AI & SOCIETY
https://doi.org/10.1007/s00146-022-01572-2

OPEN FORUM

Public procurement of artificial intelligence systems: new risks and future proofing

Merve Hickok[1]

Received: 7 June 2022 / Accepted: 19 September 2022

Abstract

Public entities around the world are increasingly deploying artificial intelligence (AI) and algorithmic decision-making systems to provide public services or to use their enforcement powers. The rationale for the public sector to use these systems is similar to private sector: increase efficiency and speed of transactions and lower the costs. However, public entities are first and foremost established to meet the needs of the members of society and protect the safety, fundamental rights, and wellbeing of those they serve. Currently AI systems are deployed by the public sector at various administrative levels without robust due diligence, monitoring, or transparency. This paper critically maps out the challenges in procurement of AI systems by public entities and the long-term implications necessitating AI-specific procurement guidelines and processes. This dual-prong exploration includes the new complexities and risks introduced by AI systems, and the institutional capabilities impacting the decision-making process. AI-specific public procurement guidelines are urgently needed to protect fundamental rights and due process.

Keywords AI systems · Algorithmic accountability · Public procurement · Human rights · AI governance · Transparency · Due process

Public entities around the world are increasingly deploying AI and algorithmic decision-making systems to support public services or use their enforcement powers. The rationale for the public sector to use these systems is similar to the private sector-increase efficiency and speed of transactions and lower the costs (UK Government 2016; GSA 2020). However, public entities are first and foremost established to meet the needs of the members of society and protect the safety, fundamental rights, and wellbeing of those they serve. Their existence is justified by the promise of such service and protection. People agree to abide by rules knowing they will be served in return, the decisions will not be arbitrary and there are means of redress and assigning responsibility when a harm occurs. Therefore, public entities are held to a higher level of accountability and transparency than private ones which are profit-driven and might not necessarily have the interests of public as a priority.

Currently AI systems are deployed by the public sector at various administrative levels without robust due diligence, monitoring, or transparency. This results in a growing entanglement between the private vendors and public actors, and a blurring of the lines of accountability and responsibility. Public sector actors are also keenly aware of the gap between their existing internal capability and capacity compared to what is needed to properly procure and manage these systems (Executive Order 2020; OECD 2020). This paper critically maps out the challenges in procurement of AI systems by public entities and the long-term implications necessitating AI-specific procurement guidelines and processes. This dual-prong exploration includes both new complexities and risks introduced by AI systems, and the institutional capabilities impacting the decision-making process. AI-specific public procurement guidelines are urgently needed to protect fundamental rights and due process.

✉ Merve Hickok
merve@lighthousecareerconsulting.com

[1] AIethicist.org, Ann Arbor, MI, USA

Published online: 02 October 2022

1 Literature review

When a public entity deploys an AI system to provide a public service or to enforce its powers, the choice for the individual members of the public to opt-out from the use of the system or being subjected to its use is limited. An individual who is in an unbalanced power relationship against a government entity cannot easily challenge the procurement and implementation of a system. Individualizing the harms and impact of a system can also make it difficult to distinguish between personal experience and group-level collective harms. To a certain extent, this power imbalance is corrected with transparency and accountability mechanisms available in public procurement process which obliges the public actor to provide access to information. The entity may be required to conduct assessments, disclose the details and findings, be ready to share further information if requested and answer to the public. The public and civil society, on the other hand, can use this information to understand the impact of the system, on certain groups, society, or environment. Such insight can help public to also challenge the system's fairness, request modification or termination. This evaluation may also result in consequences for the public actor. However, the ability of the public entity to effectively share information and the society to benefit from the process and hold the entity accountable can be significantly impacted with the introduction of complex algorithmic systems. This impact is compounded when these AI systems are proprietary systems.

In United States, after civil society and legislators voiced concerns over privacy and bias in facial recognition technology (Buolamwini and Gebru 2018; NAACP 2022), Internal Revenue Service (IRS) limited the use of ID.me, a biometric identity verification software. News headlines show algorithms found to be biased against African American defendants in prediction of recidivism used in sentencing and bail process (Angwin et al. 2016), leading to false arrests (Hill 2020), or downgrading of student results from underperforming schools (BBC 2020).

As these examples continue to grow, accountability concerns grow in parallel. It is now customary to list all the algorithmic bias cases at the beginning of each research paper to draw attention to how ubiquitous algorithmic systems became and how these systems might be biased. However, despite its implications on fundamental rights and due process, the literature covering the nuanced challenges of AI in public systems is still growing slowly. This paper highlights the current research and practice gap focusing on public procurement guidelines for AI systems.

Literature review of this paper covers policy documents, academic research, and civil society reports. Several policy and regulatory developments are envisaged to govern public and private use of AI systems, such as European Commission's draft AI Act which proposes bans on certain AI systems. The draft bill requires providers developing, and public entities using high-risk AI systems to assess their AI systems, engage in ongoing risk management and register their assessments and documentation in a public database (European Commission 2021). Council of Europe is also working on a legally binding transversal instrument, which proposes certain AI systems and practices used by public actors to be banned. Council's Ad Hoc Committee on AI recommends human rights impact assessments to be conducted for AI systems which might have a negative impact on health, safety, and fundamental rights (Council of Europe 2021). Government of Canada requires public entities to conduct impact assessments prior to production of an algorithmic system (Government of Canada 2020), while UK regulator provides guidance to organizations on how to explain AI practices (Information Commissioner's Office 2020). France parliament requires all algorithms used by the government be made open and accessible to the public (L'Assemblée nationale 2016). United States executive branch establishes principles for the use of AI in the Federal government (Executive Order 2019, 2020), while National Institute of Standards and Technology drafts AI Risk Management Framework (NIST 2022).

In addition to these regulatory discussions, academic researchers surface the impact of algorithmic systems in public sector and call for algorithmic accountability (Barocas and Selbst, 2016; Calo and Citron 2021; Cooper et al. 2022; Crump, 2016; Diakopoulos, 2014; Eubanks, 2018; Kroll et al. 2017; O'Neil 2016; Pasquale, 2015; Richardson et al. 2019; Schwartz, 1992; Veale et al. 2018; Young et al. 2019), and impact assessments to be mandatory (A Civil Society Statement 2021; Ada Lovelace Institute 2021; Kaminski and Malgieri 2019; Reisman et al. 2018). A robust literature identifies the need for transparency and public disclosures. Such disclosures can be in the form of transparent procurement documentation, mandated human rights impact assessments, registries, as well as specification documents detailing the qualities of the datasets used and the design decisions embedded in the AI systems (Bender and Friedman 2018; Gebru et al. 2021; Hind et al. 2019; Holland et al. 2018; Metcalf et al. 2021; Shin 2020).

Some civil society organizations directly call on governments to take action. Center for AI and Digital Policy's "AI and Democratic Values Index 2020" report (CAIDP 2022) provides an analysis of the national AI strategies and practices across 30 countries. One of the five recommendations of the Index for national governments is "Countries must commit to the principles of fairness, accountability, and transparency in the development, procurement, and implementation of AI systems for public services." In the second edition of

AI and Democratic Values Index (CAIDP 2021), the analysis is extended to 50 countries. Results show some countries are deploying responsible practices in their use of AI in public sector. However, there is still an outsized number of governments using AI systems which were not developed, procured, and implemented transparently or managed in a way benefiting fundamental rights and society first. AI Now Institute provides analysis of algorithmic decision-making (ADM) system uses by government in US (AI Now 2018). AlgorithmWatch, a European civil society organization, in its Automating Society Report, provides a similar analysis of use cases by European governments. The report recommends "Without the ability to know precisely how, why, and to what end ADM systems are deployed, all other efforts for the reconciliation of fundamental rights and ADM systems are doomed to fail (AlgorithmWatch 2020).

2 Methodology

Public entity systems and decisions are subject to various requirements for fairness, transparency, and accountability. However, the ability to meet these requirements might change with introduction of AI systems. The complex nature of algorithmic systems introduces new and emerging risks when applied in different social, political, or economic contexts. This paper uses a dual-prong exploration of both new complexities introduced by AI systems, the institutional capabilities impacting the decision-making process and the long-term implications. Such dual-prong approach is necessitated by the intertwined nature of risks and implications in general, and the new complexities added by AI systems in specific. The exploration brings together academic research with the broader policy discussions. Analysis of emerging risks and challenges creates pathways to then identify the impact on fundamental rights and due process. When the complexities and implications are mapped together, researchers and policymakers can use this framework to critically interrogate the existing public agency practices and develop new guidelines and processes. The methodology includes analysis of policy and advocacy documents, investigative journalism articles, and literature review of the concepts of responsibility, fairness, accountability, and transparency in the context of both algorithmic systems and public entities. The author also draws from discussions with numerous public sector practitioners and advocates globally about institutional challenges and the complexities of governing AI systems in the public sector.

