Move forward or break ranks: Workshopping re-idealized explanation obligations to foster fundamental change and resistance to oppression

Aviva de Groot, Tilburg Institute for Law, Technology, and Society (TILT), Tilburg University, The Netherlands; Research Associate at The African Centre for Epistemology and Philosophy of Science (ACEPS), Johannesburg

This paper presents a set of 'duties of care' that can be used to guide decision and explanation processes in complex socio-technical-legal environments. The duties support and strengthen existing explanation laws across legal domains. Informed by critical legal, philosophical, and technological insights, the model supports the conscientious challenges of decision makers and explainers who wish to avoid complicity in orchestrated harms. The duties support them and anyone working on fair decision making to tease out their knowledge-and-information needs and whether these are organized for, and which knowledge/deficits should raise alarm and are cause for resistance. The model was workshopped in two public agencies to test its 'fit' and usefulness in this salient decision-making sphere. This paper reports on the first impression of what it takes to bring fundamental research directly into a practice context, and offers the workshop design to anyone interested to use in their own work.

Keywords: transparency, explanation, justification, justice, epistemic injustice, racism, discrimination, resistance, ADM

Reference Format:
Aviva de Groot. 2025. Move forward or break ranks: Workshopping re-idealized explanation obligations to foster fundamental progress and resistance to oppression. In *Proceedings of the Fourth European Workshop on Algorithmic Fairness (EWAF'25)*. Proceedings of Machine Learning Research, pages.

1 Introduction

1.1 An urgent need to address old problems

The increasingly overt racist and populist political climate in the EU regulatory space should raise concerns about the governance of 'AI' and automation practices, policies and ambitions. Efforts of recent years to improve Member States' and EU policies in spheres like immigration[77], the digital welfare state [52] [78] and policing[22], informed by decades of hard-won evidence of automated racism, discrimination and marginalization stand to be undermined. What becomes of hopes that those in power will act on the findings, recommendations, and judicial achievements of lawyers, scholars, activists, journalists and NGO's?[76] This paper argues to let these concerns inform fundamental equity and justice orientations of the 'algorithmic fairness' project by addressing public servants in research and outreach directly where-ever possible. Government employees, committed to their oath to uphold public and constitutional values face tough conscientious questions. How to serve the public when populism rules? What does it mean to uphold constitutional values if racist policies do not breach constitutions or even human rights frameworks or, when they *are* found to do so, simply continue?[72] With regard to automation specifically, straightforward

Authors' Contact Information: Aviva de Groot, Tilburg University, The Netherlands, aviva.degroot@tilburguniversity.edu

EWAF'25, June 30–July 02, 2025, Eindhoven, NL

guidance can be hard to find in the complex regulatory landscape comprised of tech-agnostic legal norms and a multitude of (local to supra-national) ethical frameworks, detailed 'digital' laws and now an AI Act that even struggles to define its subject usefully.[74] Rather than adding more norms to this, this paper argues to support these key players with research output that helps them to use (and where necessary, refuse) the available guidance in better, critical ways. The space to do this becomes visible by zooming out, rather than in on tech-oriented governance. If the current political climate reveals anything, it is the need for fundamental change at the ideological levels of technology *and* the ethics and (human) rights spheres that are relied on to protect people from harm.[5, 32, 47]

1.2 Bringing fundamental critical guidance to local policy levels: the real complexity

The force of law to protect against technologically supported, institutionalized state harms is historically weak, not least in the European 'value space.'[28] The EU's seeming endorsement of the industry's 'novelty and complexity' narrative with regard to regulating tech-driven harms is problematic for this reason, but also for how it suggests that these harms live in an entirely unregulated space–this is not true either. Before section 2 explains this further, a short engagement with what can be described as the 'technological complexity argument' helps to make this paper's case for transformative engagement with existing legal guidance. Claims of technological complexity are a heritage way to exclude public scrutiny and divert regulatory efforts, especially in the sphere of explanation and justification. Tech-companies and physicians long before them have used the complexity argument to keep their practices 'free' from effective oversight, retain self-governance and standard setting, and exclude so-called 'laymen' from these spheres – saliently, groups that are negatively affected by their output. [27:260, 35] At the local level of public decision making, what has in fact become complex is the sheer amount of ADM tools and data scatterings, tech contracts, people involved in and on decision chains[7, 53]; and simultaneously applicable legal and ethical frameworks. This landscape has been 'topped up' with a layer of progressive guidance consisting of not-yet-implemented (or simply ignored) critical reports from e.g., NGO's, investigative journalism and commercial auditors. This complexity itself is abused by the Dutch state that always calls for even *more* evidence of how tech causes problems, and more assessments of the legal status of these problems.[38]

This paper presents an effort to improve legal guidance 'on the ground' in the sphere of transparency, explainability and justification. It offers a model of tech-agnostic explanation duties informed by critical legal, philosophical, and technological insights. The model first of all supports the aforementioned conscientious challenges of civil servants. It helps them to tease out their knowledge-and-information needs, whether these are organized for, and what kinds of knowledge/deficits should inspire further action, possibly in the form of protest or resistance. The value of the model and the needs it caters to are of course not restricted to workers in public institutions.[61] For those working on fairness and transparency in automated decision making, the model helps to explicate the kind of things that users should be (made) aware of, adding value to any available audit, impact, or risk assessment models. For legal and ethics work and research, the model helps to draw law and ethics together to improve both. The model can be used to tease out whether the 'right' things are putatively justified and explained in any decision domain to align with what's at stake for decision subjects. In interdisciplinary work, all these functions come together. To test the contents and operability of the model, it was workshopped in two public policy spheres. This paper reports the first

impressions of what it takes to bring fundamental research directly into a practice context, and offers the workshop design to anyone interested to use it in their own work.

Of course, not all decision makers will be eager to be engaged this way. But in The Netherlands where this project is located, public servants' resistance against racist and otherwise oppressive policy is increasingly out in the open and scholarly attention for this phenomenon is revived.[10, 58, 66] Not all resistance is directly tech-related, but it does not need to be to 'hit' the subject of automation. For example, resistance against state complicity with Israel's genocidal efforts[30, 60] includes AI-sustained crimes.[31] Resistance against state ambitions in the spheres of social statistics [79] covers the kind of racist profiling it allows to do.[55] The paper proceeds as follows: Section 2 introduces the model and summarizes the background research that informed it. It discusses deficits of foundational explanation law, explains how epistemic in/justice research helps to analyze and address these deficits, and models these insights into a set of four explanation duties. Section 3 reports on the first workshops that were done with the model in a public policy context and meant to serve as a testing ground to inform later implementation trials. Section 4 concludes with the value of this work, its limitations, and calls for collaborative future research.

2 A re-idealized take on explanation law

2.1 Obscure ADM harms and the legal response: off on the wrong footing

As many scholars in the 'FAccT' fields will be aware of, the proliferation of automated decision-making technologies in Europe triggered fundamental, dignitarian arguments at various law-making moments and levels. Calling upon Europe's automation history in the Second World War[9, 33], the Council of Europe and the EU claimed that new rules were necessary to protect citizens from new totalitarian treatment in the form machine-powered management of people as if they were numbers or, more modern, 'mere temporary aggregates'[20]. Prominent among the regulatory responses were individual rights for decision subjects to know the reasons behind decisions made about them, reasons that had been disappearing from view in 'black box' systems [18, 69] Research on how to tease out these explanations and whether that was easy or impossible ensued.[11, 34, 56, 57, 68] Pre-existing rules of especially administrative law were made exemplary of what explanation law has to offer and what should be met in cases of ADM, too.[51] One could say a legal field was born.

