ALGORITHMIC RULE

AI AND THE FUTURE OF DEMOCRACY IN SWEDEN AND BEYOND

EDITED BY
SIMON VINGE
MAJA FJAESTAD

ALGORITHMIC RULE

AI AND THE FUTURE OF DEMOCRACY IN SWEDEN AND BEYOND

Book published in November 2025 by:

FEPS
FOUNDATION FOR EUROPEAN PROGRESSIVE STUDIES

FOUNDATION FOR EUROPEAN PROGRESSIVE STUDIES (FEPS)
Avenue des Arts, 46 – 1000 Brussels, Belgium
T: +32 2 234 69 00
www.feps-europe.eu
@FEPS_Europe

Akademikerförbundet SSR

AKADEMIKERFÖRBUNDET SSR
Box 128 00, 112 96 Stockholm, Sweden
T: +46 8 617 44 00
https://akademssr.se/
@akademssr

Editors: S. Vinge, M. Fjaestad
FEPS Project Coordinators: M. Aal, L. Saez Jimenez, V. Zwisele
English language editor: R. Cowie
Design, layout and printing: O. Wydawnicza (ASPRA-JR)
Cover design: E. Grandin (Zie)
Translation: Eurominds

This book was produced with financial support from the European Parliament. It does not represent the view of the European Parliament.

DIGITAL PROGRAMME

This book is published as part of the FEPS-Nordic Digital Programme 2025-2026, a collaboration between the FEPS and its members and trade unions from Denmark, Sweden, Norway and Finland. It is based on the Swedish version "AI & makten över besluten" and was translated into English and complemented by three additional chapters.

KBR deposit number: D/2025/15396./32
ISBN: 978-2-39076-037-5

ALGORITHMIC RULE

AI AND THE FUTURE OF DEMOCRACY IN SWEDEN AND BEYOND

Edited by
Simon Vinge
Maja Fjaestad

TABLE OF CONTENTS

Foreword

It is difficult to go a day without reading, hearing or seeing something related to artificial intelligence (AI). What is still often referred to as "emerging technology" has become ubiquitous in our lives, both our professional and private ones. From tangible examples of AI, such as ChatGPT, to obscure AI systems deployed by governments and companies alike that affect us from a distance, this technology is influencing our lives in a myriad of ways.

The rapid (and not always successful) adoption of AI across Europe has also forced us to reckon with the impact of this technology on our society. Looking in our own backyard here in Brussels, we saw how EU legislators responded by passing the world's first comprehensive AI law, the AI Act, which established rules for the development and use of AI across all sectors. New scholarships, reports and other publications emerge daily, joining a crowded corpus in which it is sometimes difficult to discern work that offers a novel insight or perspective.

This book does exactly that. Simon Vinge and Maja Fjaestad have curated a selection of essays that offer readers a series of diverse and multidisciplinary reflections on AI in our lives. The authors argue that, while big tech might keep our minds busy speculating about doomsday scenarios, the real change is going on beneath the surface, within our public administrations and at our workplaces.

The volume moves from an analysis of developments from Sweden, a country with some of the world's strongest institutions and highest levels of freedom of information, to reflect on the impact that AI will also have on the rest of Europe. It shows how democratic institutions are fundamentally affected and challenged by this transformation, for example, when the use of AI in public services negatively impacts the trust that citizens have in the state.

By taking us through real-life examples of the challenges posed by the usage of AI, the book offers us a warning as well as a clear call to action. Without deliberate and strong action to shape the trajectory of AI in a progressive way – one that centres the public interest and human wellbeing – we risk becoming increasingly intertwined with and dependent on

technology over which we have no meaningful control. The good news is that a positive future for AI is possible, and this book provides inspiration, alternatives and concrete suggestions to make it a reality.

Dr. László Andor
Secretary General,
Foundation for European Progressive Studies

Simon Vinge and Maja Fjaestad

Introduction: From algocracy to democracy

Many people probably think of Sweden as both digital and open. A country with well-developed infrastructure and technology, where transparency, the principle of public access and freedom of information are key to both administration and governance. This perception aligns with Sweden's self-image and is, to some extent, accurate. However, this book highlights how the digital transformation raises entirely new questions for public administration, governance and democracy – questions that Sweden and other countries have yet to answer. At the same time, technological developments threaten the very openness, transparency and principle of public access that have been carefully constructed over the centuries.

The chapters in this book highlight a range of different challenges confronting democratic societies today. As public administration becomes digitalised, there is a risk of turning bureaucracy into *algogracy* – a concept we will return to shortly. Even in countries that explicitly and constitutionally seek to guarantee transparency in state actions, the introduction of new technologies tends to violate the rule of law, unless these principles are actively safeguarded. The opening examples are well known, yet they merit renewed attention in a Europe shaken by numerous scandals involving biased, discriminatory and opaque use of artificial intelligence (AI) and algorithms. Johan Hirschfeldt opens the book by addressing the questions raised by one such high-profile algorithmic scandal – questions crucial for countries committed to upholding fundamental legal principles. Readers across nations will find useful insights from Swedish reflections on these international cases.

Much of the public debate about AI has focused on the technology itself and its applications within the private sector. In this book, however, we want to focus on the public sphere, and specifically state administrations and institutions whose rules govern our everyday life – as well as our work life. Algorithms, sometimes AI-driven, are becoming an increasingly integral part of public administration in Sweden and beyond. Even

citizens' routine encounters with bureaucracy are frequently mediated by chatbots, as discussed here by Anne Kaun. While this might enhance measured efficiency, it also raises important questions about what else is at stake in these encounters. This new technology carries the seed of hope, promising improvement and more streamlined operations. Yet, the ongoing technological transformation also brings significant complications. We emphasise the urgent work ahead for countries committed to integrating this new technology into the public sector in a responsible way – a task that, so far, barely has begun.

Algorithms

Almost every human action can be seen as an algorithm. An algorithm comprises a set of instructions that are arranged in a time series to solve a particular problem or perform a particular task. In a way, algorithms are more a matter of *structure* rather than *technology*. Analogue algorithms, such as a laminated flowchart put on a wall to help visualise a decision-making process, may also be considered. It can be argued that all bureaucracy is an algorithm; a codified decision rule that leads to certain actions or assessments. The very nature of bureaucracy includes a desire for predictability, where similar cases should result in the same decision. Predictability can be recreated in *rule-based* algorithms but is scarcely possible in *generative* AI, which are built on large language models.

The exercise of authority is based on legitimacy and trust, and we delegate our civic right to make decisions to elected representatives and institutions for them to make those decisions on our behalf. But we also rightly expect that it should be possible to see and understand what has been delegated and how decisions have been made in our name. A layer of complexity is added when public decisions are placed in digitalised processes with a programmed sequence. When algorithms are transformed into code, they can be scaled up and become complex to the point that they go beyond simple transparency or explainability. We argue that this introduces something *qualitatively new*, but through a *quantitative method*: many small steps eventually become one big step. With AI technology, which can be described as systems of algorithms that have been machine trained by a training algorithm, the challenge of explainability increases by orders of magnitude.

Anyone looking for an educational example of the power of algorithms in public administration and their consequences need look no further than Charlotta Kronblad's text on school placements in one of Sweden's larger municipalities. A cross between *The Trial* and *Catch-22* may sound entertaining, but the consequences are too serious to be amusing.

Algocracy challenges public transparency

In her text, Ivana Bartoletti approaches the concept of *algocracy* as a description of the experience of being trapped in an opaque system. Bureaucracy is subject to the law, but algocracy hides behind a veil of incomprehensibility. That incomprehensibility, in turn, obscures another aspect too rarely discussed – the importance of who writes the rules. Algocracy too often means bureaucracy handed over to private, profit-driven stakeholders in the form of major corporations, and it would be naive to assume that rules and laws can be neutrally translated into code. The public sector, unlike the private sector, operates under rules that are shaped by what we demand and expect of a democracy. It must be possible to explain its decisions and justify its processes; hence, the rules of society need to be predictable, specific and comprehensible. They must also be open to participation and change through the tools offered by democracy.

Using algorithms to support decision-making, with a human "in the loop" making the final decision, may look like an attractive option, but this too poses challenges. An official is obliged to be able to explain decisions to citizens, regardless of the amount of support provided when decisions were made – and, by definition, be able to understand them. The term "machine bias" describes our human tendency to trust whatever machines produce. And even if an algorithm merely offers a *proposed* decision, it is often difficult – or sometimes impossible – for an individual employee to actively reject the proposed decision generated automatically. In practice, this makes the difference between decision-making and decision-making support non-substantive.

Finally, we uphold what the Scandinavian countries refer to as the principle of public access to official records. While similar provisions exist in many countries through different freedom of information acts, their scope, strength and enforcement differ. The importance of being

able to scrutinise and understand the exercising of public power has already been touched upon, but we wish to further emphasise the necessity of this in relation to the sovereign power to whom we delegate our civic authority. This is the only way to build trust in the institutions of society.

In Sweden, the principle of public access is formulated as a requirement for the activities of government and other state and municipal authorities to be conducted as openly as possible, under public and media scrutiny. As a general rule, everyone has the right to view documents held by public authorities. A document is public if it has been received or compiled by a public authority and is kept there, and it must be disclosed and disseminated without delay when a request is submitted. We can see that technological change within the administration threatens this principle, with this book providing examples of this – even when no laws have changed. We have seen first-hand that algocracy, where regulations are translated into code, results in the abandonment of this principle in reality. A decision cannot be explained, because it was made by an algorithm.

Our aim is to exemplify the power of algorithms, without trivialising the issue. Technical questions about the development of AI are frequently asked, but too few are drawing attention to democratic questions. It should also be noted that most of the AI scandals that have occurred have not really been about technology. How could the mistakes discussed in this book happen? Programmers or technicians are unable to supply the answers. Instead, the answers lie within the democratic institutions that managed the technology: institutions that were clearly unprepared for the algorithms, or that failed to assume their democratic responsibilities.

Technology is politics

Technology is always part of a social context and is never free of values. Both conscious and unconscious expectations can be built into technology, and the field of science and technology studies underlines that distinguishing between technology and politics is difficult, or impossible. Technical decisions can easily appear to be apolitical, when made by engineers and other groups assigned expert status. However, these decisions can dramatically shape the scope for future political decisions,

which makes it impossible to draw a clear line between politics and technology.[1]

The apparent neutrality of technology can thus become a fallacy. Technology is always built by someone, in a certain way, with a particular purpose in mind; and this purpose will also be reflected in the design of the technology. If the political dimensions of technology are not taken into account, decisions that ought to be made in democratic assemblies risk being rendered invisible. To counteract this, we need to be aware that technological development is not self-sustaining, but it can and should be subject to the principles of democracy. We therefore return time and time again to the question of who writes the code.

Values are seldom more clearly embedded in technology than in the case of AI and algorithms. An unconscious bias, or simply the choice of data used to the train the AI model, can lead to distorted and skewed results that risk being implemented far beyond the original purpose and scope of the algorithm. As a result, algorithm and data output reproduces the pattern that was originally used to train it, including the same flaws. We can note and deal with the biases that we are aware of, but what can be done about unknown, unconscious biases and conclusions that are based on incomplete data? This applies in particular to large language models, which entered the scene in the late autumn of 2022. In these models, the relationship between training data and output is even more opaque.

The devil is in the data

Due to embedded values, algorithms cannot be studied in isolation, even in media contexts, but must be viewed as part of a technological and social entity.[2] This is why we may see unexpected and unimagined consequences for institutions that we take for granted. In her contribution, Ulrika Björkstén discusses how algorithms can govern news selection in science journalism, and how professions are forced to relate to the power of (click) algorithms.

1 For example: G. Hecht (1998) *The Radiance of France: Nuclear Power and National Identity after World War II* (Cambridge, MA: MIT Press); and L. Winner (1986) *Do Artifacts Have Politics? in The Whale and the Reactor. A Search for Limits in an Age of High Technology* (Chicago: University of Chicago Press), p. 32.

2 For example: J. van Dijck (2013) *The Culture of Connectivity: A Critical History of Social Media* (Oxford: Oxford University Press).

In the digital world that surrounds us, every day we are exposed to suggestions, recommendations and decisions that have been made by AI and algorithms. This can include anything from what movies streaming services recommend (romantic comedies for women; action for men) to what profiles you see on a dating site (based on your past activity, as well as aspects such as where you live or your income) and what websites you can use for shopping ("our credit-scoring algorithm has flagged your profile"). Experiences like this can range from convenient and amusing to worrying and unsettling. So-called filter bubbles and information distortions are well-known risks when similar social media algorithms shape the public discourse. AI-powered recommendation systems can propagate and amplify polarising views, which, in turn, can undermine democratic debate and decision-making.[3]

Moreover, it should be noted that data from apps we have come to rely on, and often install on phones belonging to our employers, have been the subject of major leaks or completely repurposed. Since this book was published in Swedish, events presented by Marcin de Kaminski in his text – leaks of GPS data from fitness apps used by the Secret Service in the USA – have been reiterated in Sweden's national security service. This points to society's continued naivety regarding the consequences of digitalisation.

Technology is power

The influence that several major tech corporations have had on the development of digital tools is generally acknowledged today, not least given the geopolitical developments of the past year. This influence extends to what is seen – and is allowed to be seen – in our new public spaces in social media. Their governing algorithms are black boxes, designed to maximise the amount of time we spend in front of the screen. But the same influence can also be seen in public administration. Exploiting the public sector's wealth of data is evidently a commercial interest.

The need to raise public awareness about the power of private companies has been discussed by a number of writers in recent years. Marietje

3 This process is already well documented, but we would like to draw attention to a report by Jutta Heider and Olof Sundin, „Algoritmer i samhället", which was written back in 2015.

Schaake, for instance, points out how tech companies wield not just immense economic power: by stepping into roles normally carried out by democratically elected bodies, they also gain political power in shaping public functions from cybersecurity to systems used for policing, elections and military defence policy.[4] Governments frequently shirk political responsibility, as they outsource more and more critical state functions to tech companies.

Another aspect concerns how the data that drive automated decision-making are generated. The amount of data generated by each citizen is not evenly distributed. Those who rely more on high-tech gadgets or drive instead of using public transportation tend to leave a larger "data footprint". This is just one example of how a data-driven society risks perpetuating trends and social divisions. There is also a geographical aspect in this regard, where we can see that the data used in AI contexts is heavily skewed towards a handful of wealthy countries.[5]

The possibility to measure virtually everything is one consequence of the ever-increasing use of algorithms throughout society. And – what gets measured gets managed, a fact that is particularly evident in large parts of the labour market. Tasks that could not be measured before can now generate thousands of data points, which means that companies are lining up to evaluate every second of lost time from hours worked. Data, such as location, body temperature, keystrokes, emails sent and meetings scheduled, can be collated in an algorithm designed to direct and allocate work. But is it possible to increase productivity by micromanaging and controlling every last element of the working day? Evidence, not least macroeconomic productivity figures, points in a different direction.[6] At the same time, mass surveillance of labour can result in degraded, more stressful work, as well as a loss of freedom in the very place where most people spend the majority of their waking time – the workplace.

4 Schaake, M. (2024) *The Tech Coup: How to Save Democracy from Silicon Valley* (Princeton, NJ: Princeton University Press).

5 Benchmarking AI performance in more than half of the datasets from 26,000 research papers came from just 12 elite institutions and technology companies in the USA, Germany and Hong Kong (China). Frequency of data use by country. Research by: B. Koch, E. Denton, A. Hanna and J. Foster (2021) "*Reduced, Reused and Recycled: The Life of a Dataset in Machine Learning Research*". Also, in "2022 Internet Health Report 2022". Mozilla Foundation,

6 See Acemoglu, D. (2024) „Don't believe the AI hype". Project Syndicate, 21 May; or Acemoglu, D. (2024) "The simple macroeconomics of AI". Working Paper 32487. National Bureau of Economic Research, May. DOI: 10.3386/w32487

Algorithmic work management issues should be at the top of the agenda for both international organisations and trade unions. The impact on our working lives and the declining bargaining power of workers cannot be overstated.

Democratic participation should not require expertise

The last few years have brought about many calls for increased AI literacy and technical proficiency for both workers and citizens. These calls deserve support and endorsement to increase knowledge and engagement with technological aspects of societal change. Public education has been one of the labour movement's key projects since its formation. However, demands for literacy must never be a substitute for explainability. Democratic participation should never be dependent on technical expertise.

For several years, the concept of "explainable AI" has been at the heart of discussions in ethics and implementation research. Technological development calls for the principle of explainability, in the sense of comprehensibility and accountability.[7] Many of the chapters make it clear that transparency is a prerequisite for participation, but it is far from enough on its own, and this is also why open-source code by itself is not enough to provide civic influence. It is extremely difficult to scrutinise the values, rules and consequences that are embedded in algorithms. For true democracy, we need popular education, active citizens and media that are both free and adequately resourced.

What happens if decisions cannot be made transparent? The ability to understand and gain insight into the process itself and the pathways to decisions is perhaps as crucial as fair outcomes and the possibility for appeal. If the algorithmic decision-making process is difficult to understand – or, with AI, even incomprehensible by design – then, by definition, it cannot be a democratic decision. As Karim Jebari writes in his chapter: even if the toss of a coin produces the same outcome as a fair trial, this does not mean that we can accept making decisions by flipping a coin. Given the importance of the process, Jebari asks whether AI algorithms

7 Floridi, L., J. Cowls, M. Beltrametti et al. (2018) „AI4People – an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations". *Minds & Machines*, 28: 689-707.

should be allowed to make decisions at all, and he finds it difficult to answer affirmatively. AI will produce biased results if its learning is based on historical outcomes. Just because a particular model of car is overrepresented in speeding statistics, it does not mean all owners of that model should be fined.

Decisionmakers have a duty to understand and be able to explain the power entrusted to them by society, and they can never delegate this responsibility to third-party developers. At the same time, democracy is the opposite of expert rule. The claim that only experts in technology should be allowed to govern that same technology is a common power strategy that undermines discussion and participation. This expert objection is frequently heard in relation to technology; more rarely is it argued, for example, that only people with experience of farming should be allowed to influence agricultural policy. Decisionmakers' access to technical expertise needs to be institutionalised and adequately funded so that democratic institutions do not constantly lag behind other tech stakeholders in terms of knowledge.

Another common objection is that the regulation of technology impedes innovation. This is a rhetoric that we caution against. We also reject the false dichotomy that pits regulation against innovation. Promoting sustainable and inclusive innovation, while preventing inappropriate use of AI by public stakeholders, requires legislation, regulatory frameworks and transparency. Only through these measures can development be steered towards benefiting the majority, rather than concentrating control in the hands of a few. Throughout history, regulation in its various forms has been an indispensable element of the democratic exercise of public power in the face of new technologies – and it has enabled many of the innovations we rely on today.[8]

Public administration belongs to citizens

Power in the public sector rests with the people, and we alone decide how society should use technology. Technology is created by people, and no natural laws govern how it is shaped or how it should change our lives. This premise is particularly relevant to public administration, and

8 Mazzucato, M. (2013) *The Entrepreneurial State: Debunking Public vs. Private Sector Myths* (London: Anthem Press).

society has faced similar issues in every technological leap. The ownership of communications infrastructure, such as telephone lines, or how to balance commercial interest against private ones with the emergence of mass media, are but a couple of historical examples.

As stated, the best way to ensure that algorithms used in public administration address the right issues is for the administration itself to design and code them from scratch. Ideally, this would be done using open-source code, which makes the technology more transparent and open to co-creation, and less dependent on individual companies' proprietary solutions. In this book, Marcus Matteby demonstrates how a medium-sized Swedish municipality can work with innovative solutions in the public sector instead of buying off-the-shelf packages from tech giants. By itself, open-source code does not guarantee sufficient transparency, but it does provide a solid foundation for scrutiny. Ensuring that the programmes in which we invest taxpayer funds can be shared between public stakeholders is also reasonable. Ultimately, this approach reduces the risk of reliance on obsolete software that discourages standardisation and integration with other systems.

Politicians and citizens should act upon the fact that AI is part of society, not something that opposes it. Therefore, it is necessary to actively reflect on where we do *not* want to use AI (the so-called question zero). Such reflection is largely absent from the debate. Regulation and guidelines are needed to ensure that the use of AI does not threaten democratic rights. This can never be satisfactorily achieved through self-regulation by tech companies. Regulation should include transparency, accountability and protection of users' privacy. In his chapter, Stefan Larsson writes about how this can be accomplished, even though in practice it is like trying to hit a moving target.

A viable algorithmic future

One way forward is the notion of what we would refer to as *algorithmic auditing*. Comprehensibility must never be dependent on whatever technology the public sector happens to be using at a given time. And even for experts, reviewing millions of lines of code to scrutinise an algorithm is impossible. This is why an impartial function is needed that, like the audit role in limited liability companies, ensures that the algorithm does what it says it does – and nothing else. Establishing such a framework

is no small undertaking. In Sweden alone, the traditional auditing industry employs tens of thousands of highly trained professionals who safeguard the accuracy of corporate accounts. Yet, for anyone who believes in the potential of technology, it is clear that the cost of such an imposition would be only a fraction of the benefits to be gained as the role of algorithms expands.

Many of the chapters in this book highlight the importance of society itself owning the entire source-code process, particularly for its most critical functions. However, the vast majority of countries would find it difficult to free themselves of external suppliers, consultants and foreign technology companies in one fell swoop. This is why Christina Colclough and Hanna Johnston emphasise how the public sector can, and should, act to uphold democratic principles, even when software is procured from a third party.

We cannot predict all the consequences of our impending algorithmic future, but the progressive movement needs to tap into its potential. We create algorithms together – they are shaped by human hands – therefore, we as a society hold the key to our future. Society has every opportunity to open the black box of algocracy and allow democracy to enter. Let us take that chance. Although the technology is new, the tools are the same as they have always been – tried and tested since the emergence of Greek city states: law and politics.

Charlotta Kronblad

Algorithmic injustice – a very miscoded school placement algorithm

The first time Gothenburg used an algorithm to allocate school places to the city's children was in the spring of 2020, but something went wrong. All of a sudden, hundreds of children had been placed in schools on the opposite side of the river that divides Sweden's second-biggest town, with hour-long commutes as a result. Some children unexpectedly ended up in schools over ten kilometres from their homes, while others who could see "their" schools from the kitchen window were not allocated places there, against all expectations. Bizarre placements were uncovered all over the city, and no one in the school authority seemed to know what had happened or wanted to accept responsibility for the decisions.

A year later, it was clear that the municipal authority, the City of Gothenburg, had made a number of errors in its implementation of an algorithmic decision-making system, and that the coding of that system was not compliant with applicable law.[1] Nevertheless, hundreds of children were still forced to spend their lower secondary years in schools that the algorithm had assigned them to by mistake.

Evidently something had gone wrong, but neither the city – nor any other branch of public administration – managed to put things right. Instead, an algorithmic injustice was created and reinforced. A kind of injustice that risks being exacerbated as more and more authorities and public agencies use algorithmic decision-making systems to distribute welfare.

So, what actually happened in Gothenburg? How could the city be so blind to the impact of incorrect algorithmic decision-making? And why did no one take responsibility for the children who were affected? This chapter addresses these issues and suggests what might be done to create a fair digital future in compliance with the rule of law.[2]

1 Report by the city auditors of Gothenburg.

2 This chapter is based on my own experiences, my own data collection and ongoing research projects in collaboration with Lisen Selander, Sirkka Jarvenpaa, Anna Essén and Magnus Mähring. I am grateful for their assistance in mapping and analysing the case

The digital future requires a digitalised public sector

That the future is digital is a fact. We are undergoing a digital transformation in which both industry and society are being reshaped, and where algorithms are playing an increasingly influential role in our lives. This is also true of the public sector, where millions of decisions are made by algorithms every year. This is fundamentally a good thing, and necessary if we are to meet both current and future challenges. We need to work in the smartest way possible and make careful use of society's resources. That's why digitalising administrative tasks such as the allocation of school places is a positive step for society. Using digital tools provides opportunities to optimise placements and find solutions that humans might not automatically see or think of. Algorithmic decision-making thereby constitutes an opportunity for public administration. In 2020, algorithmic decision-making had been legal for government agencies in Sweden for some time. However, for municipal agencies, this was not yet allowed in 2020. Instead, algorithmic decision-making systems were only allowed as support for human-made decisions. An amendment to the legislation entered into force in July 2022, allowing for automated decision-making in municipal authorities (e.g., school authorities and/or other local agencies). This means that municipal authorities today are allowed to use automated systems to make decisions. Accordingly, decision-making systems can now be used both as decision support tools and to make final decisions regarding school placements.

How digital decision-making systems for school placements work

The results generated by a school placement system are dependent on the instructions given to the system and on the data entered into it. The system allocates placements to students on the basis of information about available school spots and schools' addresses, together with information about the children's preferred schools (possibly ranked in order) and where the children live. How the decision-making process then takes place is dependent on the regulatory framework of the municipality

described in this chapter, as part of our education and outreach efforts. However, I bear sole responsibility for this text.

in question. The Swedish Education Act states that children should be placed in the schools selected by their parents, unless this would prevent another child obtaining a placement at a nearby school. Legal practice states that if geographical parameters are used by a city to define placements, the walking or cycling distance should determine the order of priority, rather than the straight-line distance. Hence, a lawful school placement algorithm should include some form of optimisation that takes into account parents' preferences. Furthermore, if school places are in short supply, children who live within a short walking distance of a school should be given priority over children who live further away.

Decision-making systems for the allocation of school places are characterised by the fact that limited resources (school places) are to be allocated, and that all decisions are made simultaneously. That's why each and every decision is contingent on the outcome of all other decisions, so the placements influence one another. A vast number of combinations of correct and lawful decisions can be created, depending on the data entered into the system and what the system is instructed to do. However, if the algorithm – or any part of the data – contains errors, the number of resulting combinations of incorrect placements is almost infinite.

The Gothenburg case demonstrates precisely this kind of systematic impact, where errors were multiplied several times. Without access to the algorithm itself, it becomes almost impossible to tell whether a decision is right or wrong, and proving it is even more difficult. That's because the decision may appear justified on the face of it, but it could still differ from the decision that would have been made if the regulatory framework had been applied lawfully and the code had been correct. This is one reason why the majority of errors in the Gothenburg case were never corrected: it is simply impossible to correct errors that cannot be seen or understood.

What actually happened in Gothenburg in 2020

Parents in Gothenburg were informed of the school placements for their children on Walpurgis Night[3] in 2020. The decisions were posted on the city's website at four in the afternoon, and on this particular day, thousands of parents were surprised, and many were downright annoyed, when

3 A traditional spring celebration that falls on 30 April each year.

they found out what decisions had been made. I was one of those parents, and I remember the shock of seeing a different school to the one we had selected; the one we were sure our son would end up attending. We had assumed that, unless we actively chose another school, he would end up in a large lower secondary school where all the children from my son's primary school had been placed for over a decade. But without warning, and with no justification whatsoever, the City of Gothenburg and its school authority had implemented a new digital system that produced completely different placements compared with previous years, when student placements were allocated manually by administrators using spreadsheets.

That same evening, we started receiving calls from parents of our son's classmates, who were all experiencing the same thing. The children in his class had been placed in eight different lower secondary schools dotted all over the city. Several of them had ended up placed in schools on the other side of the river, and some were placed in schools more than ten kilometres from home. Four children living in the same block, just metres apart, were placed in four different schools. "This can't be right", we said to one another. "They must have made a mistake." The placements were simply too strange; there was no way they could be right.

