UNIVERSITY OF AMSTERDAM

# UvA-DARE (Digital Academic Repository)

## Procedural Justice and Judicial AI

*Substantiating Explainability Rights with the Values of Contestation*

Metikoš, L.; van Domselaar, I.

**DOI**
10.59490/jhtr.2025.3.8163

**Publication date**
2025

**Document Version**
Final published version

**Published in**
Journal of Human-Technology Relations

**License**
CC BY

Link to publication

**Citation for published version (APA):**
Metikoš, L., & van Domselaar, I. (2025). Procedural Justice and Judicial AI: Substantiating Explainability Rights with the Values of Contestation. *Journal of Human-Technology Relations*, *3*. https://doi.org/10.59490/jhtr.2025.3.8163

**General rights**
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)*

Download date:05 Sep 2026

TU Delft OPEN

JHTR
Journal of Human-Technology Relations

# PROCEDURAL JUSTICE AND JUDICIAL AI

## Substantiating Explainability Rights with the Values of Contestation

Ljubiša Metikoš l.metikos@uva.nl
University of Amsterdam, Faculty of Law - Amsterdam, The Netherlands, ORCiD 0009-0003-9436-8737

Iris van Domselaar i.vandomselaar@uva.nl
University of Amsterdam, Faculty of Law - Amsterdam, The Netherlands, ORCiD 0000-0001-6119-5559

Article type: Research article

Review process: Double-blind peer review

Topical Collection: Ethics and Normativity of Explainable AI: Explainability as a Social Practice (Guest editors: Tobias Matzner, Suzana Alpsancar, Martini Philippi, Wessel Reijers)

This open-access article is published with a Creative Commons CC-BY 4.0 license
https://creativecommons.org/licenses/by/4.0/

DOI: 10.59490/jhtr.2025.3.8163
ISSN: 2773-2266
Submitted: 6 May 2025 Revised: 20 October 2025 Accepted: 15 December 2025
Published: 30 December 2025

How to cite (APA): Metikoš, L., Van Domselaar, I. (2025). Procedural Justice and Judicial AI: Substantiating Explainability Rights with Values of Contestation. *Journal of Human-Technology Relations*, *3*(1), pp.1-34. https://doi.org/10.59490/jhtr.2025.3.8163.

Corresponding author: Ljubiša Metikoš

©2025 Ljubiša Metikoš and Iris van Domselaar, published by TU Delft OPEN on behalf of the authors.

**Keywords**

Explainability; Procedural Justice; AI; Legal Tech; Digital Justice; Courts; Adjudication; Transparency; Contestation

**Abstract**

The advent of opaque assistive AI in courtrooms has raised concerns about the contestability of these systems, and their impact on procedural justice. The right to an explanation under the GDPR and the AI Act could address the inscrutability of judicial AI for litigants. To substantiate this right in the domain of justice, we examine utilitarian, rights-based (including dignitarian and Dworkinian approaches), and relational theories of procedural justice. These theories reveal diverse perspectives on contestation, which can help shape explainability rights in the context of judicial AI. These theories respectively highlight different values of litigant contestation; it has instrumental value in error correction, and intrinsic value in respecting litigants' dignity, either as rational autonomous agents or as socio-relational beings. These insights help us answer three central and practical questions on how the right to an explanation should be operationalized to enable litigant contestation: should explanations be general or specific, to what extent do explanations need to be faithful to the system's internal behavior or merely provide a plausible approximation, and should more interpretable systems be used, even at the cost of accuracy? These questions are not strictly legal or technical in nature, but also rely on normative considerations. Finally, this paper also evaluates what theory of procedural justice could best safeguard contestation effectively in the age of judicial AI. Thereto, it provides the first building blocks of an AI-responsive theory of procedural justice.

**Plain-language Summary [TBC]**

# 1. INTRODUCTION

Artificial Intelligence (AI) is increasingly used by courts across the world to assist judges in adjudicating cases. AI systems can provide judges with analyses of both the facts and the applicable laws of cases, which can help them reach their verdicts (Fabri, 2024; Sourdin, 2021b; Stern et al., 2021). Profiling systems can, for example, generate analyses of the character of litigants (Metikoš, 2024). AI has also been used across jurisdictions to analyze, summarize, and predict case law (ADELE, 2025). In addition, generative AI systems that rely on Large Language Models (LLMs) are able to provide quick answers to legal questions and even write entire verdicts (Osa & Remolina, 2024; Surden, 2024).

The use of these various systems has been lauded by some commentators for their potential to improve the scalability and quality of adjudication (Volokh, 2019). Similar arguments in the defense of judicial AI can be seen in the fast-paced adoption of AI in Chinese courts (Stern et al., 2021). It is said that AI promises to improve the speed of judicial decision-making and lessen the costs of judiciaries, by helping to manage the workload of judges and the paralegals that support them. The phrase '*better, faster, cheaper'* justice emphasizes these core considerations as to the efficiency benefits of the use of judicial AI (Re & Solow-Niederman, 2019).

At the same time, the implementation of assistive AI in courts has been severely criticized by legal scholars. Judicial AI can produce outputs that are inequitable, inaccurate, or even discriminatory (Re & Solow-Niederman, 2019). In addition, these errors might remain uncontested. For various reasons, judicial AI can be inscrutable for individuals subjected to AI-supported decision-making procedures (Bayamlioglu, 2018; Selbst & Barocas, 2018; Yeung & Harkens, 2023). For instance, if a litigant does not know how they were profiled, or why a certain text or legal argument was generated, they are unable to voice their opinion about whether the AI system relied on correct and reasonable grounds. As a way to address this opacity, scholars have argued that a '*right to an explanation*' could enable individuals to contest AI (Bayamlioglu, 2018; Sarra, 2020). Kaminski and Urban (2021) argue, for example, that contestation without an explanation is a '*meaningless endeavor'*, stressing the contingency of contestation on the explainability of (judicial) AI.

Art. 6 of the European Convention of Human Rights (ECHR) provides litigants with a right to a reasoned judgment and a right to adversarial proceedings, requiring basic transparency on any information that might be in the hands of the court (*H. v. Belgium*, 1987, para. 53). EU law furthermore addresses the need for transparency in regard to (judicial) AI, and provides individuals with a right to an explanation about certain AI systems under the General Data Protection Regulation (GDPR) and the AI Act (AIA) (Metikoš & Ausloos, 2025).

Both in the legal debate that has arisen about the latter two laws, as well as in the computer science field of explainable AI (XAI), scholars have discussed different ways to technically operationalize the right to an explanation (Brkan & Bonnet, 2020; Nišević et al., 2024). For the context of judicial AI, we hold that three questions are central to further substantiating the contestation-enabling right to an explanation: how general or specific should explanations be, how faithful must they be to the system's internal behaviors (or could a plausible approximation also be allowed), and should more faithfully interpretable systems be used even at the cost of accuracy?

The answers to these questions are not purely legal or technical in nature. Rather, they require both a normative and, importantly, a procedural perspective on the value of a contestation-

enabling right to an explanation in the context of judicial AI.[1] A reasoned answer to these questions is intimately linked to the values that are to be realized by the contestation of judicial AI in the context of a legal procedure.[2] Scholars discussing the explainability and contestability of AI in decision-making have proposed several values in this regard. One such value has been accuracy; contesting errors or biases present in an AI system has been defended as vital to improving the accuracy and quality of a decision (Doshi-Velez et al., 2019; Selbst & Barocas, 2018; Zarsky, 2013). In addition, safeguarding contestation has also been defended as valuable in and of itself, regardless of whether it makes any difference to the final judgment (Kaminski & Urban, 2021; Miao, 2022; Naudts & Vedder, 2025; Selbst & Barocas, 2018).

To assess the merits of these arguments in the context of *judicial* AI, we will use the lens offered by four legal-philosophical approaches to procedural justice, which provide a moral view on the procedural arrangements of legal decision-making (Meyerson & Mackenzie, 2018). We discuss utilitarian (specifically a law and economics approach), rights-based (including a Dworkinian outcome-based approach and a dignitarian process-based approach), and relational theories of procedural justice to help us reason about why and how contestation should be safeguarded through a right to an explanation in the context of judicial AI.[3] More specifically, we show how these approaches can help to identify the underlying normative assumptions in the legal and XAI debates on the concept of a contestation-enabling right to an explanation, and help to identify the underlying normative arguments on the need for explainability that have been put forward in academic debates.

By using theories of procedural justice in the context of judicial AI, this paper also fills a gap in the legal-philosophical discourse on procedural justice. This scholarly debate has so far hardly addressed the emergence of digital technologies in the justice system.[4] Moreover, legal theory, in general, has not been sufficiently concerned with overcoming the enormous justice gap. It focuses too much on how ideal procedures should look like, rather than genuinely engaging with the state of the current justice system (Van Domselaar, 2022b; Susskind, 2019).

The structure of this paper is as follows: in section 1 we shall first discuss the legal basis for the right to an explanation and its goal of contestation, under the GDPR and the AIA. We show how these rights apply to the judicial context, as well as how the right to an explanation has been formulated to promote the contestability of decision-making procedures. In section 3 we introduce, based on this discussion, three central questions on how explainability must be operationalized that we have deduced from the literature on explainable AI (XAI). Subsequently, we aim to answer these questions in sections 4, 5 and 6, by applying different theories of procedural justice. Finally, in section 7, after synthesizing these different views, we propose that a relational approach to procedural justice could potentially best safeguard the contestation of

[1] For a more expansive look on the methodology and importance of adopting philosophical perspectives in legal-doctrinal research see: (Taekema & Burg, 2020). For similar arguments being raised in the field of value-based design see: (Buijsman et al., 2025).

[2] It is salient to note that the AI Act barely addresses procedural matters. The right to an explanation (art. 86) was a late addition during its legislative process (Metikoš & Ausloos, 2025). Apart from art. 85 of the AI Act , which grants the right to lodge a complaint, no other procedural rights are included. Similar gaps exist in the wider field of AI ethics, which has focused more on substantive justice rather than procedural justice (Kitchin, 2017). This paper therefore helps elaborate on these rights and the wider field of AI ethics, by connecting them with insights from the field of procedural justice.

[3] Some of the theories that we discuss here concern views from common law scholars or specific kinds of adjudicatory practices, such as the civil, criminal, or administrative context. Nevertheless, these theories do contain normative outlooks on what constitutes a just procedure in *general*. We will therefore incorporate these theories to the extent that the comments made by their authors are applicably more widely.

[4] For example, the edited volume on Procedural Justice and Relational Theory (Meyerson et al., 2021), which covers a wide range of contemporary concerns, gives little attention to the role of digitization.

judicial AI by focusing on the practical effectiveness of the right to an explanation. We then conclude and summarize our findings in section 8.

# 2. THE LEGAL FOUNDATION OF THE RIGHT TO AN EXPLANATION AND CONTESTATION

In Europe, the core provision on the transparency of judicial decision-making can be found under the right to a fair trial. This right has been laid down in art. 6 of the European Convention of Human Rights (ECHR). This includes the right to a reasoned judgment and the right to adversarial proceedings. These two obligations, respectively, entail that judges must substantiate their verdicts with reasons, and disclose all relevant documentation in the hands of the court (*H. v. Belgium*, 1987, para. 53; *Ruiz-Mateos v. Spain*, 1993, para. 63).[5] While these provide a potential basis for some transparency requirements for judicial AI (Hendrickx, 2025), to date, there is no case law that clarifies that clarifies their implications for the use of judicial AI systems.

