

AI Literacy and Procedural Justice in Algorithmic Systems: A Conditional Account on Data Evaluation Practices and Human Oversight

Carmen Loefflad, Jens Grossklags

Professorship of Cyber Trust & MDSI
Technical University of Munich
carmen.loefflad@tum.de, jens.grossklags@in.tum.de

Abstract

Procedural justice is a central lens to understand how individuals evaluate algorithmic procedures. It involves different dimensions, reflecting both relational concerns about how people perceive they are treated, and instrumental concerns about whether individuals perceive control within algorithmic processes. Focusing on data evaluation practices (rule-based vs. AI-based) and human oversight (human-in-the-loop vs. human-out-of-the-loop) as two distinct algorithmic procedures, we advance a conditional account of procedural justice. We posit that the relationship between algorithmic procedures and perceived procedural justice depends on individuals' AI literacy and on the dimension of procedural justice considered. We test these arguments using data from a large-scale vignette experiment in a behavioral regulation context ($N = 3,203$), and find that distinct aspects of AI literacy operate asymmetrically across algorithmic procedures and dimensions of procedural justice. AI awareness conditions how data evaluation practices translate into procedural justice perceptions, but plays a substantially weaker role for evaluations of human oversight. Human-out-of-the-loop oversight increases perceived control, but simultaneously reduces relational perceptions. These findings demonstrate that aggregate evaluations of procedural justice can conceal important heterogeneity, as algorithmic procedures may produce opposing effects across perceived procedural justice dimensions or across individuals with different levels of AI literacy. Our study contributes to the debate on the ethics of algorithmic systems by showing that perceived procedural justice depends on how individuals understand algorithmic procedures, and that it is shaped by the multidimensional nature of procedural justice.

Introduction

Algorithmic systems increasingly determine access to opportunities and resources. They process data into knowledge about individuals' preferences or likely behaviors, which are then used to shape how individuals are evaluated or treated (Just and Latzer 2017). The procedures underlying these systems vary substantially in how data is collected, processed, and translated into decisions. Such variations include distinct algorithmic models used to evaluate data (Durán and Jongsma 2021), as well as the form of human involvement in decision-making processes (Nakao et al. 2022).

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Understanding how individuals evaluate distinct algorithmic procedures has become a central concern, and the concept of *procedural justice* has emerged as an important lens to understand how people judge the fairness of algorithmic procedures (Lee et al. 2019). Yet, despite growing attention to procedural justice in algorithmic contexts, empirical findings about how algorithmic procedures translate into perceptions of procedural justice remain inconsistent. For example, in some studies, human involvement in algorithmic procedures increases perceived procedural justice (Nagtegaal 2021; Newman, Fast, and Harmon 2020), while others find it does not (Yurrita et al. 2023). Similarly, some studies suggest that simpler rule-based systems are perceived as more procedurally just than more complex AI-based models (Wang et al. 2023), whereas others find little or no difference (Wang, Harper, and Zhu 2020).

We argue that these inconsistencies arise because prior research might have overlooked two fundamental aspects. First, procedural justice is not a uniform construct. From the social psychology perspective, procedural justice includes both *relational* concerns about respectful treatment (Tyler, Rasinski, and Spodick 1985; Tyler and Lind 1992), and *instrumental* concerns related to perceived control within decision-making processes (Thibaut and Walker 1975). Different algorithmic procedures may activate these components in different ways. For example, aggregate effects may appear weak or inconsistent when procedures simultaneously affect relational and instrumental evaluations in opposing directions.

Second, evaluating the procedural implications of algorithmic systems might depend on individuals' ability to understand and interpret algorithmic decision-making processes. We study these interpretive capacities through the concept of *AI literacy* (Long and Magerko 2020; Wang, Rau, and Yuan 2023). Individuals affected by algorithmic decisions may interpret identical procedural features differently depending on their level of AI literacy. For example, individuals with higher AI literacy may be better able to recognize the implications of AI-based rather than rule-based decision-making, or of different levels of human involvement. In contrast, individuals with lower AI literacy may not differentiate between distinct procedural configurations. Consequently, treatment effects averaged across different levels of AI literacy might obscure substantial heterogeneity in how proce-

dural features shape perceptions of procedural justice.

Conceptual Setup and Research Contributions

This study develops and empirically tests a conditional account of procedural justice in algorithmic systems (Figure 1). As a first contribution, we move beyond treating procedural justice as a uniform perception, but conceptualize it as a multidimensional construct, which contains both *relational* and *instrumental* components. While prior work has conceptually translated distinct procedural justice components to the language of algorithmic systems (Lee et al. 2019), empirical research has largely treated procedural justice as one uniform dimension within larger justice frameworks (Juijn et al. 2023; Yurrita et al. 2023). This neglects that the relational and instrumental components of procedural justice might be differentially affected by algorithmic procedures. Our framework, therefore, provides a more fine-grained account of procedural justice by recognizing that the relationship between algorithmic procedures and procedural justice depends on whether relational or instrumental aspects of justice are evaluated.

Second, we argue that understanding the implications of different algorithmic procedures requires specific AI literacy competencies, which might, subsequently, shape how procedures are evaluated by individuals who vary in these competencies. Against the backdrop of increased research attention to AI literacy (Celik 2023; Wang et al. 2025), recent studies have established a multi-dimensional AI literacy scale that conceptualizes the construct as literacy in AI usage, AI awareness, AI evaluation and AI ethics (Wang, Rau, and Yuan 2023). By exploring the moderating role of these four AI literacy components, we offer a potential explanation for why identical algorithmic procedures may elicit different procedural justice perceptions across individuals.

Third, we focus on two algorithmic procedures that have received limited attention in prior research, namely data evaluation practices and human oversight. Data evaluation practices refer to the ways in which input data are transformed into knowledge and decision outputs (Wang, Harper, and Zhu 2020; Wang et al. 2023). Human oversight captures distinct ways through which algorithmic procedures remain subject to human involvement and intervention. By examining these two fundamental procedural configurations, we extend prior work that has extensively focused on procedural features such as transparency and explainability (Hase and Bansal 2020; Loefflad, Chen, and Grossklags 2024b), and demonstrate that their implications for perceived procedural justice are conditional on both the specific justice dimensions, as well as on the AI literacy of those affected by algorithmic decisions. Focusing on data evaluation practices and human oversight, we ask:

1. How do algorithmic procedures relate to instrumental and relational dimensions of procedural justice?
2. To what extent are these relationships shaped by distinct aspects of AI literacy?

We examine the research questions using a large-scale dataset ($N = 3,203$) from a vignette-based survey study previously conducted by the authors (Loefflad and Grossklags

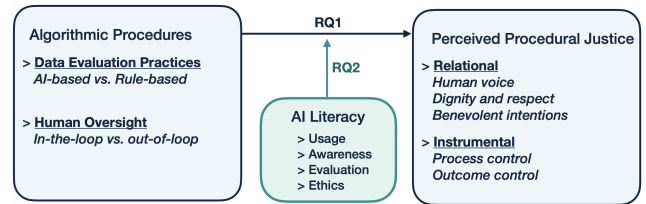


Figure 1: Conditional account on the effects of algorithmic procedures on perceived procedural justice.

2026). That study investigated people’s perceptions of algorithmic systems that adopt a behavioral steering function, aimed at incentivizing people to behave in certain ways, for example in more climate-friendly or healthy manners. Participants evaluated the procedural justice of hypothetical scenarios, which presented them with algorithmic systems that collected behavioral data, evaluated it against a behavioral target, and issued targeted interventions, for example in the form of algorithmic nudges or behavioral scores (Loefflad, Chen, and Grossklags 2024b; Chen, Bogner, and Grossklags 2025). These systems varied in their procedural configurations, namely their data evaluation practice (rule-based or AI-based) and the level of human oversight (in-the-loop or out-of-the-loop). While the original study examined how intervention designs (e.g., nudges vs. scores) and procedural configurations shape various perceptions, the present paper conducts a secondary analysis that focuses exclusively on the conditional relationship between algorithmic procedures and procedural justice, examining how the effects of procedural configurations vary across different dimensions of procedural justice, and how they are shaped by different aspects of AI literacy.