Over the long weekend, we all got together and tried to work out what had happened. We thought that the authority would immediately resolve the matter if we just showed them the errors they had made. And up until that point, we believed that only our school was impacted by these incorrect decisions. Soon, however, we began receiving information that other areas were also affected. For instance, about a hundred children from Majorna, an area in the city centre, had been placed on Lindholmen, at the other side of the large river. These places look close together on the map, if you can fly or swim; in practice though, it would now be impossible for these children to get to school on foot. Instead, they would have to take a tram, bus and boat, changing several times on the way. The media picked up on some of the more extreme cases, and in the spring, Gothenburg's school placement disaster made newspaper headlines and was covered on TV and radio. It was clear that something had gone wrong, but the magnitude of the errors seemed to surprise both the compulsory education authority and the politicians. In the early summer, however, the authority admitted that it had ignored the City of Gothenburg's own rule of placing children no more than eight kilometres from their schools and subsequently offered to rectify the situation for the 450 children assigned placements further away.

A year later, the City of Gothenburg's auditors confirmed that the compulsory education authority had made a number of errors in its implementation of the decision-making system. Their report[4] directed strong criticism at the school authority, pointing out that it had ignored warnings from headteachers and software developers that the system had been coded to calculate straight-line distances instead of walking distances, and that the system had not taken sufficient account of parental preferences. It emerged that incorrect addresses had been used for some schools, and the report highlighted serious shortcomings in communication between the authority and the education committee (consisting of politicians responsible for the schools). It also turned out that the authority had actually been surprised by the extent of the errors. This was due to the fact that the decisions had not been scrutinised before they were issued: instead, the authority had merely conducted a few spot checks. Furthermore, the decisions that went against parental preferences had not been escalated to a supervisor in accordance with the authority's regulatory framework. This meant that the person who had formally made the decisions had not even seen them.

The algorithm ends up in court

In May 2020, I sued the City of Gothenburg for the unlawful implementation of the automated decision-making system for school placements. I asked the Gothenburg Administrative Court to examine the legality of the decision-making process and the underlying code. It was important to apply for a legality review of the code, rather than appealing individual decisions, as individual appeals and corrections would not address systemic errors. I was also interested in the issue from a research perspective. What happens when algorithms end up in court, become the subject of various kinds of legal proceedings and are used as evidence? Can the court restore justice in instances where algorithmic decision-making in the public sector has gone wrong?

Thus, this court case had nothing to do with me trying to correct my own child's placement. Instead, I was trying to get the court to rule that the code itself was unlawful, which would mean that the entire decision-

4 „Grundskolenämnden – granskning av verksamhetsåret 2020". City of Gothenburg, 17 March 2021.

making process would be invalid, and all placements would have to be reallocated. The aim, therefore, was to explore whether it would be possible to correct the placements of all affected children in Gothenburg. This is a reasonable ambition in terms of justice, rather than simply correcting the placements of children whose parents happen to have the know-how, expertise and – not least – the will to appeal.

My argument in court was based on the fact that the decisions had been fully automated, and that automated decision-making at a municipal level was not permitted at that time. I also pointed out that parents' preferences had not been taken into account, that the authority had allocated placements to children on the basis of straight-line distance to school and that the errors had multiplied, affecting placements all over the city because of a domino effect. To support my case, I submitted statistical reports of the actual outcome, arguing how the algorithm must have been structured given the actual placements that were allocated. I also stated that I had never got access to the algorithm itself, despite requesting it repeatedly; nor did I receive a decision rejecting my request to access the code.

No review

The city defended itself by claiming that the decision-making was not automated and insisted that the system had merely served as a decision support tool, and that they had done their best to take parents' preferences into account. They submitted no evidence to support these claims. To my surprise, however, it turned out that they did not need to. The Administrative Court's judgment[5] states that:

> in circumstances where the parties do not agree on the relevant facts of the case, it is incumbent upon [Kronblad as] the complainant to disprove the claims of the authority and prove that the placement decisions were made by computer software and not by municipal delegates.

In other words, it was up to me to prove that the decisions had been made unlawfully, and analyses of the actual outcome did not amount to sufficient evidence. This meant that the algorithm itself would have been needed as evidence, but the court never looked at it. Since I was not given

5 Ruling of the Administrative Court in case 6541-20, p. 6.

access to the algorithm, I was also unable to submit it as part of my evidence. And because I failed to provide the algorithm, the court simply chose to dismiss the case on the ground that the burden of proof rested with me, despite the fact that the court could certainly have exercised its *duty to investigate* by requesting the material it deemed necessary to assess the facts. In addition, it is not uncommon in Swedish court proceedings to shift the burden of proof in cases where only one party has the opportunity to gain access to certain evidence, in which case it is no longer up to the complainant to prove certain facts; instead, it falls to the opposing party to provide evidence to the contrary. However, the court did not do this either. Hypothetically, it would have been much easier in this case for the school authority to prove that the decision-making system was lawful, compared with my ability to prove that an algorithm – which I was never allowed to see – was unlawful. Legal precedent shows that in other cases the municipality has been required to account for the information forming the basis of the decision under appeal.[6]

I appealed the ruling, arguing that the administrative court had failed in its official duty to investigate and that the burden of proof should have been shifted. Unfortunately, the administrative court of appeal did not grant an application to retry the case, which led to yet another appellate procedure. This time, I challenged the administrative court of appeal's refusal to grant a retrial by taking the case to the Supreme Administrative Court. As grounds for a new hearing, I argued that it is

> impossible for citizens to defend their rights when they are subjected to automated and systematised decisions made by public agencies. Without access to the decision-making algorithm and the information entered into the system, it is impossible for individuals to prove their case, and furthermore, it becomes impossible for the courts to make a legal assessment.7

I also pointed out that the case carried strong precedential value, given the fact that new legislation was being drafted at the time that would legalise automated decision-making at municipal level, and that guidance in this area was therefore greatly needed.

While the case was pending before the Supreme Administrative Court, the City of Gothenburg's auditors published their report, which supported my position. There was now an independent public review which indi-

6 From the Swedish Supreme Administrative Court: HFD 2024 ref. 62

7 Appeal against the ruling of the administrative court of appeal in case 6541-20, 20201208, p. 3.

cated that the decisions had been automated, that straight-line distance between home and school had been used incorrectly (instead of walking distance), and that the authority had not sufficiently taken into account parents' preferences regarding school placements. I promptly submitted the auditors' report as new evidence in the hope of securing the review of the case. Unfortunately, the Supreme Administrative Court disagreed, issuing a standard response stating that leave to appeal had not been granted. Such decisions do not need to be justified by the court, which means I am still at a loss as to why they chose not to examine the facts of the case. The school placement algorithm escaped scrutiny, therefore, and its legality was never examined by any court.

Algorithmic injustice comprises both social and legal injustice[8]

The court ruling may actually have been the best possible outcome for my research, as I can now argue that the blindness of the courts extends all the way to the top. For the children of Gothenburg, however, this meant that no one was held accountable and that most of the errors remained uncorrected. Several hundred children had to complete their entire lower secondary education in schools in which they would never have been placed had the algorithm been coded lawfully.

What happened in Gothenburg was an algorithmic injustice with no possibility of redress. This algorithmic injustice comprises two distinct components. Firstly, there is a social injustice, where the incorrect placements resulted in unfair distribution of public resources in terms of school places. Secondly, there is a legal injustice in that the legal system was incapable of correcting the social injustice that had occurred. Of the two, I believe that the legal injustice is the more significant. It is inevitable that errors can and will occur when public administrations operate in cutting-edge technological environments. But the consequences are much more serious when the courts are incapable of correcting the errors that are made. There is no longer a means of redress at that point, and as the rule of law collapses, so does our trust in it. The formal opportunity that indi-

8 See C. Kronblad, A. Essén and M. Mähring (2024) „When justice is blind to algorithms: Multilayered blackboxing of algorithmic decision making in the public sector". *MIS Quarterly*, 4(48): 1637-1662.

viduals have to appeal rulings means little if courts continuously fail to review, and do not understand, the algorithms involved in legal disputes and used as evidence.

From public administration to public obscurity

So, what can we do about it? In recent years, I have been focusing on these issues (in light of the Gothenburg case) together with my colleagues, Lisen Selander and Sirkka Jarvenpaa, at the University of Gothenburg and the University of Texas, and Anna Essén and Magnus Mähring at the House of Innovation, Stockholm School of Economics. In two different research projects, we have highlighted how difficult it is for the people affected by algorithmic decision-making even to realise that algorithmic decisions are being made. We have also developed theories that explain why the institutions of society (such as the compulsory education authority and the courts) failed to recognise and address the consequences of algorithmic decision-making. The remainder of this chapter presents insights from these different projects, thereby helping to shed light on how blindness to algorithmic systems emerges and what this means for justice in the digital age.

Firstly, it is incredibly difficult for anyone affected by an algorithmic decision to detect whether anything has gone wrong, or whether something unlawful has happened – or even to know whether the decision was automated.[9] In the Gothenburg case, the authority did not tell people that they had used an algorithmic system, which meant that the recipients of the decisions were unaware of that fact. This means that it is almost impossible for individuals to realise that errors have been coded into the system that generated them. That's because individuals are only aware of a small fraction of the data entered into the system – their own school placement preferences and their home address – and will only be privy to a tiny aspect of the outcome (the school placement decision relating to their own child). To even detect that an automated decision-making system has had an impact, and that this has contributed to systematic errors, we must band together and collectively compare and analyse the

9 See L. Selander, S. L., Jarvenpaa and C. Kronblad (2023) „Awakening to algorithmic transgressions: Non-users discovery of algorithmic decision making". *Academy of Management Proceedings 2023*(1), 19344. DOI: 10.5465/AMPROC.2023.17bp

decisions received. This means that considerable resources are needed to form a complete picture of the system. Not only that, but technical resources are required to be able to work out what happened and understand how the system arrived at its decisions. However, social, emotional and communication resources are also important to be able to collect data from others affected, to attract media attention and to contribute to the public debate. For individuals, therefore, it is virtually impossible to evaluate or challenge an algorithmic decision, as individuals are unable to "see the big picture" on their own. Instead, our research shows that resources need to be mobilised at the group and community levels.

Avoidance, concealment and denial[10]

If we instead turn our attention to the public stakeholders and institutions that use algorithmic systems, and that should reasonably be capable of "seeing" and understanding them, we realise that they, too, frequently struggle to do just that. We can conclude from the Gothenburg case that institutional stakeholders often choose to turn a blind eye and simply ignore certain information. This aligns with previous research and theory on *ignoring practices*,[11] which highlights the fact that stakeholders do sometimes choose to ignore information if it does not fit with their worldview, or consists of information that is too inconvenient to take into account and deal with in practical terms. When such ignoring practices are repeated, or when they occur simultaneously among many different stakeholders, this can lead to "macro-ignoring", where both the organisation to which the stakeholder belongs and the surrounding context become blind to what is happening.

In the Gothenburg case, this happened because the school authority initially dismissed certain information, such as warnings from the software company that the results would be anomalous if the authority decided to apply straight-line distances in the system. The situation was made worse because the authority chose not to review the results (the

10 See C. Kronblad, A. Essén and M. Mähring (2024) „When justice is blind to algorithms: Multilayered blackboxing of algorithmic decision making in the public sector".

11 For a review of various theories on ignoring, functional stupidity, macro-ignoring etc., see M. Alvesson and A. Spicer (2019) *Dumhetsparadoxen* (Stockholm: Fri Tanke); L. McGoey (2019) *The Unknowers* (London: Zed Books); M. Knudsen (2011) "Forms of inattentiveness". *Organization Studies*, 7(32): 963-989.

individual placement decisions) before sending them out to parents. According to the audit report, this was due to a concern that administrators might be tempted to alter individual decisions. In other words, the authority had such blind faith in the system – which it did not understand – that it chose "not to look at" the results. This avoidance behaviour, which meant that they failed to address the issue initially, made the problem worse, as the errors were allowed to multiply. Ultimately, the authority was surprised by the results. This is how the social injustice came about.

The school authority subsequently concealed the algorithmic decision-making from the public by simply withholding certain information. Despite repeated requests, the authority failed to disclose the algorithm to the people who asked for it and provided inadequate information about what it had done to the responsible politicians. The authority then (wrongly) claimed that the system was used merely as a support tool. This is how information about the system was withheld and obscured. It is difficult to scrutinise the system or appeal the decisions without accurate information about what happened and how the algorithm works.

Ignoring at multiple levels[12]

This is the point at which we reach the administrative court, which is meant to uphold the rule of law and ensure that incorrect public decisions are rectified. Here, we encounter another kind of ignoring, which is likely to be rooted in society's limited understanding of what algorithmic decision-making involves. Rather than updating the law with respect to digital administration and algorithmic decision-making, the court continues to apply the same procedural rules and approaches as before. Rules intended for a more analogue form of administration. One example of this is when the court places the burden of proof on the party with no access to the system. This ignores the agency of the system and its outcomes, as well as the new power imbalances brought about by the growing use of algorithmic decision-making in the public sector. This blindness is demonstrated by the fact that the courts missed the opportunity to investigate the issue itself and chose not to request the algorithm or any additional information about the system.

12 See C. Kronblad, A. Essén and M. Mähring (2024) „When justice is blind to algorithms: Multilayered blackboxing of algorithmic decision making in the public sector".

At a societal level, we hold there is a lack of insight into how digitalisation has transformed the conditions under which public power is exercised. We need to update our old solutions to achieve societal goals (that remain unchanged), such as transparency, justice and rule of law. The Gothenburg case demonstrates that a new public decision-making method has been introduced without ensuring that the legal instruments that we deploy are still fit for the purpose. The possibility of fair review and redress is eliminated by denying the need to update the law to reflect digitalisation. This allows legal injustice to perpetuate the social injustice brought about by flawed decision-making.

In Gothenburg, thousands of children were placed in the wrong schools, and neither the city nor the court accepted responsibility for correcting the situation. What happened, therefore, was the result of an interaction between the behaviour of the school authority, which rendered both itself and others blind to the errors, and an institutional context that had not been updated to take into account the risk of new types of algorithmic errors emerging. This combination of avoidance, withholding and denial creates a blindness to algorithmic systems.

Describing how various forms of ignoring are interlinked makes it possible to understand how social injustices can emerge and persist over many years, as exemplified in the Australian and Dutch cases presented by Johan Hirschfeldt in this book. There is a risk of new types of social injustices emerging and being repeated, as neither the public authority using algorithmic systems, nor the people affected by their decisions, have access to all the relevant information. What can be noted from the Gothenburg case is that the technology involved is not particularly complex. The algorithm itself is not hard to understand. The school placement algorithm in question is not based on artificial intelligence and has no self-learning components, unlike in the Dutch case, for example. The main error in Gothenburg was simply due to the fact that the system had been programmed to issue placements to children on the basis of straight-line distance from a bird's-eye view, rather than walking distance. This error and its implications for a city divided by a river are easy to comprehend. Despite this fact, the majority of affected children received no redress, and the situation was not corrected. We argue that this is because algorithmic systems can be "black-boxed", not only by technology, but by social and institutional behaviours. In other words, the decision-making system becomes concealed (put in a black box) by actions made by the surrounding authorities and institutions. In this case, the situa-

tion is more akin to a white box being wrapped up in black paper, which makes a relatively simple technology impossible to access. Our research shows that it is extremely difficult for individuals to assert their rights when institutional stakeholders are permitted to obscure the systems by disregarding and withholding information, particularly when the legal infrastructure is outdated.

How we can create a fair future in compliance with the rule of law

Achieving justice, both social and legal, requires more than simply demanding that algorithmic decision-making systems be explainable and transparent. While such actions would certainly be an improvement, they place almost impossible demands on individuals to interpret and understand complex systems. Instead, we need to recognise that the position of the individual has been weakened in relation to the digital apparatus of public authority. The exercise of digital authority is harder to observe and control than its analogue counterpart. That's why it's reasonable to reverse the burden of proof in algorithmic cases and ensure that the people using the systems also take responsibility for their implementation and commit to understanding them, rather than blinding themselves and others to the consequences of algorithmic decisions. In addition to this, we should implement a number of other updates to the legal infrastructure with a view to providing genuine opportunities for legal review and redress. For instance, it should be possible to examine the legality of the code, rather than the courts being able to rely solely on an individual review of specific decisions. The ability to assert one's rights should not be reliant on the know-how, expertise or will of the individual, or their parents, to appeal against decisions made by the public sector.[13]

The changes that we propose do not necessarily relate to the need for more complex regulation of the technology in and of itself, as can be seen from the artificial intelligence regulation discussed by Stefan Larsson in his contribution to this book. Rather, it is more a matter of ensuring

13 For a framework aimed at improving legal certainty with respect to algorithms, see C. Kronblad, A. Essén and M. Mähring (2024) „When justice is blind to algorithms: Multi-layered blackboxing of algorithmic decision making in the public sector".

that the rule of law is still in function, that is, that the basic procedural rules and legal remedies designed to protect individuals are still effective and, if they are not, making sure to update them.

Johan Hirschfeldt

Algorithms and social benefit systems should follow constitutional law

Individuals have ended up in trouble, whistleblowers have been silenced and trust has been undermined when poorly designed systems lacking transparency and proportionality have been allowed to make algorithm-based decisions without other brakes being applied. This chapter outlines how and what lessons can be learnt from previous missteps.

The EU's AI Act governs large systems where a slew of personal data is used by public authorities to provide, for example, social security benefits. Such systems are recognised as belonging to a particularly high-risk area in need of enhanced protection. One fundamental issue relates to how rule-of-law oversight should be structured.[1] The rule-of-law requirements set out in the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR) and the Charter of Fundamental Rights of the European Union are, of course, paramount in this regard. However, these are not addressed in detail here, nor is the General Data Protection Regulation (GDPR) discussed. Instead, the focus is primarily on national regulation at the constitutional level in Sweden. My contribution aims to serve as a reminder of something that is quite obvious, if you think about it: alongside EU law, the Swedish Constitution is what ultimately sets out the rule-of-law requirements in the field of artificial intelligence (AI).

The two cases

The discussion begins by recalling two cases involving recovery procedures in social benefit systems. The cases, one Dutch and one Australian, are no longer current, but they have not received the international

1 See, for example: J. Chamberlain and J. Reichel (2023) „Supervision of artificial intelligence in the EU and the protection of privacy". *FIU Law Review*, 2(18): 267. DOI: 10.25148/lawrev.17.2.5

and public attention that they deserve. They are included here, as they so clearly illustrate the far-reaching consequences that shortcomings in AI systems and their application can have on a number of societal levels – socially, politically, legally and administratively.

The systems used algorithms – with machine learning, in the *Dutch* case – where new decisions were gradually produced with AI support based on earlier decisions that were generated by the system. That system included options for selecting suspect cases for review, supported by an AI-based, self-learning function for risk classification. To train itself, the system used the risk classification of previous applications that had been processed in the system. These applications had been categorised as correct or incorrect. The more a case resembled a previous one that was rejected, the higher the risk category to which the case was assigned. Different weights were assigned to the various risk factors. The citizenship of welfare recipients was one of several dozen indicators for scoring risk (gender, religion, living conditions, address, etc.). Foreign (i.e., non-Dutch) citizenship was a particular risk factor in the review.

The system selected cases for recovery on the basis of the indicators and their weighting for different situations. The intention was to examine the high-risk cases manually, but no information was provided on why the system selected the case. However, recovery took place without further investigation. It was, of course, important to effectively combat and limit the crime that had actually been detected, and that was costing society huge amounts of money. That was why the tax authorities decided, in a later phase of the application, on a very broad strategy of general recovery in the cases. These recoveries were then carried out routinely on the basis of pure groupthink: "80% of applications are wrong and 20% are right".[2]

A key point that has attracted attention in the case of the *Australian* Robodebt scheme is that before the decision was made on the system, the government had failed to follow the advice of its legal experts, who had said that the scheme had no basis in law. Moreover, warnings from whistleblowing employees who understood how the system operated and were familiar with the effects were not taken seriously.

In both countries, the application of the systems resulted in significant

2 „Netherlands - opinion on the legal protection of citizens". CDL-AD(2021)031. Venice Commission, 18 October 2021, p. 19.

negative societal consequences. The aim was to fight crime. However, the harmful impact of the systems on large numbers of low-income welfare recipients, clients or "customers" was enormous.

Besides the harm done to individuals, this also led to reams of additional work for supervisory authorities, courts and specially appointed commissions. The failures also cost money in the form of additional costs for the public in defending the systems, correcting the system errors and compensating the people who had suffered harm. Ultimately, all of this led to the systems gradually losing their legitimacy. In the end, the failures became politically unmanageable for government and parliament.[3]

These two recovery cases are not unique. The American MiDAS case regarding automated recovery of unemployment benefits in Michigan has also garnered attention.[4] Another case, albeit slightly different in nature, is the British Post Office Scandal, where more than 500 subpostmasters were required to make payments because of accounting errors that were wrongly calculated by a computer system developed by a world-leading computer company and commissioned in 1999.[5] Payment demands, dismissals and court cases followed. This resulted in a great deal of human suffering and enormous costs for the public, including payment of damages. A special committee of inquiry with the authority to obtain documentation and conduct hearings under oath is still working on the case. The question of financial compensation has not yet been settled.[6]

Here, these cases are intended merely to serve as a background for illustrative purposes. The circumstances of these cases are not set out in greater detail than has already been presented. Instead, reference is made to commission reports and other literature that shed light on the cases. However, an opinion regarding the Dutch case by a special legal

3 Regarding the policy implications in the Netherlands, see: J. Henley (2021) „Dutch government over child benefit scandal". *The Guardian*, 14 January; J. Henley (2021) "Dutch government resigns over child benefits scandal". *The Guardian*, 15 January; T. Erdbrink (2021) "Government in Netherlands resigns after benefit scandal". *The New York Times*, 15 January; S. Berends (2021) "European Social Policy Network, Dutch child benefit scandal: Origin and latest developments". ESPN Flash Report 2021/51.

4 Ranchordás, S. (2022) „Empathy in the digital administrative state". *Duke Law Journal*, 6(71): 1376.

5 „British Post Office Scandal". Wikipedia.

6 „Searchable transcripts of the Post Office Horizon IT Inquiry hearings". Post Office Inquiry website.

body within the Council of Europe, the European Commission for Democracy through Law (the Venice Commission), is presented below.[7]

The Venice Commission's review of the Dutch case

The Venice Commission is a review body within the Council of Europe. It comprises nationally appointed professors, judges and other experts who prepare and publish reports on constitutional and other public law and political science issues relating to legislation and the application of law in member states. Its remit also includes the separation of powers and governance. When it comes to legislative matters, these reports may bear some resemblance to the advisory opinions of the Swedish Council on Legislation. They are not legally binding for member states in the same way as judicial rulings of the Court of Justice of the European Union and the European Court of Human Rights in Strasbourg. Instead, they are recommendations to the circle of member states of the Council of Europe (a wider circle than the EU member states). These reports are used to develop pan-European standards of a kind. They are rooted in European law, including the ECHR and EU law, as well as in the common and diverse constitutional cultures of the member states. That is why it is interesting to look at the Venice Commission in this context.

The Venice Commission delivered its opinion on the Dutch case in 2021.[8] The background to this was a request for review from the Dutch Parliament. The Commission began its opinion by addressing the legitimacy of combating fraud. It then conducted a comprehensive critical review of the case from a constitutional and legal perspective. The following sections discuss some of the legal viewpoints on the application of AI that I believe are particularly important.

The Venice Commission raised the matter of the risk classification model mentioned earlier. Here, the applicants' nationality had been used

7 "Netherlands - opinion on the legal protection of citizens". Venice Commission. See also: "Report of the Childcare Allowance Parliamentary Inquiry Committee". CDL-REF(2021)073. Venice Commission, 14 September 2021; "Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal". Amnesty International, 25 October 2021. This report contains a large number of references. For the Australian case, see "Report of the Royal Commission into the Robodebt Scheme". Commonwealth of Australia, 7 July 2023.

8 „Netherlands - opinion on the legal protection of citizens". Venice Commission.

as one of the criteria for selecting individuals to be investigated, without further evidence that they had committed fraud. This was a violation of applicable law, including Article 22 of the GDPR.[9] The Commission also recalled the general prohibition of discrimination laid down in the ECHR in relation to the enjoyment of any right set forth by law. The ECHR also states that no one shall be discriminated against by any public authority.[10] This prohibition of discrimination is applicable in cases where there was no objectively acceptable reason to treat people differently. This may include situations in which public authorities have acted at their discretion or unintentionally engaged in discrimination, even indirectly. According to the Venice Commission, where the EU or a member state permits the exclusive use of an algorithm on which to base decisions, EU law requires appropriate measures to be taken to safeguard the rights, freedoms and legitimate interests of the individual concerned. No decisions made exclusively by an algorithm may be based on "sensitive" personal data (i.e., data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, as well as genetic data, biometric data, health data, or data relating to an individual's life or sexual orientation). In the Dutch case, the use of nationality as a possible criterion for identifying fraud was regarded as falling under this prohibition. Discriminatory methods had been systematised by means of algorithms.

The Commission also commented on the trend towards self-learning AI systems, stating that it could become almost impossible to detect partiality and bias in such sophisticated systems. Since those systems were being populated with huge volumes of training data, it was much harder to identify which criteria were being used.

The Commission also raised concerns about weaknesses in the legislative process and in parliamentary scrutiny of the executive (the government and its closely related public authorities in the Netherlands). It also addressed the administrative procedure with its rigid interpretation of legal provisions and the lack of transparency within and from the tax authorities in relation to the government and other public authorities. Whistleblowers' warnings from the shop floor had been ignored. It was noted how the complaints procedures for individuals were complex and formal.

9 This article contains provisions on automated individual decision-making, including profiling.

10 Article 1 of Protocol No. 12.

Finally, the role of the courts was addressed. General administrative law in the Netherlands explicitly recognises the principle of proportionality. This is also an important legal principle in both EU law and the ECHR. In the Netherlands, however, the Supreme Administrative Court had overruled the lower courts and found that the principle of proportionality did not apply to the specific legislation regarding benefits. This interpretation was then upheld in the courts between 2012 and 2019. Ultimately, however, the Supreme Court broke with its own earlier practice. This happened during the review of an individual case where the Court of Justice reversed its position on the applicability of the "all or nothing" rule concerning recovery. The court now required the general principle of proportionality to be applied, the various interests to be weighed against each other and the impact of a decision not to be disproportionate to the purpose of that decision.

The EU now issues an annual Rule of Law Report, with subsections on each member state. The 2024 report sets out the measures implemented in this area in the Netherlands.[11]

Some reflections from the perspective of Swedish constitutional law

Of course, the Venice Commission's fundamental reflections on the necessary relationship between the legal system and algorithmic decision-making systems are relevant to Sweden as well. The following section therefore presents a number of reflections from a Swedish perspective, along with a few references that are also accessible to readers outside Sweden.[12]

11 „2024 Rule of Law Report: Country chapter on the rule of law situation in the Netherlands". SWD(2024) 819 final. European Commission, 24 July 2024.

