MIS Quarterly

SPECIAL ISSUE

# WHEN JUSTICE IS BLIND TO ALGORITHMS: MULTILAYERED BLACKBOXING OF ALGORITHMIC DECISION-MAKING IN THE PUBLIC SECTOR[1]

**Charlotta Kronblad**
Department of Applied Information Technology, Gothenburg University,
Gothenburg, SWEDEN {charlotta.kronblad@ait.gu.se}

**Anna Essén and Magnus Mähring**
Department of Entrepreneurship, Innovation, and Technology, House of Innovation,
Stockholm School of Economics, Stockholm, SWEDEN {anna.essen@hhs.se} {magnus.mahring@hhs.se}

*Both research and public discourse have recently drawn attention to the downsides of algorithmic decision-making (ADM), highlighting how it can produce biased and discriminatory outcomes and also pose threats to social justice. We address such threats that emanate from but also go beyond algorithms per se, extending to how public agencies and legal institutions respond or fail to respond to the consequences of ADM. Drawing on a case study of the use of an ADM system in public school administration, we explore the practices through which public institutions avoided engagement with the detrimental consequences of ADM, leading to injustice. We provide a conceptual model outlining how organizational ignoring practices can lead to social and institutional blackboxing of an ADM system, engendering both social and legal injustice. Our work paves the way for interdisciplinary research on the multilayered blackboxing of ADM. We also extend algorithmic injustice research to include a legal dimension and provide practical implications in the form of a legal framework for ADM in the public sector.*

**Keywords:** Algorithmic decision-making, social justice, organizational ignoring, institutional blindness, multilayered blackboxing, legal justice, public sector

## Introduction

We live at a time when welfare states are becoming increasingly digitalized (United Nations General Assembly, 2019) and algorithms are exerting more and more influence over our lives (Benjamin, 2019; Eubanks, 2018; Zuboff, 2019). In the public domain, digital technologies are being implemented to promote welfare by increasing efficiency, quality, and equal treatment (United Nations General Assembly, 2019). However, ethics and accountability have become growing problems, as algorithms are sometimes unfair and inscrutable, causing harm and diminishing citizens' rights (Martin, 2019; O'Neil, 2016). It is critical that we become aware of these increased algorithmic risks, as we otherwise risk "stumbling, zombie-like, into a digital welfare dystopia" (United Nations General Assembly, 2019, p. 2). Indeed, a growing body of research within information systems (IS) addresses how digital technologies can threaten social justice, showing, for instance, how biases in algorithmic decision-making (ADM) can have discriminatory outcomes (Athey, 2017; Gianfrancesco et al., 2018; Giermindl et al., 2022).

[1] Min-Seok Pang, Atreyi Kankanhalli, Margunn Aanestad, Sudha Ram, and Likoebe M. Maruping were the accepting senior editors for this paper. Dubravka Cecez-Kecmanovic served as the associate editor. Online transparency materials can be found at https://osf.io/fgqbk. They consist of: (1) examples of coding and empirical material (nine tables), (2) selected media headline references and reports in study (two tables), (3) summary of legal court case arguments (one table).

While this literature draws attention to the social injustices resulting from the often opaque decision-making of algorithms, we identify the need to address the potential threats to social justice that emanate from but also go beyond algorithms per se and explore how public institutions (public agencies and the legal system) address and cope with algorithmic agency (Rinta-Kahila et al., 2022) and its negative consequences. Put differently: When algorithms do wrong, will our public institutions put things right?

Clearly, recent rapid advances in the use of algorithms for decision-making have partly unforeseen and partly negative consequences for consumers (Khandani et al., 2010), patients (Faraj et al., 2018), citizens (Cecez-Kecmanovic, 2022), criminal suspects (Sheehey, 2018), and job applicants (Giermindl et al., 2022). These developments also place new demands on the legal system to address the unjust outcomes of algorithmic agency. In a public administration setting, this connects social justice (Aanestad et al., 2021; Sandel, 2020; Tyler, 2012) tightly to legal justice (Goldman & Cropanzano, 2015; Tapp & Kohlberg, 1971) and the legal system's avenues for recourse and restitution in relation to errors made by ADM systems. One recent example concerns the Dutch tax authority's use of a self-learning algorithm to detect suspected social welfare fraud, erroneously demanding repayments that pushed thousands of innocent families into poverty, leading to these families losing their homes and to more than a thousand children being placed in foster care (Amnesty International, 2021; Heikillä, 2022). After more than six years, and in the face of mounting media attention, the Dutch court of The Hague finally determined that the governmental use of the algorithm violated the right to private and family life, issuing a court order suspending its use (Vervloesem, 2020). While the court finally put an end to this unjust practice, this example shows that wrongs perpetrated by algorithms can go uncorrected for excruciatingly long periods of time, with public institutions effectively turning a blind eye or engaging in practices of ignoring (Essén et al., 2022; McGoey, 2019; cf. Rinta-Kahila et al., 2022).

Furthermore, it has been noted that courts confronted with injustices related to algorithmic agency have avoided taking a stand on the ADM-related issue (for instance, whether software code should be considered public record), instead determining the outcome of individual cases based on other, case-specific factors (Kaun, 2021). Together, these observations suggest that public institutions might be ill-equipped to restore social justice when algorithms do wrong, leading us to ask the following research question:

*How do public institutions avoid addressing and correcting unintended and detrimental consequences of algorithmic decision-making and what implications does this have for social and legal justice?*

We pursue this research question through a grounded analysis (Charmaz, 2011; Urquhart, 2022) of practices and outcomes of ADM in the context of public school placements in Gothenburg, Sweden, where the use of ADM resulted in widespread breaches of applicable regulations and legislation, assigning thousands of children to schools in violation of relevant rules. Despite massive protests, most violations remained unaddressed, first by the Public School Administration (PSA) and subsequently by the court system. Our study combines firsthand observations and publicly available data with a deliberate field intervention (Zuiderent-Jerak, 2015) in the form of a lawsuit against the City of Gothenburg by the first author, intended to explore the legal system's ability to deal with ADM. As our analysis developed and our engagement with existing theory evolved, we drew on the literature on organizational ignoring (Gross, 2007; Knudsen, 2011; McGoey, 2019) and institutional blindness (Brussig & Knuth, 2013; Krumer-Nevo et al., 2016; McGoey, 2020), in combination with recent theorizing on blackboxing (Anthony, 2021; Lebovitz et al., 2022; Rai, 2020) and algorithmic injustice (Marjanovic et al., 2022).

Based on our analysis, we provide a theoretical explanation of how social injustice implications of ADM can remain unaddressed (and uncorrected) by public institutions. Our conceptualization contributes to the literature on artificial intelligence (AI) and social injustice by showing how algorithmic injustice (Marjanovic et al., 2022) can be rooted in organizational ignoring practices (Gross, 2007; McGoey, 2019, 2020) that contribute to making actors blind to the ADM and its consequences. We also build on recent work suggesting that opacity related to AI can be partly traced to practices surrounding technology (Anthony, 2021; Lebovitz et al., 2022) by demonstrating how blackboxing (Christin 2020; Pasquale, 2015; Rai, 2020) can be extended beyond the technology itself in layers of social and institutional blackboxing that prevent actors from seeing and addressing ADM errors. Furthermore, we extend existing theory on algorithms and social (in)justice to also encompass the legal domain by showing that ignoring practices and layered blackboxing can engender legal systems to sustain and reinforce algorithmic injustices instead of correcting them.

## Theoretical Background

Social justice is often broadly defined as fairness, moderation, and equality in the distribution of rights and resources in society (Lettinga & van Troost, 2015; UN, 2006). Scholars have argued, however, that algorithms require us to rethink social justice (Dencik et al., 2018; Gillingham, 2019; Hoffmann, 2019; Keddell, 2019; Završnik, 2020) and that the

current spread of algorithms constitutes a societal transition period (sometimes referred to as "abnormal times," see Fraser, 2008; Marjanovic et al., 2022). Below, we discuss ADM and the nature of algorithms, particularly in relation to social (in)justice, connect social justice to legal justice, and introduce organizational ignoring as a theoretical lens.

## Algorithms and Social Justice

Empirical research addressing the relationship between algorithms and social justice is nascent (Kordzadeh & Ghasemaghei, 2021), but it nevertheless provides important insights into how algorithms can reinforce existing and create new forms of social injustice. For example, studies suggest that algorithms can exacerbate present socioeconomic inequalities (Kellogg et al., 2020; Möhlmann et al., 2021; Wiener et al., 2021; Zheng & Wu, 2022), lead to historical biases becoming solidified (e.g., if trained on poor data) (Rosenblat et al., 2014; cf. Kordzadeh & Ghasemaghei, 2021), and reinforce discrimination (e.g., by profiling individuals based on race, ethnic group, gender, or socioeconomic status) (Boyd & Crawford, 2012; Chouldechova, 2017; Fu et al., 2021; Marabelli, Vaast, & Jingya, 2021).

The use of algorithms can also create new forms of injustice. For instance, algorithms can personalize services yet simultaneously restrict people's options and choices (e.g., by excluding them from offers and information), affecting certain groups in unexpected and negative ways and creating new forms of marginalization and inequality (Constantiou & Kallinikos, 2015; Galliers et al., 2017; Marjanovic & Cecez-Kecmanovic, 2017; Zuboff, 2015, 2019). The novelty or abnormality of the use of algorithms can partly be traced to the specific characteristics of ADM, such as datafication, technological inscribing, and algorithmic inscrutability, which make consequences systematic in nature and change the "what" of social justice, as well as the "who" of the victims of social injustice (Marjanovic et al., 2022). Potential injustices produced by algorithms thus involve not only maldistribution (unjust distribution of resources) but also misrecognition (reinforced inequalities in the status of people) and misrepresentation (unequal access to democratic institutions and means of social redress) (Fraser, 2008; Marjanovic et al., 2022). Not only are groups that have historically been considered vulnerable and marginalized being exposed to these algorithmic risks, but anyone participating in the data economy is subject to them (O'Neil, 2016).

## Algorithmic Opacity and Social Injustice: Technology-Centered and Sociotechnical Perspectives

A key aspect of how algorithms can affect social injustice and whether and how people can preempt or react to algorithmic consequences concerns the ways in which algorithms can be opaque or inscrutable (Faraj et al., 2018; Rai, 2020). A logical starting point for addressing the problem of humans not understanding what algorithms do is an engineering-oriented view (Dolata et al., 2021). This view, also referred to as *trustworthy AI* (e.g., Kaur et al., 2022; Vassilakopoulou et al., 2022) or, more critically, as *solutionism* (Marjanovic et al., 2022), emphasizes technological design as the way to prevent undesirable and unjust consequences of the use of ADM. For instance, algorithms can be better designed to become more bulletproof against biases, and insufficient or partial data can be improved to remove causes of injustice (Barocas & Selbst, 2016; Feuerrigel et al., 2020; Fu et al., 2021; Martin, 2019). In particular, structurally and functionally complex algorithms are identified as leading to "models that are *inherently* uninterpretable to human users" (Rai, 2020, p. 138, emphasis added). This, in turn, points towards technology choices (e.g., static vs. dynamic algorithms; see Marabelli, Newell, & Handunge, 2021) as well as design solutions (e.g., post hoc interpretability techniques that approximate deep-learning models with simpler interpretable models to turn blackboxed models into glassboxed ones—Du et al., 2018; Marabelli et al., 2021; Rai, 2020).

Other studies, employing a more distinct sociotechnical view, emphasize that solutions to algorithmic opacity and associated risks for unjust decisions also include measures such as "humans-in-the-loop" to facilitate human intervention and ensure that the logic of ADM systems is complemented with consideration of its wider consequences (Dolata et al., 2021; Grønsund & Aanestad, 2020; Krakowski et al., 2023; Marjanovic et al., 2021, 2022; Teodorescu et al., 2021). Notably, recent studies have questioned the technology-centric perspective on algorithmic blackboxing, showing, for example, that social practices, such as unengaged users who do not interrogate the algorithm (Anthony, 2021; Lebovitz et al., 2022)[2] or enforced secrecy when code is seen as intellectual property (Burrell, 2016), may contribute significantly to the opacity of ADM systems. So far, however, much of this work has focused on how humans can be "looped into" concrete AI design and usage contexts (Anthony, 2021; Lebovitz et al., 2022; Teodorescu et al., 2021).

[2] The idea of lack of engagement relates to a different conceptualization of blackboxing from actor-network theory as invisibility rather than opacity. Here, blackboxing is seen as a consequence of the success of a technology, with routinization of technology-related practices leading to the technology fading from view and awareness (Latour, 1999; Pinch, 1992; Pinch & Bijker, 1984).

Overall, research on algorithms and social justice has, to date, focused primarily on how to *prevent* errors, with considerably less effort directed towards theorizing how society *reacts* when errors occur. Recent research has called for built-in societal responses to the wider consequences of AI (Dolata et al., 2021; Kordzadeh & Ghasemaghei, 2021) and argued that "new institutions are needed for staging and provisionally resolving … [algorithm-related] disputes democratically" (Marjanovic et al., 2022, p. 6). However, this literature has not investigated how societal structures and capabilities securing social justice beyond an ADM-using organization can be developed, nor does it unpack the dimensions of algorithmic injustice affected by the absence of such capabilities. We argue that both social and legal dimensions of algorithmic injustice are important in this context, as discussed below.

## The Social and Legal Dimensions of Algorithmic Injustice

So far, the role of legal systems in relation to algorithmic (in)justice remains largely unaddressed in the IS literature, and the distinction between social and legal justice is thus relatively unacknowledged. Yet the agreed-upon principles of social justice in any context require a legal system capable of upholding and defending those principles in relation to the new threats posed by ADM (Goldman & Cropanzano, 2015; Stahl, 2008). Social justice is tightly connected to societal norms and value judgments, that is, contextual perceptions of what is fair and what is an equal distribution of resources and rights in any context (Angelle & Torrance, 2019). It contains a moral element and aims toward a fair distribution of material resources in society (Stahl, 2008). In contrast, legal justice makes no material promise but refers only to procedural rights. As such, legal justice provides individuals with access to legal institutions and potential restitution, safeguarding that institutional actors will adhere to the law and be held accountable if they do not (Goldman & Cropanzano, 2015; Lens, 2007). Hence, legal justice is concerned with the ability of an independent court system to assess and uphold justice by following the set procedures of the law in the given jurisdiction (Bingham, 2008; Tapp & Kohlberg, 1971), with legality and accountability serving as guiding principles (Sannerholm, 2022).

