Testimonial Injustice in Governmental AI Systems

Catriona Gray

Abstract: *This chapter analyses the application of AI systems which test and/or contest the accounts of human subject(s), and which are applied within the course of governmental decision-making. It argues that the rise in these decisional practices demands thorough interrogation of the ways in which testimony is elicited, offered, and received as an element of AI systems. This enables critical inquiry beyond narrowly conceived ethical categories, allowing for more comprehensive accounts of the range of harms – material and epistemic – produced by systems which bypass, undermine, and challenge the testimony of their targets. I identify the three evidentiary manoeuvres by which testimony figures in various governmental AI technologies: obviation, diminishment and impugnment, and apply the concept of epistemic justice to illuminate the different ways in which harm is produced through their enactment. I argue for a sociotechnical approach which recognizes that resulting testimonial injustices are not easily addressed by the cultivation of more virtuous practices and instead require alternative governance responses. This enables much-needed analysis at the intersection of ethics, epistemology and politics which better equips us to identify new vectors of domination and marginalization, and to imagine and realize less violent alternatives.*[1]

Though often presented as novel and innovative, the use of digital technologies and automation to support governmental decision- and policy-making has emerged over several decades (Henman 2010). These practices have been marked by a recent turn toward greater use of artificial intelligence (AI) systems (Berryhill et al. 2019). This turn has been driven by increases in computational power, higher levels of internet use, alongside increasing computing ubiquity, and greater availability of large datasets and large cloud

1 This work was supported by the Engineering and Physical Sciences Research Council grant EP/S023437/1.

storage facilities. AI systems encompass a range of software-based technologies including machine learning, as well as symbolic approaches based on manipulation of abstract representations of objects and relations (Garnelo/Shanahan 2019). Machine learning, which is now widely used in governmental decision-making (Veale/Brass 2019), includes a range of techniques that allow computational systems to learn directly from data and experience rather than pre-programmed rules. In public-sector contexts, large administrative datasets are used to build machine learning models that in turn support policy and operational decisions, including those which are highly consequential for individuals and their rights. These transformations can be observed at different levels of government across the globe – from municipalities to international organizations (ITU 2021).

Despite this growing adoption, the precise definition of artificial intelligence in law and policy remains contested. The European Union's proposed AI Act (European Commission 2021) is the first attempt anywhere in the world to legislate comprehensively on AI. It broadly defines AI to cover systems which can generate outputs including predictions, recommendations, content and decisions, and which are developed using three approaches listed in its Annexes: (a) machine learning, (b) logic and knowledge-based approaches, and (c) statistical approaches, Bayesian estimation, search and optimization methods. As part of legislative negotiations, member state governments have put forward amendments seeking to narrow the regulation's scope by limiting the definition of AI to systems designed with a level of autonomy to achieve a given set of objectives using machine learning or logic and knowledge-based approaches (Bertuzzi 2022). Civil society and human rights organizations have argued that this move would exclude many rudimentary software systems which can nonetheless have significant impact on people's lives and pose significant risks to their rights (AlgorithmWatch 2021). Rather than limiting inquiry to advanced computational techniques, this chapter considers AI in an expansive sense to include all three approaches listed above, and without a requirement of autonomy.

A more capacious approach is consistent with a sociotechnical analysis which apprehends AI systems not as discrete products or services but as dynamic and constituted by complex webs of actors, code, data and infrastructure. The potential social and ethical implications of a given AI system will depend not just on its technical attributes but on its changing input data, adaptations in its use, and integration with other systems. Many of the risks AI poses to individuals, groups and societies are not inherent in the technology;

they are shaped too by complex development processes and applications across changing contexts. For example, facial recognition technology when used to automate the operation of a coffee machine in a person's home would present a very different range of possible impacts from use of the same product in the policing of political demonstrations.

In this chapter, I analyze the application of AI systems that have three common features. First, they produce *individualized* outputs – including determinations of status, eligibility, entitlements and application of sanctions. Second, they are *adversarial* in that they test or contest the accounts of the human subject (or subjects) to whom they relate. Third, they are applied within the course of *public policy*, including in legal and administrative decision-making and service delivery. I argue that these legal and public policy decisional practices demand thorough interrogation of the ways in which testimony is elicited, offered and received as an element of governmental AI systems. This in turn enables identification and analysis of multiple dimensions of injustice, including, as I will elaborate, epistemic injustice.

Decision-supporting AI systems may be designed, integrated and used within policy practices in ways that undermine particular epistemic values (such as scrutability and explainability) and that give rise to unjust harms against specific individuals and groups in their capacity as knowers. As various thinkers within social epistemology have elucidated, some forms of injustice arise from wrongs against people in their specific capacity as knowers (Fricker 2007; Pohlhaus 2014). These are *epistemic injustices*. Drawing on these theoretical currents, I address a major shortcoming in dominant conceptualizations of ethics and responsibility in AI. Attending only to narrowly conceived ethical and legal categories (including those of fairness and bias) results in a failure to account for the full range of (material and epistemic) harms produced by systems which bypass, undermine and challenge the testimony of their targets. There is a nexus of ethics, epistemology and politics in need of critical attention from scholars concerned with the social implications of AI and related technologies.

The chapter proceeds as follows. First, I introduce the concept of epistemic – including testimonial – justice. After identifying the three main ways in which testimony figures in various adversarial governmental AI systems, I illuminate the different ways in which epistemic injustice can be produced through the development and deployment of AI systems in public policy settings. Whilst the cultivation of epistemic virtues (such as open-mindedness) is often proposed as a means of addressing epistemic injustices, I show that this will fall short when it comes to governmental decision-making supported by AI.

1. Epistemic injustice and AI

The reality that social life increasingly unfolds across digital environments requires us to rethink theories and applications of epistemic justice (Origgi/Ciranna 2017). Miranda Fricker's pathbreaking book *Epistemic Injustice* (2007) offers an account of how unequal social relations inflect what gets to count as knowledge and who gets to count as a credible knower or epistemic agent.[2] These unequal social relations may include, for example, positional categories such as gender, or the relationship between a public authority and a person in receipt of social assistance. Fricker distinguishes between what she terms *testimonial injustice* and *hermeneutical injustice*. The former occurs when one attempts to convey knowledge ("prejudice causes a hearer to give a deflated level of credibility to a speaker's word") and the latter at a prior stage when we attempt to make sense of our own social experiences ("a gap in collective interpretive resources puts someone at an unfair disadvantage when it comes to making sense of their social experiences") (2007: 1).