But there was no explanation crisis. To be sure, many identified 'black box' concerns were alarming enough. But the resulting harms were neither new nor dependent on specific technologies, depending on who you would ask. Nor was law's failure to protect against them. The 'new problems' narrative in that sense itself obfuscated how the characterizations of harms represented long-term experiences of less privileged groups in societies. To summarize some main concerns: the institutional inequalities that ADM helps to bring about are true enough[3, 50], but run along pre-existing lines[2, 5, 65]. Such harms indeed elude individuals' ability claim rights in response, but the inadequacy of individual claims-based law is neither new nor accidental.[47, 67] Also called out as problematic is how inexplicable decisions insult the human ability to reason, a so-called foundation of law.[20, 62] This ignores how who counts as rational, reasoning subject in law has historically functioned as a tool to exclude many kinds of people along racist and discriminatory lines from claiming their rights.[23, 42] A next concern pertains to a growing lack of insight

into what makes up a decisional process, which is problematic for decision subjects as well as decision makers. For subjects this certainly results in a lack of safe and meaningful process participation[21] but again, this is by far not a new problem for those in less privileged societal positions. Medicine is infamous in this respect [15, 35]. Public administrations are typically on the forefront with regard to the implementation of novel data science methods, and those over whom they exert the most power are the first to be affected.[5, 9, 41] For decision makers, it is certainly true that the duty to motivate how decisions are made becomes harder (or even impossible [13]). But in light of the preceding remarks on the historical positions of explainees this can hardly be called a new problem either. Even if that *were* the case: if those charged with giving reasons are disabled at their jobs because ADM produces a lack of insight into what goes wrong, why would explanation laws allow the use of such methods in the first place?

2.2 What's missing from explanation law?

A high level, rule-of-law-democracy view on explanation places it in service of ideals of transparent governance [29]. More particularly –yet still tied to rule-based decision making– explanation laws are supposed to allow decision subjects and others (e.g., supervising bodies, judges, legal scholars) to assess whether a decision was reached in agreement with rules and other norms that govern a particular decision sphere. For this, decisions as well as the norms that govern them need to be understandable [8, 39] Some argue how proper explanation practices naturally cater to that need: explaining the application of a rule necessarily explains the interpretation of a rule and so, the rule.[54] The ethics *of* such reasoning needs to be scrutable too.[17:216] More fundamentally, and recognized in less rule-based spheres of decision making too (e.g., health care), explanation norms are attached to the attribution of decision authority including 'powers of expertise'. Legal and (referred) ethical and professional norms mean to ameliorate the power and information inequalities that exist between those invested with (decision) authority and their 'subjects' (e.g., civilians, patients, clients). Framed in terms or rights and/or duties, explanation norms are supposed to play a key role against the *abuse* of such powers in decision making. If no justification exists, decision makers can be held to account on the basis of their (bad) reasons. The other side of the fundamental coin frames a human right to explanation as something that follows from human dignity. The argument easily leads to discussions on law's (lack of) respect for some humans' dignity and (unhelpful) human essentialism which would require more space to deal with properly – the reader is kindly referred to this paper's underlying research which discusses this in detail. [27 section 2.1.3]

Across the legal domains they exist in, specific explanation rules express what lawmakers find of interest to know about decisions (or other influential conclusions, e.g., diagnoses) and what needs to be justified to whom, by who, and how. By setting these 'need to know' benchmarks explanation rules also prescribe what kind of decision processes we allow ourselves to have in the first place; in practice if not explicitly. Critical analysis of explanation rules however reveal rather limited views on what is of interest to know, limited views on the level of explainer end explainee understanding to assume and aim for, and related and other obstacles to challenge decision rules and norms behind a decision in the first place.[27] To take the Duch Administrative domain as an example: decades of reasoned administrative decision making in the Dutch Childcare benefits scandal [70, 78] illustrate the mootness of the administrative explanation paradigm in terms of the above considerations. In the Scandal, the livelihoods of many

thousands of families were irreparably harmed after being falsely accused of fraud, fined, and forced to repay benefits. The gross obscurity and injustice of the Scandal produced a constitutional crisis[71] and fueled ongoing reforms of administrative laws. From the initial decisions of the Tax Authority and other institutions that the families needed deal with to the administrative supreme courts that rules such decision making, the explanation regimes that governed the decision procedures had failed to bring sufficient relevant facts to light. In their reflections, [71] administrative court judges stated how "critical information about the knowledge and methods [the Tax Authority] used was hidden from us". They deplored how they failed to see that "the messy case files of the Tax Authority were no accident." Instead of being triggered to improve their information positions, the judges "did not investigate victims' situations well enough." Even when, in retrospect, they admit that "it was clear that something was systematically wrong, and that many more victims would follow." They consider that their motivations inevitably "reasoned away foundational principles" of proper administration. But did they? Among reasons for their own decision making at the time, they cited how they abided by legal principles such as legal certainty and equal treatment that allowed to ignore the institutional nature of the wrong-doing by treating cases as a string of incidents. Crucially, the governing administrative law principle of proportionate decision making was not applied to these cases – an interpretation of the relation between applicable laws that divided administrative scholars even after the supreme court finally reversed this interpretation.[6, 19, 59]

The rules are being reformed, but that process is slow. It is yet to be seen whether a more elaborate, stronger explanation regime will materialize. In response to (assumed) ADM explanation problems, bureaucratic inscrutability, and mixed-methods crises such as the Benefits scandal, the go-to political direction for improvement has become to re-humanize and de-formalize the public sphere. The creation of informal spheres in unequal relations however easily invites the abuse of powers. There is also no good reason to assume that more discretion will make for better decision making. As the Dutch State commission against discrimination and racism found, demonizing 'formality' misses the mark in contexts where people don't know, or don't want to know how to avoid to do the wrong thing.[75] A last set of reflections from the Benefits scandal judges supports the point for laws that are better as well as firmer with regard to justification, and that include clear red lines. Seeing things differently and acting on it, the judges reflected, required to break ranks. "Why did we not make more of a fuss?" they wondered, and "[w]hat do you do when the lawmaker wants 'A', but that conflicts with everything that you feel, think, and know. Are you allowed to, are you brave enough to decide 'B'?"[71]

2.3 Epistemic Justice for Explainees

The proper thing to do in face of continued and grave injustices that are 'not supposed to happen' under governing ideals is to start by accepting the harms as a consequence of the failure of these ideals. Starting from, and sticking with, the identified types of injustice can inform the necessary re-idealizing process. [43, 45] Even before properly starting this effort, one big 'known' can be operationalized on the basis of the discussion in the previous sections. The weakness of individual rights against institutionalized harms arguably informs to alleviate subjects' burden of proof. This entails a shift of focus from subject self-advocacy to obliging decision makers to explain how they did the right thing. The right thing being, at least, to understand how your decision-making avoids institutional wrong-doing, and

act upon such knowledge *or lack of* knowledge. And in catering to decision subjects' information positions, a richer understanding of human thriving, cognition and meaning making needs to be promoted [16:52, 44:41]

