AI & SOCIETY (2024) 39:155–167
https://doi.org/10.1007/s00146-023-01731-z

MAIN PAPER

Just accountability structures – a way to promote the safe use of automated decision-making in the public sector

Hanne Hirvonen[1]

Received: 4 October 2022 / Accepted: 12 July 2023 / Published online: 12 August 2023

Abstract
The growing use of automated decision-making (ADM) systems in the public sector and the need to control these has raised many legal questions in academic research and in policymaking. One of the timely means of legal control is accountability, which traditionally includes the ability to impose sanctions on the violator as one dimension. Even though many risks regarding the use of ADM have been noted and there is a common will to promote the safety of these systems, the relevance of the safety research has been discussed little in this context. In this article, I evaluate regulating accountability over the use of ADM in the public sector in relation to the findings of safety research. I conducted the study by focusing on ongoing regulatory projects regarding ADM, the Finnish ADM legislation draft and the EU proposal for the AI Act. The critical question raised in the article is what the role of sanctions is. I ask if official accountability could mean more of an opportunity to learn from mistakes, share knowledge and compensate for harm instead of control via sanctions.

Keywords Automated decision-making · Artificial intelligence · Official accountability · Public accountability · Accountability structures · Sanctions · Safety research

1 Introduction

1.1 Digital catastrophe

The child benefits scandal that escalated in 2021 in the Netherlands could be described in this way. To put it briefly, the Dutch tax authority had been wrongly accusing families of fraud, forcing them to repay benefits. The use of an algorithm that turned out to be discriminatory lasted years and led to serious monetary problems and human suffering –even "unemployment, divorces, and families losing their homes" according to Bits of Freedom, the Dutch digital rights foundation (2021).[1] This unfortunate example illustrates some of the risks of the increasing use of automated decision-making (ADM) by the public sector.[2] In addition, the example shows that there is a general need to find ways to avoid unintended consequences in relation to the ADM and

Hanne Hirvonen
hanne.hirvonen@helsinki.fi

[1] Doctoral Programme in Law, University of Helsinki and Legal Counsel at Turre Legal LLC, Helsinki, Finland

[1] Similarly, Amnesty International in its blog post (2021). See also for example The Guardian news "Dutch government faces collapse over child benefits scandal" written by Jon Henley (2021). Albert Meijer and Stephan Grimmelikhuijsen have also raised this case while describing the downsides of new technologies (2020). Examples from other countries can be found from many sources like the Automating Society Report 2020 (although not every case is a failure and automation can also lead to a great deal of good) (Chiusi et al. 2020). Furthermore, the progress shown in that report has led Michele Loi et al. to use the term "automated society", which also illustrates the extent of the phenomenon (2021). Hence, there is no doubt that the deployment of ADM systems is happening throughout society, and not less in the public sector.

[2] Of course, many of the problems related to the use of ADM are similar in the private and public sectors but the public sector has certain significant features like being able to exercise public power which distinguish it from the private sector (Loi et al. 2021; Backes and Eliantonio 2017). Also, note that some writers use the term algorithmic decision-making or even algorithmic or computational systems and refer to the same practice.

algorithms (Meijer and Grimmelikhuijsen 2020). Of course, this ADM is only one dimension of the use of automated and algorithmic decision-making tools and is an example of "automated prediction". Automated prediction differs from "automated individual decisions" as it is a "form of statistical analysis used to identify individuals from a broader group" and often a base for human-made decisions (Widlak et al. 2020). As in the example, automated prediction or profiling is often used as an enforcement tool (Widlak et al. 2020; Kuziemski and Misuraca 2020). Similarly, Zalnieriute Monika et al. divided automation into "human-authored pre-programmed rules (such as expert systems)" and "tools that derive rules from historic data to make inferences or predictions (often using machine learning)" (2019). This also illustrates that it is possible to make distinctions based on technology. However, it is unclear whether the technological choices behind certain systems form a reasonable base for differentiating between accountability rules for ADM in the public sector.[3] Hence, in this article, the definition of ADM is not limited to certain technologies like artificial intelligence (AI)-based systems (separated from rule-based systems).

Automated decision-making can also be defined and divided according to the degree of automation. In this article, the focus has not been limited to the fully automated processes through which the decision is given without any direct human control.[4] Focusing only on those would leave many problematic situations without attention. As in the Netherlands example, an algorithmic system was used as a management tool and civil servants were also involved in the decision-making procedure (Amnesty International 2021). Finally, "decision-making", which is automated, may also generally refer to many types of individual decisions. In this sense, the notion of mass administration is used to describe the most current field of application of automation (Schartum 2016). The scope of this article covers both automated individual decisions and automated prediction in the public sector. Thus, the scope includes both written decisions by which an authority has ruled on an administrative matter that concerns a person's (natural or legal) right, obligation or interest and actual exercise of power, like a police officer stopping someone on the street if based on ADM. In many cases, the definition of ADM needs to be more limited than this. However, in the context of this article, ADM has been used broadly.

As there have been problems around ADM, there have also been reactions. Several regulatory projects concerning the use of technological solutions like ADM or more generally AI are now aiming to bring control over automated administration.[5] Safety and reliability are among the things being aimed at, with a clearer legal responsibility regime and with accountability as one of the keywords in policy papers.[6] Accountability structures that could cover the use of ADM do not arise out of nowhere but are based on previous regulations and legal thinking. The understanding of the concept of accountability generally and official accountability specifically has grown a lot in recent years in the context of ADM. However, the disruption that ADM causes to the administrative and constitutional law would allow even deeper re-thinking of suitable control mechanisms (Liu et al. 2020).

Law is flexible and often legal concepts can be interpreted so that new rules are not needed. However, technological developments may change the regulatory needs as they reshape society. ADM may itself shape our understanding of the rule of law, and interpretation and application of concepts such as accountability, may change (Zalnieriute et al. 2019). The rule of law, and inter alia accountability, governs the relationship between the state and its citizens. ADM causes changes in the manner public administration institutions are organised and consequently affects this relationship (Kaun 2022). In addition, there might also be regulatory solutions that are not fully in line with the empirical knowledge of the current time. The timely question is whether society should promote the safe use of ADM systems in the public sector by relying on and reinforcing traditional accountability structures or if rethinking is needed.

For this article, I examined which legal accountability structures could increase the safe use of ADM systems in light of safety research. In other words, I asked whether the concept of official accountability and especially the dimension of sanctions, could be reimagined to bring about a more

[3] Though in some cases rule-based systems may be simpler and causality between different actors easier to show.

[4] Cobbe et al. have made the same choice in this regard and refer to ADM as "automated processes [that] can either directly produce a decision or produce information on which a human decision-maker subsequently bases their decision in whole or in part" (2021). Schartum uses the term "complete automation" when referring to activity where "both the collection of data describing case-relevant facts and the processing of these data are automated" (Schartum 2016).

[5] Many of them do not concern only the public sector. In the EU, the regulator has made a proposal for the EU AI draft. In addition, Canada and UK have also just suggested their own AI regulation proposals. See An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts (Canada) and Data Protection and Digital Information Bill (UK).