To this end, we first present background information on procedural justice in algorithmic systems, and on AI literacy as an important interpretive ability in the era of AI. Subsequently, we introduce the two procedural configurations studied in this paper and, for each, outline the possibility that their relationship with the relational and instrumental dimensions of procedural justice is conditional on individuals’ AI literacy. Subsequently, we present our methodological approach and results, and discuss our findings in detail.

Procedural Justice in Algorithmic Systems

Procedural justice research examines how decision-making processes by governance systems, institutions, or authorities shape individuals’ evaluations of these governance bodies (Tyler 2006). Procedural justice is a multifaceted construct that comprises both *instrumental* components – those that relate to strategic considerations about how procedures relate to decision outcomes (Thibaut and Walker 1975) – and *relational* components – those that refer to feelings about procedures (Tyler and Lind 1992).

Accounting for the distinct relational and instrumental dimensions of procedural justice in algorithmic systems is important, as algorithmic processes may differentially affect these distinct components. Uncovering the relationship between procedures and components of procedural justice

would allow a more fine-grained understanding of the motives underlying individuals' judgment of procedures, and would yield important insights into the ethical design and implementation of algorithmic procedures.

Below, we expound upon the distinct instrumental and relational aspects of procedural justice in algorithmic systems.

Instrumental Judgments These arise when individuals evaluate procedures based on the extent to which they enable influence over decision-making processes and outcomes. This is captured in the notion of control. *Outcome control* describes a situation in which people feel they can influence the outcome. *Process control* refers to the feeling of being able to influence processes. Often, outcome control is considered a consequence of process control, as influence over procedures also implies influence over outcomes (Thibaut and Walker 1975). In algorithmic systems, both forms of control might be contingent on individuals' ability to interpret system processes, which highlights that control over processes and outcomes also requires a certain understanding of these processes (Mittelstadt et al. 2016).

Relational Judgments They comprise whether people perceive that decision-makers act with *benevolent intentions*, and whether people feel treated with *dignity and respect* when being subject to decision-making. Relational components also capture whether people feel they have a *voice* in decision procedures, defined as the ability to express one's views (Lind, Kanfer, and Earley 1990).

We adopt the feeling of being treated with respect, and benevolent intentions as central relational aspects in algorithmic procedures. In contrast, voice in algorithmic systems is often not exercised directly by affected individuals, but mediated through human intermediaries who translate individuals' concerns into the design and operation of algorithms – that is, through human experts that oversee algorithmic procedures. Therefore, we conceptualize *human voice* as the extent to which individuals perceive that humans can represent their interests within algorithmic procedures. Human voice thus promotes a relational feeling of being heard. At the same time, it addresses the concept of contestability, understood as the ability to contest or influence decisions, which reflects an instrumentally rooted motive (Leventhal 1980). Therefore, *human voice* in algorithmic systems comprises both instrumental and non-instrumental considerations (Lind, Kanfer, and Earley 1990).

These five distinct components of procedural justice provide the analytical basis for examining how specific procedures in algorithmic systems translate into distinct perceptions of procedural justice. An overview is given in Table 1.

AI Literacy

AI literacy refers to the competencies necessary to evaluate, use, and collaborate with AI technologies (Long and Magerko 2020). Prior work suggests that individuals' reactions to algorithmic decision-making processes depend partly on their familiarity with and understanding of AI systems (Gonzalez et al. 2022; Schoeffer, Machowski, and Kuehl 2022; Ullstein et al. 2022), suggesting that AI liter-

acy might play a key role in shaping how individuals interpret and evaluate differences in algorithmic procedures.

Yet, the nature of these effects is underexplored from several perspectives. First, recent work has presented AI literacy as a multidimensional construct that captures different competencies related to AI (Wang, Rau, and Yuan 2023). However, empirical evidence on which aspects of AI literacy shape judgments of algorithmic procedures is scarce. Second, related work has primarily studied the role of AI literacy in more general evaluations of algorithmic systems (Schoeffer, Machowski, and Kuehl 2022; Gonzalez et al. 2022), such as fairness and trustworthiness, rather than how it influences perceptions of procedural justice specifically. Third, prior studies that examined the moderating role of AI literacy have largely contrasted human and AI-based decision-making, which provides only limited insights into how AI literacy shapes evaluations of different procedural configurations *within* algorithmic systems themselves, for example, how data is processed into decision outputs, or the level of human involvement within these processes.

We address this problem space by examining the relationship between algorithmic procedures, AI literacy, and perceived procedural justice from a more comprehensive perspective. Beyond treating procedural justice as a multidimensional construct, we also account for distinct aspects of AI literacy, relying on a recently developed scale, which distinguishes AI awareness, usage, evaluation, and ethics (Wang, Rau, and Yuan 2023). With this approach, we evaluate how different aspects of AI literacy might shape procedural justice perceptions indirectly, by influencing how individuals interpret the implications of algorithmic procedures. We next present the aspects of AI literacy as developed by (Wang, Rau, and Yuan 2023), and explain their relation to the components of procedural justice.

AI Awareness. Literacy in AI awareness describes the cognitive ability to identify, comprehend, and understand AI technology. Awareness is particularly relevant for evaluations of procedures, because understanding how algorithmic systems operate constitutes a prerequisite for assessing how decisions are produced (Celik 2023). In this sense, AI awareness may be especially important for perceptions of control, as individuals can only perceive influence over procedures and outcomes when they understand algorithmic processes (Gonzalez et al. 2022).

AI Usage. Literacy in AI usage describes the ability to use AI technology to complete or accomplish tasks in a proficient manner. It captures the ability to integrate AI technologies into work tasks and everyday activities (Wang, Rau, and Yuan 2023). This familiarity with using AI systems through repeated usage experiences may increase individuals' perceived efficacy in navigating algorithmic procedures. Therefore, AI usage might be relevant for procedural justice dimensions related to process and outcome control.

AI Evaluation. Literacy in AI evaluation describes the ability to critically analyze and evaluate AI technologies, and to formulate informed judgments about them (Wang, Rau, and Yuan 2023). Such evaluative capabilities may be

Dimension	Conceptual Definition	Adaptation to Algorithmic Systems	Survey Item
Process Control	Perceived ability to influence decision-making procedures	Ability to understand and influence how data inputs are processed within algorithmic systems	I can influence the process through which the algorithmic systems makes decisions.
Outcome Control	Perceived ability to influence decision outcomes	Perceived capacity to affect outcomes through interaction with or understanding of system processes	I can influence the decision outputs of the algorithmic system.
Dignity & Respect	Perception of being treated with respect and dignity by decision-makers	Perception that algorithmic systems treat individuals with dignity and respect	In the way the algorithmic system is implemented people are treated with dignity and respect.
Benevolent Intent	Belief that decision-makers act with good intentions	Perception that algorithmic systems act in the best interest of affected individuals	The algorithmic system operates in the best interest of citizens.
Human Voice	Opportunity to express views in decision-making processes	Perceived ability of human experts to represent or incorporate individual concerns in algorithmic procedures	Introducing a human component through which algorithmic decisions are reviewed, feels reassuring.