Engaging with the domains of epistemic in/justice usefully further informs what ideals to chase for in light of what explanation law currently fails to address. In short, epistemic in/justice research investigates the political nature of the practices through which knowledge is made and made use of. The core concern for social and informational, or 'power and knowledge' inequality is certainly shared in (the ideals of) explanation laws. Contrary to these laws though, the *relations* of power and knowledge are centralized and taken seriously. The literature explores how knowledge and its practices can respect or fail to respect (groups of) persons as participants in, and subjects of knowledge and decision-making along central themes such as "authority, credibility, justice, power, trust, and testimony"[37]. Research from these fields helps to understand and approach explanation as a knowledge practice that is alive to the same dynamics. In explanation, conclusions are drawn about norms, methods and information that went into decision-making such as rules, 'data', private and public interests, diagnostic concepts – in short, products of other knowledge practices. Taking this angle inevitably brings critical analyses of law itself. Among other things for how laws codify wrongful notions about people in the realm of race, sex, dis/ability, or any other (imagined) category, and how such notions are or aren't challenged over time in legal reasoning.

Back to the task of explainers: taking an epistemic approach to explanation emphasizes how explainers of decisions are always also explainees themselves. Their understanding needs with regard to what happens 'upstream' in decision processes will have to be described, and met. At the same time, they are dependent on their explainees for knowledge about things like the 'downstream' effects of decisions, and about explainees' understanding of the decision process. In the process of decision-making explainers and explainees both bring knowledge and experience to the table– but also inclinations to trust and distrust, believe or not believe each other and/or the process. In that sense one can say they are engaged in *negotiations.* [27:91] This also helps to understand explanation as an interactive process, even absent an in-person exchange. E.g., when a subject receives a written statement only by (e)mail for example, the addressee's understanding of the statement will influence not just their understanding, but also their thinking on any further course of action, such as filing for appeal. A brief further elaboration of some core concepts hopefully fosters a bit more detailed appreciation of how this literature can inform law to do better justice to explainees; not least in ADM environments. Much named in the field is the term testimonial injustice to signal how decision subjects (patients, suspects, welfare recipients) are wrongfully treated as unreliable sources of knowledge [25, 35] Analyses also investigate how knowledge practices that exclude, or otherwise fail to respect certain (groups of) people end up causing harms: such knowledge incorporates 'hermeneutical' problems.[44] An obvious example is medical research's prioritization of the white male body[49]. Translated to a medical explanation-setting, that example also warns for naïve assumptions of explainers and explainees being part of a shared knowledge sphere. The persistent problems of the medical fields also illustrate how the unchecked attribution of epistemic authority allow practitioners to surveil and experiment (on) people they don't mean to serve[14]. Histories and critiques of the fields of data science and AI point out similar vulnerabilities [5, 12, 36, 41]. Some dangerous ideological directions persistently pollute knowledge practices *and* shared knowledge spheres; eugenics is a case in point [4, 26, 73]. Failures to call this out and design adequate policy responses facilitate low maintenance oppression. Law's role in

sustaining harmful power structures within its own field and that of any other expertise is highly problematic in this sense.[28, 32:16] For those on the wrong end of 'bad knowledge', much energy goes into dealing with excessive burdens of proof towards others[63] and fighting internalized harms.[16:231]. When (groups of) people are marginalized this way, it becomes understandable (or even advisable) that they would reject knowledge, refuse to provide it, or otherwise engage in some form of epistemic resistance [44:50]. Persistent injustices in any kind of decision making (medicine, ADM, bureaucracy) understandably inspires people to stay away from services, help or benefits even when such abstinence will harm them, too. [21] This underscores the need for *institutional* efforts.[1] Recent US examples illustrate the force of institutional efforts in the negative: knowledge about (the influence of) racism and other marginalizing thought is effectively being suppressed with world-wide effects on knowledge practices. How can better explanation rules help to cultivate the necessary discomfort for explainers such as the civil servants his paper is particularly concerned with?

Inspiration for epistemically care-full explanation obligations can be found in various authors' descriptions of a pair of values that ideally govern any information-related authority. The pair includes dimensions of knowledge and dimensions of conduct. Code's *due care & intellectual honesty* [16] corresponds with Williams' later *accuracy & sincerity* (or "take care and do not lie") [64] and Fricker's' c*ompetence and trustworthiness.* [24] Accuracy, due care, competence –the 'left side' of the pair– are especially saliant with regard to achieving the right understanding. Important here is are the value of chasing insight, truth, and facts, as non-oppressive representations of people and situations: to resist unfair knowledge spheres [44:48, on the value of truth chasing see also 46], even if this would count as 'irresponsible' in the eyes of one's own community[16:56]. Sincerity, intellectual honesty, and trustworthiness –the 'right' side of the pair– especially see to the behavior while doing this, and to communicating the achieved knowledge. Explainers need clear legal instructions in both these dimensions to strive for meaningful information positions for themselves and for their explainees. This means they need to be equipped with the necessary reflective and investigative capabilities with regard to their decision domain, their methods, and with regard to the social-informational and experiential positions of their explainees. They also need to be able to interact with explainees in ways that respect explainees' needs. Last, but certainly not least, they need to rebel when they are not able and/or not facilitated to perform these duties. Put differently, if they don't or can't know, they already have reasons to refuse to do their job.

2.4 Epistemic care, modeled to explanation

The model that incorporates epistemic justice insights frames the re-idealized obligations as care duties. Care duties are obligations to meet in the pursuit of a particular goal or the execution of particular task, but that does not necessarily make them an 'effort only' achievement. E.g., if the point is to keep someone safe, the fact that they weren't safe can be taken to prove that the duty wasn't met, regardless of how the duty bearer interprets safe, and regardless of what the circumstances allowed to do (although these may inform culpability.) Care duties are also instated when a result can be hard to establish, in which case evidence can be required to demonstrate how a goal was aimed for. The model's care duties include both these characteristics and all duties come with a putative protest function for when a duty cannot be met. The model also serves person with influence on upstream dimensions of

decision-making by showing what, in the end, will need to be justified and therefore justify-able.[48] The domain and technology agnostic character of the model makes it applicable to explanation paradigms across decision domains.

The research that underlies this paper applied the model to analyze the legal explanation paradigms of health and administrative law in The Netherlands. [27 ch. 4 and 5] In both domains, no conflict was found with general and domain-specific high level explanation principles. Indeed, some surprising levels of alignment were identified. Sometimes the care duties complemented the principles, or would encourage to interpret them differently. In both cases it became clear however that law(makers) sometimes 'abused' the open ended-ness of high-level norms as well as their hiatuses by setting restrictive rules. Administrative law rules for example exclude crucial preparatory phases of decision making from explanation regulation, which nowadays includes things like profiling 'signals.' In Health Law, the ethical duty to verify patients' understanding was not codified in the recent legal overhaul that meant to align the law with 'shared decision making' practice – a major, missed opportunity to "check the checker".[40]

The model has four elements. Each element describes a dimension, or phase, of explanation. The order is intuitive rather than putative, and mainly helps to separate and therewith explore the different aspects and activities. In reality, the phases will overlap. Some details in the descriptions may be more applicable than others, depending on the decision context.