12 The legal issues in this area are discussed in the committee report *Juridik som stöd för förvaltningens digitalisering (SOU 2018:25*): see summary on p. 27. See also two collected volumes: L. Colonna and S. Greenstein (eds) (2022) *Nordic Yearbook of Law and Informatics 2020–2021: Law in the Era of Artificial Intelligence* (Stockholm: The Swedish Law and Informatics Research Institute); K. de Vries and M. Dahlberg (2022) *De Lege, Law, AI and Digitalisation* (Uppsala: Faculty of Law). Other useful references include P. Wahlgren (2018) "From Lex Scripta to Law 4.0". *Scandinavian Studies in Law*, 65: 159; R. Karlsson (2020) "Den digitala statsförvaltningen – Rättsliga förutsättningar för automatiserade beslut". *Förvaltningsrättslig tidskrift*, 1: 51. Also, the works of Markku Suksi, including "Concluding reflections on the digitalization of government functions through automated

Public authorities and their agencies, in Sweden and elsewhere, must be held to high standards of accountability for the systems, their design and their use. In that case, AI can contribute to efficiency, equal treatment and the rule of law; or, in other words, provide major benefits to society. Ultimately, system design must be compliant with the legislation. Then there is the operation of the systems and their handling of individual cases. In other words, this is a matter of the application of the law. For the application of the law, legislation imposes important requirements on how public authorities handle cases and make decisions, and on how courts assess and review them.

The use of AI must be compliant with observation of the *principle of legality* (that public power is exercised under the law) and the *principle of proportionality* (that legislation and action by a public authority must be likely to lead to the intended result, the action must never go beyond what is necessary and it may be taken only if the intended result is reasonably proportionate to the inconvenience that is likely to be caused to the individual against whom the action is directed). Furthermore, *transparency requirements* must be observed (with the right of access to information for the parties involved and transparency towards the media and the public, who must be able to stay informed about the systems and their functionality).[13]

According to the principle of legality, a fundamental requirement of legislation is that it must be predictable. Therefore, it must also be transparent. Furthermore, to be acceptable under the rule of law, legal regulation must be structured in a specific hierarchy (constitution, law, ordinance, official regulations), whereby a subordinate rule is covered by a superior rule at each level.[14] If a subordinate rule is inconsistent with a superior rule, it may not be used when a case is examined. Courts and administrative authorities are responsible for this review of legality in their application of the law in individual cases. This is ultimately done via what is known as a judicial review, which also includes examining compatibility

decision-making, with some wider AI issues" and "Formal, procedural, and material requirements of the rule of law in the context of automated decision-making", both in Suksi, M (ed.) (2023) *The Rule of Law and Automated Decision-Making: Exploring Fundamentals of Algorithmic Decision-Making* (Cham: Springer), pp. 65 and 209.

13 Chapter 1, Section 1, and Chapter 2, Sections 11 and 21, of the Instrument of Government, the main constitutional act of Sweden, and Chapter 2 of the Freedom of the Press Act, which also has constitutional status.

14 Chapter 8 of the Instrument of Government.

with the constitution and international conventions. All courts and public authorities share this responsibility, with the Supreme Court and the Supreme Administrative Court holding ultimate responsibility.[15]

A computerised system is designed to process individual cases. This will result in proposals for decisions that will be examined by a human case officer or, in some cases, automated final decisions. Every stage must comply with the law in this process, too. Therefore, when a decision-making process that ultimately has to comply with the requirements of the legislation is designed to be fairly automated in its application, the applicable legal rules must first be translated into the system's software.

In other words, the technical design of a system must align with the legal requirements set out in the legislation for the application of the law that will follow in decisions being made in each individual case. This concerns the exercise of public power, which, according to the very first paragraph of the Swedish Constitution, the Instrument of Government, must be exercised under the law (the principle of legality).[16] Accordingly, the design and application of the system must also clearly align with the relevant legal rules and, not least, be compliant with fundamental rights and freedoms and the Administrative Procedure Act.[17] A general requirement expressed in the constitution is that systems and their application must be based on objectivity, impartiality and equal treatment; this requirement is set out in the first chapter of the constitution on the foundations of the system of government.[18]

Decisions should not only be based on the rules; they should also be understood by the person receiving them. That is why there is a requirement to provide the grounds for decisions; in other words, there is a duty to state reasons.[19] Application must also be compliant with the requirements of objectivity and proportionality, as also set out in the Adminis-

15 Chapter 11, Section 14, and Chapter 12, Section 10, of the Instrument of Government. See „The Instrument of Government (1974:152)".

16 Chapter 1, Section 1, third paragraph of the Instrument of Government.

17 Chapter 2 of the Instrument of Government and the Administrative Procedure Act (2017:900). See generally: C. Muller (2020) „The impact of artificial intelligence on human rights, democracy and the rule of law". Ad Hoc Committee on Artificial Intelligence (CAHAJ), Council of Europe, CAHAJ (2020)06-fin.

18 Chapter 1, Section 9, of the Instrument of Government.

19 Section 32 of the Administrative Procedure Act.

trative Procedure Act.[20] In this context, a reasonableness assessment of whether the outcome of a decision is reasonable may need to be performed before the decision is made.

Application in individual cases also requires transparency towards both the individual party through communication or the right of access to information for the parties involved[21] and the responsible public authorities. This is a matter of fair trial or fair treatment.[22]

However, it is clear from the Australian and Dutch cases that this is not enough. Effective requirements for access to information and control also need to be set up. The principle of public access to official documents and supervision exercised by the Parliamentary Ombudsmen (JO) and the Swedish National Audit Office (*Riksrevisionen*) are important instruments already set out in the constitution.[23]

System development and the need for legal controls

Before any new decision is taken, machine learning can be used to assign scores to various factors that have been significant or have otherwise appeared in previously documented decisions. Such factors may be based on statistical sampling and profiling. An arrangement of this kind must be compliant with the GDPR and may be effective and appropriate in some cases. However, this moves beyond the realm of "conventional" application of law, which is based on legal analysis of the specific circumstances of each individual case. It may also involve risks. If earlier decisions were materially incorrect (i.e., blatantly wrong, disproportionate or unreasonable), there is a risk that assigning scores will compound such errors when they are used to make new decisions.

To avoid this when dealing with anything other than the most routine forms of large-scale applications, there must be scope for proportionality assessments in individual cases. This, in turn, requires qualified case officers and opportunities for manual procedures. Case officers must not place blind trust in data generated automatically. Furthermore, warnings from internal critics and whistleblowers must be taken seriously.

20 Section 5 of the Administrative Procedure Act.
21 Section 25 of the Administrative Procedure Act.
22 Chapter 2, Section 11, of the Instrument of Government.
23 Chapter 2 of the Freedom of the Press Act, and Chapter 13, Sections 6 and 7, of the Instrument of Government.

Of course, protection must be built in during system development. We must not lose the ability to realise what might happen, even in the early stages of the development phase.

Without transparency, courts and public authorities are unable to answer the question of whether the system is structured in accordance with legal requirements (i.e., the general rules), so that fully or partially automated application in individual cases can be lawful and subject to review. It must be possible for the public authority's case officers to understand the decisions generated by the systems applied, and to justify or explain those decisions. They also have to be comprehensible to the individuals concerned so that they can exercise their rights by requesting a review or appeal.

This also requires transparency in relation to the companies that develop or operate the systems on behalf of public authorities, although trade secrets must also be respected. Systems must be open to scrutiny by clients, higher courts, audit institutions and supervisory authorities. Similar demands may also be made by the media and the general public on the basis of the demand for transparency and the duty to provide information or, as in Sweden, through the constitutionally protected principle of public access to official documents. In the Dutch case, even parliament was denied access to crucial information and transparency for a long time. Such shortcomings ultimately risk undermining trust in institutions and democracy.

The intentions and outcomes of decision-making systems based on algorithms may be very well-meaning. In an ideal world, perhaps computerised systems can make fairer decisions than officials. But it is still necessary to ask the following question: is it technically and administratively possible to embed the conditions for this within the framework of the legal system? This is undoubtedly achievable for some systems; whereas for certain decision-making situations, it is not. If transparency cannot be guaranteed, control and the rule of law are compromised, which is unacceptable. What is known as the "black box problem" with self-learning systems, with results that cannot be understood, poses a threat that has to be limited or overcome through new methodological developments.

Summary

To finish, here are a few simple conclusions about what national law requires in the Swedish context.

Our constitution sets out certain solemn core values. These must not be regarded as “empty words”, but must be respected as starting points for all legislation and the work of the public administration. This involves legality, all people being of equal value, the freedom and dignity of the individual, equal treatment and non-discrimination, objectivity, impartiality, and proportionality.

Systems must be designed to meet these requirements and then deliver predictable results, even at the level of individual decisions. This requires a system architecture that is compliant with the law at every stage of the regulatory system, from the constitution to official regulations. The need for access to information and transparency has to be met for decisions to be understood. They must be capable of being reviewed and legally scrutinised.

Failures of the kind discussed here may have far-reaching social, legal, economic and political consequences for people and public institutions; consequences that may extend all the way up to the government and parliament. All stakeholders in the companies and public authorities involved in system design, development, operation, management and decision-making must be urged to remain vigilant and take responsibility for developments. EU regulation is of crucial importance in this regard. Its application ultimately rests with our politicians.

Marcin de Kaminski

Digital threats and risks as a formative guiding principle

Most of my life has revolved around the impact of technology on democracy and human rights. Sometimes my involvement has been professional; at other times, it has taken on a more activist character. I have often been faced with the same questions, regardless of which "hat" I was wearing at the time. One of the main questions I am asked is why experiences of technical vulnerabilities in faraway countries should also matter to us here in Sweden. Time and again, I have been forced to explain that issues relating to privacy and integrity are not only relevant to people who are concerned about "having something to hide". On the contrary: nowadays everyone – both individuals and organisations – has data that is worth protecting, as I outline in this text.

Of course, where the boundaries lie between digital security, algorithms and artificial intelligence (AI) – a much-hyped concept of late – is debatable. However, I believe that basic security issues related to data and data management are fundamental, even when it comes to more innovative developments. It is easy to be dazzled by the possibilities afforded by either generative AI or algorithmic data management. Similarly, there appears to be a widely held belief that digital development in the public sector should take place in the same way as experimental digitalisation in the commercial sector. As more public organisations seem to want to invest in things like test labs, agile working methods and digital platforms – where we are all viewed as customers rather than citizens – we cannot afford to ignore the past mistakes.

A blurred boundary

At the time of writing, there is ongoing debate in Sweden and abroad about the TikTok app and whether its use should be permitted on mobile phones used by employees in the public sector. Several US states have completely banned the use of the app on work phones, fearing that the

app could "spy" on public sector employees and send information to China. Of course, there may be cause for such concern for reasons related to defence policy, and certainly for reasons related to the protection of sensitive infrastructure and industry. If there is a risk that information about the movement of public sector employees could be leaked to a foreign power, being reluctant to share it may be perfectly reasonable. The arguments are strengthened if there is reason to suspect that the browsing habits of public sector employees, combined with TikTok's aggressive algorithms, could increase the risk of foreign influence.

At a time when the distinction between the use of personal and work-related devices remains blurred, there is every reason for concern; not least because many organisations continue to allow employees to use work phones freely for personal purposes. Even though solutions may be available that allow private and work-related calls to be separated for benefit-related reasons, people are often free to install private apps on their devices. That kind of freedom is convenient for the employee, but problematic for the employer. If the employer is a public body, this may pose problems with respect to both security and confidence in decision-making. The deliberate privacy leaks, aggressive algorithms and security vulnerabilities of apps are neither new nor limited to TikTok.

To complete the overall picture, a reminder of some early warning signs may be in order. In the spring of 2018, the US military was rocked by the mapping of secret military bases.[1] Bases that were already on record were described online in considerably more detail than before, and certain secret facilities were suddenly exposed. This came about because US military personnel had been using the Strava app to log their daily workouts. The social features of this popular fitness app allowed information to be shared about employees' runs around secure installations. Updates from other social media about overseas deployments and identities made it possible to map the secret bases and their staff based on publicly available information.

Strava responded by disabling some public features, while also giving users the option to disable map displays of certain preselected locations, as well as the option to hide start and end points. It seems obvious that such precautions may be important even for people who are not stationed at military bases.

1 Hern, A. (2018) „Fitness tracking app Strava gives away location of secret US army bases". *The Guardian*, 28 January.

Whether other fitness apps have also done this is unclear; what is certain, however, is that data that is often collected voluntarily and presented as part of a "gamified" app experience may also end up being used in more dubious contexts.

An employer could argue that responsibility for this lies entirely with the employee. TikTok and Strava are very rarely mission-critical apps, and of course their use during working hours can be easily regulated too. At the same time, however, the waters are muddied by the fact that more employers are introducing various types of codes of conduct that also apply to online environments, and employees often represent their employers outside working hours. The boundaries are suddenly less clear when a digital misstep in an employee's free time could harm the employer's reputation.

In my opinion, the question of where an employee's professional sphere ends and the personal sphere begins is highly relevant in a working environment that is becoming increasingly digitised. Evidently, this has implications for data protection, as no one wants employees to accidentally leak their employer's information, but it also has a bearing on privacy: the sanctity of a private life is important precisely because of work, as well as for other highly personal reasons.

Unwelcome leaks in private life

The closer we get to matters of an intimate and physical nature, the more apparent the need for respect for privacy and integrity. Over the last decade, online dating has become the most popular way of finding a partner, bringing with it a host of challenges. Ask around in your social circle and you will no doubt hear horror stories about dates who turned out to have used fake profile pictures or said something in their profiles that turned out to be untrue. Yet the risks are greater for some users than for others. The Grindr app is primarily aimed at an LGBTQI+ user base. By seeking out potential partners in an app tailored to LGBTQI+ users, people can avoid the discomfort of standing out from the crowd in a heteronormative, cisnormative app. All very unproblematic. But Grindr's history has been marred by data leaks and design problems. Grindr was rocked by major data leaks in the mid-2010s. Millions upon millions of data points belonging to the app's users appeared more or less publicly on the Internet, which meant that users who did not actively choose to

remain anonymous could be directly mapped and tracked. This was embarrassing for many, but potentially life-threatening for anyone living in a country where homosexuality is stigmatised or illegal.

Time and again, people using Grindr to look for either a partner or a casual encounter have been lured into relatively unsophisticated traps where – for example – a potential date turned out to be a gang of thugs out for a spot of "gay bashing". Use of the app increased during the pandemic years, and presumably because of this – albeit anecdotally – state repression against LGBTQI+ people also increased in several countries, often using evidence directly gathered through surveillance on Grindr. This led to arrests, or even imprisonment, often involving violence. The fact that this happened in non-democratic countries should be serious enough, but a few years ago, the same kind of mapping took place in the USA as well – and was even more sophisticated. A religious magazine had been planning for some time to reveal the sexual identity of a Catholic priest, as it was suspected that he had been involved in a number of homosexual relationships. The magazine succeeded in publishing a relatively well-substantiated investigative article, reportedly based on data that had been purchased from commercial organisations, which was then cross-referenced with location data from what was thought to be the priest's Grindr account. This case highlights the commercial data economy, where the all-too-common disclaimer "data may be shared with third parties" has, in some cases, resulted in personal data being shared far beyond what was originally intended; and far beyond what users either wanted or could have imagined.

Although mobile phones are now the most personal digital items we have, some things are extremely personal regardless of the level of digitalisation. For a long time, one dominant subculture among health apps has been the market targeting women's health, the main category being ovulation and menstruation apps. Entering information about the menstrual cycle allows users to access a wide variety of apps that offer other kinds of health predictions. Deviations from a normal cycle can be reported via the app, as can assumptions about fertility – even the somewhat questionable notion of "safe" periods. That is fair enough: it is reasonable for health apps to target a female user base. And it is certainly fair to assume that the developers of such apps, which collect some of the most privacy-sensitive data that women can submit, have a solid grasp of data protection and privacy issues.

Recent research shows how British period-tracking apps, along with other apps that track women's health trends, routinely leak personal data

to third-party services that then sell the data on to potentially unethical advertisers. Apps aimed at women are a lucrative market, the retail market even more so. Accurately targeted marketing leads to more responsive customers, and missing a period has resulted in ads for baby clothes. That same user data is suspected to have been leaked in several high-profile cases in the USA, where women's personal data was unfortunately cross-referenced with the mapping of abortion clinics – at a time when abortion rights are increasingly being called into question. This was also fuelled in several instances by the fact that abortion clinics themselves had added various types of statistical tools to their websites in good faith, to gain an insight into their usage. Those same statistical tools have been shown to collect information on visitors' browsing habits before and after visiting the abortion clinic's website: when data has been leaked, it has been possible to identify people looking for information about abortions, and when they did so. Rarely does data capitalism come so uncomfortably close as when such connections are revealed.

In such cases, it is easy to say that users have entered the information in the relevant apps of their own volition, without coercion. Blaming users is a common and recurring digital defence mechanism. In the cases above, it can also be argued that these are examples of entertainment apps that are used for reasons linked to pleasure and curiosity, rather than anything critical to personal health. At the same time, competition in the digital healthcare market is rapidly increasing, with app-based healthcare solutions available in both the private and public sectors. It is tempting to use a healthcare provider that offers fast responses around the clock via a user-friendly app, but unfortunately, how the information we enter is handled is not as easy to understand. Organisations can limit their liability through contractual clauses that users must accept to be able to use the service. But in a digital age where users in general arguably suffer from "consent fatigue", and where it is quicker to click past cookie notices and terms than to actually find out what the cookies and terms mean, I think this is irresponsible idleness. Actors that are either part of, or closely connected with, the public sector's offering to citizens should not shift data-related risks onto users. This is where the public sector must be prepared to take the lead.

It is difficult to describe in simple terms how digital organisations should manage data-driven security risks, which are often linked to algorithm-driven user incentives. The direct recommendation to all organisations handling any kind of sensitive information is to implement a data separa-

tion standard, where employees have to use devices that are restricted to work-related purposes and managed centrally. While risks may certainly remain, especially the risk of private devices being linked by means of repeated co-locations via location services, this approach helps to create more secure isolation of data. The risk of accidentally sharing sensitive business data is greatly reduced, and I would argue that we can no longer rely on the ostrich approach and claim we were unaware.

The state as a data stakeholder

New ways of understanding and assessing threats and risks have emerged in today's commercialised digital reality. It is difficult nowadays to find any organisation that does not advertise on social media to reach out with marketing or information. Digital advertising tools, or advertisers' interfaces on social media platforms, often provide more clues than we might imagine. Anyone wanting to place an ad on a platform such as Facebook, X (formerly Twitter) or YouTube can choose from a wide range of parameters for more successful and accurate marketing. For example, it is not difficult to aim an ad at a target group, such as "males, aged 35-45, living in one- or two-dwelling houses near to large or medium-sized towns, parents of at least two children and interested in public affairs, cars and shooting sports". Anyone who has ever seen the number of variables available for advertisers to pick and choose from is likely to become more sceptical about the data-driven surveillance economy that we are increasingly unable to avoid.

It is not hard to see how this soon raises questions in the reader's mind about how the above examples can or should be linked to the public sector. The answer is every bit as simple as it is obvious: it is standard practice for the private sector to develop services for the public sector. That is why it is essential for the public sector to recognise and analyse the risks of outsourcing almost all development of digital solutions to the private sector. We also see this for the state as an employer when procuring algorithmic management systems with intrusive and covert surveillance of work as the default setting.

Of course, Swedish bureaucracy was interested in citizens' data long before social media platforms and digital surveillance became a reality. Many of the challenges regarding state centralisation of data are not primarily the result of technological innovation in and of itself. In my view,

the handling of protected personal data is a case in point. The state's responsibility to protect its most vulnerable citizens is regularly put to the test when it emerges that the Swedish Tax Agency's own procedures are one of the reasons why sensitive data is disclosed. There are times when rudimentary process problems, such as having to post items to people with protected identities, causes difficulties. It is difficult to work out how physical items, such as letters and parcels, can be delivered to people whose whereabouts and addresses are regarded as critically sensitive. At the same time, that sensitive data is processed in systems that the Swedish Tax Agency's own staff have struggled to manage securely at times. I spent some time advising a number of women's refuges on security issues. Even after the refuges implemented a wide range of security measures for their safe houses, there was still a great deal of uncertainty about what could be expected from the state's protection mechanism. These questions arose at around the same time as the Swedish Transport Administration outsourced the management of police personal protection to Serbia; meanwhile, the provider contracted by 1177, the Swedish medical helpline, for its healthcare calls turned out to have posted recordings on the Internet and allowed them to be accessed by the public.

The fact that the private sector and its data solutions are insecure at times poses a genuine threat. However, it is understandable, and possibly even acceptable, that commercial technological development is driven by maximisation of profits. This is a line of reasoning that can and should be questioned; nevertheless, it is something that permeates large parts of our society, even beyond digital innovation. The public sector should be driven by something else, where the value of the activity cannot always be monetary. This is particularly true in contexts in which the capital managed by the public sector is not financial, but human.

Responsible contracting

One striking example of this, where data protection challenges already encountered by the private sector could be assumed to have been resolved, was the wide range of apps launched in partial response to the issues raised by the global pandemic. We saw international examples of extremely intrusive COVID apps, some of which forced citizens to take regular selfies to upload to outsourced cloud solutions as part of infection control initiatives. Sweden was no exception: during the pandem-

ic, an app was launched here in connection with a research project at a major Swedish university. Targeted social media marketing encouraged citizens to enter a variety of health data into forms uploaded to foreign cloud servers. It was claimed that the data would be anonymised, as users merely identified themselves using parts of their postcode. However, the digital trails that were left by entering the data were ignored, so, as the previous examples show, the data-driven surveillance economy means that they may reveal more about users than anyone had anticipated.

One interesting feature of the COVID-19 app described above was that it was initially presented to Sweden's Public Health Agency, which was responsible for infection control and tracking during the pandemic. The agency showed little interest, however, as no needs analysis linked to the development of such an app had ever been conducted. Instead, Lund University was persuaded to incorporate the app into a research project, which, given the university's official status, could easily be interpreted as state-sanctioned data collection. The app element of the project, it seems, was soon dropped, but no information is available on how much data was collected or how it was used.

Another revealing example is the City of Stockholm's major investment in *Skolplattformen*, the School Platform, an app used collectively by all local schools. To date, the price tag stands at more than 1 billion Swedish krona (around €100 million at the time of procurement), and the cost per student has outstripped the cost of schoolbooks in some years.[2] Despite the cost, the app has been heavily criticised for being difficult to use, and some of its features have not even been fully implemented – and by the time this book is published, the app will have been taken out of service completely. This is probably one of Sweden's most widely discussed and scrutinised public procurement procedures, yet all it took was for one inquisitive parent to change the platform's URL, and suddenly it was possible to access personal data about every child in the entire Stockholm region. *Skolplattformen* also came in for criticism by the Swedish Data Protection Authority (now the Swedish Authority for Privacy Protection, IMY) when a General Data Protection Regulation review of the system revealed that an unjustifiably large number of people were able to access the protected personal data of particularly vulnerable pupils.

2 Lidbom, O. (2024) *Skolplattformen: En sann historia om miljarder, konsulter, föräldrar och barn* (Stockholm: Fri Tanke)

Accountability and ignorance

There are many reasons to continue advocating for innovative digitalisation of public services. The private sector, with its capacity for innovation, can lead the way on multiple fronts. However, the public sector – the fundamental remit of which is to provide services and welfare to citizens – has to learn from the mistakes made in solutions developed by the private sector. It is not for public actors to replicate problems that have already been identified. Over the years, the digitalisation of society has been challenged by both private actors and the solutions developed by them. It is not hard to understand that there is a certain pressure on the public sector, which must simultaneously prioritise the stability and development of society. Private actors, who do not have the same fundamental societal accountability, can use their experimental innovation to be disruptive. The key here is to tread carefully. The responsibilities that rest with the public sector cannot be delegated to other actors, unless the issue of accountability is also clearly regulated.

The rapid escalation of digitalisation in recent years has brought the issue of accountability to the fore. In many ways, it seems like an impossible dilemma. Public actors have been experimenting with various solutions and models to find ways to facilitate digital activities that keep pace with rapid global technological change. Unfortunately, time after time, this has meant that issues relating to sectors that have already been opened to competition or privatised have once again become testing grounds. The efficiency drive that today's societal developments appear to demand means that digitalisation will be a permanent feature of the landscape going forward. Difficult issues, such as data protection and digital security, must underpin this work if we are to ensure that such developments benefit society and its citizens.

Public authorities should not be drawn into innovation initiatives that fail to prioritise citizens' rights and freedoms. Launching or permitting digital solutions that are important to society without understanding how the technology is structured or how the solutions handle citizens' data means deliberately ignoring the instruments of power – the algorithms – that are embedded in the new technology. It also overlooks the impact they will have on citizens' faith in the ability of society to evolve in harmony with the digital age in which we live.

Anne Kaun and Maris Männiste

"What should I do?" Interacting with local government in the age of AI

As a citizen, you meet the state numerous times in your everyday life – for mundane and life-changing matters. How does one apply for financial assistance? What documents are required for a building permit? When, and through which app, should an application for pre-school or school placement be submitted? How is the rubbish collected?

Such encounters with local government largely revolve around knowledge and information: clarifying bureaucratic procedures; navigating decision-making processes; and understanding rights and responsibilities. With technological development, however, these "bureaucratic encounters" have been reshaped. Interactions between citizens and the state that took the shape of personal exchanges with individuals now default to digital interfaces. We have shifted from speaking to civil servants to navigating automated systems and artificial bots. Many chapters in this book describe the algorithms "behind" these encounters. This chapter focuses instead on what happens during the encounter itself – and on the organisational changes this transformation inevitably and sometimes unpredictably generate.

Introducing chatbots to organise citizens' knowledge and information gathering is not an inevitable development. Rather, it is both a technological and social transformation based on specific practices, imaginaries about the potential future that technology might bring and material aspects of the technology implemented itself. Often, the process for implementing technology into organisations is contradictory and controversial, as a recent Swedish debate on whether the initial assessment within the healthcare system should be automated using a chatbot illustrates.[1] This specific chatbot introduction was motivated by the claim that triaging would become more efficient and safer. This imaginary is countered with fears of discrimination against vulnerable groups as the public debate around the chatbot shows. The public debate centred on potential risks,

1 «Chattbot införs på 1177 – trots kända risker för patienter". *Dagens Nyheter*, 5 October 2023.

increased inequality between social groups with different needs and fundamental societal values regarding access to the healthcare system. This case points to the fact that the introduction of technical systems is often accompanied by controversy, debate and strong feelings.[2]

How many chatbots are used in public administration?

It is difficult to estimate how extensively chatbots are used in general, but a market survey estimates the global chatbot market at $7.76 billion, with an expected growth of over 23% in the next seven years.[3] In the Swedish public sector, many government agencies and, increasingly, municipalities with high levels of citizen interaction have introduced chatbots. A report from AI Watch 2020 also states that the dominant form of automated and artificial intelligence (AI) based systems used in the public sector are chatbots (compared to other uses of AI, such as automated citizen profiling or use of generative AI in citizen-state interactions).[4]

The chatbots in use vary in complexity and accessibility – from advanced frequently asked questions (FAQ) lists to interactive, machine-learning-based applications based upon commercial large language models, such as OpenAI's ChatGPT, Google's Gemini, Microsoft's Copilot, Perplexity and DeepSeek, and live chats. The most common chatbots in the public sector are service-triage applications used to categorise knowledge and information needs and structure the search process accordingly, partly integrating natural language processes to analyse prompts.