Table 1: Procedural justice dimensions in algorithmic systems, distinguishing between instrumental dimensions (**in green**), relational dimensions (**in red**), and human voice, containing both instrumental and relational dimensions.

particularly relevant for assessing whether algorithmic procedures are perceived as reasonable, and appropriate. AI evaluation may thus be especially important for the relational dimensions of procedural justice.

AI Ethics. This aspect refers to the awareness of the responsibilities and risks arising from the use of AI. Awareness of ethical issues is a requirement to appropriately engage with AI. It may be particularly relevant for relational evaluations of procedural justice, as it shapes sensitivity toward whether algorithmic procedures are perceived as appropriate and aligned with normative expectations about fair treatment. AI ethics literacy may further influence how individuals evaluate the dignity, respect, and benevolent intent communicated through algorithmic procedures.

Having discussed how individuals differ in their capacity to interpret and evaluate AI systems, we can now examine the specific procedural configurations whose effects on procedural justice likely depend on these capabilities.

Procedures in Algorithmic Systems

Algorithmic procedures represent a *throughput* dimension, in which data inputs are transformed into decision outputs (Starke and Lünich 2020; Custers and Vrabec 2024). We focus on two key procedural configurations, namely (i) how data is processed into knowledge (*data evaluation practices*), and (ii) the extent and form of human involvement in decision-making (*human oversight*). Despite their centrality in both academic and regulatory debates, prior research has paid limited attention to these procedural aspects. Moreover, these configurations have distinct implications for the interpretability and contestability of algorithmic decisions. As such, they provide important, but distinct cases to examine how procedural characteristics interact with individuals’ AI literacy in shaping perceptions of procedural justice.

Data Evaluation Practices

Data evaluation practices refer to the processes through which algorithmic systems transform input data into actionable knowledge, which informs decisions about individuals (Kitchin 2017). Due to our study context of behavioral steering, we use the term *data evaluation* instead of processing to highlight that these practices not only transform behavioral data, but also lead to a judgment regarding how well people align their behaviors with a behavioral target.

In this section, we contrast AI-based and rule-based data evaluation practices. To clarify the differences between these two practices, we first introduce the types of input data available to algorithmic systems, as well as the forms of knowledge and decision outputs generated from them. Subsequently, to highlight their potential relationship with AI literacy, we examine how different data evaluation practices shape the transparency, interpretability, and contestability of the processes through which input data are converted into evaluative knowledge and lead to decisions.

Input Data Data input is the raw data fed into algorithmic systems, and can be made available to algorithmic systems in different ways (Patil, Wenzel, and Grossklags 2026). Individuals can actively provide data to algorithmic systems, for example, by using social media (Linders 2012), or through allowing algorithmic systems to access self-tracked data (van Zeben and Kamphorst 2020). On the contrary, data input can also be passively gathered without the data subject’s active involvement, for example, by technologies such as CCTV systems (Marx 1998; Fontes et al. 2022).

Knowledge Creation and Decision Outputs Input data is processed to generate actionable knowledge about individuals, often in the form of inferences or profiles. This knowledge subsequently informs algorithmic decisions (Kitchin 2017). Inference refers to generating new information and

insights about individuals' specific attributes. These inferences can range from simple derivations to more sophisticated approximations of sensitive characteristics, for example, behavioral tendencies or character traits (Custers and Vrabec 2024; Engelmann et al. 2022). Profiling goes one step further, and attaches evaluations, assessments, or judgments to produced knowledge. In this sense, inferences are the result of the process through which new attributes are derived from input data, and profiling is the result of evaluating inferred information, which guides assessments, judgments or interventions directed at individuals (Hildebrandt 2008).

The practices through which data are processed determine the extent to which outputs are traceable, understandable, and interpretable, thereby defining the degree to which the relationship between inputs and outputs can be reconstructed. We distinguish rule-based and AI-based data evaluation practices as two analytical boundary scenarios for assessing how different processing logics shape perceptions of procedural justice.

AI-based vs. Rule-based Methods Both rule-based and AI-based systems can generate *knowledge* from data, in the form of inferences or profiles. Yet, they fundamentally differ in their logic through which they generate knowledge. Rule-based algorithms act deterministically, relying on traceable and pre-specified rules, and are based on clearly defined assumptions. AI-based algorithms, in contrast, deploy sophisticated AI or machine learning techniques, and rest on a probabilistic, discovery-driven approach through which they create knowledge without explicit guidance (Custers 2004; Hildebrandt 2008; Durán and Jongsma 2021).

These differences have important implications for the transparency and interpretability of algorithmic processes. Through their defining characteristics, rule-based algorithms offer greater understandability of how data relates to produced knowledge. AI-based methods, in contrast, are often characterized as black-box systems, which blur the relationship between input and output, and make it difficult to understand how data is converted into knowledge (Wang et al. 2023; Durán and Jongsma 2021; Pasquale 2015; Innerarity 2021). AI-based methods can generate sophisticated *evaluations* from diverse input data, enabling the approximation of latent or unobservable attributes and traits, and the creation of comprehensive profiles. In such systems, it is not only opaque how input data is turned into knowledge, but often also what kind of inferences or profiles are constructed (Custers and Vrabec 2024).

As such, perceptions of procedural justice might be affected by distinct data evaluation practices. As rule-based systems allow people to better follow data evaluation logics, they might increase perceptions of control, or enhance the feeling of being treated with respect (Loefflad, Chen, and Grossklags 2023; Mittelstadt et al. 2016). We, therefore, explore to which relational and instrumental components of procedural justice data evaluation practices relate.

Importantly, these distinctive characteristics have implications for the interpretive demands they place on individuals to follow decision logics. Rule-based systems enable decision subjects to understand the relationship between

inputs, processing steps, and outputs (Arrieta et al. 2020; Colaner 2022). AI-based systems, in contrast, provide only indirect forms of traceability, for example, through model-level explanations rather than transparent rules (Hase and Bansal 2020). Hence, understanding how they work requires greater ability to understand algorithmic processing logic. Interpreting such systems may, therefore, depend on distinct aspects of AI literacy (Celik 2023). Therefore, we explore whether distinct aspects of AI literacy moderate the association between data evaluation practices and distinct components of procedural justice.

Human Oversight

Human oversight refers to the involvement of natural persons in the supervision or correction of AI system outputs (Koulu 2020). As it requires that humans can identify errors and intervene in the system's operation, it is often operationalized by letting dedicated experts oversee or review algorithmic decision-making processes. A crucial aim of human oversight is to ensure that algorithmic decision-making processes are judged or reviewed by a human expert who is capable of understanding and assessing these processes and their associated decision outputs (Sterz et al. 2024). With this, human oversight is supposed to mitigate or prevent algorithmic harm (Green 2022), and, therefore, purportedly legitimizes algorithmic decisions (Martin and Waldman 2022). Yet, while distinct forms of oversight exist, it is largely unclear how those subject to algorithmic decisions perceive these distinct forms, particularly in-the-loop and out-of-the-loop human oversight.

In-the-loop oversight implies that an automated decision is reviewed by a human expert before it translates into legal consequences. This implies that outputs are subject to prior validation. In procedural terms, this form of oversight provides an anticipatory form of human intervention, as it allows human experts to validate outcomes before they materialize into actual consequences for decision subjects.

Under *out-of-the-loop* oversight, in contrast, human intervention is ensured only after the output of an AI system has become effective. As such, decisions can be reviewed, contested, and revised after an algorithmic decision has been made. Crucially, this retrospective expert review must be actively initiated by those affected by a decision, providing a reactive form of oversight (Lazcoz and de Hert 2023).