3 Procurement at a glance

In a normal public procurement scenario, a public actor might announce a Request for Information/Price (RFI / RFP) to gather information on vendor, product, service, technical specifications, or pricing from private entities. The procurement team(s) might then complete due diligence as per the needs of the entity and award a contract to a vendor. So far this is a normal process which repeats itself across many countries. In a better scenario, there might be a requirement to conduct an impact assessment (such as environmental or human rights impact assessments) or consideration on how the product fulfils public policy objectives. The results of the assessments are shared publicly, so a discussion can take place before a determination to deploy a system can be made. These publicly available impact assessments allow the interested parties and impacted communities to engage with the process, raise concerns, provide feedback, and at times, stop a system from being implemented if the impact level is unacceptable. The transparency in this process helps the parties to question and verify information and to hold organizations accountable. In an optimum scenario, the public is also engaged in the oversight of some of the systems so the practitioners can be checked against conformity with the rules of engagement. For example, a civilian oversight council, or a citizen oversight board, might be a governing body assessing the engagements of a department of public safety or law enforcement at a State or City level.

4 Challenges regarding responsibility

The challenge of the distributed state of responsibility in the case of AI systems can emerge across three different layers. First one is due to the nature of different administrative levels within a country, the second is due to multiple actors involved in the design, development, and use of AI systems and the third layer is the increasingly complex and opaque nature of AI and big data systems. This draws from Nissenbaum's concept of 'many hands', one of the barriers to accountability. Complexity can refer to datasets and models; organizational/institutional layers without clear cut responsibilities; different systems and datasets interacting with each other, and finally the nature of operating systems that contribute to the functioning of the whole system. Nissenbaum suggests that any and all of these levels of 'many hands' problems can operate simultaneously, further obscuring the source of issue (Nissenbaum 1996).

US, for example, is a country with federal system. It has a distributed level of responsibility and engagement across its federal, state, city and even town level administrative structures. In non-federal systems, the levels are different but can be still distributed as national systems versus city or town municipalities. Each entity at each level has its own policy agendas and budget and procurement priorities. As of 2020, only at US federal government level, 157 AI tools across 64 different national government entities were documented (Engstrom et al. 2020; Coglianese and Lehr 2017;

Coglianese and Lampmann 2021). This number might not reflect the accurate count since there is no consolidated public registry with a single definition of AI systems. Total number of AI tools across different administrative levels is unknown. Even at horizontal level, for example law enforcement in different cities might use a variety of AI products. Variances in implementation and operationalization of AI systems can also cause a layer of complexity in assigning responsibility.

The development and maintenance of AI systems themselves also necessitate different stakeholder involvement in the process. From collecting the data to developing of AI models to securing the infrastructure to maintaining the systems, multiple actors make decisions through the lifecycle of an AI system. So even setting aside the distributed responsibilities within the public entities, the AI development process is also complicated to pinpoint to a certain decision which might have caused a harmful outcome.

AI systems are used to analyze very large sets of data to make predictions, classifications, recommendations, decisions, etc. The complexity of the datasets and some of the more advanced techniques used for AI models make these systems opaquer, at times to the point where neither the developers nor the users understand how a certain outcome was produced. In addition, some techniques allow the AI models to continuously learn from new data and user interaction. This means that even if there was an understanding of the initial model, the situation might change in time due to new learning or adversarial attacks to the systems. Paraphrasing Weizenbaum, complexity distributes responsibility (Weizenbaum 1976).

In the context of AI systems used by public sector, this multi-layered complexity can also mean that the public actor itself does not understand the system it is procuring and deploying. Institutional capacity limitations, both on procurement and implementation phases, may result in discriminatory or faulty systems embedded in core function of the entity. A great number of current regulatory efforts as well as the technical research focus on a requirement of explainability of AI systems (Adadi and Berrada 2018; Dwork et al. 2012; Forsythe 1995; Haijan and Domingo-Ferrer 2013; Ribeiro et al. 2016). Explainability usually focuses on technical transparency of the components of AI systems. The assumption is if the behavior of the model and the outcomes can be explained for different parties, then the system can be scrutinized for accuracy, mathematical definitions of fairness and model behavior. Other studies analyze effect of explainability in AI on user trust and attitudes toward AI (Shin 2021). However, technical transparency might not always be available. US Federal Acquisition Regulation (FAR), which is the primary document, and agency acquisition regulations, gives the government unlimited rights in data except for copyrighted works. FAR "specifically excludes the source code, algorithms, processes, formulas, and flow charts of the software" from the Form, Fit, Function data (US FAR 2022). Even if all information was available, as Busuioc remarks, "significant technical expertise asymmetries run to the detriment of [public sector] users, further compounded in the public sector by resource shortages and cut-back pressures on public services, often driving the adoption of algorithms in the public sector" (Busuioc 2021). In short, the ability of public procurement teams to understand the accurate functioning of algorithmic systems is constrained by informational asymmetries, multiple sources of bias (Hickok et al. 2022; Brown et al. 2021), current procurement guidelines, human biases in perception (Shin 2022) and multi-layered complexities detailed above. These constrains then create a butterfly effect on how the algorithmic systems impact society.

5 Challenges regarding fairness

Fairness in algorithms has been discussed in different public system use cases such as welfare eligibility (Eubanks 2018; Lecher 2018), immigration detention (Koulish 2016), or recidivism (Angwin 2016; Larson et al. 2016). In the absence of AI-specific public procurement guidelines and lower levels of institutional capabilities (Dunleavy et al. 2007), public actors may implement AI systems which result in unintentional, negative impact on individuals or society. Public actors interact with society (Sloane et al. 2021). They might procure a proprietary system developed without consideration for existing policy motivations, values, regulatory rules, or fundamental rights. A four-part formula can help explain how AI systems may magnify or deepen the existing inequities and biases within society.

Values + Data + Algorithmic Models = Outcomes

Humans encode their values within all the systems and structures they build. Value encoding, which does not consider the diversity of perspectives and experiences, results in empowering and privileging one group's values and perspectives over others. Value misalignment, on the other hand, means what we want the AI systems to do and what AI system does may be very different, leading to serious unintended consequences (Birhane et al. 2022). So even when we intentionally try to code certain values, we might get it wrong.

Data which trains the AI models are collected by humans, shaped by humans and they are about humans. "Every data set involving people implies subjects and

objects, those who collect and those who make up the collected. It is imperative to remember that on both sides we have human beings (Onuoha 2016)." Every such dataset reflects historical and structural inequities.

Algorithmic models work on mathematical definitions and functions. They optimize the given functions. There are multiple definitions of algorithmic fairness (Verma and Rubin 2018). Shin points out the meaning of algorithmic fairness is context dependent and that there is no widely accepted definition (Shin 2020). Sometimes different definitions of fairness cannot be simultaneously achieved (Berk et al. 2018; Chouldechova 2017; Friedler et al. 2021; Kleinberg et al. 2017; Mitchell et al. 2021). The issue is compounded due to dependency on 'only' mathematical formulations of fairness. What cannot be formalized to ensure fairness or equity cannot be part of an automated system. A corporate vendor developing technological solutions will end up simplifying the problems. Public policies will be translated into what can be quantified and what can be coded.

A public entity must have the means to interrogate an AI system and understand the consequences of deployment. These risks must be examined at procurement stage. However, AI systems are sociotechnical systems, in other words, they are made up of both social and technical elements. They interact with their environments. Their behavior and outcomes are shaped by their interactions with humans and environment. In return, they shape their environments and change the behavior of those around them (Dobbe et al. 2018; Eckhouse et al. 2019; Sculley et al. 2015). These systems are used to reduce the complexity of human nature and interactions to collectible data points. These data points are expected to represent humans and society through abstractions and constructed labels. However, value encoding through abstraction may 'render technical interventions ineffective, inaccurate, and sometimes dangerously misguided when they enter the societal context' (Passi and Barocas 2019; Selbst et al. 2018). Development of AI systems requires interactions to be datafied via proxies. It also requires legal and philosophical concepts to be converted into mathematical formulations. All these actions require developers to make choices and design decisions (Sculley et al. 2015). A corporate vendor might claim its AI system is a solution to a societal problem or a public service. However, such claims are easily made without an understanding of the context of a public service, the rationale and history of policies, the impacted people and communities, and the interactions of these systems with stakeholders and implications of the public service in the greater welfare of society.