Chatbots are part of a broader category of technology that has recently been conceptualised as communicative AI. Over the past 20 years, many communicative technologies, programs and devices have assumed the role of communicator and knowledge provider, either by interacting with humans or by performing communicative tasks on their behalf. Communicative AI includes conversational agents, social robots and software for automated writing, and varies in how it functions as a knowledge in-

2 Marres, N., M. Castelle, B. Gobbo et al. (2024) «AI as super-controversy: Eliciting AI and society controversies with an extended expert community in the UK». *Big Data & Society*, 2(11). DOI: 10.1177/20539517241255103

3 "Chatbot market size & trends". Grand View Research.

4 Misuraca, G. and C. Van Noordt (2020) *"AI Watch - artificial intelligence in public services: Overview of the use and impact of AI in public services in the EU". JRC Research Reports*, JRC120399.

termediary – from interpersonal conversation partners to content producers. Chatbots are computer programs that interact with users based on natural language processing (NLP) models. They can be any software application or machine agent capable of engaging in a conversation or interaction with humans via text or voice.

Chatbots are a form of narrow AI that extract meaningful information from free text based on user input and help "find the intent behind the question a user asks and deliver an appropriate response".[5] Chatbots are used for a wide range of purposes, such as providing information, customer service, entertainment, or even for social companionship or therapy. They can be standalone applications or provided by customer service on specific websites.

Chatbots are expected to solve many of welfare's challenges

In the public sector, AI is expected to solve several key problems for public services: a control problem (ensuring compliance of the public administration with regulations); a cost problem (how to meet demand with reduced resources or increased demand with the same resources); a convenience problem (how to meet growing citizen expectations); and a contact problem (how to maintain trust and mutual empathy towards the public administration). However, not all of these issues can be solved through the use of chatbots.

Previous research shows that chatbots can reduce the administrative burden on public sector organisations and improve and standardise communication between the state and citizens. Digitised self-service tools such as chatbots, which are available 24 hours a day, 365 days a year, can offer more convenient services for citizens when interacting with the state. Research on chatbots used in Norwegian welfare services has shown that the bots can handle inquiries under heavy load equivalent to the capacity of 220 service employees.[6] In a time of labour shortages in

5 Goyal, P., S. Pandey and K. Jain (2018) *Deep Learning for Natural Language Processing: Creating Neural Networks with Python* (New York: Apress), p. 19.

6 Følstad, A. and N. Bjerkreim-Hanssen (2023) "User interactions with a municipality chatbot–lessons learnt from dialogue analysis". *International Journal of Human–Computer Interaction*, 18(40: 4973-4986. DOI: 10.1080/10447318.2023.2238355

the public sector, chatbots are thus also seen as time-saving tools for a reduced workforce.

Research has also shown that citizens choose to use chatbots in the public sector primarily for pragmatic reasons, as a simple way to get answers instead of making a phone call or reading large amounts of text. Verne and colleagues show that citizens interact with chatbots as if they were talking to humans, using long sentences containing multiple keywords and greeting phrases.[7] This can be problematic, as it increases the risk of the chatbot misinterpreting the user's intent and failing to find a good match between the question and an appropriate response.

However, most public sector chatbots today lack creative translation capabilities, as they can only compare user input with the predefined keywords entered into the system. This can lead to instances of failure and the need for repair, or what we consider digital frictions in the interaction, where the parties must identify and resolve the misunderstandings that have occurred. In such cases, users are often given fallback responses intended to keep them engaged in the conversation and refine their query (the ubiquitous "I didn't understand the question" etc.). These fallback responses prompt citizens to rephrase their questions, which can be difficult without domain knowledge.

Previously, caseworkers took on the role of mediator to help citizens identify the correct category where their information needs fit. With the use of self-service systems and chatbots, this burden instead shifts to the citizen. Now the citizen must have knowledge to use the correct keywords in their queries and how to correctly interpret the chatbot's responses, which is what we would call domain knowledge in this instance.

Moreover, chatbots are often organisation-specific for a municipality or public agency and can typically only respond to questions relevant to that specific organisation's scope. This means that citizens must know which topics and questions fall under which institution's responsibility. In other words, users need specific interaction knowledge to communicate effectively with chatbots that differs from interactions with other humans that include clues from body language.

7 Verne, G. B., T. Steinstø, L. Simonsen et al. (2022) «How can I help you? A chatbot's answers to citizens' information needs». *Scandinavian Journal of Information Systems*, 2(34): 7.

Kringla: A specialised municipal chatbot

Hello and welcome to Södertälje Municipality's chatbot!
My name is Kringla and I'm here to help you with questions about municipal services, general information about the municipality and guidance. What can I help you with? 😊

- Application – mother tongue
- Question about building permits
- Question about preschool
- Phone hours

Question: Why are you called Kringla?
Response: I'm sorry, I don't quite understand what you mean. Œ
It might be because I need more information to give a correct answer or that your question is not a municipal matter.
Unfortunately, I can't carry on a lively discussion or respond to your specific case. Please try asking your question in another way. Thank you for chatting with me! 😊

This is what a conversation with the chatbot Kringla from Södertälje Municipality in 2025 might look like. The example illustrates that the chatbot has limited flexibility when it comes to answering unexpected questions, which has to do with its technical software infrastructure. The chatbot has been integrated into the contact centre's routine communication work, which includes a reception at the city hall, a call centre, an email service and the chatbot. A specialist group works about two hours per week to maintain and improve the chatbot.

In practice, users of the chatbot receive answers that resemble FAQs compiled on the website and in the internal database used by information officers during live interactions at the call centre or visitor centre in the city hall. In many cases, the chatbot responds with links to relevant parts of the municipality's website and related forms, such as in the case of school placement applications where the client receives links to the relevant information page. However, the chatbot does not provide personalised service, such as logging in or integrating digital ID solutions.

The interactions between clients and the chatbot are based on matching client questions with a database of potential, preprogrammed answers. AI – or in this case, NLP – is used on the user side to analyse queries or "prompts" and then link them to prewritten responses. All responses sent to clients were formulated and formatted (sometimes in-

cluding emojis or other special characters) by the municipality's service staff. The AI "training" referred to above consists of collecting, updating and maintaining the responses in the database.

Imaginaries: The chatbot as AI

The general technical conditions are translated and connected to imaginaries about what the chatbot is and does. On the website and in various official documents, the chatbot is presented as an AI chatbot. However, speaking of AI evokes specific expectations regarding complexity and the level of interactivity, both among clients and the service staff in the municipal office. The software itself is not based on machine learning and would not, by some definitions, be classified as AI at all. But whether the chatbot is AI-based or not is not the central issue here. Rather, it is the fact that the municipality presents the chatbot as AI-driven and thereby positions itself within an ongoing AI discourse that is much broader than chatbots alone.

The chatbot's implementation is justified by imaginaries surrounding its design as an assistant or servant, striking us as a public administration idea closely moulded on private sector precedents. The most well-known such agent is probably Apple's voice assistant "Siri", which can be used for tasks such as selecting songs, reading emails and making appointments using the company's smartphone. Our interviewees particularly emphasise improved service and access to the municipality as the main purpose of the chatbot. Kringla was implemented with a vision and expectation of 24/7 remote access to public services, something one of our informants strongly emphasised. The chatbot is available outside of office hours and is not tied to a specific location.

Our interviewees repeatedly emphasised that the focus is on citizen access, rather than the replacement of staff, to counter public concerns about automation reducing the number of caseworkers. Instead, the additional service that the chatbot will enable is highlighted, avoiding explicitly discussing critical issues such as reduced resources or efficiency targets.

The chatbot is available around the clock, and the unit manager proudly noted increased use of the chatbot outside normal working hours, which now accounts for approximately 30% of chatbot interactions. However, overall usage remains comparatively low. Since the new chatbot was in-

troduced in Södertälje, there have been around 200 chatbot interactions per month, compared with 1,800 emails and 7,500 phone calls during the same period. The municipality now plans an information campaign to raise awareness about the chatbot and increase the number of users.

In general, the chatbot, and other projects included in the municipality's broader digitalisation strategy, indicate that the implementation of AI also functions as a form of public relations. Södertälje Municipality has been particularly ambitious, launching several Internet of Things and open data initiatives, including live updates on lake temperatures and real-time monitoring of bicycle traffic over one of the town's bridges. The contribution to the positive framing and branding of the municipality as future-oriented and heavily invested in smart technologies is not just a welcome side effect, but potentially the most important outcome of these implementation projects.

The opportunities that chatbots offer in the public sector are largely based on applications already common in the private sector. These chatbots are designed to act as "assistants and civil servants within the public sector", while citizens take on the role of users who follow learned scripts for how to interact and communicate with specific systems–in order, as one of our Swedish interviewees put it, "to help citizens faster".

Citizens are becoming increasingly experienced and knowledgeable users of chatbots, but each chatbot currently provides a unique experience because they are trained in slightly different ways, even when the underlying software infrastructure is the same. At the same time, expectations shaped by prior interactions with other, perhaps more advanced chatbot solutions, as well as through media coverage of chatbots, rarely match users' actual experiences with municipal chatbots like Kringla.

Implementing a chatbot

Implementation projects involve not only material conditions and imaginaries, but also practices; in other words, the actual work of "making" and "maintaining" a chatbot. This includes not only programming practices but also how the implementation is organised. In addition to implementation practices, there is also the citizen perspective where their practices change in response to the new communication infrastructure.

The implementation of municipal chatbots like Kringla is tied to cycles of public procurement, with contracts lasting up to five years. These time

intervals can lead to rapid changes when a contract expires. In our case, the switch of supplier was presented as a major advantage, giving the municipality more autonomy in how the chatbot database is updated and adapted. At the same time, the switch meant that service staff had to be trained in the new infrastructure, and the system had to be rebuilt from scratch. Although constructing the new chatbot required a considerable investment of time, our interviewees emphasised that the transition was not perceived as problematic. Staff are accustomed to regularly switching to new software systems. Most of them already work in three to four different systems simultaneously. Learning a new program was seen as just another part of their everyday work routine, and already the Kringla we met was the second iteration of the chatbot. Nonetheless, organisational structures are needed to support such an implementation project. We could also note that dysfunctionality regarding the digital work environment is almost taken for granted by municipality workers, something addressed by Colclough and Johnston in their chapter.

Digital change brings organisational change

In connection with the introduction of the Kringla chatbot, the municipality also restructured and reorganised parts of its operations. The chatbot was intended to enable more time-efficient communication with citizens, which, in turn, would free up more time for service staff in the contact centre. As a result, staff were assigned new service tasks related to other departments. The contact centre, which had previously been primarily responsible for citizen contact, was now also tasked with administrative duties for other municipal departments, such as archiving and handling incoming documents. To meet the expanded assignment, the duties of the contact centre's service staff were both broadened and standardised for greater efficiency. In short, the implementation of new technology as a consequence brings institutional and organisational change.

The chatbot's implementation is thus embedded within a broader data and organisational infrastructure that is a larger environment linking different systems, which are rarely fully compatible with each other.

Beyond the broader organisational structure, several actors support the implementation project. These include companies providing off-the-shelf software solutions, digital consultants, municipal service workers

and information officers. Most do not have specialised training in public administration and bring different professional values and identities to the project. These diverse backgrounds partly led to conflicts of interest and differing priorities. For example, there were tensions between maintaining long-established administrative routines and introducing new work methods associated with the chatbot, such as greater standardisation of citizen-municipality contact and clearer communication pathways throughout the municipal organisation. The chatbot, as a newly standardised and standardising gateway to the municipality, became part of this broader transformation.

In Södertälje, this meant that the contact centre was not only responsible for the practical development of the chatbot but also became an ambassador for the broader transformation effort, for example, by convincing other parts of city hall that the chatbot would bring benefits to their daily work. In this context, a specialist group was also formed with technical responsibility for the chatbot, which also served as a "digital ambassador" within the municipal organisation.

Conclusion: Shifts in meeting the state

Bureaucratic encounters between citizens and the state have changed with the introduction of new communication technologies, such as chatbots, which can be described as communicative AI. This has also transformed the way citizens acquire knowledge and information in relation to the public sector. Citizens now need to learn how to communicate effectively with chatbots to receive accurate answers, for example, by formulating their questions and problems briefly and using relevant keywords. In other words, a form of chatbot literacy or AI literacy is currently required, including domain knowledge on how to write effective prompts. This might change with new iterations of chatbots and large language models, but we currently see a pressure to adjust to the needs of municipal chatbots that are still rudimentary.

At the same time, the previous guidance provided by municipal service staff has shifted from explaining bureaucratic processes and decision-making chains to focusing more on how *digital* services and platforms work and how to navigate interfaces. This represents a knowledge shift among public employees, moving from understanding bureaucratic processes and administrative structures to more narrow technical expertise.

The examined chatbot project is presented alongside ambitious promises for the future, such as simple and inclusive access to public services. There are, however, also underlying values, for example, increased efficiency, driving investments and resource allocations intended for the development of datafication infrastructures like chatbots. Digitally enhanced accessibility improves municipal services for some individuals, while others – including those who have difficulties processing digital information – do not experience the changes as improved service. Here, the benefits for some are contrasted with universal access. This can create digital frictions related to values concerning technological development and implementation more broadly. Municipal employees working with the chatbot implementation play an important role in negotiating and mitigating these digital frictions. They either emphasise efficiency gains for other units or focus on future potentials for citizens that might one day be realised.

Towards the data welfare state?

Municipal chatbots must also be viewed within a broader context of data welfare, namely, the increased introduction of data-based methods in welfare provision. The data welfare state and the automation of welfare refer to a fundamental social transformation toward using data-rich technologies to shape people's daily well-being. These technologies can both enhance quality of life and potentially introduce new vulnerabilities for all parties involved concerning data, digital technologies and automation. However, there are a number of requirements to achieve data welfare that are not in place at the moment, as the Kringla case illustrates. In the following, I make two broader propositions that allude to what the data welfare state should be about:

The data welfare state has to be grounded in the lived experiences of the people. Rather than finding problems to technological solutions that are decoupled from people's needs and capabilities, we need to turn the datafication process on its head and start from people's perspective. What is meaningful and necessary in their lives? How can data and digital infrastructures be supportive? But also, how can they be designed in ways that are intuitive and accessible for people?

This re-centring of people, which has so far not been achieved by applications like Kringla, in the automation and datafication process im-

plies the need for a shift in the methods and forms of implementation, where citizens must play a larger role. Focusing on individuals involves considering both the target audience for potential automated solutions and supporting active data activists who voluntarily contribute their data for informed decision-making in the public sector. This also entails raising citizens' awareness about data collected about them for automated decision-making purposes. Focusing on individuals involves not only accommodating the preferences of the target audience for data solutions but also acknowledging the potential vulnerability of data experts and decisionmakers who must make accurate and equitable decisions in uncertain situations when utilising data solutions.

The data welfare state needs to be built on public infrastructures that are owned, maintained and developed by public actors. The starting points for the technologies that are used within welfare systems cannot be based on profit-maximising principles but need to follow public values. Consequently, the overly strong reliance on Big Tech players – in the Swedish public sector, for example, primarily Microsoft – needs to be critically reconsidered. All components of data centres, from the cable infrastructures to the cloud solutions, need to run on public values rather than commercial interests. Opting out, sometimes proposed for users of social media platforms and other digital infrastructures that live on our data, is not an option in the context of the welfare state. In addition to the envisioned ideal of creating public data infrastructures according to the ideals of EU data spaces,[8] which establish not only norms, principles and examples of best practices in data sharing, it is crucial to create favourable conditions for the establishment of shared data cooperatives based on individual initiatives[9].

8 A European Strategy for Data, European Commission, 2024

9 Further reading: Guzman, A. L. and S. C. Lewis (2020) Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society*, 1(22): 70-86. DOI: 10.1177/146144481985869 and Suchman, L. (2023) "The uncontroversial 'thingness' of AI". *Big Data & Society*, 2(10). DOI: 10.1177/20539517231206794

Gerard Rinse Oosterwijk and Simon Vinge

What gets measured gets algorithmically managed

Miners once carried canaries into tunnels, carefully observing them for signs of dangerous gases. When the birds died, the workers would promptly exit the subterranean shafts. In today's labour market, platform workers may occupy the most precarious position – and we argue that their exposure to management practices today is strikingly similar to the canaries' exposure to noxious air. Specifically, this concerns *algorithmic management (AM)*. Before the widespread adoption of such technology, the entire labour market could be likened to underground workers.

Platform workers are employed through apps, where algorithms decide where they go, the work they do and how much they will be paid. This treatment is not confined to the most visible platforms (such as food delivery and ride hailing) but is also becoming the norm in more traditional sectors, such as warehouses and service industries.[1] Moreover, algorithms are expanding beyond these sectors as well; AM is gaining traction in almost any sector where there is enough data.

The digital transformation, AI revolution and the COVID-19 pandemic, which normalised remote work, has radically increased this data availability and accelerated the transformation. Today, AM can be found in virtually every corner of society. Fields such as healthcare, education and public administration – previously comparatively insulated from the most precarious effects of technological change – are now in the direct line of fire. This will have massive consequences not only for the labour market but also for the everyday experience of work we all perform. This chapter aims to describe this development and to outline what is needed to alter the path going forward.

1 Gent, C. (2024) *Cyberboss: The Rise of Algorithmic Management and the New Struggle for Control at Work* (London: Verso); J. Wrangborg and M. T. Jensen (2024) "Algoritmen som chef konsekvenser av algoritmiskt styrt arbete". Policy study. FEPS, June; C. Håkansta, R. Lind, P. Strauss-Raats et al. (2024) "Algorithmic management: Experiences and responses". Policy study. FEPS, December.

AM demands worker surveillance

AM refers to a type of software application that monitors, instructs, manages or evaluates workers using automated systems. The AM systems are fed with vast amounts of data from software tracking and monitoring everything from keystrokes and eye or mouse movements to body temperature, Wi-Fi networks and GPS positions. The software tools are programmed with algorithms that can make automated decisions that affect workers based upon this data. They can perform the simple instruction of which task to do next in a warehouse, and they can use surveillance data for more complex tasks, such as aggregating and predicting performance scores, and determining most if not all issues facing management.

We often speak of artificial intelligence (AI) in the workplace and this can include AI-driven systems, but it does not necessarily have to. We see both rules-based models, where the algorithm is still programmed by humans, and a machine learning approach, where the algorithm makes its own rules based on the output and data used to train the model. This distinction is important to note, and is further developed the chapter by Karim Jebari.

What gets measured will get (algorithmically) managed

It is only through the digital transformation of the shop floor that it became possible to gather the necessary data to run the algorithms that generate accurate instructions or predictions. This "datafication" of work enables the automation of certain tasks and activities that human managers previously did. In the ubiquitous warehouse example referred to, the algorithm orders shelf picking, but the technology can be readily deployed for any tasks where we have data. Which, today, given the digital tools almost all of us use, means most workplaces.

This development is not something we expect in the future – it is already here. Estimates vary, but the most recent studies indicate that in 70-80% of European workplaces, one or more types of AM tools can be found.[2] Examples of AM range from automatic shift scheduling in several

2 Milanez, A., A. Lemmens and C. Ruggiu (2025) „Algorithmic management in the workplace: New evidence from an OECD employer survey". OECD Artificial Intelligence Papers,

sectors to workflow management in government agencies or driver-assist systems in the transportation sector. Additionally, in all sectors, including white-collar sectors such as finance, AM is gaining traction, evaluating workers in real time based on their performance metrics. In office jobs, generally, there are tools in the Microsoft 365 software package used for performance tracking of workers.

Labour market balance of power is at risk

This new development has many consequences for workers and their job quality. It is a slightly different kind of wave of automation than the ones that came before, because it directly automates or at least changes the worker-manager relationship. In some cases, like with platform workers and warehouse workers, the manager is entirely replaced by the machine when instructing workers. In other cases, AM is assisting the manager and providing detailed lines of action to execute. An optimised shift schedule or a detailed data-driven performance report with benchmarks for other workers and targets is easily validated and difficult to overrule by the human manager. This also reflects a shift in the balance of power between workers and management in favour of the latter.[3] Management always knows more, and managers are provided with new, detailed arguments or proof that can be used against workers if needed. All of this is entirely automated and requires minimal effort by the manager and incurs minimal costs for the company. Additionally, preliminary results from worker surveys found that it can lead to a loss of autonomy, decreased job motivation, and an increase in stress and job insecurity for workers.[4] This makes the rapid spread of AM across European workplaces alarming.

As with other algorithmic use cases, a key issue is that of AI bias. This can have worrying effects, not least through the frequent use of algorithms for recruitment, where the existing bias in the data risks being replicated when relying on it for the selection of candidates. Lacking safeguards and checks and balances, whole groups in society could be

no. 31, February.

3 Cox, T. and G. R. Oosterwijk (2024) „Algorithmic management in the workplace: Case studies on the impact of algorithmic technologies in seven sectors in the Nordics". Policy study. FEPS, October.

4 Jensen, M. T., G. R. Oosterwijk and A. S. Nørgaard (2024) „Computer in command: Consequences of algorithmic management for workers". Policy study. FEPS, June.

excluded because of certain traits that are not viewed favourable by the algorithms, which are usually trained with historical data from the existing workforce. Not rarely, this happens within the algorithmic "black box", hiding the bias or even making it impossible to see. Despite many known cases where this went wrong, many companies still rely on AI tools for the selection of candidates.[5]

Transparency is not enough

Moreover, while one of the often-preached solutions of algorithmic transparency can be part of the answer, it is worker influence that seems to be a strong moderating factor to reduce the adverse effects of AM on workers.[6] Here, transparency is of course a necessity for influence, but it is far from enough. Involving workers in the process of implementing and running AM systems is a crucial precondition for achieving not only humane working conditions, but also to reach the objective of increased efficiency gains.[7] The paradox is that while AM weakens the position of workers and trade unions, it is in the interest of employers and companies to actively engage both their workers and unions.

As with most tools, even artificially intelligent ones and algorithms, it is how they are used that decides the outcomes. In the platform economy, the manager is completely substituted by in-app AM, whereas in traditional sectors, it is more commonly used to enhance the manager's ability to make better decisions. At least that is the theory, because the manager or the worker will be inclined or motivated to follow the instructions given by the AM system. The phenomenon of *machine bias* – that humans tend to trust machines, even if we know they are biased or take limited concerns – is well described in the literature. Overruling the prescribed driving route or the work capacity evaluation recommendation will raise questions. A change in the premade schedule or judgement will depend on the willingness of the manager to take personal conditions into

5 Goodman, R. (2018) „Why Amazon's automated hiring tool discriminated against women". ACLU, 12 October.

6 Jensen, M. T., G. R. Oosterwijk and A. S. Nørgaard (2024) „Computer in command: Consequences of algorithmic management for workers".

7 Cox, T. and J. Anttila (2024) „Algorithmic management and workplace digitalisation in Finland: Insights from the transport and logistics and retail sectors". Policy study. FEPS, September.

account, on the time available, and the resources and experience of the worker to actually raise the issue. When evaluating the performance of a worker, an automated report substantiated with numbers and scores on key metrics will reduce the leeway of even the manager to judge a worker on more qualitative measures.

Who controls the code decides the outcome

We are optimists and while it is crucial to outline the risks and dead ends with new technology, it is of course important to envision the ways forward. There are many applications, from automating tedious and monotonous tasks to helping the human eye find deviations or areas for improvements in workflows, where AI – and even AM tools – have massive potential. However, there are some almost insurmountable obstacles to achieving such a path of development in a worker-friendly way.

The main issue here is the lack of control by either the worker or management over the coding of these systems and the basic ways the algorithm functions. In most cases, the AM systems that companies deploy are bought off the shelf from tech providers directly. Often, these are large US-based firms and the tools are implemented with the help of external consultants. From this setup, the people in the organisation that work with the AM system lack the knowledge and capacity to deploy the algorithms in a worker-friendly or even efficient way.[8] The algorithm that becomes part of the work process is a black box for them, but meaningful transparency means that within the organisation there are people with the capability to understand the systems and have the power to tweak and adjust them to get a more positive outcome. We also see that workers need to be involved early – already in the development phase of AM tools – which is not possible without access to the source code and having a fundamental grasp of how such systems work.

For the developers of AI systems, especially those based in the USA, adopting a cooperative, consensus-based approach to labour and industrial relations is not top of mind when programming and designing such systems. In the current context of a Silicon Valley emboldened by the

8 Juego, B., T. Ø. Kuldova and G. R. Oosterwijk (2024) „Algorithms by and for the workers: Towards a fair, democratic and humane digitalisation of the workplace". Policy study. FEPS, January.

Trump administration, we see them attack European tech legislation rather than adapt and cater to European industrial relations and social dialogue in the services they provide. It is naive to think that just by using European subsidiaries, the final say will not be had by the US headquarters. This difference in worldview clashes with the context of the European Social Model, where we need to have adaptable AM systems that worker representatives can negotiate on and can adapt to the needs of both the workers and the work process. Off the shelf, one-size-fits-all solutions have their limits and workers should demand a say on these technical aspects that have a material impact on their working conditions. One solution for this would be to develop homegrown, European tech solutions, where AI in the workplace tools become a strand of the effort to strive for more tech sovereignty.

Contrary to what is often heard in the current debate, this could boost productivity, since much research focuses on how "the tacit knowledge"[9] of the task at hand is key for useful digitising systems – not surveillance.

Rules and supervision have the potential to steer AM development

One way to help this development is to strengthen the legal framework that protects workers, also to give clear guidance on the kind of innovation we expect in Europe in this area. As Mariana Mazzucato argues, regulations should be seen as tools for directing economic activity towards societal goals by proactively shaping markets to achieve specific missions, like addressing climate change, or in this case improving working conditions in times of AM.[10]

The labour market and occupational health and safety are fields of strict regulation. For a large part, this is arranged at the level of EU member states, with a role for social partners to implement the protection given. These existing rules and safeguards need to be applied to the new situation. With an understanding that the AM tools are not developed to respect European workers' rights. The troves of data that are collected for AM purposes also make it possible to follow the letter, if not the in-

9 Acemoglu, D. (2023) *Power and Progress* (John Murray Press)
10 Mazzucato, M. (2013) *The Entrepreneurial State* (London: Anthem Press).

tention, of the law, for example, by using proxy data to act on to achieve a discriminatory outcome on illegal grounds. If the details of the algorithm and the data that lead to a decision are not disclosed to supervisors, it will be impossible to prove any wrongdoing. That is why the code should be made accessible – and in an understandable way – on request to government agencies.

AM is not the first wave of innovation and automation, and it does not require rewriting the whole basis of labour market regulation. It might require authorities like labour inspectorates to invest time and energy in new guidelines, but also to ensure that labour rights are respected, despite swift changes in the labour environment. Furthermore, we cannot depend on inspecting authorities to check all workplaces, so we need worker representatives and trade unions to actively negotiate the terms under which AM is implemented and play a role in motivating employers to keep the practices within the guardrails of occupational health and safety requirements.