While these dimensions of justice have not been theorized in the IS literature, Rinta-Kahila et al. (2022) reported on how governmental actors, after an extended delay, responded to socially unjust effects of ADM system use with a court ruling forcing the decommissioning of an ADM system. That study, however, focused on the actions of a public agency, not on the legal system, and attributed the slow response to the negative consequences of ADM to the public agency's limited vision, confirmation biases, and managerial myopia. Beyond that study, empirically based work on how public (administrative and legal) institutions respond to and are potentially implicated in ADM-related social and legal injustices is scarce. Indeed, despite calls for research on the role of institutions in dealing with the negative consequences of ADM (Dolata et al., 2021; Kordzadeh & Ghasemaghei, 2021) and indications that institutions are slow to respond (Rinta-Kahila et al., 2022; Vervloesem, 2020), theorizing that explains the processes through which institutions can avoid responding is underdeveloped. To this end, we use organizational ignoring (Gross, 2007; McGoey, 2019, 2020) as a theoretical lens.

## Nonresponse to Algorithmic Injustice: An Organizational Ignoring Lens

Research on organizational ignoring challenges the idea that ignorance results from lack of information and knowledge, showing instead that we sometimes actively make efforts to prevent ourselves from knowing (Brunsson, 1998; Gross, 2007; Knudsen, 2011; McGoey, 2019, 2020; Vaughan, 1997), independent of the amount of information or knowledge available (Alvesson & Spicer, 2012). Ignorance can thus be intentional, making it different from unknowability (Bovensiepen & Palkmans, 2020; McGoey, 2020). While the notions of strategic ignorance (Gross & McGoey, 2015; McGoey, 2019, 2020), functional stupidity (Alvesson & Spicer, 2012), and willful ignorance (Schaefer, 2019) suggest self-interest and goal-directed deliberation (related to an individual, a group, or an organization), acts of ignoring can also be reactive, triggered by potentially uncomfortable signals that threaten current practices and norms.[3]

The organizational ignoring literature addresses self-directed acts of ignoring, wherein individuals and organizations make themselves blind to awkward or destructive information (Ashforth & Anand, 2003; Essén et al., 2022; Knudsen, 2011), as well as acts by which organizations strategically prevent others from knowing, thus creating blindness in the environment (McGoey, 2019, 2020; Oreskes & Conway, 2011; Proctor, 2012). Actors have been shown to engage in such practices out of fear of consequences (i.e., if they or others started to know), to avoid liability, and to normalize unexpected observations and maintain institutions (Lok & de Rond, 2013; Oliver, 1991; Proctor, 2012; Rayner, 2012). For

[3] Contrasting these terms, willful blindness in the field of criminal law refers to deliberate denial of knowledge or understanding of specific circumstances that makes someone *legally culpable* (Robbins, 1990; Williams, 1961). We are avoiding this legal term in favor of established concepts from organizational studies (Schaefer, 2019) since our study does not focus on legal culpability.

example, uncomfortable experiences and knowledge can be denied, dismissed, diverted, excluded, or displaced (Klintman, 2019; Knudsen, 2011; Rayner, 2012); consequences (or the lack thereof) can be neglected by decoupling visions and practices (Schaefer, 2019; cf. Brunsson, 1982); and facts can be distributed across systems and units, which in turn prevents actors from connecting the dots (Heimer, 2012). As a result, actors prevent themselves (an act referred to as self-imposed blindness) (Knudsen, 2011) and others from knowing.

Acts of ignoring, and the blindness they produce, are easy to associate with the absence or *lack* of effort. Yet ignoring also involves and requires effort (Knudsen, 2011), and particularly when there are expectations of being knowledgeable, acts of ignoring require creative justification (Essén et al., 2022). Ignoring can also become normalized and unchallengeable (Ashforth & Anand, 2003), with such "macro-ignorance" being a "sedimentation of individual ignorance into rigid ideological positions or policy perspectives that obscure … mistaken assumptions from adherents, leading to new patterns of individual micro-ignorance" (McGoey, 2020, p. 200).

The notion of institutional blindness similarly points to patterned and collective acts of ignoring that have become part of an institutional fabric and/or result from how an institution operates (Brodkin & Marston, 2013; Brussig & Knuth, 2013; Krumer-Nevo et al., 2016). For instance, Krumer-Nevo et al. (2016) accounted for the repeated failure of correctional services to see young people involved in the drug trade, which in turn exacerbates their social exclusion (cf. Brodkin & Marston, 2013). Similarly, Brussig and Knuth (2013) studied a public institution's failure to recognize and respond to the special circumstances of Germany's migrant populations, distinguishing this as institutional blindness. These studies point at structurally determined ignoring and resulting blindness. Such blindness can also be actively reinforced by maintaining (rather than adapting) institutions in the face of changing circumstances. For instance, through buffering (decoupling activities from external contact) and concealing activities (for instance through window dressing), organizations can reduce external evaluation and thereby maintain autonomy and the legitimacy of current arrangements (Oliver, 1991; cf. Meyer & Rowan, 1977). Organizations can also protect current arrangements by acting as if disruptive events or activities had not happened and/or were unimportant (Lok & de Rond, 2013). Thus, intentionality in acts of ignoring can be understood as inhabiting "a spectrum where deliberate unawareness is at one end of the spectrum, and normalized blindness or unconscious disregard on the other" (Bovensiepen & Palkmans, 2020, p. 388).

In summary, the literature on algorithms and their consequences, in combination with the ignoring literature, provides important building blocks for our grounded analysis of how public institutions can avoid addressing and correcting the detrimental consequences of ADM. These literatures enabled us to explore the social processes through which ignoring and nonengagement with an ADM system can persist in our field setting. Next, we outline the methods we used in this pursuit.

## Methods

We conducted a grounded longitudinal case study (Charmaz, 2014; Glaser & Strauss, 1967; Urquhart, 2022) of an ADM implementation and its consequences, combining field notes with a range of publicly available documents and media reports. Leveraging an interdisciplinary author team, including both IS scholars and legal expertise, our method also encompassed a deliberate field intervention (cf. Zuiderent-Jerak, 2015), with one of the authors filing a lawsuit against a public institution for using an ADM system in an unlawful manner. We used the field intervention as an approach to engaged scholarship (van de Ven, 2007), gaining new insights into a complex, under-researched area in which empirical cases are still scarce (Marjanovic et al., 2022) and allowing us to capture unique data.

### Empirical Context

We focused on the implementation of an ADM system for public school placement by the PSA in the Swedish city of Gothenburg. The PSA manages the largest school district in Sweden, with overall responsibility for the public education of about 50,000 children in grades 0-9. Each year, 12,000 decisions on school placements are made. Until 2019, placements were done manually by administrators, using Excel sheets. In 2020, an ADM system was implemented to improve the decision-making process, including making it fairer due to the purported objectivity of the algorithm [City Auditors Report, 2021].[4] According to Swedish legislation, children should be placed at the school of their parents' choice, unless this violates another child's rights to placement at a nearby school [Swedish School Law, 2010:800, ch. 10 §30]. When schools reach capacity, children should be placed at other nearby schools. When considering what schools are nearby, legal praxis stresses that a walking/biking perspective should be applied. In 2020, municipalities were not allowed (according to municipality law) to make automated decisions, but ADM systems could be used to support the decision-making process.

The ADM system was provided with data on children (home

[4] Sources within brackets are part of the empirical material, see Table 1.

address and school preference) and data on schools (address and available slots) and employed a rule-based algorithm to allocate children to available schools. The ADM system was instructed (1) to place the children at the closest school (from a bird's eye perspective) and (2) if that school was full, to place the children at another, nearby school (defined by local regulation as 4 kilometers for the lower grades and 8 kilometers for seventh grade and above). While the intention was to comply with all laws and regulations, in practice, the use of the ADM system violated national law as well as local regulations, producing thousands of erroneous placement decisions, most of which went uncorrected. Our study traced and analyzed how this occurred and how the consequences played out within the ADM-using organization, in the public eye, and through the court system.

## Data Collection and Field Intervention

We drew on multiple data sources collected between 2020 and 2023, as summarized in Table 1, including notes, email documentation, publicly available documents (audits, legal documents, public statements by the city, published reports and media reporting—for details, see online transparency material).[5] The field intervention (conducted by the first author) consisted of initiating and pursuing a legal process, namely filing a lawsuit against the PSA with the purpose of forcing the court system to assess the legality of the ADM system and its use.[6] The administrative legal process in Sweden occurs solely in writing, and all legal correspondence and supporting evidence from both parties was collected. Acting as plaintiff, the first author also captured notes from meetings and emails concerning the case. This involvement in the field was vital for our study, not only in terms of access to data but also for triggering the legal system to act and observing how this process played out. Additional details on data and methods are available in the online transparency material.

The workshop sessions, including legal and public administration officials and legal IT experts, were conducted to corroborate our understanding of the legal system and to capture reflections on our emerging interpretations of the findings. These workshops were distributed in time allowing us both to collect additional data about the legal system and to receive feedback on successive iterations of our analysis.

## Data Analysis

The focus of this study—the ignoring practices of public institutions in relation to ADM and their implications for social and legal justice—emerged following insights gained through our involvement in and analysis of the focal case (Charmaz, 2014; Glaser & Strauss, 1967). In line with recent studies of ADM and social justice (e.g., Rinta-Kahila et al., 2022), we approached the case inductively. In doing so, we were informed by an overarching sociotechnical perspective (Lee, 2004; Sarker et al., 2019), but initially without a distinct theoretical lens, aiming to understand the ADM implementation in its social context. Relatively early, this open-ended exploration was supplemented with an interest in implications for social justice (Kordzadeh & Ghasemaghei, 2021; Marabelli et al., 2021; Marjanovic et al., 2022). While the specific theoretical angle initially remained open, several of the authors had been sensitized to theories (Urquhart, 2022) on organizational decision-making, institutional stability and change, and irrationality in organizations through their earlier work. All three authors were deeply involved in data analysis and theory building to ensure multiple viewpoints on the process and to ensure reflexivity related to the different roles of the authors (Urquhart, 2022).

As we analyzed the data, we discovered that the involved public institutions (the PSA and the court system) seemed to engage in practices that enabled them to refrain from investigating and responding to—and even from seeing—all the systematic errors produced by the ADM system and its use. This triggered our interest in how this inability to see ADM agency and consequences occurred, as well as its (in)justice implications.

Successive iterations of our research question, ultimately focusing on how public institutions avoid addressing and correcting unintended and detrimental consequences of ADM, as well as the social and legal (in)justice implications thereof, guided our analysis of the data. Our coding scheme, developed over several rounds, was inspired by grounded theory method (Glaser & Strauss; 1967; Urquhart, 2022) and used NVivo software for manual coding. Initially, we reviewed multiple sources of data to construct a case narrative. In this process, we arranged empirical observations and other data on a timeline to identify critical events. We identified two partly overlapping phases that differed in terms of which public institution was primarily facing (but refraining from engaging with) the negative consequences of the ADM system. In the first phase, the ADM system was implemented and managed by the PSA; in the second phase, the legal process unfolded within the administrative court system.

[5] The transparency material can be found at https://osf.io/fgqbk.

[6] The first author discovered the case as one of her children was in a class being split up and placed all over the city. While the lawsuit addressed the systematic nature of the ADM system, she separately sought individual redress for her child's placement. The family matter was thus pursued and resolved separately from the lawsuit.

**Table 1. Data Overview**

| Method/source | Details | Issue captured |
|---|---|---|
| Media reports and public documents<br><br>2020-2021 | • 30 news articles from national and local media houses/press (2020-04-27–2020-10-23)<br>• 11 television/radio broadcast news stories (2020-05-07–2020-10-29)<br>• Internal Auditors Report 2020-08-07 (by the school administration) (28 pages)<br>• External Auditors Report by Ernst & Young (2022-12-17, 29 pages)<br>• City Auditors Report 2021-03-16 (42 pages)<br>• Internal evaluation of the school placement process (2021-06-24, 66 pages)<br>• Parent Group Statistical Report 2020-06-07 (statistical analysis of school placements, 52 pages)<br>• Parent Group Report on Consequences 2020-08-14 (analysis of school placement consequences, 24 pages)<br>• Social media posts (tweets) by the City of Gothenburg (2020-06-18) | How the ADM system was implemented and used at the administration and the resulting outcome and consequences (used in the construction of the case narrative |
| Observations and notes<br><br>2020-2021 | • 14 email threads between the plaintiff and other affected parents about understanding the case and the lawsuit<br>• 20 emails from the plaintiff to the school administration and to school politicians<br>• 8 email conversations with media representatives Notes from meetings with the school administration PowerPoint slides constructed to visualize algorithmic enactments, for presentations (2021-10-19, 2022-05-06, 2022-06-17) | Details about the case and background information to the lawsuit, as well as detailed documentation on the lawsuit and the court's rulings |
| Documentation generated in the lawsuit concerning the ADM system<br><br>2020-2021 | • Applicable Swedish legislation: School law (2010:800), Municipality law (2017:725), Administrative process law (1971:291), etc.<br>• Proposal for new legislation for automated algorithmic decision-making by municipalities (SOU 2021:16)<br>• Primary data generated at the field (total of 60 pages) including:<br>o Decision letter (2020-04-30)<br>o Summons application to the Administrative Court of Gothenburg (2020-05-17)<br>o Statement of defense from the PSA (2020-05-26)<br>o Plaintiffs' clarification of statement (2020-06-22)<br>o Defendants' clarification of statement (2020-07-01)<br>o Verdict from the Administrative Court (2020-08-27)<br>o Letter of appeal to the Administrative Court (2020-09-27)<br>o Decision from the Administrative Appellate court not to retry the case (2020-11-16)<br>o Letter of appeal to the Administrative Appellate court (2020-12-08)<br>o Letter to include new vital evidence (2021-03-18)<br>o Decision from the Supreme Administrative Court not to retry the case (2021-06-11) | How the legal system viewed and dealt with errors made by ADM systems and how blindness can be created and upheld by public institutions |
| Expert workshops used for input on interim analysis results<br><br>2021-2023 | • Expert commentaries and feedback collected at six workshops (2021-10-19, 2022-05-06, 2022-06-17, two workshops on 2022-12-13, and 2023-03-23), each 2-3 hours long. About 300 legal and public administration and information technology (IT) experts were consulted, including judges at an administrative appellate court, at a network of chief judges, the IT department of the Swedish National Courts Administration, and senior officials of these public institutions. The PSA case was presented and discussed. Workshop topics included: What constitutes an error in times of algorithmic decision-making, how do courts handle algorithmic issues and algorithms as evidence, how can legal justice be ensured in digital contexts? Feedback and comments were collected through the interactive seminar tool Wooclap, and field notes were taken. | Experts were used as sounding boards to corroborate the soundness and meaningfulness of how case data were interpreted and analyzed. |