In addition to these general procedural obligations, specific AI-based transparency obligations can now be found in the form of *explainability rights*. In the EU, we can find these rights to an explanation in the General Data Protection Regulation (GDPR), adopted in 2016, and the AI Act (AIA), adopted in 2024. Under the GDPR, the right to an explanation has been deduced in scholarship and case law from a combined reading of several provisions, most notably art. 13, 14, 15, and 22 GDPR. Based on these provisions, the Court of Justice of the EU (CJEU) has ruled that data subjects have *'a genuine right to an explanation as to the functioning of the mechanism involved in automated decision-making'* (Metikoš & Ausloos, 2025; Dun & Bradstreet Austria, 2025, para 57). The CJEU (Dun & Bradstreet Austria, 2025, para 55) has also stated that the '*main purpose'* of the right to an explanation is to enable contestation.

At first sight, the GDPR could be a valuable legal instrument for litigants wishing to contest judicial AI. However, as the GDPR's right to an explanation is limited to *solely* automated decision-making, this right would only apply to so-called 'robot judges' (Metikoš, 2024). Currently, however, AI development for the justice sector has been focused far more on the supportive role of legal tech, rather than the full replacement of human judges (Fabri, 2024; Sourdin, 2018, 2021b; Tim Wu, 2019). Examples include the OLGA system used in Germany, which helps with case categorization, metadata extraction, and the search for specific information within thousands of cases (Schindler, 2025), the ADELE system, which is a pilot developed for Italian and Bulgarian judges that can help predict the outcome of cases (ADELE, 2025), or profiling systems such as OxRec, which is used in Sweden to predict the risk recidivism risk of inmates (Metikoš, 2024). Generative AI systems based on LLMs have also been implemented in a pilot in the Netherlands to help judges write verdicts in criminal cases (District Court of Rotterdam, 2025). Considering therefore the more widespread usage of *supportive* rather than *replacive* AI systems (Sourdin, 2021b), the GDPR's right to an explanation seems ill-equipped to regulate the judicial sector and to provide a contestation-enabling right to an explanation.

[5] We do not focus on the role of art. 6 ECHR in this paper, although there is a wider ongoing debate in the literature about its relevance for (transparent) judicial AI (Hendrickx, 2025; Metikos, 2024; Palmiotto, 2021). This paper focuses specifically on how contestability is safeguarded through explainability requirements for judicial AI, not on broader procedural transparency standards. Explainability rights under the GDPR and the AI Act deal, in this regard, more directly with the transparency issues plaguing judicial AI. This is not to deny the valuable insights further analysis of art. 6 ECHR may offer, but rather to clarify the scope of our analysis to be more concretely focused on legal provisions that directly formulate explainability requirements.

Under the AIA, a more directly applicable right to an explanation exists under art. 86 (1) AIA. This provision does include AI systems that assist (rather than replace) judges in researching and interpreting the facts or applicable laws of a case (Metikoš, 2024). It prescribes that a decision subject should receive '*meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken*'. This mimics much of the wording of the GDPR's right to an explanation. Recital 171 AIA provides some clarification as to the goal of this right. It states that an explanation '*should provide a basis on which the affected persons are able to exercise their rights*'. This also mimics how the GDPR's right to an explanation has been understood by the CJEU; as instrumental to enabling an individual to contest the outputs of AI. Consequently, the right to an explanation, both under the GDPR and the AIA, can be understood as an '*action-enabling*' type of transparency (Selbst & Barocas, 2018) that makes it possible for litigants to contest judicial AI systems.

However, no further clarification as to *why* contestation should be safeguarded is mentioned within the Recitals of these two regulations. Questions, therefore, remain as to what contestation-safeguarding explanations should actually enable. Should they, for instance, help identify errors, increase participation of decision-subjects, foster trust, or communicate respect?

It is difficult to envision what kind of explanation would be needed when these regulations do not state any clear normative goals, besides contestability. Various, sometimes conflicting, answers have consequently been raised by legal and computer science scholars as to what an explanation should look like under the GDPR (Brkan & Bonnet, 2020; Edwards & Veale, 2017; Wachter, Mittelstadt, & Russell, 2017). Before we can discuss the contours of these different types of explanations we shall first formulate three central questions in the field of XAI. These questions need to be answered first, if we wish to define how a right to an explanation can safeguard litigant contestation of judicial AI.

# 3. EXPLAINABILITY QUESTIONS: SPECIFICITY, FAITHFULNESS, AND TRADE-OFFS

The right to an explanation under both the GDPR and the AIA is formulated quite broadly and allows for different methods to achieve explainability. However, defining explanations *in abstracto* is challenging, and perhaps even impossible, as Brkan and Bonnet (2020) argue. Hence, below, we raise three questions as to what kind of information should be provided to a litigant who wishes to contest a judicial AI system, and how this might be (technically) achieved. We shall also argue that these questions, which thus far have been mostly conceptualized as legal-doctrinal or computer science debates, are highly normative in nature and intimately linked to a view on what values a legal procedure should realize.

## 3.1 GENERAL AND SPECIFIC EXPLANATIONS

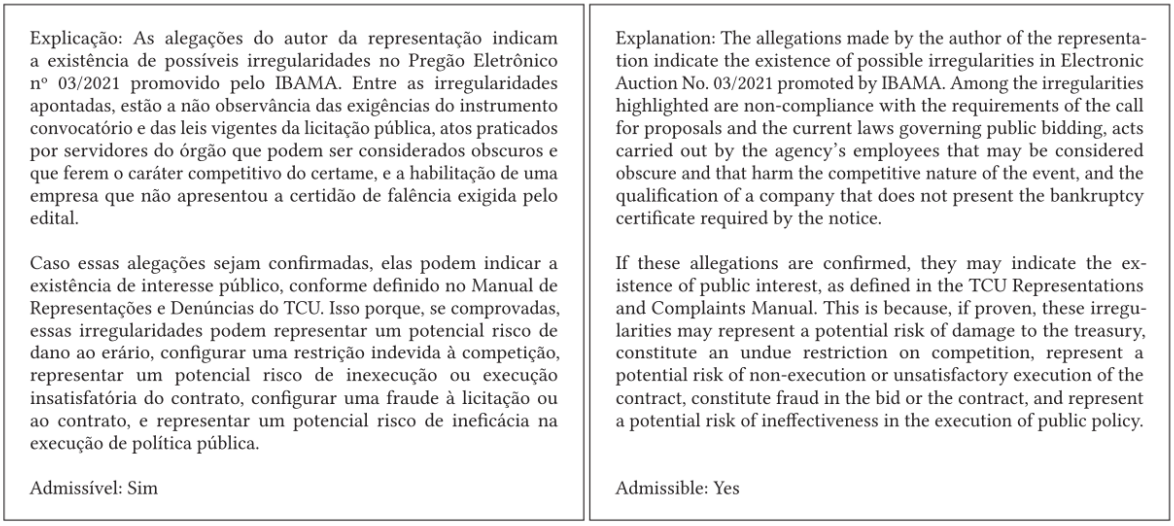
Legal scholarship on the GDPR's right to an explanation, as well as XAI scholarship more broadly, has asked whether explanations must show how an AI system works in *general*, or how a *specific* output came to be. This distinction, between disclosing general and specific explanations, has also been termed as '*weak*' versus '*strong*' explanations (de Laat, 2022), or '*system functionality*' versus '*specific decisions*' type explanations (Wachter, Mittelstadt, & Russell, 2017).

General explanations aim to provide, as the name implies, general information about a system (de Laat, 2022). Under this category, information can be included as to how the system typically processes inputs, its performance metrics, or how the model was audited and tested. Moreover, this information can also be granted *ex ante*, before the system has been applied to the case at

hand. To give an example of this type of explanation in the context of judicial AI, we can look at the ADELE (2025) system (Analytics for DEcisions of LEgal cases). This system has been developed as an aid for judges in Italy and Bulgaria to help analyze case law and even predict the outcome of cases. But the system cannot explain how exactly a specific prediction came to be. Rather, users are provided with the following general explanation:

> *"This result was obtained through the use of algorithms based on artificial intelligence and is founded on machine learning conducted on datasets on the outcome of the case in 230 decisions of Italian courts on claims for infringements of rights on trademarks and patents. To train the system, the decisions are reviewed and annotated by legal experts, taking into account the relationships between the parties' requests, claims and arguments, the court's motivation, and the final outcome of the litigation. This prediction of the outcome of the case is not binding and is not intended to replace professional legal advice. Neither the partners in the ADELE consortium nor the European Commission, with whose financial support the ADELE project is implemented, can be held responsible for its provision."*

Specific explanations, on the other hand, aim to explain how a specific output came to be. Such explanations are presented *ex post*, after the system has been applied to a particular case. This has been described by various legal commentators as more useful for those wishing to contest an AI system, as it addresses how the system processes the particularities of an individual's case (de Laat, 2022; Wachter, Mittelstadt, & Russell, 2017). To illustrate this, we can look at the Brazilian INACIA system. This system uses an LLM to select cases for judicial review by assessing whether a case will succeed based on its legal merits. The system categorizes cases into three categories: 'inconclusive', 'not grounded in law', and 'grounded in law' (Pereira et al., 2024). It is also instructed to then explain this decision by providing the specific reasons that support it, which can be seen in the following image:



Explicação: As alegações do autor da representação indicam a existência de possíveis irregularidades no Pregão Eletrônico nº 03/2021 promovido pelo IBAMA. Entre as irregularidades apontadas, estão a não observância das exigências do instrumento convocatório e das leis vigentes da licitação pública, atos praticados por servidores do órgão que podem ser considerados obscuros e que ferem o caráter competitivo do certame, e a habilitação de uma empresa que não apresentou a certidão de falência exigida pelo edital.

Caso essas alegações sejam confirmadas, elas podem indicar a existência de interesse público, conforme definido no Manual de Representações e Denúncias do TCU. Isso porque, se comprovadas, essas irregularidades podem representar um potencial risco de dano ao erário, configurar uma restrição indevida à competição, representar um potencial risco de inexecução ou execução insatisfatória do contrato, configurar uma fraude à licitação ou ao contrato, e representar um potencial risco de ineficácia na execução de política pública.

Admissível: Sim

(a) Brazilian Portuguese version.

Explanation: The allegations made by the author of the representation indicate the existence of possible irregularities in Electronic Auction No. 03/2021 promoted by IBAMA. Among the irregularities highlighted are non-compliance with the requirements of the call for proposals and the current laws governing public bidding, acts carried out by the agency's employees that may be considered obscure and that harm the competitive nature of the event, and the qualification of a company that does not present the bankruptcy certificate required by the notice.

If these allegations are confirmed, they may indicate the existence of public interest, as defined in the TCU Representations and Complaints Manual. This is because, if proven, these irregularities may represent a potential risk of damage to the treasury, constitute an undue restriction on competition, represent a potential risk of non-execution or unsatisfactory execution of the contract, constitute fraud in the bid or the contract, and represent a potential risk of ineffectiveness in the execution of public policy.

Admissible: Yes

(b) English version.

The choice between general and specific explanations raises a normative question as to which type of explanation should be provided to litigants. Should the right to an explanation prioritize general explanations to, for instance, enhance overall trust in the system, or should it mandate specific explanations that enable individuals to contest the particular decision affecting them? In the case of the latter, we will see in the next section that the provision of specific explanations can be technically difficult to achieve in practice.