2.4.1 Element one: investigation of explainer authority

Duty: This element obliges explainers to investigate their own social-epistemic positions. Is their understanding of the decision aims, process, and methods of sufficient quality to support their authority and trustworthiness? Whoever is accountable for the resulting decisions is obliged to ensure that explainers are epistemically equipped for their task. If an explainer can't meet any part of the duty, they need to rebel.

Aim: The point of the element is to avoid to become an instrument of unjust knowledge and decision practice, and to be able to explain any 'avoidance strategies' to explainees. The extent of such explanations is best determined in a decision domain's context, informed by injustices in it, and also needs to be informed by the next element. In any case, the point is to communicate *how*, and not just *that* explainers are trustworthy. Rebelling is necessary when explainers for example lack access to justificatory sources or aren't afforded the time, means, or authority to investigate.

2.4.2 Element two: engaging with the social-epistemic positions of explainees

Duty: Explainers are obliged to investigate the social-epistemic positions of explainees. Can explainees be expected to responsibly provide (or have provided) the necessary input, understand the output, and generally partake in the decision process responsibly i.e. to their benefit? When this is not the case or it can't be properly investigated, explainers need to rebel.

Aim: Like element one, the aim is to 'prepare the table' for the negotiation of the how's and why's of decision outcomes; to ensure that explainees will have an explanation experience that does them justice. For this, explainers should become knowledgeable about their explainees' situatedness in relation to the larger system that the decision under scrutiny is inevitably part of. Explainees may have more insight into this than explainers have, but also less

knowledge of what happens on levels that are not accessible to them. **Note:** In seeking input from explainees from 'affected communities' (as is widely promoted in ADM times), explainees should be facilitated (financially, socially and practically). It may be hard to find connection with subjects who are not yet aware of the impact a (type) of system may have on them as well as with groups that face barriers for getting involved. The need to rebel exists when explainers feel their explainees are in no position to participate in the decisional process to their benefit.

2.4.3 Element three, practicing interactional justice

Duty: Explainers are obliged to recognize explainees as knowers and rights-holders. Information should be provided proportionate to explainees' pre-investigated (element two) and incidental needs. Explainees' knowledge and understanding of the decision process needs to be discussed with them with the aim of promoting responsible (dis)trust, sustained by accessible evidence from outside of the authoritative setting. (The point of) subjects' explanation rights need to be argued, and other rights (like contestation) explained. Any social overpowering needs to be avoided by e.g. allowing explainees to bring allies or make recordings.

Aim: The duties of this element describe behaviors that need to be given an explicit place in the testimonial process. If any description goes beyond what a process is seen to need, this should be argued (e.g. in protocol or records). The inclination in public decision making to treat much practiced or voluminous decision processes as simple, self-evident, routine and predictable has led to sub-optimal explanation practices and much ADM-related wrongs. **Note:** there will be a lack of training for these kind of conversations in various contexts. Health care history for example teaches how talking about medical knowledge has been as necessary as it was avoided (saliently, as 'medical technical'). This puts patients at a disadvantage. Translated to the algorithmic context, people will have a hard time expressing their experience, needs, and interests if they don't know how they are (now or typically) algorithmically processed.

2.4.4 Element four, creating records

Duty: Explanation practices need truthful accounts of the exchanges that are prescribed under element three. Such records should provide evidence of how the previously described duties were attended to, or else provide reasons for why this evidence is missing. The records ideally have input from explainees and are in any case shared with them. They are also made available for outside scrutiny in accordance with rules that govern the decision domain and relevant privacy and data protection regulation.

Aim: These record-related duties are meant to produce more comprehensive accounts than the statements of reasons that are typically the outcome of decision processes. This acknowledges how 'explanation', like decision making, is a place of possible oppression. Comprehensive records can sustain progressive development of decision and explanation practices across time and domains.

3 Workshopping fundamental research: a work in progress

3.1 Aims, preparations, and workshop design

3.1.1 Aims and preparations

The workshop series is part of a project titled ‘van zorgen over uitleg naar zorgvuldige uitleg’, which in Dutch expresses the route from explanation concerns to care-ful explanation practices. (In Dutch, the word *zorg* means concern as well as care.) The first workshop series, planned to run in public policy and health care contexts, serve as a testing ground to learn about the modeled duties’ perceived usefulness and gather feedback on what could be improved – both with regard to the model itself as with regard to workshopping it in a practice context. These tests, that are partly done at the time of writing, will inform the second phase in which the model can be particularized for a particular decision process together with practitioners. This in turn will inform the third phase that seeks to trial it in (that) actual practice. This paper reflects on the first phase workshops given in a public policy context.

The main challenge in this first round was to translate the elaborate and theory-heavy model into a practical tool. It needed to be explainable in a relatively short time without belittling the complex transformations it means to achieve. A domain-agnostic, ‘shorthand’ version of the model was prepared. First succinct descriptions of the elements’ duties and aims were horizontally lined up in a spreadsheet. Rows were then fitted underneath and filled with ‘practice links’ that related the elements to the chosen public policy context in different ways. One row added relevant explanation principles and rules of the field, another row referred to recent public or scientific discussions about particular decision practices, yet another added examples of how the model’s duties and/or the existing legal rules or principles weren’t met, citing sources. Some inter/national ADM developments and cases were added. This preparatory work ideally builds on earlier preparations that necessarily proceed entering a context, i.e. gaining a basic critical understanding of the field’s decision making and justification traditions. As a third step, a mapping of adjacent applicable value frameworks such as the ‘proper public conduct’ norms of the National Ombudsman institute was added in additional rows, to show how the model elements resonate with the larger normative sphere the domain exists in. From this quite elaborate spreadsheet, a selection was made for presentation and printed on a handout that participants could use during the workshop (for which a row with guiding questions was added) and take home. The remaining schedule was taken with to the context in case there would be room to discuss more and to help out with examples and references during exercises. This last option turned out to be a major help – nothing speaks like examples.

3.1.2 Selection and roles of participants

In the first public agency, the workshop was offered to a group who in the preceding part of the week had participated in various activities around transparency and explainability of decision making. In the second agency, a recruitment letter was distributed through internal channels of the institution’s information office and algorithm governance department. The letter explained the workshop was open to anyone who contributed in any way to decision making or the governance of decision making for which explanation and justification obligations apply. The letter stated how explanation law is under (legal)development, introduced how the workshop meant to sustain this development, and

that participants' feedback would be used to improve the project's potential to do so. Two groups of participants self-selected for the available dates. In total, three workshops were given over the two agencies. Participants' professional roles included legal professionals; privacy officers; state council (attorney); software engineers, developers, and testers; legal and policy rule experts, automation of legal and policy rules experts; project, process, and quality leads; product owners and data quality professionals.