**Note**: Dates are in the international date format: YYYY-MM-DD.

Throughout the research process, we corroborated different data sources (data from the legal intervention, firsthand observations, and publicly available documents) iteratively to develop an increasingly thorough understanding and stable narrative of the studied process, its actors, and the ADM system (its use, agency, and consequences). We used the City Auditors' Report (2021) as an authoritative retrospective account of the process. This report, seen as highly credible by a wide range of stakeholders and having a substantial impact on developments in the PSA and the city administration, contained observations from inside the PSA and insights into the ADM system, tests of the system with actual data, and a detailed analysis of its consequences, including an account of the different types of errors that the ADM system caused and how they diverged from legislation and regulations (cf. the Appendix). This informed us of the types of errors that remained unseen and unaddressed. We also used feedback from the workshops for the analysis (van de Ven, 2007). While the collected insights supported the overarching focus of our analysis, they also inspired new questions about how and why ADM errors were ignored, which we used to interrogate the data further.

To understand how the PSA and the court avoided seeing, addressing, and correcting errors, and the injustice implications thereof, we engaged in open coding close to the empirical observations. We iteratively and selectively sorted our open codes to form subcategories and categories concerning the PSA's and the court's (non)acknowledgment of errors produced by the ADM system, as well as the dimensions of injustices that were produced and reproduced as a result. To develop the emerging categories and to relate them to existing theory (Charmaz, 2014; Urquhart, 2022), we consulted several streams of literature, such as technological frames (Orlikowski & Gash, 1994) and institutional maintenance (Dacin et al., 2010; Lawrence & Suddaby, 2006), converging on using organizational ignoring (e.g., Gross & McGoey, 2015; McGoey, 2019, 2020), together with key insights on institutional blindness (Brussig & Knuth, 2013; McGoey, 2020), opacity and blackboxing (Anthony, 2021; Lebovitz et al., 2022; Rai, 2020), and algorithmic injustice (Marjanovic et al., 2022) in our further analysis and theory building. We used these to conceptualize the nonresponse[7] to the ADM issues as the result of effort and to detangle how such efforts involved acts directed internally as well as externally (thereby making oneself and others blind) (Bachrach & Baratz, 1963; Knudsen, 2011).

Iterating between our data, analysis, and extant theory, we ultimately grouped the categories into three interrelated theoretical dimensions (ignoring practices, layered blackboxing, and algorithmic injustice implications). We found that the ignoring practices made the relatively simple algorithm of the ADM system, and information about it, unavailable, initially to the organization implementing and using the ADM, but subsequently also to other actors and to the public. We also developed the insight that the failure of the legal system to scrutinize, or even to enable scrutiny of the ADM, was important for how social and legal injustice occurred and was maintained, thus adding an institutional blackboxing layer, making recourse and restitution for affected public service recipients practically impossible (see Figure 1, and Figure 2 in the Discussion section).

The online transparency material contains our coding and empirical examples linked to the three theoretical dimensions: ignoring practices, layers of blackboxing, and algorithmic justice implications. Based on the categories and theoretical dimensions developed through the open and selective coding (see Figure 1), we further explored the relationships between three theoretical dimensions, leading to a set of theoretical insights and the development of a conceptual model (see the Discussion section) inspired by established practices in grounded theory building within IS (Urquhart, 2022).

## Findings

In this section, we provide a temporally structured account of our case, focusing on the practices that explain how the ADM system could remain unscrutinized by the focal public institutions in our case—the PSA and the administrative court system—despite its legality and implications being repeatedly contested by parents, the media, and external auditors. We structure our narrative into two overlapping phases—tracing, first, the implementation and use of the ADM at the PSA and, second, the legal process. The second phase shows how the court failed both to assess the algorithm and to provide an avenue for legal recourse for affected public service recipients (i.e., the legal system failed to see and address the wide range of errors made by the PSA and the resulting social injustice). Tables with illustrations, empirical examples, codes, and sources are provided in the online transparency material.

[7] We inferred nonresponses from the multiple sources of documentation suggesting that the public organizations were exposed to / had received information (verbally, email, reports) that was potentially relevant, but that they had not engaged with (in-depth scrutinized and responded point by point to the content of) the early indications and warnings about the ADM (Bachrach & Baratz, 1963; Knudsen, 2011). We focused on organizational-level nonresponses in terms of the absence of social behaviors (rather than psychological/cognitive responses) that engaged with the criticisms, e.g., from statements by individuals representing the public organizations in the media and from identified organizational-level nonactions in reports about the case. The continued official support of the ADM served as an additional indication of this nonresponse since this ran against what the warnings suggested.

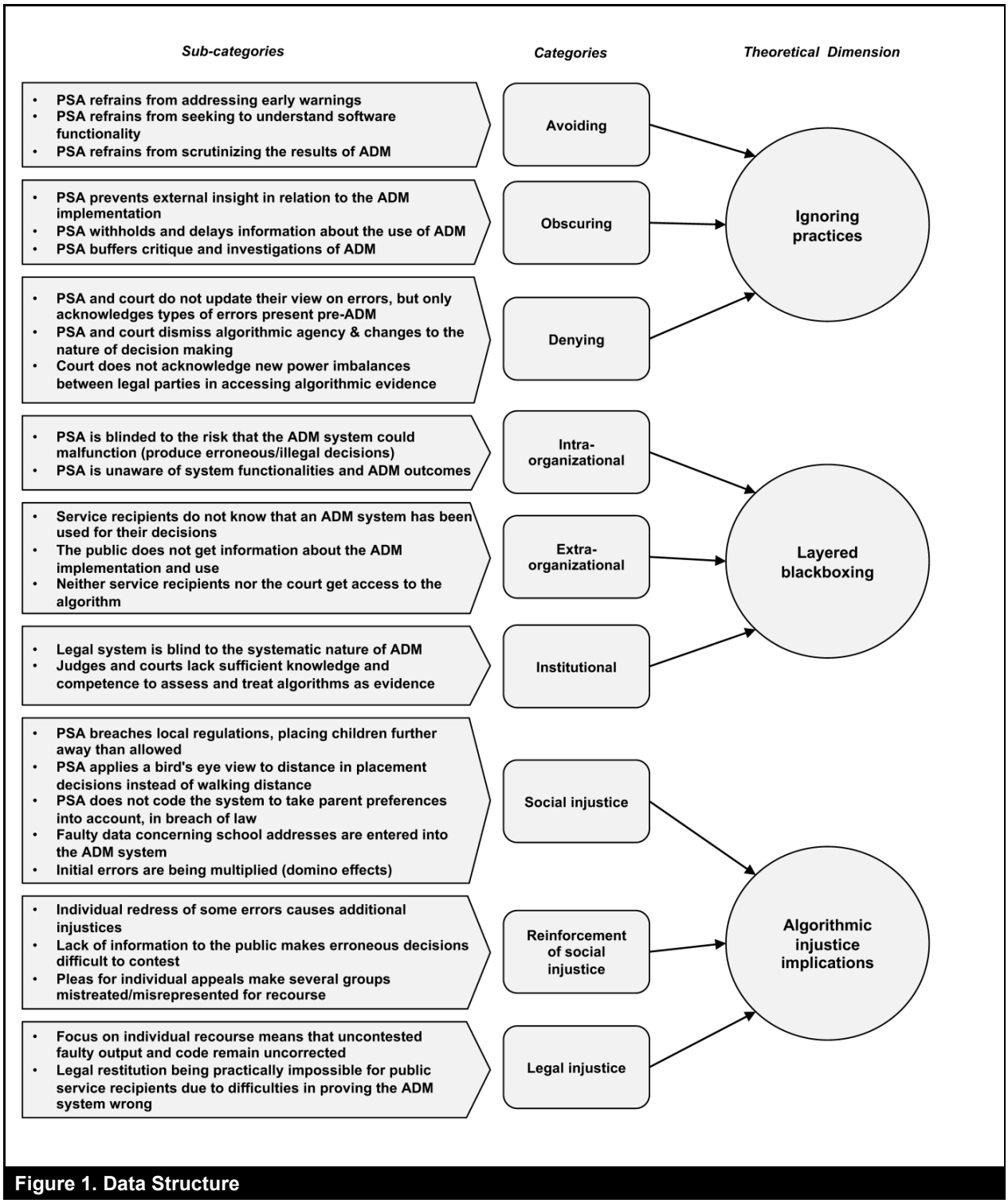



**Figure 1. Data Structure**

### Phase 1: The Implementation and Use of an ADM System in Public Administration

In the fall of 2019, the PSA saw an opportunity to utilize digital technology to improve the school placement process, which was considered outdated. This was internally motivated as a means of ensuring "efficiency" and to obtain "equal treatment for all children in all parts of the city" [City Auditors Report, 2021, p. 14]. Automation was framed as an improvement over an existing process that relied on potentially subjective and cognitively constrained judgments of individual administrators, who could also potentially fall prey to undue influence from parents.

A software firm was chosen to deliver the software program, which was configured according to the instructions of the school administration, allegedly using "the same [decision] rules" [City Auditors Report, 2021, p. 14] as had been applied the year before. Data concerning all children (ages and addresses) were fed into the system, as were school-related data and available slots. Throughout this process, the PSA "did not anticipate that this would result in any fundamentally different decisions than the year before" [City Auditors Report, 2021, p. 20]. The ADM system was configured to place each child at the closest school whenever possible or else place them at another, nearby school (defined as 8 kilometers for children going into seventh grade and above, and 4 kilometers for younger children). Test runs were executed, with adjustments to parameters in between (for instance, regarding the allowed distance for placements, and how many open school slots there were at different schools). Since the PSA experienced a shortage of available school slots, it tweaked the definition of what was nearby by using 8 and 4 kilometers as approximates, rather than absolute limits. After a final test, the aggregated outcome was compared with individual placement preferences from parents, showing an 87% correspondence. The administrators concluded that this was sufficient and "locked" the outcome [City Auditors Report, 2021, p. 19].

Early on, however, the software firm warned that the system was not being configured and deployed in the same way as in other cities [Internal Auditors Report, 2020]. For instance, the software firm raised concerns about the use of a bird's-eye view measure of geographical distance between residences and schools, which they argued could lead to odd or irregular placement outcomes. Similar warnings had come from PSA employees reacting to the rapid implementation of a system they did not understand and from school principals who had received preliminary placement lists, noting that some children being placed at their schools lived very far away. The PSA did not engage with this critique. It did not ask for clarifications, did not investigate further or seek further knowledge about the ADM procedure, nor did it halt and/or modify the software code or the implementation process.

At this point, the placement process was running late. It was already the end of April; placement decisions should have been delivered months ago. The PSA found there was "no possibility [of allowing a human administrator to] control every decision made" [City Auditors Report, 2021, p. 19] and went ahead with officially communicating the decisions produced by the ADM to the parents. This meant that the senior manager who had "formally approved some of the decisions [and who was formally responsible for any divergent decisions] had not even seen them" [City Auditors Report, 2021, p. 19]. The PSA later explained this practice, i.e., of communicating decisions that no administrator had reviewed, as: "There was a risk tied to a better insight into the process, as this could lead to administrators and parents becoming inclined to interfere and affect the outcome" [p. 20]. Thus, the 12,000 placement decisions produced by the ADM were effectuated without modifications. Each decision was formally attributed to the same human PSA administrator.

The placement decisions were communicated on April 30, 2020. Of the 12,000 children, 1,400 (13%) had failed to get any of their preferred choices, compared to only 4% the year before, when the manual process was used [Internal Auditors Report, 2020]. Soon, thousands of parents were reaching out to the PSA, and the media started to report on children being placed in schools far away from their homes, contrary to their expressed preferences and in violation of national and local regulations (illustrative media headlines are provided in the online transparency material). For instance, several hundred children were placed in schools where natural boundaries (rivers, fjords, major highways) restricted routes and extended commuting. The media focused on the most severe, single cases of erroneous placements, and this focus was maintained by the PSA, as it did not know of or acknowledge the severity and scale of the problem. In TV interviews, the head of the PSA blamed existing regulations and technology in general, referring only to the individual cases of erroneous placements and remaining silent about any details of the ADM system. The head of the PSA further responded to critical journalists by referring to undesirable placements as appropriate to address on a case-by-case basis, thus directing attention away from the systematic nature of the ADM system's outputs:

> ***Journalist:*** *This example with one child who would travel 45 minutes, 11 km one way to school, is that ok?*
>
> ***Head of PSA:*** *No, it is not, and that was never our intention. And I strongly advise the [child's parents] to apply for a school change and to appeal the decision. And then we will do everything we can to fix this [individual decision].* (Swedish Television, SVT News, May 7, 2020)

The massive news coverage triggered school politicians from all parties to demand that the PSA investigate what had gone wrong [*Dagens Nyheter*, May 8, 2020]. In response, the PSA instructed discontented parents to submit individual appeals and to apply for individual relocation of their children to desired schools. By the end of May, over 1,000 placement decisions had been individually appealed, and 2,500 individual requests for relocations had been submitted [Internal Auditors Report, 2020]. Still, the PSA offered no explanations or apologies. However, the PSA initiated an "internal audit to find out what had gone wrong and what would be important to consider for future placement processes" [p. 5] and, on June 18, the city tweeted, "several errors have been conducted in this year's school election and placement process #thecityofGothenburg." The PSA now admitted that 450 placements were formally wrong because of noncompliance with the city's own regulations, based on distance (children being placed farther away than 8 and 4 kilometers, respectively). The PSA regarded these decisions as formal errors and appropriate for individual redress (i.e., appropriate to solve through the relocation of individual children, without any acknowledgment of systematic errors). Simultaneously, the PSA became aware that it had registered incorrect addresses for certain schools. This was also acknowledged as a formal error, and the PSA attempted to correct decisions based on this error by manually relocating affected children. However, since schools were already full, only a few of these children were successfully relocated.