This time the technology is different

It is clear that the AI revolution and autonomous AM systems bring new challenges and raise new questions that did not exist in previous waves of innovation. One difference with AI is that the system gets some form of agency; it becomes an actor that can take decisions that affect workers. While data has been used to motivate decisions, this data is transformed from passive information to an active component through AI. Amassing the data to build a file for firing a worker used to be a painstaking task for a manager, but is now something that can be generated in seconds. There are obviously automatic decision-making processes with a significant effect on workers. Under Article 22 of the General Data Protection Regulation, workers (and other types of "data subjects" affected by such processes) have the right not to be subjected to such decisions.[11] However, the burden of proof here makes this right difficult to enforce in practice; thus, it follows that regulation is needed at both the software and data levels.

The recent transversal regulation of AI in Europe, the AI Act, also recognises the implications of these developments, by placing the workplace

11 Article 22 GDPR.

in the high-risk category. Within the act's risk-based approach, it requires precautions to be taken and a risk assessment to be made before deploying AI tools in an employment setting. What this practically means for the deployment of AM, we will have to see and will depend on the implementing acts and guidelines of the EU AI Office. These documents will be what, in reality, steer the concrete actions that companies need to undertake to comply with the rules.

European legislator is working on further guardrails

Given the AI Act is to cover all aspects of the risks of AI deployment in Europe, and that at the moment we see little appetite for further sectoral legislation, it brings some cause for worry. At the time of writing, we do however see some signs of a specific directive on AM coming from the European Commission, which is an initiative that would have great potential. Following recent achievements in regulating AI in Europe, we can be optimistic about the prospect of the EU taking a lead in also regulating the impact of AM on workers.

When drafting this additional legislation on AM, we can expect the EU legislator to take the Platform Work Directive as a basis. The recently approved rules for workers in the platform economy deal with the aspects of AM by providing rules on the use of workers' data, transparency, human oversight and review, plus a general limitation of worsening occupational health and safety through AM systems. All of this could equally apply to workers in traditional sectors, rather than being limited to defined platforms.

One aspect that is interesting in the draft regulation for platform workers is that worker representatives are allowed to consult an external expert. The cost will be borne by the employer, who will also assist them in analysing and negotiating on the content and consequences of the algorithms. A similar provision for all workers around Europe would give trade unions the means and tools to meaningfully assess the AM systems that are deployed in the companies where their members work. These kinds of provisions can push towards worker consultation, which would be a helpful move towards levelling the playing field between management and workers on these highly technical systems.

Europe's workers, industry and leaders must shape the AM revolution in a productive direction

The current development of AM is not on a compatible path with strengthening the European social model. For some of the reasons outlined in this chapter, such as the lack of control and influence over algorithmic and AI systems, it will be a challenge to rebalance the direction of travel. The fact that the current technological development is weakening the power balance of workers in favour of management will make it harder for workers to negotiate a more favourable outcome. This makes the swift adoption of AM across sectors in Europe worrying, because it is creating a new reality that will be hard to transform.

At the same time, Europe can seize an opportunity by developing a different kind of AI and AM; a more cooperative form of the technology that works in favour of both management and workers. Utilising worker input, creativity and knowledge of the production process to innovate and adapt has the most significant potential for real productivity gains. This would require that both European companies and their workers take control over the algorithms, and that the software produced for European workplaces is designed with this possibility of co-determination in mind. The chapter by Christina Colclough and Hannah Johnston outlines some of the necessary principles when in-house design is not feasible.

But sensible procurement and more in-house design of algorithms is still not enough, but must be complemented with regulation steering innovation in a desired direction. We cannot expect the Big Tech US developers to start making these products out of the goodness of their hearts. Rather, Silicon Valley has obvious stakes in current developments, and why we need European and in-house software developers to meet a demand that can be created by new regulatory requirements for AI use in the workplace.

European competitiveness on a global scale will depend on the way we shape our industry and our industrial relations in times of swift AI development and adoption. Seeing legislation that protects workers as red tape is missing the potential transformative and steering potential of regulation. It fits the European social model to adapt the potential of AM to support rather than undermine working conditions. In the end, we need both workers and management to come together to innovate; for this, a healthy workplace democracy is necessary. When the influence

over algorithms can be shared by more than a handful of American tech companies and their CEOs, this will increase productivity to the benefit of all Europeans.

Ulrika Björkstén

The algorithms of science journalism before and during the COVID-19 pandemic

Public service journalism is not part of the public sector. In fact, the fundamental idea behind public service journalism is that its activities are carried out by independent media organisations, and hence, does not include state-sanctioned or controlled media. And yet, or more correctly *because* of this, public service journalism is a cornerstone of the public sphere in liberal democracies. Free but publicly funded journalism was one way of addressing the ubiquitous impact of radio as a new medium in the 1920s and, of course later, that of television. In the age of algorithm-driven, click-based journalism, the relevance and need for independent journalism with secure funding has hardly diminished.

The idea behind public service journalism is to create media organisations that are independent of commercial media's need to constantly maximise their reach to acquire paying readers and/or generate revenue from advertisers. Thanks to its stable funding, public service media should be able to highlight even the news that people did not know they needed (or wanted) to know, rather than being forced to prioritise those stories that attract the most readers at any given moment. Well, that is the theoretical idea at least. And yet public service journalism also operates, at least in part, under the same laws as commercial mass media. If the content fails to attract a big enough audience, there is a risk of losing long-term trust, relevance and – not least – public willingness to fund it.

The craft of journalism is often described in terms of gut instinct. The point of focusing on a specific body part that is *not* the brain is of course to underline that journalism is not a science. Perhaps my lecturers felt it was particularly important to stress this on a course in journalism aimed at academics, and as a science journalist, I soon learnt to accept that colleagues on the news desk would occasionally repeat the point. The question, then, is what indeed *is* journalism? How does the gut inform us what news and stories to report, and what are they supposed to tell us about

how we should go about it? Perhaps it is actually a set of internalised algorithms that we use to sort through the stream of events to find what could potentially become journalistic stories. In this chapter, I draw on my own experience in science journalism, specifically as head of the science news desk at *Sveriges Radio* (Swedish Radio) during the COVID-19 pandemic, as an example of how the sorting algorithms used in journalism may be structured and how they shape the flow of news. I also reflect on the way in which the internalised algorithms of the profession – that journalistic gut instinct – are now, at least in part, being pushed aside by a completely different kind of algorithm, both in commercial media and in public service: click algorithms.

As we know, artificial intelligence (AI) has also made inroads into journalism in recent years. Or, perhaps, in the spirit of logic, we should call it artificial gut instinct (AGI)? Both public service and commercial media have, for several years already, been measuring the value of news in terms of numbers of "clicks". Simple algorithms can ensure that the news stories generating the most clicks (by being shared on social media, for example) are continually bumped up to the top, or to the most visible position on media organisations' websites. This is an easy way to generate more clicks. More advanced tools now also allow us to analyse how long a reader or listener stays with a text before moving on to the next news story or webpage. Where do they lose interest? Where do they switch off? These answers are now readily accessible. This has given us a new perspective on storytelling, and in particular, it has altered the tone of radio. The first few minutes of a programme in today's digital media landscape, where every radio programme also has to function as a podcast that people choose and listen to purely on its own merits, need to be very different from the slow introductions and jingles that were common less than a decade ago, when the radio would simply be on in the car or the kitchen, and the radio channel itself was the brand.

For science journalism, this shift has made news work even more challenging. Another old journalistic truth is that "the best news is 25% new and 75% old". The point is that the audience needs to have some prior understanding of the context to actually perceive the information as news, and even more so to find the information interesting and relevant. In all the years I worked at the science news desk of Swedish Radio, we fought a constant battle with the introductions to each programme. To highlight the news that we judged to be most interesting and relevant to the public, we often found ourselves needing to provide background information at

the same time as we presented the story itself. We were unable to assume that our listeners had the necessary background knowledge to grasp the relevance of the story. This is what sets science journalism apart from event journalism, and it is one of the reasons why I have personally always questioned whether the traditional news format is suitable for science journalism at all. The ability of science to complicate and question our view of reality can help make people wiser – and this should be the goal of science journalism. Turning the latest results in a long research series into simple headlines without context simply doesn't do the trick. To manage this contradiction, science journalists have developed their very own sorting algorithms; their own version of a collective gut instinct. There is a tradition of a freer approach to news in science journalism, compared to the more externally driven event journalism. Science journalists are generally used to setting their own agendas. Rather than placing itself at the heart of the news flow, good science journalism alternates between two positions: one is well ahead of the general news flow, and the other is slightly behind it. On one hand, science journalism depicts the development of knowledge long before it is influencing people's everyday lives, while, on the other, there is a reflective side to science journalism that, in the aftermath of news events, seeks out scientific knowledge to provide a deeper understanding of events and developments.

Before the COVID-19 pandemic, in this manner, it was easy to define the sorting mechanisms deployed by science journalists using two tracks; two ways of sorting what deserved to become a news story. Science journalism's own algorithms, if you will.

The first sorting track involves following scientific progress and reporting on discoveries and new fields of research that may have real significance only in a few years' time – or maybe never. It involves keeping the audience up to date with the research front. In this track, we report on new, often uncertain and always – in some sense – provisional findings. After all, a new scientific discovery is always provisional as it awaits follow-up, even when it has been peer-reviewed and published in a scientific journal. The challenge for science journalists in this case is to present the news in a way that sparks interest, but without overstating the importance of the individual study being reported, while also making it clear why this particular study, of all the possible studies that we never report on, deserves attention.

The second track, which dominates in feature programmes and longer articles, focuses on highlighting scientific knowledge that can, in dif-

ferent ways, deepen our understanding of various events in society. It might involve how to cure various diseases or how to deal with energy supply and climate change with the help of both new technologies and innovations in the social sciences; it might involve educational research or neuroscience to shed light on education issues from a variety of angles, or sociology and criminology to provide new insights into gang crime. In this case, science journalism involves identifying relevant fields of knowledge and integrating them to shed light on a given, non-scientific issue. In other words, in this track, the news itself is something that has happened or is happening in society, and the science we seek out to help us deepen our understanding does not necessarily have to be new. On the contrary, it is an advantage if the findings we report on in this track are not too recent, but have had time to mature into reliable scientific knowledge.

So, the two sorting mechanisms in science journalism can be summarised as follows:

1) **Starting point**: the agenda of science itself.
   The role of the science journalist: to follow the flow of peer-reviewed articles.
   Outcome: new, often provisional research findings are presented.
2) **Starting point**: the agenda of society.
   The role of the science journalist: to search for scientific commentary on the general news flow.
   Outcome: knowledge that has matured is presented in a new context.

The ability to strike the right note in these two tracks, to make news and programmes that audiences perceive as interesting and relevant, can of course also be described as "gut instinct". The cognitive algorithms I describe should be viewed as an attempt to elucidate the sorting mechanisms that have evolved as an unspoken agreement within the profession – the collective gut instinct.

The question now is whether news media's algorithms of *artificial* gut instinct actually accept science journalism's alternative sorting methods. What does the audience miss out on if they do not, but instead systematically push the most relevant scientific reporting further down in media organisations' digital feeds? More specifically, the question I want to ask in this chapter is whether our experiences from the pandemic can help science journalism to better adapt its internal algorithms to the demands of artificial gut instinct.

The pandemic brought an immediate and major shift in the way we worked in the science newsroom at Swedish Radio. I do not primarily refer to the fact that, like so many others, we suddenly switched our working methods from meeting every day in person for editorial discussion into remote work; all of us alone, connected only by screens. Above all, the very content and method of our journalism changed.

In simple terms, you could say that the two established sorting tracks of science journalism merged. Moreover, the pace of scientific knowledge production became so rapid that the provisional nature of new findings became even more prominent. All of a sudden, we had to deal with a rapid flow of new, highly provisional information that was available in scientific reports that had not yet been peer-reviewed – or, in some cases, findings that had not been formally documented at all – dealing with the dominant societal issue of the day: just how dangerous was COVID-19, and how should we protect ourselves, as a society and as individuals? Our new way of working could be summarised in a third, new, sorting algorithm:

3) **Starting point**: to convey any knowledge that can help us understand the dominant societal issue of the moment.
   The role of the science journalist: to follow a rapid flow of potentially interesting but incomplete research findings and assess their reliability.
   Outcome: provisional research findings are presented in a constantly shifting context.

The pandemic thus meant that science journalism suddenly found itself in the middle of the news flow and became pure news journalism. We had to adopt a new way of working where we were constantly referencing new – even brand new or incomplete – science that had not yet been reviewed by other researchers. The very studies that our previous sorting algorithms would have ruled out were now, by necessity, brought into the news flow by our new overarching algorithm. Moreover, science journalism suddenly found itself at the centre of attention of our colleagues at the digital commissioning desk, who are tasked with feeding journalism's *AGI* algorithms. Remember – the algorithms of artificial gut instinct have little to do with sorting the relevance of incoming news; it is all about steering the output of journalism: to maximize the digital impact of the news.

Two days after the pandemic was declared, on Friday 13 March, the science newsroom – in line with this logic – had been commissioned

to launch an entirely new journalistic format: a daily short-form podcast focusing on the new virus. The starting point was to answer the public's questions, regardless of whether any clear answers were available. Our role was to use our science journalism expertise to search for the best answers that were available at any given time, including all the uncertainties. Our very first short podcast addressed the big question of the moment – what role did children play in the spread of the disease, and would closing schools help? A total of 73 short podcasts were produced, with the highest production rate during the first three months of the pandemic. At the outset, we were expected to produce one 8-minute podcast every day. That specific duration was based on what the gut instinct algorithms had shown to be the optimal listening time. In this way, we invented a new question-driven news journalism. Instead of basing our reporting on the weight of the answers, as we normally would when deciding which scientific reports qualify as news, we now did exactly the opposite. Every day, we discussed what would be the most interesting question to ask about the pandemic – and the answers were allowed to be just as certain, vague or diverse as they could be on that specific day. The challenge in this regard was to be as transparent as possible about the degrees of certainty and uncertainty in available knowledge, rather than asking those questions that were the most likely to produce definitive answers. That is why we often returned to the same question over the course of the year during which we worked with this short podcast format (which eventually tapered off to a couple of episodes per week as listener interest waned). As knowledge about the new virus grew, so did the answers we found.

Another change in our working methods was driven by how the authorities changed their way of communicating with the public. Directly after the outbreak of the pandemic, the Public Health Agency of Sweden, together with other relevant agencies, such as the Swedish Civil Contingencies Agency (MSB) and the National Board of Health and Welfare, began holding a series of daily press briefings. One thing that surprised us – and most other Swedish journalists, I believe – was how the press briefing soon became a kind of campfire for the whole country; the sheer number of Swedes who watched or listened to the press briefings every day. Journalists are used to using press briefings as opportunities to gather information, which we then process, check with other sources and then report on to the public. It is also worth noting that in Sweden, there is a tradition of individual journalists saving their "best" questions for

one-on-one interviews. You do not want to ask your smartest questions right under the noses of your competitors: instead, you save them for individual interviews after the press briefing in the hope of securing exclusives containing newsworthy responses. But we quickly realised that the press briefings had a much larger audience and impact than our own programmes and news articles. As a public service newsroom, our most important job was to deliver answers to relevant journalistic questions to as large an audience as possible. Therefore, we chose to abandon the notion that a news story had to be broken initially in our own broadcasts, and instead began to treat the press briefings as a journalistic arena in its own right.

A new part of our workflow soon involved holding an editorial meeting every day, well ahead of the scheduled press briefing, to discuss the most relevant questions to ask. These might relate to current developments, but increasingly we began to use the press briefings to publicly ask questions based on our own research into the pandemic: precisely the kind of questions that, according to earlier journalistic logic, we would never have dreamed of asking in front of our competitors. Although this meant that other media outlets were sometimes the first to report on the basis of our research, we reached more listeners as a result, and the science newsroom of Swedish Radio was often quoted by other media, which gave us an acknowledgement for our work.

This new approach made an impact when the Swedish Institute for Media Studies (*Institutet för mediestudier*) reviewed what degree of independence Swedish journalists displayed during the authorities' press briefings. According to this analysis, the science newsroom of Swedish Radio was the one news outlet with national coverage that asked the highest proportion of critical questions. Almost 40% of our questions fell into this category, according to the Swedish Institute for Media Studies. Local media were roughly at about the same level in their critical question, followed by "foreign media" in third place. Other national media was far behind.

In other words, the pandemic compelled our science newsroom to invent two new science journalism formats, both of which were question-driven. First the daily short podcast, and then the use of the authorities' press briefings as a platform for journalistic output. From a journalistic perspective, the former was far less controversial than the latter. The podcast was a direct result of the listener behaviour analysis, which is now routinely included when new programme formats are launched, guided

by media organisations' algorithm-based feedback loops. The form and length of the format was tailored to reach as many listeners as possible with the most topical issue at any given time. However, the impact of the authorities' press briefings came as a surprise to newsrooms all over Sweden. By investing journalistic effort into our participation there, the science newsroom of Swedish Radio achieved impact in various ways – at the press briefings themselves, in other media that quoted us and, eventually, in the form of a positive evaluation by the Institute for Media Studies. However, this impact did not generate any direct algorithmic reward in terms of Swedish Radio's artificial gut instinct, as it did not measurably contribute to bringing new listeners to our own digital platforms. At the same time, the decision to prioritise the press briefings was also "data-informed": it was based on knowledge of the large number of listeners and viewers that followed the press briefings, and the knowledge that a substantial percentage of the audience continued to listen or watch during the journalists' Q&A session. Whether our coverage of the press briefings can be regarded as algorithm-based is perhaps ultimately a matter of definition.

The COVID-19 pandemic has been associated with an "infodemic", meaning that there was a parallel world of information and disinformation on social media. The positioning adopted by many academics and researchers online meant that many potential experts became activists. As a result, there emerged a clear need for public service journalists to engage with and critically examine the claims made by independent or self-appointed experts. Helping the public to work out which expert knowledge is of relevance to a particular issue is always a key role for science journalism. This task demanded a great deal of time and effort during the COVID-19 pandemic; not least because of the impact of different interpretations of the pandemic that, regardless of their actual quality, could spread quickly in the new algorithm-driven media landscape. Besides analysing and commenting on official statements, science journalism took on the important role of examining the basis on which various experts, in their roles as activists, were making their claims. Disinformation and various types of pressure and smear campaigns on social media were also part of this reality that we as science journalists had not been accustomed to.

For the science journalism profession, being embedded in the algorithm-driven logic of news journalism turned out to be something quite new. It demanded innovative ways of working, very different from our tra-

ditional approach to set the agenda and shape the editorial algorithms ourselves. However, there is no doubt that the pandemic has strengthened science journalism in a way that appears to be permanent. A number of newsrooms have made major investments in new science journalism positions. Science journalism has matured as a field and become more integrated with news journalism, public affairs reporting and commentary. But if science journalism is to maintain this new position, it will have to continue to satisfy the digital algorithms of artificial gut instinct. At a time when disinformation has become a ubiquitous social problem and a tool in all kinds of warfare – both real and cultural – news journalism also has a great deal to learn from the methods and criteria for evaluating knowledge that are deployed in science journalism. Protecting the working practices and values that professional science journalism actually managed to uphold, even in the midst of the pandemic's infodemic, has never been more important. Above all, this involves safeguarding specialisation and in-depth knowledge among staff and editorial teams. Only in this way can journalism's collective gut instinct continue to serve the public by assessing and imparting the knowledge needed to enable each and every person to make informed decisions – on both personal and political levels.

Marcus Matteby

On a ministerial visit, institutional entrepreneurs and municipal perspectives

One morning in 2023, I received a visit from the Swedish Minister for Public Administration. I waited outside my office and watched him walking round the corner, just far enough away to give me time to feel a bit nervous before we greeted one another. What should I actually say to a Minister, and what would he say to me? I thought back to a news headline that had appeared a few months earlier: "Minister for Public Administration seeks magic wand for public sector digitalisation". The same newspaper claimed to know where it was, and that my employer's approach was the solution for Sweden. I have no idea why they put me at the top of their list of people with the most influence over the country's digitalisation strategy. But it was clear evidence that I was doing something right in trying to drive digital transformation in the toughest environment there is, and perhaps the last analogue stronghold in Sweden: municipalities.

We walked into my office, and the first thing the Minister saw was our development factory, the teams of system developers that make us unique. I explained how we have all our expertise under one roof, helping our municipality and all its companies to adapt and meet the expectations of business owners and residents. We needed this end-to-end approach because digital transformation requires control. Without it, the race towards digitalisation would be lost before the starting gun had fired, because the alternative – outsourcing combined with accelerating development – leads to lock-in, dependence on providers and erosion of in-house understanding of the very conditions that this kind of development relies on. But if you invest in control instead, the odds will be stacked in your favour.

In any case, my opening remarks for the meeting were well rehearsed. I started off with the most important thing of all. "What I want visitors to remember is the importance of trust. If we use trust as a foundation, stacks of digital solutions can be built – but without trust, they collapse."

This is not about me; it is about what we can create together if we trust one another. What the municipalities of Ånge and Sundsvall are doing has never been tried before and demands an extraordinary level of trust.

The local government sector stands at a crossroad that will shape society going forward. Perhaps for the first time ever, we have the opportunity to use artificial intelligence (AI) and digital transformation to create an equal society regardless of where people live, and to maintain our level of welfare while strengthening democracy, independence and privacy. If we ensure that algorithms are used correctly, with control over decision-making processes and full transparency for residents, it is possible to create the trust that underpins all successful transformation.

Unfortunately, many municipalities are heading in a completely different direction; a direction that, in the long run, will lead to reduced welfare levels[1] and weakened democracy as a direct impact of outsourcing algorithm design. Some people have made conscious choices, but most remain passive, waiting for guidance from the government or the Swedish Association of Local Authorities and Regions (SALAR; the municipal employer's organisation). Guidance that they are highly unlikely to receive. SALAR has chosen not to take action and define standards, but instead leaves such decisions to its members, all in the name of municipal self-government. But is self-government a sufficient reason for its 310 members to each invent their own solutions, when they have similar needs? This results in individual purchases that cost a fortune, with no real bargaining power against a handful of giant providers, and what we might, in all honesty, call pretty shoddy systems.

So why did the Minister for Public Administration visit the municipality of Sundsvall, and what was I hoping to gain from his visit? Well, he was curious to know how Sweden's leading municipality for digitalisation is transferring its success to another municipality by means of collective digital transformation. A digital merger between municipalities lays the foundation for equal access to welfare and a more secure life for residents. The state has long dreamed of seeing small municipalities giving up and merging with others: something that has certainly happened in the past. But according to researchers at the Ministry of Finance, this is unlikely to happen again, at least not by choice.[2] Could democratic nor-

1 Magnusson, J., E. Bragsjö, E. Rådinger et al. (2021) "Sveriges kommuners digitaliseringsstrategier". Digital Government Research Consortium.

2 „Steg mot stärkt kapacitet". SOU 2024:6. Regerinskansliet.

malisation initiated by members be the way forward for equal access to welfare? When asked what the government could do better, I pointed to the lack of central government control. However, at the same time, this is a complicated balancing act. The lack of central government control also allows something to emerge from the ground up, something that is based on trust and solidarity rather than control. Change happens on the periphery, when "the centre cannot hold", as the poet Yeats put it.[3] AI will reveal how municipalities and regions all too often relinquish control over the data of residents and businesses. Charlotta Kronblad presents a highly topical example from another municipality in this book. Sovereignty is a worryingly rare concept when public organisations procure digital solutions, and what I see is a mix of ignorance and unmanageable business models. The consequences are already concerning; but with the rise of algorithms, the headaches will become unbearable – and difficult to cure.

Control over data

Does having control over your data really matter all that much? Cash-strapped local governments should at least listen to the argument that it could save them a lot of money in managing their cities and municipalities.[4] Algorithms play a crucial role in tasks such as machine learning, where they are used to train models to recognise patterns and make predictions or decisions based on this very data. Algorithms are also used in areas such as natural language processing, computer vision and many other sub-areas, all of which are already widely used in the public sector. Access to data is vital for the use of algorithms and AI, and its quality, quantity and relevance have a direct impact on the quality and efficiency of automated decision-making.

Unfortunately, access to data presents a major challenge for Sweden's municipalities and regions. The root cause is what we in Sundsvall call a systemic flaw in the municipal sector, where the digital development of 290 municipalities is largely controlled by an oligopoly of private sector system providers; in some cases, with no competition at all. This places

3 Yeats, W. B. (1920) „The Second Coming". *The Dial*.

4 „Slutrapport: Uppdrag att främja offentlig förvaltnings förmåga att använda artificiell intelligens". DIGG, 23 January 2023.

significant limits on the digital legacy, with the data normally locked in or restricted to systems that are specific to a particular area of activity, such as education or planning permission. Each area of activity usually has just one or two system providers, all of them more or less sharing the same lock-in factor; combined with a lack of structure and standardisation, this causes a persistent headache that many municipalities find difficult to remedy. Of course, there will be individual solutions where many parties have helped to build up enough general data over the years to be able to utilise AI as a support tool. However, these limited solutions are far from harnessing the potential of the new technology.

Many municipalities and regions hope to address the systemic flaw by improving the procurement process, where more guidance is given from Christina Colclough and Hannah Johnston in this book. However, an example from Sundsvall shows the challenges coming from procurement, even when done "by the book". We recently procured a new IT system to meet an urgent need to comply with new requirements under the Employment Protection Act (LAS), and duly included access to collected data and calls for open application interfaces (APIs) among the requirements.[5] Both are essential if we are to comply with the strategy on sustainable digital development that was agreed at the political level by the municipal council. These requirements may be regarded as rather rudimentary for a modern organisation. But imagine my surprise when the providers responded by saying "we could meet those requirements in two to three years, at the earliest". Not only were there no acceptable alternatives on the market, but the development time needed to meet our requirements was unreasonably long.

The way forward

So, how do we break free from the stranglehold created by decades of flawed procurement and non-existent development? How do we move forward in a way that creates truly sustainable long-term solutions, instead of continuing to spend taxpayers' money on the latest fad? If we want to avoid ending up in the same municipal blind alley time and time again, our work needs to start from the ground up, just like when we con-

5 Persson, P. (2021) „Sundsvalls kommuns API-strategi". Sundsvalls kommun.

struct buildings or railways. As in this case, the foundation is the national digital infrastructure, and all digital solutions must then rest on this foundation. If we apply a number of common principles when we develop this infrastructure, it can also be shared between municipalities and become a national asset for the public sector. The core principles of this infrastructure must then be designed to enable:

- control over data, as I have argued for in this chapter;
- the open standards that so many providers are reluctant to follow, such as the APIs referred to above; and
- "public money, public code" for transparency, in line with the principle of public access to information, that is, open-source code should be the default in public administration.

If we can achieve this, we will contribute to both increased legal certainty and a stronger democracy, as direct effects of an open national digital infrastructure. We will also encourage actors to unleash their creativity in ways that we cannot predict. If, as developers, we follow these principles, not only will we create the conditions that will allow us to benefit from the potential of AI going forward, but we will also create shareable, transparent solutions with insights into every decision. It is vital that we actively raise awareness about AI solutions at an early stage to effectively promote a fair and equal welfare system. For several hundred local governments with similar needs and the same mandate, it goes without saying that we should capitalise on the expertise, data and successes from every one of us.

The fact that municipalities and regions were virtually incapable of producing software was not regarded as a particularly major issue – until algorithms began to be used. IT systems were procured that more or less met the needs of the services provided, despite some vocal criticism at times.[6] However, developments in AI are shifting the narrative, and we are facing challenges that can only be resolved through radical change. I am convinced that this contributes to the growing sense of powerlessness that many people feel in relation to these issues throughout much of the welfare sector.