3.1.3 The workshop design

The workshop comprised of a theoretical part, a series of exercises, and rounds of reflection. The theory was scheduled to take up a maximum of 25 minutes in a total schedule of 3.5 hours. This first part started off with a presentation of the workshop's aims: to help participants gain insight into the roles of various kinds of knowledge(s) in justification processes; and to raise awareness with regard to participants' own influence and responsibilities in what are typically very large networks. After that a brief iteration of fundamental legal explanation aims and duties in their domain was given. An introduction to the model's theoretical background, and an element-by-element run-by of the itself in its pre-prepared version concluded the theoretical part. This order of the presentation was tweaked along the way, as will be explained further on.

The exercise series involved several rounds of dealing with questions. These were designed to foster reflection on the part of participants with regard to their knowledgeability of the decision processes they worked on. Mapping questions and answers should reveal ways in which the information positions of those involved, and the process that determines these positions could be improved. Posters with the model's elements, represented by icons, were pinned to flip-overs. In round one, participants (divided in groups) were asked to run by each element, recover one or two relevant things they knew about the decision process and pin their answers to the right element on the poster. Round two asked them to jot down one or two relevant things they *needed to know* about the process in order to do their work well, again per element. The printed handout with the model, its practice links, and the guiding questions for each element supported the exercises. E.g., for element 1 (position of the decision maker/explainer), round one (knowledge to offer), the guiding question was: "what part of my understanding of the decision process should the decision maker be aware of as well?" and for round two (knowledge needs): "what do I need to understand about the decision process so that I can avoid to contribute to unjust decision making?" A guiding question for round one with regard to the position of the decision subject (element 2) was "what part of my understanding of the decision process are decision subjects ideally aware of as well?" And a guiding question for element 3 (explainer-explainee interaction), round two (knowledge needs), "what do I need to know about the explanation interaction in order to serve that process, e.g. with evidence"? In the third and fourth exercise round, participants were asked to study their wall of sticky notes together and for each 'piece' of knowledge that was either on offer or required, to write down any takers or providers that were known to them. Providers could be people, departments, databases, or anything else. If there were none, this too would be noted.

The third part of the workshop was reserved for reflection, discussion and feedback. After discussing group and individual reflections, participants were asked to write down their key take-aways and fill in an elaborate

feedback form. These materials, the posters with the sticky notes, and observations recorded on paper in the exercise and reflection rounds were studied in between and after the workshops.

These first workshops were planned as an iterative process, where each workshop would be improved based on the experience of the preceding one. Several adjustments were made based on observations (of myself and workshop assistants) and on feedback from participants. The adjustments saw to the order in which the theoretical parts were presented, the examples that were used, and the depth of engagement in specific parts. In the second workshop, more time was spent on the power/knowledge axis: on epistemic in/justice theory and how that mapped onto the public decision-making domain. This landed well and even more easily than legal parts, which was an unexpected finding. This inspired to flip the order of presentation in the third workshop and present the background theory first. The exercises in the practical part remained largely unchanged since these worked quite well – even in cases where they were considered hard to do. The ensuing discussions on their functionality was (mostly) considered to be insightful for both sides.

3.2 Results and discussion

3.2.1 The guiding value of elaborate preparations

These first workshops were set up to test the model's resonance with different public agency workers. The design assumed a mixed audience in terms of precise roles, expertise, and numbers of participants. In light of this and because the point was to gather inspiration for a more bespoke development of the model later on, the workshop exercises did not revolve around one pre-designed, fictional case. Instead, participants were asked to imagine their own typical 'case load' in the workshop exercise rounds. All teams reported that this was hard and that working on a particular case would have made the exercises easier. All other preparations for the workshop turned out to be both useful and necessary to deal with this. The collected practice links, case-based examples, knowledge of fundamental and ongoing legal developments in the field helped to clarify and answer participants' questions. The acquired understanding also provided an important backdrop for observing participants' work and for example allowed to ask clarifying questions from them before wrapping up their materials for later study. When 'testing' the model with less time for preparation, teaming up with experts from the field in designing the exercises and assist during the workshop could probably cover these needs.

3.2.2 Tuning interaction to (sensitive) group needs

The more informal, clarifying 'resonance' conversations worked best in the second and third workshop because of the lower number of participants. These conversations in turn helped to create a comfortable setting in which to work with what for some participants was quite sensitive subject matters. Many public agency workers are aware of the disparate impact of the Childcare benefits scandal and 'algorithmic mishaps' since then; some voiced concerns about the far-right turn of the government in place. The extent to which participants were interacted with during the exercises as a whole also differed. The large number of participants in the first workshop limited the capacity for it, but they also set to their tasks quite independently. Some were colleagues in daily life and all of them

were in an energetic 'workshop mode' as one assistant observed. In workshop two and three, interaction was gradually increased. Not just in terms of clarifying guidance, but also because some participants reported to feel pressure to 'do the right thing' in their jobs, inspired by the workshop. Surfacing which subjects or questions brought on such pressure together with participants helped to deal with this. One interesting discussion dealt with on who to turn to when the 'protest barrier' was clearly met, which was related to the civil servant oath that they had all taken. For these and other reasons (e.g. the combination of disciplines), that group decided to stick together and continue their discussions on the larger subject: "we found each other now, let's make use of this."

This pressure was not reported in the bigger group. Most of these participants had less individual influence on end decisions, which may have been a factor. The larger workshop however produced more productive theoretical discussions in the reflective rounds, e.g. on the relation of law and ethics ,and on what 'legitimacy' means. The protest duty was less mentioned in their discussion. This is perhaps partly to blame on lack of elaboration of this function in the first presentation. It was however marked in fierce pink, also on the handout.

3.2.3 Practical and moral clarifications in a complex legal landscape: positive resonance

Some participants, especially those with less legal training struggled with the complexity of the legal landscape at the start of the presentation (contrarily, one participant was struck by the comparability of their work on legal analysis, especially of policy goals and what the model asked.) This complexity is a very old and persistent problem of the administrative domain, which makes this an expected finding. The model was overall experienced as clarifying and improving the existing explanation paradigm.at the same time. Some specific feedback: the model was described as an overdue "ethical check" on "strictly legal" explanation demands; the workshop helped to realize how guidance for doing justice in decision making is lacking in law and policy; the model encouraged to look beyond rules and regulations to make the "system world" connect with real world experience of decision subjects. Several participants reported to find the 'duties of care' framing clearer, and stronger, than the/a rights-oriented framing. The model's focus on decision makers' obligations *while* keeping the subject in the loop at all times was appreciated. These findings, like those under the next header, should perhaps inspire to emphasize how the model is set up to strengthen legal rules in place (and is not in conflict with them) – i.e., to emphasize how law *is* always already an ethical practice, too, and they can use the model (and that argument) to fend for improvement from within.

3.2.4 Saving law from itself

Several participants whose daily work revolved around transparency/explainability juxtaposed their understanding of "legal transparency" to the "explainability" that they understood the model to promote. They also mapped this difference to the difference between what they called a "legally acceptable" decision, and a "genuinely just" decision. Their perceptions of incommensurability in each of the framings made for productive discussions. Several participants had jotted knowledge about the difference between legality, legitimacy, and justifiability on sticky notes in the 'knowledge needs' exercise round. Some of the group reported how they had come up with the idea of an explanation "disclaimer" in a previous workshop that did feature a fictional case. The disclaimer was meant to function as a legal cover for explainers who wished to give decision subjects a "real" explanation about how a decision in their case was

taken. The disclaimer would say that they could not use the real explanation in any legal claim. The rather surprising (and illegal) solution and the perceptions of problem it assumes to address resonates with the earlier mentioned tendency in Dutch policy circles to blame formal requirements in law and policy for the inability to do individual cases justice, and casting informal procedures with added discretionary powers as the solution to this. In fact, ongoing legal reforms are supposed to put an end to 'informal review' procedures that left decision subjects without *any* explanation of the new decision. Another reason to make clear that the real cure to suboptimal explanation law needs to be *part of* law.