Human Oversight and Procedural Justice Expert-mediated human oversight might provide an institutionalized form of human voice by ensuring that individual circumstances can be considered, and that decisions can be reviewed or corrected. While prior research suggests positive effects of such oversight on perceptions (Nagtegaal 2021), less attention has been directed to the possibility that distinct forms of human oversight could differentially affect distinct dimensions of procedural justice. In-the-loop oversight may enhance perceptions of process control and reassurance, as human judgment through dedicated experts is continuously present during decision processes. Out-of-the-loop oversight, in contrast, may strengthen perceived outcome control, as it provides individuals with the possibility

to have algorithmic decision outputs challenged or reversed. At the same time, both forms may contribute to relational evaluations, for example, when expert-level oversight of algorithmic outputs signals respectful treatment. Yet, the nature of these effect remains empirically underexplored.

Human Oversight and Aspects of AI Literacy Importantly, the relationship between human oversight and perceptions of procedural justice may depend on individuals' ability to engage with distinct forms of oversight, and to evaluate their distinct implications. First, while human oversight has been portrayed as an important procedural safeguard, it has also been criticized for providing a false sense of reassurance if it is implemented merely symbolically (Kyriakou and Otterbacher 2023). Individuals with higher AI literacy may be better able to recognize these implications and distinguish between meaningful and merely symbolic forms of oversight, whereas those with lower AI literacy may be less likely to do so. Second, in out-of-the-loop oversight, taking the initiative to have technologically produced artifacts challenged might be more plausible when individuals perceive a certain level of self-efficacy in navigating algorithmic technologies (Igbaria and Iivari 1995), which is closely linked to AI literacy (Türker and Pala 2020). When AI-literate people feel more equipped and confident to contest outcomes through out-of-the-loop oversight, they might more positively react to this form of oversight. Conversely, individuals with lower AI literacy may perceive retrospective contestation mechanisms as less accessible, which could weaken perceptions of procedural justice. Therefore, we explore whether the effect of different forms of human oversight on procedural justice may depend on the AI literacy of those subject to algorithmic decisions.

Summarizing, we posit that the way procedural configurations in algorithmic systems translate to perceptions of procedural justice might be affected by the AI literacy of those affected by algorithmic decision, which plays different roles across procedural configurations. For data evaluation practices, AI literacy primarily enables individuals to interpret how inputs are transformed into outputs, and, therefore, shape perceived control. For human oversight mechanisms, AI literacy is more closely related to individuals' ability to recognize and activate contestation opportunities, particularly in settings where intervention depends on the initiation of individuals themselves.

Method: Vignette-based Survey Study

Data Collection

We conducted a vignette-based survey study to examine how two procedural configurations (data evaluation practices and human oversight) in algorithmic systems shape perceived procedural justice. The study presented an algorithmic system in a *behavioral regulation context*, where algorithms collected behavioral data, and processed it to yield evaluations about individuals' behavioral compliance with a pre-defined target (Loefflad 2024). These evaluations were the basis for interventions to influence behaviors, for example, in the form of algorithmic nudges or behavioral scores.

These interventions aimed at improving a specific collective action challenge, for example, increasing population health.

Participants were recruited via Prolific Academic. The study employed a between-subjects design in which each participant was randomly assigned to one vignette describing a single algorithmic system. We collected data from participants residing in the United States. No additional socio-demographic restrictions were imposed. After applying attention and comprehension checks, the final sample consisted of 3,203 participants.

The mean survey completion time was 9.27 minutes ($SD = 2.97$). Participants received, on average, £1.85, which is equivalent to an average hourly wage of £12.

Our institution does not require formal ethics approval for survey-based studies. Nevertheless, we followed standard ethical procedures, including informed consent and anonymized data collection.

Vignette Design

Participants were first presented with a short scenario which described a collective action problem in one of three domains: climate, health, or social cohesion. In each scenario, an algorithmic system was introduced to collect behavioral data, to evaluate it against a behavioral target, and to decide on an intervention to influence behaviors. For example, in the climate context scenario, interventions were described that steered people to adopt more pro-environmental behaviors (see Online Supplement (OS); Tables OS-1 and OS-2 (Loefflad and Grossklags 2024a)).

Following the scenario description, participants were shown a factorial vignette, which described the procedural configurations of the algorithmic system. Each participant was presented with one specific data evaluation practice (rule-based or AI-based), and one form of human oversight (in-the-loop vs. out-of-the-loop) (see Online Supplement; Tables OS-3 and OS-4 (Loefflad and Grossklags 2024a)). Besides distinct procedural variables, the study contained two *design variables*, which manipulated how and when interventions were communicated and implemented. One design variable related to timing, namely in whether interventions were channeled in the form of a nudge, before behaviors take place (ex ante), or in the form of a score-mediated reward, after behaviors occur (ex post). Another design variable (framing) captured whether interventions were positively or negatively framed (Loefflad and Grossklags 2026).

While the implications of distinct intervention designs are examined in detail in a related working paper (Loefflad and Grossklags 2026), we here focus on the procedural aspects (data evaluation practices and human oversight) through which data is turned into knowledge and decisions. However, we maintain the distinct contextual and design variables in our dataset, to situate the procedural manipulations within a broad range of distinct design and contextual settings. This operationalization increases the external validity of our study, and allows us to assess whether the effects of algorithmic procedures are robust across different contexts and intervention designs.

In summary, we employed a between-subjects vignette design in which each participant was presented with a single

scenario, characterized by one collective action context, one intervention design (including timing and framing), and one combination of procedural features (human oversight and data evaluation practice). Participants evaluated the scenario on the five dimensions of procedural justice.

To ensure treatment validity, the survey included comprehension checks on all manipulated procedural features, and provided participants with corrective feedback in case of wrong answers. An additional attention check was included, and only respondents who passed all checks were retained for the subsequent analysis.

Measures

Prior to the vignette, we retrieved participants' literacy measured by AI awareness, AI usage competence, AI evaluation competence, and AI ethics-related understanding, based on a recently established and widely adopted scale (Wang, Rau, and Yuan 2023). After the vignette, we retrieved several perceptions. We here focus exclusively on perceived procedural justice, which was measured over three relational dimensions (human voice, benevolent intentions, and respectful treatment), and two instrumental measures related to process and outcome control (5-point Likert scale). All items were adapted from prior research on procedural justice and algorithmic decision-making contexts (Thibaut and Walker 1975; Tyler, Rasinski, and Spodick 1985; Loefflad, Chen, and Grossklags 2023). Example items that capture these different dimensions of procedural justice are shown in Table 1. The wording of the items was adapted to the specific behavioral steering context of our study (see Online Supplement (OS), Table OS-5 (Loefflad and Grossklags 2024a)).

Data Analysis

We first estimated several ordinary least squares (OLS) regression models with perceived procedural justice as a composite variable that averages the five subdimensions (Table OS-2). We subsequently expanded the model. First, we regressed procedural justice on human oversight and data evaluation practices (Model 1). In Model 2, we explored their interaction to assess whether these procedural dimensions jointly shape perceptions of procedural justice. In Model 3, we added the four dimensions of AI literacy as individual-level covariates. In Model 4, we further included interaction terms between the components of AI literacy and human oversight, as well as between the components of AI literacy and data evaluation practices. As our data includes treatment differences for policy design and context, we assessed the robustness of these relationships. In Model 5, we included controls for policy design features that arise from the factorial vignette manipulation, and in Model 6, we further controlled for differences across distinct contexts.

To identify whether procedures differentially affect distinct dimensions of procedural justice, we repeated the previously described modeling approach for each subdimension of procedural justice, namely, dignity and respect, benevolence, human voice, process control, and outcome control.

Results

Our sample ($N=3,203$) contained 1,532 observations for in-the-loop, and 1,671 for out-of-the-loop oversight. We had 1,536 observations in the AI-based data evaluation condition, and 1,667 observations in the rule-based condition.