The public actor must also have the capability and capacity to monitor such embedded AI systems on an ongoing basis. Even a perfectly designed and deployed system (if one exists) shapes the behavior of its institutional users. Users change their expectations according to the data the system can collect and process, and according to the outcomes the system can produce. The decision-making environment changes. This issue is compounded by models using machine learning where AI models change their behaviors and outcomes due to new data and interactions. With such changes in the environment, the corporate product slowly fades into the background and becomes part of the institution without being questioned. If proper training is not provided and institutional capability is not present, the results of AI systems may be taken as objective truths which do not necessitate further review or questioning. Separately, the institutional priorities and agenda of a public entity might change in time. The policy goals may impact how a system is used. For example, the spirit of a policy might be to provide access to resources to everyone who is in need or eligible in the most efficient way. An algorithmic system is deployed for that goal. However, with changing agendas and priorities, the system might eventually be used to catch fraud, to limit the number of people accessing the systems, or to criminalize individuals from different backgrounds with lower income or education.

In most cases public entities are interested in AI systems due to resource and skills constraints. The entity determines a need for a solution to use resources more efficiently and respond to changes quicker. Big data collection and processing capabilities with AI systems become attractive. However, the stakes become higher in cases where such systems make determinations or even recommendations impacting people's life and liberty, access to opportunities and benefits, their rights such as rights to privacy, expression or association and due process. In those situations, a more robust process is necessary to determine whether a system is the needed solution, whether it has impact on human rights and due process, and who gets to make those decisions. The same skill constraint which created a need in the first place can mean there is not enough internal resources and skill to critically interrogate vendor solutions.

One extreme example of above-mentioned complexities is predictive policing. The system depends on individual or location risk profiling and combines multiple data sources to make predictions about who might commit a crime (person-based or where a crime will be committed (location-based). The approach is not only an affront to presumption of innocence and due process, but it also lacks any scientific validity. The predictions depend on historical policing data which is racially and socioeconomically biased in many parts of the world (Barocas and Selbst 2016; Ensign et al. 2018; Kroll et al. 2017; Richardson et al. 2019). The tool offers the possibility of unconsciously privileging quantifiable metrics (Selbst et al. 2018). The predictions and heatmaps from these systems change policing practices. They change how resources are allocated, how police should

interact with people, what data they should collect, how police will be incentivized by the collected data or which crimes should be prioritized (Brayne 2020). The outcomes become the starting point for certain policing practices. The private system changes the way a public service is rendered. Predictive policing systems are dubious and discriminatory but are nevertheless implemented by police departments across many states and countries. We see the lines becoming blurred regarding who is accountable the consequences when an AI system harms a person or a group. Behind a false veil of objectivity, these systems can have irreversible impact on members of the society (Ananny and Crawford 2018). Such systems might harm a person physically, emotionally, mentally, or financially. Either by intent or due to unintentional consequences, AI systems might end up discriminating against people who share similar characteristics (Buolamwini and Gebru 2018; Obermeyer et al. 2019; O'Neil 2016). Calo and Citron write "agencies are turning to systems in which they hold no expertise, and which foreclose discretion, individuation, and reason-giving almost entirely." The writers also suggest "the systems the U.S. government is increasingly procuring yield results no human can justify." (Calo and Citron 2021).

6 Challenges due to scale, speed and connectedness of algorithmic systems

Public actors have been procuring private software and technology for decades. What is different and so concerning with AI systems and big data infrastructures? Over the last couple of decades, the technologies to collect, store, process and connect data improved exponentially. Such change in hardware and software made it easier, faster, and cheaper to build and use these systems. These improvements made it easier for both public and private actors to acquire and use data at unprecedented scale. When AI systems process data and provide algorithmic outcomes, these decisions are made on a scale which cannot be matched by humans. At one hand, this means more public service requests can be handled, at scale and speed which was not previously possible. On the other hand, it also means if an AI system is biased, providing harmful outcomes or being intentionally used for discriminatory purposes, the results will impact a larger portion of the society. When an application or a request is received, a human public employee reviews the case and decides one by one, whereas an AI system can review the data points from thousands of cases and apply the rules encoded in the system to thousands of applications in a matter of seconds and minutes. Therefore, if there is a value misalignment in the code, or if the case is nuanced and requires contextual information, the speed and scale of harm will also be a lot greater than that of a human review. For example, Michigan's MiDAS system (Michigan Integrated Data Automated System) was introduced in 2013 to detect fraudulent applications for unemployment benefits, and thus reduce state's benefits spending. The system built by private vendors upon request resulted in wrongfully accusing more than 40,000 individuals of unemployment fraud. Not only public was not informed about how the system worked, but state provided minimal resources to answer questions about false accusations. Thousands saw lost their houses, filed for bankruptcy, or had their credit scores ruined. Across US, litigation further shows how little is known about these algorithmic systems and how the algorithmic outcomes can be arbitrary (Barry v Lyon 2016; Cahoo v SAS Analytics 2019; Arkansas DHHS v Ledgerwood 2016; K.W. v. Armstrong 2015; Matter of Lederman v King 2016; Loomis v. Wisconsin 2017). The Michigan auditor general investigation later found 93 percent were falsely accused of fraud.

As these systems are deployed by one public actor after another, eventually they will be connected too. The outputs of one algorithmic system will become inputs to another. Records across health, education, labor, credit, justice systems for example will all act as a giant public database. One wrong case decision, or wrong information in an AI system will cause a domino effect of harms on those impacted by these decisions.

7 Challenges on transparency

One of the consequences of private vendors collecting data through online behavior, sensors, personal devices, applications and other technologies is data becomes available to other actors too. Such data can be used to provide more personalized services in some cases. However, it can also be utilized by some public entities where the entity could not collect such information directly, such as civil and criminal enforcement, monitoring, or adjudication domains. Data collected with publicly owned cameras, sensors, drones, can be combined with data collected by private actors. Such aggregation of previously unconnected databases can generate inferences (at times superfluous correlations). Inferences from these systems can then be used to allocate resources to certain geographical areas or neighborhoods, or to concentrate on certain crime activity. For example, The U.S. Immigration and Customs Enforcement (ICE) purchases datasets from data brokers (Biddle 2021), car telematics data (including location, speed, idle time) from car data application vendors (Brewster 2021; Newman, 2019; Talla 2019), or subscription data from utility companies (Aleazis and Haskins 2020; Faife 2021). Law enforcement can request video surveillance footage from doorbell cameras of private citizens (Lyons 2021; Priest 2021), or use facial recognition system like Clearview even though the system is deemed

illegal in multiple countries (Ryan-Mosley 2021). There are fewer regulatory protections on acquisition of data through these means since public actor is not collecting data itself directly. The actor is either purchasing the already collected data or having a private vendor purchase and process the data on its behalf. Less reporting and transparency requirements and less respect for purpose limitations means individuals or public have no insight as to who is using which data for what purposes. Freedom of Information Act (FOIA) exemptions protects "trade secrets and commercial or financial information obtained from a person [that is] privileged or confidential." (Department of Justice 2004) This exemption opens the possibility for datasets within proprietary AI systems to be also exempted from FOIA requests. A recent report from Georgetown Law Center on Privacy & Technology shows how US Immigration and Customs Enforcement (ICE) accesses the personal information of hundreds of millions of Americans through private data brokers (Wang et al. 2022). AI systems which can process such large data points thus become part of crucial infrastructure. In response to the requirements of Executive Order 13,960, federal agencies in US recently published the inventory of AI uses cases. While it is a great step for transparency, due to freedom given to each agency on how to complete an inventory and what level of detail to disclose, the inventories vary significantly. For example, in contrast to Georgetown Law Center's report, Department of Homeland Security Immigration and Customs (ICE) discloses only four AI systems in use for its inventory. Department of Justice's (DOJ) public AI use case inventory is limited a single page, also with only four use cases, providing single word details. Neither DOJ nor ICE disclosures include above mentioned examples of agency practices or vendors, nor do they provide any actionable information to public (NAII 2022).