## 3.2 FAITHFULNESS AND PLAUSIBLE APPROXIMATIONS

An important issue in the field of explainable AI (XAI) concerns the *faithfulness* of specific explanations. Faithfulness, also called correctness, reliability, accuracy (Mohseni et al., 2021), sincerity (Babic & Cohen, 2023), or fidelity (Miró-Nicolau et al., 2024), refers to the extent to which an explanation reveals how the AI system *actually* functions, or whether it merely offers a potential *plausible* explanation (Babic & Cohen, 2023; Rahnama & Hossein, 2025; Stevens & De Smedt, 2024).

The unfaithfulness of explanations arises, in part, because certain types of AI systems can be difficult to interpret for computer scientists who wish to develop an explanation about the system's behavior. For example, data-driven AI systems that rely on complex 'neural networks' can become highly complex and large (Nišević et al., 2024).[6] Moreover, they can behave in ways that are difficult to understand and follow (Selbst & Barocas, 2018). For example, certain self-learning models can rely on patterns and correlations that have been found in datasets that are not intuitive for humans trying to inspect the system. They can also behave widely differently in similar instances, which makes finding a stable logic difficult. Or, they can be continuously self-adapting, making it difficult to discern any stable logic from the system's behavior (Bayamlioglu, 2018; Yeung & Harkens, 2023).

Because directly interpreting such models is not always feasible, various explainability methods have been developed that try to interpret AI systems from the 'outside' without inspecting the system's internal processes. These methods can be characterized as 'model-agnostic' because they do not provide direct insight into the internal workings of the model itself (Molnar et al., 2022). That is to say, these methods try to probe and analyze the model without opening it up. These include, for example, the well-known LIME method (Local Interpretable Model-agnostic Explanations), where one creates a simple model to approximate why a specific portion of an opaque model behaved the way it did. With LIME, one slightly changes input data, checks how the output changes, and highlights the most important features to provide an explanation (Ribeiro et al., 2016).

However, while advances have been made in the field of XAI, the reliability of model-agnostic methods leaves much to be desired. Often, interpretations of a specific region of an AI model might produce locally faithful explanations of how the system operates in that specific instance (Edwards & Veale, 2017). But the logic we uncover in such instances cannot always be generalized to the model as a whole, or in other specific instances (Babic & Cohen, 2023).

Another challenge that might arise has been termed the 'Rashomon effect'. Often, many different plausible, but contradicting, explanations can arise from the same opaque model (Molnar et al., 2022). Such issues show the unfaithfulness of the explanations that computer scientists can currently provide for certain types of AI models. Rudin (2019) therefore argues that 'explanations' are a misnomer when we try to explain opaque models, and rather argues in favor of the term 'approximations'. Consequently, scholars such as Babic and Cohen (2023)

[6] The same can, however, also hold true for simple rule-driven systems, if they rely on excessively large decision trees (Stohl et al., 2016).

describe XAI methods as a 'fool's gold'. They state that these explanations are 'insincere-by-design', and are of little value to those interested in addressing AI-made mistakes.[7]

This raises the normative question as to whether explanations that are merely plausible (i.e., they *might* be true) are sufficient, or whether they should be (fully) faithful (i.e., actually representing the model's true reasoning)? Or, to what extent is contestation negatively affected by unfaithful explanations?

## 3.3 ACCURACY AND CAPABILITY TRADE-OFFS

Arguably, one way to resolve the lack of faithfulness that can arise from model-agnostic explanations is to use models that are 'intrinsically interpretable' in lieu of opaque models. Intrinsically interpretable models can be directly interpreted because of their relatively simple structure and smaller size. Relying on these models, instead of more complex opaque models, makes it more feasible to provide explanations that are faithful (Babic & Cohen, 2023; Rudin, 2019).

Authors such as Rudin (2019) have, for example, argued that judicial profiling systems used for the prediction of recidivism risk should rely on these types of models. She critiques the complexity and opacity of black box models used to profile litigants in the criminal justice system, arguing that in such contexts, we should rather rely on intrinsically interpretable models. This would make it possible for computer scientists to more reliably and faithfully explain how these systems function.[8]

At the same time, the use of these simpler models can sometimes lead to a trade-off in the accuracy and capabilities of the AI system (Bayamlioglu, 2018). This is because certain advanced development techniques cannot be used, as this would again produce a highly complex and opaque model (Molnar et al., 2022). More opaque AI models might be able to ascertain certain patterns in datasets that are not findable for humans, leading them to ascertain correlations that can prove highly useful in case law analysis or in the prediction of litigant behavior. Relying on simpler systems might, in this regard, limit our ability to find such useful patterns (Yeung & Harkens, 2023). Such trade-offs might also occur with Large Language Models (LLMs), as they fundamentally rely on large and complicated neural networks (Liao & Vaughan, 2023). Necessitating the use of interpretable models could then potentially wholly ban the use of LLMs in legal proceedings.

Moreover, the construction of interpretable models might require additional efforts from computer scientists who have to manually construct complicated models and resolve computational problems that would have been resolved by an opaque self-learning system (Rudin, 2019). This can significantly raise development costs on sometimes already overstrained judicial budgets. This raises a normative question as to what values should take priority when we develop judicial AI: explainability and contestability vs. accuracy and costs.[9]

[7] Still, it has to be said that advances are being made in this active field of research, with different types of XAI methods showing different levels of faithfulness (Miró-Nicolau et al., 2024).
[8] We focus in this regard on black-box systems that are not faithfully explainable because of their technical complexity. Nevertheless, it is important to mention that, besides technical limitations, AI models can also be difficult to explain because of organizational and institutional obstacles. See: (Metikoš, 2024; Porębski, 2024).
[9] Nevertheless, such trade-offs do not occur in every instance and can differ between different types of AI and use-cases (Bell et al., 2022). For example, Rudin (2019) has argued that intrinsically interpretable models can be as accurate and capable as their opaque counterparts in predicting the risk for recidivism, although development costs could become higher.

### 3.4 THE RELEVANCE OF THEORIES OF PROCEDURAL JUSTICE

Should explanations be general or specific? Plausible or faithful to the system's reasoning? Should accuracy take precedence over explainability? As stated above, the GDPR and AIA formulate explainability requirements, but they do not concretely answer these questions. A reasoned answer to these questions will, arguably, be intimately linked to the normative goals that litigant contestation in legal procedures is to realize. But in this regard too, the GDPR and AIA remain largely silent. In the following section, we shall therefore explore these questions from the perspective of different normative theories of procedural justice. Additionally, we will illustrate how these theories provide a useful lens to also analyze the variety of arguments and positions found in the literature on the right to an explanation. And lastly, these theories will also help us understand the wider impact of opaque judicial AI on the procedural justice of trials. We will start, in this regard, with the utilitarian theory of procedural justice.

## 4. THE UTILITARIAN (OR LAW AND ECONOMICS) APPROACH

A utilitarian approach to procedural justice holds that the extent to which citizens should be afforded certain procedural rights, such as the right to an explanation, is subject to a cost-benefit analysis. This approach can be classified as *instrumental*: it considers procedural rights to be a means to an end. In this case, their net utility (Meyerson & Mackenzie, 2018). There are, however, various ways to perform such a social cost-benefit analysis (Solum, 2004). For this paper, we look at the 'law and economics' theory of procedural justice.[10] This theory focuses on the monetary costs and benefits of legal procedures. It is based on the assumption that all the relevant costs of adjudication can be expressed in terms of prices (Bone, 2017; Hylton, 2017, pp. 335–337).

From a law and economics perspective, adequate legal procedures are based on a rational trade-off, or 'balancing act', between the economic costs of certain procedures on the one hand, and the economic value they realize on the other by efficiently enforcing substantive law (Bone, 2017, pp. 143–170). Posner (1973) provides one account of this approach. He distinguishes between two types of costs: error costs and direct costs, and argues that the central aim of legal procedures is to keep both these costs as low as possible.[11]

Error costs refer to the societal costs associated with judicial error. For Posner, substantive laws must contribute to economic efficiency. Incorrect judicial decisions, such as a wrongly imposed liability or a wrongful conviction, will reduce that efficiency. Wrong decisions will harm legal certainty and, in the context of civil law, scare off economic activity as the state is unable to provide a stable investment environment (Hylton, 2017, pp. 335–337; 340; Posner, 1973, pp. 399–400). In the context of a criminal procedure, for example, these error costs consist of the societal costs of false convictions (the disutility and the stigma effect of imprisonment) and the expected costs of false acquittals (reducing deterrence, and increasing crime).

The direct costs are the costs to society of having certain legal procedures in place. These include, for example, the time and resources required for such proceedings, such as the salaries of judges and lawyers, the upkeep of the courthouse, the costs of supportive technology, etc. (Posner, 1973). If the direct costs of a particular procedural arrangement exceed the expected *net* value of the increased accuracy of judicial decisions, such an arrangement would not be

[10] See for a different utilitarian or law-and-economic approach: (Kaplow & Shavell, 1994).
[11] As Posner puts it: "The economic goal is thus to minimize the sum of error and direct costs." Posner, (1973, pp. 401).

worth implementing. A particular level of accuracy of adjudication is therefore not a moral right that the parties involved have, but it will depend on this balancing act (Dworkin, 1981, p. 73).

From a 'law and economics' approach to procedural justice, the value of judicial AI lies in the idea that it could reduce direct costs by, for instance, reducing the judicial workload. Indeed, the need to tackle huge backlogs has been put forward as an important reason to use AI in courts (Sourdin, 2021a). In China and Kenya, for example, policymakers and academics have underlined the need to combat the cost of long deliberations by judges (Ogonjo et al., 2021; Shi, 2022; Stern et al., 2021). Besides reducing direct costs, judicial AI could also foster the correctness of legal outcomes and effectuate individuals' substantive rights by better informing judges with additional information, which (arguably) creates more accurate verdicts (Tim Wu, 2019). Conversely, and in the same spirit, exclusively human legal proceedings are regularly evaluated in utilitarian terms too: as mainly time-consuming, costly, and unpredictable (Osztovits, 2021).

At the same time, this (over)emphasis on potential efficiency and quality gains has been critiqued as a blind 'AI faith' (Gentile, 2024). As Donoghue (2017) notes, governments believe 'that digital technologies will provide the 'transformative' panacea for improving efficiency'. Moreover, judicial AI can err; it can contain (discriminatory) biases (Barocas & Selbst, 2016), produce incorrect or non-existing legal references (Merken, 2023), or otherwise reason in ways that are illogical or undesirable (Pasquale & Malgieri, 2024; Surden, 2024). In brief, the use of judicial AI can also seriously increase the risk for error costs.

These error costs may be especially high given the large-scale adoption of judicial AI tools. This technology will, after all, not only be used in a few select cases. Rather, its efficiency benefits are especially noticeable at a larger scale. Consequently, the errors generated by judicial AI tools will affect many litigants simultaneously. The error costs of judicial AI will go beyond the regular risk that a singular erroneous judge may pose to the law. We have to, therefore, consider that there is a heightened need for effective contestation of such tools when they are used *en masse.*

But from a law and economics perspective, the question whether citizens should be granted a contestation-enabling right to an explanation, and in what way, will depend on its instrumental ability to reduce the risk for erroneous verdicts while not prohibitively increasing the direct costs of judiciaries. Different scholars have emphasized the instrumental value of explanations (Kaminski & Urban, 2021; Sarra, 2020; Selbst & Barocas, 2018; Wachter, Mittelstadt, & Russell, 2017).[12] Selbst and Barocas (2018) argue, for example, that the value of explanations lies, in part, in their ability to enable individuals to evaluate decisions. Zarsky (2013) mentions in the same vein that involving outsiders, such as decision-subjects, could lead to 'crowdsourcing' critical insights, lessening the chance that any errors or mistakes in the system have been overlooked by its user or developer.