For Fricker, the central case of systematic testimonial injustice rests on what is deemed "identity-prejudicial credibility deficit" (2007: 28). A credibility deficit can be a result of innocent error which is "both ethically and epistemically non-culpable" (2007: 21) but it will not meet the threshold of an *injustice* unless there is an element of the ethical wrong of prejudice.[3] In essence, testimonial injustice describes the accordance of lower credibility to a speaker or a knower because they belong to a group that has been typecast in some way. Alert to the dehumanizing effect of testimonial injustice, Fricker argues that to cause a person to suffer it is to degrade them not just qua *knower* but also, symbolically, qua *human*. Moreover, she suggests that persistent and systematic cases may also "genuinely inhibit the development of an essential aspect of a person's identity" (2007: 54) and even exercise social

2 It must be noted that Fricker was also working with ideas drawn from a feminist standpoint, and postcolonial and other critical theories (Spivak 1987; Hill Collins 1998).

3 A prejudice can either be one which is relatively incidental and localized or one which 'tracks' the subject across different contexts (e.g., homophobia), giving rise to different forms of injustice.

constructive power in a way that "constrains who the person can be" (2007: 58). Other thinkers seeking to understand epistemic injustice (or violence) have placed less emphasis on prejudice, and instead conceptualized, for example, "pernicious ignorance" (Dotson 2011).

Turning to the other limb of epistemic injustice, hermeneutical injustice, its features appear quite different, but it is also constituted by a wrong to a person in their capacity as a knower. It precedes testimonial injustice and is less directly a matter of credibility determination. The underlying rationale is that, where experiences are occluded, a knower is unable to make sense of them in order to give testimony and to come up with ways of challenging the prevailing social order. Whereas the powerful have at their disposal coherent interpretations of social reality which work to their advantage, others are marginalized in multiple ways by their social location, making their resistance less possible.

Digitally mediated environments are not just spaces in which epistemic injustices can occur or become exacerbated; they can generate novel and distinctive epistemic injustices. Noting the relative neglect within wider critical analysis, Symons and Alvarado argue for scholarly light to be shed on the specifically *epistemic* harms and injustices arising from the "design, development and deployment of complex and opaque data-driven technologies such as machine learning, deep neural networks, and big data analysis" (Symons/Alvarado 2022). According to the authors, the main reason hermeneutical and testimonial epistemic injustices arise through the operations of AI and related technologies is "their opacity and their inability to permit corrective recourse" (2022: 92). Earlier scholarship had already identified the "epistemic opacity" of computational systems and processes (Humphreys 2004) – a feature which has become even more pronounced in some, though not all (Felzmann et al. 2020), AI systems, leading in some cases to epistemic injustice (Alvarado and Humphreys 2017). Taking up these insights, I examine three evidentiary manoeuvres which are reconfiguring the role of testimony (namely obviation, diminishment and impugnment), and the epistemic justice implications of their enactment.

2. The concept of testimony in AI

Although philosophers may disagree about its exact nature – including whether it generates or merely transmits epistemic properties – testimony can be distinguished from other ways of justifying knowledge such as memory, perception and reason (Audi 1997). Within epistemology, scholars have attempted to distinguish testimony from mere assertion. Coady (1992) suggests that, to count as testimony, a speech act must meet several conditions, including: that it is being offered as evidence of a proposition; that the speaker has competence, authority and credentials to make such a statement; and that it is relevant to some unresolved question about which those hearing the testimony are in need of relevant evidence. A revised version of this definition which hinges on the speaker *purporting to convey information* is put forward by Graham (1997). According to this view, what matters is that the speaker *intends* their audience to believe they embody the relevant qualities, and that *the speaker believes* their statement of a proposition to be relevant to a question they believe is unresolved and about which they believe those hearing it are in need of relevant evidence.[4] As Freiman and Miller (2020) suggest, however, a charge of anthropocentrism may be levelled against this human cognition-dependent conception of testimony and assertion. According to them, some instrumental (including AI-generated) outputs can also constitute a kind of *quasi-testimony.*

The adoption and use of AI to produce outputs relating to specific people requires us to re-evaluate not just the role of the speaker (the person or entity giving testimony), but that of the hearer (the person or entity receiving testimony). If we consider the status of an artificial entity as a potentially epistemically and ethically responsible agent – what we might refer to as the "problem of technological agency" (Rosenberger 2014) – we know that a machine cannot *hear* speech as testimony, nor can it alone truly form a belief (as opposed to producing an output or reaching a conclusion). However, it does not follow that we can analyse only the epistemic activities of the *human* agents who design, use and interpret AI systems. Instead, we can undertake a more sociotechnical form of analysis (Sartori/Theodorou 2022) which privileges neither the human nor the machine (nor the data) and which regards these elements, and the epistemic properties they enact and transmit, as interrelated.

As is now well known, there has been an expanded use of digital technologies within processes designed to verify and evaluate evidence, and to adjudicate claims made by citizens regarding, for example, their eligibility for a ser-

4 The concept of 'bent testimony' has been developed to capture scenarios in which, because of disputed norms of communication, we may treat search algorithms as if they were asserting true content at the top of a search page – even in the absence of assertion (Rini 2017; Narayanan/De Cremer 2022).

vice or benefit. Testimony often features in these systems in numerous complex ways, often in combination. Below I cite several examples of AI technologies applied in public policy settings – including in migration, social security and policing – and describe their treatment of testimony within the three-part typology referred to above. This is not intended to be exhaustive but rather to highlight that analysis of the epistemic properties and ethical stakes involved in the use of AI systems requires consideration of their purposes and operations. Whilst these three categories are not entirely discrete, each corresponds to a specific evidentiary manoeuvre. In the first (obviation), testimony is largely discarded as a source of evidence. In the second (diminishment), testimony is not discarded but rather devalued in the face of alternative or contradictory evidence. In the third (impugnment), there may be no alternative or contradictory evidence at all, and the system is used instead to evaluate the quality of testimonial evidence.

2.1 Obviating testimony

In this grouping, direct testimony is largely sidelined. Rather than eliciting self-reported accounts of a subject's identity, biography or predicament, these systems tend to generate or extract information about a subject without their direct input and/or awareness. In such cases, sources of knowledge justification other than testimony are used. This may include biometric information derived using computer vision systems which purport to recognize an individual's intimate characteristics such as sexuality, personality traits or emotional state. This information is often based on analysis of data inputs deriving from the body (such as gait, facial geometry, eye gaze), and electroencephalogram (EEG) (Seo/Laine/Sohn 2019), and training data based on the socially conditioned human perceptions of these traits. For the most part, these systems have been developed in research settings, and have not been deployed by democratic governments. Many have been described as pseudoscientific and a form of physiognomy (Stark/Hutson 2022; Kaltheuner 2021). However, even less controversial classification systems, such as those which claim to identify a person's gender, rely on the notion that "social systems of classification have an essence; an intrinsic, biological substrate that can be detected via the face once-and-for-all with the assistance of 'objective' automated tools" (Scheuerman/Pape/Hanna 2021).