For a private AI vendor, the priority is not usually the alignment of values and policies within their system and the protection of fundamental rights and rule of law. The objectives are usually creating demand for product, increasing profit and market share, and avoiding liability (a risk to the business) as much as possible in the process. Even when the public actor has an established need and is requesting vendor responses to such need through a formal procurement process, lack of institutional capability to critically analyze AI systems and use of non-AI specific procurement guidelines will result in liability for public entities. AI-specific procurement guidelines need to drive in-depth due diligence and robust processes to understand the organizational, societal, and individual impacts can reduce the knowledge and capability gap. Otherwise, a regular procurement analysis will fall short of addressing all implications specific to big data and AI systems, and the multistakeholder engagement necessary. Additionally, subcontracting arrangements can also obscure the real actors involved and diminish the usefulness of transparency. In 2012, Europol, EU's law enforcement agency, signed an agreement with a French multinational Capgemini to create Europol Analysis System. Capgemini subcontracted the work to Palantir (European Parliament 2020). Even when Europol had issues with Palantir's software Gotham and considered litigation, a full disentanglement was not possible.

Increasingly the public entities are engaging with private data or AI system vendors in more direct and non-transparent ways. When a private vendor interacts with a public entity, the quickest and easiest point of entry is preferred over a public discussion on what their AI system might mean for the community or society. For example, there might be times of crisis when a need for a quick action and solution might be used as an excuse to skip the regular procurement process and obligations. In UK, NHS onboarded Palantir in March 2020 to help develop NHS Covid-19 Data Store, with a no-bid contract valued at £1 between NHS and Palantir. The contract was awarded using what is called the G-Cloud 11 Framework, an accelerated procurement system for minor contracts and does not require a tender to be published. The contract was only revealed after questions from data privacy activists. It is still not clear if impact assessments have been conducted. The cost of continuing with Palantir, however, was clear when contract was extended at £23.5 m at the end of the trial period. In Greece, another zero-cost agreement between Palantir and Greek government was revealed. The agreement which gave Palantir access to vast amounts of health data to help manage COVID-19 crisis was not registered in public procurement system, nor was a mandated data impact assessment was conducted (Black, 2021; Howden et al. 2021).

Alternatively, a vendor might have direct engagement with senior decision-makers or might supply its products for free or discounted prices to the public entity to slowly build the need for the product. Instead of providing a solution for the public actor's established need, such engagements mean that a need is created for a product. Direct engagement and entry point means skipping several layers of stakeholders, internally and externally, who should have been involved in the process. One such case was Palantir gifting its predictive policing system to New Orleans. A secretive arrangement between the company and the mayor, combined with some unilateral powers of the executive, allowed for the system to be implemented and used without public knowledge. The agreement was also unknown to many officials, and it never had to pass through a public procurement process, which would have required public debate and the signoff of the city council (Winston 2018).

The incentive for the AI vendor, in such cases, is the ability to collect data, or train its AI models, or use the organization as a reference for a further sale, or simply to establish itself within the organization (Laperruque 2017)

for a prolonged contract. The more connected an AI system becomes within the organization, the harder it becomes to decouple it later. Palantir is used as an example for multiple angles of the same question in this paper. The vendor is transparent in suggesting 'The systemic failures of government institutions to provide for the public will continue to require both the public and private sectors to transform themselves' and it wants to become 'the central operating system not only for individual institutions but for entire industries.' (Palantir 2020). However, the company is by no means the only example where theoretical concerns about AI systems in the public domain turn into reality. As Marietje Schaake, the director of Stanford's Cyber Policy Centre, warns "We're building a software house of cards which is sold as a service to the public but can be a liability to society. There's an asymmetry of knowledge and power and accountability, a question of what we're able to know in the public interest. Private power over public processes is growing exponentially with access to data and talent." (Howden et al. 2021).

Public disclosures are recommended to enhance transparency. However, as detailed above, they are not always available. Effective oversight and enforcement mechanisms are crucial to enforce transparency and shed light on the actions of public actors. However, we also need to treat such transparency as a means to an end. While very useful in its own right, the focus on technical parts and outcomes misses an understanding of the social elements (Wieringa 2020). We also need to deliberate values and choices, enforce responsibility and accountability.

8 Challenges for accountability

When a private vendor is engaged to provide public services or access to public services, two issues emerge. First, such contracting means the public entities, intentionally or unintentionally, transfer some of their responsibility to private company. Such public entity which should carry a higher duty of care outsources its services to a profit-driven entity through AI systems, except some of the obligations to the public disappear in the transfer. Mulligan and Bamberger refer to "procurement as policy" whereby the algorithmic systems "frequently displace discretion previously held by either policymakers charged with ordering that discretion, or individual front-end government employees on whose judgment governments previously relied…When the adoption of those systems is governed by procurement, the policies they embed receive little or no agency or outside expertise beyond that provided by the vendor: no public participation, no reasoned deliberation, and no factual record. Design decisions are left to private third-party developers. Government responsibility for policymaking is abdicated." (Mulligan and Bamberger 2019). Through procurement conditions and contractual arrangements, a public entity can ensure the vendor carries responsibility and liability for system outcomes or malfunction. However, this still leaves the vendor only answering to the public entity and leaves the affected individuals and communities having to deal with private entities. In cases where a vendor does not have competition in the market, it can also use its power to deflect any accountability and liability if a harm occurs. In 2021, Internal Revenue Service (IRS) signed an $86 million contract with ID.me to provide biometric identity verification services. The arrangement required taxpayers submit their biometrics in the form of a selfie to authenticate their identity. ID.me claims to already serve 27 states and multiple federal agencies (Rappeport and Hill 2022). If the service does not perform equally and equitably across different demographics due to skin tone, age or gender, a taxpayer might be penalized for the error. National Institute for Standards and Technology observes that rates of false positives for Asian and African American faces relative to images of Caucasians can range from a factor of 10 to 100 times (NIST 2019). Alternatively, if ID.me databases are breached, the taxpayer might be subject to identity theft at highest level since one cannot change their biometric identifiers (Buolamwini 2022). Although this arrangement between IRS and ID.me was put on hold after advocacy groups pushed back, the vendor still has contracts across multiple jurisdictions as a public entity partner to verify unemployment insurance applications and still impacting millions of individuals (ACLU 2022a; Metz 2021).

The second emerging issue is the ability of the private vendors to hide behind IP protections. In the absence of a regulation or a contractual requirement which mandates disclosure, a private company does not have any incentive to share its design decisions or code with any actor. This makes it impossible to analyze how these systems work, audit their validity, reliability, or accuracy, or have an ongoing debate about whether they should be in use. Busuioc, analyzing limitations of algorithmic systems and the implications such limitations pose for public accountability, calls on the emerging accountability gap. Busuioc, referring to Pasquale's work, highlights how he traces 'a shift in this context from "legitimacy-via-transparency to reassurance-via-secrecy"" (Busuioc 2021; Pasquale 2011).

As Moss et al. argue even "voluntary commitments to auditing and transparency do not constitute accountability… [as] they do not meet the standard of accountability to an external forum (Moss et al. 2021)." Currently, most of the public entities are not subject to any governance mechanisms which requires a transparent internal and external management of all AI systems used by an entity. Although several policy examples are emerging globally (Central Digital and Data Office 2021; City of Amsterdam 2020; City of New York 2020; Executive Order 2020; Government of Canada

2020; Government of New Zealand 2020; L'Assemblée nationale, 2016; Seattle, 2017; UK Office for AI 2020) in most cases even a public entity itself does not have a full picture of its entanglement with a private AI vendor. For example, even a city level law enforcement entity may not know exactly all the systems used across its different departments, how data is integrated, how the outcomes are shaping their practices and policy. This makes it hard for public and civil society to engage with the right partners, find information and hold anyone accountable. In his definition of accountability, Bovens requires five integral parts: (1) an actor, (2) a forum, (3) a relationship between the two, in which the (4)actor is obliged to explain and justify its conduct, the forum can pose questions and pass judgement, and the actor might face (5) consequences (Bovens 2007). In a situation where the actor(s) cannot be properly assigned due to distributed and transferred responsibilities, and vendors are not obliged to explain the behavior of their AI systems, it becomes extremely hard to assign any accountability and consequences when AI systems harm individuals or groups, or infringe upon human rights. In their 2017 article, Kroll et al. write "accountability mechanisms and legal standards that govern decision processes have not kept pace with technology" (Kroll et al. 2017).