In parallel, the PSA insisted that it was not relevant to review the algorithm itself. On several occasions during the spring, individuals and the media requested—orally and in writing—that the algorithm be made accessible to those requesting it (claiming that the principle of public record applied to the code). The PSA, however, never presented the algorithm or gave any information concerning its implementation, nor did it present any explanation of why the code was not provided.

Reacting to the slow and partial response from the PSA, parents in areas that were particularly affected by faulty decisions came together to understand what had happened. Several groups of parents went beyond trying to ensure corrective relocation of their own children to attempt to understand what had caused the situation. Requests to multiple PSA officers to share detailed information about the algorithm went unheeded. However, in response to (the equivalent of) a Freedom of Information Act request, the PSA provided anonymized lists of its placement decisions, with school locations and children's home addresses. Using the data, one parent group analyzed the decisions, finding that systematic domino effects had multiplied scarcity of school slots six times and thereby affected school placements all over the city [Parent Group Statistical Report, 2020]. During this time, the software firm also divulged to these parents that only geographical considerations had been coded into the system and that no regard had been given to parents' school preferences. These findings were shared on a dedicated Facebook page, at organized public demonstrations, and in several reports [Parent Group Statistical Report, 2020; Parent Group Report on Consequences, 2020]. The reports were subsequently submitted as evidence in the court case (see section on Phase 2, below).

The growing public criticism of how the PSA handled the situation led the PSA to initiate an external audit, conducted by Ernst & Young in the fall of 2020. This audit was devoted to how the crisis had been handled but not its causes. The PSA also launched a citizens' dialogue (focused interviews with parents and a series of workshops with different parties) to restore trust in the PSA and to improve its communication around the placement process for the coming years. This initiative did not allow for in-depth discussions on the ADM system but rather focused on how to mitigate the negative emotions triggered by its implementation.

### Analytical Summary of Phase 1

In this phase, the PSA *refrained from addressing* early warnings about ADM risks and consequences raised by the software firm and school principals. Since the PSA had also *refrained from seeking to understand the software functionality* in sufficient detail and *refrained from scrutinizing the results of the ADM* before communicating the decisions to parents, both the scale and the scope of errors *came as a surprise*. In other words, the PSA chose to avoid confronting critical information about the ADM and its consequences. Thus, it *blinded* itself (*intraorganizational blackboxing*) to the ADM system. As a result, the PSA *enabled social injustice* since societal resources (school slots) were not distributed fairly.

The PSA also acted to *prevent insight into the ADM implementation*, not making a formal (public) decision to procure the ADM system. Despite requests from external actors (the media, parents, and the general public), the PSA *withheld and delayed information on how the ADM had been used*. These practices initially shielded the algorithm from scrutiny and the organization from criticism. To this end, the PSA also attempted to maintain control of the situation by *buffering critique and investigations*: by issuing several audits, as well as a citizens' dialogue, addressing the placement process while deliberately restricting the scope of these audits and discussions, thus negating the need to find the root causes of the problems. Together these practices blackboxed the ADM in relation to extraorganizational stakeholders. The PSA effectively blinded others from seeing and contesting the ADM system, reinforcing social injustice.

The PSA justified its nonresponse to the calls for increased information about the ADM with the argument that the ADM had *changed nothing in regard to algorithmic agency and the nature of decision-making*—since a human was still behind each decision. Calls for increased insight into the ADM thereby became irrelevant. While the PSA admitted that mistakes had been made, it only acknowledged such formal errors that were compatible with the pre-ADM situation, *arguing that existing categories of errors and ways to handle them were sufficient*. Thus, the PSA ignored the systematic consequences stemming from errors inscribed into the code and emphasized individual solutions: in effect *supporting existing legal procedures as appropriate for correcting algorithmic decisions*. Since the PSA denied any need to adapt its view on errors and ways to handle them, the new categories remained unseen and uncorrected (see the Appendix).

### *Phase 2: How the Court Avoided Assessing the Legality of the ADM System*

In response to the PSA's lack of engagement with the agency and the systematic nature of the ADM system (and driven by the assumption that individual appeals would never solve a systematic error), the first author of this paper decided to take the algorithm to court. The lawsuit was based on an early suspicion that something had gone wrong in the automation of school placements, that the PSA had breached the law in the process of making automated decisions, and that breaches of law were encoded into the ADM system. The lawsuit was filed on May 18, 2020, asking the Administrative Court[8] of Gothenburg to assess the legality of the ADM system and the decisions it had produced.

The main argument in the summons application was that the school placement decisions had been automated and that the placements had only been based on geographical parameters of school locations, where children who applied (as well as those who had not submitted any preferences) had been ranked on their absolute closeness to the school, with slots being allocated without regard to expressed parental preferences. It stressed that allowing only for individual, rather than system-wide, corrections also constituted an injustice, as this primed children with particularly resourceful parents to get redress, arguing that "you cannot solve a systematic problem with individual corrections; it has to be systematically addressed" [Summons Application, 2020]. The summons application also stated that access to the algorithm had been requested on several occasions but had not been provided by the PSA. As evidence, the plaintiff instead submitted the Parent Group's Statistical Report and their Report on Consequences (including analysis of actual placements and reverse-engineering of the algorithm) as well as the Internal Auditors Report, in which the PSA admitted that several errors had been made in the decision-making process.

The main counterargument forwarded by the PSA was that the ADM system had not been used to automate decisions but merely as a decision support system: "To use an IT system for public administration is common and normal. This does not mean that the IT system made the decisions, rather, it is still the named public official who is responsible for each [individual] decision" [Statement of Defense, 2020]. Moreover, the PSA argued that parent preferences had been addressed and claimed that "it is the goal of the administration that to the highest extent possible ensure parental preferences, in line with the applicable legislation." The PSA did not provide any evidence in support of its claims.

Additional pleas by the parties followed over the summer of 2020, with the plaintiff arguing that "being responsible for a decision is simply not the same as being active in making a decision by looking at all relevant material and applying regulation" [Plaintiff's Clarification of Statement, 2020] and pointing out that 12,000 decisions executed over a very compressed time frame (one afternoon) suggests that one official could not realistically have been actively involved in making all of them. The PSA, in turn, repeated that "the named official is responsible for each decision" [Defendant's Clarification of Statement, 2020] and stressed again that the ADM system was solely used as support. The PSA also highlighted its social justice ambition and belief in equality through ADM: "It should be added that a centralized process means that all pupils in the city have been treated in an equal manner."

On August 27, 2020, the court ruled in favor of the PSA. Noting that the parties disagreed on whether (1) the decisions were automated and (2) parental preferences had been sufficiently taken into account, it concluded: "Under circumstances where the parties do not agree on the matters of a case being tried for legality, it is up to the plaintiff to disprove the claims of the administration and to prove that the decisions have been made by a computer program and not by administrators" [Court Verdict, 2020]. The burden of proof was thus placed on the plaintiff, while the submitted evidence (reports by the PSA internal auditors and by the parent group) was deemed insufficient. (The online transparency material provides an overview of arguments in the court case.)

[8] In Sweden's civil law system, disputes between individuals (citizens and noncitizens) and governmental actors are handled through the administrative court system. As the first instance, local administrative courts settle legal matters. These courts have as part of their brief a supporting inquisitory role, helping individuals to frame their legal questions, prepare the matter, and investigate relevant circumstances.

The court verdict was appealed to the Administrative Appellate Court, with the plaintiff citing the failure of the lower court to perform its investigative duties (asking to see the algorithm) or to reverse the burden of proof. The Appellate Court decided not to try the case, without further motivation. This decision was in turn appealed to the Supreme Administrative Court. At this time (March 2021), the city auditors published their report on the case, based on a thorough independent testing of the ADM system, including scrutiny of the algorithm [City Auditors Report, 2021]. The audit criticized the PSA's implementation of the ADM system and concluded that the ADM system had been coded incorrectly, in breach of laws and local regulations. The report also concluded that, had the ADM system been deployed correctly, almost all children would have been placed at the school of their choice [City Auditors Report, 2021]. The report confirmed the claims of the plaintiff and was therefore submitted as additional evidence to the Supreme Administrative Court, with an appeal for renewed consideration of the case in the light of new, publicly available evidence that strongly indicated that the initial verdict of the lower court was incorrect. The Supreme Administrative Court decided not to retry the case, without further motivation. This blocked the case from being considered, and the ADM system and its use remained legally unassessed.

### Analytical Summary of Phase 2

In Phase 2, the courts' responses to and treatment of the lawsuit contributed to a situation in which the algorithm was never scrutinized, which in turn made it impossible to determine legally whether the ADM system entailed unlawful elements that had caused systematic errors (which was a clear conclusion in the 2021 City Auditors Report). Systematic errors, affecting hundreds of children, remained unacknowledged and uncorrected (See the Appendix for an overview). Similar to the PSA, the court only acknowledged *existing categories of errors and regarded existing ways to handle them as sufficient*. Thus, despite the fact that ADM fundamentally alters the execution and consequences of decisions, the court applied the legal framework no differently than it would have in a case of manual decision-making. *Dismissing algorithmic agency and changes to the nature of decision-making* that came with ADM, while *supporting existing legal procedures as appropriate for correcting errors* resulted in a *failure to hold the public institution implementing the ADM (the PSA) accountable*.

However, the courts displayed not only an institutional *blindness in regard to new categories of errors* but also blindness to *new power imbalances* between the public institution using the ADM (with access to and control over it) and public service recipients without such access. This is further illustrated in the Catch-22 situation of the court demanding that the plaintiff should provide proof of ADM agency and consequences, while not facilitating access to it, by requiring the PSA to present this evidence. By Swedish law, nothing would have prevented the court from reversing the burden of proof and demanding that the PSA provide evidence supporting its position. The court could also have used its legally defined investigative duty to request that the PSA share the algorithm or demand that the PSA clarify the details concerning the use of the ADM system. By *refraining from using these options*, the court never had to rule on the legality of the algorithm. In this sense, the court *failed to acknowledge new systematic power imbalances*: If plaintiffs are required to present algorithms as evidence while not having any means to access these algorithms, unjust consequences of ADM will repeatedly persist. This indicates an *institutional blackboxing* of ADM systems where legal frameworks and norms are still compliant with a pre-algorithmic societal context and, hence, the production of *legal injustice*.

## Epilogue: How Nonengagement with the ADM System Produced Injustices

The PSA not only implemented an ADM system that produced thousands of erroneous decisions, but it also failed to correct the majority of them and obscured information about the situation (i.e., the ADM process and its outcome) in court. In a similar fashion, the court system failed to recognize the changes to decision-making that come with the implementation of ADM in the public sector. In the end, the court sustained and reinforced the injustice that the PSA had caused by not scrutinizing the legality of the algorithm and failing to hold the PSA accountable.

This injustice is likely to persist. To date, both the PSA and the court system only have procedures for identifying and addressing individual errors and appeals, consistent with a manual decision-making process; they have neglected datafication, technological inscribing, and the systematic nature of ADM consequences (Marjanovic et al., 2022) while also remaining blind to algorithmic agency and new power imbalances.

On the flip side, the court of public opinion and the work performed by the city auditors ultimately triggered change. A number of officials were fired from the PSA due to misconduct, the algorithm was reconfigured to become more compliant (but also more complex and more opaque), and the public became more aware of algorithmic risks.

In mid-2022, the Swedish municipality law was amended to allow for automated algorithmic decision-making. However, no structural, legal, or procedural changes have been implemented to ensure future accountability and to safeguard against injustices stemming from algorithmic decisions. As of June 2023, over 600 children still attend schools where they should not have been placed had the ADM system been configured, coded, and used according to Swedish law and local regulations.

## Discussion

Previous work on ADM and social justice has predominantly focused on the discriminatory outcomes and inequalities produced by algorithms through their design and use (Boyd & Crawford, 2012; Chouldechova, 2017; Fu et al., 2021; Marabelli, Vaast, & Jingya, 2021; Marjanovic et al., 2022; cf. Kordzadeh & Ghasemaghei, 2021). In some cases, such consequences are the flipside of intentional value capture conducted by powerful actors, such as the marginalization or exploitation of gig workers on online labor platforms (e.g., Wiener et al., 2021; Zheng & Wu, 2022). More often, however, the literature points to consequences that are unintended, unpredictable, harmful, and potentially wide-ranging (Marabelli, Vaast, & Jingya, 2021; Marjanovic et al., 2021; Rinta-Kahila et al., 2022).

Extending this literature, we foreground how institutional public actors—also beyond the organizations directly using ADM—make themselves and others unable to see and address such unintended consequences of ADM, thereby reinforcing, expanding, and prolonging the resulting injustices. More specifically, we provide a grounded conceptual model (Figure 2) explicating how three distinct ignoring practices produce a multilayered blackboxing of ADM, adding layers of opacity and inaccessibility of the ADM system and its unjust consequences. In doing so, we also highlight how institutions can contribute to the blackboxing of ADM and point at the active efforts underpinning the related creation of blindness (Knudsen, 2011; McGoey, 2019). This enables and reinforces algorithmic injustice (Marjanovic et al. 2022), including both social and legal dimensions.