As well as AI systems deployed to extract information directly from the body, many AI systems make inferences about a person based on biographical and circumstantial details. Angwin et al. (2016) investigated use of such a system by state authorities in the United States which aimed to predict the recidivism risk associated with specific individuals based on scores derived from 137 questions answered by defendants or taken from their criminal records. They found significant racial bias. As well as inference, in this type of encounter there is also a strong element of memory as an epistemic source, as it is heavily dependent on data related to a person which is stored and then retrieved. In this scenario, even though subjects are being asked for their direct input to complement information already held about them – that is, they are allowed to speak – the process is conducted in a highly mechanical and circumscribed way. Participants may be able to guess how their answers will effectively count against them and expose them to harsher treatment, but they are not informed about how their responses will be analysed or what inferences are being made about them. Even if there is a small element of verbal testimony, the exchange is highly imbalanced, and respondents are not given the opportunity to know what it is they are being asked and why. As Hildebrandt notes, being profiled by a machine, without access to the knowledge used to categorize us, means we "cannot adequately anticipate the actions of those that know about us what we may not know about ourselves" (Hildebrandt 2008: 17). A person being asked, for example, simply how often they used to get into fights at school is not able to give contextual information or to know which unresolved question is regarded by the hearers as in need of evidence (and so their answer would not qualify as testimony in the schema outlined by Graham [1997] above). The opacity of many AI systems and their outputs can effectively curtail opportunities for subjects (and their legal representatives) to review decisions (Cobbe 2019) or to adapt legal and advocacy strategies based on clear decision-making criteria. The latter problem arises particularly where statistical and logical models may be manipulated by public servants, leading to irrational outcomes (Koulish/Evans 2021).

In the examples of systems which deal with testimony in ways that effectively bypass it in favour of perception, memory and inference, it may not be immediately obvious where the epistemic injustice arises. If testimony is not being elicited and someone is not engaged in their capacity as a knower, one might assume no epistemic harm is done. But it is in these situations that a particularly grave form of epistemic injustice may occur. AI systems which generate and process information about a person – but which do not allow for acknowledgement of them as a knowing subject – produce what many would regard as epistemic objectification. That is, people are treated as sources of information, rather than informants. Here, our knowledge of ourselves and the world is not so much regarded as unreliable as it is deemed superfluous, unfathomable or even irrelevant.

The above describes a failure (by public authorities) to respect the dignity and autonomy of the targets of AI systems *in the process of producing knowledge*. However, epistemic harms cannot be separated from the character of *the knowledge itself*. In many cases, particularly where machine learning is used, the knowledge production enabled by an AI system is based on probability, not actuality. Rather than being designed to enable the formation of a genuine belief about a person, the purpose is to produce calculative inferences about that person which can serve as the basis for action. This type of knowledge is based on prediction rather than theorization and is less concerned with whether something is true than with the cost of acting as if it were so (Joque 2022). As Origgi and Ciranna put it, people are "epistemically diminished as individual knowers: the knowledge of themselves objectifies them in a new way. Their identity becomes a virtual object, a 'statistical double'" (2017: 305).

Vallor (2021) argues that certain AI technologies are being developed to 'scrape' the body for unconsented and supposedly unmediated emotional 'truth', in a way that mirrors the *basanos* of Ancient Greece (a method of extracting truthful testimony through torture from the bodies of enslaved people). Such technologies, which include emotion recognition systems, are typically deployed against those without the power or knowledge needed to consent (and, conversely, to refuse). In Vallor's reading, rather than testimony being ignored entirely, there is a displacement from speech and communication toward the body as a site of truth.

Departing from *objectification* as the lens for understanding the wrongs of epistemic injustice, Cusick (2019) argues that the concept of *derivization* is more productive and precise. Here, following Cahill (2011) and Pohlhaus (2014), the core wrong is not the treating of others as mere (bodily) objects (as objectification entails), but the "active, willful misinterpretation of the evidence from victims' own bodies and lives and a derivatizing of them as persons for others rather than for themselves" (Cusick 2019: 112). This allows analysis of how bodies (and not just words) are treated as sources of information to serve others' ends. Bypassing verbal testimony to read information from a body is not always wrong (for example, in some medical settings). However, it can become wrong when listeners treat themselves as the only active participants in testimonial exchanges. Derivization, then, may occur when AI systems are used in ways that obviate verbal testimony to treat persons as for others rather than for themselves. Though the possibility of derivization is already present in adversarial adjudicative settings without the introduction of AI, the intersubjective qualities of decision-making (Bergman Blix 2022) are reduced, or even eliminated, by automation. AI limits potential opportunities for subjects and their representatives to intervene with more narrative accounts and context-driven argumentation.

2.2 Diminishing testimony

Critics of techno-solutionist thinking contend that it results in over-reliance and excessive trust in quantified methods and outputs. In this vein, Broussard (2019) identifies a pervasive tendency to assume that technological solutions are inherently better than their alternatives, a tendency she calls *technochauvinism*. What is less theorized, however, is the extent to which the excessive credibility accorded to and reliance on AI systems and their outputs correspond (necessarily or in effect) with a deflation of the testimony of their targets.

Empirical studies into decision-making in human–algorithm interaction have identified distinct biases in how outputs are processed by decision-makers. Automation bias, according to Alon-Barkat and Busuioc (2022), consists in the "human propensity to automatically defer to automated systems, despite warning signals or contradictory information from other sources" (2022: 2). The British Post Office scandal provides an illustrative example of the potential consequences of excessive trust in or reliance on automating digital technology, including how this may diminish the relative position of testimony. Over the course of several years, hundreds of workers were wrongly prosecuted for theft, false accounting and fraud because of discrepancies arising from a flawed software system. The courts later found that the Horizon system was wrongly represented as reliable and its outputs as incontrovertible – effectively reversing the burden of proof so that the onus was on the defendants to prove that no losses had occurred (Wallis 2021). Whether or not all actors truly believe in the integrity and veracity of their outputs, we can identify discursive processes that give AI-enabled systems a type of social power (Beer 2017) and that in turn may make it easier for actors to deliberately rely on systems which appear to be flawed or for which malfunction is foreseeable. Where computational systems are widely regarded (or at least treated) as embodying a special, authoritative form of reliability or trustworthiness, any countervailing testimonial evidence will hold less power.

Distinct from automation bias, *selective adherence* is the propensity "to adopt algorithmic advice selectively, when it matches pre-existing stereotypes about decision subjects (e.g., when predicting high risk for members of negatively stereotyped minority groups)" (Alon-Barkat/Busuioc 2022: 2). Whilst automation bias has been identified in studies in social psychology, in this study it was found that public servants do not necessarily tend to defer to automated system outputs (Alon-Barkat/Busuioc 2022). However, the authors found that decision- makers may be likely to adhere to decision recommendations when they align with existing group stereotypes and disadvantage minority groups (hence *selective adherence*). This was identified in the case of the infamous discriminatory AI enabled decision-making system which was used by the Dutch government to flag potentially incorrect or fraudulent childcare benefit claims. Decision-makers operated with the perverse incentive of ensuring more money would be retrieved through the scheme than the system itself would cost (Amnesty International 2021).