9 Limitations and future research directions

There are several limitations to this research work. The first one refers to the information availability and asymmetry. The research is limited to publicly available documents such as academic literature, government reports and registries, investigative journalism, litigation text, and private discussions with practitioners. Both the public actor procuring the AI systems and the vendor developing an AI system currently contribute to the unavailability of easily accessible information. The public actor may have an interest in keeping the details of its intelligence or enforcement systems behind a wall of protections. This interest might drive from legitimate concerns about counter actions and possibility for malicious actors to game the system (Veale et al. 2018). Alternatively, the agency itself might not have access to proprietary algorithms due to prior contractual commitments, or current exemptions in procurement regulation. The vendor, on the other hand, contributes to the information asymmetry by benefiting from legal protections for trade secrets (Katyal 2019). The vendor might also be concerned about liability or an employee backlash it might receive if the details of its cooperation with government was to become public (Campbell 2018; Shane and Wakabayashi 2018).

A different set of limitations relate to reproducibility and replicability of the outcomes of AI systems. Even with full access to these machine-learning systems and technical literacy, it might still be impossible to trace back a particular decision of the AI system and reproduce the exact same result. This creates a situation where an individual whose rights are infringed (or an entity acting on behalf of the individual) may not be able to trace back or replicate the discriminatory decisions.

Another limitation refers to the incentives embedded into the design, development, procurement, and implementation of AI systems. Both the public actor and the vendor can state what problem(s) they are solving with the AI system. However, such public statements or disclosures usually do not include the organizational incentives impacting decisions. Developers or procurement officials may be incentivized to complete due diligence in less time, spending less resources, or in ways possibly contradicting with responsible design and development, or in-depth due diligence.

This research mapped out the challenges in procurement of AI systems by public entities and the long-term implications necessitating AI-specific procurement guidelines and processes. Future research can provide an analysis of benefits and limitations of transparency, especially in the form of public disclosures. How do public disclosures contribute to the governance of AI systems? What are limitations of disclosures? Can emerging technology be used in new ways to contribute to meaningful participation of society in the debates impacted fundamental rights and due process? This kind of inquiry and in-depth analysis can be replicated across many jurisdictions globally as every country has different procurement regulations, infrastructure and governance mechanisms.

In any procurement environment, humans ultimately conduct the due diligence and review the available documents and make judgements. Any transparency method, whether in the form of disclosures, datasheets, explainability reports, or notices needs to be understood accurately by its audience. This means capacity, capability, and perceptions of public procurement officials play a crucial role. A robust literature analyzes human biases, communication approaches and requirements for different types of explanations of AI systems. However, future research can also focus on the sense-making and perception of public officials operating within the confines of government bureaucracy and politics and how they judge current AI principles and the social value of the systems they are involved in procuring.

10 Conclusion: recommendations for public agencies

As this paper mapped out the challenges and risks in procurement of AI systems by public entities and the long-term implications, recommendations for what an AI-specific

public procurement guideline and process should include is also necessary.

The issue at hand is not a single AI system or device, but the whole ecosystem. AI-specific procurement guidelines and governance must be applicable to devices used by the public entity and its agents; to externally collected and acquired data; to the AI software fed by these datasets; to AI platforms which connect disparate software; to the cloud-based hosting infrastructures. Already public agencies are moving their data and communications infrastructures to cloud-based hosting systems. These systems are owned by a handful of major technology companies. This creates an inevitable dependency on private vendors. Public sector will not be able to sustain its own infrastructure. If not governed with public interest and fundamental rights in mind, this eventual entanglement will mean some vendors will become too big to fail. They will be too powerful and will set the terms of the engagement. The situation becomes more concerning when vendors are involved in very high-stake decisions like law enforcement, border management, intelligence, or health and benefit systems. Even if the public entity is interested in severing its relationship, as exampled in the Europol Analysis System case (European Parliament 2020), or if the system is not working as expected, the entity might not be able to easily terminate its contract and disentangle itself from the relationship.

If private AI systems are deployed within public sector, human rights rule of law, and commitment to principles of fairness, accountability, and transparency must be required. Otherwise, public actors will have embedded systems without an independent capability to maintain these systems, or skills to monitor system's performance. Alternative oversight and accountability mechanisms will also not be available due to the initial lack of transparency or subcontracting arrangements. A note of caution is necessary here. Most of the issues explained above about corporate AI systems is also applicable to systems built inhouse by public entities. These systems are still sociotechnical systems. The motives and values of the developers and the institution will still be embedded in these AI systems. The need for governance mechanisms and accountability structures is still there. Therefore, the solution to the corporate entanglement and dependency is not building these systems internally. Obligations and documentation within an AI-specific procurement process must be applicable for both external and in-house development cases. A recent case in point was a lawsuit claiming Immigration and Customs Enforcements (ICE) created a "secret no-release policy" and manipulated the risk assessment algorithm to recommend only one decision. *Velesaca v. Decker* case challenged the automatic and indefinite incarceration virtually all of the thousands of people ICE arrested between 2017 and 2020 for alleged immigration offenses. The algorithm used to recommend an arrestee be released or detained until a hearing was changed in 2015 and again in 2017, removing the ability to recommend release, even for arrestees who posed no threat (Robertson 2020). The detainees were not subject to due process and never had any change at recourse. The settlement in the case in March 2022 secures the right to a fair release assessment for everyone arrested by ICE in New York (ACLU 2022b; Velesaca v Decker 2020). The example of how a risk profiling system forced the Dutch government to resign should be a reminder for all public entities. Systeem Risico Indicatie, SyRi, an algorithm used by Dutch government to detect possible social welfare fraud, was found to be discriminatory against people with dual nationality and low income. The authorities started claiming back benefits from families who were flagged by the system, without proof they had committed such fraud. The claims pushed tens of thousands of families to poverty and separated more than a thousand children from their families into foster care. Some victims committed suicide (Heikkila 2022). District Court of Hague found that "under article 8 of the ECHR, the Netherlands did not strike a fair balance between privacy and the benefits of the use of new technologies to prevent and combat fraud because Syri was "insufficiently clear and verifiable" (Court of Hague 2020). A parliamentary report into the childcare benefits scandal found institutional bias and authorities hiding information or misleading the Parliament about the facts (Dutch Parliament 2020). In response, Dutch parliament adopted a motion in April 2022 to make it mandatory to conduct human rights impact assessment before using algorithms when algorithms are used to make evaluations or decisions about people, and where possible, to make impact assessments public (Dutch Parliament 2022). In May 2022, The Netherlands Court of Audit found that six out of nine algorithms it audited did not meet basic requirements and exposed the government to various risks: from inadequate control over the algorithm's performance and impact to bias, data leaks and unauthorized access (Netherlands Court of Audit 2022).

Another requirement in the public procurement process is to ensure whether an AI system is the right solution to a need or problem. We need to be aware of techno-solutionism and focus on the structural causes of an issue, not the parts of the issue we can collect data about and patch algorithmic systems over them. Such a determination must be made by engaging, internally and externally, multidisciplinary public officials and impacted communities in the decisions (Hickok, 2021). Voices of the impacted communities must be heard and respected. The obstacles preventing them from participating in such engagements must be removed. Especially for cases where a system makes determinations about a person's life and liberty, ability to practice fundamental rights, or access to resources, impact assessments and documentation must be mandated. In parallel, public entities must engage

with impacted communities and civil society in a transparent, multi-stakeholder manner which respects participation parity, to agree on AI-specific procurement guidelines, reporting and disclosure requirements.

Public must have access to relevant information in a way that facilitates meaningful engagement. In October 2021, Eric Lander and Dr Alondra Nelson, by-then White House Office of Science and Technology Policy Director and Deputy Director, stated 'Powerful technologies should be required to respect our democratic values and abide by the central tenet that everyone should be treated fairly…country [US] should clarify the rights and freedoms we expect data-driven technologies to respect…enumerating the rights is just a first step. Possibilities include the federal government refusing to buy software or technology products that fail to respect these rights, requiring federal contractors to use technologies that adhere to this "bill of rights," or adopting new laws and regulations to fill gaps. States might choose to adopt similar practices.' (Lander and Nelson 2021). In the same way, where decisions have serious implications for individuals, algorithms can neither be secret (proprietary) nor uninterpretable (Busuioc 2021; Rudin 2019). AI systems are developed by humans; however, these systems are usually mistakenly perceived as independent, objective, unquestionable technologies. Therefore, the outcomes of these systems should not be used as substitute to other steps in a due process. Both the public and private actors must be held accountable for decisions and outcomes. Procurement, development, and implementation must be subject to robust governance and enforcement mechanisms. These mechanisms necessitate both initial internal capacity building and an ongoing capacity enhancement as the science and technology advances. Data generated, collected, processed, and used by humans can never be bias free. An appreciation of such fact, an understanding of the risks specific to AI systems and their socio-technical aspects should make public actors pay even more attention to due diligence. Procurement regulations should be updated to include obligations for the developers to share details of data qualities, model decision decisions, optimization techniques and processes when required by the public entity. Additionally, procurement guidelines should require a capable internal workforce to be in place before a procurement decision is made for an algorithmic system.