3.2.5 More is less: the value of elaborate explanation (rules)

The model's systematic "unraveling of what explanation is" into the four phases/elements received much positive feedback. The exercises with the elements were appreciated for how they stimulating "combined thinking" about knowledge on offer, knowledge needs, and knowledge 'leads': how to find takers and providers of knowledge, and how to organize for that. These stimulations were described as a useful way to raise awareness about "the processes one is involved in." One interesting remark was how the model was seen to address a lot more than explanation law does, yet the sum total was not more but "more clear". Perhaps the explicit connection between *duties* and *aims* in the model, which many described as of much value, helped produce this outcome. This also inspired to flip the order presentation in the start of the workshop after the first iteration. The more emphasis the goals got, the clearer the duties became, which feeds the argument for more and more explicit codification of high-level guidance, and the codification of principles. The presentation had explicitly mentioned these high-level norms and how knowing about them is legally required in public decision making. Participants were asked who in fact had knowledge about them and nearly all participants had shaken their heads. Yet hardly anyone jotted down 'legal principles' in the sticky-notes rounds for 'knowledge needs'.

3.2.6 Major knowledge gaps: explainer and explainee needs

A major theme in participants' feedback pertain to element 3's emphasis on explanation as/in interaction. Awareness about (the influence and need for mitigation of) power inequalities was raised, and about the inevitably interactive character of explanation – raising questions about "how to organize for this." Reflecting on their 'knowledge on offer', many participants reported how they had gained insight about the *relevance* of their expertise, yet lacked knowledge about the information needs and positions of explainers and explainees to understand to what extent they could benefit from this. Or, they knew this was knowledge that explainers and explainees did not have and would appreciate, but not how to get it to them. This included knowledge about how law translated to code, and work done on explainability. Apparently, the only takers for such knowledge existed on other parts of the decision chain. The fact that many public agencies had started to increase their focus on digital information provision and chatbots, rather than in-person modalities was also called out as problematic. This lack of contact with citizens, lack of knowledge about their needs and how to cater to these, and also a lack of knowledge about the effects of policies on citizens was considered a major problem by many participants. Some wondered whether the model could be used to reorganize "the system" along more direct contact lines, or inspire process reorganizations more generally. The value of

explainee contact was also emphasized by a participant who was more familiar with the complex experiences of citizens in bureaucracy than some of their colleagues. They explained how those confrontations had led to a profoundly deeper understanding of their work. On the feedback form, in the space for the question 'who would you want to do this workshop with' several participants answered "explainers and explainees"

3.2.7 Concluding observations

The results encourage to continue the research process towards an implementable model without much -if any- changes to the model's contents at this point. The lessons that can be drawn are mostly in terms of *how* to work with the model in the more bespoke, pre-designed situations envisioned for phase two. For others who would be interested in doing a workshop with the model as it currently stands, the following main points are useful to take on board. With regard to the epistemic in/justice underpinnings of the model: there were several advantages to letting this take up more space than planned and starting the workshop with it. The theory, that all participants were unfamiliar with, helped to understand the 'situation' of explanation and what is at stake and at play in it. Not just for explainer and explainee, but for anyone working on decision processes. 'Epistemic' relations became visible – including saliently *missing* relations. The deeper understanding of what happens in explanation in turn helped to make more sense of the rules at play in the field, and qualify them. This was also true for participants who were quite familiar with the legal explanation landscape. Put differently, the 'fit' of the theory made it into a useful thinking tool. The same can be said about the model. The way the model unraveled subjects, actions, and actors allowed participants to find their 'location' in decision making, to understand their own needs and how their contributions are or could be of use to others. The space that opened up this way allowed for creative thinking. But that also raises a more complicated point. For the model to be accepted, embedded and make a difference in existing practice environments, it will also need to be accepted as a 'legal tool'. A tool that makes the most *of* law and the principles that (can) guide it. Engaging with participants' understanding and confusions about legality, legitimacy, in/formality and in/justice is therefore encouraged. The protest button is there for when this progressive effort is undermined – not to break all things before this work has even started.

4 Conclusion

This paper called attention to the fundamental aims and values that underly explanation rules in law, namely to ameliorate power and information inequalities between decision makers and their 'subjects' and prevent that the former misuse their powers. By specifying what needs to be justified about decision processes, such rules also prescribe what kind of processes we allow ourselves to have. The paper argued how these rules have historically proven to be a weak force against –especially– institutionalized racism, discrimination and marginalization in decision making; harms that have been boosted by the proliferation of ADM. This is all the more problematic in light of the racist and populist political climate that we are now again witnessing in Europe. This climate puts pressure on the consciences of public servants who wish to avoid complicity; a task that has not become easier in what have become complex socio-legal-technical ADM environments. Not so much because of the 'complexity' of technology but because straightforward guidance can be hard to find in a complex national and supra-national regulatory landscape,

including a multitude of ethical tools and frameworks (that have failed to compensate for laws weaknesses). This paper argued to support their decision making – *at* their jobs, and *about* their jobs – in critical algorithmic fairness research directly. The paper introduces a 'perfectly legal' yet transformational framework of explanation duties that can be used to orient such research output. The framework is neither restricted to any decision domain, nor any kind of decision method. It models insights from epistemic in/justice research onto four dimensions of decision-and-explanation processes. This brings in attention for the power/knowledge relations that existing explanation laws fail to address. The model helps decision makers, explainers and others with any kind of influence on decision making to understand what they need to be able to understand and explain, and which knowledge deficits should inform them to raise alarm or if necessary, break ranks. Several workshops were done to test the 'resonance' model in a public decision-making context. These workshops validated the value and fit of the framework and will inform its further development for more bespoke application. The workshop design is included in the paper and can also be done in research environments, for example to inform interdisciplinary work in the 'fairness fields'. On a more theoretical level, the paper hopefully inspires to engage (further) with epistemic in/justice research in any work that pursues fundamental equity and justice in decision making.

Acknowledgements

Thanks to Shazade Jameson for acting as a seasoned and enthusiastic sparring partner for the workshop design. Thanks Merel Noorman and Gijs van Maanen for comments on the first setup. Thank you, Shazade (1st ws), and Anne Vaendel (all ws's) for invaluable help and observations during the workshops. Thanks to the anonymous reviewers for their highly engaged and valuable feedback.