Participants were diverse in terms of age, gender, education, and ethnicity. Regarding age, 5.44% of the participants were below 25 years, 12.68% between 25 and 30, 27.27% between 31 and 40, 23.18% between 41 and 50, 17.43% between 51 and 60, and 13.90% above 60 years old. 53.14% identified as female, 45.22% as male, 1.23% as diverse, and 0.42% did not disclose their gender. Regarding education, 0.72% had not completed high school, 25.12% had a high school degree, 51.41% held a college degree, 17.94% held a master's degree, and 4.01% held a doctoral degree. 25.70% identified as European, 53.90% as American, 6.35% as Asian-Pacific, 6.07% as Hispanic, 5.86% as multicultural, and 2.21% did not reveal their ethnicity.

We next report the results of the regression analyses, a detailed overview of which is provided in the Online Supplement (Tables OS-7–OS-12).

AI Literacy

Regarding AI literacy, the ability to use AI technologies is positively and consistently associated with perceived procedural justice ($\beta = 0.15$, $p < 0.01$ in Model 6, Table OS-7). This finding is consistent across distinct model specifications. Other aspects of AI literacy relating to awareness, ethics, and evaluation are not associated with perceived procedural justice.

Human Oversight and AI Literacy

Aggregate Procedural Justice We first evaluate the models that considered procedural justice as a combined construct (Table OS-7). In the partially specified models (Models 1–4), out-of-the-loop (OL) oversight is associated with a significant decrease in perceived procedural justice. However, this effect becomes attenuated and statistically insignificant once the intervention design (Model 5) or contextual variations (Model 6) are controlled for. We did not find evidence that distinct aspects of AI literacy moderate the association between human OL oversight and perceived procedural justice.

Components of Procedural Justice The effects of human oversight differ substantially across model specifications and procedural justice components. In partially specified models (see Model 3), OL oversight is negatively associated with relational aspects, including respectful treatment ($\beta = -0.16$, $p < 0.01$; Table OS-8), perceived benevolence ($\beta = -0.19$, $p < 0.01$; Table OS-9), and perceived reassurance through human voice ($\beta = -0.53$, $p < 0.001$; Table OS-10). At the same time, OL oversight is not significantly related to perceived outcome or process control.

However, some of these associations are attenuated once intervention design and contextual factors are controlled for (see Model 6). In these fully specified models, the negative associations between OL oversight and relational dimensions (benevolence and respectful treatment) become statis-

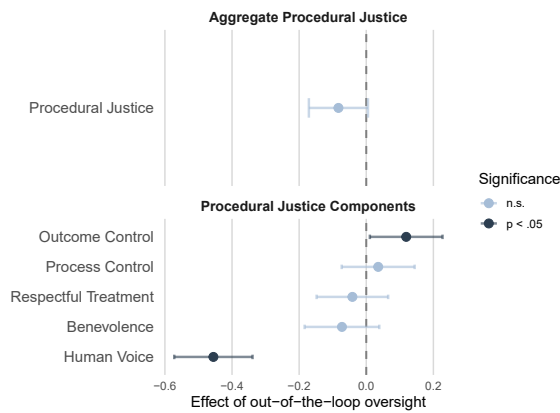


Figure 2: Effects of out-of-the-loop oversight on perceived procedural justice, and components of procedural justice. The abbreviation n.s. refers to not significant ($p > 0.05$).

tically insignificant. Yet, what remains stable is the strongly negative association between OL oversight and perceived reassurance through human voice ($\beta = -0.46$, $p < 0.001$; Table OS-10). In addition, in fully specified models, OL oversight is positively associated with outcome control ($\beta = 0.12$, $p < 0.01$; Table OS-12). Thus, when context and policy design are accounted for, OL oversight is associated with increased perceived outcome control, and, simultaneously, with a decrease in reassurance through human voice. These effects are illustrated in Figure 2.

Across all models, we observe three stable patterns. First, distinct forms of oversight are not significantly associated with process control. As such, human oversight mechanisms do not alter perceptions of influence over how decisions are made. Second, OL oversight is consistently negatively associated with perceived reassurance through human voice. Third, distinct facets of AI literacy do not moderate the effect of OL oversight on perceived procedural justice.

Data Evaluation Practices and AI Literacy

Aggregate Procedural Justice Across distinct model specifications, we do not find evidence of a consistent main effect of rule-based versus AI-based data evaluation on perceived procedural justice.

However, we find a significant interaction between AI awareness and data evaluation practices, which implies that the effect of AI awareness depends on which data evaluation method is used, and vice versa. When systems rely on AI-based evaluations, a one unit increase in AI awareness is positively associated with perceived procedural justice ($\beta = 0.12$, $p < 0.05$; Model 6; Table OS-7). However, as indicated by a negative interaction between rule-based evaluation and AI awareness ($\beta = -0.18$, $p < 0.01$; Model 6; Table OS-7), this association is reversed in rule-based systems. Here, the marginal effect of AI awareness becomes slightly negative when systems operate in a rule-based manner ($\Delta\beta = -0.06$). Therefore, an increase in AI awareness is associated with an increase in perceived procedural justice in AI-based systems, but with a decrease in rule-based

systems. Reversely, this same interaction shows that at average levels of AI awareness, there is no effect of using rule-based vs. AI-based models on perceived procedural justice. However, as AI awareness increases, perceptions of procedural justice of rule-based systems marginally decrease, as indicated by the negative interaction effect ($\beta = -0.18$, $p < 0.01$; Model 6; Table OS-7), resulting in a negative marginal net effect ($\Delta\beta = -0.13$).

Components of Procedural Justice We next examine which components of procedural justice drive the interaction between data evaluation practice and AI awareness. We find that the significant interaction between AI awareness and data evaluation is driven primarily through a negative joint effect on *process control* ($\beta = -0.33$, $p < 0.001$, Model 6; Table OS-11). We observed no significant associations for outcome control, benevolent intentions, or respectful treatment, and a moderate impact on reassurance through human voice ($\beta = -0.19$, $p < 0.05$, Model 6; Table OS-10). As such, the crossover pattern observed for aggregate procedural justice is primarily driven by perceived control over decision-making processes. Substantively, individuals with higher AI awareness perceive greater process control under AI-based evaluation systems, whereas individuals with lower AI awareness perceive reduced process control under the same conditions.

Estimated Marginal Means We examine how differences in aggregate procedural justice and in perceived process control between rule-based and AI-based systems vary across different levels of AI awareness. Therefore, we computed estimated marginal means at low ($-1SD$), average, and high ($+1SD$) levels of AI awareness using the `emmeans` package, and based on the fully specified model (Model 6). The results show that at low levels of AI awareness ($-1SD$), rule-based data evaluation practices are associated with higher perceived procedural justice ($M = 3.29$, 95% CI [3.19, 3.39]) compared to AI-based data evaluation practices ($M = 3.07$, 95% CI [2.96, 3.17]). At the mean level of AI awareness, the difference between conditions is small, with slightly higher perceived procedural justice in the rule-based condition ($M = 3.26$ vs. 3.21). At high AI awareness ($+1SD$) AI-based evaluation is associated with higher perceived procedural justice ($M = 3.36$, 95% CI [3.25, 3.47]) than rule-based evaluation ($M = 3.23$, 95% CI [3.13, 3.33]). The results thus show a crossover interaction pattern. While AI-based evaluation are perceived as less procedurally just for those with low AI awareness, they are perceived as more procedurally just for those with high AI awareness. Here, the overall effect is mainly driven through the effect on *process control* (see Tables OS-13 and OS-14 in the Online Supplement). The estimated means of procedural justice as a composite construct are shown in Figure 3.