As we have seen, the classic case of testimonial injustice put forward by Fricker is one of identity-based credibility deficit i.e., a situation in which a person's testimony is accorded decreased (or, in some cases, increased) credibility not because of any relevant factors, but on the basis of prejudices. In Fricker's original theorization of testimonial epistemic injustice, interpersonal discrimination within localized interactions was taken as a basic premise. However, if we follow subsequent contributions (Coady 2017; Medina 2011) in viewing credibility as a relational or distributional (and even finite) epistemic good, excessive trust or reliance in technologies and their outputs may engender changes to the handling of subjects' testimony. In such cases, testimonial injustice arises from the assumption that technological outputs are more valid – as in automation bias (Alon-Barkat/Busuioc 2022; Symons/Alvarado 2022). Where decisions confer access to finite resources or are constrained by management targets and incentives, decision-makers may also experience pressure to adjust their evidentiary burden to, for example, allow only the most straightforward or convincing claims to succeed, or to lower the thresholds for flagging potential fraud. In this way, the superficial *objectivity* (Porter 1995) of AI systems may generate credibility and (at least temporarily) ward off scrutiny. As this reminds us, the credibility given to quantified methods and reasoning is not just an effect of perceptions about technological capability and reliability – it is deeply political (Rose 1991). As Alon-Barkat and Busuioc (2022) suggest, in the wake of high-profile events such as the Dutch childcare benefit fiasco, public servants' attitudes and behavior are likely to reflect greater scepticism and even diligence with regard to AI decision-making tools.

2.3 Impugning testimony

In some cases, AI tools are used to directly evaluate and verify the integrity of a speaker's testimony. This is subtly different from diminishment in that it relates to the quality of the testimony, rather than how much weight should be attached to it. This treatment of testimony is typified by the iBorderCtrl project which aimed to develop systems with the ability to perform "deception detection" based on facial recognition technology and the measurement of micro-expressions by finding so-called "biomarkers of deceit" on the bodies of people attempting to cross borders. The controversial project – funded by the European Commission (European Commission 2022) – was designed with the aim of speeding up border control of third-country nationals crossing EU borders by providing authorities with information to support decision-making. Psychologists have, however, refuted the premise of such a system and argued that, without definitive and reliable cues to deception, its validity is highly questionable (Jupe and Keatley 2019).

Sánchez-Monedero and Dencik (2020) situate iBorderCtrl within a lineage of lie detection and data-driven deception detection technologies, arguing that these systems perform not just technical but distinctly political functions. Other systems which use behavioral data (such as keystrokes) or language to assess credibility and detect fraud have been developed and continue to attract investment (Bittle 2020). Although many of them only flag cases for further attention – rather than offering a final determination of honesty or deception – their use signals an intensification of the level of scrutiny applied to testimony within individualized decision-making. Despite well-grounded concerns about their validity and accuracy, these systems at least claim to make available a wider range of behaviors, physical features and communicative traces which can be used to test the credibility of testimony offered by their targets.

Many other instances of decision-making in public policy are characterized by suspicion and high burdens of proof. This is particularly apparent in asylum adjudication processes, where applications often hinge on the credibility accorded to the asylum seeker's account of their own identity and biography. Information including state-produced country reports is often accorded more authority than the testimony of the asylum seeker and is used to impugn their narrative accounts (Haas and Shuman 2019). This has been accompanied, at times, by widespread use of invasive and rights-violating practices to determine credibility – particularly in asylum applications related to sexual orientation or gender identity (Spijkerboer 2013). As this suggests, the introduction of digital technologies to assess credibility often takes place against a backdrop of unjust epistemic practices by public authorities which impugn the accounts of individuals (Sertler 2018).

As noted already, a point of disagreement amongst scholars about the nature of testimonial injustice is whether credibility can be assessed by viewing one speaker at a time. According to Fricker's original theorization, credibility is not a finite resource and so credibility *deficit* is not *inversely proportional to credibility excess; it can be looked at in relative isolation.* Medina (2013), however, convincingly argues that we cannot separate subjects from their own social positionality. This is because our judgements about what is "normal" tend to take shape in "comparative and contrastive" ways (Medina 2013: 66) – for example, queer identities becoming known in part through a series of oppositional binaries (Sedgwick 2008). For this reason, a subject may be unfairly assessed as lacking credibility largely because of a comparison with the normal (Medina 2013: 63). Indeed, this process of comparison between objects is a core feature of many models based on machine learning through which "the 'other' is algorithmically produced as anomaly" (Aradau/Blanke 2018: 1).

In systems which purport to detect deception using machine-learning algorithms trained from large datasets, epistemic injustice may also arise from the exploitation of one's data – even when accorded the appropriate level of credibility. If my testimony is justifiably deemed credible and personal data about me (such as my facial expression at the time) is then then used to help build a model to impugn the testimony of others, we might think of this unconsenting exchange as a kind of *epistemic extraction* (Pasquinelli/Joler 2021). People whose data correspond to what is (or will be) deemed 'normal' are compelled to offer testimony that may be combined and used to train an AI system which could automate and reinforce further epistemic and material harms. Here, the qualities of 'credible' testimony are alienated from its content, and from the speaker, and used to generate markers of credibility through ongoing calculations of similarity and anomaly. As a result, it becomes impossible to refuse testimonial participation in the production of unjust determinations and decisions. AI systems which are used to impugn testimony therefore add to the epistemic injustices already present in, for example, asylum adjudication processes by making these contrastive relationships between excess and deficit – or normal and anomalous – both more pronounced and less open to contestation.

3. Epistemic virtue and governmental AI

To address instances of epistemic injustice, a set of correctives rooted in virtue ethics has been proposed. According to Fricker, a virtuous hearer must exercise critical awareness so that they can identify the impact of identity power in their credibility judgement. This must address both the identity of the speaker and the hearer's own position and be an ongoing, reflexive process: "someone whose pattern of spontaneous credibility judgement has changed in light of past anti-prejudicial corrections and retains an ongoing responsiveness to that sort of experience" (Fricker 2007: 97). This set of sensibilities, taken together, makes up the hybrid virtue of testimonial justice.[5]

As for the corrective to hermeneutical injustice, a virtue is again proposed – *hermeneutical justice*. However, the task of overcoming unequal power relations that cause situations of hermeneutical marginalization (and injustice), Fricker admits, "takes more than virtuous individual conduct of any kind; it takes group political action for social change" (2007: 174). Medina (2013) takes up this suggestion, arguing that in order to work toward epistemic justice, we must cultivate sensibilities that encourage us to be open to, and actively in search of, sources of contestation and epistemic friction. To counteract the epistemic vices of arrogance, laziness and closed-mindedness, virtues of humility, diligence and open-mindedness are proposed instead.