6 Söderström, J. (2022) *Jävla skitsystem! Hur en usel digital arbetsmiljö stressar oss på jobbet – och hur vi kan ta tillbaka kontrollen* (Stockholm: Karneval förlag).

Take back control of system development!

Algorithms are rapidly replacing decision-making processes. Unfortunately, this is happening in the traditional way, and often with a lack of competition because of the systemic flaws in the municipal sector's provider market. In general, each area of activity has just one or two providers, which results in little or no competition.[7] The pace at which IT systems are upgraded and AI is implemented is glacial when there is no competition, and hence, no real need for development. At the same time, the market's system providers are living with their own digital legacy in the form of products that were developed to digitise paper-based processes within organisational silos. All of a sudden, these legacies are now expected to act as digital process tools in an ecosystem filled with other legacies. That was never their intended purpose, and it will not be possible to adapt these systems to meet our current demands.

IT systems typically lack integration capabilities via APIs, which means that data remains out of reach of both processes and IT systems. This, in turn, leads to lock-in effects and entrenches organisational silos, which makes it impossible to scale up systems. It may sound abstract, but there are examples where data from healthcare processes that ended up in these types of "silos" has resulted in patient deaths. One such example involves the transfer of patients between municipal wards and regional care. There is obviously a need here for a cross-cutting system where data can move between different levels of the organisation, but the market does not offer such a solution. Can a single provider be tasked with providing the entire solution? No. And one of many telling municipal examples of why this approach fails is the total collapse of the City of Stockholm's *Skolplattformen* system, which Marcin de Kaminski also discusses in his contribution to this book. The conclusion is simple: local government need to completely rethink its approach to developing these kinds of digital solutions.

One common concern is the economic viability of in-house development. It costs money to maintain control over the development and lifecycle management of digital solutions. Developing a specific solution in house for an individual administration may well be more expensive than a procurement procedure – initially. But when the solution is scaled up to meet other needs, the marginal cost becomes very low, almost zero.

7 Persson, P. (2024) „Toward citizen-centered digital government: Design principles guided legacy system renewal in a Swedish municipality". ScholarSpace, 3 January.

This approach creates a kind of reverse multiplier, a form of leverage that leads to positive outcomes: the more actors involved, the greater the benefit for everyone. That is why there is a need for a shared and open digital infrastructure.

As a field, AI is currently riding high on the hype curve, and so, there is a significant risk that individual prestige-driven projects will overshadow the potential for creating sustainable solutions. By developing partnerships and tools that give all stakeholders the same opportunity to benefit from AI, we ensure that the benefits can be replicated by new users and that the leverage from our efforts is brought to bear across all 290 municipalities and 21 regions.

These algorithms need to have certain key attributes and principles:

- open – anyone should be able to inspect and purchase them;
- modular – the needs of different customers/partners may vary in scale; and
- flexible and simple – the needs of different customers/partners may vary in nature.

How to harness the positive effects of algorithms

To begin with, it is necessary to ensure that the platforms providing AI services in a public setting are based on a common standard. No matter which public authority a citizen interacts with, they should be able to expect the same level of transparency. No one denies that AI holds enormous potential to benefit the public sector, but with that said, there is a risk that the market's business models will create even greater lock-in effects, further reducing transparency. The leading AI solutions today are black boxes, which offer absolutely no insights into how decisions are made. However, this need not be the case if we succeed in steering development towards the use of a more open AI infrastructure that provides transparent algorithmic decisions.

The public sector should therefore aim to achieve full transparency with regard to the exact factors that form the basis of the conclusions drawn by AI services. Every interaction with an AI service should generate data that describes the specific conditions on which the automated conclusion is based.

Provided below is an example of general minimum information requirements for each response or decision:

- the question asked or the task entered in the AI service;
- the prompt submitted with the question;
- settings in the AI service (e.g., temperature; that is, how "creative" you want the language model to be);
- the embedding model used to transform the question or task into a vector;
- the source identified by the model as containing the answer;
- the dataset identified by the model as containing the answer;
- the language learning model (LLM) used to translate the answer; and
- the final conclusion translated by the language model into an answer.

The following is an example of how a decision made by an AI service could appear when exported, giving the end-user full insight into what formed the basis of the decision made by the AI service at the exact time the decision was made. Each response is published in an open format, and to build trust in AI, every end-user interacting with AI services in the public sector should also be able to download the decision-making data in an open format.

Here is a table showing an example of a data package from the AI service:[8]

LLM	SOLAR-0-70b-16bi (https://huggingface.co/upstage/SOLAR-0-70b-16bit)
LLM settings	Temperature=0.4
	multilingual-e5-large (https://huggingface.co/intfloat/multilingual-e5-large)
Question	What are the opening hours at *Solskenets förskola* (preschool)?
Prompt	You work for the municipality of Sundsvall as an assistant and guide to all residents and visitors to the municipality. You help them by answering their questions based on the information to which you have access. You always use the information provided to you and never deviate from it.

8 „Transparens i beslut som fattas av en AI-tjänst". Sundsvalls kommun, 27 September 2023.

Data source	sundsvall.se/solskenet
Data	Solskenets förskola Adressvägen 123 123 45 Sundsvall Opens: 06:00 Closes: 18:00
LLM translation	The opening hours of *Solskenets förskola* are 6 am to 6 pm.

To summarise, there are obvious risks of a democratic deficit when several hundred local governments have to create their own standards and procurement procedures for the algorithmic decision-making systems we now face. In the absence of national coordination, there is only one option for responsible municipal administration: open-source code with modular and flexible solutions shaped by an API approach. These solutions should be shareable to promote greater equality, in line with economic common sense. Division is cheaper than multiplication; by which I mean that an essentially common solution can be shared by all. The concept of openness involves not only sharing, but also transparency and democracy.[9] That is what I told the Swedish Minister for Public Administration in the autumn of 2023. Hopefully, when the next Minister for Public Administration visits, Sundsvall will no longer be an outlier.

9 Koponen, J. (2022) „Manifest för en hållbar digital utveckling i kommunsektorn". Sundsvalls kommun.

Ivana Bartoletti

Fair AI: Utopia, aspiration or deception? The human cost of automated efficiency

Artificial intelligence (AI) is advancing at an unprecedented pace, and its influence across society is profound. In healthcare, for instance, AI is reshaping diagnostics, revolutionising treatment protocols and enhancing patient care, all while streamlining drug discovery and improving operational efficiencies. The promise of reduced costs and smarter systems makes AI compelling not only in medicine but across nearly every sector, where it is optimising administrative tasks, helping professionals work more effectively and improving the quality of daily services. The potential of AI to transform public services is especially significant.

By automating repetitive processes, AI can make government systems faster, cheaper and more accessible. Better use of data allows agencies to identify those most in need of support, targeting assistance more accurately and, in theory, reducing waste and inefficiency. For policymakers struggling with limited budgets and growing demand, AI holds out the hope of doing more with less.

Yet, history warns us that every technological revolution is a double-edged sword, and this is even more true for AI. Left unchecked, AI in public services could just as easily harm as help, and many chapters in this book points to the challenges we face. Algorithms trained on incomplete or biased datasets can replicate, and even magnify, social inequalities. Welfare or housing allocation systems may inadvertently penalise marginalised groups, reinforcing gender, racial or socio-economic disparities rather than correcting them. And because public services often involve society's most vulnerable, such errors and biases carry especially heavy consequences, eroding both trust in institutions and faith in fairness.

Cultural narratives about AI reflect this duality. In popular culture, we often see AI depicted through a troubling lens, particularly in gendered representations. From the seductive fembots in influential films like *Blade Runner* and *Ex Machina* to the discontented computer-wife in *SpongeBob*

SquarePants, these portrayals can perpetuate a narrative of victimhood and entrapment surrounding technology.[1]

These stories matter because they reflect and shape our anxieties about control, autonomy and power. More importantly, they remind us that technology is never neutral; it mirrors the societies that create it.

The risks are already visible in the digital realm. A global survey[2] of more than 8,000 young women and girls across 180 countries found that over half had experienced online abuse, a sobering illustration of how digital technologies can become breeding grounds for misogyny, harassment and violence. Similarly, tools like location-sharing apps, designed for safety and convenience, have also enabled stalking, surveillance and coercion. These patterns reveal how easily innovations meant to protect or empower can be co-opted to harm.

AI systems, by drawing on biased or incomplete training data, risk hardcoding such inequities into the very fabric of public decision-making. One study,[3] for example, showed that biased outcomes in algorithmic recruitment elicited less moral outrage than identical biases displayed by humans, suggesting a troubling lower standard of accountability when discrimination comes from a machine. In the public sector, this tendency could normalise unfair practices in welfare, healthcare or education systems, all under the guise of efficiency and objectivity.

These dangers intersect with labour-market dynamics, too. Because women disproportionately perform routine and administrative tasks, they face higher risks of job displacement through automation, compounding structural inequalities already present in society.[4] Thus, questions of equity and justice are inseparable from discussions about AI's role in government.

So, what does this mean? Does it mean we should step back and shy away from technology? Absolutely not.

Technologies are not themselves emancipatory, and AI perhaps even less, as it is built on existing data. But if technologies are accompanied by an emancipatory project, they can have the exact reverse effect.

---

1 Watercutter, A. (2015) „*Ex Machina* has a serious fembot problem". Wired, 9 April.

2 Nelson, S. (2020) „We need a safer online world for women and girls". Web Foundation Blog, 9 April.

3 Bigman, Y. E., D. Wilson, M. N. Arnestad et al. (2023) „Algorithmic discrimination causes less moral outrage than human discrimination". *Journal of Experimental Psychology*, 1(152): 4-27.

4 McDonald, C. (2024) „Women face greater risk of job displacement from automation". Computer Weekly, 29 August.

This chapter delves into the essential criteria for acceptable AI use within the public sector, underscoring the necessity for the EU to embrace these technologies responsibly, leveraging them to drive progress while safeguarding against their underlying risks.

The issue of fairness

Algorithms are transforming the way decisions are made, analysing vast quantities of data faster and often more reliably than humans. They already shape many aspects of our daily lives: from the adverts we see and the products we are offered, to the routes our GPS recommends and the news stories that surface in our feeds. As societies increasingly rely on algorithmic systems, understanding how they work – and where they fail – becomes essential.

At the heart of the concern lies **algorithmic bias**. Bias occurs when systems inadvertently incorporate the prejudices of their designers or the skew of their training data. Far from being harmless glitches, these biases can produce real-world harm. Women, particularly those from minority backgrounds, may be denied loans, while facial and speech recognition tools may misinterpret or fail to register people with darker skin tones or certain accents. As scholar Sofiya Noble describes,[5] this kind of bias isn't just a minor hiccup in an otherwise fair system: it is a systemic issue that is deeply embedded in the technology that drives search engines and digital platforms.

Importantly, bias does not arise only from data. It can creep in at every stage of an algorithm's life cycle: from design choices to implementation and everyday use. This complexity means tackling bias requires more than cleaning datasets – it demands oversight, accountability and a commitment to equity at every level of development and deployment.

Awareness of these issues is growing. Public debate and academic research have pushed "algorithmic discrimination" into the mainstream. A 2022 survey found[6] that over one third of companies had faced chal-

5 See: Noble, S. (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press); V. Ceia, B. Nothwehr and L. Wagner (2021) "Gender and technology: A rights-based and intersectional analysis of key trends". Oxfam Research Backgrounder.

6 Davis, J. (2022) „The cost of AI bias: Lower revenue, lost customers". Information Week, 19 January.

lenges or direct consequences linked to AI bias, ranging from financial losses and legal costs to reputational damage and loss of customer trust. In the private sector, these risks are taken increasingly seriously. In the public sector, where the stakes are often higher, they must be confronted urgently.

Two recent cases illustrate the consequences of unchecked bias in government services, adding to the example provided by Johan Hirschfeldt that opens this book:

- **Germany**: At the State Office of Transportation in Hamburg, a woman's administrative application was disrupted when the required biometric system failed to recognise her face. While officials denied a technical fault, local staff confirmed that such errors disproportionately affected people with darker skin tones, a stark reminder of how facial recognition technology can perpetuate discrimination.[7]
- **Austria**: The *AMS Algorithm* was introduced to profile jobseekers and predict their chances of reintegration into the labour market. While designed to improve efficiency and reduce costs, the system drew heavy criticism for disadvantaging women and older jobseekers. A later study found that while counselling processes became more "efficient", the model entrenched inequalities; it channelled funding toward the "middle group" of jobseekers, while offering little protection against structural discrimination. Strikingly, the system's design had included virtually no procedures to identify or mitigate bias.[8]

These examples underscore the stakes when fairness is sidelined. In each case, citizens suffered tangible harm: exclusion from services; wrongful accusations; or diminished opportunities. More troublingly, individuals often had no clear way to understand, challenge or appeal the decisions made by these systems. The opacity of algorithmic governance leaves people vulnerable to injustice, while shielding institutions behind claims of technological objectivity.

7 For more on this, see: J. Wulf (2022) „Automated decision-making systems and discrimination: Understanding causes, recognizing cases, supporting those affected". AlgorithmWatch, June, p. 8. See also: J. Buolamwini and T. Gebru (2018) „Gender shades: Intersectional accuracy disparities in commercial gender classification". *Proceedings of Machine Learning Research*, 81: 77-91.

8 „Governing algorithms: Perils and powers of AI in the public sector". Digital Future Society, May 2021. available at Governing_algorithms.pdf

The lesson is clear: algorithmic bias in public services is not a marginal issue but a systemic risk. It erodes trust, deepens inequality and undermines democratic legitimacy. Unless fairness and accountability are embedded into AI systems from the outset, we risk sliding into what some call "**algocracy**" – rule by algorithm – where opaque systems wield power over people's lives without transparency or recourse.

But can bias be "resolved"?

Developing and deploying ethical AI has become a political priority, particularly in Europe. The EU has positioned itself as a global leader, embedding fairness, accountability and human rights into its AI strategy. Similar sentiments were briefly echoed in the USA, where former President Joe Biden introduced a blueprint for an "AI Bill of Rights". Yet, political shifts soon followed: under the Trump administration, equity and fairness in technology slipped down the national agenda.

Still, AI does not operate in a legal void. Algorithms used to hire or fire workers, allocate welfare benefits, or determine loan eligibility remain subject to existing legal standards rooted in privacy, non-discrimination and fundamental rights. This principle is often overlooked in the rush to scale up AI deployment. While it remains uncertain whether today's legal frameworks can fully address the harms posed by algorithmic decision-making, given that traditional grounds for discrimination do not always align with the complexities of algorithmic bias, the case for comprehensive oversight has never been stronger.

A striking example comes from Amsterdam's *Smart Check* initiative.[9] The system was designed to evaluate welfare applications, detect errors and flag potential fraud. City officials believed they could build a tool that balanced efficiency with fairness, and they invested significant time and resources into doing so. *Smart Check* was even piloted on live applications. Yet, despite adherence to emerging best practices, the system

9 A full account of the Smart Check case is available here: E. Guo, G. Geiger and J.-C. Braun (2025) „Inside Amsterdam's high-stakes experiment to create fair welfare AI". MIT Technology Review, 11 June. It is important to note, and positively welcomed, that "in response to a public records request, the city disclosed multiple versions of the Smart Check algorithm and data on how it evaluated real-world welfare applicants, offering unique insight into whether, under the best possible conditions, algorithmic systems can deliver on their ambitious promises".

failed: its outputs could not guarantee fairness, nor did they prove effective. Ultimately, the project was abandoned, underlining just how difficult it is to create AI systems that are both equitable and reliable.

Why did it fail? Part of the difficulty lies in the very concept of fairness. Dozens of mathematical models and frameworks exist to measure it, but many are mutually incompatible. An algorithm deemed "fair" under one definition may violate another. This tension was evident in Amsterdam: attempts to embed fairness into design instead produced discriminatory outcomes in practice.

This raises a critical question: can such systems ever be deployed responsibly in real-world contexts where the stakes are high and outcomes deeply affect people's lives? Human decisionmakers, of course, are far from neutral. They display their own prejudices and blind spots. But this reality does not excuse AI from scrutiny. On the contrary, algorithmic systems demand *greater* accountability and transparency, precisely because those affected often have no way to understand, challenge or appeal decisions made by machines.

As Raphaële Xenidis and I argued in a report for the Council of Europe,[10] algorithmic discrimination introduces a set of distinct challenges. Machine-supported decisions are made at a scale far greater than human judgment alone. Bias can be hidden in complex interactions between data and design, making sources of discrimination hard to pinpoint. "Cleaning" datasets is both technically challenging and context-dependent, while proxies for protected characteristics – such as postcode standing in for race – make bias harder to eliminate. Predictive systems are especially problematic: they draw on historical inequalities (like the gender pay gap) as though they were causal, creating feedback loops that perpetuate disadvantage. At the same time, opacity in AI systems makes them difficult to explain, while responsibility for discrimination is often diffused or denied.

The crucial point is that because the roots of these biases are not purely technological, they cannot be resolved by technology alone. Tackling algorithmic discrimination requires more than better models or cleaner datasets; it requires political will, institutional accountability and

10 „Study on the impact of artificial intelligence systems, their potential for promoting equality, including gender equality, and the risks they may cause in relation to non-discrimination". Council of Europe, August 2023.

a proactive commitment to preventing structural inequalities from being reinforced through data.

In light of these challenges, we must ask what criteria should govern the use of AI in public services, and how can governments ensure that fairness and justice are upheld in practice?

A way forward

The EU's AI Act delineates specific criteria for high-risk AI, highlighting its potential impact on fundamental rights, including access to essential services. This designation brings a framework of controls surrounding AI systems employed in significant decision-making scenarios. While this emphasis on oversight is vital, I argue it does not address the core challenges we face.

Firstly, it's essential to recognise that the EU AI Act is not a catch-all solution for AI regulations – rather, it serves as a preliminary framework governing the entry of AI products into the European market. Pre-existing equality and privacy laws do remain at the forefront, especially for high-risk AI applications, as they are crucial to ensuring fairness in automated decision-making.

This is why one must question the fundamental fairness of these technologies. The notable opposition to the EU's AI Act raises eyebrows, especially since it doesn't introduce many new requirements beyond those already established in existing legislation. While the bureaucratic processes to ensure compliance may seem daunting for some stakeholders, advancements in technological tools can substantially alleviate these administrative burdens.

The pressing issue, therefore, is the inherent fairness of the AI tools themselves. The recent discontinuation of the Smart Check system in Amsterdam highlights the dilemmas inherent in the deployment of AI in public service. This outcome is arguably regressive for both camps, those advocating against any AI usage in critical decision-making and those who champion AI's potential to enhance transparency and fairness. Activists suggest that, particularly in contexts affecting individuals' lives, such as social services, public sector reliance on AI should be curtailed.

Nevertheless, Amsterdam's intentions were grounded in a desire to strike a balance. The city's efforts aimed to renew public trust: "We've

learned from the things that happened before us", emphasising a commitment to fairness and ethical governance.

Ignoring the setbacks this initiative faced serves no one's interests. Instead, it is crucial to dissect what went awry and cultivate lessons from this experience. Such challenges are not isolated; they are issues that every government and local authority must confront as they navigate the complexities of AI integration.

To this end, it is vital for EU governments, alongside local administrations, to promote best practices, conduct thorough impact analyses and prioritise transparent information sharing. Cities must open their algorithms and methodologies to public scrutiny; this transparency is essential for rebuilding trust and ensuring the ethical usage of AI in public services. By fostering an environment of accountability and collaboration, we can pave the way for responsible AI deployment that honours our commitment to fairness, equity and the fundamental rights of all citizens.

Karim Jebari

Artificial intelligence and public justice

Being male or living in a low-income neighbourhood may be statistically associated with a higher risk of committing crime. But basing a public decision, such as denying parole, on such grounds would be a form of discrimination. Moreover, withholding the reasons for such decisions from the person concerned and their lawyer would cause outrage.

The example may seem absurd: it runs completely counter to our idea of justice and the public exercise of power. But we may be moving towards this reality, if current trends persist. These days, decisions are increasingly being made with the support of sophisticated machine learning (ML) algorithms. In some cases, decisions are made solely by the tool. Recently, ML has also begun to be used by public authorities and other public decisionmakers.

ML is a type of artificial intelligence (AI) that has made major advances of late, both technically and in terms of usage. The use of ML is now widespread in software that adopts certain criteria to recommend one or more options from a larger dataset. Google Search and YouTube, for instance, rely on ML-based recommendation algorithms to suggest search results and videos. ML is a kind of AI, but it differs from earlier forms of AI, such as expert systems. Such early AI systems consist of two main parts: a database of structured information and a set of logical rules ("if condition A is not met, then the person is not entitled to measure B"). One advantage of this older generation of AI is its relative transparency. The rules are coded by programmers and can be regarded as codified laws and practices. The example given in Charlotta Kronblad's text of the school placement algorithm would fit in this category. ML, by contrast, builds a mathematical model for decision-making based on what are known as training data.

The most common form of ML technique is "supervised learning". Here, the algorithm is presented with labelled training data. This might, for instance, involve a number of facts about a person and information

indicating that the person failed to repay a mortgage.[1] For each object, the algorithm learns to associate a certain pattern of characteristics with a particular category, such as "good mortgage customer" and "bad mortgage customer". After being shown a large number of objects, the algorithm is capable of classifying objects that are not included in the dataset used for training.[2] This means that ML categorises an object according to whether it shares characteristics with other objects that have already been categorised. For example, a person may be categorised as having a "high risk of reoffending" if they share certain characteristics with other people who have reoffended. So, while expert systems only categorise individuals according to a number of clear rules, ML categorises people on the basis of statistical similarity.

There is significant potential for the use of ML in public decision-making, but its inherent properties pose a significant practical and normative challenge for decisions that should be subject to the principle of publicity. These include the principle of equality before the law, the principle of accountability for public decisionmakers and the principle of publicity. This chapter focuses on the latter.

Principles of the exercise of power

We should distinguish between the consequences of using AI in public decision-making processes and the consequences of using AI as a tool for private actors, such as companies and consumers. Although these cases share some problems and challenges, the use of AI in public decision-making poses a particular challenge: these users must follow principles of legitimate decision-making processes. I argue here that the democratic legitimacy of public decisions made or supported by ML risks being undermined.

By "public decision", I mean an act, or set of acts, that results in a citizen being treated in a certain way by public authorities. This may involve a person being deprived of liberty under the Act on Preventive Detention of Intoxicated Persons, a person receiving student finance or a person re-

1 Agrawal, A., J. Gans and A. Goldfarb, A. (2018) *Prediction Machines: The Simple Economics of Artificial Intelligence* (Brighton, MA: Harvard Business Review Press).

2 Russell, S. and P. Norvig, P. (2020) *Artificial Intelligence: A Modern Approach*, 4th ed. (London: Pearson).

newing their driving licence. In many cases, the final decision is preceded by a series of earlier decisions. To be eligible for student finance, for example, a student must have obtained a certain number of higher education credits in the previous semester. Whether or not a student passes is a decision made by the teaching faculty at a particular university. That decision is then used as a basis for the decision to grant student finance. ML algorithms can be used for all these decisions, in theory, or for some of them.

According to some of the most influential democratic theories, citizens have a fundamental interest not only in being treated fairly, but also in *seeing how justice is done*. This idea is central to what makes democracy a legitimate form of government. In the literature, this idea is referred to as "the principle of publicity" (which is distinct from "the principle of public access") and is a cornerstone of both national and European legal and constitutional frameworks. Public decision-making processes should follow the principle of publicity, which has two requirements:

1) Reason-giving: public decisions concerning citizens should be justified by reasons, which means that a decision should be based on certain objective and specific circumstances of the case in question, as well as rules, laws and regulations.
2) Accessibility: this means that the reasons should be accessible to both the party affected directly and other relevant stakeholders. Accessibility means that it should be possible to access and comprehend the decisions.

The characteristics of ML mean that decisions made using this technology may be incompatible with both of these requirements. ML gives reasons that are "statistical", and therefore, not specific to the individual case; nor can the reasons be said to be accessible, as the functions of ML cannot be fully explained.

Note that the principle of publicity should not be confused with the requirement for transparency in public decisions. According to this view, publicity is not the same as state transparency. Publicity does not mean that political and legal institutions are transparent across the board. Full transparency is not always desirable. Confidentiality may be important in some cases, such as when children are defendants in court cases. Transparency refers to whether information is publicly available. The principle of publicity, on the other hand, concerns the relationship between citizens and public authorities. Its purpose is to protect the interests of citizens as individuals subject to the law.

Challenges of ML in public decision-making

The principle of publicity requires decisions not only to be fair, but to be made in the right way. If a judge were to find someone guilty on the toss of a coin, this would not be acceptable even if the accused did actually commit the crime. Part of what it means to make a decision in the right way is that it must be possible to justify it with the right kind of reasons. This involves making decisions based on relevant facts in the individual case and based on the will of the democratic assembly, as expressed in democratically enacted rules and laws. When the problem of non-transparency is combined with the problem of reason-giving, ML appears to pose a unique challenge to the possibility of legitimate exercise of public power.

Decision-making systems based on ML algorithms may conflict with the requirements for the right kind of reasons. Firstly, to be correct, a reason must concern the specific case in relation to laws, rules and procedures. For instance, a motorist may be fined for speeding if the traffic police are able to prove that the recorded speed of the car exceeded the speed limit. Penalising a motorist simply for driving a model of car that is statistically associated with speeding would mean violating the reason-giving component of the principle of publicity. As ML typically categorises an object specifically on the basis of statistical similarity, this assessment – if applied unquestioningly by a public official – could violate the reason-giving component of the principle of publicity. If a court were to use an algorithm to assess a defendant's risk of reoffending, that assessment would be based not on that specific individual's risk of reoffending, but on the individual's statistical similarity to people who have committed new offences and been arrested for them. This has occurred in a number of US states, where an algorithm known as COMPAS makes a statistical assessment on the basis of variables that are not related to the individual's own actions. Moreover, the algorithm does not take into account actual reoffending, but only whether the individuals in question have reoffended *and* been convicted. This means that if there is a statistically significant difference between people who commit new crimes but manage to avoid the legal system and people who do not, this difference will affect who is deemed to be at "high risk" of reoffending.

Secondly, a decision can only be justified by reasons that are regarded as relevant in a specific case. The example in the introduction

discussed how low-income neighbourhoods may be statistically associated with a higher risk of committing crime, but that basing a public decision – such as denying parole – on such grounds would be a form of discrimination. This poses a general challenge to the democratic legitimacy of the use of ML. It also risks further penalising already disadvantaged groups.

Of course, the risk of public decisions not being based on reasons relevant to the specific case is not unique to decisions supported by ML. Public officials may make decisions out of habit without taking into account the specific details of the case in question. However, we ought to distinguish between a state where individuals fail to follow democratic principles, and when a state where the systematic procedure for making decisions is contrary to such principles. When ML is used in public decisions, it is more akin to the latter rather than the former.