Discussion

In this paper, we investigate the relationship between algorithmic procedures, AI literacy, and perceived procedural justice. Focusing on two procedural configurations, namely, data evaluation practices and human oversight, we develop a conditional account of how individuals evaluate procedural

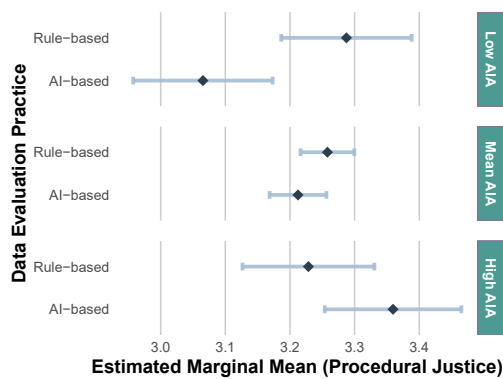


Figure 3: Predicted means of perceived procedural justice at mean AI awareness (AIA), low AIA (-1 SD), and high AIA ($+1$ SD).

justice in algorithmic systems. Accounting for the relational and instrumental components of procedural justice, we theorize that the relationship between algorithmic procedures and perceived procedural justice depends on individuals' AI literacy, as AI literacy shapes how individuals interpret and evaluate procedural choices. For data evaluation practices, AI literacy is relevant through its association with computational thinking (Celik 2023), whereas for human oversight, AI literacy relates to individuals' perceived ability to question and contest algorithmic decisions.

Procedural Justice Components: Nuanced Effects

The effect of procedures on judgments becomes more clearly visible once distinct dimensions of procedural justice are taken into consideration. For human oversight, we found that the apparent null effect on the composite construct of perceived procedural justice is due to its differential effect on distinct components of procedural justice. When human oversight is implemented out-of-the-loop, people's perception of being able to control *outcomes* increases. However, in that same configuration, people's perceived reassurance through human voice within decision-making processes decreases. As such, out-of-the-loop oversight positively relates to instrumental outcome control, and negatively relates to the feeling of being heard. These mechanisms are concealed when procedural justice is averaged over its components.

For data evaluation practices, we find that the effects procedural justice are mainly driven by a positive joint effect of data evaluation practices and AI literacy on *process control*. As such, perceptions are greatly affected by the association between data evaluation practices and one specific instrumental aspect of procedural justice.

Taken together, these findings suggest that null effects in algorithmic decision-making research might reflect offsetting effects across distinct dimensions of a broader construct. This might explain why prior research has not found consensus when researching the procedural configurations studied here (Wang, Harper, and Zhu 2020; Wang et al. 2023; Yurrita et al. 2023; Koulu 2020). At the same time, effects on perceived procedural justice might be subject to

one particularly pronounced underlying driver. Together, our work thus demonstrates that procedural justice in algorithmic systems is conditional on the components of justice affected by specific procedural configurations, and highlights the importance of accounting for both instrumental and relational components of procedural justice.

AI Literacy: Selective Importance

Procedural justice effects might also be conditional on AI literacy. As shown in this work, the apparent null effect of distinct forms of data evaluation arises because AI-based data evaluation practices have distinct effects on people with different levels of AI awareness. When people have high AI awareness, the use of AI-based data evaluation practices increases perceptions of procedural justice. In contrast, when people have low AI awareness, the use of AI-based data evaluation practices decreases these perceptions.

Selective Moderation. Importantly, AI literacy moderates only data evaluation practices, but not on human oversight. This suggests that AI literacy, particularly awareness, *selectively* moderates how individuals evaluate procedural elements of algorithmic systems, depending on the degree to which a procedural feature requires computational and interpretive capacities. In a procedural configuration characterized by strong interpretive demands, such as delineating the implications of AI-based and those of rule-based evaluation practices, AI awareness appears as an important variable that shapes how the justice of these procedures is judged.

By contrast, AI awareness is not decisive for initiating review and contestation of AI decisions, as exemplified by its non-significance in human oversight. As such, our findings suggest that AI literacy does not uniformly shape evaluations of algorithmic procedures, but that its influence depends on the computational opacity of a procedural feature.

Aspects of AI Literacy. Our results highlight the importance of distinguishing between distinct aspects of AI literacy, as not all of its components play an equal role in shaping procedural justice perceptions. Across distinct models, AI usage is consistently positively associated with perceived procedural justice. This suggests that individuals who are more familiar and comfortable integrating AI technologies into daily practices tend to evaluate AI-based systems more favorably, independent of a specific procedural configuration. This likely reflects general technological acceptance and self-efficacy in adopting and engaging with AI systems.

In contrast, AI awareness is the key driver of the moderation effect for data evaluation practices. Awareness captures the ability to recognize and understand AI technologies and their functional implications (Wang, Rau, and Yuan 2023). Therefore, this dimension appears directly relevant for distinguishing between different computational logics within algorithmic decision-making procedures.

Finally, AI evaluation and AI ethics do not exhibit systematic effects in our models. One possible explanation is that these dimensions capture normative reflection and judgment, and refer less to an understanding of computational procedures (Wang, Rau, and Yuan 2023), which might render their influence on *procedural* configurations less salient.

Yet, as these aspects capture important dimensions of engaging and interacting with AI, future research should identify for which algorithmic features individuals' capacity to evaluate and judge AI play a more important role.

Implications for AI Ethics

Our findings have implications for the ethical design of algorithmic systems, particularly regarding the trade-offs between different procedural justice dimensions. In data evaluation practices, observed effects were primarily driven by perceived process control. Other justice components were considerably less impacted. This suggests that aggregate evaluations of procedural justice may conceal the specific mechanisms through which algorithmic procedures shape perceptions of justice.

Such tensions become particularly apparent when distinct components are affected in opposing directions, as exemplified by the relationship between procedural justice and human oversight. Here, we observed a clear tension between increased perceptions of outcome control under out-of-the-loop oversight and reduced perceptions of being reassured through human voice. This highlights that procedural choices might lead to trade-offs, as some choices, which strengthen perceived instrumental control over outcomes, can simultaneously reduce relational judgments. Importantly, this trade-off becomes visible only when we distinguish between different dimensions of procedural justice. Thus, designing ethical algorithmic systems requires balancing potentially competing aspects of procedural justice. The relative importance of relational versus instrumental concerns is likely context-dependent and may vary across individuals and application domains. At the same time, some relational expectations, such as being treated with respect, might function as more stable normative baselines.

A key insight of our study is that procedural justice in algorithmic systems cannot be understood independently of individuals' ability to interpret procedural structures. In algorithmically mediated environments (Dencik et al. 2018), important procedural characteristics may remain epistemically inaccessible to large parts of the population (Lutz 2019). As a result, identical algorithmic procedures can generate fundamentally divergent perceptions of procedural justice across individuals with different levels of AI literacy. These findings underscore the growing importance of AI-relevant competencies for participating in algorithmically governed societies. If interpretive capacities are unevenly distributed, algorithmic procedures may contribute to new forms of inequality in how individuals understand, evaluate, and respond to algorithmic systems (Aydin 2021).

With this, our study contributes to broader discussions on digital divides and algorithmic inequality (Lutz 2019; Van Deursen and Van Dijk 2014). Prior research has primarily focused on how unequal decision outcomes (Wang, Harper, and Zhu 2020; Loefflad and Grossklags 2024b; Loefflad, Chen, and Grossklags 2024a), or different participation levels (Loefflad, Chen, and Grossklags 2025) create perception differences. Our findings suggest that disparities may also emerge at an interpretive level. In particular, AI awareness shapes how individuals judge data evaluation

practices, which implies that procedural justice also depends on interpretive capacities. With this, our work underscores that, depending on the procedural configurations of algorithmic systems, inequalities in AI literacy translate directly into differences in how algorithmic procedures are perceived and evaluated.