A functioning democracy in which both fundamental rights and rule of law are prioritized, society first needs and agreement, a social contract, on what kind of systems should be allowed or banned. As Gabriela Ramos, Assistant Director-General for the Social and Human Sciences of UNESCO, suggests 'AI technologies can be used to strengthen government accountability and can produce many benefits for democratic action, participation, and pluralism, making democracy more direct and responsive. However, [such technologies] can also be used to strengthen repressive capabilities and for manipulation purposes' (Ramos 2022). An engaging public debate and discourse should result in a basic agreement about which systems should be prioritized and which systems should never be implemented.

References

A Civil Society Statement (2021). An EU artificial intelligence act for fundamental rights. https://edri.org/wp-content/uploads/2021/12/Political-statement-on-AI-Act.pdf

ACLU (2022a) Three key problems with the government's use of a flawed facial recognition service. ACLU Florida. https://www.aclufl.org/en/news/three-key-problems-governments-use-flawed-facial-recognition-service

ACLU (2022b) Settlement secures the right to consideration of release for people arrested by ICE in New York. ACLU New York. https://www.nyclu.org/en/press-releases/settlement-secures-right-consideration-release-people-arrested-ice-new-york

Ada Lovelace Institute, AI Now Institute and Open Government Partnership (2021) Algorithmic accountability for the public sector. https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/

Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6:52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

AI Now Institute NYU (2018) Automated decision systems: examples of government use cases. https://ainowinstitute.org/nycadschart.pdf

Aleazis H, Haskins C (2020) DHS authorities are buying moment-by-moment geolocation cellphone data to track people. BuzzFeed News. https://www.buzzfeednews.com/article/hamedaleaziz/ice-dhs-cell-phone-data-tracking-geolocation

AlgorithmWatch (2020) Automating society report 2020. https://algorithmwatch.org/en/automating-society-2020/

Ananny M, Crawford K (2018) Seeing without knowing. New Media Soc 20(3):973–989. https://doi.org/10.1177/1461444816676645

Angwin J, Larson J, Mattu S, Kirchner L (2016) Machine bias: there's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Arkansas Department of Human Services v. Ledgerwood, 530. Arkansas Supreme Court (2017) Casetext. https://casetext.com/case/ark-dept-of-human-servs-v-ledgerwood-1

Barocas S, Selbst AD (2016) Big data's disparate impact. California Law Review 104: 671. https://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf

Barry v. Lyon, 834. 6th Circuit (2016) Casetext. https://casetext.com/case/barry-v-lyon-2

BBC (2020) A-levels and GCSEs: How did the exam algorithm work? https://www.bbc.com/news/explainers-53807730

Bender E, Friedman B (2018) Data statements for natural language processing: toward mitigating system bias and enabling better science. Trans Assoc Comput Linguist [Online], 6: 587–604. https://aclanthology.org/Q18-1041/

Berk RA, Heidari H, Jabbari S, Kearns M, Roth A (2018) Fairness in criminal justice risk assessments: the state of the art. Sociol Methods Res. https://doi.org/10.1177/0049124118782533

Biddle S (2021) LexisNexis to provide giant database of personal information to ICE. Intercept. https://theintercept.com/2021/04/02/ice-database-surveillance-lexisnexis

Birhane A, Kalluri P, Card D, Agnew W, Dotan R, Bao M (2022) The values encoded in machine learning research. *FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency*. June 2022.173–184. https://doi.org/10.1145/3531146.3533083

Black C (2021) Revealed: data giant given 'emergency' covid contract had been wooing NHS for months. The Bureau of Investigative Journalism. https://www.thebureauinvestigates.com/stories/2021-02-24/revealed-data-giant-given-emergency-covid-contract-had-been-wooing-nhs-for-months

Bovens M (2007) Analysing and assessing accountability: a conceptual framework. Eur Law J 13(4):447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x

Brayne S (2020) Predict and surveil: data, discretion, and the future of policing, 1st edn. Oxford University Press

Brewster T (2021) These companies track millions of cars—immigration and border police have been grabbing their data. Forbes. https://www.forbes.com/sites/thomasbrewster/2021/04/01/these-companies-track-millions-of-cars-immigration-and-border-police-have-been-grabbing-their-data/

Brown S, Carrier R, Hickok M, Smith AL (2021). Bias mitigation in data sets. https://doi.org/10.31235/osf.io/z8qrb

Buolamwini J (2022) The IRS should stop using facial recognition. The Atlantic. https://www.theatlantic.com/ideas/archive/2022/01/irs-should-stop-using-facial-recognition/621386/

Buolamwini J, Gebru T (2018) Gender shades: intersectional accuracy disparities in commercial gender classification. Proc Mach Learn Res 81:1–15

Busuioc M (2021) Accountable artificial intelligence: holding algorithms to account. Public Adm Rev 81(5):825–836. https://doi.org/10.1111/puar.13293

Cahoo v. SAS Analytics Inc., 912. 6th Circuit (2019). Casetext. https://casetext.com/case/cahoo-v-sas-analytics-inc

Calo R, Citron DK (2021) The automated administrative state: a crisis of legitimacy. Emory Law J 70(4):797–845

Campbell AF (2018) How tech employees are pushing Silicon Valley to put ethics before profit. Vox. https://www.vox.com/technology/2018/10/18/17989482/google-amazon-employee-ethics-contracts

Center for AI and Digital Policy (2021) AI & Democratic Values Index 2020. www.caidp.org/reports/aidv-2020/

Center for AI and Digital Policy (2022) AI & Democratic Values Index 2021. www.caidp.org/reports/aidv-2021/

Central Digital and Data Office (2021) Algorithmic transparency standard. https://www.gov.uk/government/collections/algorithmic-transparency-standard

Chouldechova A (2017) Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. Big Data 2(2017):153–163. https://doi.org/10.1089/big.2016.0047

City of Amsterdam (2020) Amsterdam algorithm register. https://algoritmeregister.amsterdam.nl/en/ai-register

City of New York (2020) Public oversight of surveillance technologies act. https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3343878&GUID=996ABB2A-9F4C-4A32-B081-D6F24AB954A0

Coglianese C, Lampmann E (2021) Contracting for algorithmic accountability, Adm Law Rev Accord, vol. 6, 175. https://scholarship.law.upenn.edu/faculty_scholarship/2311/

Coglianese C, Lehr D (2017) Regulating by robot: administrative decision making in the machine-learning era. Faculty Scholarship at Penn Carey Law. 1734. https://scholarship.law.upenn.edu/faculty_scholarship/1734

Cooper AF, Laufer B, Moss E, Nissenbaum H (2022) Accountability in an algorithmic society: relationality, responsibility, and robustness in machine learning. FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency. June 2022. 864–876. https://doi.org/10.1145/3531146.3533150

Council of Europe (2021) Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law. https://www.coe.int/en/web/artificial-intelligence/work-in-progress

Crump C (2016) Surveillance policy making by procurement, 91 Washington Law Review. 1595 (2016). https://digitalcommons.law.uw.edu/wlr/vol91/iss4/17

Department of Justice (2004) FOIA guide, 2004 edition. https://www.justice.gov/archives/oip/foia-guide-2004-edition-exemption-4

Diakopoulos N (2014) Algorithmic accountability reporting: on the investigation of black boxes. Tow Center for Digital Journalism Publications, Columbia Journalism School

District Court of Hague (2020). SyRI. https://perma.cc/DS89-K477. English explanation: library of Congress (March 13, 2020). Court prohibits government's use of AI software to detect welfare fraud. https://www.loc.gov/item/global-legal-monitor/2020-03-13/netherlands-court-prohibits-governments-use-of-ai-software-to-detect-welfare-fraud/

Dobbe R, Dean S, Gilbert TK, Kohli N (2018) A broader view on bias in automated decision-making: reflecting on epistemology and dynamics. Workshop on Fairness, Accountability and Transparency in Machine Learning during ICML 2018, Stockholm, Sweden. https://doi.org/10.48550/arXiv.1807.00553

Dunleavy P, Margetts HZ, Bastow S, Tinkler J (2007) Digital era governance: IT corporations, the state, and e-government. Oxford University Press