Epistemic injustice cannot be separated from unjust social structures more broadly – including, for example, racism, socioeconomic inequalities and disablement. For Medina (2011), it is at the level of the *social imaginary* that epistemic injustice is most deeply rooted, and countering it is largely a task of re-imagination and of determining what can count as epistemic alternatives. Credibility deficits affecting marginalized groups, then, are not always underpinned by some form of prejudice; there are several *structural* causes. These include differential access to markers of credibility, particularly through educational and distributive inequalities. Examining only the local properties of interactions and decisions therefore may not reveal all injustices. This is because cognitive biases can be *transactionally* innocent but nevertheless act as vehicles for the spread of *structural* injustices. As Anderson argues, just as individuals are accountable for how they act independently, they are accountable

5 Correcting for prejudice is hybrid because it is necessary in order both not to miss out on the truth and to avoid commission of unjust practices to a person in their capacity as a knower, according to Fricker (2007).

for how they act collectively: "Epistemic virtue is needed at both individual and structural scales" (Anderson 2012: 171). She argues for a move away from what she regards as Fricker's preoccupation with individual epistemic virtue toward a consideration of epistemic justice as a virtue of what she terms *social institutions*. Institutions, she argues, may have the power to prevent or correct problems that virtuous individuals cannot solve on their own – particularly where unjust outcomes result from complex and cumulative inputs and operations. Others have similarly argued that organizations can cultivate virtues of open-mindedness, courage, integrity and humility (Choo 2016). However, from a social scientific perspective, *pace* Anderson, the concept of institution ought not to be used synonymously with either *organization* or *social structure* (Fleetwood 2008). One must also be careful not to commit what Archer (1995) calls *upwards conflation*, i.e., the attribution of causal efficacy only to agents but not to structure (which is instead conceived of as a mere aggregation of agentive forces).

Nevertheless, if epistemic virtues can and should be cultivated as collective and organizational values, we can inquire into how this might extend to governing arrangements involving AI to support decision-making. In the following section, I show that efforts to cultivate virtues such as open-mindedness and empathy as a means of regulating the operation and minimizing negative effects will encounter four key constraints. These constraints emanate from the complex embeddedness of AI in governing practices (Carmel 2019). AI cannot be understood as a type of isolated, standalone product or service. Rather, the ethical and political implications will vary greatly depending on the context, purpose(s), operation and use. Where the responsible parties procuring, deploying and operating systems are public authorities, the effects on the relationships between citizens and the state, in particular, are significant. I set out four dimensions that require consideration when addressing the epistemic injustices outlined above. These relate to: the political economy of AI adoption; the distribution of agency and responsibility; the nature of bureaucratic rule; and decision-making knowledge in government.

3.1 The political economy of AI adoption

Though many thinkers already agree that epistemic injustices are a structural problem and, as such, in need of a structural response (Samaržija/Cerovac 2021), this proposition alone does not identify which structures are relevant and in which contexts. While it is widely recognized that social structures, including oppressive social relations such as those of race (Mills 2017) and gender, shape epistemic injustices, there is a need to attend at the same time to specific political economic contexts. These technologies which reconfigure the role of testimony within the exercise of public power are not outcomes of discrete and spontaneous policy decisions. They emerge under particular conditions, with particular political rationales, and are often constituted by complex interactions between public and private interests. This further complicates the task of imbuing systems, their use, and their oversight and accountability mechanisms with epistemic virtue. Attending only to epistemic practices leaves unaddressed many *extra-epistemic* forces involved in the production of epistemic harms outlined in the typology above (obviation, diminishment, impugnment). This is where democratic politics (rather than epistemic virtue alone) may be a more suitable means of addressing injustices.

3.2 The distribution of agency and responsibility

Much thinking and writing on epistemic justice proceeds on the basis that some knowers are dominantly situated in relation to others. It is widely accepted that technologies are never apolitical or value-neutral (Winner 1980), and AI's propensity to reproduce social inequalities is well documented (O'Neil 2017; Noble 2018; Buolamwini/Gebru 2018). However, where beliefs are formed – or conclusions established – not simply by individuals who hear testimony and weigh it up with other sources of knowledge, but in processes distributed across sociotechnical systems, questions of domination are less clear cut. Digitalization and automation disorganize established decision-making practices and procedures in ways that may make it difficult to locate ethical and political agency and sources of harm. Ananny and Crawford (2018) highlight the *assembled* nature of algorithmic systems which, rather than being reified objects, are made up of a network of human and computational actors. Given that agency and responsibility are distributed and relational, any attempts to foster virtue must target not just developers, operators, users or the system components, but the entire assemblage. The complexity of AI supply chains complicates this task. Many systems are derived from multiple sources or come already embedded in goods and services. Allocation of legal responsibility throughout the AI supply chain and lifecycle is central to current debates about the regulation of AI. These considerations could be expanded to include sites (and even chains) of epistemic responsibility.

3.3 The nature of bureaucratic rule

As I have suggested, the character of the institutional entity responsible for making decisions – in most of the cases discussed, state agencies – is significant. Indeed, there is a specifically *bureaucratic* character to many of the epistemic harms discussed in this chapter. This comes with important analytical implications. As Du Gay (2000) notes, bureaucratic conduct is frequently regarded as irremediably unethical. This tendency is perhaps most acute in the work of Hannah Arendt (1969), who argued that bureaucracy can be thought of as *rule by nobody*. Similarly, in *Modernity and the Holocaust*, Bauman (1989) examined the nature of bureaucracy as representative of modernity, and as characterized by its functional division of labor and separation of the technical from the moral. Bureaucracy, for Bauman, is dehumanizing because it produces distance between the (ethical) conduct of decisions and their effects and outcomes. It has even been observed that machine learning and bureaucracy are both modes of goal-oriented, rational ordering that claim neutrality and objectivity through detached abstraction (McQuillan 2020). While we might not go as far as to claim they are coextensive, there is, as this suggests, much promise in understanding the application of machine learning techniques as cohering with bureaucratic rule. The ignorance, deflation or disbelief of citizen testimony is not simply a result of negative aggregated attitudes or inadequate hermeneutical resources. The widespread use of automating technologies means subjects are also *epistemically* disadvantaged by their *materially* subordinated position as people who rely on bureaucratic formations to meet their basic needs. It is only through extension of substantive political and socioeconomic rights, not just improved epistemic practices, that these unjust power asymmetries can be redressed.