Such differences have important implications, as they can directly translate into differences in engaging with or adopting algorithmic systems (Loefflad, Chen, and Grossklags 2025), which ultimately shapes individuals' ability to effectively navigate and benefit from these systems (Al-Hujran et al. 2015). Together, this points toward a new form of digital divide rooted in the ability to trace and interpret algorithmic systems.

Importantly, differences in AI literacy should not be understood as inherently given, but as arising from, for example, differential access to technologies (Aydin 2021). Prior work has illustrated that AI literacy directly reflects digital divides to a large extent (Celik 2023). Together with our findings, this suggests that algorithmic systems may reproduce or even amplify existing inequalities through unequal interpretive capacities, which raises the importance of finding strategies to mitigate these novel forms of digital divide.

Conclusion

In this paper, we developed and empirically examined a conditional account of procedural justice in algorithmic systems. We argued that evaluations of algorithmic procedures depend both on individuals' AI literacy and on the specific dimensions of procedural justice under consideration. Our findings support this perspective. While AI literacy moderated the relationship between data evaluation practices and procedural justice perceptions, human oversight exhibited opposing effects across distinct components of procedural justice. These findings demonstrate that seemingly null or inconsistent effects may conceal more differentiated underlying dynamics. By decomposing procedural justice into relational and instrumental components, and by treating AI literacy as a multidimensional construct, our study uncovers patterns that may remain obscured in approaches relying on aggregated scales. More broadly, our work highlights that procedural evaluations of algorithmic systems are not uniform across individuals, but depend on people's capacity to interpret and engage with algorithmic procedures. As algorithmic systems increasingly shape access to opportunities and social participation, understanding these interpretive differences becomes essential for both research and the ethical design of algorithmic systems. We hope that future research further investigates the implications of distinct interpretive capacities, and finds strategies to mitigate that differences in interpretation create novel forms of digital divides.

Declaration of Generative AI Use

During the preparation of this manuscript the authors used ChatGPT in order to improve the wording of specific sentences. After using this tool, the authors reviewed and edited the content as needed. The tool did not contribute to the study design, data collection, or analysis.

Acknowledgments

We thank the anonymous reviewers for their constructive comments and the participants of our study for their contributions. Carmen Loefflad gratefully acknowledges support through the Laura Bassi Scholarship Program by the Bavarian State Ministry of Science and the Arts.

References

- Al-Hujran, O.; Al-Debei, M. M.; Chatfield, A.; and Migdadi, M. 2015. The Imperative of Influencing Citizen Attitude toward E-government Adoption and Use. *Computers in Human Behavior*, 53: 189–203.
- Arrieta, A. B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58: 82–115.
- Aydin, M. 2021. Does the digital divide matter? Factors and conditions that promote ICT literacy. *Telematics and Informatics*, 58: 101536.
- Celik, I. 2023. Exploring the Determinants of Artificial Intelligence (AI) Literacy: Digital Divide, Computational Thinking, Cognitive Absorption. *Telematics and Informatics*, 83: 102026.
- Chen, M.; Bogner, K.; and Grossklags, J. 2025. A Comparative Analysis of Two Data-Driven Mechanisms for Behavioral Change: Nudging and the Chinese Social Credit System. *Digital Society*, 4(1): 1–24.
- Colaner, N. 2022. Is Explainable Artificial Intelligence Intrinsically Valuable? *AI & Society*, 37: 231–238.
- Custers, B. 2004. *The Power of Knowledge: Ethical, Legal and Technological Aspects of Data Mining and Group Profiling in Epidemiology*. Wolf Legal Publishers.
- Custers, B.; and Vrabec, H. 2024. Tell Me Something New: Data Subject Rights Applied to Inferred Data and Profiles. *Computer Law & Security Review*, 52: 105956.
- Dencik, L.; Hintz, A.; Redden, J.; and Warne, H. 2018. Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services. Project report, Open Society Foundations.
- Durán, J. M.; and Jongsma, K. R. 2021. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. *Journal of Medical Ethics*, 47(5): 329–335.
- Engelmann, S.; Ullstein, C.; Papakyriakopoulos, O.; and Grossklags, J. 2022. What People Think AI Should Infer From Faces. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 128–141. Association for Computing Machinery.
- Fontes, C.; Hohma, E.; Corrigan, C. C.; and Lütge, C. 2022. AI-powered public surveillance systems: Why we (might) need them and how we want them. *Technology in Society*, 71: 102137.
- Gonzalez, M. F.; Liu, W.; Shirase, L.; Tomczak, D. L.; Lobbe, C. E.; Justenhoven, R.; and Martin, N. R. 2022. Ally-ing with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes. *Computers in Human Behavior*, 130: 107179.
- Green, B. 2022. The Flaws of Policies Requiring Human Oversight of Government Algorithms. *Computer Law & Security Review*, 45: 105681.
- Hase, P.; and Bansal, M. 2020. Evaluating explainable AI: Which algorithmic explanations help users predict model behavior? In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, ACL '20, 5540–5552. Association for Computational Linguistics.
- Hildebrandt, M. 2008. Defining Profiling: A New Type of Knowledge? In Hildebrandt, M.; and Gutwirth, S., eds., *Profiling the European Citizen*, 17–45. Springer.
- Igbaria, M.; and Iivari, J. 1995. The effects of self-efficacy on computer usage. *Omega*, 23(6): 587–605.
- Innerarity, D. 2021. Making the Black Box Society Transparent. *AI & Society*, 36: 975–981.
- Juijn, G.; Stoimenova, N.; Reis, J.; and Nguyen, D. 2023. Perceived Algorithmic Fairness using Organizational Justice Theory: An Empirical Case Study on Algorithmic Hiring. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, 775–785. Association for Computing Machinery. ISBN 9798400702310.
- Just, N.; and Latzer, M. 2017. Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet. *Media, Culture & Society*, 39(2): 238–258.
- Kitchin, R. 2017. Thinking Critically About and Researching Algorithms. *Information, Communication & Society*, 20(1): 14–29.
- Koulu, R. 2020. Proceduralizing Control and Discretion: Human Oversight in Artificial Intelligence Policy. *Maasticht Journal of European and Comparative Law*, 27(6): 720–735.
- Kyriakou, K.; and Otterbacher, J. 2023. In Humans, We Trust: Multidisciplinary Perspectives on the Requirements for Human Oversight in Algorithmic Processes. *Discover Artificial Intelligence*, 3(1): 44.
- Lazcoz, G.; and de Hert, P. 2023. Humans in the GDPR and AIA Governance of Automated and Algorithmic Systems. Essential Pre-requisites Against Abdicating Responsibilities. *Computer Law & Security Review*, 50: 105833.
- Lee, M. K.; Jain, A.; Cha, H. J.; Ojha, S.; and Kusbit, D. 2019. Procedural Justice in Algorithmic Fairness: Leveraging Transparency and Outcome Control for Fair Algorithmic Mediation. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW): 182.
- Leventhal, G. S. 1980. What Should Be Done with Equity Theory? New Approaches to the Study of Fairness in Social Relationships. In Gergen, K. J.; Greenberg, M. S.; and Willis, R. H., eds., *Social Exchange: Advances in Theory and Research*, 27–55. Springer. ISBN 978-1-4613-3087-5.
- Lind, E.; Kanfer, R.; and Earley, P. 1990. Voice, Control, and Procedural Justice: Instrumental and Noninstrumental