### How Ignoring Practices Drive Social Injustice Consequences of Algorithmic Decision-Making: A Conceptual Model

The ignoring practices and associated layers of blackboxing depicted in our model can operate not only in relation to algorithms that are "inherently uninterpretable" (Rai, 2020, p. 138) but also in relation to those that (as in our case) are rule based and possible to scrutinize, that is, glassboxed or whiteboxed (Loyola-González, 2019; Rai, 2020; Rudin & Rudin, 2019). What makes such algorithms and their consequences opaque and blackboxed is not their complex learning mechanisms and hidden predictive calculations but rather the social and institutional wrapping around them. In our model, we visualize this argument with three boxes covering the ADM system. The ADM system (its technical features) is depicted as an inner box, which is traditionally seen as a technological blackbox (Rai, 2020; cf. discussion in Dolata et al., 2021). Below, we define and discuss the identified ignoring practices, how they blackbox the ADM, and how this produces and reinforces algorithmic injustices, relating our contributions to extant research about organizational ignoring, ADM, and social (in)justice.

The first ignoring practice, *avoiding* confrontation with ADM risks and consequences, refers to how the ADM-using organization ignores information that questions or complicates its ADM use, such as early warnings of design flaws and potential unintended consequences. This practice thus involves active efforts by the ADM-using organization to prevent information about ADM from being internally recognized and articulated, something the literature recognizes as self-imposed blindness (Bovensiepen & Palkmans, 2020; Essén et al. 2022; Knudsen, 2011), willful ignorance (Schaefer, 2019), functional stupidity (Alvesson & Spicer, 2012), and as "being aware and (at least publicly) unaware of something at the same time" (Zerubavel, 2006, p. 3). By "looking the other way" and remaining silent (p. 3) in relation to voiced concerns, the organization circumvents the need to scrutinize, evaluate, and internally communicate risks related to ADM. Indeed, through the first ignoring practice, the ADM-using organization turns the ADM risks into a latent secret: "information that is not yet information; it is something that avoids observation" (Knudsen, 2011, p. 966).

This self-directed ignoring practice contributes to the blackboxing of ADM risks within the organization (*intraorganizational blackboxing*). As a result, existing (and potentially underdeveloped) ADM practices are normalized within the organization, enabling the organization to proceed with ADM use in ways that violate rules and regulations (Lok & de Rond, 2013; cf. Ashforth & Anand, 2003; Heimer, 2012). This builds on and extends previous work suggesting that knowledge workers sometimes refrain from interrogating and thus trying to understand the algorithms they face in daily work (Anthony, 2021; Lebovitz et al., 2022). In particular, as observed in other ADM settings (Ballantyne, 2019; Benjamin, 2019; Eubanks, 2018; O'Neil, 2016), there were, in our case, early indications that the ADM enabled social injustice in terms of the maldistribution (Fraser, 2008; Marjanovic et al., 2022) of public resources (school slots). The first practice extends the insight that ADM can contribute to such unjust distribution of public resources by identifying how ADM-using organizations prevent themselves from seeing and acknowledging these unjust consequences.

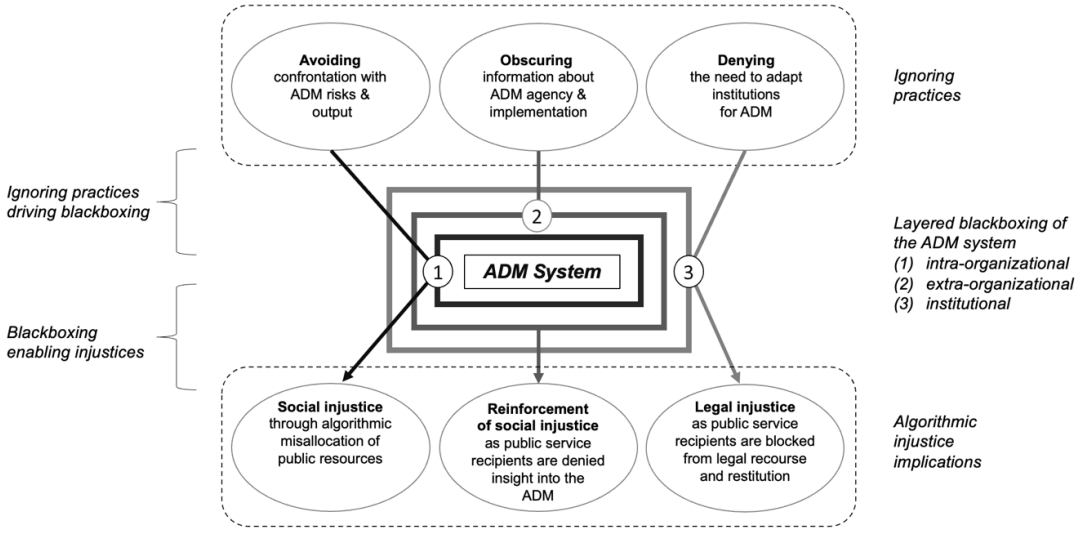



**Figure 2. Conceptual Model of Multilayered Blackboxing and Its Algorithmic Injustice Implications**

The second ignoring practice, *obscuring* information about the ADM and its implementation, refers to the ADM-using organization's production of blindness in its environment; that is, intentional efforts to shield the algorithm and its related use practices from external stakeholders, known as other-directed forms of strategic ignorance (Gross, 2007; McGoey, 2019; Oreskes & Conway, 2011; Proctor, 2012). Here, vital information about the ADM is withheld (e.g., when the public calls for transparency and requests access to the code) (cf. concealing; Oliver, 1991), delayed, and buffered by the ADM-using organization (e.g., by defining a narrow scope for audits that only serves as pseudo-transparency) (cf. Klintman, 2019; Proctor, 2012; Rayner, 2012). By preventing transparency, the organization shields the ADM system from being seen and scrutinized by the recipients of public services (those subject to the algorithmic decisions) as well as by the media and other institutions and organizations in the external context. This means that public service recipients are misrepresented in that they are excluded from participating in and influencing the implementation and use of ADM systems (Marjanovic et al., 2022).

This second ignoring practice thus produces an additional layer of blackboxing, making the ADM system, its implementation, and its socially unjust consequences increasingly inaccessible for external stakeholders (*extraorganizational blackboxing*). This blackboxing protects the organization, as well as its members, from external criticism and potential political intervention (Essén et al., 2022; Lok & de Rond, 2013; Oliver, 1991), thus reinforcing social injustice.

The third ignoring practice, *denying* the need to adapt institutions to ADM, refers to the reinforcement of macrolevel ignorance (McGoey, 2020) and institutional blindness (Krumer-Nevo et al., 2016) through which the structural disregard of potential problems is maintained (Bovensiepen & Palkmans, 2020; cf. Oliver, 1991). Here, explicit denial of the need to modify established rules (regulations, laws, or guidelines) and procedures contributes to a self-imposed blindness (Knudsen, 2011) at the institutional level. In our case, this is manifested in the courts (and individual judges) denying the need to update procedures to allow systematic redress and formally preventing themselves and other stakeholders from accessing evidence necessary for scrutinizing the ADM's legality and assigning accountability. This showcases a pattern of microlevel acts that contribute to macrolevel, or institutional, ignoring (McGoey, 2020). By maintaining that established institutional rules and procedures are appropriate and sufficient, this third ignoring practice protects and normalizes current institutional arrangements (Lok & de Rond, 2013; Oliver, 1991).

Notably, the third ignoring practice not only prevents an appropriate legal response to the immediate social injustices enabled by the ADM—in our case, the maldistribution of public resources (Marjanovic et al., 2022)—but also enables new injustices in terms of the misrepresentation of public service recipients. While previous work has pointed to the risk of misrepresentation in terms of these recipients being voiceless and not given rights to participate (or be represented) in the design and implementation of ADM systems (Marjanovic et al.,

2022; cf. Fraser, 2008), we identify an additional form of misrepresentation occurring in the legal arena: one in which public service recipients face significant barriers to understand whether and how they have been affected by ADM and are challenged to present evidence of defective or malfunctioning algorithms to the court. In combination with new power imbalances in which only the ADM-using organization controls access to and information about the ADM system, the absence of formal requirements to release the ADM code in explainable form leaves the affected individual voiceless. Our conceptualization thus showcases a new form of legal injustice. By not acknowledging that new power imbalances follow from ADM implementation and by denying the vital importance that access to the code has for the possibility of recourse and restitution courts can fail to safeguard the legal rights of public service recipients (Goldman & Cropanzano, 2015) subjected to ADM-enacted injustices.

In addition, the requirement placed on individuals to appeal a faulty decision creates additional injustices related to misrecognition (Marjanovic et al., 2022), as already-marginalized groups may not have the social or cultural standing and literacy to contest the decisions of public sector ADM. These forms of misrepresentation and misrecognition imply that the opportunities for public service recipients to participate democratically in the legal system—to get ADM errors corrected and to hold the responsible ADM-using organization accountable—are significantly constrained.

The third ignoring practice contributes to an institutional inability to address adverse consequences of ADM and to correct algorithmically performed (Marjanovic et al., 2022) social injustices through formal, legal means (*institutional blackboxing*). That is, the third ignoring practice, and the institutional blackboxing it implies, exacerbates already reinforced social injustices (here, the maldistribution of resources), by adding a legal dimension of algorithmic injustice (misrepresentation and misrecognition), hampering the ability of legal institutions to uphold justice (Tapp & Kohlberg, 1971).

To summarize, our model and related theorizing show why public institutions might not put things right when algorithms do wrong—and, indeed, how they might even exacerbate and prolong both social and legal injustice. The three ignoring practices reinforce each other, resulting in mutually supportive layers of blackboxing. For example, the deliberate nonconfrontation with ADM risk in the first practice is supported by the second practice's obscurement of information about ADM agency and consequences towards external stakeholders, which reduces external pressure internally to recognize (confront) these novel issues. By withholding and delaying ADM information (thus constraining available evidence, i.e., the very thing the court can assess), the second practice facilitates the third practice of denying the need to adapt institutional frameworks. Finally, the legal nonscrutiny of the ADM and the absence of institutional reforms resulting from the third ignoring practice also implicitly downplays ADM risks (and the legal risk of ignoring ADM risks) among ADM-using organizations, thus potentially facilitating the first practice. Therefore, in combination, the three practices constitute a self-reinforcing blackboxing dynamic in which the opportunities and incentives for relevant actors to scrutinize the ADM are increasingly narrowed (see, e.g., Sydow et al., 2009). This social and institutional blackboxing significantly diffuses accountability across a set of actors, colluding to maintain a status quo that can be detrimental to social and legal justice outcomes.

## Theoretical Contributions

Our conceptualization makes several contributions to the literature on algorithms and social justice, which predominantly theorizes how the design and use of ADM can lead to ethically problematic consequences (Giermindl et al., 2021; Kordzadeh & Ghasemaghei, 2021; Marabelli, Vaast, & Jingya, 2021; Marjanovic et al., 2021; Ransbotham et al., 2016; Rinta-Kahila et al., 2022; Schlagwein et al., 2019; van den Broek et al., 2021), proposing terms such as digital scars (Marabelli, Vaast, & Jingya, 2021) and algorithmic pollution (Marjanovic et al., 2021) to underscore how ADM can enable social injustices beyond the ADM-using organization. The literature also acknowledges the need for further research on how society responds to these new threats of potential algorithmic injustice (Dolata et al., 2021; Kordzadeh & Ghasemaghei, 2021; Marjanovic et al., 2022).

Specifically, Marjanovic et al. (2022) called for research into the legal basis on which citizens are being "denied rights to participate in key decisions about [ADM],[9] to know how [ADM] works, and [to know] how decisions are made about their requests for welfare" (p. 11). Our conceptual model and related theorizing provides an answer to this question, outlining how the maldistribution produced by the ADM is blackboxed through and for institutions, where the misrepresentation and misrecognition of affected groups primes them for legal injustice, implying new forms of injustice in the legal context. We thus also provide an answer to the question of how manifest cases of ADM-related social injustice can remain unaddressed.

[9] Marjanovic et al. (2022) refer to digital welfare systems but discuss them as examples of automated algorithmic decision-making.

Building on a sociotechnical perspective (e.g., Dolata et al., 2021; Lee, 2004; Mumford, 2006; Sarker et al., 2019), our notion of layered blackboxing thus moves beyond the technology-centric view of blackboxing widespread in the literature on ADM and social justice (Dutta et al., 2020; Loyola-González, 2019; Rai, 2020; Rudin & Rudin, 2019). In doing so, we emphasize that the inscrutability of algorithms is not (only) an attribute of the technology as such (Lebovitz et al., 2022). Specifically, we extend recent work suggesting that opacity is partly produced by work practices surrounding AI (Anthony, 2021; Lebovitz et al., 2022). We do so by showing how blackboxing is produced in intraorganizational, extraorganizational, and institutional domains. We also do so by theorizing how ignoring practices produces layers of blackboxing that obscure ADM and its consequences, leading to injustice implications. Relatedly, we show how layered blackboxing prevents actors from addressing and contesting the consequences of ADM, even to the point of legal systems being blind to such consequences. When blackboxing becomes institutional, avenues to recourse and restitution become blocked, leading public institutions to exhibit traits of macro-ignoring (McGoey, 2020) of ADM as a phenomenon requiring institutional adaptation.

We also observe that features of ADM, such as technological inscribing, datafication, and systematic effects tied to algorithmic social injustice (Marjanovic et al., 2022; cf. Dencik et al., 2018) are prone to being ignored and becoming blackboxed because they are not readily visible, understood, or expected. This ties in with the notion of algorithms changing the nature of social justice in abnormal times (Fraser, 2008; Marjanovic et al., 2022).