Dutch Parliament (2020) Eindverslag onderzoek kinderopvangtoeslag overhandigd. https://www.tweedekamer.nl/nieuws/kamernieuws/eindverslag-onderzoek-kinderopvangtoeslag-overhandigd?msclkid=f1d677c5ae8311ecaaa202cbd2ef5f6d

Dutch Parliament (2022) otie van de leden Bouchallikh en Dekker-Abdulaziz over verplichte impact assessments voorafgaand aan het inzetten van algoritmen voor evaluaties van of beslissingen over mensen. https://www.tweedekamer.nl/kamerstukken/moties/detail?id=2022Z06024&did=2022D12329

Dwork C, Hardt M, Pitassi T, Reingold O, Zemel RS (2012) Fairness through awareness. ITCS '12: Proc. of the 3rd Innovations in Theoretical Computer Science Conference, 214–226. https://doi.org/10.1145/2090236.2090255

Eckhouse L, Lum K, Conti-Cook C, Ciccolini J (2019) Layers of bias: a unified approach for understanding problems with risk assessment. Crim Justice Behav 46(2):185–209. https://doi.org/10.1177/0093854818811379

Engstrom DF, Ho DE, Sharkey CM, Cuéllar M (2020) Government by algorithm: artificial intelligence in federal administrative agencies. Administrative Conference of the United States. https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf

Ensign DL, Friedler SA, Neville S, Scheidegger CE, Venkatasubramanian S (2018) Runaway feedback loops in predictive policing. FAT.

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

European Parliament (2020) Parliamentary question: E-000173/2020h https://www.europarl.europa.eu/doceo/document/E-9-2020h-000173-ASW_EN.html

European Commission (2021) 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 (2021) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

Executive Order of President 13859 (February 11, 2019). Maintaining American Leadership in Artificial Intelligence. https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence

Executive Order of President 13960 (2020). Promoting the use of trustworthy artificial intelligence in the federal government. https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government

Faife C (2021). Utility companies will no longer share data with ICE — but many loopholes remain. The Verge. https://www.theverge.com/2021/12/9/22826271/utilities-ice-data-sharing-thomson-wyden

Fields-White M, Graubard V, Rodríguez Álvarez A, Zeichner N, Robertson C (2020) Unpacking inequities in unemployment insurance. New America. https://www.newamerica.org/pit/reports/unpacking-inequities-unemployment-insurance/a-focus-on-fraud-over-accessibility-the-punitive-design-of-ui

Forsythe DE (1995) Using ethnography in the design of an explanation system. Expert Syst Appl 8(4):403–417. https://doi.org/10.1016/0957-4174(94)E0032-P

Friedler SA, Scheidegger C, Venkatasubramanian S (2021) The (Im) possibility of fairness: different value systems require different mechanisms for fair decision making. Commun ACM 64(4):136. https://doi.org/10.1145/3433949

De La Garza A (2020) States' automated systems are trapping citizens in bureaucratic nightmares with their lives on the line. Time. https://time.com/5840609/algorithm-unemployment/

Gebru T, Morgenstern JH, Vecchione B, Vaughan JW, Wallach HM, Daumé H, Crawford K (2021) Datasheets for datasets. Commun ACM 64(12):86–92

General Services Administration (2020) Artificial intelligence in federal procurement. https://www.youtube.com/watch?app=desktop&v=XJsgbGk8BIw

Government of Canada (2021) Directive on automated decision making. Modified on January 2021. https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592

Hajian S, Domingo-Ferrer J (2013) Direct and indirect discrimination prevention methods. In: Custers B, Calders T, Schermer B, Zarsky T (eds) Discrimination and privacy in the information society studies in applied philosophy, epistemology and rational ethics, vol 3. Springer, Berlin, Heidelberg

Heikkila M (2022) A Dutch algorithm scandal serves a warning to Europe. Politico. https://www.politico.eu/newsletter/ai-decoded/a-dutch-algorithm-scandal-serves-a-warning-to-europe-the-ai-act-wont-save-us-2/

Hickok M (2021) Lessons learned from AI ethics principles for future actions. AI Ethics 1:41–47. https://doi.org/10.1007/s43681-020-00008-1

Hickok M, Dorsey C, O'Brien T, Baur D, Ingram K, Chauhan C, Gamundani A (2022) Case study: the distilling of a biased algorithmic decision system through a business lens. https://doi.org/10.2139/ssrn.4019672. https://osf.io/preprints/socarxiv/t5dhu/

Hill K (2020) Wrongfully accused by an algorithm. New York Times. https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html

Hind M, Mehta S, Mojsilovic A, Nair RG, Ramamurthy KN, Olteanu A, Varshney KR (2019) Increasing trust in AI services through supplier's declarations of conformity. IBM J Res Dev 63(4/5):6:1-6:13. https://doi.org/10.1147/JRD.2019.2942288

Holland S, Hosny A, Newman S, Joseph J, Chmielinski K (2018) The dataset nutrition label: a framework to drive higher data quality standards. *ArXiv, abs/1805.03677*

Howden D, Fotiadis A, Stavinoha L, Holst B (2021) Seeing stones: pandemic reveals Palantir's troubling reach in Europe. *The Guardian*. https://www.theguardian.com/world/2021/apr/02/seeing-stones-pandemic-reveals-palantirs-troubling-reach-in-europe

https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement

Information Commissioner's Office (2020) Explaining decisions made with AI. https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/

National Institute of Standards and Technology (2022). Draft AI risk management framework. https://www.nist.gov/itl/ai-risk-management-framework

National Institute of Standards and Technology (2019). Study evaluates effects of race, age, sex on face recognition software. NIST. https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software

National Artificial Intelligence Initiative Office (2022) Agency inventories of AI use cases. https://www.ai.gov/ai-use-case-inventories/

K.W. v. Armstrong, 180, Class action (2016). Casetext. https://casetext.com/case/kw-ex-rel-dw-v-armstrong-5

Kaminski ME and Malgieri G (2019) Algorithmic impact assessments under the GDPR: producing multi-layered explanations. International data privacy law, 2020, forthcoming. *U of Colorado Law Legal Studies* Research Paper No. 19–28

Katyal S (2019) Private accountability in the age of artificial intelligence. 66 UCLA Law Rev 54:125

Kleinberg J, Mullainathan S, Raghavan M (2017) Inherent trade-offs in the fair determination of risk scores. Proc Innov Theoret Comput Sci 43(1):23

Koulish R (2016) Immigration detention in the risk classification assessment era. Connecticut Public Interest Law Journal, Vol. 16, No.1. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2865972#

Kroll JA, Huey J, Barocas S, Felten EW, Reidenberg JR, Robinson DG, Yu H (2017) Accountable algorithms. University of Pennsylvania Law Review. 165(3). 633 https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3

L'Assemblée nationale (2016) French digital republic act. https://www.vie-publique.fr/eclairage/20301-loi-republique-numerique-7-octobre-2016-loi-lemaire-quels-changemen

Lander E, Nelson A (2021). Americans need a bill of rights for an AI-powered world. Wired. https://www.wired.com/story/opinion-bill-of-rights-artificial-intelligence/

Laperruque J (2017) Taser's free body cameras are good for cops, not the people. Wired. https://www.wired.com/2017/04/tasers-free-body-cameras-good-cops-not-people/

Larson J, Mattu S, Kirchner L, Angwin J (2016) How we analyzed the COMPAS recidivism algorithm. ProPublica. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Lecher C (2018) What happens when an algorithm cuts your health care, The Verge. https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy

Loomis v. Wisconsin, 38 Supreme Court (2017) Casetext. https://casetext.com/case/state-v-loomis-22

Lyons K (2021) Amazon's ring now reportedly partners with more than 2000 US police and fire departments. The Verge. https://www.theverge.com/2021/1/31/22258856/amazon-ring-partners-police-fire-security-privacy-cameras.