3.4 Decision-making knowledge in government

Recent scholarship has begun to examine the knowledge practices of street-level bureaucrats (Lipsky 2010) who use algorithmic decision tools (Snow 2021), including how they exercise (or withhold) individual judgement. Ranchordás (2022) argues that *empathy* ought to be seen as a key value within administrative law, and, accordingly, that the digitalized state ought to consider multiple viewpoints as well as individual circumstances (2022: 45). She argues that there ought to be a duty to forgive errors in some cases – a duty not well served by automation and data-driven decision support. Governmental AI technologies are often accompanied by the curtailment or removal of administrative discretion (Bovens and Zouridis 2002; Eubanks 2018), and their design and development may overlook the plurality of forms and practices of knowledge involved in policy- and decision-making processes, including implicit or tacit knowledge (Polanyi 1962). Attempts to automate or support decision-making using algorithmic outputs may challenge capacities to adapt and improvise in unpredictable circumstances, and potentially subordinate idiosyncratic and uncodified ways of knowing and acting. As a result, there may be less scope for decision-makers to work with knowledge in more reflexive and virtuous ways. In other words, just as fairness cannot be automated (Wachter et al. 2021), nor can virtue.

Similarly, we cannot assume those deploying AI aim to accumulate as much knowledge as possible, or that such knowledge has a direct correlation with power. The work of Linsey McGoey (2012), and that of other scholars concerned with the sociology of ignorance, has illuminated the ways in which "strategic unknowns" can be harnessed as resources which enable knowledge to be deflected, obscured or manipulated to expand the scope of what remains unintelligible. Institutions employ strategies to keep "uncomfortable knowledge" at bay (Rayner 2012: 107). Rather than being the result of a flawed design or ineffective decision-making, the exclusion of countervailing testimonial knowledge may be driven by strategic imperatives.

Conclusion

Much of the critical literature on the use of AI in government has detailed the ways in which the design and use of technologies that automate decision-making processes can produce significant harmful, discriminatory and rights-violating outcomes for their targets (Eubanks 2018; O'Neil 2017). As well as these individualized and collective adverse outcomes, the use of automated decision-making in the public sector is altering how power, authority and knowledge are arranged within and between institutions. Scholars have begun to argue for a shift away from narrowly conceived notions of AI ethics toward wider consideration of justice (Gabriel 2022). In this chapter, I have advanced this further by showing that the use of AI is reconfiguring how testimony is elicited, offered and received, and that this is giving rise to specifically epistemic injustices. Through the use of AI technologies, people are made legible to the state (Scott 2020) in ways that silence, derivize and extract from them. By variously obviating, diminishing and impugning the testimony of individuals targeted by decisions, these systems are not just harming them by undermining their dignity as knowing persons; they are reshaping the place of testimony in public policy, with potentially far-reaching implications for democratic citizenship.

Whereas much existing scholarship posits the cultivation of epistemic virtue as one way to prevent and address epistemic injustices, the nature of (AI-supported) governmental decision-making means this is insufficient as a mode of prevention or redress. The drivers and effects of epistemic injustice in these cases extend well beyond unjust structures of knowledge. AI systems are being used in processes already designed to be impersonal, highly rationalized and not always amenable to epistemic virtues. Furthermore, the nature of many systems precludes the identification of a single locus of bias (or epistemic vice). Whether demanded of the human in the loop or the "institution" in the loop, reflexive and virtuous practices cannot resolve these problems.

How then might we attempt to reposition testimony within governmental decision-making to prevent and address such injustices? Many of those most implicated by these practices are already politically disempowered, and digitalization and automation can produce further exclusions which limit their capacity to contest and advocate. Nevertheless, as recent political and legal mobilizations and initiatives have demonstrated, there are opportunities for collective redress, resistance and refusal (Ganesh/Moss 2022; Dent 2022), as well as formal regulation. Alongside these movements for greater democratic oversight and control of AI, various public accountability measures such as public registries and impact assessments have been proposed and implemented to different degrees (Ada Lovelace Institute et al. 2021). To address the types of injustices outlined in this chapter, however, AI policy-making and regulation must go further in centering the perspectives and guaranteeing the rights of decision subjects themselves.

Bibliography

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. Last access: 8 October 2022.

AlgorithmWatch. 2021. EU policy makers: Protect people's rights, don't narrow down the scope of the AI Act! *AlgorithmWatch*. https://algorithmwatch.org/en/statement-scope-of-eu-ai-act/. Last access: 5 October 2022.

Alon-Barkat, Saar and Madalina Busuioc. 2022. Human–AI interactions in public sector decision making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice. *Journal of Public Administration Research and Theory*. https://doi.org/10.1093/jopart/muac007.

Alvarado, Rafael and Paul Humphreys. 2017. Big Data, thick mediation, and representational opacity. *New Literary History* 48:729–749. http://doi.org/10.1353/nlh.2017.0037.

Amnesty International. 2021. Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal. *Amnesty International*, 25 October 2021. https://www.amnesty.org/en/documents/eur35/4686/2021/en/. Last access: 3 October 2022.

Ananny, Mike and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20:973–989. https://doi.org/10.1177/1461444816676645.

Anderson, Elizabeth. 1995. Feminist epistemology: An interpretation and a defense. *Hypatia* 10:50–84. http://doi.org/10.1111/j.1527-2001.1995.tb00737.x.

Anderson, Elizabeth. 2012. Epistemic justice as a virtue of social institutions. *Social Epistemology* 26:163–173. https://www.tandfonline.com/doi/abs/10.1080/02691728.2011.652211.

Angwin, Julia, Jeff Larson, Surya Mattu and Lauren Kirchner. 2016. Machine bias. *ProPublica*, 23 May 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Last access: 3 October 2022.

Aradau, Claudia and Tobias Blanke. 2018. Governing others: Anomaly and the algorithmic subject of security. *European Journal of International Security* 3(1):1–21. https://doi.org/10.1017/eis.2017.14.

Archer, Margaret. 1995. *Realist Social Theory: The Morphogenetic Approach*. Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9780511557675.

Arendt, Hannah. 1969. A special supplement: Reflections on violence, *New York Review*. 27 February 1969. https://www.nybooks.com/articles/1969/02/27/a-special-supplement-reflections-on-violence/. Last access: 3 October 2022.

Audi, Robert. 1997. The place of testimony in the fabric of knowledge and justification. *American Philosophical Quarterly* 34:405–422.

Bauman, Zygmunt. 1989. *Modernity and the Holocaust.* Cambridge: Polity.

Beer, David. 2017. The social power of algorithms. *Information, Communication & Society* 20**:**1–13. https://doi.org/10.1080/1369118X.2016.1216147.

Bergman Blix, Stina. 2022. Making independent decisions together: Rational emotions in legal adjudication. *Symbolic Interaction* 45(1):50–71. https://doi.org/10.1002/symb.549.