[6] Accountability is inter alia one of the seven key requirement of trustworthy AI application (Communication from the Commission; Building Trust in Human-Centric Artificial Intelligence. Brussels, 8.4.2019. COM (2019) 168 final).

stable legal framework in relation to ADM. I have done this by summarising some key findings from the research literature on safety and then evaluating two separate legislative projects from this perspective.[7] Another of these regulates the use of ADM in the public sector in Finland and yet another is the European Commission's proposal for the AI Act (AIA). I focus on the public sector, official accountability and sanctions,[8] and I have pointed out where the findings of safety research and traditional legal thinking are not necessarily compatible. At the same time, I discussed whether an accountability structure informed by safety research could be reconciled with the rule of law value of accountability of the public exercise of power. I chose this approach because of the nature of the ADM. As I later explain, it can be described as a socio-technical system or a human–computer interaction (HCI). ADM combines both human and technical aspects, which are also relevant in the field of safety research.

The article has been organised in the following way. Next, I elaborate on previous research regarding ADM and accountability to provide more context for this article. In the third section, I briefly present the field of safety research, including the concept of just culture, and explain why blaming and sanctioning generally hinder safety. In the fourth section, I first use the regulatory example from Finland to illustrate how accountability can be assigned to individual civil servants and evaluate this model. Then, I move on to the EU level to study the AI Act draft and the accountability structures it contains. In addition, I present some alternatives to sanction-based accountability structures that possibly could promote the safe use of ADM in the public sector. In the last section, I discuss the findings of this article further and summarise my work.

2 Safety research as a tool of rethinking

Safety and security are broad subjects and relevant to all aspects of society. Thus, safety-related research is conducted in many fields, with well-known examples including patient safety, traffic safety and nuclear safety. Common to all of these is the high-risk nature of the activity. Digital administration is also at least partly "disaster sensitive" by nature or, in other words, in some contexts creates "high risk" as the proposed EU AI Act puts it.[9] In ADM, tasks that were previously handled by humans are solved by the use of data and algorithms. The changing role of human and human–computer interaction is also an important similarity between many well-known areas of safety research and the use of ADM. Of course, the risks can be very different in administrative decision-making and areas such as traffic. While safety most often means physical safety in the latter context, in this article the safe use of ADM is understood more broadly and refers to avoiding harmful consequences of all kinds. It should be noted that one article is not enough to cover all the aspects of the safety and safe use of ADM, and further discussion of the concept would be useful.

In addition, it is important to note that rule of law principles such as accountability both legitimise the exercise of public power and protect citizens from it. Accountability has a strong connection to the essential legal basis of state governance as the notion of accountability may be included in the principle of the rule of law (Heringa 2017). The requirement that the administration is both procedurally and substantively accountable before the courts is central to the rule of law (Craig 2012). The connection between accountability and the rule of law differs from the legal protection and accountability structures that have been created to protect air passengers or patients (for example). This affects the comparison between ADM in the public sector and traditional fields of safety research—not all the possible accountability structures necessarily fulfil the requirements of the rule of law. Still, as the phenomenon of automation is wide and many experiences are available, safety research might offer new perspectives.

Jaana Hallamaa has suggested that safer AI design could be promoted by means of safety research, but she has not studied this in legal terms. Hallamaa points out that "although safety has been an object of intensive study for more than a century and the results of the work have had a great practical impact in different fields, it is a neglected

[7] The material that was used to go about the study was the preparatory work of the Finnish law proposal concerning ADM (HE 145/2022 vp) and the preparatory work concerning the proposed AI Act (Proposal dated 21.4.2021 COM(2021) 206 final, Presidency compromise text 29.11.2021 and Presidency compromise text 15.6.2022. (See references). This material, and especially drafted rules concerning accountability, sanctions and information sharing, was evaluated in the light of safety research (key findings from literature). The material of the article is based on the situation that was 5 January 2023, when the corrected version of the article was submitted. After this, the Finnish law drafting project has continued, ended and the new regulations have entered into force. In the EU, the preparation continues. This development in Finland is briefly updated to the article but it is a subject of another study to fully focus on the rules that actually came into force.

[8] A special focus on the context of public sector is needed, especially as the AIA proposal does not distinguish between private and public law.

[9] In the automated process, problems may be rapidly replicated, and a vast number of people may suffer from the same mistake being made. In Finland, the parliamentary ombudsman has investigated cases where errors in automated taxation affected thousands of taxpayers (original decision EOAK/3379/2018 in Finnish, see also Yle News 2019 in English).

source in the present context" (2021). As an exception, she mentions Sammarco's study (Hallamaa 2021; Sammarco 2005). In addition, Riikka Koulu has approached safety in the context of AI by referring to the "human factors" which originally come from aviation security. Koulu's focus in her article was on the human oversight required for ADM, which she critically evaluated in the light of human–computer-interaction studies. According to Koulu, current policy decisions over automation are insufficient, as earlier HCI research has vividly demonstrated the human limits in controlling automation (2020a). It must be added that combining safety and legal liability is not new for legal scholars who have focused on subjects other than ADM. Hannamari Helke's dissertation "Maritime Safety Investigation in the Legal Field and Its Relationship to the Legal Determination of Criminal Liability" is an example of this type of research. To solve the tension between criminal law and maritime safety investigation, Helke suggests decriminalisation of certain negligent acts or omissions (2022).[10]

Accountability, rule of law and ADM have already been the subject of interest of several other researchers. Accountability has been handled in connection with transparency. Michèle Finck has highlighted "the need to ensure that the public law concepts recognised in liberal democracies are respected as a whole" (2019). Zalnieriute et al. have also researched transparency and accountability, among other core rule of law concepts and brought out academic literature related to ADM and the rule of law in detail (2019). Jennifer Cobbe has approached the use of ADM specifically in the public sector and from the administrative law perspective and stated that the use of ADM does not shift the responsibility and accountability for the lawfulness of the decision-making from public bodies (2019). Later Cobbe et al. introduced a reviewability framework that serves the development of meaningfully accountable ADM processes in practice (2021). Joshua A. Kroll has distinguished the various ways accountability is used and what it refers to in the context of AI (2020). Responsibility for official actions is one dimension of the supervision of ADM systems and the use of power behind those systems. As Kroll states, "notion of accountability as normative fidelity demonstrates that accountability can serve as a governance mechanism" (Kroll 2020). Reuben Binns handles the question of sufficient justifications as one dimension of accountability and suggests that the notion of public reason could "answer to the problem of reasonable pluralism in the context of algorithmic decision making" (2018).

Maranke Wieringa's systematic literature review on algorithmic accountability also helps in defining the debate over ADM and accountability. Wieringa recalls the importance of consequences without which there would be a danger to fall "into the trap of virtue-washing or ill-defined expectations about the system's accountability requirements" (2020). In turn, Alan F. T. Winfield and Marina Jirotka have noted that collective accomplishment and interdependent systemic properties should be considered in accountability structures regarding autonomous systems (2017).[11] Moreover, Nikolaus Poechhacker and Severin Kacianka handle the problem of causality as another dimension of accountability, in their article "Algorithmic Accountability in Context. Socio-Technical Perspectives on Structural Causal Models" (2021b). This review describes the diversity, and also the fragmentation, which is related to research on the topic.