- Concerns in Fairness Judgments. *Journal of Personality and Social Psychology*, 59: 952–959.
- Linders, D. 2012. From E-Government to We-Government: Defining a Typology for Citizen Coproduction in the Age of Social Media. *Government Information Quarterly*, 29(4): 446–454.
- Loefflad, C. 2024. Automated Decision-Making Systems for Behavioral Regulation: Understanding Perceptions and Behavioral Reactions. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, 28–31.
- Loefflad, C.; Chen, M.; and Grossklags, J. 2023. Factors Influencing Perceived Legitimacy of Social Scoring Systems: Subjective Privacy Harms and the Moderating Role of Transparency. In *Proceedings of the International Conference on Information Systems*, ICIS '23, 1–17. Association for Information Systems.
- Loefflad, C.; Chen, M.; and Grossklags, J. 2024a. Reputational Discrimination and Fairness in China's Social Credit System. *Digital Government: Research and Practice*, 5(4): 40.
- Loefflad, C.; Chen, M.; and Grossklags, J. 2024b. Social Scoring Systems for Behavioral Regulation: An Experiment on the Role of Transparency in Determining Perceptions and Behaviors. In *Proceedings of the 7th AAAI/ACM Conference on AI, Ethics, and Society*, AIES '24, 891–904. Association for the Advancement of Artificial Intelligence.
- Loefflad, C.; Chen, M.; and Grossklags, J. 2025. Citizen involvement in algorithmic regulation: How awareness and participation shape public opinion, civic behaviors and social trust. *Human Behavior and Emerging Technologies*, 2025: 9184190.
- Loefflad, C.; and Grossklags, J. 2024a. AI Literacy and Procedural Justice in Algorithmic Systems: A Conditional Account on Data Evaluation Practices and Human Oversight – Supplementary File. <https://osf.io/4gq6d>. Last accessed on August 7, 2026.
- Loefflad, C.; and Grossklags, J. 2024b. How the Types of Consequences in Social Scoring Systems Shape People's Perceptions and Behavioral Reactions. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, 1515–1530. Association for Computing Machinery.
- Loefflad, C.; and Grossklags, J. 2026. Algorithmic Regulation as a Policy Approach to Collective Action: Conceptual Framework and Citizen Perspectives. Manuscript under review.
- Long, D.; and Magerko, B. 2020. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–16. Association for Computing Machinery.
- Lutz, C. 2019. Digital Inequalities in the Age of artificial intelligence and Big Data. *Human Behavior and Emerging Technologies*, 1(2): 141–148.
- Martin, K.; and Waldman, A. 2022. Are Algorithmic Decisions Legitimate? The Effect of Process and Outcomes on Perceptions of Legitimacy of AI Decisions. *Journal of Business Ethics*, 183: 653–670.
- Marx, G. T. 1998. Ethics for the New Surveillance. *The Information Society*, 14(3): 171–185.
- Mittelstadt, B. D.; Allo, P.; Taddeo, M.; Wachter, S.; and Floridi, L. 2016. The Ethics of Algorithms: Mapping the Debate. *Big Data & Society*, 3(2).
- Nagtegaal, R. 2021. The Impact of Using Algorithms for Managerial Decisions on Public Employees' Procedural Justice. *Government Information Quarterly*, 38(1): 101536.
- Nakao, Y.; Stumpf, S.; Ahmed, S.; Naseer, A.; and Strapelli, L. 2022. Toward Involving End-users in Interactive Human-in-the-loop AI Fairness. *ACM Transactions on Interactive Intelligent Systems*, 12(3): 1–30.
- Newman, D. T.; Fast, N. J.; and Harmon, D. J. 2020. When Eliminating Bias isn't Fair: Algorithmic Reductionism and Procedural Justice in Human Resource Decisions. *Organizational Behavior and Human Decision Processes*, 160: 149–167.
- Pasquale, F. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press. ISBN 9780674368279.
- Patil, S.; Wenzel, T.; and Grossklags, J. 2026. Investigating How Types of Data Associated With Smart Home Devices Influence Privacy Concerns and Perceived Benefits. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, CHI '26, 1–21. Association for Computing Machinery.
- Schoeffler, J.; Machowski, Y.; and Kuehl, N. 2022. Perceptions of fairness and trustworthiness based on explanations in human vs. automated decision-making. In *Proceedings of the 55th Hawaii International Conference on System Sciences*, HICSS '22, 1095–1102. Association for Information Systems.
- Starke, C.; and Lünich, M. 2020. Artificial intelligence for political decision-making in the European Union: Effects on citizens' perceptions of input, throughput, and output legitimacy. *Data & Policy*, 2: e16.
- Sterz, S.; Baum, K.; Biewer, S.; Hermanns, H.; Lauber-Rönsberg, A.; Meinel, P.; and Langer, M. 2024. On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT'24, 2495–2507. Association for Computing Machinery.
- Thibaut, J. W.; and Walker, L. 1975. *Procedural Justice: A Psychological Analysis*. L. Erlbaum Associates. ISBN 9780470858684.
- Türker, P. M.; and Pala, F. K. 2020. The effect of algorithm education on students' computer programming self-efficacy perceptions and computational thinking skills. *International Journal of Computer Science Education in Schools*, 3(3): 19–32.
- Tyler, T. R. 2006. *Why People Obey the Law*. Princeton University Press. ISBN 9780691126739.
- Tyler, T. R.; and Lind, E. 1992. A Relational Model of Authority in Groups. *Advances in Experimental Social Psychology*, 25: 115–192.

- Tyler, T. R.; Rasinski, K.; and Spodick, N. 1985. Influence of Voice on Satisfaction With Leaders. Exploring the Meaning of Process Control. *Journal of Personality and Social Psychology*, 48: 72–81.
- Ullstein, C.; Engelmann, S.; Papakyriakopoulos, O.; Hohen-danner, M.; and Grossklags, J. 2022. AI-Competent Individuals and Laypeople Tend to Oppose Facial Analysis AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '22. Association for Computing Machinery.
- Van Deursen, A. J.; and Van Dijk, J. A. 2014. The Digital Divide Shifts to Differences in Usage. *New Media & Society*, 16(3): 507–526.
- van Zeben, J.; and Kamphorst, B. A. 2020. Tracking and Nudging through Smartphone Apps: Public Health and Decisional Privacy in a European Health Union. *European Journal of Risk Regulation*, 11(4): 831–840.
- Wang, B.; Rau, P.-L. P.; and Yuan, T. 2023. Measuring User Competence in Using Artificial Intelligence: Validity and Reliability of Artificial Intelligence Literacy Scale. *Behaviour & Information Technology*, 42(9): 1324–1337.
- Wang, C.; Wang, H.; Li, Y.; Dai, J.; Gu, X.; and Yu, T. 2025. Factors Influencing University Students' Behavioral Intention to Use Generative Artificial Intelligence: Integrating the Theory of Planned Behavior and AI Literacy. *International Journal of Human-Computer Interaction*, 41(11): 6649–6671.
- Wang, G.; Guo, Y.; Zhang, W.; Xie, S.; and Chen, Q. 2023. What Type of Algorithm is Perceived as Fairer and More Acceptable? A Comparative Analysis of Rule-driven Versus Data-driven Algorithmic Decision-making in Public Affairs. *Government Information Quarterly*, 40(2): 101803.
- Wang, R.; Harper, F. M.; and Zhu, H. 2020. Factors Influencing Perceived Fairness in Algorithmic Decision-Making: Algorithm Outcomes, Development Procedures, and Individual Differences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–14. Association for Computing Machinery. ISBN 9781450367080.
- Yurrita, M.; Draws, T.; Balayn, A.; Murray-Rust, D.; Tintarev, N.; and Bozzon, A. 2023. Disentangling Fairness Perceptions in Algorithmic Decision-Making: The Effects of Explanations, Human Oversight, and Contestability. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, 1–21. Association for Computing Machinery. ISBN 9781450394215.