Thirdly, when ML is part of the decision-making process, these problems may be exacerbated due to the lack of clarity surrounding how ML works. ML systems sometimes make decisions in ways that differ fundamentally from human decision-making. Human decision-making is characterised by what the literature calls "graceful degradation". This means that when human judgement is impaired, the deterioration in the quality of the outcome is proportional to the severity of the failure; unlike in the case of typical ML systems, where even a minor failure can cause total breakdown. This is particularly worrying in systems involving critical life-or-death decisions, such as decisions about which patients with pneumonia should be admitted to hospital and which can be sent home. Another aspect that makes ML systems difficult for humans to understand is the fact that failures in ML systems are sometimes the result of data patterns that are not apparent to a human observer. For example, adding a few pixels (which are invisible to the human eye) to an image can dramatically alter the ML system's ability to identify an object. For that reason, ML systems can be said to be psychologically non-transparent, which means it is difficult for a human to intuitively understand how an ML system makes decisions.

Fourthly, many algorithms are owned by companies and are regarded as trade secrets, thereby rendering them inaccessible for scrutiny by relevant stakeholders. This means that even elements of a decision that could be public, such as input data, are deliberately withheld on occasion. So, in certain cases, ML is legally non-transparent: its characteristics make it difficult for the general public to access information about

a given algorithm, the data used and how the algorithm was applied in a particular decision-making process. One example of this is how politicians in Gothenburg chose not to be transparent about how they had used an algorithm to place students in the city's schools; see Charlotta Kronblad's contribution to this book.

Fifthly, there are other ways in which ML algorithms are non-transparent. *Observational* non-transparency makes it difficult to know whether an algorithm will be successful in a particular case, even if it has performed well on training data. *Theoretical* non-transparency renders it impossible to explain how an algorithm reached a decision at each step. *Sociological* non-transparency means that the general public lacks the knowledge necessary to assess algorithmic decision-making.

This is related to the extent to which ML actually replaces human activity, which is reliant on the institutional practices and norms of decision-making. The lack of transparency in ML algorithms may hinder the ability to monitor, understand and explain cases in which ML is used in public decision-making. Therefore, tasks that are not designed to be performed by an automated process may inadvertently be performed by that very process.

Humans may fail to intervene and monitor machines. The vehicle manufacturer Tesla, for instance, has a self-driving system that has been involved in a number of high-profile accidents. These have often been caused by drivers behaving as though the self-driving system were fully autonomous; by watching films while driving, for example. Overestimating algorithmic capability is particularly worrying when algorithms are introduced with the expectation that they will reduce the costs of public administration. That ambition may, in turn, result in increased workloads for public officials, which may make it impossible for a human decisionmaker to make independent decisions, forcing them instead to delegate decision-making to the algorithm. That is why we should consider the actual use of a particular software package, and not just the intended use.

ML is powerful technology with the potential to revolutionise many aspects of society, from public planning to logistics and the analysis of social trends. Its ability to analyse and interpret large volumes of data can lead to significant improvements in the efficiency of resource allocation across society, particularly with regard to public services. Generative AI already appears to have helped to increase productivity in a variety of sectors. But for these benefits to be realised legitimately and sustainably

in the public sector, public authorities must use the technology responsibly and in accordance with the law and the principles that distinguish democratic states governed by the rule of law from authoritarian mafia states. This means safeguarding the principle of publicity and fairness when algorithms are used.

Stefan Larsson

Flexibly unpredictable? The AI Act as a regulator of a moving target

Introduction: On flexibility in technology regulation

Is technology moving too fast? And what happens when legislation seeks to regulate a moving target? The tension between transformative technological innovation and regulation has long been the subject of both political and academic interest, often positioned in a tug of war between legal fixity and flexibility. It is sometimes expressed as a "pacing problem", where technological development is said to be fast and the law too slow.[1] This idea has gained renewed momentum with respect to the European regulation of artificial intelligence (AI), as formulated in the recently established AI Act.[2] But this is a false dichotomy: even a field as dynamic as AI requires regulation capable of ensuring that certain socially relevant risks are addressed. In fact, as with much standardisation and unified assessment of anything from aeroplane safety to communication protocols, adequate regulation tends to drive technological adoption. And normative guardrails have, in most jurisdictions, been developed over time to enable a much-needed balancing of interests in anything from media distribution to labour markets and democratic institutions, all of which are increasingly affected by AI deployment in various ways. That said, it is important to understand the impact and consequences of a *flexible* variant of regulation that struggles to capture the very definition of the main object of regulation, here AI; perhaps especially in relation to

1 For a detailed analysis of the AI Act from this perspective, see S. Larsson, J. Hildén and K. Söderlund (forthcoming) „Implications of regulating a moving target: Between fixity and flexibility in the EU AI Act". *Law, Innovation and Technology*.

2 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

the large language models (LLMs) that have come to extend beyond both text and language, and beyond any specific domain of use. It is worth recalling that what we now term *generative AI* did not exist as established terminology when the European Commission first published its proposal for the AI Act in April 2021. Now, however, the focus of AI discourse is specifically on generative AI (although the AI Act uses the term *general-purpose AI* as the main terminology for this), as more general AI models are being used to generate high-quality text; images; software code and, increasingly, videos on the basis of relatively simple instructions (prompts) from users. The need to regulate what is known as *agentic AI* is also an emerging discussion, that is, the risks and ethical issues linked to the development and use of relatively autonomous AI systems,[3] including the issue of whether the AI Act adequately addresses this evolving use of AI.[4] Overall, the field is dominated by global tech giants because of the need for large data volumes, (energy-consuming) processing power and cutting-edge computer science expertise; this also emphasises the geopolitical perspective embedded in the relationship between the USA, China and Europe. The combination of rapid AI development and the drafting of EU-wide legislation that aims to impose the same requirements on all member states provides an interesting interaction between policy and technology, which invites further analysis and reflection.

Purpose of the chapter

This chapter has its origins in a Swedish version but has been updated through studies conducted as part of a Nordic research project on the 0governance of AI and automated decision-making in the public sector,[5] and within the framework of a research cluster on the AI welfare state.[6]

3 Gabriel, I., A. Manzini, G. Keeling et al. (2024) „The ethics of advanced AI assistants". arXiv:2404.16244

4 Oueslati, A. and R. Staes-Polet (2025) «Ahead of the curve: Governing AI agents under the EU AI Act». The Future Society, June.

5 *The Automated Administration: Governance of Automated Decision-Making in the Public Sector* (PI: Larsson). The project is funded by the research programme Future Challenges in the Nordics.

6 The cluster is coordinated by the universities of Södertörn, Lund and Karlstad in Sweden, and supported by the Wallenberg AI, Autonomous Systems and Software Program – Humanity and Society (WASP-HS).

The analysis aims to deepen our understanding of how European AI regulation is grappling with regulating a moving target as exceptional as AI development, and to show how these insights are relevant from a technology governance and supervisory perspective. In the chapter's conclusion, I highlight the need for structures for skills development, collaboration between public authorities, and active interpretation and communication of both existing legislation and future application of the AI Act. The emphasis here is on the flexible elements of the AI Act that defer certain normative issues by broadening the Commission's mandate, or by commissioning standards to be developed to assist regulatory compliance. Particular attention is paid to generative AI, as this is a fast-moving field of technology that has significantly influenced the legislative process.

AI governance as a pacing problem

In its simplest form, the pacing problem expresses the challenge posed by the difference in the pace of development between regulation and technological innovation, which has led to proposals for more adaptive ways of governance.[7] The idea of an ongoing pacing problem between governance and AI development was clearly expressed in the "moratorium letter" published in March 2023. The letter, which had over 31,000 signatures, called for a pause of at least six months in training the most powerful AI system (GPT-4) at the time the letter was written.[8] The signatories argued that developments were proceeding too quickly, with the risk of losing control over AI technology unless developments were deliberately slowed. Other researchers, oriented towards social-science-based research, have pointed to issues with bias and discrimination related to general AI systems, observing that stakeholders may avoid responsibility.[9] A third perspective was set out by a large group of companies concerned that the European

7 See: J. Herkert, G. Marchant and B. R. Allenby (eds) (2011) *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (Dordrecht: Springer). For an analysis of the dynamics in terms of "future-proofing legislation", see S. Ranchordás and M. van 't Schip (2020) "Future-proofing legislation for the digital age", in S. Ranchordás and Y. Roznai (eds) *Time, Law, and Change: An Interdisciplinary Study* (Oxford: Hart Publishing).

8 „Pause giant AI experiments: An open letter". Future of Life Institute, 22 March 2023.

9 „Five considerations to guide the regulation of 'general purpose AI' in the EU's AI Act". AI Now, 14 April 2023.

AI Act risks stifling innovation through its requirements.[10] The European legislator has attempted to position itself along similar lines of complexity; particularly in relation to the last two perspectives, recognising on one hand the need to regulate the risks of undesirable effects such as discriminatory AI systems, scalable disinformation campaigns and malfunctioning or overused surveillance technologies, and on the other hand trying to deal with the fast-moving landscape in terms of what technologies are being developed and what concepts are needed to describe them. The answer appears to have been an attempt to balance legal fixity with built-in flexibility. Both sides present challenges as regulatory strategies. On one hand, there is no certainty that the chosen path of relatively general product safety regulation is the most beneficial for Europe, where everything that can be called AI (there are also significant definitional issues)[11] is to be assessed according to a risk scale with varying degrees of requirements for evaluations before such systems are permitted to be placed on the market. On the other hand, flexible elements are accompanied by varying degrees of unpredictability, and ultimately – taken to the extreme – legal uncertainty.[12] The legal literature also emphasises any law's necessary clarity in its conceptual framework and the importance of moderation in amendments,[13] and hence, also predictability in terms of normative stability over time.[14] Essentially, legal fixity is necessary for imposing requirements on stakeholders, with fines and supervision for compliance, and for market stakeholders of various kinds to be able to adapt their efforts and strategies or simply get used to a new norm. Too much change is costly and confusing. However, some level of flexibility is needed because the legislator cannot know how the field will evolve in terms of innovation, or how markets will develop and be structured. Below, I elaborate in more detail what this flexibility entails.

10 Herijgers, L. (2023) „150 European companies write open letter on dangers of AI Act". *Techzine*, 30 June.

11 For the evolving nature of the AI concept over time, see S. Larsson (2021) „AI in the EU: Ethical guidelines as a governance tool", in A. Bakardjieva Engelbrekt, K. Leijon, A. Michalski et al. (eds) *The European Union and the Technology Shift* (Cham: Palgrave Macmillan); for the AI Act's definitional issues, see N. A. Smuha and K. Yeung (2024) "The European Union's AI Act: Beyond motherhood and apple pie?" DOI: 10.2139/ssrn.4874852

12 Larsson, S., J. Hildén and K. Söderlund (forthcoming) „Implications of regulating a moving target: Between fixity and flexibility in the EU AI Act".

13 Popelier, P. (2000) „Legal certainty and principles of proper law making". *European Journal of Law Reform*, 3(2): 321.

14 Super, D.A. (2010) „Against flexibility". *Cornell Law Review*, 6(96): 1375.

Technological leaps and legal struggles

Towards the end of 2022 and in the midst of the process of drafting the AI Act, the legislator was struck by the leap towards generative AI, as evidenced by the launch of ChatGPT, which has influenced its design. European AI legislation was first proposed by the European Commission in April 2021, following strategic work on ethical guidelines and a White Paper.[15] The Commission's proposal was followed by the proposals set out by the European Council in December 2022 and the European Parliament in June 2023. Inter-institutional negotiations took place in the autumn of 2023, which led to a provisional agreement on 9 December 2023. The final legislative text was adopted by EU institutions between February and April 2024. The legislation entered into force on 1 August 2024, with some parts applicable after just six months, and the remaining main parts applicable over the following 12, 24 and 36 months.

From AI ethics to AI law – a new policy wave

Soft governance, through various types of non-legislative principled or ethical guidelines, has been important and a hallmark of governance in the field of AI.[16] This is also true of Europe, most clearly expressed in the AI High-Level Expert Group appointed by the Commission in 2019, which published ethics guidelines for trustworthy AI that same year. Explicit reference to these guidelines is also made in Recital 7 of the AI Act. Developments in generative AI, with more general-purpose AI models, have led to new governance challenges and conflicts of interest that have prompted a number of stakeholders, from 2023 onwards, to develop policies and recommendations on how the field ought to be governed. The OECD, for instance, published a policy document in September 2023, in which it pointed to risks of misinformation, reproduction of bias – discriminatory structures, that is – and issues regarding copyright issues related to training data.[17] In June

15 S. Larsson (2021) „AI in the EU: Ethical guidelines as a governance tool"; N. A. Smuha and K. Yeung (2024) „The European Union's AI Act: Beyond motherhood and apple pie?"

16 See: S. Larsson (2020) „On the governance of artificial intelligence through ethics guidelines". *Asian Journal of Law and Society*, 3(7): 437.

17 „Initial policy considerations for generative artificial intelligence". OECD Artificial Intelligence Papers, 18 September 2023.

2023, the ACM Technology Policy Council highlighted the need for regulation and the introduction of safeguards, with a human in the loop, challenges concerning intellectual property ownership, data protection, and the need for opportunities and mechanisms to correct inaccuracies.[18] In the autumn of 2023, the Biden administration published a comprehensive executive order on safe, secure and trustworthy AI (which was revoked by the Trump administration in January 2025), while the G7 countries have published both AI guidelines and a code of conduct for AI providers. An international AI safety report was published under the leadership of Yoshua Bengio at the AI Action Summit in Paris in February.[19] The entire report focuses on general-purpose AI, with an emphasis on risks and how these can be mitigated. It is clear that the generation of LLMs and more generic AI models gives rise to a significant need for a better understanding of how best to regulate and govern the development and use of these systems. Before examining the EU AI regulation process in more detail (which has clearly been influenced by these developments), I should first clarify what aspects of AI development have taken a significant leap forward over the last few years.

Generative AI

A major breakthrough in language modelling came with what was known as the transformer architecture in 2017, which required far less training than other technologies to achieve unprecedented results in language translation. The increase in capacity was widely noted with Open AI's GPT-3, the third version of a *generative pre-trained transformer* (GPT), which was launched in June 2020. Globally, a race is underway amongst some of the largest technology stakeholders concerning the capabilities of LLMs. While ChatGPT – with its public launch in November 2022 – stunned the world in terms of how quickly it was adopted by millions of users and incorporated into Microsoft's Bing search engine through collaborations, this is merely the tip of an iceberg that also includes other language models, such as Meta's LLaMA; Google's Bard/Gemini

18 Jain, R., J. Matthews and A. Saucedo (2023) „Principles for the development, deployment, and use of generative AI technologies". ACM Technology Policy Council, 27 June.
19 Bengio, Y., S. Mindermann, D. Privitera et al. (2025) „International AI safety report". arXiv:2501.17805

and variants from Chinese tech giants such as Baidu; and, not least, the DeepSeek-R1 model, which sent shockwaves through the world's stock markets in January 2025 when it demonstrated its capabilities despite limited access to the hardware available to American developers. Video models such as OpenAI's Sora were unveiled in February 2024 and became available to a wider audience in December 2024. The multi-modal models (as they are known), which are capable of combining expression across images, audio and video, began to demonstrate their potential to a wider audience during 2024. As stated, agentic AI has increasingly come into focus; particularly in 2025, where method development and evolving autonomy in solving increasingly complex tasks is being driven by all the major model developers, such as Anthropic, Microsoft and Meta.[20]

The European AI Act: Risks and definitions

The process of developing a regulation as comprehensive as the European AI Act is therefore particularly interesting given the rapid pace of change in the field of AI. Some of the main flexible elements of the AI Act that can be viewed as a response to the pacing problem of AI include regulatory sandboxes,[21] standardisation processes and what are known as delegated acts. These will allow the Commission to propose amendments at a later stage with a view to clarifying or specifying certain ambiguities in the legislation.[22]

A product safety regulation

The AI Act is designed in line with the EU's *New Legislative Framework*. This means that it is a kind of product safety regulation, which is particularly tangible in the case of high-risk AI systems where European Con-

20 See: Oueslati, A. and R. Staes-Polet (2025) „Ahead of the curve: Governing AI agents under the EU AI Act".

21 For a comprehensive overview, see C. Novelli, P. Hacker, S. McDougall et al. (2025) „Getting regulatory sandboxes right: Design and governance under the AI Act". DOI: 10.2139/ssrn.5332161

22 Larsson, S., J. Hildén and K. Söderlund (forthcoming) „Implications of regulating a moving target: Between fixity and flexibility in the EU AI Act".

formity (CE) marking is a key element.[23] The categorisation of risk levels is a primary structure in the act for organising and allocating different requirements. The various risk levels are described as *unacceptable risk*, *high risk*, *limited risk* and *no risk*. AI systems that pose a danger to health, safety and fundamental rights are classified as *high-risk AI*. According to the act, all high-risk AI systems are required to be evaluated before being placed on the market and throughout their entire life cycle, which makes the high-risk category perhaps the most important part of the regulation. Whether or not an AI system is included in this category has significant implications. In the impact assessment that preceded the proposal for the AI Act in April 2021, the European Commission concluded that it was most appropriate to regulate AI technology *itself* in relation to high-risk applications, rather than specific behaviours, regardless of the technology.[24] As a result, the act has inevitably placed enormous importance on the definition of what AI actually *is*. To clarify, the definition is important in this context for at least two main reasons:

1. It determines the scope of the regulation. If the definition is very broad or general, the AI Act comes across as an almost generic technology regulation aimed at risk areas, which could result in a cumbersome document structure. Conversely, if it is very narrow and combined with costly requirements, there is a risk of what the literature on pacing problems refers to as *evasive entrepreneurship*, where avoiding regulation may become a clear corporate strategy.
2. This is a clear sign of the pacing problem, which is relevant to the phenomenon addressed here. If innovation is advancing so rapidly that the concepts to be formalised in the regulation need to be substantially reworked during the legislative process itself, this indicates that the regulation needs to find ways to allow flexibility while the law is used to fix the applicable regulatory framework.

The three proposals for an AI Act from the Commission, the Council and Parliament approached the definition of AI in slightly different ways. One aspect, important from a pacing perspective, lies in the Commission's ability to modify the definition in retrospect, thus adding further technologies to be classified as AI under the act through what are known

23 Smuha, N. A. and K. Yeung (2024) „The European Union's AI Act: Beyond motherhood and apple pie?"

24 „Impact assessment of the regulation on artificial intelligence". European Commission, 21 April 2021.

as delegated acts or implementing acts.[25] The possibility of granting the Commission powers to adopt "non-legislative acts" within the AI Act is intended to amend or supplement the *non-essential* parts of the legislation.

A technical leap in the midst of the legislative process

"Generative AI" was, as stated, not explicitly addressed in the Commission's proposal for the AI Act from April 2021. "Generative AI systems" are mentioned in the Council's subsequent version dating back to late 2022, but it was not until Parliament's version from June 2023 that the rapid development of LLMs was properly addressed through various requirements relating to "foundation models" and "general-purpose AI". The requirements were mainly concerned with transparency and a focus on data management, risk management, model evaluation, energy efficiency and quality control, documentation in a central EU database, and the provision of a detailed summary of the use of training data protected under copyright. According to one assessment, at that point, there were few providers who supplied any information about aspects such as the copyright status of training data (including OpenAI's GPT-4, Google's PaLM 2 and Meta's LLaMA), perhaps partly to obstruct claims from rights holders.[26] Providers also reported inconsistently on carbon emissions from training, or on their strategies for measuring emissions. The regulation of general AI models became a controversial element in the negotiations in the autumn of 2023, with France, Germany and Italy reportedly not wanting this to be included in the AI Act. Yet that is what happened (for "general-purpose AI", i.e., GPAI): some documentation requirements apply to all providers of GPAI models, while a tiered system imposes more stringent requirements on the most powerful providers; those that, according to the legislation, may pose a "systemic risk" due to adverse effects on health and safety, fundamental rights or society as a whole, which could propagate throughout the value chain. Of interest from the pacing perspective is the fact that the threshold for providers posing systemic risk

25 For a more comprehensive analysis, see S. Larsson, J. Hildén and K. Söderlund (forthcoming) „Implications of regulating a moving target: Between fixity and flexibility in the EU AI Act".

26 Bommasani, R., K. Klyman, D. Zhang et al. (2023) «Do foundation model providers comply with the draft EU AI Act?" Stanford University, June.

is determined by the total amount of computing power used to train the AI model: the threshold can be adjusted and other criteria may be developed at a later date; this further highlights the flexible and somewhat unpredictable way in which GPAI is being regulated.

Discussion: Regulating a moving target

Striking a reasonable balance quickly is difficult when it comes to technology regulation. Pace in itself presents a dilemma for both innovation and regulation. The approach adopted by the supervisory authorities at both member state and central European level will therefore be a crucial issue in the implementation of the AI Act.[27] Soft governance in the form of guidelines, often framed in terms of "ethics", has – as noted above – played a significant role in the development and use of AI in recent years. The field of general-purpose AI models that can be adapted to many different applications and allow for casual, everyday use by almost anyone, has prompted (sic) a new wave of identified policy needs with respect to the risks of spreading misinformation, bias/discrimination and the copyright status of training data.

The problem with nebulous phenomena: What is AI?

There is an inherent conceptual uncertainty in "AI" as a governance concept, since it tends to be used as an umbrella term for a host of rapidly evolving methods and technologies. This is applicable to "AI" itself, as well as to newer concepts such as "foundation model"; "generative AI"; and "general", "multi-purpose", "multi-modal" or "agentic" AI. This leads to legal challenges that have to balance, on one hand, principled demands for legal certainty in terms of conceptual clarity, predictability and the exercise of public authority in accordance with the law, aligning with the market's need for stable market conditions; and, on the other, an innovation landscape that is changing so rapidly that interpretive uncertainties and the need for more adaptive approaches arise.

27 Much of the supervision of high-risk AI falls to member states, whereas supervision of general-purpose AI falls to the newly established AI Office. See: K. Söderlund and S. Larsson (2024) „Enforcement design patterns in EU law: An analysis of the AI Act". *Digital Society*, 2(3): 41.

On power balances and skills

From a democratic perspective, it is problematic that the flexibility mechanisms used by the EU legislator risk shifting power balances away from democratically elected decisionmakers towards the Commission, technical committees and standardisation bodies. Furthermore, AI regulation gives rise to a need for more active interpretation of applicable law in light of the new AI-influenced practices among relevant supervisory authorities in member states. How well prepared they are for this is very likely to differ. A frequently noted pacing problem concerns information asymmetries between legislators and technology developers; in other words, it may be difficult for legislators to fully understand the new technologies themselves, let alone identify key regulatory needs when these technologies are introduced into society. Learning, therefore, is not just a matter for the innovator, but also for the public authority representing a particular area. It is likely that this collaborative dynamic with respect to AI issues will need to be strengthened *between* public authorities as well. Thus, the need for collaboration is also evident from a supervisory perspective, given the breadth of the field of AI, ranging from issues relating to discrimination and human rights to technically advanced matters potentially involving financial markets, cybersecurity, competition, consumers, medical and healthcare, and data protection issues. As no single expert or supervisory authority can be expected to grasp such a complex set of problems, the issue inevitably comes down to a need for coordination and cooperation – and interpretation.

On unpredictable flexibility

Essentially, the field of AI – viewed as a pacing problem between innovation and governance – does not lend itself to the simple notion of "speeding up legislation". This is a dynamic field that impacts power balances. It involves conflicting perspectives on governance under legal principles that relate to the market's need for predictability, while regulation must take an adaptive approach to what technology can do in human hands. The way in which the AI Act addresses this dynamic is to include flexible and partially unfinished elements to be clarified and amended later, within the framework of risk categorisation resembling product safety regulation. However, one hope with the AI Act is that harmonisation of

legislation in all EU countries will promote the beneficial use of these technologies. At the same time, how effectively the regulation works in practice will depend on how well it is interpreted and clarified by public authorities.

Christina Colclough and Hannah Johnston

Promoting digital sovereignty and fundamental rights: Six principles for inclusion in public service procurement contracts

Introduction

The relatively stable political and economic landscape that has characterised the European continent and European diplomatic relations since World War II has fallen into turbulent times. This became unmistakable in 2025 when – following threats from the USA that it would cut off military support to its longstanding allies – the EU bolstered its own defence spending with double-digit increases.[1] The European response has widely been viewed as a step towards ensuring Europe's sovereignty, but, while this particular episode garnered extensive media attention, it was only the most recent episode of a longer struggle over how power and control will be exerted between Europe and the USA. In recent years, this quest for control has largely concerned the digital sphere and questions of data rights and governance.

Under Ursula von der Leyen's current Commission, Europe's digital sovereignty – or its "ability to act independently in the digital world" is a key priority.[2] This notion is predicated on the idea that, rather than remaining under the spell of American hegemony and reliant on American corporations, Europe should determine for itself the governance of data, information flows and other digitally mediated activities – especially given the impact and transformative potential of digitalisation for the economy and society. This approach has garnered criticism from American com-

1 Psaropoulos, J. T. (2025) "Europe leads global defence spending rise, awakening to security deficit". Aljazeera, 28 April.

2 Velliet, M. (2023) „Digital sovereignty: European policies, American dilemmas". IFRI, 31 January.

panies, whose global dominance can be at least partially attributed to decades of minimal regulation and who provide billions of euros worth of digital services to European clients. Within Europe, while there is ample support for this proposal, this shift may prove difficult; even the European Commission has privately conceded that "when it comes to dominant players such as the US, 'decoupling is unrealistic and cooperation will remain significant across the technological value chain'".[3] Indeed, mere months after von der Leyen advocated for greater European investment in tech, the Commission awarded US-based Amazon Web Services a contract for €500 million to provide cloud computing services.[4] Amazon alone has now received over €1.3 billion of European public procurement contracts.[5]

Outsourcing threatens digital sovereignty

The frequency and value of public service digital service procurement contracts raises a fundamental question about digital sovereignty that reaches beyond geography: why – given the risks that outsourcing presents to digital sovereignty and public service integrity – are public institutions spending so much money on procurement contracts for digital services at all?

Public procurement within the EU has grown by 900% since 2009. Indeed, public procurement now totals €2 trillion annually, and accounts for over 14% of the EU's GDP.[6] Recent news stories about the UK's €380 million contracting of Palantir to develop a data platform for the National Health Service[7] and the aforementioned EU contracts with Amazon to provide cloud services confirm that the amount of public money spent on

3 Pollet, M. (2025) "EU views break from US as 'unrealistic' amid global tech race". Politico, 30 April.

4 „For Christmas, the Commission awarded Amazon a contract for cloud services worth half a billion euros". Parliamentary question E-001982/2025, 16 May 2025.

5 "As US-EU trade war escalates, UNI Europa calls for ending Amazon's access to EU public contracts". UNI Europa, 2 April 2025.

6 „Public procurement in the EU: Less competition for contracts awarded for works, goods and services in the 10 years up to 2021". European Court of Auditors, 2023; "'So long and thanks for all your taxes' How much Amazon makes from public contracts in Europe". UNI Europa.

7 Campbell, D (2023) "Patient privacy fears as US spy tech firm Palantir wins £330m NHS contract". *The Guardian*, 21 November.

private procurement contracts for digital tools and services is not negligible. Procurement has thus become an integral part of how many organisations, including public services, meet their digital needs, despite the fact that they – like a growing number of organisations – have come to rely on digital innovations, tools and strategies for even their most basic operations.