3 Safety research: how to control activities and promote safety?

3.1 Notion of an "accident" or "human error"

Safety as a general concern and especially occupational safety began to raise interest in the nineteenth century due to industrialisation (Swuste et al. 2021). Now safety forms a wide-ranging field of research; for example, the *Journal of Safety Research* "provides for the exchange of scientific evidence in all areas of safety and health" and "focuses on basic and applied research in unintentional injury and illness prevention" (the journal's webpage). Another journal, *Safety Science,* in turn "extends from safety of people at work to other spheres, such as transport, energy or infrastructures, as well as every other field of man's hazardous activities" (the journal's webpage). Furthermore, Paul Swuste et al. started their recent book "From Safety to Safety Science" by asking "how do accidents and disasters occur? How has knowledge of accident processes evolved (Swuste et al. 2021)?" These questions also reflect the kind of information the 'safety' research field can produce, although the review is very general.

One can approach safety, risk avoidance and harm in many ways. Section 1 in the Finnish Safety Investigation Act illustrates well the different starting points of law and safety investigation. According to the section, the purpose of a safety investigation is to promote general safety and prevent accidents and incidents and of losses resulting from accidents. The same section also states directly that a safety

[10] See also as an example "How do we move from a blame culture to a learning culture in policing?" (The Police Foundation and KPMG 2018).

[11] In relation to the collective accomplishment, see also how Salo-Pöntinen describes the design process of joint cognitive systems (2021).

investigation is not conducted to allocate legal liability. These differences are not as neutral as it first seemed. Sidney Dekker, a professor who has researched widely human factors and safety, reminds that there are typically no such legal concepts like the notion of an "accident" or "human error" (2016). According to Dekker, out of many alternative approaches, legal tradition offers only "one perspective on a case of failure" and "one language for describing and explaining an event" (2016).[12] Subrahmanyam Radhakrishna makes a similar observation on legal language and missing notions related to accidents and regarding patient safety, states that "error can also easily be dressed up by the legal system and made to look like a punishable offence" (2015). Maurizio Catino separates "individual blame logic" and "organisational function logic" approaches to accidents and states that the first is in agreement with the Western legal system (2008). This tension is handled more closely in the next subsection.

3.2 From blaming to an alternative accountability framing

Safety research has evaluated the legal way of approaching accidents quite critically. Especially problematic is the aim to control activities by the thread of personal sanctions and a strong focus on individuals and their accountability. Dekker is critical of the criminalisation of human error and explains why criminalisation is actually harming safety (Dekker 2016; Dekker and Breakey 2016). According to his research, the criminalisation of human error does not serve any original purpose, like prevention, in a judicial system. Charging and convicting individuals will not lead to learning from mistakes and improving, and accounting for mistakes should therefore be carried out in some another way (2016). Furthermore, legal proceedings tend to oversimplify accidents as faults by a few or a single individual, even though accident causation in complex, dynamic systems is known to be more complicated. Dekker reminds us that, "many factors, all necessary and only jointly sufficient, are needed to push a basically safe system over the edge into breakdown" (2016). In the medical field, it is claimed that as long as disclosure of medical errors is likely to result in lawsuits, discussion of errors will remain limited. The problem is again placed on the liability that is limited to individuals added with provider service contracts that are easily terminated (Liang 2002). Respectively, in research regarding railway safety, data have suggested that "the culture of blame was creating an atmosphere where frontline staff were reluctant to report adverse events directly to superiors" (Jeffcot et al. 2006). Jeffcot et al. pointed out that the operational staff almost one voice stated the need to maintain a confidential incident reporting system (Jeffcot et al. 2006).

In addition, many times the role of the organisation can be ignored, as the focus is on the individuals. As Dekker puts it "Putting the front-end operator on trial is an example of single-loop learning, which focuses on the first part (possibly a human) that can be connected to the failure and replacing or otherwise dealing with just that part" (2016). Blaming individual users of the technology will not help to develop safety improvements as this approach overlooks the important systemic issues and shifts responsibility from chiefs and vendors, which Green points out in relation to human oversight and algorithms (2022). Even the sharpest criticism focuses on individual accountability law may hinder discussing on errors also on the organisational level. In their review, Milch and Laumann found that "many studies address the existence of a blame culture in which organisations blame each other if something goes wrong" (2016). They refer to an interview with a railway director who explains that contractual relationships and operation of the adversarial legal system tend to lead to a blame culture, whereby companies have a "you should have done that, we were right, you were not" type of dialogue (Milch and Laumann 2016; original source Jeffcott et al. 2006).

3.3 The concept of just culture

It is important to note that safety science does not propose getting rid of all accountability. Instead, it has developed the concept of just culture, which also answers problems concerning blaming free culture and blaming only the system (Dekker and Breakey 2016). Sidney Dekker and Hugh Breakey recommend a shift from retributive justice to restorative justice, which respects social and legal obligations but avoids blame by offering alternative forms of accountability (2016). If accountability is seen as a forward-looking practice, accountability will not be linked to sanctions or punishment anymore. Instead, "it focuses on the collaborative work necessary for change and improvement and connects organisational accountabilities and community expectations to such changes" (Dekker and Breakey 2016). As Hallamaa has summed up the research in the field, the idea of a just culture is to remove barriers to error reporting and learning by creating a work environment in which people willingly report safety-related issues (2021). This is needed in order to improve the prevailing practices and boost organisational resilience (Hallamaa 2021).

Even though combining safety research and traditional legal thinking has certain challenges, it is done in some sector-specific EU legislation. Perhaps the best-known example concerning codifying the findings of safety research into the legislation comes from the aviation sector. Regulation (EU)

[12] See Dekker's homepage to get general introduction to safety science and his work.

No 376/2014 on the reporting, analysis and follow-up of occurrences in civil aviation names the just culture directly several times. According to paragraph 37 in the recital, a 'just culture' should encourage individuals to report safety-related information but not absolve individuals of their normal responsibilities. Furthermore, according to paragraph 37, employees and contracted personnel should not be subject to any prejudice on the basis of information provided pursuant to this Regulation. Only in cases of wilful misconduct or when there has been manifest, severe and serious disregard with respect to an obvious risk and profound failure of professional responsibility, might the treatment differ.[13] As this example from the aviation sector shows, it is possible to lighten the legal rules on personal accountability, at least in the private sector, and move in a direction that is less focused on blame and punishment and is keener on learning and prevention. Whether this would be useful and possible in relation to the use of ADM is evaluated in the following sections.

4 ADM, personal accountability and sanctions

When failures like the one in the Netherlands occur, the concept of official accountability and questions about its current legal forms come up. Many may ask *what* the legal consequences are. Sanctions are one inherent part of many interpretations of accountability (Bovens 2007; Lindberg 2013). This seems rational, as the threat of sanctions should prevent wrongdoings and negligent conduct. The most far-reaching example of sanctions-based regulation is the personal dimension. As the duty to hold office carefully is generally a part of official accountability regulation, it is safe to state that some type of personal, individual accountability usually concerns public officials when acting in office and making administrative decisions.