References

[1] Elizabeth Anderson. 2012. Epistemic Justice as a Virtue of Social Institutions. Social Epistemology 26, 2 (April 2012), 163–173.
[2] Agathe Balayn and Seda Gürses. 2021. Beyond-Debiasing: Regulating AI and its inequalities. EDRi.
[3] Solon Barocas and Andrew D. Selbst. 2016. Big Data's Disparate Impact. California Law Review 104, 3 (2016), 671–732.
[4] J. Beckwith and R. Pierce. 2018. Genes and Human Behavior: Ethical Implications. In Molecular-Genetic and Statistical Techniques for Behavioral and Neural Research, RT Gerlai (ed.). Elsevier Academic Press.
[5] Ruha Benjamin. 2019. Race After Technology. Polity Press.
[6] Lukas van den Berge. 2021. Bestuursrecht na de toeslagenaffaire: hoe nu verder? Over het rechtskarakter van het bestuursrecht. Ars Aequi 2021, (November 2021), 987.
[7] Reuben Binns and Michael Veale. 2021. Is that your final decision? Multi-stage profiling, selective effects, and Article 22 of the GDPR. International Data Privacy Law 11, 4 (2021).
[8] Patrick Birkinshaw. 2011. Transparency as a Human Right. In Transparency: the key to better governance?, Christopher Hood and David Heald (eds.). Oxford University Press, Oxford, UK.
[9] Edwin Black. 2001. IBM and the Holocaust: The Strategic Alliance Between Nazi Germany and America's Most Powerful Corporation. Crown Publishing Group.
[10] Wouter Boonstra. 2025. 'Ambtenaar heeft morele ondergrens nodig.' Binnenlands Bestuur. Retrieved January 14, 2025 from https://www.binnenlandsbestuur.nl/carriere/ambtenaren-moeten-morele-ondergrens-bepalen
[11] Kiel Brennan-Marquez. 2017. Plausible Cause: Explanatory Standards in the Age of Powerful Machines. Vanderbilt LawReview 70, (2017), 1249–1301.

[12] Meredith Broussard. 2023. More Than a Glitch: Confronting Race, Gender, and Ability Bias in Tech. MIT Press, Cambridge, Massachusetts.
[13] Ryan Calo and Danielle K Citron. 2021. The Automated Administrative State: A Crisis of Legitimacy. Emory Law Journal 70, 4 (2021).
[14] Havi Carel and Ian James Kidd. 2014. Epistemic injustice in healthcare: a philosophial analysis. Med Health Care Philos 17, 4 (November 2014), 529–540. https://doi.org/10.1007/s11019-014-9560-2
[15] Havi Carel and Ian James Kidd. 2017. Epistemic Injustice in medicine and health care. In The Routledge Handbook of Epistemic Injustice (paperback edition), Ian James Kidd, José Medina and Gaile Pohlhaus, Jr. (eds.). Routledge.
[16] Lorraine Code. 1987. Epistemic Responsibility. Brown University Press.
[17] Felix Cohen. 1931. The Ethical Basis of Legal Criticism. The Yale Law Journal 41, 2 (1931), 201–220. https://doi.org/10.2307/791552
[18] Council of Europe. 2018. Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data (Convention 108+).
[19] K J De Graaf and A T Marseille. Exit willekeurstoets. Bestuursrechterlijke toetsing aan het evenredigheidsbeginsel na 2-2-'22. Ars Aequi 2022, April .
[20] EDPS Ethics Advisory Group. 2018. Towards a digital ethics.
[21] Virginia Eubanks. 2018. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (1st edition ed.). St. Martin's Press, Macmillan Publishers.
[22] European Union Agency for Fundamental Rights (Ed.). 2024. Addressing racism in policing. Publications Office, Luxembourg. https://doi.org/10.2811/971074
[23] Charles Foster. 2011. Human Dignity in Bioethics and Law. Hart Publishing.
[24] Miranda Fricker. 1988. Rational Authority and Social Power: Towards a Truly Social Epistemology. Proceedings of the Aristotelian Society, New Series 98, (1988), 159–177.
[25] Miranda Fricker. 2007. Epistemic Injustice: Power and the Ethics of Knowing. Oxford University Press.
[26] Timnit Gebru and Émile P. Torres. 2024. The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence. First Monday (April 2024). https://doi.org/10.5210/fm.v29i4.13636
[27] Aviva de Groot. 2023. Care to explain? A critical epistemic in/justice based analysis of legal explanation obligations and ideals for 'AI'-infused times. Tilburg University.
[28] Aviva de Groot and Siddharth Peter De Souza. 2025. Law-Making, Knowledge-Making, World-Making: Reading the EU AI Act Through an Epistemic In/Justice Lens. In Regulating human behaviour, regulating data, Bart van der Sloot, Georgio Monti and Friso Bostoen (eds.).
[29] David Heald. 2011. Variaties of Transparency. In Transparency: the key to better governance?, Christopher Hood and David Heald (eds.). Oxford University Press, Oxford, UK.
[30] Independent International Commission of Inquiry on the Occupied and Palestinian Territory, including East Jerusalem, and Israel. "More than a human can bear": Israel's systematic use of sexual, reproductive and other forms of gender-based violence since October 2023. United Nations Human Rights Council. Retrieved March 17, 2025 from https://www.ohchr.org/en/press-releases/2025/03/more-human-can-bear-israels-systematic-use-sexual-reproductive-and-other
[31] Amjad Iraqi. 2024. 'Lavender': The AI machine directing Israel's bombing spree in Gaza. +972 Magazine. Retrieved March 17, 2025 from https://www.972mag.com/lavender-ai-israeli-army-gaza/
[32] Nani Jansen Reventlow. Radicale rechtvaardigheid – De wereld die we nodig hebben. Retrieved November 7, 2024 from https://radicalerechtvaardigheid.nl/
[33] Meg Leta Jones. 2017. The right to a human in the loop: Political constructions of computer automation and personhood. Soc Stud Sci 47, 2 (April 2017), 216–239.
[34] Margot E. Kaminski. 2018. The Right to Explanation, Explained. (June 2018).
[35] Jay Katz. 2002. The Silent World of Doctor and Patient.1984 (Johns Hopkins edition ed.). Johns Hopkins University Press.
[36] Yarden Katz. 2020. Artificial Whiteness: Politics and Ideology in Artificial Intelligence. Columbia University Press.