This holds true in the domain of justice as well. From a law and economics perspective, a lack of explainability could harm economic efficiency, as litigants might be unsure whether AI-assisted verdicts will not be filled with factual and legal inaccuracies that they would be unable to properly contest.[13] One of the questions that arises then would be whether general or specific explanations are more beneficial to contesting errors. As we stated before, general explanations could showcase information such as how the system was tested or trained, or its performance metrics. To a certain extent, a litigant could then contest the use of the system if they deem that the system has too high a risk of producing inaccurate information. But a general explanation would not truly help a litigant to contest how the AI system arrived at a certain conclusion in

[12] This is not to say that these scholars explicitly espouse a law and economics approach to procedural justice.
[13] As we will see, this instrumental focus on the improvement of verdicts is also shared with the Dworkinian approach.

their particular case (de Laat, 2022; Wachter, Mittelstadt, & Floridi, 2017; Wachter, Mittelstadt, & Russell, 2017). Arguably, for litigants trying to address concerns related to their specific situation, general explanations would therefore be less helpful than specific explanations.

Nevertheless, it might be so that substantial costs arise from developing explainable judicial AI systems. Kaminski and Urban (2021) discuss that the costs of making AI contestable 'might make the use of AI unwieldy'. For example, fully disclosing all relevant parameters might be highly costly, as this might harm the business interests of the private developers who have created the system in question (de Laat, 2022; Dor & Coglianese, 2021). IP rights and trade secrets have in the past been raised as an obstacle to the exercise of the right to an explanation in a number of cases (de Laat, 2022; Metikoš & Ausloos, 2025). Therefore, to prevent such obstacles from occurring in the first place, contractual safeguards might need to be negotiated before judiciaries cooperate with external developers (Metikoš, 2024). However, these developers could choose not to work together with judiciaries who put forward such strenuous demands, or they could ask for higher development fees. Alternatively, judiciaries could develop such systems themselves. But this too might require significant investment in hiring and training staff.

The *production* and *provision* of explanations might also raise costs. Zarsky (2013) mentions that writing up, editing, collecting, and disseminating information about the AI system will take resources from any organization implementing transparency measures. A right to an explanation can also call for active research to be done by the developers or users of an AI system, necessitating additional labor costs. In short, the provision of a right to an explanation is contingent on the world's 'finite resources' (Miao, 2022). Considering all of this, explainable judicial AI might prove to be too high a direct cost to be justifiable from a utilitarian point of view.

The question also arises whether providing an explanation is even useful from a utilitarian perspective. After all, explanations can be unfaithful. This severely hinders the usefulness of contestation to mitigate the actual errors present in the system (Babic & Cohen, 2023). If litigants are not provided with the real reasoning of the system, they could, for example, focus on unimportant parameters that barely play a role in a system's assessment (Barocas et al., 2019).

Alternatively, we might therefore use intrinsically interpretable models, which provide more faithful explanations. This improves the usefulness of contestation. However, while this might hold true, the use of such systems is not always desirable. As indicated above, using more complex and opaque models might showcase patterns in datasets that are undetectable for humans who develop interpretable models, who can only process a limited amount of information (Bayamlioglu, 2018; Yeung & Harkens, 2023). Using a faithfully interpretable model might therefore limit the accuracy of the outputs made by the system. Nevertheless, this trade-off does not occur in every instance, and more intrinsically interpretable AI systems can sometimes be as accurate as their more opaque counterparts (Bell et al., 2022; Rudin, 2019). Still, if the trade-off in accuracy is too high, a utilitarian could not justify the use of such systems. Moreover, the higher development cost that could arise with the reliance on interpretable models must also be considered (Rudin, 2019).

Another important point is that, although it might be difficult to contest the errors of non-faithfully explainable judicial AI, there could still be a net social benefit in the use of such systems, because of the potential efficiency gains. The use of more advanced, but opaque, systems could reduce the direct costs of adjudication, as judges spend less time researching and drafting cases. This could, for instance, be the case with profiling systems, which could reduce labor costs. Judges would not need to delve too deeply into the case file of a litigant, but rather receive a handy summary about the litigant in a time-efficient manner. The issue that litigants

would not receive faithful, specific explanations about these systems, inhibiting contestation, could be overlooked if the reduction in direct costs is high enough.

Utilitarians would also consider the difference between high- and low-value disputes in this regard. In more landmark cases that would produce an important legal precedent, the societal stakes are higher to produce an accurate verdict. Meanwhile, more menial court cases that address (allegedly) less significant matters, such as small traffic violations, would be less in need of an explanation. Moreover, utilitarians can also raise the question whether all uses of judicial AI should be explained, including mere case translation or document-filing tools.

Dyoshi-Velez et al. (2019) give a version of this utilitarian argument and state that the higher impact of a decision, the more deserving it is of an explanation. Henin and Le Métayer (2021) similarly argue that contestability cannot be required in any context; rather, one must 'take into account the potential impacts of the decisions on individuals and on society in general'. This provides us with a dual conceptualization of the impact of judicial AI: the impact on the litigant, and the wider impact on society as a whole. From a utilitarian perspective, we can argue that cases with a wider societal impact are more worthy of an explanation, as this will affect more people. Still, Aastrup Munch et al. (2024) argue that even low-stakes decisions can have a large aggregate effect, which must be taken into account when assessing the need for a right to an explanation.

In short, the provision of an explanation must be done in such a way that the risk for increasing direct and error costs is kept as low as possible. Explainability requirements that severely harm the accuracy of an AI system, by, for example, imposing the use of intrinsically interpretable models, might not be justifiable under the utilitarian view. Moreover, this is even the case for erring and opaque AI systems, which cannot be evaluated or contested by those subjected to their outputs, when net social utility arises from their implementation. The striving towards evermore efficiency, therefore, can be justified, to the detriment of procedural rights, such as the right to an explanation.

Hence, the utilitarian view does not address the value of explainability as a *right* that is independent of any cost-benefit balancing exercise. In the next section, we will see how procedural rights can be conceptualized as actual rights that cannot simply be waived away if their costs exceed their benefits for society.

# 5. RIGHTS-BASED APPROACHES

Utilitarians posit that procedural requirements must contribute to a net benefit for society. Strictly speaking, from this perspective, individual litigants are not entitled to procedural *rights*, nor to a certain level of accuracy in verdicts (Dworkin, 1981, p. 73). For instance, a system that would have a substantive chance of erring in low-value cases, but would sufficiently decrease the direct costs of the judiciary, could still be justified under a utilitarian approach. Rights-based theories differ from the utilitarian view in that citizens are guaranteed genuine procedural rights, regardless of their net consequences for social welfare. These theories can be divided into outcome-based and process-based theories. The first is premised on the idea that procedural rights have as their main objective to secure the accuracy of the legal decision. As such, they offer an instrumental approach to procedural justice (Meyerson & Mackenzie, 2018). The process only matters to the extent that it contributes to an accurate outcome.

By contrast, process-based theories argue that certain procedural arrangements have an intrinsic moral value. These theories emphasize that harm can also be done by disregarding certain procedural values, even if the outcome of the legal procedure is accurate (Summers, 1974). In the next two sections, we look at a specific outcome- and a process-based theory: the Dworkinian and dignitarian approaches to procedural justice.

## 5.1 THE DWORKINIAN OUTCOME-BASED APPROACH

One version of an outcome-based theory of procedural justice can be found in the work of Ronald Dworkin.[14] Dworkin holds that citizens have a *moral* right to have their substantial rights upheld in court and, consequently, also have a moral right to a correct judicial decision, regardless of the societal consequences. As he puts it: 'someone is entitled to win a lawsuit if the law is on his side, even if society overall loses thereby' (Dworkin, 1981, p. 94).

By contrast, an incorrect outcome would cause injustice to the litigating parties and result in *moral* harm. This *moral* harm, that stems from an incorrect judicial decision, is to be distinguished from the *bare* harm that the parties involved experience as a matter of subjective empirical fact. Whereas, from a utilitarian perspective, the bare harm—such as the suffering, outrage, or resentment experienced by a wrongly convicted defendant—can be weighed against the net societal benefit of more efficient legal procedures, this kind of balancing is not permitted within a Dworkinian framework (Dworkin, 1981, p. 81). Citizens who are involved in a legal procedure should thus be guaranteed procedural rights to the extent that these rights are necessary for arriving at correct outcomes.

Nevertheless, Dworkin does not defend the position 'that an individual has a right to the most accurate procedures possible' (Dworkin, 1981, p. 73). Legal procedures do not have a categorical priority over other moral and societal concerns. According to Dworkin, we can and should not maximally put society's resources into funding the judiciary. Dworkin leaves the decision of how important it is to prevent certain moral harm to democratic decision-making procedures. Not because these decisions are necessarily right, but because this is a fair way to decide matters about which there can be reasonable disagreement (Dworkin, 1981, pp. 78–79).[15]

From a Dworkinian point of view, the use of judicial AI can be quite laudable, as it has been ascribed the potential of strengthening judicial decision-making, not only in its efficiency but also in its quality (Tim Wu, 2019). AI could improve verdicts by providing additional insights, as well as more objective and uniform analyses of the law (Re & Solow-Niederman, 2019). Still, these systems are not infallible, and they can, of course, err. In such instances, the moral harm that arises must be contestable for litigants. Or in other words, when judicial AI errs and leads to the wrong application of the law or the reliance on false facts, an explanation must enable a litigant to find and contest those errors.

But how then would the Dworkinian approach answer the questions we raised in section 2? Arguably, this approach focuses on the ability of explanations to ensure the accurate application of the law in a given case by enabling litigants to contest any AI-made errors. To that end, specific explanations would be preferred, rather than general explanations about the system's average functioning. This is because a Dworkinian theory of procedural justice is concerned with the moral harm that may occur in the *individual* case of a litigant, as it primarily aims to uphold the substantive rights of individual litigants.

Nevertheless, the question arises whether litigants (or their lawyers) are indeed likely to improve the verdict, when given an explanation. Arguably, because of the potential lack of technical skills of the average litigant and their lawyers, it is questionable to what extent litigants will indeed be able to contest the system effectively and improve the accuracy of the

[14] The most straightforward exposition of his theory of procedural justice can be found in his essay *Principle, Policy, Procedure,* as published in *A Matter of Principle*. A discussion of procedural rights can also be found in Part V of *Justice for Hedgehogs*. See: R. Dworkin (2013). *Justice for Hedgehogs,* The Belknap Press.

[15] This holds true at least to the extent that these procedures honor citizens' right to equal concern and respect, which means that the risk of moral harm must be equalized over the entire population (Dworkin, 1981, pp. 78).

verdict. Scholars have, consequently, argued that the right to an explanation relies on a 'transparency fallacy' (Edwards & Veale, 2017). This argument raises the point that most litigants are not highly skilled computer scientists, nor are the lawyers representing them likely to be well able to audit an AI system for any bias or mistakes (Ananny & Crawford, 2018; Binns, 2022; Edwards & Veale, 2017). Therefore, using explanations as a tool to address technically complex issues found within these systems might be only possible for computer scientists, and not for lay people (Selbst & Barocas, 2018).