In this light, it is interesting to evaluate how it was planned for ADM to be brought under the rules of official accountability in Finland.[14] In the terminology of the Finnish law proposal, neither "safety" nor "safe use" come up directly. However, the question of who bears responsibility for the use of ADM forms one important dimension of the proposal. Actually, one of the main points of the whole process is to clarify the official accountability carried personally by every civil servant and employee working with ADM (Suksi 2020). Of course, this is a relevant question in many other jurisdictions, even though solutions may vary and the need to find specific people to hold accountable varies too. For example, it has been stated that public organisations should carry the responsibility for administrative decisions themselves, and the responsibility should not be channelled to individual civil servants (Widlak et al. 2020). This seems to suggest a somewhat different regulatory solution from the Finnish one.

The reason for the Finnish approach under which personal accountability is so strongly emphasised lays in the Finnish doctrine of official accountability. The Constitution of Finland establishes the basis of this approach as its Section 118 states that a civil servant is responsible for the lawfulness of his or her official actions. […] Everyone who has suffered a violation of his or her rights or sustained loss through an unlawful act or omission by a civil servant or other person performing a public task shall have the right to request that the civil servant or other person in charge of a public task be sentenced to a punishment and that the public organisation, official or other person in charge of a public task be held liable for damages, as provided by an Act. Public bodies may be liable to compensate for harm caused, and the Chancellor of Justice and Parliamentary Ombudsman often directs the public entities instead of focusing on single civil servants, but otherwise there are no real collective forms of official accountability in Finland.

What is relevant is the importance of the criminal responsibility as one of the forms of official accountability in Finland. The fact that existing accountability structures of administrative law are closely attached to the criminal forms of liability and thus, to the threat of punishment, could be seen in the section mentioned above and more closely in the Finnish Criminal Law. Thus, it has been important to determine what official tasks are related to the use of ADM that public officials are responsible for. It was proposed that the legislative changes would be placed in a new chapter that would be added to the Act on Information Management in Public Administration. In addition, a new chapter would be added to the Administrative Procedure Act (and this was the final structure of these legislative changes). The proposed chapter of the Act on Information Management in Public Administration included quite extensive obligations on the development, introduction and monitoring of an ADM system. Authorities are expected to document all the phases and identify the people who have participated in the work, and the division of their tasks. Development documents should also state who approved them and the person responsible for testing should also be named (HE 145/2022). Although the law was subsequently adopted as

[13] See also definition of "just culture" in article 2, and in addition Regulation (EU) 2018/1139 on common rules in the field of civil aviation and establishing a European Union Aviation Safety Agency.

[14] To get a more detailed overview see Suksi's article "Administrative due process when using automated decision-making in public administration: Some notes from a Finnish perspective" (2020). In addition, note that this article is based on the preparatory work. The content of the law was only confirmed on March 23, 2023, and the regulations entered into force on May 1, 2023 in Finland.

amended and the obligations were finally regulated in a little less detail, the main principles presented here did not change.

These examples from the proposal illustrate how personal accountability was attached to the use of ADM in the public sector in Finland. To sum up, the aim is to make sure that there are individuals who can be held accountable for the ADM. At the same time, this means that there should be people to blame if something goes wrong. This seems to illustrate the idea of a careful person who does not make mistakes – the human perception that underlies the Finnish doctrine of official accountability. While the chosen regulatory model is the only reasonable way forward within the current Finnish legislation, some have raised the question of whether accountability should be formulated differently in relation to ADM (Hirvonen 2022).[15] However, these considerations have not been the subject of a more detailed discussion during the legislative process.

If looked at from the safety research perspective, it could seem somewhat problematic how the means to control the use of ADM are to be tightly linked to personal official accountability. There is a risk that this sort of accountability structure, including the risk of even facing criminal sanctions, will prevent people from openly talking about their mistakes or near misses. However, as explained in Sect. 3.3 the "just culture" can be understood in a way that the idea is not to protect in situations in which there is wilful negligence or when an obvious risk has been clearly and seriously ignored. Thus, as long as only more serious acts lead to criminal liability, there would seem to be no real contradiction. This means that an accountability structure informed by safety research could be reconciled with the rule of law value of accountability by tuning the penalty threshold if needed. Moreover, fulfilment of the official accountability is a question of legitimisation as the connection to the Constitution shows. The legitimacy of the use of ADM may be based at least partly on the social agreement on the opportunities to hasten, cheapen and improve the public sector's work by means of technology (Catanzariti 2021). Yet, even if our understanding of the rule of law and legitimate exercise of public power change over time, factors such as efficiency should not determine the acceptability of the use of power. This advocates keeping and developing reasonable individual accountability in relation to the use of ADM.

The use of ADM is always a matter of human–computer interaction and ADM systems always also include human aspects, such as humans designing and building the ADM systems (Zalnieriute et al. 2019; Ranerup and Henriksen 2022). Zalnieriute et al. argue that "the transparency and accountability of outputs [of ADM systems] hinges on the accountability of those designing the system" and in order for designers to be held accountable, there must be defined standards against which their actions can be evaluated (Zalnieriute et al. 2019). Hence, it would seem to be meaningful to approach accountability from the perspective of people's responsibilities over their work. This actually fits well with the just culture thinking which values clear duties and responsibilities. The Finnish legislative changes did clarify what are the "normal responsibilities" of civil servants working with the ADM systems. The new rules state what is required of civil servants and thus also protect them from unfair treatment. Somewhat opposite to this approach would be the idea of ADM as a product and thus, building accountability structures on the base of product liability, which is happening now in the EU (see next section). An ADM system or "decision making program" should not be seen only as a product, especially in the public sector, where the question of exercising public power is involved.

In addition, a lot depends on how the regulations would actually be enforced. It is possible that the personal accountability in the ADM context remains mostly symbolic. It is evidently difficult to connect any causality-based accountability to complex ADM-related harm in real life. As Meijer and Grimmelikhuijsen summarise, "algorithmization is not limited to the use of merely one technological system in an organization" (2020). Among other things, the "problem of many hands" comes often up in relation to the ADM systems in which many people, even from different organisations, co-operate to create and use the systems and also data that is used may come from a range of sources (Meijer and Grimmelikhuijsen 2020). Hence, practical reasons in relation to missing evidence, etc. may lead to the shift from repressive to reparative justice in the context of ADM, even if the legislation were to be based on personal accountability and including possible punishments.

The question of clear duties and personal accountability relates to the somewhat blurry need to keep "the human-in-the-loop" in the use of ADM (Koulu 2020a; Koulu 2020b and Green 2022). In this article, there is no opportunity to deal more deeply with the problems associated with human-in-the-loop models. However, it is necessary to refer briefly to those. First, people responsible for overseeing and controlling the ADM system may act in ways that do not meet the system designer's expectations (Binns 2022). Second, ADM systems that were originally designed to assist the human decision-maker may eventually have the final word in many cases, as it has been shown that civil servants using these systems are hesitant or unable to question the suggested outcomes (Kuziemski and Misuraca 2020; Wagner 2019). In addition, it has been argued that with human-technical systems, there are both human and machine contributions to decision making and thus, the false assumption that

[15] In this context Hirvonen refers to Koivisto, Koulu and Suksi and continues the debate.

either a human *or* a machine must be at fault should be corrected (Wagner 2019). Furthermore, the mere process that produces individual culprits is not enough to ensure that the ADM system itself and its use is in accordance with social values (Waldman 2019). Hence, when the accountability structures related to ADM systems are sketched, one should also look at the organisational implementation of the ADM systems (Meijer and Grimmelikhuijsen 2020). In Sect. 5, I evaluate the European Commission's AI Act proposal and how it emphasises the role of the organisation over individuals and aims to operationalise accountability in this way.