[37] Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr. 2017. Introduction to The Routledge Handbook on Epistemic Injustice. In The Routledge Handbook of Epistemic Injustice (paperback edition), Ian James Kidd, José Medina and Gaile Pohlhaus, Jr. (eds.). Routledge.
[38] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. 2025. Kamerbrief met kabinetsreactie op rapport "Etnisch profileren is overheidsbreed probleem" van Amnesty International - Kamerstuk - Rijksoverheid.nl. Retrieved March 17, 2025 from https://www.rijksoverheid.nl/documenten/kamerstukken/2025/02/03/kamerbrief-met-kabinetsreactie-op-rapport-etnisch-profileren-is-overheidsbreed-probleem-van-amnesty-international
[39] B.J. Koops. 2013. On decision transparency, or how to enhance data protection after the computational turn. In Privacy, due process and the computational turn, M Hildebrandt and K de Vries (eds.). Routledge, Abindon, 169–220.
[40] J. Legemaate. 2018. Aanpassingen van de WGBO. Tijdschrift voor Gezondheidsrecht 42, 6 (December 2018), 556–564.
[41] Jill Lepore. 2020. If Then: How the Simulmatics Corporation Invented the Future (Illustrated edition ed.). Liveright, New York.
[42] C. McCrudden. 2008. Human Dignity and Judicial Interpretation of Human Rights. European Journal of International Law 19, 4 (September 2008), 655–724.
[43] Amade M'charek. 2022. Race and sameness: on the limits of beyond race and the art of staying with the trouble. Comparative Migration Studies 10, 1 (April 2022), 13. https://doi.org/10.1186/s40878-022-00287-z
[44] José Medina. 2017. Varieties of hermeneutical injustice. In The Routledge Handbook of Epistemic Injustice (paperback edition), Ian James Kidd, José Medina and Gaile Pohlhaus, Jr. (eds.). Routledge.
[45] Charles W. Mills. 2005. "Ideal Theory" as Ideology. Hypatia 20, 3 (2005), 165–184.
[46] Veli Mitova. 2021. A New Argument for the Non-Instrumental Value of Truth. Erkenntnis 2021, (June 2021), 1–23.
[47] Samuel Moyn. 2019. Not Enough: Human Rights in an Unequal World. Harvard University Press, Cambridge, Massachusetts.
[48] Daniel Neyland. 2016. Bearing Account-able Witness to the Ethical Algorithmic System. Science, Technology, & Human Values 41, 1 (January 2016), 50–76.
[49] Jessica Nordell. 2021. The bias that blinds: why some people get dangerously different medical care. The Guardian. Retrieved October 15, 2021 from https://www.theguardian.com/science/2021/sep/21/bias-that-blinds-medical-research-treatment-race-gender-dangerous-disparity
[50] Cathy O'Neill. 2016. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown, New York.
[51] Marion Oswald. 2018. Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 376, 2128 (September 2018), 20170359.
[52] Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights in the case of NJCM /De Staat der Nederlanden (C/09/550982/HA ZA 18/388). 2019. Brief by the United Nations Special Rapporteur on extreme poverty and human rights as Amicus Curiae in the case of NJCM C.s./De Staat der Nederlanden (SyRI) before the District Court of The Hague (case number: C/09/550982/HA ZA 18/388). Retrieved October 22, 2019 from https://www.ohchr.org/Documents/Issues/Poverty/Amicusfinalversionsigned.pdf
[53] Devansh Saxena, Charles Repaci, Melanie D Sage, and Shion Guha. 2022. How to Train a (Bad) Algorithmic Caseworker: A Quantitative Deconstruction of Risk Assessments in Child Welfare. In CHI Conference on Human Factors in Computing Systems Extended Abstracts, April 27, 2022. ACM, New Orleans LA USA, 1–7.
[54] Frederick Schauer. 1995. Giving Reasons. Stanford Law Review 47, 4 (April 1995), 633–659.
[55] Gerwin van Schie. 2022. The Datafication of Race-Ethnicity. Utrecht University.
[56] Andrew D. Selbst and Solon Barocas. 2018. The Intuitive Appeal of Explainable Machines. 87 Fordham Law Review 1085 (2018).
[57] Andrew D. Selbst and Julia Powles. 2017. Meaningful information and the right to explanation. International Data Privacy Law 7, 4 (November 2017), 233–242.

[58] Dr Thom Snaphaan and Dr Julien van Ostaaijen. Ruimte voor tegenspraak: Hoe meer ruimte te creëren voor het tegengeluid van rijksambtenaren.
[59] Marcel Ten Hooven. 21 January. De verzorgingsstaat is grimmig geworden. De Groene Amsterdammer. Retrieved June 4, 2021 from https://www.groene.nl/artikel/de-verzorgingsstaat-is-grimmig-geworden
[60] Titus Verster. 2024. Ambtenaren roepen voor 42e keer op tot staakt-het-vuren in Gaza. The Rights Forum. Retrieved March 17, 2025 from https://rightsforum.org/ambtenaren-roepen-voor-42e-keer-op-tot-staakt-het-vuren-in-gaza/
[61] Kate Vredenburgh. 2022. Freedom at Work: Understanding, Alienation, and the AI-Driven Workplace. Canadian Journal of Philosophy 52, 1 (February 2022).
[62] Jeremy Waldron. 2011. How Law Protects Dignity. Social Science Research Network, Rochester, NY. Retrieved September 5, 2018 from https://papers.ssrn.com/abstract=1973341
[63] Margaret Urban Walker. 2010. Truth Telling as Reparations. Metaphilosophy 41, 4 (2010), 525–545.
[64] Bernard Williams. 2002. Truth and Truthfulness. Princeton University Press.
[65] Patrick Williams, Adi Kuntsman, Emeka Nwankwo, Danella Campbell, and Leah Cowan. Surfacing Systemic (In)justices: A Community View. Systemic Justice.
[66] Berber van der Woude. 2024. Ambtenaren, kom in verzet! Jacobin. Retrieved January 13, 2025 from https://jacobin.nl/ambtenaren-kom-in-verzet/
[67] Shelley Wright. 2001. International Human Rights, Decolonisation and Globalisation: Becoming Human. Routledge, London.
[68] John Zerilli, Alistair Knott, James Maclaurin, and Colin Gavaghan. 2019. Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard? Philos. Technol. 32, (2019), 661–683.
[69] 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.
[70] 2020. Ongekend Onrecht: Verslag van de Parlementaire ondervragingscommissie Kinderopvangtoeslag. Tweede Kamer der Staten-Generaal, Den Haag.
[71] 2021. Recht vinden bij de rechtbank: lessen uit kinderopvangtoeslagzaken. Werkgroep reflectie toeslagenaffaire rechtbanken.
[72] 2024. Etnisch profileren is overheidsbreed probleem. Nederlandse overheid moet burgers beschermen tegen discriminerende controles. Amnesty International. Retrieved April 2, 2024 from https://www.amnesty.nl/content/uploads/2024/03/Amnesty-2024-Rapport-Etnisch-profileren-is-overheidsbreed-probleem-2.pdf?x90465
[73] 2024. Legacies of Eugenics: An Introduction. Los Angeles Review of Books. Retrieved March 8, 2025 from https://lareviewofbooks.org/article/legacies-of-eugenics-an-introduction
[74] 2025. Maximizing Fundamental Rights Protections in the AI Act: Liberties' Opinion on Definition and Prohibitions. Retrieved March 17, 2025 from https://www.liberties.eu/en/stories/ai-paper-consultation/45277
[75] 2025. Voortgangsrapportage - Discriminatie in dienstverlening. Ervaringen en lessen om discriminatie in publieke dienstverlening te voorkomen en te bestrijden. Staatscommissie tegen Discriminatie en Racisme.
[76] 2025. Liberties Rule of Law Report 2025. Retrieved March 17, 2025 from https://www.liberties.eu/en/stories/rolreport2025-main/45330
[77] Joint statement - A dangerous precedent: how the EU AI Act fails migrants and people on the move. Access Now. Retrieved March 17, 2025 from https://www.accessnow.org/press-release/joint-statement-ai-act-fails-migrants-and-people-on-the-move/
[78] Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal. Amnesty International.
[79] Iedere Nederlander gelijkwaardig: maak de motie-Becker ongedaan. Retrieved March 17, 2025 from https://petities.nl/petitions/iedere-nederlander-gelijkwaardig-maak-de-motie-becker-ongedaan?locale=nl