Moreover, from a Dworkinian perspective, we can question how useful specific explanations of opaque models can be if these explanations do not faithfully showcase the real reasoning of the judicial AI system (Babic & Cohen, 2023). If litigants cannot assess the actual logic behind a particular output, but were only presented with a plausible guess, they might be unable to contest any of the actual errors present.

In this regard, intrinsically interpretable models, whose small scale and simplicity would make them more faithfully explainable to an individual litigant, could be used instead of complex and large AI models (Rudin, 2019). Still, as discussed before, there might be a trade-off between accuracy and capability[16] when an intrinsically interpretable AI model is used instead of a more complex model (Bell et al., 2022). If such a trade-off were indeed to substantively increase the risk of moral harm, it would not be justifiable from a Dworkinian approach.[17]

In short, from a Dworkinian, outcome-based theory of procedural justice, specific explanations are preferred, but their usefulness can be critiqued on the basis of the transparency fallacy, the issue of faithfulness, and the potential trade-off in accuracy that arises from the reliance on intrinsically interpretable systems.

One counterargument that can be raised against both the utilitarian and Dworkinian approaches, however, is that they do not take into account the non-instrumental value of procedural justice. A process-based approach to a right to an explanation emphasizes the value of having a contestable legal procedure *on its own merits*, which we will discuss in the next section.

## 5.2 THE DIGNITARIAN PROCESS-BASED APPROACH

Contrary to Dworkin's outcome-focused approach, a process-based approach to procedural justice stresses the intrinsic value of certain procedural rights, separate from their instrumental value. Process-based theories of procedural justice predominantly focus on the value of *dignity* that is honored by granting citizens certain procedural rights (Meyerson & Mackenzie, 2018; Summers, 1974). These dignitarian theories assert that citizens subjected to the law should always be treated with the respect that is due to all moral agents, understood here in a Kantian sense as free and rational agents. Waldron argues, in this regard, that certain traditional participatory procedural rights are justified by the idea that a citizen who is involved in a legal procedure is not a mere object or thing to be decided upon. Rather, a litigant is an autonomous and responsible person. They should be informed and consulted about how relevant facts are to be determined by a judge and how the law is to be applied to their case (Waldron, 2011).

From this perspective, the increased role of AI in legal decision-making has been critiqued as it can harm the dignity of individuals. For instance, profiling systems have been critiqued with the argument that they reduce individuals to mere data points, dehumanizing them (Dao, 2020; Hildebrandt, 2019; Yeung & Harkens, 2023). Smuha (2021) argues that 'all individuals have an

[16] We refer here back to section 3.3., where we discussed that certain types of models, such as LLMs, could not be used when faithfulness requirements are imposed. Consequently, there is a restriction as to what capabilities a judicial AI system may have.
[17] However, to reiterate, this trade-off does not occur in every instance. See: (Bell et al., 2022; Rudin, 2019)

inherent dignity and merit being treated with respect for their own multifaceted individuality'. She, accordingly, warns about the loss of meaning that can occur when data-driven AI systems engage in processes of 'data categorization'. This process becomes all the more problematic when such systems are opaque, as litigants cannot contest the profiles that have been made about them, or partake in legal procedures that affect their lives (Dao, 2020; Kaminski & Urban, 2021; Selbst & Barocas, 2018).

Moreover, the fact that AI is involved in decision-making in ways that decision-subjects cannot make sense of, can be considered a standalone type of *hermeneutic harm*. This harm occurs regardless of whether the system in question errs (Rebera et al., 2025). This is because individuals deserve, as dignified beings, to be able to understand and contest how impactful decisions are being taken about them.

From a dignitarian perspective, procedural rights can function as a crucial counterweight to opaque judicial AI systems, offering a 'forum for the expression of human dignity' (Dao, 2020). Kaminski and Urban (2021), for instance, argue that 'affording a right to contest affords a form of respect to individual people in the system. It permits participation. It establishes agency.' In this vein, a right to an explanation could be seen as an indispensable epistemic good that allows individuals to take part in decision-making that seriously affects their lives (Selbst & Barocas, 2018).

Considering all of this, how would the questions we raised in section 2 be answered from a dignitarian approach to procedural justice? Compared to the previous two theories, at first sight, a dignitarian approach seems to offer less clear guidance in this regard. Rueda et al. (2025), for instance, warn of the vagueness of dignity as a concept within AI ethics and the difficulty of applying it in practice. Selbst and Barocas (2018) similarly argue that operationalizing the dignitarian view on the right to an explanation is difficult in practice. After all, what does it concretely mean to provide an explanation that respects the dignity of an individual?

In their account of procedural justice, dignitarians' core focus is on the rights of litigants to deliberate and participate in the legal proceeding before them. Miao (2022), Jongepier and Keymolen (2022), and Zerilli et al. (2019) provide an analysis of the role of explanations in enabling individuals to engage in decision-making deliberations.[18] Explanations of AI should serve to enable a person's rational autonomy and defend their deliberative agency by providing the individual with the ability to critically reflect on the system in question (Jongepier & Keymolen, 2022; Miao, 2022). The kind of explanation warranted from the dignitarian view, therefore, should provide the right epistemic conditions for an individual to participate in the proceedings before them, by making it possible to contest the reasons based on which the judge uses the AI system's output.

[18] Miao takes on an explicitly dignitarian account, but the accounts of Jongepier and Keymolen and Zerilli et al. do not, explicitly, have such a foundation. However, Jongepier and Keymolen do rely on the concept of a 'deliberative agent', which they define as a 'rational being', which is similar to the participatory and deliberative focus of the dignitarian approach. Zerilli et al. also rely on a rational-oriented conceptualization of human decision-subjects. They discuss how humans engage in practical reason and explanation-giving, with particular regard to the 'intentional stance', as developed by Dennett (1998). Their paper therefore also relies on the view that individuals are rational beings that engage in rational and reasoned debates. In this regard, Jongepier and Keymolen and Zerilli et al. share in common with the dignitarian notion, the normative goal that explanations should serve to enable the rational participation of decision-subjects in decision-making debates. Therefore, without arguing that these authors are themselves dignitarians, we do nevertheless argue that we can apply their views as a method to extrapolate how one can participate in dignitarian rational debates that occur during a trial.

To enable litigants to participate and deliberate, specific explanations would be preferable over general explanations. The former would better allow a litigant to engage in the legal debate that occurs within their own case. Providing general information would not be conducive to fostering participation within the current proceedings at hand, but would only help the litigant understand the general functioning of the system *in abstracto*. This would not foster a practice of contestation that would be conducive to participation and deliberation in their current trial.

But would the dignity of a litigant be harmed by giving specific explanations that are unfaithful? Full faithfulness may be excessive for dignitarian aims, as meaningful participation, not causal accuracy, is the priority. Zerilli et al. (2019) argue that such transparency requirements impose a double standard, since judges' true motivations—often shaped by upbringing or mood—are not explicitly reflected either in their reasoning. Litigating parties primarily contest the facts and reasons as they are provided in legal judgments, but not necessarily the internal causes of such judgements. Miao (2022) argues similarly with the help of a dignitarian autonomy-based test that explanations must provide the same level of information about an AI system that a person would also receive if a decision had been fully made by a human.

Based on the argument that transparency standards should apply similarly to humans and machines alike, Zerilli et al. propose that model-agnostic explanations of specific sections of a model are sufficient for the kind of reasoning and debate that most individuals engage in, including during a legal process. As we stated before, such explanations are not consistently, faithfully, and reliably representative of the actual behavior of the system, because we cannot reliably inspect an opaque system's internal behavior (Babic & Cohen, 2023; Burrell, 2016). But we also do not have direct access to the internal psychological processes of humans. This is, however, not necessarily a problem. Human judges provide explicit reasons to underpin their verdict, which in turn can be contested by a litigant. Zerilli et al. (2019) specifically focus in this regard on the judicial standard, which they describe as 'the most procedurally, evidentially, statutorily, and precedentially constrained form of official reasoning that exists'. In this judicial context, the provision of reasons, not internal motivations or brain inspections, sufficiently ensures the participation of a litigant. Consequently, from a dignitarian perspective, this same standard should also apply to AI-assisted proceedings. The relevant focus would then lie on the reasons the judge provides for relying on an AI tool. From this point of view, the risk that an explanation might not be faithful and unlikely to showcase any of the actual errors present in the system is not necessarily an issue. Therefore, from a dignitarian point of view, the risk that an explanation might not be faithful and unlikely to showcase any of the actual errors present in the system is not necessarily an issue.[19] What matters for participation is not transparency into the system's internal operations, but whether the justification offered by the judge enables meaningful engagement with the decision.[20]

Contrarily, instrumentalist approaches, such as the utilitarian and Dworkinian theories of procedural justice, imply that unfaithful explanations are not a reliable basis to find errors in an AI system. But for dignitarians, it is not their principal aim to improve the accuracy of the verdict with procedural rights. Hence, explanations that allow the litigant to understand the reasons why a judicial AI system produced a certain output would be valuable on their own terms, as the litigant would be better able to participate in a legal procedure. Moreover, human judges too can be unfaithful, agnostic or even insincere about their true motives for arriving at a particular

[19] This is not to say that scholars such as Miao (2022) Jongepier and Keymolen (2022) and Zerilli et al. (2019) would fully disregard faithfulness as a standard for explainable judicial AI. Rather, we argue that a solely process-based focus on procedural justice, does not take seriously enough the need for procedural rights to enhance correct outcomes.

[20] See also (Hildebrandt, 2019), who discusses the need for justifications rather than (overly technical) explanations.

judgment. Still, we debate about the reasons that they do put forward, providing a basis for communication, debate, and contestation.

This is, of course, not to say that judges, judiciaries, or the computer scientist who support them, may simply invent an explanation out of thin air.[21] Rather, a plausible approximation of a model, which gives the litigant reasons that they can discuss and deliberate on, suffices to enable rational debate. Moreover, the judge operating an AI system that is not faithfully explainable will themselves also not be able to receive a faithful explanation about that system. Still, the judge too deliberates on the trustworthiness and accuracy of the system on the basis of the information that they do have. In this regard, a dignitarian approach wishes to open up the judge's deliberative process to the litigant, but not necessarily ensure that these explanations are (fully) faithful to the actual behavior of the AI system.[22]

Instrumentalists, such as utilitarians or Dworkinians, would likely critique this view. They could raise that unfaithful explanations achieve nothing. No one in this process is effectively auditing the system, and making sure that it does not produce mistake after mistake in the future. Indeed, a valid criticism of the dignitarian approach is that it could engage in a form of *procedural fetishism* (Bagley, 2019; Kaminski & Urban, 2021; Zalnieriute, 2021). Arguably, a right to an explanation, then, does not only throw away valuable time and resources, but could also serve to legitimize risk-prone AI systems that should not be legitimized (Peters & Visser, 2023). In this regard, a dignitarian right to an explanation can also risk to legitimize problematic kinds of government action that rely on erring judicial AI (de Fine Licht & de Fine Licht, 2020). Think of discriminatory profiling systems, which could be (unfaithfully) explained to have no bias, all the while ethnically profiling litigants. Kaminski and Urban (2021) argue against this instrumentalist critique, however, and state that procedural rights alone are not a panacea, but that does not mean that they are not valuable in their own right.