Rather than focusing only on algorithms as inherently interpretable or uninterpretable (cf. Rai, 2020) and on algorithms by themselves causing algorithmic injustice (cf. Marjanovic et al., 2022), we suggest that injustices—and the layered blackboxing of them—can also be traced to institutional arrangements. This view of blackboxing suggests that the practices around an algorithm can render even simple rule-based, interpretable algorithms shielded from access, interpretation, contestation, and remedy. Further, this demonstrates the relevance of extending the concept of algorithmic injustice (Marjanovic et al. 2022) to include not only a social but also a legal dimension of injustice and to distinguish the two to understand their interrelationship.

Finally, we also provide a contribution to the literature on organizational ignoring and institutional blindness, which accounts for how institutions (Krumer-Nevo et al., 2016; Lok & de Rond, 2013; Oliver, 1991) and organizations can blind themselves (Alvesson & Spicer, 2012; Essén et al., 2022; Knudsen, 2011; Schaefer, 2019) or others (Gross, 2007; McGoey, 2019, 2020; Oreskes & Conway, 2011; Proctor, 2012) to phenomena. This literature tends to theorize these processes in isolation, often focusing only on their consequences for the organization. In contrast, we show how the combination of self-directed (Essén et al., 2022), other-directed (Gross, 2007; Proctor, 2012), and institutionally prescribed (Lok & de Rond, 2013; McGoey, 2020) ignoring practices can lead to multilayered blackboxing of unethical or unjust consequences of new technology and how this can lead to sustained injustices for which accountability is suspended.

## *Limitations and Future Research*

Our study addresses the introduction of ADM into a public-sector setting. Other sectors, with different levels of maturity concerning ADM and other governance principles might exhibit other patterns. Furthermore, our study is situated in a civil law setting with high trust in public institutions, high levels of transparency in public administration, and low levels of corruption. Thus, our findings might not correspond to a different institutional context, such as a common-law context. In addition, our case did not involve advanced algorithms using, e.g., deep learning. We would argue, however, that the particularly uncomplicated characteristics of the algorithm, and the favorable institutional context it is applied in—making this a best-case scenario in several regards—strengthens the relevance and importance of our findings. In other words, if this can happen with a technologically simple algorithm, what might be the consequences when the algorithm itself is undeniably opaque and inscrutable? If this can happen in a context with high transparency and high trust in public institutions, what might happen in a setting that is more authoritarian and/or corrupt? We hope that questions such as these will inspire further research.

Moreover, our study accounts for (non)responses to ADM consequences among public organizations, but we did not capture the individual sensemaking processes of the involved organizational members (Essén et al., 2022; Vaughan, 1997; Weick, 1993). While such considerations by individual actors were outside our scope, understanding why individuals ignore warnings and refrain from seeking knowledge about destructive ADM consequences is an important topic for further research. Our work also indicates that the ambition to protect the integrity and legitimacy of institutions plays a role in actors ignoring ADM. This suggests that it might be useful to explore how institutional maintenance (Lawrence & Suddaby, 2006; Lok & de Rond, 2013) operates in relation to the introduction of ADM and how this, in turn, could impede the institutional adaptation needed to realize the constructive potential of technology fully (Sahay et al., 2019).

A grounded case study is not designed for theory testing or statistical generalization but for theory building. While this should be seen as an inherent characteristic of the research approach and not a limitation as such, it nevertheless opens the way for further studies to advance our understanding of how various forms of organizational ignoring and layered blackboxing in relation to ADM can marginalize as well as benefit individuals and organizations.

Our study also indicates that the idea of ensuring there are "humans-in-the-loop" (Dolata et al., 2021; Grønsund & Aanestad, 2020; Marabelli, Newell, & Handunge, 2021; Teodorescu et al., 2021) can be further interrogated and developed. In our case, the humans involved did not add safeguards, fairness, or transparency to the use of the ADM system. In fact, the opposite occurred. Thus, extending the idea of engaging a wider set of stakeholders and views in the training and monitoring of ADM systems (society-in-the-loop; Rahwan, 2018), we stress the need to study institutional safeguards-in-the-loop; that is, additional layers of protection against algorithmic misconduct beyond corrective algorithmic design and human (users)-in-the-loop. We propose this as an important subject for research as well as practice.

Finally, we are encouraged by the outcome of our field intervention (Zuiderent-Jerak, 2015), which generated data that would otherwise not have been available. Active interventions as a form of engaged scholarship (van de Ven, 2007) might be particularly warranted to research new digital phenomena with unknown or uncertain and potentially wide-ranging consequences. Based on our field intervention, we see potential in using active interventions in other settings to pursue a deeper understanding of algorithmic agency and its consequences before widespread diffusion has occurred and institutional responses have been developed.

## Practical Implications

At the start of this paper, we posed the rhetorical question: When algorithms do wrong, will our public institutions put things right? The answer emerging from our study is: no, at least not reliably. Hence, our study points to the need to develop strategies to counter the ignoring practices and resulting types of blackboxing that we identify—in particular, those related to institutions that are not adapted to new algorithmic circumstances and dynamics (United Nations General Assembly, 2019). Moreover, our study implies that calls for explainable AI (Asatiani et al., 2020, 2021; Loyola-González, 2019; Rudin & Rudin, 2019) and related regulatory initiatives, such as the EU Artificial Intelligence Act (European Commission, 2021) might be missing an important part of the problem. While lower complexity algorithms are often perceived as correlated to lower risks (European Commission 2021; Nagtegaal, 2021), we show that detrimental ADM outcomes and related impediments to legal restitution can be substantial even when technical complexity and opacity is low. In short, uncomplicated ADM technology is no safeguard for justice. Moreover, our insights into the role of ignoring practices and layered blackboxing for algorithmic injustice prompt the consideration of new measures to prevent and counteract these practices and outcomes. Specifically, we need to consider principles that allow us to counteract and address different layers of blackboxing. We therefore propose a legal framework (Table 2) that incorporates layers of blackboxing, (including technological blackboxing, for completeness) and articulates guiding principles and recommended actions to help to make ADM more just and ethical, with a particular focus on structural remedies such as regulations, rules, and governance mechanisms.

Concerning technological blackboxing, algorithmic *explainability* is commonly proposed in both IS and legal communities (Nagtegaal, 2021; Rai, 2020; Rudin & Rudin, 2019). Additional remedies could also include responsibility among technology developers and public agencies using ADM to consider other means for increasing *transparency*, such as disclosure of key variables (Marabelli, Vaast, & Jingya, 2021) and providing access to information about the ADM system. For intraorganizational blackboxing, we see *engagement* and *responsibility* as key considerations, with particular emphasis on how public agencies using ADM can be held accountable for articulating the nature of ADM use, risks, consequences, and regulatory compliance.

Extraorganizational aspects include addressing the increasing power distance between actors controlling the use of ADM systems and those subject to their outcomes. By holding public agencies accountable for how they communicate and provide *access* to information about ADM, the *power balance* can be improved. Concrete measures here include declarations, overviews of how rules are translated into code, and requiring public agencies to be ready and able to prove the legality of algorithms.

Furthermore, we need to acknowledge that when ADM systems fail and erroneous outcomes are produced by public-service agencies, it is ultimately up to legal institutions to safeguard the rights of individuals, showing the crucial value of addressing legal *recourse and restitution*. We need legal institutions to be empowered—through laws as well as competencies—to assess the legality of algorithms and their use and to hold actors accountable when ADM systems fail. This requires courts to see and understand algorithms, their agency, and their consequences and to determine their legality and accountability (Sannerholm, 2022; Stahl, 2008). We also see a need to hold legislative actors accountable for adapting institutions to ensure fair legal proceedings for individuals subjected to public-sector ADM, with particular care taken to address the systematic nature of ADM consequences.

**Table 2. A Framework for Preventing, Counteracting, and Remedying Detrimental Consequences of ADM in Public Administration**

| Blackboxing layer | Guiding principles | Recommended actions: regulations, rules, governance mechanisms | Implications for society/service recipients |
|---|---|---|---|
| Technological: inscrutability of algorithms (cf. Rai, 2020; Rudin & Rudin, 2019) | Explainability/ transparency | • Technology developers are responsible for developing and providing open/self-explanatory code and relevant documentation<br>• Technology developers are responsible for providing representation of input variables, and mechanisms underlying the functioning of ADM in a way interpretable to users<br>• Public agencies are responsible for ensuring that the points above are included in public procurement agreements of ADM (not allowing code to be a trade secret) | Fundamental to enable critical scrutiny by any party |
| Intra-organizational: inscrutability of ADM system(s) within the ADM-using organization | Engagement/ responsibility | • Public agencies using ADM carry accountability (at individual and organizational level) for explaining ADM use, risks, and consequences. This encompasses formal responsibility to monitor the regulatory compliance, risk, and consequences of ADM systems<br>• Public agencies using ADM are responsible for verifying that the code and processes of ADM systems correspond to applicable regulations and make public translations of regulations into code | Ensures engagement with ADM and enables both prevention and early detection and correction of unintended detrimental ADM consequences |
| Extra-organizational: inscrutability of ADM system(s) outside the ADM-using organization: public service recipients, the media, the general public, and other external actors | Access/ power balance | • Public agencies using ADM are required to issue declarations of use of ADM ("we use ADM for this decision")<br>• Public agencies using ADM are responsible for providing code and ADM system information to service recipients, as well as to the media and external parties (e.g., external auditors) on request according to the Freedom of Information Act or equivalent legislation<br>• Public agencies using ADM are responsible for providing an overview that summarizes what the algorithms do (for instance a schema/graphic illustrating in overarching terms what the operation of the software does: data used, code, regulations complied with, etc.)<br>• Public agencies using ADM with information on and access to implemented ADM systems should be obliged to prove their legality in contestations and court hearings | Enables critical scrutiny of ADM systems and implementations by external actors (e.g., the media and the public)<br>Empowers public service recipients to investigate and take action |
| Institutional: inscrutability of ADM system(s) caused by courts, laws, regulations, and norms | Recourse/ restitution | • Legislative actors should update laws to acknowledge the systematic nature of ADM decisions and their consequences by allowing systemic appeals and legality assessments of code (e.g., by instituting preliminary decisions with an appeals period before ADM decisions are finalized and executed)<br>• Legislative actors should update laws to require a systematic review and systematic correction of ADM outcomes when an ADM is proven to give erroneous outcomes<br>• Courts and judges should be given resources and training to improve their ability to assess ADM systems<br>• Legislative actors should (when applicable) establish legal procedures that shift the burden of proof of ADM legality from the plaintiff (e.g., service recipient) to the public agency using the ADM<br>• Governmental actors should institute an *ombudsperson* (public advocate) for effective support to the public to contest ADM outcomes | Enables access to effective hearings and potentially allows redress/compensation for service recipients who experience harmful ADM consequences<br><br>Incentivizes compliance by user organizations with the above guidelines (as noncompliance can have legal implications) |

The proposed framework highlights that blackboxing is not only a problem to solve for technology developers or the public agencies using ADM systems; instead, it is of particular importance that legislators and legal professionals also proactively consider how laws and regulations need to be adapted to an algorithmic world, for instance in regard to burden-of-proof reversals when public agencies employ ADM systems or through appellate processes targeting algorithms that allow for systematic reversal of decisions taken in bulk.

These suggestions are, however, not likely to be fully effective in ensuring that particularly vulnerable groups can pursue legal recourse and restitution. We therefore suggest the creation of an *ombudsperson* for algorithmic justice to help social service recipients safeguard their rights and serve as a counterweight to the growing power imbalance between ADM-using public organizations and public service recipients. An ombudsperson would be especially helpful for groups that face multiple risks of exposure to algorithmic injustice, such as being overrepresented in the need of public welfare (and the risk of maldistribution) and/or being marginalized in terms of cultural status, with increased risks of misrepresentation and misrecognition (Fraser, 2008; Marjanovic et al., 2022). Such an ombudsperson could be created on various administrative levels, from local to national arenas.

## Conclusion

Previous work on ADM and social justice has predominantly focused on outlining the discriminatory outcomes and inequalities produced by algorithms through design and use within specific organizational contexts (Marabelli, Vaast & Jingya, 2021; Marjanovic et al., 2022; cf. Kordzadeh & Ghasemaghei, 2021). We provide a theoretical explanation for how unjust consequences of ADM can go unaddressed by public institutions, foregrounding how organizational practices of ignoring (Gross, 2007; Krumer-Nevo et al., 2016; McGoey, 2019, 2020) can result in intraorganizational, extraorganizational, and institutional blackboxing of ADM, with these multiple layers of inaccessibility reinforcing each other. We also demonstrate the relevance of nuancing descriptions of (in)justice, proposing that the social justice consequences of ADM relate to legal justice, including the (in)ability of legal institutions to correct unjust or unlawful ADM consequences and assigning accountability for algorithmic malpractices. Thus, our study provides important insights into how a "digital welfare dystopia" in public services (United Nations General Assembly, 2019, p. 2) might emerge and elucidates the nature of the problem we need to address to harness the power of algorithms better in societal domains.

## Acknowledgments

We are most grateful for the excellent guidance provided by the special issue senior editor team (with Margunn Aanestad as lead on this article) and the associate editor and for the helpful feedback and suggestions from the three anonymous reviewers. We would also like to extend our thanks for valuable input from conference and research seminar participants at Cambridge Judge Business School, Stevens Institute of Technology, the 2022 EGOS conference (Subtheme 23), the "AI and Emerging Technologies: Beyond the Hype" conference at LUISS (May 2023), a Swedish Center for Digital Innovation workshop, and several seminars at the House of Innovation at Stockholm School of Economics. We are also grateful for individual input and advice from many colleagues, particularly Cathy Urquhart, Robert Gregory, and Stanley Greenstein. All remaining errors are our own. Financial support for this research from the Marianne and Marcus Wallenberg Foundation (Grant 2021.0074), the Knut and Alice Wallenberg Foundation (2020.0243), the Swedish Research Council for Health, Working Life and Welfare (Grant 2021-00582), the Jan Wallander and Tom Hedelius Foundation (Grant P21-0067), and the Erling-Persson Foundation is gratefully acknowledged.