A path to mandating digital sovereignty and public service integrity

Within public services, the use of digital tools and technologies is accelerating, presenting both opportunities and challenges. Proponents of recent technological innovations, such as artificial intelligence (AI), big data and cloud computing, promise a more efficient and cost-effective delivery of public services, and these can be procured through multitudes of private, third-party developers and corporations. The focus of many critics, meanwhile, has been on the attendant risks of many of these technologies – for example, bias, safety and security, and data privacy.[8] Yet, in addition to the risks of the technologies themselves, the very notion of procuring digital services for public services is risky. Beyond digital sovereignty, procuring privately owned or developed digital tools raises concerns about public service quality, autonomy, workers' rights, and the protection of workers' and the public's fundamental "digital rights".[9] These concerns arise from the often opaque instructions and code embedded in digital systems, the risks associated with data extraction, and the increasing dependence on third-party control over data and data-driven technologies.

8 Noble, S. U. (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press); J. Buolamwini (2023) *Unmasking AI: My Mission to Protect What Is Human in a World of Machines* (New York: Random House); S. Viljoen (2024) "The broader lessons of privacy law. *Boston University Law Review*, 4(104): 1131-1149.

9 We draw on the EU's concept of „digital rights" to invoke the principle that technology should serve and benefit and empower people, and not infringe upon their fundamental rights or security, as a starting point. This requires putting people and their rights at the centre of the digital transformation; supporting solidarity and inclusion; ensuring freedom of choice online; fostering participation in the digital public space; increasing safety, security and empowerment of individuals (especially young people); and promoting the sustainability of the digital future.

If the call for greater digital sovereignty by European leaders is to be taken seriously, radically new measures are needed to strengthen the autonomy of European Public Services. While the most comprehensive way to ensure sovereignty and avoid the risks discussed above is to adequately fund public services and encourage in-house development of digital tools and services, this "solution" is not easily achieved on the heels of decades of outsourcing. Thus, European regulators and member states are faced with a choice of how to procure digital services and tools externally. They can either continue to prioritise a cost-minimisation strategy that is likely to result in a race to the bottom with regards to safeguarding rights, protecting the environment and ensuring decent work, or they can align procurement laws and practices with public interests, sustainable development and fundamental rights. This chapter argues for the latter.

Multiple paths to digital sovereignty

Addressing what ought to be mandated for inclusion in procurement contracts is timely: the EU is revising the EU's Public Procurement Directive and a proposed draft from the Commission is expected in 2025.[10] The most comprehensive way to ensure public sector autonomy, uphold the digital rights of the public and public service workers, and promote sustainable solutions would be to revise the directive to include mandatory non-pecuniary "quality-based award criteria" that affirm these values and principles. This approach would diverge from current public service procurement practices, which are predominately awarded on the basis of price alone.[11] While a revised directive along these lines would have the broadest impact and reach, it is not the only avenue to promote digital sovereignty through procurement.

Irrespective of the outcome of the directive revision process, the very nature of procurement offers opportunities for intervention. The public service, and public service workers and unions, for example, can work within their own spheres of influence to use social dialogue mechanisms to revise procurement processes or negotiate contract terms with exter-

10 "Public procurement directives – evaluation". European Commission.

11 Packroff, J. (2023) „Most public contracts still chosen based on price alone, study shows". Euractiv, 26 October.

nal providers that detail the conditions under which services or goods will be rendered. Although this approach may lead to a more fragmented regulatory landscape, it nonetheless presents opportunities to include measures that promote and uphold public service autonomy and digital rights by mandating their inclusion within industrial relations practices or individual procurement contracts.

What follows are six principles that promote quality-based award criteria that should be mandated in the revised Public Procurement Directive and can be further embedded within public services though workplace and supplier contracts. The inclusion of these principles ensures that the use of externally sourced digital technologies within public services is sustainable; it aligns with democratic values, workers' rights and public interests; and it promotes digital sovereignty.

Six principles to ensure digital rights through procurement award criteria

1 Transparency and explainability

The principles of transparency and explainability are cornerstones in the use of any digital system. Indeed, understanding how digital systems function is foundational if workers and the public are to secure other digital rights regarding how such systems are governed and controlled. Within the EU, transparency is a key principle of the General Data Protection Regulation (GDPR), requiring that data processors inform data subjects about what data is collected about them and how it is processed, but when it comes to procurement contracts, the principles of transparency and explainability should be extended.

Not only should the procurement process itself be transparent with respect to who, within and external to the public service, is involved; how bids are solicited and evaluated; and how service providers are selected – the principles of transparency and explainability should also be written into the contract that is rendered. Additionally, beyond GDPR requirements, transparency and explainability provisions required of service providers and outlined in procurement contracts should explain the purpose of the system and how the system will be deployed in the public service, as well as clearly detailing how the system will be overseen and which department or manager is responsible.

A key benefit of including these kinds of provisions in procurement contracts is that they promote trust in public services. Transparency – including plain and easy-to-understand information on the purpose of the system(s), the sources of data/input and how data is processed to generate predictions, content or recommendations – improves understanding of these systems. In turn, malfunctions (like discrimination and bias) can be more readily identified, and an ethical use of technology that aligns with the public interest is more easily achieved.

Transparency and explainability requirements cannot stand alone though. To be meaningful, they must be linked to principle 3 below on inclusive governance.

2 Environmental impact and fundamental rights impact assessments

As digital systems become more integrated into public services, a structured process must be developed to evaluate how a particular technology may affect fundamental rights and freedoms – such as the freedom of expression, privacy, non-discrimination and the right to effective remedy – of individuals. This practice is in the EU AI Act, article 27, referred to as a fundamental rights impact assessments (FRIAs). In addition, there must be a mandatory requirement to complete an environmental impact assessment (EIA), which evaluates how technologies affect the planet and its resources.

Assessments should be conducted prior to deploying technologies (ex-ante) and then on a regular basis by competent authorities to ensure that these rights are not impeded. When assessments are undertaken, they should be completed in cooperation with representatives of the subjects concerned (see principle 3 below). They should at all times be available to the public, including the preventive and remedial measures that will be taken to mitigate risks and harms identified, as well as the arrangements for internal governance and complaint mechanisms.

The rapid expansion of AI and cloud computing is powered by an ever-growing number of data centres. The International Energy Agency estimates that, from 2024 to 2030, data centre electricity consumption will grow by around 15% per year, more than four times faster than the growth

of total electricity consumption from all other sectors.[12] In Europe, it is estimated that data centres could account for up to 6,600 million cubic metres in water consumption in 2027.[13] Environmental impact accounting must therefore include calculations of the carbon footprint, water consumption and resource usage of digital systems in the short, medium and long term. Including this demand in public procurement will help ensure that digital tools procured for public services are sustainable and do not contribute to long-term ecological harm.

FRIAs and EIAs are not merely political fads, but pertinent necessities, and should cover the entire supply and value chain of digital systems. These provisions will additionally ensure that contractors who receive procurement contracts operate within ethical and environmental standards, reducing the risks of human rights abuses or environmental degradation.

3 Mandatory inclusive governance of digital systems

Public procurement policies must require that the digital systems procured are inclusively governed by providers and the public service. Inclusive governance means that representatives of those who are subjected to the systems (e.g., members of the public and/or workers) are included in governance processes. A comprehensive list of affected stakeholder parties can be identified via a thorough impact assessment.

Although inclusive governance requires time and resources, engaging in dialogue with affected parties in an effective way to improve the public service's understanding of digital systems; similarly, by responding to the needs and concerns of affected stakeholders, the public service can also improve its ability to transparently and responsibly deploy such systems. For example, inclusive governance may result in the introduction of restrictions on the types of data collected, what it can be used for and by whom (see further elaborations on this in principle 5 below for one example of possible restrictions).

Engagement should begin before a system is introduced and continue throughout its use, ensuring that the interests of all affected parties are

12 „Energy demand from AI". International Energy Agency, April 2025.

13 Illmavirta, J., P.-Z. Kow and S. K. Sahoo (2023)"Unique continuation for the momentum ray transform". arXiv:2304.00327

considered in the development and deployment phases. These kinds of contract provisions can ensure that humans remain the main drivers of decision-making. Inclusive governance is also an effective way to ensure that care and consideration is observed with respect to all parts of the data life cycle – from what data is collected, to how it is stored, analysed and accessed, to how it is disposed of.

Inclusive governance can be undertaken in a range of ways, including by prioritising companies with collective agreements as partners. In some contexts, works councils or a representative group of managers and employees can be effective governance partners. The public service should further extend participation to citizens to include them in the oversight and governance of digital systems. Inclusive governance builds trust, helps prevent abuses and ensures that the technology serves the public interest. It also helps ensure that digital systems are not introduced at the expense of workers' rights or of the public's interest.

4 Public service rights of access and control

Public services should operate in the public interest, and ensuring that this happens in the long term requires that the public service retains control over how public institutions function and how decisions are made. Although many digital systems are designed by external, third-party developers and rely partially on externally provided data, public services must still demand a right to access and control how these digital systems function. Not only must the public service understand how these systems function (principle 1), but they must also be able to access the data and programming that comprise digital systems to shape the system's functions and operations as needed. Access and control are vital for public services to be able to amend or alter algorithms if negative or unintended outcomes arise or to avoid potentially harmful situations.

Linked to this, language should be added in procurement contracts concerning joint data control between the public service and the private party(ies). This to ensure that public services retain control over the information that is required to serve in the interests of the public.

A growing number of AI-driven tools, such as automated hiring systems and large language models, operate with a high level of opacity. Thus, without demands for access and control over digital systems, the public service risks becoming dependent on external forces and private

tech companies. Ensuring that procurement contracts include language that ensures the rights of access and control over digital systems mitigates the risk that public services will develop a dependency on proprietary technologies, or concede decision-making autonomy to corporations.

System opacity is also a reason that unions should consider including provisions for the "reverse burden of proof". When workers or the public have concerns about how AI systems and digital tools function, they should be able to voice these concerns without fear of reprisal. The national partners can work together to develop an institutional mechanism to receive such concerns. Given the lack of transparency in many digital systems and tools, developers must provide access to any relevant data and programming and both they and the deploying public service should be required to prove that systems are functioning as intended and in respect of the rule of law. When unions or the public allege that digital systems have an adverse effect, the burden to prove otherwise should fall on the deployer – as occurs in some jurisdictions when employment practices lack transparency. It follows the principle that "where the facts and events at issue lie wholly, or in part, within the exclusive knowledge of the authorities or other respondent, the burden of proof should be regarded as resting on the authorities, or the other respondent, respectively".[14]

5 Prohibition of repurposing and/or selling data to third parties

Whilst the GDPR offers some sound data protection rights, it does not challenge the global business model of data trading and all of the potential rights violations stemming from that. Repurposing data or selling it to third parties means that data can end up being used by other entities in ways that may not serve the public interest, the public service or the workers. For this reason, procurement contracts should include stringent data management clauses. These should limit data use to the explicit reasons set out in the procurement contract, and the sale of data – or analyses generated from such data – to third parties should be prohibited. For

14 European Union Agency for Fundamental Rights and Council of Europe (2018) *Handbook on European Non-discrimination Law* (Luxembourg: Publications Office of the European Union), p. 239.

example, third parties should be prevented from using data derived from the procurement task to train their own large language models, or for any other means. This does more than promote the privacy of stakeholders: it ensures that the data generated by the public service remains within the public domain and cannot be exploited for private profit and that how it is used now, and in the future, is in the interest of the public good. Contract amendments can always be permitted if circumstances warrant it, but a good rule of thumb is to limit the use of data to parties that have been specifically licensed by the public service to use it.

Procurement policies should additionally require contractors to store the data they collect and manage as part of the procurement contract in Europe where European laws and regulations prevail. Existing rights under the GDPR normatively permit the movement of data only to jurisdictions where data protections are deemed to be "adequate". Given that GDPR provides greater data protections than many regions of the world, bilateral framework agreements – such as between the EU and the USA have been negotiated to ensure adequacy for the purpose of receiving EU data. However, recent geopolitical upheaval and the interruption and non-compliance with longstanding trade and other cooperation agreements cast doubt on the tenure of any negotiated deal, including one having to do with data protection.

6 Managerial and worker competencies

To actualise all of the above, procurement policies must include provisions related to ensuring that all relevant managers and workers have the competencies required to procure, understand, govern, use and independently analyse digital technologies. Building on the AI Act, article 4 on AI literacy, capacity building must be included as a required provision in procurement contracts. This will encourage accountability at all levels, lead to a safer and more ethical use of digital systems, and protect public service autonomy.

All managers involved in the procurement process, including defining the criteria for successful bidding, external party contract development and oversight, and using and analysing the outcomes of any procured digital system or technology must have the necessary competencies to be effective in this role. This is essential to protect public service autonomy, quality public services and ethical standards.

External developers of digital systems must partake in this competence building by disclosing information about the training data, system instructions, by disclosing their FRIA (see principle 2) and by engaging in the co-creation of the digital systems required with public services. It is the responsibility of the public service to ensure that all relevant managers have the competencies required to enable the realisation of the provisions in this document.

Workers can be subjects of digital systems that are developed as part of the procurement. For example, a cybersecurity system can monitor the workers' online activities. Workers can also be required to use a digital system as part of their job functions. They must therefore receive adequate training to ensure they understand the digital systems, know how to use them, can explain the outcomes to the public and can meaningfully participate in inclusive governance (see principle 2). With adequate competencies, workers will be able to protect their dignity, sovereignty over digital systems and their rights. They will also have the competencies to co-govern and suggest modifications to the systems, identify bias and other harms and will understand how to review automated decisions.

Conclusion

Public procurement has long been a method for implementing public policy across a wide array of fields,[15] and the procurement of digital tools and services is now embedded in public services, societies and labour markets. Amidst the contemporary challenges, Europe stands at a crossroads: it can either remain on the current procurement path that prioritises cost savings above all else, or it can take a bold step toward digital sovereignty that aligns procurement practices with public interest, sustainable development and the protection of fundamental rights. This chapter advocates for the latter and proposes six key principles to guide this shift.

The principles address several pressing issues. First and foremost, they promote an approach to digital sovereignty that is consistent with the core values of the European Social Model, where economic and so-

15 Grandia, J. and J. Meehan (2017) „Public procurement as a policy tool: Using procurement to reach desired outcomes in society". *International Journal of Public Sector Management*, 4(30): 302-309. DOI: 10.1108/IJPSM-03-2017-0066

cial progress are seen as inseparable. In a digitalised society and labour market, true social progress is intertwined with democratic participation and the safeguarding of fundamental rights. Our proposed measures ensure that Europe can make digital advances without sacrificing the very principles that underpin its democratic institutions.

Secondly, the principles ensure the autonomy and continued capacity of public services to uphold public interests, rather than conceding democratic control to the whims of commercial companies. By laying down stringent limitations on how third parties can capitalise on public service data and by demanding joint access to and control over digital systems, public services will regain control over the means of governance and retain their authority to make decisions that prioritise the welfare of citizens.

Finally, while proponents of digitising public services often tout productivity and efficiency gains as primary benefits, this narrative is incomplete without considering the social and environmental costs associated with the use of digital technologies. If these costs are not factored into the equation, the promises of digitalisation may not only be overly optimistic – they may be outright unsustainable.

Europe's path forward must ensure that its digital transformation is not driven by corporate interests but rather by a commitment to the public good. The principles in this chapter provide a blueprint for integrating sustainability, democracy and digital rights into public procurement, securing both the autonomy of public services and the protection of fundamental rights. By embedding these principles into the very fabric of procurement practices, Europe can shape a digital future that is as inclusive, transparent and sustainable as it is innovative.

About the authors

Dr. Anne Kaun is Professor of Media & Communication Studies Sweden and director of the Hub for Digital Welfare Research and the WASP-HS funded research cluster the AI Welfare State at Södertörn University, Stockholm. She is the co-author of Prison Media (2023 together with Fredrik Stiernstedt, MIT Press) and Data Welfare State (2025 together with Anu Masso, Sage). In 2021, she was named Wallenberg Academy Fellow funded by the Knut and Alice Wallenberg Foundation, and she is the co-host of MKV-podden – a podcast by and for media and communication scholars.

Dr. Charlotta Kronblad is a former lawyer with ten years of experience in the legal field, and holds a PhD from Chalmers University of Technology. She has completed a postdoc. at Stockholm School of Economics and is currently conducting research on digital transformation at the Department of Applied IT at the University of Gothenburg. Her research focuses on the intersection of law and technology, particularly the implementation of algorithmic decision-making in the public sector and its implications for social and legal justice, as well as issues related to cybersecurity and defence. Her work has been published in leading international journals. Charlotta is also a sought-after speaker on digital transformation in both the private and public sectors. In addition, she serves as a member of the Legal Expert Group at AI Sweden and is affiliated with the Swedish Center for Digital Innovation.

Dr. Christina J. Colclough is founder of the Why Not Lab – a boutique value-driven consultancy that equips workers and their unions across the world with the skills and tools to ensure collective rights in the digital age. She has a background in labour market research and in the global labour movement, where she led on future of work policies, advocacy and strategies for a number of years. Christina is a board member of the Partnership on AI and a fellow of the Royal Society of Arts in the UK. She is a member of the UNESCO #Women4EthicalAI Platform; the OECD.AI Network of Experts; and is affiliated to FAOS, the Employment Relations Research Center at Copenhagen University.

Gerard Rinse Oosterwijk is a political advisor to Alex Agius Saliba MEP, Vice President of the S&D Group in the European Parliament in charge of the Digital Agenda, communication and campaigns. Gerard is also a FEPS Expert Consultant and was FEPS Policy Analyst on Digital from 2023 to 2025. Before FEPS, Gerard served as a political secretary and director of the social-democratic Vooruit movement in the Brussels Parliament, as well as a senior policy advisor in the European Parliament for Paul Tang MEP. His experience in the private sector includes roles as an operational manager and marketeer at ING Bank, and he was co-founder of a political start-up. Gerard holds a master's in European social policy from the London School of Economics and an LLM in Private Law from Leiden University.

Dr. Hannah Johnston is a professor in the School of Human Resources Management at York University in Toronto. Her research focuses on the digitalisation of work and union responses to technological change. She has a PhD in geography from Queen's University (Canada) and was previously a postdoctoral fellow at Northeastern University in Boston. Internationally, Hannah has worked for the International Labour Organization and with trade unions and workers' organisations on issues related to collective organising, algorithmic management and technological change.

Ivana Bartoletti is an internationally recognised expert in privacy, artificial intelligence (AI) governance and responsible technology. She has held senior roles in both the private and public sectors, leading on innovation, digital transformation and the ethical adoption of AI. An expert for the Council of Europe, she co-authored a landmark study on the impact of AI on gender equality. Ivana is the author of An Artificial Revolution: On Power, Politics, and AI (Indigo Press) and A Digital Union: Based on European Values (FEPS), and co-editor of The AI Book (Wiley). She is a sought-after voice on technology, policy and ethics worldwide.

Johan Hirschfeldt is a former president of the Svea Court of Appeal in Stockholm. He has formally been Under-Secretary for Legal Affairs first at the Ministry of Employment and then at the Prime Minister's Office. He has also been Chancellor of Justice and a substitute member of the European Commission for Democracy through Law (the Venice Commission). Johan previously served as the Parliamentary Ombudsman and was a legal advisor at the National Tax Board and Under-Secretary for Legal Affairs at the Ministry of Employment.

Dr. Karim Jebari is a researcher at the Institute for Futures Studies. He defended his doctoral thesis in December 2014 at the Royal Institute of Technology (KTH). Karim is writing a book on the climate crisis and the future of humanity, and his interests range from the social impact of emerging technologies to long-term risks to human civilisation. He has published several articles in the field of existential risks, bioethics, the philosophy of AI and human enhancement.

Dr. Maja Fjaestad is an associate professor at the Royal Institute of Technology (KTH), affiliated with the AI Policy lab at Umeå University and employed as senior advisor to the president at Karolinska Institutet. Her research has revolved around the interplay between technology and society, as well as science and policy during crisis management. Maja was State Secretary for Health in Sweden during 2018-2022 and held the same position at the Prime Minister's office. She has strong international engagement and was among other commissions of trust appointed by the European Commission to its working groups on science diplomacy in 2024. She is also a member of the Royal Swedish Academy of Engineering Sciences.

Marcin de Kaminski is Vice President of Programs at the Open Tech Fund (OTF), leading efforts to advance internet freedom technologies, innovation and global human rights. He previously served as a policy specialist at the Swedish International Development Cooperation Agency (Sida), where he focused on information and communications technology, internet freedom, and organising the Stockholm Internet Forum. Most recently, he was Director of Security and Innovation at Civil Rights Defenders, overseeing protection programs and rapid response mechanisms for at-risk human rights defenders. With a background in internet research and social innovation, Marcin also served on OTF's Advisory Council, guiding support for the internet freedom community.

Marcus Matteby is an executive doctoral student and a member of the SCDI Digital Government Research School. He has a long history in IT and AI management and currently serves as CIO/CDO for the municipalities of Ånge and Sundsvall, known for their pioneering digital merger as well as their use of open-source software and open data protocols. In 2024, Marcus was named CIO of the Year, and under his leadership Sundsvall received the award for Digital Municipality of the Year. He has also served as a delegate in the Swedish AI Commission, a government committee

promoting the use and development of artificial intelligence. His research focuses on the collective digital transformation, with a particular emphasis on leadership and strategies that enable enhanced digital transformation in the public sector, primarily within local government.

Dr. Maris Männiste is a lecturer on critical data studies at the University of Tartu. Previously, she worked as a postdoctoral researcher at Södertörn University, where she was part of the CHANSE funded project Automating Welfare – Algorithmic Infrastructures for Human Flourishing in Europe (AUTO-WELF). In that role, she contributed to comparative research on automated decision-making systems in Sweden and Estonia, and co-authored studies on public sector chatbots and citizens' attitudes toward automation in welfare provision. Her research focuses on the intersection of critical data/algorithm studies, media and communication, and public administration.

Simon Vinge is Chief Economist at the Swedish trade union Akademikerförbundet SSR, with a focus on artificial intelligence and algorithms. He holds a master's degree in economics from the Stockholm School of Economics and previously worked at the Ministry of Finance, the Prime Minister's Office and Kommunal. Simon is currently the Swedish representative on the Economic Committee of the European Trade Union Confederation, and he also works on issues related to global taxation and welfare policy.

Dr. Stefan Larsson is a senior lecturer and Associate Professor in Technology and Social Change at Lund University, Sweden, Department of Technology and Society. He is a lawyer and socio-legal researcher with a PhD in the sociology of law as well as a PhD in spatial planning. He leads a multidisciplinary research group on AI and Society that focuses on social and normative implications of artificial intelligence and adaptive technologies in both the private and public domains, ranging from public sector decision-making to mammography and social robotics.

Dr. Ulrika Björkstén is Secretary-General of Public & Science Sweden, a non-governmental organisation that works to promote dialogue between academia and the general public. Ulrika has a background in solar cell technology research, with a PhD in physical chemistry from the Swiss Federal Institute of Technology at Lausanne (EPFL), and has worked for almost three decades as a science writer and producer, including more

than ten years as editor in chief of the science newsroom at Swedish Radio. Alongside her journalistic work, she has a particular interest in science and technology studies and spent a year as a visiting fellow at Cambridge University writing about the interaction between genetic research and wider cultural and societal issues. In 2023, she was awarded an honorary doctorate from the Faculty of Arts at Uppsala University.

About the FEPS-Nordic Digital Programme

This book is published as part of the FEPS-Nordic Digital Programme, a collaboration between the Foundation of European Progressive Studies (FEPS) and its members and trade unions from Denmark, Sweden, Norway and Finland. The 2025-2026 Programme focuses on the impact of AI on labour productivity and the public sector, as well as trade union capacity building in relation to algorithmic management and other aspects of workplace digitalisation.

Members of the FEPS-Nordic Digital Programme 2025 - 2026 Consortium are:

FEPS
FOUNDATION FOR EUROPEAN PROGRESSIVE STUDIES

Foundation for European Progressive Studies (FEPS)
Avenue des Arts, 46 – 1000 Brussels, Belgium
www.feps-europe.eu
@FEPS_Europe

Friedrich Ebert Stiftung

Friedrich-Ebert-Stiftung Nordic Countries (FES)
Barnhusgatan 10 – 111 23 Stockholm, Sweden
https://nordics.fes.de/
@FES_Nordics

Kalevi Sorsa Foundation

Kalevi Sorsa Foundation
Siltasaarenkatu 18-20 C, 6. Krs – 00530 Helsinki, Finland
https://www.sorsafoundation.fi/
@SorsaFoundation

Akademikerförbundet SSR

Akademikerförbundet SSR
Box 128 00 - 112 96 Stockholm, Sweden
https://akademssr.se/
@akademssr

The Economic Council of the Labour Movement (Arbejderbevægelsens Erhvervsråd - AE)
Reventlowsgade 14, 1. Sal – 1651 Copenhagen, Denmark
https://www.ae.dk
@tAEnketank

The Swedish Trade Union Confederation (Landsorganisationen i Sverige - LO Sweden)
Barnhusgatan 18 – 105 53 Stockholm, Sweden
https://www.lo.se/
@LOSverige

The Swedish Commercial Employees' Union (Handelsanställdas Förbund - Handels)
Sveavägen 90 – 111 81 Stockholm, Sweden
https://www.handels.se/
@handelsfacket

Negotia
PB 9187 Grønland - 0134 Oslo, Norway
https://www.negotia.no/
@Negotia.no

Supporting unions of the FEPS-Nordic Digital Programme 2025-2026 are:

Service Union United (Palvelualojen ammattiliitto - PAM)
Siltasaarenkatu 18 A - 00530 Helsinki, Finland
https://www.pam.fi/
@pam.fi

The Central Organisation of Finnish Trade Unions (Suomen Ammattiliittojen Keskusjärjestö - SAK)
Pitkäsillanranta 3 - 00531 Helsinki, Finland
https://www.sak.fi/
@sak.fi

Industrial Employees (Teollisuuden Palkansaajat - TP)
John Stenbergin ranta 6 - 00530 Helsinki,
https://tpry.fi/
@TP_palkansaajat

Danish Union of Metalworkers (Dansk Metal)
Molestien 7 – 2450 Copenhagen, Denmark
Danskmetal.dk
https://www.danskmetal.dk/
@danskmetal

Algorithmic rule

AI and the Future of Democracy in Sweden and beyond

The future is here and it is not simply digital; it is algorithmic. Algorithms now shape not only what we see and hear through the media, but also how public administrations function and how work is managed. Decisions about health, employment and even children's education are increasingly outsourced to automated systems. This book explores the algorithmic present, and what can sometimes be described as algocracy – rule by algorithms. To navigate this new reality, we must ask not only technological but also democratic questions.

What national algorithmic scandals have already unfolded, and what lessons can be drawn from them? How is the future of work being reshaped, as employers rely on algorithmic management tools powered by workplace surveillance? Should artificial intelligence be allowed to make decisions in the public sector? Where can we find examples of local governments working for inclusion and transparency through open-source programming? How are institutions handling the massive and sensitive data flows that structure daily life? And what does it mean for democracy if the predominant face of the state becomes a ubiquitous chatbot?

The book also looks ahead to a progressive algorithmic future and to alternatives promotes sovereignty in contrast to Big Tech. Taking back control means reclaiming not just oversight, but also the real promise of technology. The book offers practical insights in how to control of the code, fight the surveillance of work, and retain primacy of democracy over technology.

FEPS
FOUNDATION FOR EUROPEAN PROGRESSIVE STUDIES

Akademikerförbundet SSR
DIGITAL PROGRAMME

978-2-39076-037-5