5 Alternative model—Emphasis on organisational accountability

5.1 ADM and accountability through sanctioning the organisation

As with many agencies, the European Commission has safe use of AI as a regulatory objective. According to the explanatory memorandum concerning the proposal for the AI Act, the Commission aims to "ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights" and "facilitate the development of a single market for lawful, safe and trustworthy AI applications". The horizontal approach of the AIA is different compared to the Finnish case presented above and focusing directly on the ADM in the public sector. However even if AI and not ADM is the central term in this law proposal, the draft AI regulation would also cover the use of ADM which is in the scope of this article. However, there are uncertainties as to what "AI" will be. The definition of AI has been elaborated in order to limit classic software programs out of the scope of the regulation (Presidency compromise text 2021 and 2022). Smuha et al. suggests changing the name of the regulation to the 'Algorithms Act' or the 'Software Act' or limiting the scope to machine learning systems (2021). This could clarify the scope depending on the direction regulator wishes to go in. In any case, many of the ADM systems used in the public sector would be covered by the regulation and potentially these systems can be "high-risk applications" according to the proposals systematisation. Thus, safety being one of the regulatory objectives, it is meaningful to see how the EU is actually trying to promote safety in this context.

The proposal for the EU AI Act builds accountability from somewhat different elements compared to the idea of personal accountability including sanctions. This occurs for several reasons that touch on the scope of the legislation, the legislator's power and the doctrinal base of the proposal. Despite differences between the EU proposal and the example from Finland, the prevention element seems to be ultimately based on sanctions in both cases.[16] Article 71 in the proposal handles penalties and according to it "Member States shall lay down the rules on penalties, including administrative fines, applicable to infringements of this Regulation". In the proposal, the key roles are "provider" and "user". Both of these have obligations and failing to meet those obligations may lead to the penalties. According to the definitions in Article 3, "provider" means a natural or legal person or public authority that develops or that has an AI system developed with a view to placing it on the market. In the same article, "user" is defined to mean any natural or legal person or public authority using an AI system under its authority, except when the AI system is used in the course of personal non-professional activity. The position of a natural person is not considered any deeper in the proposal. Paragraph 53 in the recital explains that it is "appropriate" that a specific natural or legal person, defined as the provider, takes the responsibility for placing on the market or putting into service of a high-risk AI system, regardless of whether that natural or legal person is the person who designed or developed the system. Moreover, the need to set responsibilities to users is elaborated only generally.

In practice, it would probably be unlikely that a natural person would carry the role of user or provider. In the light of safety research, emphasis on the accountability of legal entities seems to be justified but depending on the jurisdiction, it is not necessarily sufficient in terms of constitutional requirements in relation to ADM and the exercise of public power. In addition, with regard to the accountability of legal entities, it still is worth considering whether sanctions would lead to learning from mistakes or just complex legal processes, which could shift the focus from actual safety promotion.[17] At the same time, it would be important to know whether and in which scenarios it is actually meaningful to hold individuals accountable for AIA-related infringements, as it is also possible to hold natural persons accountable within the proposal's framework. The need for secure, blame-free information sharing also comes up from the position papers of The Future of Life Institute (FLI) and The Future Society (TFS) as both of them recommend connecting the Directive (EU) 2019/1937 protecting whistle blowers to the AIA (FLI 2022; TFS 2022). While this is worthwhile, it does not fully solve the tension between personal accountability (broadly understood) and sharing information about one's own or colleagues' mistakes and infringements.

[16] Even though monitoring the use of an AI / ADM system and other aspects like that are considered in both proposals too.

[17] The political opinion on sanctions varies (Bertuzzi 2022) and some changes to them have already been proposed (Presidency compromise text 2021 and 2022).

Besides direct penalties, other types of consequences are in place in the proposal. The way that the AIA is constructed is not actually that far from authorisation procedures. AI systems are seen as products that need to fulfil the product safety requirements. High-risk AI systems should bear CE marking to indicate their conformity with the AIA, and in certain conditions, the market surveillance authority may demand the withdrawal of a product from the market. Hence, the proposal provides an incentive this way for the providers to produce safe ADM, as recalls are harmful both for private and public producers. This approach to "sanctioning" might increase safety more than fines, as the focus is to remove dangerous systems from use. Thus, in this regard ADM accountability structures benefit from product liability thinking.[18] However, it is important to note that the proposed model is distinguished from pharmaceutical regulation, in which the authority performs the evaluation and grants permission to place the product on the market, as Michael Veale and Frederik Zuiderveen Borgesius have pointed out (2021). In the following subsection, I leave direct and indirect penalties and sanctions to see what other means there are to promote safety in the AIA. I still base my lens on the findings of the safety studies and especially on information sharing and learning.

5.2 Other ways to promote safety within the AIA

So far, my focus has been on the forms of accountability that aim to control the use of ADM by the thread of sanction. Yet, this is only one way to promote objectives such as safety. It is important to note that the AIA proposal also includes accountability mechanisms other than punishment. The Finnish proposal could also be presented from this point of view, but due to limitations of space, I have focused on the EU level. Another reason for this limitation is that it remains to be seen what leeway the draft regulation leaves for national legislation. In the worst-case scenario painted by Veale and Zuiderveen Borgesius, the extraordinarily broad scope of the draft will "restrict legitimate national attempts to manage the social impacts of AI systems' uses" (2021).

The proposed AI regulation contains its own chapter on the monitoring and reporting obligations for providers of AI systems with regard to post-market monitoring and reporting and investigating AI-related incidents and malfunctioning (Title VIII of the AIA). According to the proposal:

> "AI providers will be obliged to inform national competent authorities about serious incidents or malfunctioning that constitute a breach of fundamental rights obligations as soon as they become aware of them, as well as any recalls or withdrawals of AI systems from the market. National competent authorities will then investigate the incidents/or malfunctioning, collect all the necessary information and regularly transmit it to the Commission with adequate metadata. The Commission will complement this information on the incidents by a comprehensive analysis of the overall market for AI" (the explanatory memorandum concerning the proposal for the AI Act).

Thus, sharing information on incidents and malfunctioning is recognised in the proposal, which is also justified in light of the safety studies referred to. However, there are several limitations in this area of the regulation which cannot be fixed only with the guidance that the commission develops later (see Article 62) to facilitate compliance with the reporting obligations.