## References

Aanestad, M., Kankanhalli, A., Maruping, L., Pang, S., & Ram, S. (2021). Call for papers: MISQ digital technologies and social justice. *MIS Quarterly*. https://misq.org/skin/frontend/default/misq/pdf/CurrentCalls/SI_DigitalTechnologies.pdf

Alvesson, M., & Spicer, A. (2012). A stupidity-based theory of organizations. *Journal of Management Studies*, *49*(7), 1194-1220. https://doi.org/10.1111/j.1467-6486.2012.01072.x

Amnesty International. (2021). *Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal*. Index Number: EUR 35/4686/2021. https://www.amnesty.org/en/documents/eur35/4686/2021/en/

Angelle, P. S., & Torrance, D. (2019). The significance of context in the enactment of social justice. In P. S. Angelle & D. Torrance (Eds.), *Cultures of social justice leadership* (pp. 195-207). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-10874-8

Anthony, C. (2021). When knowledge work and analytical technologies collide: The practices and consequences of black boxing algorithmic technologies. *Administrative Science Quarterly*, *66*(4), 1173-1212. https://doi.org/10.1177/00018392211016755

Asatiani, A., Malo, P., Nagbøl, P. R., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2020). Challenges of explaining the behavior of black-box AI systems. *MIS Quarterly Executive*, *19*(4), 259-278. https://doi.org/10.17705/2msqe.00037

Asatiani, A., Malo, P., Nagbøl, P. R., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence systems. *Journal of the Association for Information Systems*, *22*(2), 325-352. https://doi.org/10.17705/1jais.00664

Ashforth, B. E., & Anand, V. (2003). The normalization of corruption in organizations. *Researching Organizational Behavior*, *25*, 1-52. https://doi.org/10.1016/s0191-3085(03)25001-2

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, *355*(6324), 483-485. https://doi.org/10.1126/science.aal4321

Bachrach, P., & Baratz, M. S. (1963). Decisions and nondecisions: An analytical framework. *American Political Science Review*, *57*, 632-642. https://doi.org/10.2307/1952568

Ballantyne, N. (2019). The ethics and politics of human service technology: The case of predictive risk modelling in New Zealand's child protection system. The Hong Kong *Journal of Social Work*, *53*(1), 15-27. https://doi.org/10.1142/s0219246219000044

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, *104*(671), 671-732. https://doi.org/10.15779/Z38BG31

Benjamin, R. (2019). Race after technology. *Polity Press*. https://doi.org/10.3917/res.229.0255

Bingham, L. B. (2008). Designing justice: Legal institutions and other systems for managing conflict. *Ohio State Journal on Dispute Resolution*, *24*(1), 1-50. https://escholarship.org/uc/item/8ph0g1zz

Bovensiepen, J., & Pelkmans, M. (2020). Dynamics of wilful blindness: An introduction. *Critique of Anthropology*, *40*(4), 387-402. https://doi.org/10.1177/0308275X20959

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662-679. https://doi.org/10.1080/1369118x.2012.678878

Brodkin, E. Z., & Marston, G. (2013). *Work and the welfare state. Street-level organizations and workfare politics.* Georgetown University Press. https://doi.org/10.1353/book26636

Brussig, M., & Knuth, M. (2013). Good intentions and institutional blindness: Migrant populations and the implementation of German activation policy. In E. Z. Brodkin & G. Marston (Eds.), *Work and the welfare state* (pp. 185-208). Georgetown University Press. https://doi.org/10.1353/book26636

Brunsson, K. (1998). Non-learning organizations. *Scandinavian Journal of Management*, *14*(4), 421-432. https://doi.org/10.1016/S0956-5221(98)00002-5

Brunsson, N. (1982). The irrationality of action and action rationality: Decisions, ideologies and organizational actions. *Journal of Management Studies*, *19*(1), 29-44. https://doi.org/10.1111/j.1467-6486.1982.tb00058.x

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data and Society*, *3*(1). https://doi.org/10.1093/oso/9780199206285.003.0003

Cecez-Kecmanovic, D. (2022, June 25-28). *Algorithmic automation of knowledge work—A sketch for exploring the becoming of knowledge work*. Presented at the 13th International Process Symposium. Cyprus.

Charmaz, K. (2011). Grounded theory methods in social justice research. *Strategies of Qualitative Inquiry*, *4*(1), 359-380.

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). SAGE. https://doi.org/10.7748/nr.13.4.84.s4

Christin, A. (2020). The ethnographer and the algorithm: Beyond the black box. *Theory and Society*, *49*(5), 897-918. https://doi.org/10.1007/s11186-020-09411-3

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, *5*(2), 153-163. https://doi.org/10.1089/big.2016.0047

Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: Big data and the changing context of strategy. *Journal of Information Technology*, *30*(1), 44-57. https://doi.org/10.1057/jit.2014.17

Dacin, M., Munir, K., & Tracey, P. (2010). Formal dining at Cambridge colleges: Linking ritual performance and institutional maintenance. *The Academy of Management Journal*, *53*(6), 1393-1418. https://doi.org/10.5465/amj.2010.57318388

Dencik, L., Jansen, F., & Metcalfe, P. (2018). *A conceptual framework for approaching social justice in an age of datafication* (Data Justice Lab Working Paper). Retrieved September 2023 from datajusticeproject.net.

Dolata, M., Feuerriegel, S., & Schwabe, G. (2021). A sociotechnical view of algorithmic fairness. *Information Systems Journal*, *32*(4), 754-818. https://doi.org/10.1111/isj.12370

Du, M., Liu, N., & Hu, X. (2018). Techniques for interpretable machine learning. *Communications of the ACM, 63*(1), 68-77. https://doi.org/10.1145/3359786

Dutta, S., Wei, D., Yueksel, H., Chen, P. Y., Liu, S., & Varshney, K. (2020). Is there a trade-off between fairness and accuracy? A perspective using mismatched hypothesis testing. In *Proceedings of the 37th International Conference on Machine Learning* (pp. 2803-2813).

Essén, A., Knudsen, M., & Alvesson, M. (2022). Explaining ignoring: Working with information that nobody uses. *Organization Studies*, *43*(5), 725-747. https://doi.org/10.1177/0170840621998720

Eubanks, V. (2018). *Automating inequality*. Picador.

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* (Document No. 52021PC0206). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, *28*(1), 62-70. https://doi.org/10.1016/j.infoandorg.2018.02.005

Feuerriegel, S., Dolata, M. & Schwabe, G. (2020). Fair AI. *Business & Information Systems Engineering*, *62*, 379-384. https://doi.org/10.1007/s12599-020-00650-3

Fraser, N. (2008). Abnormal justice. *Critical Inquiry*, *34*(3), 393-422. https://doi.org/10.1086/589478

Fu, R., Huang, Y., & Sing, P. V. (2021). Crowds, lending, machine, and bias. *Information Systems Research*, *32*(1), 72-92. https://doi.org/10.1287/isre.2020.0990

Galliers, R., Newell, S., Shanks, G., & Topi, H. (2017). Datafication and its human, organizational and societal effects. *Journal of Strategic Information Systems*, *26*(3), 185-190. https://doi.org/10.1016/j.jsis.2017.08.002

Gianfrancesco, M. A., Tamang, S., & Yazdany, J. (2018). Potential biases in machine learning algorithms using electronic health record data. *JAMA Internal Medicine*, *178*(11), 1544-1547. https://doi.org/10.1001/jamainternmed.2018.3763

Giermindl, L., Strich, F., Christ, O., Leicht-Deobald, U., & Redzepi, A. (2022). The dark sides of people analytics: Reviewing the perils for organisations and employees. *European Journal of Information Systems*, *31*(3), 410-435. https://doi.org/10.1080/0960085X.2021.1927213

Gillingham, P. (2019). Decision support systems, social justice and algorithmic accountability in social work: A new challenge. *Practice*, *31*(4), 277-290. https://doi.org/10.1080/09503153.2019.1575954

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine. https://doi.org/10.4324/9780203793206-1

Goldman, B., & Cropanzano, R. (2015). "Justice" and "fairness" are not the same thing. *Journal of Organizational Behavior*, *36*(2), 313-318. https://doi.org/10.1002/job.1956

Gross, M. (2007). The unknown in process: Dynamic connections of ignorance, non-knowledge and related concepts. *Current Sociology*, *55*(5), 742-759. https://doi.org/10.1177/0011392107079928

Gross, M., & McGoey, L. (Eds.). (2015). *The Routledge international handbook of ignorance studies*. Routledge. https://doi.org/10.4324/9781315867762

Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *Journal of Strategic Information Systems*, *29*(2), 1-16. https://doi.org/10.1016/j.jsis.2020.101614

Heikillä, M. (2022). *Dutch scandal serves as a warning for Europe over risks of using algorithms.* Politico. https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/

Heimer, C. (2012). Inert facts and the illusion of knowledge: Strategic uses of ignorance in HIV clinics. *Economy and Society*, *41*(1), 17-41. https://doi.org/10.1080/03085147.2011.637332

Hoffmann, A. L. (2019). Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society*, *22*(7), 900-915. https://doi.org/10.1080/1369118x.2019.1573912

Kaun, A. (2021). Suing the algorithm: The mundanization of automated decision making in public services through litigation. *Information, Communication & Society*, *25*(14), 2046-2062. https://doi.org/10.1080/1369118X.2021.1924827

Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy artificial intelligence: A review. *ACM Computing Surveys*, *55*(2), 1-3. https://doi.org/10.1145/3491209

Keddell, E. (2019). Algorithmic justice in child protection. *Social Sciences*, *8*(281), 1-22. https://doi.org/10.3390/socsci8100281

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, *14*(1), 366-410. https://doi.org/10.5465/annals.2018.0174

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, *34*(11), 2767-2787. https://doi.org/10.1016/j.jbankfin.2010.06.001

Klintman, M. (2019). *Knowledge resistance*. Manchester University Press.

Knudsen, M. (2011). Forms of inattentiveness: The production of blindness in the development of a technology for the observation of quality in health services. *Organization Studies*, *32*(7), 963-989. https://doi.org/10.1177/0170840611410827

Kordzadeh, N., & Ghasemaghei, M. (2021). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, *31*(3), 388-409. https://doi.org/10.1080/0960085x.2021.1927212

Krakowski, S., Luger, J., & Raisch, S. (2023). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal*, *44*(6), 1425-1452. https://doi.org/10.1002/smj.3387

Krumer-Nevo, M., Elfassi, Y., Sagy, S., & Lavie-Ajayi, M. (2016). Neither seeing nor seen: Exclusion and double exclusion in the lives of young people involved in the drug trade in Israel. *Young*, *24*(1), 36-52. https://doi.org/10.1177/1103308815613683

Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Harvard University Press.

Lawrence, T. B., & Suddaby, R. (2006). Institutions and institutional work. In S. R. Clegg, C. Hardy, T. B. Lawrence & W. R. Nord (Eds.) *The SAGE handbook of Organization Studies*. SAGE. https://doi.org/10.4135/9781848608030

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science*, *33*(1), 126-148. https://doi.org/10.1287/orsc.2021.1549

Lee, A. S. (2004). Thinking about social theory and philosophy for information systems. In J. Mingers & L. Willcocks (Eds.), *Social theory and philosophy for information systems*, (pp. 1-26). John Wiley & Sons.

Lens, V. (2007). Administrative justice in public welfare bureaucracies: When citizens (don't) complain. *Administration & Society*, *39*(3), 382-408. https://doi.org/10.1177/0095399707301006

Lettinga, D., & van Troost, L. (Eds.) (2015). *Can human rights bring social justice?* Amnesty International Netherlands. https://www.amnesty.nl/content/uploads/2015/10/can_human_rights_bring_social_justice.pdf

Lok, J., & de Rond, M. (2013). On the plasticity of institutions: Containing and restoring practice breakdowns at the Cambridge University Boat Club. *Academy of Management Journal*, *56*(1), 185-207. https://doi.org/10.5465/amj.2010.0688

Loyola-González, O. (2019). Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. *IEEE Access*, *7*, 154096-154113. https://doi.org/10.1109/access.2019.2949286

Marabelli, M., Newell, S., & Handunge, V. (2021). The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges. *Journal of Strategic Information Systems*, *30*(2021), Article 101683. https://doi.org/10.1016/j.jsis.2021.101683

Marabelli, M., Vaast, E., & Jingya, L. L. (2021). Preventing the digital scars of COVID-19. *European Journal of Information Systems*, *30*(2), 176-192. https://doi.org/10.1080/0960085X.2020.1863752

Marjanovic, O., & Cecez-Kecmanovic, D. (2017). Exploring the tension between transparency and datafication effects of open government IS through the lens of complex adaptive systems. *Journal of Strategic Information Systems*, *26*(3), 210-232. https://doi.org/10.1016/j.jsis.2017.07.001

Marjanovic, O., Cecez-Kecmanovic, D., & Vidgen, R. (2021). Algorithmic pollution: Making the invisible visible. *Journal of Information Technology*, *36*(4), 391-408. https://doi.org/10.1177/02683962211010356

Marjanovic, O., Cecez-Kecmanovic, D., & Vidgen, R. (2022). Theorising algorithmic justice. *European Journal of Information Systems*, *31*(3), 269-287. https://doi.org/10.1080/0960085X.2021.1934130

Martin, K. (2019). Designing ethical algorithms. *MIS Quarterly Executive*, *18*(2), 129-142. https://doi.org/10.17705/2msqe.00012

Meyer, J. W., & Rowan, B. (1977). Institutional organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, *80*, 340-363. https://doi.org/10.1086/226550

McGoey, L. (2019). *The unknowers: How strategic ignorance rules the world*. Zed Books. https://doi.org/10.5040/9781350225725

McGoey, L. (2020). Micro-ignorance and macro-ignorance in the social sciences. *Social Research: An International Quarterly*, *87*(1), 197-217. https://doi.org/10.1353/sor.2020.0014

Mumford, E. (2006). The story of socio-technical design: Reflections on its successes, failures and potential. *Information Systems Journal*, *16*(4), 317-334. https://doi.org/10.1111/j.1365-2575.2006.00221.x

Möhlmann, M., Zalmanson, L., Henfridsson, O., & Gregory, R. W. (2021). Algorithmic management of work on online labor platforms: When matching meets control. *MIS Quarterly*, *45*(4), 1999-2022. https://doi.org/10.25300/MISQ/2021/16280

Nagtegaal, R. (2021). The impact of using algorithms for managerial decisions on public employees' procedural justice. *Government Information Quarterly*, *38*(1), Article 101536. https://doi.org/10.1016/j.giq.2020.101536

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Penguin Random House.