In addition to this instrumentalist counterargument, a process-based argument against dignitarians can also be raised. Dignitarians can be critiqued on the basis that they conceptualize litigants in terms of rational and autonomous agents. This abstraction ignores the emotional and relational needs of actual litigants and their practical ability to contest judicial AI tools. In the next section we showcase how the relational approach to procedural justice departs from these dignitarians' abstractions, and aims to provide explanations that give litigants 'voice' in AI-assisted proceedings.

# 6. THE RELATIONAL APPROACH

Finally, an emerging strand within process-based approaches to procedural justice focuses on the normative significance of the relational aspect of institutional arrangements (Meyerson et al., 2021). Similar to a dignitarian approach, a relational approach emphasizes the intrinsic and non-instrumental value of procedures. Both approaches see procedures and outcomes as distinct sides of justice (Ceva, 2020). However, a relational approach to procedural justice, advocated by Meyerson and Mackenzie (2018), differs from a dignitarian approach in that it does not take the rational and autonomous agent as its normative starting point (Meyerson, 2020). Rather, it holds that humans are dependent creatures whose individual identities and sense of self-worth are largely constituted by their interactions with others, including

[21] We do not regard the dignitarian point of view as a form of 'pure' procedural justice, as described by Rawls (1999). Rather, we contend that dignitarian procedural rights aim to provide a basis for reasoned debate. That legal procedure does not produce by definition a final verdict that is always fully correct, however.
[22] Under this view, the right to an explanation would be quite similar to the right to adversarial proceedings as defined under the right to a fair trial of art. 6 ECHR, which necessitates the disclosure of all documents in the hands of the court. See also the case of: (Ruiz-Mateos v. Spain, 1993, para. 63). This levels the epistemic playing field to all parties involved in a trial.

interactions with public institutions, such as administrative agencies, the police, and courts.[23] A relational approach therefore also pays attention to the quality of these interactions, and to the *manner* in which citizens are treated. Whether, for instance, these interactions can be seen as dominating, disrespectful, or demeaning, with dire consequences for the sense of self-respect of the recipients of such treatment.

In their focus on the moral importance of the quality of concrete interactions within procedures, relational approaches to procedural justice find support in a large body of empirical research on *perceived* procedural justice (Lind & Tyler, 1988; Tyler & Lind, 1992). This research indicates that citizens value the fairness of a procedure independently of the actual outcome. One influential explanation for citizens' interest in the fairness of procedures, is that the way they are treated by persons with authority conveys citizens with significant information about the extent to which they are valued as members of society (Lind & Tyler, 1988; Meyerson et al., 2021; Tyler, 1988; Tyler & Blader, 2000; Tyler & Lind, 1992). A fair procedure communicates that they are valued as a member of the group to which the procedure applies. An unfair procedure suggests that one is not respected or cared about as a group member.

Supported and informed by this empirical perspective, a relational approach more specifically holds that legal procedures that are characterized by impartiality, trustworthiness, respect, and the opportunity to express one's voice, will foster litigants' sense of self-worth and self-respect. These criteria, in the words of Mackenzie (2020, p. 206), 'embody citizens' expectations of normative entitlement to be treated as social and moral equals'. Relational procedural justice theory hold that procedures that adhere to these values, will not only be *experienced* as legitimate, but are also in *themselves* normatively legitimate (Mackenzie, 2020; Meyerson et al., 2021).

From a relational perspective on procedural justice, a key concern about judicial AI is how the use of such systems may affect whether citizens are treated as equal members of society. Numerous examples illustrate how the use of opaque and inscrutable AI by public institutions can dominate, mistreat, and discriminate against groups such as immigrants, ethnic minorities, and the socially and economically disadvantaged. Examples include the Childcare Benefit scandal, the Post Office scandal, and the Robodebt scandal (Eubanks, 2017; Oldenhof et al., 2024; Van Domselaar, 2025). These scandals also teach us that seemingly easy-to-automate cases can lead to large-scale legal and ethical abuses against socially and economically vulnerable citizens.[24]

Moreover, from this same relational perspective, a more procedurally-focused concern also arises in regard to the advent of opaque AI in legal decision-making. Specifically, a lack of respect, equality, and care can be conveyed to litigants when inscrutable AI is used in legal procedures. Naudts (2024, p. 5) describes this impact as *powerlessness*. He states: 'In determining how and through which tools people are judged, ranked, classified, and categorized, (a select few) actors can establish the conditions that define the losers and victors of the digital society. Perhaps then, when viewed from this perspective, the digital society might have generated a novel category of powerless individuals, i.e., those without any authority or autonomy in the society's increased datafication.'

To address the concerns of systemic unequal treatment, powerlessness and mass-scale miscarriages of justice, of course, a right to an explanation of judicial AI will not suffice on its own.[25] Yet, fostering contestation in individual procedures could still serve procedural justice

[23] For this point Mackenzie draws on the insights from relational egalitarians such as Elisabeth Anderson (2008).
[24] This also forms a counterargument to the utilitarian view that only important or impactful cases are worth explaining.
[25] Naudts (2024) further addresses the limitations of an individual-oriented relational perspective on AI and suggest that we should also take into account collective action as a tool for remedying oppression.

(Naudts & Vedder, 2025), specifically by offering litigants *voice.* This is defined by Meyerson and Mackenzie (2018) as 'the willingness of authorities to listen attentively to a person's views, even when there is no chance of influencing the outcome'. Explanations could enable voice, in this sense, by making procedures more understandable and contestable for litigants. In turn, this enables them to express their views on the use of judicial AI, regardless of whether this serves instrumental values such as error correction (as espoused by the utilitarian and outcome-based approaches to procedural justice). Indeed, empirical research shows that individuals perceive procedures as more procedurally just, when explainability and contestability are safeguarded in algorithmic decision-making (Yurrita et al., 2023).

Considering all of this, how then would the questions we raised in section 2 be addressed from a relational perspective? To this end, we must apply the previously discussed relational elements of impartiality, trustworthiness, respect, and voice (Meyerson & Mackenzie, 2018). First, explanations must make it possible to have procedures that enable 'voice', i.e. litigants must be able to contest the system in a way that enables them to effectively express their opinions and concerns. Secondly, explanations must show an AI system's neutrality (in the sense of impartiality). Thirdly, explanations must show the trustworthiness of the usage of AI, and that it will not harm the litigant's rights and overall well-being. And lastly, explanations must also enable procedures that impart respect for the litigant as an equal member of society.

If we apply these interrelated elements to the right to an explanation, we could argue that explanations must be, most importantly, *understandable* to the litigant. Otherwise, litigants will not be able to participate, express their 'voice', and trust the usage of the judicial AI system. To illustrate, more down-to-earth questions such as 'how does the system apply social benefit rules to my specific situation' might be addressed by explaining that, in a general sense, *inputs* will be broken down into *tokens*, and processed through the massive *neural network* of a *Large Language Model*. A litigant who has concrete and particular concerns about their case, but knows nothing about LLM-based AI systems, will not be able to engage with such general, abstract, and distantiated explanations. This could alienate and confuse a litigant, limiting their voice. Moreover, this could also harm the respect they are due as an equal member of a (digitized) society, as their opinion is not worth being listened to (as opposed to that of, for example, someone more digitally literate).

Understandability is especially a salient point to raise in regard to the contestation of judicial AI, as not everyone has the required epistemic-, financial-, or time-based capital to engage with a digitized government (Ranchordas, 2021). This also ties into the more general discussion on how courts communicate to citizens, emphasizing the need to use more 'plain' language (van Domselaar, 2022a). Arguably, these factors are not properly taken into account by the dignitarian notion of the ideal 'rational' litigant (Meyerson & Mackenzie, 2018), which does not put forward such practical, and context specific, requirements for explainability rights.

How then would we answer the questions we raised in section 2 from this point of view? From a relational perspective, we could first argue that general explanations would not be very helpful to address litigants' informational needs. General explanations, namely, do not address the particularities of the case at hand, and can only provide information about how an AI system functions globally. This does not illuminate to the litigant how they can participate and express themselves in their own trial. Specific explanations, on the other hand, tackle more concretely how an AI system assessed particular aspects of a litigant's case, and are therefore less abstract and distant.[26] They give more concrete and case-specific guidance to the litigant, so that they can better express their 'voice' and understand what is going on in their trial.

[26] Nevertheless, general explanations, if understandably formulated, can still showcase how well-audited and reliable a system is. This, in turn, can fulfill the relational requirement that explanations must showcase

Nevertheless, as outlined above, explanations are not always faithful to how a system actually works. Would this not harm the relational value of trustworthiness? Meyerson and Mackenzie (2018) state that conveying trustworthiness does not necessarily entail '*honesty*'. Rather, it requires 'that people feel that authorities care about them and are motivated to be fair to them'. From this point of view, explanations would therefore not need to be faithful, as long as they are compatible with the aforementioned values of relational procedural justice: trustworthiness, neutrality, voice, and respect. This can be done, for example, by showcasing that the system has been checked, audited, and verified for any errors. Moreover, as faithfulness can be disregarded as an obstacle for relational procedural justice, the same can be said about the need to use interpretable AI models.[27]

However, as we saw under an dignitarian approach, an exclusive focus on the intrinsic values of procedural arrangements might come at the cost of valuing the importance of right outcomes. Or, in other words, it may lead to procedural fetishism (Bagley, 2019; Zalnieriute, 2021). The relational focus on *perceived* procedural justice can also lead to a manipulative use of fair process cues, arguably motivated by the assumption that the legitimacy of legal authorities must be enhanced, independently of the moral quality of their functioning (MacCoun, 2005). For instance, in the context of the use of judicial AI, a litigant can be encouraged to voice their concern about an AI system, and as a consequence feel respected through the provision of a 'comprehensible' and 'citizen-friendly' explanation (Van Domselaar, 2022a), while the explanation itself is not faithful to the actual behavior of the system, including when that system is highly problematic or simply unjust.

# 7. EFFECTIVE CONTESTATION AND AN AI-RESPONSIVE THEORY OF PROCEDURAL JUSTICE

In the previous sections, we examined how different theories of procedural justice understand the value of litigant contestation to help substantiate the right to an explanation of judicial AI. However, these theories have been developed prior to the advent of AI in the administration of justice. Hence, one potential objection against applying these procedural justice theories to the context of AI-assisted judicial decision-making is that this may insufficiently address the *specific* issues that the use of judicial AI raises. Procedural justice theories also need to be critically examined for their suitability in providing a normative foundation for the effective contestation of judicial AI.

Drawing on the foregoing discussion, we will therefore now briefly sketch the building blocks of an *AI-responsive* theory of procedural justice and how such a theory would answer the questions that we have addressed in this paper. We shall first reiterate some of the specific concerns on the adoption of judicial AI that we have discussed from a procedural justice lens in the previous sections. These are: access to justice, the risk of mass-scale error, dehumanization, and the risk of systemic unequal treatment.

First, an AI-responsive theory of procedural justice cannot ignore the fact that the growing use of judicial AI is intimately linked to the 'sorry-state' of contemporary justice systems (Genn, 2010, p. 51), particularly with regard to access to justice (Van Domselaar, 2022b). Through the

neutrality, trustworthiness and a sense of care for the litigant's wellbeing. General explanations could therefore still have a place in the explainability toolbox, from a relational perspective, although they cannot fulfill the requirements of enabling voice.

