the support of research institutes, non-profits and private sector should also be facilitated. This would incentivize and reward the beneficial deployment of AI within national governments, as well as shedding light on the controversies and risks of Automated Decision Systems in sensitive public services domain and policy areas.

This is of particular importance at a time when most societies have transformed immensely due to the rapid adoption of new digital technologies in all aspects of their lives, thanks to the combined advances in computing power, high availability of data and enhanced algorithms, further brought to the stage by the current emergency created by the Coronavirus pandemia.

However, while the positive effects of major breakthroughs possible through embracing AI in general and Machine Learning in particular are often emphasised, their potential negative consequences and risks on human conditions require that socio-economic, legal and ethical impacts are carefully addressed and anticipated.

This raises expectations for governments to play a more prevalent role in the digital society and to ensure that the potential of technology is harnessed, while negative effects are controlled and possibly avoided. In particular, the governance of AI and its underpinning data ecosystem has become a crucial policy issue as more and more innovations are based on large scale data collections from digital devices, and the real-time accessibility of information and services, contact and relationships between institutions and citizens could strengthen – or harm - democracy and trust in governance systems.

Services could be redesigned around latest technologies to make them more citizen-oriented and valuable for society. High-quality – personalized - public services and efficient public administrations could lead to higher welfare and improve the business environment as administrative procedures could be simplified. Especially businesses operating across borders will significantly benefit from increased digitalisation of public administrations and services.

Within this context, the combination of emerging technologies and paradigms such as Artificial Intelligence, Internet of Things, Edge computing, and Distributed Ledger Systems (DLTs) such as Blockchain, with topic domains such as 'Smartcities' or Application Programming Interfaces (APIs) among others, makes the governance "with and of" digital technologies decisive for the future of our societies and is recognized as an area of strategic importance and a key driver of economic growth, as it can radically improve the functioning of government and even change the way institutions are designed.

From our side, building on the initial landscaping of the regulatory approaches through selected case studies we aim to further analyze the synergies between existing data governance infrastructures and AI policies as well as discussing about the use of AI for

[17] https://ec.europa.eu/growth/industry/innovation/facts-figures/scoreboards_en.

government legitimacy and citizen empowerment, two key issues currently high on the policy agenda at global level, given the strict relationship they have with the future of our democratic systems and the need to rethink how governing in the digital age.

Disclaimer

The views expressed in this article are purely those of the authors and may not be regarded as stating the official position of the [European Commission] or any of the organisations they are affiliated to.

Acknowledgements

Work on this paper has been in part conducted under the contract CT-EX2018D339879-101 funded by the ISA[2] ELISE Action, and in support of the activities on AI for the public sector of the AI Watch, a joint initiative of DG CONNECT and the European Commission's Joint Research Centre.