First, the obligation to report is the concern only of providers and users of *high-risk* AI systems, not other types of systems. The model that applies the heaviest regulation to the highest risks may leave useful information out of the reporting as a side product. Second, only *serious incidents* and *malfunctioning* are within the scope of the reporting system. According to article 3, "serious incident" means any incident that directly or indirectly leads, might have led or might lead to the death of a person or serious damage to a person's health, to property or the environment or a serious and irreversible disruption of the management and operation of critical infrastructure. Malfunctioning refers to any malfunctioning which constitutes a breach of obligations under Union law intended to protect fundamental rights.

Obviously, this covers many situations. Nonetheless, the reporting threshold seems rather high, and it might be difficult to see when an incident is as severe as the regulation expects. Moreover, the latest Presidency compromise text contains slight changes to the definition of "serious incident" and the wording "might have led or might lead" is removed from it (2022). This may affect the interpretation of the obligation. Its own question is whether the right to good governance is recognised as a fundamental right, or whether situations that are more abstract than physical danger be recognised and reported. The reporting system starts from the idea of a single event that constitutes a thread for safety. The most evident example of this is the reference to causing death. However, ADM systems may cause cumulative indirect harm, which can be difficult to notice and hence conceptualise as a single incident or breach to be reported. In addition, the harm caused by a "faulty" ADM system can be small for a single individual party, but still have wider meaning at the societal or collective level. "Incidents" of this type may be difficult to point out if the focus is on single events or malfunctions. Furthermore,

[18] Smuha et al. criticise the AIA proposal inter alia for "putting undue faith in the effectiveness of conformity assessment and CE marking" (2021).

notification should be made after the provider has established a causal link between the AI system and the incident or malfunctioning or the reasonable likelihood of such a link. Here too, it is left to the reporting party to interpret when there is reasonable causal link. On the whole, the scope of the reporting obligation seems to be rather narrow and in the future it will be important to think more closely about what information about the ADM systems would be useful to share.

Secondly, it is important to analyse how the reporting takes place and what happens after reporting. According to article 62 subSect. 1, the provider carries the reporting obligation and should report to the market surveillance authorities of the Member States where that incident or breach occurred. Furthermore, according to Article 29, if users monitoring the operation of a high-risk AI system notice any serious incident or any malfunctioning, they should inform the provider or distributor and interrupt the use of the AI system. After receiving an incident or malfunctioning notification, the market surveillance authority should inform other relevant national public authorities and the Commission. Information sharing and learning should not be reduced to plain monitoring, and the whole field should get the full benefit of information sharing. It seems that information produced by the proposed reporting system is mostly used to supervise providers of AI systems while the real opportunities to learn from the mistakes of others remain obscure. The Future of Life Institute (FLI) makes similar observations in its position paper stating, "AI advancement in Europe would […] benefit from a clear overview of safety incidents at a European level" (FLI 2022). Therefore, FLI's recommendation is that in the spirit of the existing Seveso directive on industrial accidents, AI-related safety incidents should be reported to an EU database (2022).[19] In addition, in its position paper, the Future Society (TFS) has advised investing in the capacity to analyse information gathered from incident reports, as this would help to detect harmful AI-related macro-trends (TFS 2022).

6 Conclusion

No form of accountability will be perfect in hindering all the problems and harm in relation to the use of ADM in the public sector. As Sammarco has stated "Computer-related accidents have caused harm to the environment, injuries, and fatalities" (2005). This is also likely to be the case in the future. According to Winfield and Jirotka "All machines […] have the potential to cause harm" and "hazardous events will inevitability take place" (2017). However, this does not mean that accountability structures should not have any connection to harm prevention. On the contrary, legal rules can support the safe use of ADM. In any case, the use of ADM in the public sector forces jurisdictions to interpret and at least partly re-regulate the rules on official accountability. Thus, it is worth reflecting on the current and planned accountability structures to the findings of safety studies to evaluate their effectiveness to produce safe ADM practices.

In this article, I have examined two legislative proposals that can be seen as examples of the current tech-induced legal transformation. In the Finnish case, the aim is to provide a solid legal base for the use of ADM in the public sector. The draft EU AI Act is a much larger project than this but would cover many cases of automated administration. It is notable that both of these projects include very traditional elements of legal control—sanctions. Sanctions related to compliance with the rules are to ensure the safe use of ADM systems. Thus, in this sense, legal transformation seems modest. One can ask, when something new and potentially scary is ahead, whether the means of control are mostly supervision and sanctions. Of course, the legislation in force and the whole doctrine behind it affects the matter. There might also be practical matters behind the sanction-driven thinking. It is easier to write rules for turnover-based monetary sanctions or to refer to rules on official duty in criminal law that create new compensation and reparation systems. There are no rights of redress for individuals nor complaints mechanism included in the AIA proposal (Smuha et al 2021). Instead, the potential victims have been ignored and the legal relationship has narrowed down to only between supervisor and supervised. In relation to this, it is important to note that Hakkarainen suggests collective redress in the form of ex-ante protection as a promising way forward (2021). It remains to be seen if there will be accountability models that would better take victims into account in the future.

As also stated in earlier research, safety research brings out perspectives in the light of which human control of ADM must be treated with caution. Thus, focusing too strongly on sanctioning individuals does not seem the right way forward in relation to "accountable ADM". However, this does not mean that individual civil servants should be free from all accountability in relation to the use of ADM. On the contrary, it should be noted that the just culture concept generated in safety research values clear duties and responsibilities. Regarding the mitigation of personal responsibility, it should also be noted that the regulation of flight or patient safety is at a completely different level from the regulation of decision automation. At this stage, it is too early to use the

[19] The aim of the Seveso-III-Directive (2012/18/EU) is the prevention of major accidents involving dangerous substances.

results of the safety research as such and base accountability structures too directly on them. However, safety research can contribute to developing a framework of accountability that considers both individuals and the organisation. In other words, the results of the safety research remind us that the safe use of ADM cannot be improved by focusing on personal accountability nor sanctioning organisations alone. In the future, more attention should be paid to organisational culture, information sharing and learning from previous mistakes.

What is left for the law if looking for and sanctioning accountable parties is not the scope? The extent to which goals such as learning from mistakes, can be promoted or hindered by legal regulation should be subject to further research. Law could possibly promote the safe use of ADM by directing information sharing and building a framework for learning opportunities. As I have illustrated, the proposal for the EU AI Act has certain indirect connections to safety research, and contains sections on information sharing, even though this aspect of the draft has not raised much discussion. However, this trend could be strengthened at least in the public sector. When decision-making becomes more technical, the methods of sharing information about errors will also change. Corrections can no longer be made just by reading the decisions of courts or other supreme legality supervisors. Public entities using ADM and wider AI solutions should be able to learn from mistakes made at the system level and also share information about them. The public sector could even be a trendsetter here and learning from previous mistakes and failures could be part of good digital administration.

Acknowledgements I would like to thank the anonymous reviewers, the special issue editors and the members of our research group at the University of Helsinki Legal Tech Lab, for valuable comments on earlier versions of this article.

Funding Open Access funding provided by University of Helsinki including Helsinki University Central Hospital.

Data availability The research does not separately include data collection, but all legislative materials are openly available as public documents.