[27] In this regard, the relational approach also does not offer strong guiderails as to how to address the question on how to balance cost and accuracy trade-offs in the debate on intrinsically interpretable systems, as these values are not addressed by its focus on perceived procedural justice.

automation of ostensibly 'simple' judicial tasks, in a standardized manner, advocates of digital justice argue that adjudication can be done '*cheaper, better, quicker*' (Susskind, 2019, p. 48).

As noted above, one of the strengths of a utilitarian perspective is its ability to respond to the importance of scalability and efficiency by focusing on the net utility generated by legal procedures. Specifically, it can recognize the fact that the direct costs of certain procedural safeguards might be in tension with the need to reduce the current justice gap that legal orders face. Although we do not endorse a utilitarian approach, for reasons stated below, it nonetheless serves as a reminder that reducing this justice gap should remain a concern in any discussion that concerns the moral quality of legal procedures (Van Domselaar, 2022b).

However, we have also seen that the increased scalability enabled by judicial AI is a double-edged sword. While it may enhance access to justice, it also carries the risk of mass-scale error, as explained through our law and economics analysis of judicial AI. The ability to contest these errors is therefore crucial, not just for individual litigants but also for society as a whole.

Moreover, and as discussed in the context of the dignitarian approach, the adoption of judicial AI tools poses an inherent risk of dehumanizing individual litigants. The commensuration that occurs with the usage of profiling systems, for instance, strips litigants of their multifaceted and complex identities. The adoption of *inscrutable* and *opaque* judicial AI might, in addition, have an especially exacerbated dehumanizing effect. In such instances, litigants would be unable to assert their rational agency and defend their unique identities from the assessments of judicial AI (Miao, 2022). Procedural rights that enable litigants to contest these systems form an important counterweight in this regard. They offer, as Dao succinctly puts it; a *'forum for the expression of human dignity'* (Dao, 2020, p. 30).

Lastly, the increasing use of judicial AI is also intimately linked to the risk of systemic unequal treatment. That is to say, the brunt of mass-scale error and dehumanization often falls on particular social groups such as immigrants, ethnic minorities, and the socially and economically disadvantaged. For instance, in the section on the relational approach, we saw that large-scale AI-driven miscarriages of justice have so far affected these groups the most. In part, this is a consequence of the bias that AI systems can perpetuate.[28] Moreover, we observed the '*powerlessness'* of these affected individuals. Arguably, only a select group of judges, policymakers, and developers, will decide on who will be judged, ranked, classified and categorized by judicial AI and how. Those who will be affected most will in many cases not have any power to contest such systems (Naudts, 2024).

We argue that an AI responsive approach to procedural justice should take these concerns into account. Accordingly, such a theory should substantiate the right to an explanation in such a way that litigants can enjoy access to justice. To that end, they must be able to contest mass-scale errors, as well as their dehumanization and systemic unequal treatment.

However, how to safeguard contestation *effectively* is a recurring concern that we saw in this paper. In the sections on the law and economics and Dworkinian approaches, we already raised the question of whether litigants are sufficiently equipped to contest AI-made errors. In our discussions of dignitarian and relational procedural justice, we sidestepped this issue. We showed that a (solely) process-based approach to procedural justice emphasizes the intrinsic value of participation. But such an approach does not necessarily take into account the real-life ability of litigants to effectively contest AI-made errors.[29] This focus on process could then justify

[28] For a more extensive overview of how discrimination and unfair treatment can arise in Machine Learning-based AI systems, see: (Barocas, 2023; Solon Barocas & Andrew D Selbst, 2016).

[29] Nevertheless, we shall argue later in this section that the relational approach to procedural justice does give a normative basis for a practical and situated-account of the real-life litigant. And, consequently, we shall argue

unfaithful explanations, leaving mass-scale errors uncontested and hindering access to justice. Also, a solely process-based approach does not provide a rebuttal to the argument that providing an explanation to a litigant might be a useless endeavor, as they would be unlikely to be able to contest the errors of judicial AI tools (Edwards & Veale, 2017).

If we are to really take seriously the concerns of access to justice, mass-scale errors, dehumanization and systemic unequal treatment, we need to be attuned to the practical ability of litigants to contest judicial AI. After all, not all litigants have the epistemic-, financial-, or time-based capital to contest judicial AI. In this regard, it is important to note that a variety of *systemic* factors, (poverty, racism, sexism, etc.) (Rhode, 2001; Marchiori, 2016) may play a crucial role too in diminishing access to such capital, and in turn, access to justice. Such systemic unequal treatment may be especially worsened by the kinds of cases in which judicial AI will play a role. Will it be primarily applied in social welfare disputes, burdening poorer litigants with contesting any potential errors? Or will large companies dealing with fraud investigations have to bear the burden of contestation? The latter would, arguably, be more (financially) able to hire swaths of lawyers and AI-experts to assist in such endeavors.[30]

In our opinion, the theory that is most attuned to address the practical ability of litigants to effectively contest judicial AI, as well as these normative concerns, is the relational approach to procedural justice. It argues that courts should treat citizens as social and moral equals, considering it crucial to procedural justice (Mackenzie, 2020; Meyerson et al., 2021). Relational procedural justice theory implies that we should move beyond the dignitarian concept of the ideal, abstract and rational litigant who is able to partake in discussions before a judge (Mackenzie, 2020). It takes into account that litigants are unequal in their capacities to effectively participate in an AI-assisted trial. Consequently, and not surprisingly, relational theory has become increasingly linked to egalitarianism in discussions of digitization as well (Naudts, 2024). The relational emphasis on equality and the non-ideal litigant addresses the question of whether litigants will have equal access to justice when judicial AI is used. Moreover, it shows a care for the individuals who will be most affected by judicial AI systems, and who are also most in need of effective contestation tools to address both mass-scale errors and their dehumanization.

At the same time, we showed that the relational emphasis on *perceived* procedural justice and on the intrinsic value of procedures might lead to a situation of procedural fetishism. A litigant could be encouraged to voice their concern and receive a 'citizen-friendly' explanation, which, in truth, is wholly unfaithful to the actual internal behavior of the system. In turn, they would be unable to actually contest the true parameters that drove the system to a certain output. For a relational theory to be responsive to the specific concerns that judicial AI raises, it needs to provide a normative basis for a more effective form of contestation.

To that end, we should note that litigants also deserve an effective chance to contest any AI-made errors to ensure the correctness of the verdict. To make litigant contestation therefore *useful*, the justice system has an obligation to promote litigants' capabilities to contest any

that relational procedural justice theory can provide a normative basis for effective litigant contestation of judicial AI.

[30] It is salient to note that AI is indeed often used in cases that affect those who do not posses over the required capital to enjoy access to justice. One example of this trend is the increased usage of AI in processing migration cases. See: (Palmiotto, 2024). Outside the scope of this paper, it should also be noted that the usage of AI by lawyers also affects the quality of support that immigrants receive. For instance, in a case from the U.K., a barrister had used Microsoft CoPilot, which had hallucinated non-existent legal references in an asylum case. See: ANPV and SAPV v. Secretary of State for the Home Department, Case No. US-2025-00373 (Upper Tribunal, Immigration and Asylum Chamber); In a separate case concerning an appeal against an asylum refusal, again non-existent references had been provided and subsequently removed. See: MS (Bangladesh) v. Secretary of State for the Home Department [2025] UKUT 00305 (IA).

wrongdoing on the part of the judicial AI tool. To that end, a right to an explanation needs to take into account the relational focus on the non-ideal litigant, and their real-life ability to contest judicial AI effectively.

How then would this affect how we answer the questions we raised in section 2? Arguably, we first of all need explanations that showcase how one's case is specifically influenced by a judicial AI system, as this would better engage the litigant in their trial. Moreover, such an explanation also needs to be faithful. As we discussed previously, acts of contestation would otherwise not be effective to mitigate AI-made errors. Explanations could become mere procedural sedatives (Metikoš, 2025b) that might enhance the perception of procedural justice, while allowing substantive injustices to continue. But to ensure faithfulness, there might be a need to rely on intrinsically interpretable systems where (sometimes) a trade-off can occur between the faithfulness of explanations and the accuracy of a system. Under purely outcome-based approaches, such a trade-off would not be normatively acceptable. The main goal of procedural rights under these theories would be to safeguard the accuracy of the verdict above all.

A process-based perspective, as the one relationalists ascribe to, could nonetheless justify such a trade-off. Litigants are due certain procedural rights, independently of their influence on the accuracy of the outcome. However, our emphasis on the importance of procedures for arriving at correct outcomes within a relational approach to procedural justice adds another perspective. Namely, that a litigant who is *supported* in their ability to contest judicial AI systems, can provide new insights and improve the quality of judicial decisions. By ensuring effective contestation of judicial AI, accuracy could, ostensibly, be better safeguarded than if the litigant had given no input into their trial at all.

Importantly, such explanations are only the first step to addressing the concerns of access to justice, the risk of mass-scale error, dehumanization, and systemic unequal treatment. Individual litigants will not always be able to address these injustices effectively.[31] A discussion on the right to an explanation requires us to therefore also consider when and where we use judicial AI, as well as who has the right capital to contest such systems. We only provided here but the first building blocks for a theory of procedural justice that is more attuned to these concerns of judicial AI. We have made clear that *effective* contestation is one crucial aspect of such an approach. And, that this concept of effectiveness must take into account the non-ideal litigant, who may or may not have access to the required capital to contest judicial AI.

# 8. CONCLUSION

The right to an explanation and its role in enabling the contestation of judicial AI raise several questions in the domains of law, ethics, and computer science. To sketch the contours of what explanations should look like, we raised questions on the generality and faithfulness of explanations. We also looked at the potential use of intrinsically interpretable AI models and the accuracy trade-offs that might arise in that regard. Applying procedural justice theories revealed diverse valuations of contestation, as well as different answers to these explainability questions. Moreover, they also show the different strengths and weaknesses of these theories in the age of judicial AI.

First, utilitarian and outcome-based theories, such as those proposed by Posner and Dworkin, respectively, show how explanations can have a valuable role in addressing AI-made errors.

[31] This leads us also to the important argument that a right to an explanation is not the sole remedy to address AI-made injustice. A number of different actors play a role in this regard, such as journalists, NGO's, lawyers, judges and litigants (Naudts, 2024). Litigants should therefore not be the sole guardians of justice when judicial AI is used. However, they nonetheless still deserve an effective chance to participate and help safeguard the correctness of the outcome of their trial.

Posner's law and economics approach shows contestability's value as a tool to minimize courts' error costs, in this regard. General explanations could help show the average reliability of systems to litigants, while specific explanations could help litigants contest errors in their particular cases. However, from this utilitarian perspective, we cannot accept a right to an explanation that leads to excessively high financial and accuracy costs for courts. Moreover, concerns can be raised about the provision of unfaithful explanations, as this would hinder the usefulness of contestation to address any real AI-made errors. The use of interpretable models might address this issue, but these systems can only be relied upon if there is no net increase in error costs or direct costs. In this regard, utilitarians also take into consideration the relative (societal) impact of the case at hand to balance the net utility and costs of the right to an explanation.