Berryhill, Jamie, Kévin Kok Heang, Rob Clogher and Keegan McBride. 2019. Hello, World: Artificial intelligence and its use in the public sector. *OECD Working Papers on Public Governance,* 36. 21 November 2019. https://doi.org/10.1787/726fd39d-en.

Bertuzzi, Luca. 2022. AI Act: Czech Presidency pushes narrower AI definition, shorter high-risk list. *Euractiv,* 18 July 2022. https://www.euractiv.com/section/digital/news/ai-act-czech-presidency-pushes-narrower-ai-definition-shorter-high-risk-list/. Last access: 5 October 2022.

Bittle, Jake. 2020. Lie detectors have always been suspect. AI has made the problem worse. *MIT Technology Review.* 13 March 2020. https://www.technologyreview.com/2020/03/13/905323/ai-lie-detectors-polygraph-silent-talker-iborderctrl-converus-neuroid/. Last access: 5 October 2022.

Bovens, Mark and Stavros Zouridis. 2002. From street-level to system-level bureaucracies: How Information and Communication Technology is transforming administrative discretion and constitutional control. *Public Administration Review* 62**:**174–184. https://doi.org/10.1111/0033-3352.00168.

Broussard, Meredith. 2019. *Artificial Unintelligence: How Computers Misunderstand the World.* Cambridge, Mass.: MIT Press.

Buolamwini, Joyce and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research* 81:1–15. 2018 Conference on Fairness, Accountability, and Transparency. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf. Last access: 20 December 2022.

Cahill, Ann J. 2011. *Overcoming Objectification: A Carnal Ethics.* New York: Routledge.

Carmel, Emma. 2019. Introduction to Governance Analysis: Critical Enquiry at the Intersection of Politics, Policy and Society. In *Governance Analysis: Critical Enquiry at the Intersection of Politics, Policy and Society*, Ed. Emma Carmel, 2–24. Cheltenham: Edward Elgar Publishing.

Choo, Chun Wei. 2016. *The Inquiring Organization: How Organizations Acquire Knowledge and Seek Information.* New York: Oxford University Press.

Coady, Cecil Anthony John. 1992. *Testimony: A Philosophical Study*. Oxford: Oxford University Press.

Coady, David. 2017. Epistemic Injustice as Distributive Injustice. In *The Routledge Handbook of Epistemic Injustice,* Eds. Ian James Kidd, José Medina and Gaile Pohlhaus, 61–68. London: Routledge.

Cobbe, Jennifer. 2019. Administrative law and the machines of government: Judicial review of automated public-sector decision-making. *Legal Studies* 39(4):636–655. http://doi.org/10.1017/lst.2019.9.

Cusick, Carolyn, M. 2019. Testifying bodies: Testimonial injustice as derivatization. *Social Epistemology* 33:111–123. https://doi.org/10.1080/02691728.2019.1577919.

Dent, Anna. 2022. Disabled benefits claimants are being unfairly investigated. *Huck*, 16 March 2022. https://www.huckmag.com/perspectives/disabled-benefits-claimant-are-being-unfairly-investigated. Last access: 5 October 2022.

Dotson, Kristie. 2011. Tracking epistemic violence, tracking practices of silencing. *Hypatia.* 26:236–257. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1527-2001.2011.01177.x. Last access: 20 December 2022.

Du Gay, Paul. 2000. *In Praise of Bureaucracy: Weber, Organization, Ethics.* London: SAGE Publications Ltd.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor.* New York: St Martin's Press.

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*. COM/2021/206 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206. Last access: 5 October 2022.

European Commission. 2022. Intelligent Portable Border Control System. *Community Research and Development Information Service (CORDIS).* https://doi.org/10.3030/700626. Last access: 20 December 2022.

Felzmann, Heike, Eduard Fosch-Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. 2020. Towards transparency by design for Artificial Intelligence. *Science and Engineering Ethics* 26:3333–3361. https://doi.org/10.1007/s11948-020-00276-4.

Fleetwood, Steve. 2008. Institutions and social structures. *Journal for the Theory of Social Behaviour* 38:241–265. https://doi.org/10.1111/j.1468-5914.2008.00370.x.

Freiman, Ori and Boaz Miller. 2020. Can Artificial Entities Assert?. In *The Oxford Handbook of Assertion*, Eds. Sanford Goldberg, 414–434. Oxford: Oxford University Press.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Gabriel, Iason. 2022. Toward a theory of justice for artificial intelligence. *Daedalus* 151(2):218–231. https://doi.org/10.1162/daed_a_01911.

Ganesh, Maya Indira and Emmanuel Mossl. 2022. Resistance and refusal to algorithmic harms: Varieties of 'knowledge projects'. *Media International Australia* 181(1):90–106. https://doi.org/10.1177/1329878X221076288.

Garnelo, Marta and Murray Shanahan. 2019. Reconciling deep learning with symbolic artificial intelligence: representing objects and relations. *Current Opinion in Behavioral Sciences* 19:17–23. https://doi.org/10.1016/j.cobeha.2018.12.010.

Graham, Peter J. 1997. What Is testimony? *The Philosophical Quarterly* 47: 227–232. https://doi.org/10.1111/1467-9213.00057.

Haas, Bridget M. and Amy Shuman. 2019. *Technologies of Suspicion and the Ethics of Obligation in Political Asylum*. Athens, Ohio: Ohio University Press.

Henman, Paul. 2010. *Governing Electronically: e-government and the Reconfiguration of Public Administration, Policy, and Power.* Basingstoke: Palgrave Macmillan.

Hildebrandt, Mireille. 2008. Defining Profiling: A New Type of Knowledge? In *Profiling the European Citizen: Cross-Disciplinary Perspectives,* Eds. Mireille Hildebrandt and Serge Gutwirth, 17–30. Dordrecht: Springer Netherlands.

Hill Collins, Patricia. 1998. *Fighting Words: Black Women and the Search for Justice.* Minneapolis and London, University of Minnesota Press.

Humphreys, Paul. 2004. *Extending Ourselves: Computational Science, Empiricism, and Scientific Method.* Oxford: Oxford University Press.

ITU. 2021. *United Nations Activities on Artificial Intelligence (AI).* http://handle.itu.int/11.1002/pub/81b35851-en. Last access: 5 October 2022.

Joque, Justin. 2022. *Revolutionary Mathematics: Artificial Intelligence, Statistics and the Logic of Capitalism*. New York: Verso.

Jupe, Louise M. and David A. Keatley. 2019. Airport artificial intelligence can detect deception: or am I lying? *Security Journal* 33:622–635. https://doi.org/10.1057/s41284-019-00204-7.

Kaltheuner, Frederike. 2021. *Fake AI.* Manchester: Meatspace Press.

Koulish, Robert and Kate Evans. 2021. Punishing with impunity: The legacy of risk classification assessment in immigration detention. *Georgetown Immigration Law Journal* 36:1–72.