References

Arundel, A., Bloch, C., & Ferguson, B. (2019). Advancing innovation in the public sector: Aligning innovation measurement with policy goals. *Research Policy*.
Australian Government APSII Project. (2017). *Wrapping up the Australian public sector innovation indicators*.
Bavitz, T. C., & Hessekiel, K. (2018). *Algorithms and justice, medium, july 2018*.
[Blind For Review]. (2012). Digital Europe 2030: Designing scenarios for ICT in future governance and policy making. In *Government information quarterly, volume 29, supplement 1, january 2012* (pp. S121–S131). https://doi.org/10.1016/j.giq.2011.08.006.
Bloch, C., & Bugge, M. (2013). *Public sector innovation - from theory to measurement, structural change and economic dynamics*.
Calo, R. (2017). *Artificial intelligence policy: A primer and roadmap*.
CIFAR. (2019). *Rebooting regulation: Exploring the future of AI policy in Canada, may 2019, Canada*.
Dave, K., Systemic algorithmic harms, data & society, may 31st 2019.
Dzindolet, M., et al. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*.
Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St: Martin's Press.
European Commission. (2013). *European union public sector innovation scoreboard 2013*.
European Commission. (2016). *General Data Protection Regulation 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 of 27 April 2016, (OJ L 119 4.5.2016)*.
European Commission. (2018). *Artificial intelligence for Europe* (Vol. 2018, p. 237). COM (final).
European Commission. (2018). *The digital economy and society Index (DESI)*.
European Commission. (2019). *Data protection rules as a trust-enabler in the EU and beyond – taking stock s, 24.7.2019 COM* (Vol. 2019, p. 374) (final).
European Commission Digital Transition Monitor. (2018). *Poland "initiative for polish industry 4.0 – the future industry platform", february 2018*.
European Commission's High Level Group of Experts on AI. (2019). *Ethics guidelines for trustworthy artificial intelligence, April 2019*.
Finland Government. (2017). *Finland's age of artificial intelligence, Ministry of economic Affairs and employment*. Available at: https://www.tekoalyaika.fi/en/reports/finland-leading-the-way-into-the-age-of-artificial-intelligence/3-eleven-key-actions-ushering-finland-into-the-age-of-artificial-intelligence/.
Finland Government. (2019a). *Finnish AI Competences and how to make them stronger, March 2019*.
Finland Government. (2019b). *Implementation of the national AuroraAI programme, Ministry of finance, March 2019*.
Finland Government. (2019c). *Sustainable Europe, sustainable future - Finland's presidency programme, June 2019*.
Flyvbjerg, B. (2006). *Five misunderstandings about case-study research, qualitative inquiry*.
ePanstwo Foundation. (2019). *alGOVrithms. Are citizens governed by computer programmes?*.
Frey, C. B., & Osborne, M. A. (2013). *The future of employment. How susceptible are jobs to computerisation?* Oxford Martin School.
Friedman, B., & Nissenbaum, H. (1994). *Bias in computer systems*. Stanford, CA: Center for the Study of Language and Information, Stanford University.
Gasser, U., & Schmitt, C. (2019). The role of professional norms in the governance of artificial intelligence, forthcoming. In Markus D. Dubber, Frank Pasquale, & Sunit Das (Eds.), *The oxford handbook of ethics of AI* (p. 2019). Oxford University Press.
Government of Canada. (2017). *Annual report to parliament on immigration*. Ministry of Immigration, Refugees and Citizenship.
Government of Canada. (2018). *Responsible artificial intelligence in the government of Canada, digital disruption white paper series, April 2018, Canada*.
Government of Canada. (2019a). *Directive on automated decision making, february 2019, Canada*.
Government of Canada. (2019b). *Algorithmic impact assessment, GitHub repository*. Canada.
Government of Canada. (2019c). *Call for tenders: Artificial intelligence solution (B8607-180311/A)*.
Jiang, F., et al. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*.
Jobin, A., Ienca, M., & Vayena, E. (2019 preprint). *Artificial intelligence: The global landscape of ethics guidelines,*; principled artificial intelligence project; AI ethics guidelines global inventory. *Algorithm Watch, Zurich, CH*.
Johns, F. (2017). Data mining as global governance. In *The oxford handbook of law, regulation and technology*.
Kuziemski, M., Frahm, N., & Schioelin, K. (2019). *Zuckerberg and Warren want to fix Big Tech in different ways - and neither of them will work*. Quartz 2019. available at https://qz.com/1601587/elizabeth-warren-and-mark-zuckerbergs-differing-views-on-big-tech.
Lodge, M., & Mennicken, A. (2017). The importance of regulation of and by algorithm. In K. Yeung (Ed.), *(2017). Algorithmic regulation: A critical interrogation, regulation & governance*.
Mazzucato, M. (2016). Innovation, the state, and patient capital. *The Political Quarterly*.
McKinsey. (2016). *Digital Poland. Capturing the opportunity to join leading global economies, USA*.
McQuillan D., Rethinking AI through the politics of 1968, open knowledge, october 13th 2018.
McQuillan D., AI realism and structural alternatives, data justice Lab, June 7th 2019.
OECD. (2019). *Artificial intelligence in society*. Paris: OECD Publishing.
Misuraca, G., Broster, D., & Centeno, C. (January 2012). Digital Europe 2030: Designing scenarios for ICT in future governance and policy making. *Government Information Quarterly, 29*(1), S121–S131.
Misuraca, G., Geppert, L., & Codagnone, C. (2017). i-FRAME – Assessing impacts of social policy innovation in the EU. *JRC Science for Policy Reports Series*. European Commission's Joint Research Centre. https://doi.org/10.13039/10.2760/83089. EUR 28824 EN.
OECD. (2019). *Oslo Manual 2018 - Guidelines for Collecting, Reporting and Using Data on Innovation* (4th). .. available at. http://www.oecd.org/sti/oslo-manual-2018-9789264304604-en.htm.
Oxford Insights. (2019). *Government AI readiness Index*. https://www.oxfordinsights.com/ai-readiness2019.
Panoptykon Foundation. (2015). *Profiling the unemployed in Poland: Social and political implications of algorithmic decision-making*.
Pasquale, F., The second wave of algorithmic accountability, law and political economy, november 25th 2019.
Pew Research Center. (2018). *Public attitudes toward computer algorithms, november 2018, USA*.
Pfotenhauer, S., & Jasanoff, S. (2017). Panacea or diagnosis? Imaginaries of innovation and the 'MIT model' in three political cultures. *Social Studies of Science*.

Russel, S., & Norvig, P. (2013). *Artificial intelligence: A modern approach*.
Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency FAT* '19*.
University of Toronto, Faculty of Law and the Citizen Lab, Munk School of Global Affairs and Public Policy. (2018). *Bots at the gate: A human rights analysis of automated decision-making in Canada's immigration and refugee system.* International Human Rights Programme.
Veale, M., & Brass, I. (2019). Administration by algorithm? Public management meets public sector machine learning. In Y. Karen, & L. Martin (Eds.), *Algorithmic regulation* (p. 2019). Oxford University Press.
Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In *Proceedings of the 2018 CHI conference on human factors in computing systems (CHI'18)*.
White House. (2016). *Artificial intelligence, automation and the economy*. USA Government.
Zittrain, J. How to exercise the power you didn't ask for?, Harvard Business Review, September 2018.
Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*.