Oliver, C. (1991). Strategic responses to institutional processes. *Academy of Management Review*, *16*(1), 145-179. https://doi.org/10.5465/amr.1991.4279002

Oreskes, N., & Conway, E. M. (2011). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming* (reprint ed.). Bloomsbury Press.

Orlikowski, W., & Gash, D. (1994). Technology frames: Making sense of information technology in organizations. *ACM Transactions on Information Systems*, *12*(2), 174-207. https://doi.org/10.1145/196734.196745

Pasquale, F. (2015). *The Black Box Society: The secret algorithms that control money and information.* Harvard University Press. https://doi.org/10.4159/harvard.9780674736061

Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science*, *14*(3), 399-441. https://doi.org/10.1177/030631284014003004

Pinch, T. J. (1992). Opening black boxes: Science, technology and society. *Social Studies of Science*, *22*(3), 487-510. https://doi.org/10.1177/030631279202200300 3

Proctor, R. N. (2012). *Golden holocaust: Origins of the cigarette catastrophe and the case for abolition*. University of California Press. https://doi.org/10.1525/9780520950436

Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology*, *20*(1), 5-14. https://doi.org/10.1007/s10676-017-9430-7

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137-141. https://doi.org/10.1007/s11747-019-00710-5

Ransbotham, S., Fichman, R., Ram, G., & Gupta, A. (2016). Special section introduction—Ubiquitous IT and digital vulnerabilities. *Information Systems Research*, *27*(4), 834-847. https://doi.org/10.1287/isre.2016.0683

Rayner, S. (2012). Uncomfortable knowledge: The social construction of ignorance in science and environmental policy discourses. *Economy and Society*, *41*(1), 107-125. https://doi.org/10.1080/03085147.2011.637335

Rinta-Kahila, T., Someh, I., Gillespie, N., Indulska, M., & Gregor, S. (2022). Algorithmic decision making and system destructiveness: A case of automatic debt recovery. *European Journal of Information Systems*, *31*(3), 313-338. https://doi.org/10.1080/0960085X.2021.1994611

Robbins, I. P. (1990). The Ostrich Instruction: Deliberate ignorance as a criminal mens rea. *Journal of Criminal Law & Criminology*, *81*(1), 191-234. https://doi.org/10.2307/1143851

Rudin, C., & Radin, J. (2019). Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harvard Data Science Review*, *1*(2). https://doi.org/10.1162/99608f92.5a8a3a3d

Rosenblat, A., Wikelius, K., Boyd, D., Gangadharan, S. P., & Yu, C. (2014). *Data & civil rights: Criminal justice primer.* Data and Society Research Institute. https://datasociety.net/output/data-civil-rights-criminal-justice-primer

Sahay, S., Nielsen, P., & Aanestad, M. (2019). Institutionalizing information systems for universal health coverage in primary healthcare and the need for new forms of institutional work. *Communications of the Association for Information Systems*, *44*, 1-33. https://doi.org/10.17705/1CAIS.04401

Sandel, M. J. (2020). *The tyranny of merit: What's become of the common good?* Allen Lane.

Sannerholm, R. (2022). Responsibility and accountability: AI, governance, and the rule of law. In L. Colonna & S. Greenstein (Eds.), *Law in the era of artificial intelligence* (pp. 223-246). Swedish Law and Informatics Research Institute. https://doi.org/10.53292/208f5901.40d940a1

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The sociotechnical axis of cohesion for the IS discipline: Its historical legacy and its continued relevance. *MIS Quarterly*, *43*(3), 695-719. https://doi.org/10.25300/MISQ/2019/13747

Schaefer, S. M. (2019). Willful managerial ignorance, symbolic work and decoupling: A socio-phenomenological study of "managing creativity." *Organization Studies*, *40*(9), 1387-1407. https://doi.org/10.1177/0170840618814571

Schlagwein, D., Cecez-Kecmanovic, D., & Hanckel, B. (2019). Ethical norms and issues in crowdsourcing practices: A Habermasian analysis. *Information Systems Journal*, *29*(4), 811-837. https://doi.org/10.1111/isj.12231

Sheehey, B. (2018). Algorithmic paranoia: The temporal governmentality of predictive policing. *Ethics and Information Technology*, *21*(1), 49-58. https://doi.org/10.1007/s10676-018-9496-1

Stahl, B. C. (2008). Social justice and market metaphysics: A critical discussion of philosophical approaches to digital divides. In F. B. Tan, (Ed.), *Global information technologies: Concepts*, *methodologies*, *tools*, *and applications* (pp. 3345-3356). IGI Global. https://doi.org/10.4018/978-1-59904-939-7.ch238

Sydow, J., Schreyögg, G., & Koch, J. (2009). Organizational path dependence: Opening the black box. *Academy of Management Review*, *3*(4), 689-709. https://doi.org/10.5465/amr.2009.44885978

Tapp, J. L., & Kohlberg, L. (1971). Developing senses of law and legal justice. *Journal of Social Issues*, *27*(2), 65-91. https://doi.org/10.1111/j.1540-4560.1971.tb00654.x

Teodorescu, M., Morse, L., Awwad, Y., & Kane, G. (2021). Failures of fairness in automation require a deeper understanding of human-ML augmentation. *MIS Quarterly*, *45*(3), 1483-1499. https://doi.org/10.25300/MISQ/2021/16512

Tyler, T. R. (2012). Justice and effective cooperation. Social Justice Research, *25*(4), 355-375. https://doi.org/10.1007/s11211-012-0168-5

United Nations. (2006). *Social justice in an open world: The role of the United Nations.* https://www.un.org/esa/socdev/documents/ifsd/SocialJustice.pdf

United Nations General Assembly. (2019). *Report of the Special Rapporteur on extreme poverty and human rights*, A/74/493,

Agenda item 70(b). https://documents-dds-ny.un.org/doc/UNDOC/GEN/N19/312/13/PDF/N1931213.pdf

Urquhart, C. (2022). *Grounded theory for qualitative research: A practical guide*. SAGE.

van de Ven, A. H. (2007). *Engaged scholarship: A guide for organizational and social research*. Oxford University Press. https://doi.org/10.5465/amr.2008.34424997

van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the machine meets the expert: An ethnography of developing AI for hiring. *MIS Quarterly*, *45*(3), 1557-1580. https://doi.org/10.25300/MISQ/2021/16553

Vassilakopoulou, P., Parmiggiani, E., Shollo, A., & Grisot, M. (2022). Responsible AI: Concepts, critical perspectives and an information systems research agenda. *Scandinavian Journal of Information Systems*, *34*(2), Article 3. https://doi.org/10.17705/1SJIS.34.2.3

Vaughan, D. (1997). *The Challenger launch decision: Risky technology, culture and deviance at NASA*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226346960.001.0001

Vervloesem, K. (2020). How Dutch activists got an invasive fraud detection algorithm banned. *Algorithm Watch.* https://automatingsociety.algorithmwatch.org/report2020/netherlands/netherlands-story/

Weick, K. E. (1993). The collapse of sensemaking in organizations: The Mann Gulch disaster. *Administrative Science Quarterly*, *38*(4), 628-652. https://doi.org/10.2307/2393339

Wiener, M., Cram, W., & Benlian, A. (2021). Algorithmic control and gig workers: A legitimacy perspective of Uber drivers. *European Journal of Information Systems, 32*, 1-23. https://doi.org/10.1080/0960085X.2021.2006479

Williams, G. (1961). *Criminal law: The general part* (2nd ed.). Stevens & Sons.

Zerubavel, E. (2006). *The elephant in the room: Silence and denial in everyday life*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195187175.001.0001

Zheng, Y., & Wu, P. F. (2022). Producing speed on demand: Reconfiguration of space and time in food delivery platform work. *Information Systems Journal*, *32*(5), 973-1004. https://doi.org/10.1111/isj.12356

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*(1), 75-89. https://doi.org/10.1057/jit.2015.5

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books.

Zuiderent-Jerak, T. (2015). *Situated intervention: Sociological experiments in health care*. MIT Press. https://doi.org/10.7551/mitpress/9780262029384.001.0001

Završnik, A. (2020). Criminal justice, artificial intelligence systems, and human rights. *ERA Forum*, *20*, 567-583. https://doi.org/10.1007/s12027-019-00578-5

## About the Authors

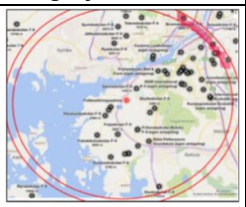
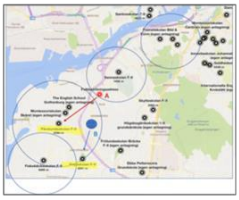
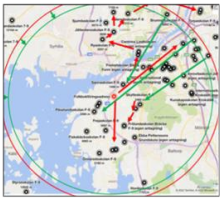
**Charlotta Kronblad** practiced law for a decade before turning to academia to complete a Ph.D. focused on the digital transformation of the legal field. Previous workplaces include law firms, courts, and legal departments within industry. She is currently working at the Department of Applied Information Technology at the University of Gothenburg and is also affiliated with the House of Innovation, Stockholm School of Economics and the Swedish Center for Digital Innovation. Her work revolves around digital technology in relation to society, law, and organizations. She has published in journals such as *Academy of Management Discoveries*, *Journal of Profession and Organization*, *Journal of Science and Technology Policy*, and *Creativity and Innovation Management*. She has also published book chapters and is a common voice in the public debate on how to digitalize society, stressing the need to ensure that legal principles and the rule of law are adhered to in the process. (ORCiD: 0000-0003-3933-4799.)

**Anna Essén** is an associate professor at the House of Innovation, Stockholm School of Economics. Her research concerns the intertwinement of digital technology and arrangements and ambitions at individual, organizational, and institutional levels of society. Part of this work focuses on the willful ignorance and silencing that characterizes digital innovation. She has published papers in outlets such as *Human Relations, Organization Studies, Sloan Management Review, Social Science & Medicine, Information & Organization,* and *MIS Quarterly*. (ORCiD: 0000-0003-3566-8409.)

**Magnus Mähring** holds the Erling Persson Chair in Entrepreneurship and Digital Innovation at the Stockholm School of Economics. He is head of the Department of Entrepreneurship, Innovation and Technology and director of the SSE House of Innovation. He is also a Fellow at Hughes Hall and Cambridge Judge Business School, Cambridge University, and co-director of the Swedish Center for Digital Innovation. His current research focuses on digital innovation and transformation, innovation ecosystems, and artificial intelligence in the workplace. He has published in journals such as *MIS Quarterly, Information Systems Research, Journal of the AIS, Organization Studies, Human Relations,* and *California Management Review*. He serves on the jury for the Swedish Tech/CIO Awards and has been appointed to the Swedish Government's Digitalization Council. (ORCiD: 0000-0002-9557-7974.)

# Appendix

**Table A1. Systematic Analysis and Visualizations of Error Categories**

| | Visualization of error category | Description of error | Responses by public institutions |
|---|---|---|---|
| 1 |  | When the PSA noted that the algorithm was not able to allocate school slots for every child within the 8 km limit, it decided to use the maximum distance of 8 km as an approximate, instead of an absolute limit, without having the authority to do so. This opened the potential for placements in schools farther away than 8 km (the outer circle). | These errors were recognized by the PSA as "formal" errors. The PSA corrected 450 of these errors (without a preceding appeal process) by assigning these 450 children new school slots within the inner circle (see Picture 6). |
| 2 |  | The PSA provided the system with faulty data (wrong addresses) for some schools. The result was that some children did not get their closest school, while other children who lived in a nearby area (but not as close) received those slots instead. | This was considered a formal error by the PSA (and the court). Thus, it was seen and treated as an error. Placements were corrected, but only if slots were available at the schools in question. |
| 3 |  | The PSA instructed the ADM system to place the children in the closest school according to a bird's-eye view, despite legal rules and praxis dictating that walking distance should determine placements. This resulted in faraway placements with long commuting times on opposite sides of rivers and fjords. | This error was acknowledged by the PSA but not as a formal error, and erroneous placements were not corrected. The court was not able to recognize these as errors if the placements were within the 8 km limit, as defined by the PSA. The PSA and the court were thus both effectually blind to these errors. |
| 4 |  | School law states that parent preferences/choices should guide school placements as long as this does not deprive another child of placement in a nearby school. This was not coded into the system. If the system had been configured to optimize for parental choices, almost all children would have received one of their five choices. | This error was not recognized by either the PSA or the court. Both denied the systematic disregard of parent preferences. Errors were not seen (if decisions still fell within the 8 km limit) and therefore not corrected. |
| 5 |  | Initial errors (such as decisions based on bird's-eye view distances or erroneous addresses) had domino effects on the placements of other children. The result was children being placed at nearby schools—but not the schools they would have been placed at if procedures had been followed. This multiplied errors in several steps, affecting families across the entire city. | Both the PSA and the court were blind to these systematic domino effects, and the erroneous placements were not corrected. |
| 6 |  | The 450 most obvious and severe formal errors were corrected by the PSA, with the affected children assigned new school slots closer to home. However, this was done without considering that other children might have been entitled to these alternative school slots. | The correction of the most severe placements errors caused additional errors that neither the PSA nor the court acknowledged or corrected. Instead, both the PSA and the court promoted individual redress, which simultaneously furthered such errors. |

**Note:** These visualizations are based on an analysis of categories of errors accounted for in the City Auditors Report (2021) and using official maps provided by the City of Gothenburg. The maps show the geographical distances to all potential schools for children, using the address of the first author for illustrative purposes.