Lipsky, Michael. 2010. *Street-level Bureaucracy: Dilemmas of the Individual in Public Services.* New York: Russell Sage Foundation.

Maynard-Moody, Steven and Michael Musheno. 2022. *Cops, Teachers, Counselors: Stories from the Front Lines of Public Service* (2nd edition). Ann Arbor, Mich: University of Michigan Press. https://doi.org/10.3998/mpub.12247078.

McGoey, Linsey. 2012. Strategic unknowns: towards a sociology of ignorance. *Economy and Society* 41:1–16. https://doi.org/10.1080/03085147.2011.637330.

McQuillan, Dan. 2020. Deep Bureaucracy and Autonomist AI. In *Deserting from the Culture Wars,* Eds. Maria Hlavajova and Sven Lütticken. Cambridge, Mass.: MIT Press.

Medina, José. 2011. The relevance of credibility excess in a proportional view of epistemic injustice: Differential epistemic authority and the social imaginary. *Social Epistemology* 25:15–35. https://doi.org/10.1080/02691728.2010.534568.

Medina, José. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations.* Oxford: Oxford University Press.

Mills, Charles W. 2017. *Black Rights/White Wrongs: The Critique of Racial Liberalism.* New York: Oxford University Press.

Narayanan, Devesh and David De Cremer,. 2022. “Google told me so!” On the bent testimony of search engine algorithms. *Philosophy & Technology* 35:22. https://doi.org/10.1007/s13347-022-00521-7.

Noble, Safiya U. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York: New York University Press.

O’Neil, Cathy. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* New York: Broadway Books.

Origgi, Gloria and Serena Ciranna. 2017. Epistemic Injustice: The Case of Digital Environments. In *The Routledge Handbook of Epistemic Injustice,* Eds. Ian J. Kidd, José Medina and Gaile Pohlhaus. New York: Routledge.

Pasquinelli, Matteo and Joler, Vladan. 2021. The Nooscope manifested: AI as instrument of knowledge extractivism. AI and Society 36:1263–1280. https://doi.org/10.1007/s00146-020-01097-6.

Pohlhaus, Gaile. 2014. Discerning the primary epistemic harm in cases of testimonial injustice. *Social Epistemology* 28:99–114.

Polanyi, Michael. 1962. *Personal Knowledge: Towards a Post-critical Philosophy.* London: Routledge & Kegan Paul.

Porter, Theodore. M. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, N.J.: Princeton University Press.

Ranchordás, S. 2022. Empathy in the digital administrative state. *Duke Law Journal* 71(6):1341–1389.

Rayner, Steve. 2012. Uncomfortable knowledge: the social construction of ignorance in science and environmental policy discourses. *Economy and Society* 41:107–125. https://doi.org/10.1080/03085147.2011.637335.

Rini, Regina. 2017. Fake news and partisan epistemology. *Kennedy Institute of Ethics Journal* 27(2): E-43–E-64. http://doi.org/10.1353/ken.2017.0025.

Rose, Nikolas. 1991. Governing by numbers: Figuring out democracy. *Accounting, Organizations and Society* 16:673–692. https://doi.org/10.1016/0361-3682(91)90019-B.

Rosenberger, Robert. 2014. Multistability and the agency of mundane artifacts: from speed bumps to subway benches. *Human Studies* 37:369–392. https://doi.org/10.1007/s10746-014-9317-1.

Samaržija, Hana and Ivan Cerovac. 2021. The institutional preconditions of epistemic justice. *Social Epistemology* 35:621–635. https://doi.org/10.1080/02691728.2021.1919238.

Sánchez-Monedero, Javier and Lina Dencik. 2020. The politics of deceptive borders: 'biomarkers of deceit' and the case of iBorderCtrl. *Information, Communication & Society* 25(3):413–430. https://doi.org/10.1080/1369118X.2020.1792530.

Sartori, Laura and Andreas Theodorou Andreas. 2022. A sociotechnical perspective for the future of AI: narratives, inequalities, and human control. *Ethics Information Technology* 24(4). https://doi.org/10.1007/s10676-022-09624-3.

Scheuerman, Morgan Klaus, Madeleine Pape and Alex Hanna. 2021. Auto-essentialization: Gender in automated facial analysis as extended colonial project. *Big Data & Society* 8(2). https://doi.org/10.1177/20539517211053712.

Scott, James C. 2020. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven, Conn: Yale University Press.

Sedgwick, Eve Kofosky. 2008. *Epistemology of the Closet.* Berkeley: University of California Press.

Seo, Jungryul, Teemu H. Laine and Kyung-Ah Sohn. 2019. Machine learning approaches for boredom classification using EEG. *Journal of Ambient Intelli-*

gence and Humanized Computing 10:3831–3846. https://doi.org/10.1007/s12652-019-01196-3.
Sertler, Ezgi. 2018. The institution of gender-based asylum and epistemic injustice: A structural limit. *Feminist Philosophy Quarterly* 4(3). https://doi.org/10.5206/fpq/2018.3.5775.
Snow, Thea. 2021. From satisficing to artificing: The evolution of administrative decision-making in the age of the algorithm. *Data & Policy* 3, E3. http://doi.org/10.1017/dap.2020.25.
Spijkerboer, Thomas (Ed). 2013. *Fleeing Homophobia: Sexual Orientation, Gender Identity and Asylum*. Abingdon: Routledge.
Spivak, Gayatri C. 1987. *In Other Worlds: Essays in Cultural Politics.* New York: Methuen.
Stark, Luke and Jevan Hutson. 2022. Physiognomic artificial intelligence. *Fordham Intellectual Property, Media & Entertainment Law Journal* 32:922–978.
Symons, John and Ramón Alvarado. 2022. Epistemic injustice and data science technologies. *Synthese* 200, 87. https://doi.org/10.1007/s11229-022-03631-z.
Vallor, Shannon. 2021. The digital basanos: AI and the virtue of and violence of truth-telling. *2021 IEEE International Symposium on Technology and Society (ISTAS)*. https://www.doi.org/10.1109/ISTAS52410.2021.9629137.
Veale, Michael and Irina Brass. 2019. Administration by Algorithm? Public Management Meets Public Sector Machine Learning. In *Algorithmic Regulation*, 121–C6.P200, Eds. Karen Yeung and Martin Lodge. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198838494.003.0006.
Wachter, Sandra, Brent Mittelstadt and Chris Russell. 2021. Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review* 41:105567. https://doi.org/10.1016/j.clsr.2021.105567.
Wallis, Nick. 2021. *The Great Post Office Scandal: The Fight to Expose a Multimillion Pound Scandal Which Put Innocent People in Jail.* Bath: Bath Publishing.
Winner, Langdon. 1980. Do artifacts have politics? *Daedalus* 109:121–136.