References

Academic research and policy papers

Backes C, Eliantonio M (2017) Administrative law. In: Hage J, Waltermann A, Akkermans B (eds) Introduction to Law. Springer, Cham. https://doi.org/10.1007/978-3-319-57252-9_9

Binns R (2018) Algorithmic accountability and public reason. Philos Technol 31:543–556. https://doi.org/10.1007/s13347-017-0263-5

Binns R (2022) Human Judgment in algorithmic loops: Individual justice and automated decision-making. Regul Gov 16:197–211. https://doi.org/10.1111/rego.12358

Bovens M (2007) Analysing and assessing accountability: a conceptual framework. Eur Law J 13:447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x

Catanzariti M (2021) Algorithmic law: law production by data or data production by law? In: Micklitz H, Pollicino O, Reichman A, Simoncini A, Sartor G, De Gregorio G (eds) Constitutional Challenges in the Algorithmic Society. Cambridge University Press, Cambridge, pp 78–92

Catino M (2008) A review of literature: individual blame vs. organizational function logics in accident analysis. J Conting Crisis Manag 16:53–62. https://doi.org/10.1111/j.1468-5973.2008.00533.x

Chiusi, F, Fischer, S, Kayser-Bril, N, Spielkamp, M (eds) (2020) Automating Society Report 2020, (AlgorithmWatch; Bertelsmann Stiftung). Available: https://automatingsociety.algorithmwatch.org visited 11.8.2022

Cobbe J (2019) Administrative law and the machines of government: Judicial review of automated public-sector decision-making. Leg Stud 39(4):636–655. https://doi.org/10.1017/lst.2019.9

Cobbe J, Lee SAM, Singh J (2021) Reviewable automated decision-making: a framework for accountable algorithmic systems. ACM Conference on Fairness, Accountability, and Transparency (FAccT ‘21), Virtual Event. Canada ACM, New York, pp 56–67

Craig P (2012) EU Administrative Law. Published to Oxford Scholarship Online. https://doi.org/10.1093/acprof:oso/9780199568628.001.0001

Dekker S (2016) Just Culture : Balancing Safety and Accountability, 2nd edn. CRC Press. Electronic book, Boca Raton

Dekker S, Breakey H (2016) ‘Just culture:’ Improving safety by achieving substantive, procedural and restorative justice. Saf Sci 85:187–193. https://doi.org/10.1016/j.ssci.2016.01.018

Finck M (2019) Automated Decision-Making and Administrative Law (August 5, 2019) Max Planck Institute for Innovation & Competition Research Paper No. 19–10 https://ssrn.com/abstract=3433684. Also in the Oxford Handbook of Comparative Administrative Law, P. Cane et al. (eds.), Oxford, Oxford University Press, 2020, which has not been available to the author.

Green B (2022) The flaws of policies requiring human oversight of government algorithms. Comput Law Secur Rev. https://doi.org/10.2139/ssrn.3921216

Hakkarainen JM (2021) Naming something collective does not make it so algorithmic discrimination and access to justice. Internet Policy Rev. https://doi.org/10.14763/2021.4.1600

Hallamaa J (2021) What could safety research contribute to technology design? In: Rauterberg M (ed) Culture and computing. Design thinking and cultural computing. HCII 2021. Lecture notes in computer science. Springer, pp 56–79. https://doi.org/10.1007/978-3-030-77431-8_4 **(LNCS 12795, ISBN 978–3–030–77430–1, (Lecture Notes in Computer Science; No. 12795))**

Helke H (2022) Maritime Safety Investigation in the Legal Field and Its Relationship to the Legal Determination of Criminal Liability.

Helsingin yliopisto. Available in Finnish https://helda.helsinki.fi/handle/10138/339161.

Heringa AW (2017) Constitutional Law. In: Hage J, Waltermann A, Akkermans B (eds) Introduction to Law. Springer, Cham. https://doi.org/10.1007/978-3-319-57252-9_8

Hirvonen H (2022) Virkavastuu ja päätösautomaatio – vastuun henkilökohtaisuus kriisissä? Lakimies 3–4(2022):386–418

Jeffcott S, Weyman A, Pidgeon NF, Walls J (2006) Risk, trust, and safety culture in U.K. train operating companies. Risk Anal 26(2):1105–1121. https://doi.org/10.1111/j.1539-6924.2006.00819

Kaun A (2022) Suing the algorithm: the mundanization of automated decision-making in public services through litigation. Inf Commun Soc 25(14):2046–2062. https://doi.org/10.1080/1369118X.2021.1924827

Koulu R (2020a) Human control over automation: EU policy and AI ethics. Eur J Leg Stud 12(1):9–46. https://doi.org/10.2924/EJLS.2019.019

Koulu R (2020b) Proceduralizing control and discretion: Human oversight in artificial intelligence policy. Maastricht J Eur Comparative Law 27(6):720–735

Kroll JA (2020) Accountability in Computer Systems. In: Dubber MD, Frank P, Sunit D (eds) The Oxford Handbook of the Ethics of AI. Oxford University Press, Oxford, pp 181–196

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

Liang BA (2002) A system of medical error disclosure. BMJ Qual Saf. https://doi.org/10.1136/qhc.11.1.64

Lindberg SI (2013) Mapping accountability: core concept and subtypes. Int Rev Adm Sci 79(2):202–226. https://doi.org/10.1177/0020852313477761

Liu HY, Maas M, Danaher J, Scarcella L, Lexer M, Van Rompaey L (2020) Artificial intelligence and legal disruption: a new model for analysis. Law Innov Technol 12(2):205–258. https://doi.org/10.1080/17579961.2020.1815402

Loi M, in collaboration with Mätzener A, Müller A, Spielkamp M (2021) Automated Decision-Making Systems in the Public Sector. An Impact Assessment Tool for Public Authorities. AlgorithmWatch. Available: https://algorithmwatch.org/en/adms-impact-assessment-public-sector-algorithmwatch/ visited 11.8.2022

Meijer A, Grimmelikhuijsen S (2020) Responsible and accountable algorithmization. How to generate citizen trust in governmental usage of algorithms. In: Schuilenburg M, Peeters R (eds) The Algorithmic Society: Technology, Power, and Knowledge, 1st edn. Routledge, Abingdon

Milch V, Laumann K (2016) Interorganizational complexity and organizational accident risk: a literature review. Saf Sci 82:9–17. https://doi.org/10.1016/j.ssci.2015.08.010

Poechhacker N, Kacianka S (2021) Algorithmic accountability in context socio-technical perspectives on structural causal models. Front Big Data 3:2021. https://doi.org/10.3389/fdata.2020.519957

Radhakrishna S (2015) Culture of blame in the National Health Service; consequences and solutions. Br J Anaesth 115(5):653–655. https://doi.org/10.1093/bja/aev152

Ranerup A, Henriksen HZ (2022) Digital discretion: unpacking human and technological agency in automated decision making in Sweden's Social Services. Soc Sci Comput Rev 40(2):445–461. https://doi.org/10.1177/0894439320980434

Salo-Pöntinen H (2021) AI Ethics : Critical Reflections on Embedding Ethical Frameworks in AI Technology. In Rauterberg, M (ed) Culture and Computing: Design Thinking and Cultural Computing. 9th International Conference, C&C 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part II (pp. 311–329). Springer. Lecture Notes in Computer Science, 12795 https://doi.org/10.1007/978-3-030-77431-8_20