Matter of Lederman v. King, 26416. New York Supreme Court (2016). Casetext. https://casetext.com/case/lederman-v-king-1

Metcalf J, Moss E, Watkins EA, Singh R, Elish MC (2021) Algorithmic impact assessments and accountability: the co-construction of impacts. ACM Conference on Fairness, Accountability, and Transparency (FAccT '21).735–746. https://doi.org/10.1145/3442188.3445935

Metz R (2021) Want your unemployment benefits? You may have to submit to facial recognition first. CNN. https://www.cnn.com/2021/07/23/tech/idme-unemployment-facial-recognition/index.html

Mitchell S, Potash E, Barocas S, D'Amour A, Lum K (2021) Algorithmic fairness: choices, assumptions, and definitions. Annu Rev Stat Appl 8:1. https://doi.org/10.1146/annurev-statistics-042720-125902

Moss E, Watkins EA, Singh R, Elish MC, Metcalf J (2021) Assembling accountability: algorithmic impact assessment for the public interest. SSRN J. https://doi.org/10.2139/ssrn.3877437

Mulligan DK, Bamberger KA (2019) Procurement as policy: administrative process for machine learning. Berkley Technol Law J 34:773. https://doi.org/10.15779/Z385X25D2W

NAACP Legal Defense Fund (2022) Coalition of civil rights groups sends letter calling for federal and state agencies to end the use of ID.me and facial recognition technology. https://www.naacpldf.org/news/coalition-of-civil-rights-groups-sends-letter-calling-for-federal-and-state-agencies-to-end-the-use-of-id-me-and-facial-recognition-technology/

Netherlands Court of Audit (2022) An audit of 9 algorithms used by the Dutch government. https://english.rekenkamer.nl/publications/reports/2022/05/18/an-audit-of-9-algorithms-used-by-the-dutch-government

New Zealand Government (2020) Algorithm charter for Aotearoa New Zealand. https://data.govt.nz/use-data/data-ethics/government-algorithm-transparency-and-accountability

Newman LH (2019) Internal docs show how ICE gets surveillance help from local cops. Wired. https://www.wired.com/story/ice-license-plate-surveillance-vigilant-solutions

Nissenbaum H (1996) Accountability in a computerized society. Sci Eng Ethics 2:25–42. https://doi.org/10.1007/BF02639315

Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. Science 366(6464):447–453. https://doi.org/10.1126/science.aax2342

OECD (2019) OECD AI principles. https://oecd.ai/en/ai-principles

OECD (2020) Integrating responsible business conduct in public procurement. OECD Publishing, Paris

O'Neil C (2016) Weapons of math destruction: how big data increases inequality and threatens democracy. Penguin, London

Onuoha M (2016) The point of collection. Data and Society. https://points.datasociety.net/the-point-of-collection-8ee44ad7c2fa

Palantir (2020) Form S-1 registration statement. SEC. https://www.sec.gov/Archives/edgar/data/1321655/000119312520230013/d904406ds1.htm

Pasquale F (2011) Restoring transparency to automated authority. J Telecommun High Technol Law 9:235–256

Pasquale F (2015) The black box society: the secret algorithms that control money and information. Harvard University Press, Cambridge, MA and London

Passi S, Barocas S (2019) Problem formulation and fairness. Proc. of the Conference on Fairness, Accountability, and Transparency. 39048. https://doi.org/10.1145/3287560.3287567

Priest D (2021) Ring's police problem never went away. Here's what you still need to know. CNET. https://www.cnet.com/home/security/rings-police-problem-didnt-go-away-it-just-got-more-transparent/

Public Oversight of Surveillance Technologies Act. https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3343878&GUID=996ABB2A-9F4C-4A32-B081-D6F24AB954A0

Ramos G (2022) Ethics of AI and democracy: UNESCO recommendation's insights. Turkish Policy Quarterly. http://turkishpolicy.com/article/1091/ethics-of-ai-and-democracy-unesco-recommendations-insights

Rappeport A, Hill K (2022) IRS to end use of facial recognition for identity verification. The New York Times. https://www.nytimes.com/2022/02/07/us/politics/irs-idme-facial-recognition.html

Reisman et al. (2018) Algorithmic impact assessments: a practical framework for public agency accountability. AI Now

Ribeiro M, Singh S, Guestrin C (2016) Why should I trust you?: explaining the predictions of any classifier. In Proc. of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. 97–101, San Diego, California. Association for Computational Linguistics.

Richardson R, Schultz JM, Crawford K (2019) Dirty data, bad predictions: how civil rights violations impact police data, predictive policing systems and justice. New York University Law Review Online. https://www.nyulawreview.org/wp-content/uploads/2019/04/NYULawReview-94-Richardson_etal-FIN.pdf

Robertson A (2020) ICE rigged its algorithms to keep immigrants in jail, claims lawsuit. The Verge. https://www.theverge.com/2020/3/3/21163013/ice-new-york-risk-assessment-algorithm-rigged-lawsuit-nyclu-jose-velesaca

Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1:206–215

Ryan-Mosley T (2021) The NYPD used a controversial facial recognition tool. Here's what you need to know. MIT Technology Review. https://www.technologyreview.com/2021/04/09/1022240/clearview-ai-nypd-emails/

Schwartz P (1992) Data processing and government administration: the failure of the American legal response to the computer, 43 Hastings Law Journal. 1321

Sculley D, Holt G, Golovin D, Davydov E, Phillips T, Ebner D, Chaudhary V, Young M, Crespo J, Dennison D (2015) Hidden technical debt in machine learning systems. NIPS 2503-2511

Seattle (2017) Washington, surveillance ordinance 123576, http://seattle.legistar.com/ViewReport.ashx?M=R&N=Text&GID=393&ID=2849012&GUID=5B7D2F80-A918-4931-9E2E-88E27478A89E&Title=Legislation+Text

Selbst AD, Boyd D, Friedler S, Venkatasubramanian S, Vertesi J (2018) Fairness and abstraction in sociotechnical systems (August 23, 2018). 2019 ACM Conference on Fairness, Accountability, and Transparency (FAT*), 59–68. https://doi.org/10.1145/3287560.3287598

Shane S, Wakabayashi D (2018) The business of war: google employees protest work for the pentagon. *The New York Times*. https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html

Shin D (2020) User perceptions of algorithmic decisions in the personalized AI system: perceptual evaluation of fairness, accountability, transparency, and explainability. J Broadcast Electron Media 64(4):541–565. https://doi.org/10.1080/08838151.2020.1843357

Shin D (2021) The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. Int J Hum Comput Stud 146:102551. https://doi.org/10.1016/j.ijhcs.2020.102551

Shin D, Lim JS, Ahmad N et al (2022) Understanding user sensemaking in fairness and transparency in algorithms: algorithmic sensemaking in over-the-top platform. AI Soc. https://doi.org/10.1007/s00146-022-01525-9

Sloane M, Chowdhury R, Havens JC, Lazovich T, Rincon AL (2021) AI and procurement – a primer. https://doi.org/10.17609/bxzf-df18

Talla V (2019) Documents reveal ICE using driver location data from local police for deportations. ACLU of Northern California. https://www.aclu.org/blog/immigrants-rights/ice-and-border-patrol-abuses/documents-reveal-ice-using-driver-location-data

U.K. Office for Artificial Intelligence (2020) Guidelines for AI procurement. https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement

UK Government Office for Science (2016). Artificial intelligence: opportunities and implications for the future of decision-making. https://assets.publishing.service.gov.uk/government/uploa

ds/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf
United States Executive Order 13960 of December 3, 2020. Promoting the use of trustworthy artificial intelligence in the federal government. https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government
United States Federal Acquisition Regulation – 2022-07. https://www.acquisition.gov/far/part-1
Veale M, Kleek MV, Binns R (2018) Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. Proc. of the 2018 CHI Conference on Human Factors in Computing Systems.
Velesaca v. Decker (2020) Casetext. https://casetext.com/case/velesaca-v-decker
Verma S, Rubin J (2018) Fairness definitions explained. In FairWare'18: IEEE/ACM International Workshop on Software Fairness. https://doi.org/10.1145/3194770.3194776
Wang N, McDonald A, Bateyko D, Tucker E (2022) American dragnet: data-driven deportation in the 21st century, Center on Privacy and Technology at Georgetown Law
Weizenbaum J (1976) Computer power and human reason: from judgment to calculation. W. H. Freeman & Co
Wieringa M (2020). What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. Proc.of the 2020 Conference on Fairness, Accountability, and Transparency. 1–18. https://doi.org/10.1145/3351095.3372833
Winston A (2018) Palantir has secretly been using New Orleans to test its predictive policing technology. The Verge. https://www.theverge.com/2018/2/27/17054740/palantir-predictive-policing-tool-new-orleans-nopd
Young MM, Bullock JB, Lecy JD (2019) Artificial discretion as a tool of governance: a framework for understanding the impact of artificial intelligence on public administration. Perspectives on Public Management and Governance 2(4). https://academic.oup.com/ppmg/article-abstract/2/4/301/5602198