Sammarco JJ (2005) Operationalizing normal accident theory for safety-related computer systems. Saf Sci 43(9):697–714. https://doi.org/10.1016/j.ssci.2005.03.001

Schartum DW (2016) Law and algorithms in the public domain. Etikk I Praksis Nord J Appl Ethics 10(1):15–26. https://doi.org/10.5324/eip.v10i1.1973

Smuha NA, Ahmed-Rengers E, Harkens A, Li W, MacLaren J, Piselli R, Yeung K (2021) How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act. https://doi.org/10.2139/ssrn.3899991

Suksi M (2020) Administrative due process when using automated decision-making in public administration: some notes from a finnish perspective. Artif Intell Law. https://doi.org/10.1007/s10506-020-09269-x

Swuste P, Groeneweg J, Guldenmund FW, van Gulijk C, Lemkowitz S, Oostendorp Y, Zwaard W (2021) From Safety to Safety Science, The Evolution of Thinking and Practice, 1st edn. Routledge, Abingdon

Veale M, Zuiderveen Borgesius F (2021) Demystifying the draft EU artificial intelligence Act (July 31, 2021). Comput Law Rev Int 22(4):97–112

Waldman AE (2019) Power, process, and automated decision-making. Fordham L Rev 88:613–632

Wagner B (2019) Liable, but not in control? ensuring meaningful human agency in automated decision-making systems. Policy Internet 11:104–122. https://doi.org/10.1002/poi3.198

Widlak A, van Eck M, Peeters R (2020) Towards principles of good digital administration. Fairness accountability and proportionality in automated decision-making. In: Schuilenburg M, Peeters R (eds) The Algorithmic Society: Technology, Power, and Knowledge, 1st edn. Routledge, Abingdon

Wieringa M (2020) What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, New York, p. 1–18. https://doi.org/10.1145/3351095.3372833

Winfield AFT, Jirotka M (2017) The Case for an Ethical Black Box. In Gao Y, Fallah S, Jin Y, Lekakou C (eds) Towards Autonomous Robotic Systems. TAROS 2017. Lecture Notes in Computer Science(), vol 10454. Springer, Cham. https://doi.org/10.1007/978-3-319-64107-2_21

Zalnieriute M, Moses L, Williams G (2019) The rule of law and automation of government decision-making. Mod Law Rev 82(3):425–455. https://doi.org/10.1111/1468-2230.12412

Official sources

Act on Information Management in Public Administration (906/2019). Available in English https://www.finlex.fi/en/laki/kaannokset/2019/en20190906. (Finland)

An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts. BILL C-27, First Reading, June 16, 2022. (Canada)

Communication from the Commission to the European parliament, the Council, the European economic and social committee and the Committee of the regions. Building Trust in Human-Centric Artificial Intelligence. Brussels, 8.4.2019. COM(2019) 168 final. (EU)

Data Protection and Digital Information Bill. Bill 143 2022–23, First Reading, Jul 18, 2022. (UK)

Directive 2012/18/EU of the European Parliament and of the Council of 4 July 2012 on the control of major-accident hazards involving dangerous substances. (EU)

HE 145/2022 vp Hallituksen esitys eduskunnalle julkisen hallinnon automaattista päätöksentekoa koskevaksi lainsäädännöksi [The government's proposal to parliament for legislation on automatic decision-making in the public administration] (available in Finnish and Swedish). Material available in Finnish https://valtioneuvosto.fi/hanke?tunnus=OM021:00/2020. (Finland)

Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. Brussels, 21.4.2021 COM(2021) 206 final. (EU)

Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. Presidency compromise text 29.11.2021. (EU)

Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union - Texte de compromis de la présidence - Articles 30–39 et 59–62. Presidency compromise text 15.6.2022. (EU)

Regulation (EU) 2018/1139 of the European Parliament and of the Council of 4 July 2018 on common rules in the field of civil aviation and establishing a European Union Aviation Safety Agency (EU)

Regulation (EU) No 376/2014 of the European Parliament and of the Council of 3 April 2014 on the reporting, analysis and follow-up of occurrences in civil aviation (EU)

The Administrative Procedure Act (434/2003). Available in English https://www.finlex.fi/en/laki/kaannokset/2003/en20030434. (Finland)

The Safety Investigation Act of Finland (525/2011). Available in English https://finlex.fi/en/laki/kaannokset/2011/en20110525.pdf. (Finland)

Other online sources

Amnesty International (2021) Dutch childcare benefit scandal an urgent wake-up call to ban racist algorithms. Blog post 25.10.2021. https://www.amnesty.org/en/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/. Accessed 11 Aug 2022

Bertuzzi L (2022) AI regulation filled with thousands of amendments in the European Parliament. EURACTIV.com. https://www.euractiv.com/section/digital/news/ai-regulation-filled-with-thousands-of-amendments-in-the-european-parliament/. Accessed 15 Aug 2022

Bits of Freedom (2021) We want more than "symbolic" gestures in response to discriminatory algorithms. Blog post 10.2.2021. https://edri.org/our-work/we-want-more-than-symbolic-gestures-in-response-to-discriminatory-algorithms/. Accessed 11 Aug 2022

Future of Life Institute (FLI 2022) FLI position on the Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). 4.8.2021. https://futureoflife.org/wp-content/uploads/2021/08/FLI-Position-Paper-on-the-EU-AI-Act.pdf?x76795. Accessed 15 Aug 2022

Henley, Jon (2021) Dutch government faces collapse over child benefits scandal. The Guardian. https://www.theguardian.com/world/2021/jan/14/dutch-government-faces-collapse-over-child-benefits-scandal. Accessed 15 Aug 2022

Journal of Safety Research webpage. A Safety and Health Research Forum. A Joint Publication of the National Safety Council http://www.nsc.org and Elsevier. https://www.journals.elsevier.com/journal-of-safety-research. Accessed 14 Aug 2022

Safety Science webpage. Elsevier. https://www.sciencedirect.com/journal/safety-science. Accessed 15 Aug 2022

Sidney Dekker homepage. https://sidneydekker.com. Accessed 14 Aug 2022

The Police Foundation and KPMG (2018): How do we move from a blame culture to a learning culture in policing? June 2018. Available https://www.police-foundation.org.uk/2017/wp-content/uploads/2018/07/from_blame_to_learning_policy_dinner.pdf.

The Future Society (TFS 2022) Proposal for a regulation - "Artificial intelligence – ethical and legal requirements". Trust in Excellence & Excellence in Trust. August 2021. https://thefuturesociety.org/2021/10/01/eu-ai-act-trust-in-excellence-excellence-in-trust/. Accessed 15 Aug 2022

Yle news (2019) "Parliamentary ombudsman: Automated tax return processing unconstitutional". https://yle.fi/news/3-11087996. Accessed 14 August 2022. Original decision (EOAK/3379/2018) in Finnish https://www.oikeusasiamies.fi/r/fi/ratkaisut/-/eoar/3379/2018

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.