A rights-based approach rejects this balancing act and contends that litigants have genuine procedural rights that cannot be traded away. This approach can be further subdivided into the outcome-, and process-based approaches. The outcome-based approach to procedural justice, as proposed by Dworkin, argues that the right to an explanation serves to improve the accuracy of verdicts and to address AI-made errors, minimizing the moral harm of an incorrect decision. Specific explanations would be preferred in this regard, as this approach enhances the ability of litigants to reach an accurate verdict in their own specific case. But, as we saw in our discussion of the utilitarian approach, unfaithful explanations do not fulfill this requirement adequately. Moreover, explainability is of limited value if litigants are themselves unable to contest AI-made errors. Using intrinsically interpretable AI might offer a solution for the unfaithfulness of explanations. However, the trade-offs in accuracy that could arise from their use cannot be accepted, as this would increase the risk for moral harm. In any case, both the utilitarian and outcome-based approaches cannot convey the procedural harm that litigants experience when they cannot contest AI-made outputs. They limit participation to those instances where contestation can be useful in addressing mistakes.

Process-based theories, such as dignitarian and relational approaches, emphasize the non-instrumental and intrinsic value of contestation. Dignitarians emphasize that we must respect litigants' dignity, autonomy, and rational thinking by letting them participate in legal debates. To this end, specific explanations that enable litigants to discuss the particularities of their own trial would be preferred. However, the question also arises whether the same transparency standard should apply here as in regard to (solely) human decision-making. One could argue that the fact that an explanation is not faithful is not necessarily an issue, as we can also not look into the minds of human judges to create faithful explanations on how they reached their verdicts. Consequently, the solution that interpretable AI offers to the issue of faithfulness might be a laudable pursuit, but it is not a necessary requirement for dignitarian procedural justice.

As part of our critique on dignitarian approaches, we have argued that dignitarians do not take sufficiently into account litigants' different abilities to understand and utilize explanations for contestation. Relational theories, by contrast, emphasize the importance of conveying impartiality, trustworthiness, respect, and voice to litigants who have different practical needs and requirements to be able to partake in legal proceedings. Specific explanations would also be of use in this regard, but these must be understandable to different types of litigants. Nonetheless, explanations do not necessarily need to be faithful to provide the perception of legal proceedings to be impartial, trustworthy, respectful, or even to enable litigants to have a voice. Consequently, the use of interpretable AI models as a solution for the lack of faithfulness of explanations may be not necessary, as seen from the relational point of view.

This relational approach, like with the dignitarian approach, might then result in explanations that are 'caring' and 'accessible', all the while AI-made errors continue to oppress and harm litigants. With both dignitarian and relational approaches, there is therefore a risk for procedural fetishism. At the same time, the law and economics and Dworkinian approaches are

also flawed as they unsufficiently honour the intrinsic values of procedures. These, as we have seen in section 4 and 5.1, make the provision of a right to an explanation contingent on litigants' practical ability to contest AI-made errors.

In short, contestation's value can extend both to addressing errors and ensuring participation. But it is also at risk of not being useful and effective at all to address the substantive errors produced by judicial AI, potentially resulting in procedural fetishism. We argued, therefore, in section 7, that an AI-responsive theory of procedural justice must ensure that a right to an explanation promotes effective contestation of judicial AI. This entails that it genuinely enables litigants to improve the outcome of their trial. In this regard, we discussed how the relational approach could provide such a normative basis.

We emphasized that an AI-responsive theory of procedural justice should not consider contestation's value as purely a process-based endeavor, regardless of its usefulness to improve verdicts. At the same time, the right to an explanation should also not be made contingent on the practical ability of litigants to contest judicial AI and improve the quality of verdicts. Rather, litigants' pervasive inability to contest judicial AI tools is an invitation to re-evaluate who can, and should, be burdened with contesting AI-made errors. Moreover, we also see a task for judiciaries to facilitate litigants' capacities to contest judicial AI.

With that goal of effective contestation in mind, litigants are first of all in need of specific and faithful explanations, based on intrinsically interpretable systems. But we need to also consider the epistemic-, financial-, and time-based capital that individual litigants rely upon to utilize such explanations. Discussions on these systemic obstacles to procedural justice will have to play a more leading role in the way we substantiate the right to an explanation, if we wish to safeguard the effective contestation of judicial AI.

## Data Access Statement
Not applicable.

## Contributor Statement
Ljubiša Metikoš is the first author. Iris van Domselaar is the second author.

## Use of AI
No AI generated text has been used in the writing of this manuscript.

## Funding Statement
The authors did not receive any financial support for the work on this article.

## Acknowledgments
The authors would like to thank their colleagues at the Institute for Information Law (IViR), the Paul Scholten Centre for Jurisprudence, and the RPA Human(e)AI at the Faculty of Law of the University of Amsterdam, for their helpful insights and critiques. In particular, the authors would like to thank prof. dr. Natali Helberger and Sanne Vrijenhoek for their helpful commentary.

## Conflict of Interest
There is no conflict of interest.

## References
ADELE (2025), 'Ethical self-assessment' <https://site.unibo.it/adele/en/publications/ethical-report> accessed 11th of February 2025, p. 11.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), Article 3. https://doi.org/10.1177/1461444816676645

Anderson, E. (2008). Expanding the Egalitarian Toolbox: Equality and Bureaucracy. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *82*, 139–160.

Osztovits, A. (2021). In Technology We Trust? The Present and Possible Future of Private Enforcement. *Acta Universitatis Sapientiae, Legal Studies*, *10*(2), 231–244.

Babic, B., & Cohen, I. G. (2023). The Algorithmic Explainability "Bait and Switch." *Minnesota Law Review*, *108*, 857–909.

Bagley, N. (2019). The Procedure Fetish. *Michigan Law Review*, *118*(3), 345–402. https://doi.org/10.36644/mlr.118.3.procedure

Barocas, S., Hardt, M., Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. The MIT Press.

Barocas, S., Selbst, A. D., & Raghavan, M. (2019). The Hidden Assumptions Behind Counterfactual Explanations and Principal Reasons (SSRN Scholarly Paper No. 3503019). https://papers.ssrn.com/abstract=3503019

Bayamlioglu, E. (2018). Contesting Automated Decisions. *European Data Protection Law Review (EDPL)*, *4*, 433.

Bell, A., Solano-Kamaiko, I., Nov, O., & Stoyanovich, J. (2022). It's Just Not That Simple: An Empirical Study of the Accuracy-Explainability Trade-off in Machine Learning for Public Policy. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 248–266. https://doi.org/10.1145/3531146.3533090

Binns, R. (2022). Human Judgment in algorithmic loops: Individual justice and automated decision-making. *Regulation & Governance*, *16*(1), 197–211. https://doi.org/10.1111/rego.12358

Bone, R. G. (2017). Economics of Civil Procedure. In F. Parisi (Ed.), *The Oxford Handbook of Law and Economics: Volume 3: Public Law and Legal Institutions* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199684250.013.003

Brkan, M., & Bonnet, G. (2020). Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: Of Black Boxes, White Boxes and Fata Morganas. *European Journal of Risk Regulation*, *11*(1), 18–50. https://doi.org/10.1017/err.2020.10

Buijsman, S., Klenk, M., & Hoven, J. van den. (2025). Ethics of AI: Toward a "Design for Values" Approach. In N. A. Smuha (Ed.), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence* (pp. 59–78). Cambridge University Press. https://doi.org/10.1017/9781009367783.005

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1). https://doi.org/10.1177/2053951715622512

Ceva, E. (2020). The many facets of procedural justice in legal proceedings. In Meyerson, D., Mackenzie, C. MacDermott, T. (Eds.), *Procedural Justice and Relational Theory*. Routledge.

Dao, A. (2020). Human Dignity, the Right to be Heard, and Algorithmic Judges. *British Yearbook of International Law*, braa009. https://doi.org/10.1093/bybil/braa009

de Fine Licht, K., & de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making. *AI & SOCIETY*, *35*(4), 917–926. https://doi.org/10.1007/s00146-020-00960-w

de Laat, P. B. (2022). Algorithmic decision-making employing profiling: Will trade secrecy protection render the right to explanation toothless? *Ethics and Information Technology*, *24*(2), 17. https://doi.org/10.1007/s10676-022-09642-1

Dennett, D. C. (1998). *The intentional stance*. MIT Press.

District Court of Rotterdam. (2025, March 26). *Rechtbank Rotterdam doet proef met Artificial Intelligence als schrijfhulp in een strafvonnis*. https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Rotterdam/Nieuws/Paginas/Rechtbank-Rotterdam-doet-proef-met-Artificial-Intelligence-als-schrijfhulp-in-een-strafvonnis--.aspx

Donoghue, J. (2017). The Rise of Digital Justice: Courtroom Technology, Public Participation and Access to Justice. *The Modern Law Review*, *80*(6), 995–1025. https://doi.org/10.1111/1468-2230.12300

Dor, L. M. B., & Coglianese, C. (2021). Procurement as AI Governance. *IEEE Transactions on Technology and Society*, *2*(4), 192–199. IEEE Transactions on Technology and Society. https://doi.org/10.1109/TTS.2021.3111764

Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., Scott, K., Schieber, S., Waldo, J., Weinberger, D., Weller, A., & Wood, A. (2019). *Accountability of AI Under the Law: The Role of Explanation* (No. arXiv:1711.01134). arXiv. https://doi.org/10.48550/arXiv.1711.01134

Dworkin, R. (1981). Chapter 3, Principle, Policy, Procedure. In *A Matter of Principle* (9th print). Harvard Univ. Press, p. 72-103, https://doi.org/10.2307/j.ctv1pncpxk.6

Eckard Schindler. (2025, February 4). *Judicial systems are turning to AI to help manage vast quantities of data and expedite case resolution*. IBM Technology. https://www.ibm.com/case-studies/blog/judicial-systems-are-turning-to-ai-to-help-manage-its-vast-quantities-of-data-and-expedite-case-resolution

Edwards, L., & Veale, M. (2017). Slave to the Algorithm? Why a Right to Explanationn is Probably Not the Remedy You are Looking for. *Duke Law & Technology Review*, *18*, 67. https://doi.org/10.2139/ssrn.2972855

Fabri, M. (2024). From Court Automation to e-Justice and Beyond in Europe. *International Journal for Court Administration*, *15*(3). https://doi.org/10.36745/ijca.640

Freyer, N., Kempt, H., & Klöser, L. (2024). Easy-read and large language models: On the ethical dimensions of LLM-based text simplification. *Ethics and Information Technology*, *26*(3), 1–10. https://doi.org/10.1007/s10676-024-09792-4

Gentile, G. (2024). *Human Law, Human Lawyers and the Emerging AI Faith* (SSRN Scholarly Paper No. 4918632). https://doi.org/10.2139/ssrn.4918632

Genn, H. (2010) *Judging Civil Justice.* Cambridge University Press.

H. v. Belgium, No. 8950/80 (ECtHR November 30, 1987). https://hudoc.echr.coe.int/eng?i=001-57501

Hendrickx, V. (2025). Rethinking the judicial duty to state reasons in the age of automation? *Cambridge Forum on AI: Law and Governance*, *1*, e26. https://doi.org/10.1017/cfl.2025.11

Henin, C., & Le Métayer, D. (2021). Beyond explainability: Justifiability and contestability of algorithmic decision systems. *AI & SOCIETY*. https://doi.org/10.1007/s00146-021-01251-8

Hildebrandt, M. (2019). Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning. *Theoretical Inquiries in Law*, *20*(1), 83–121. https://doi.org/10.1515/til-2019-0004

Hylton, K. N. (2017). Economics of Criminal Procedure. In F. Parisi (Ed.), *The Oxford Handbook of Law and Economics: Volume 3: Public Law and Legal Institutions*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199684250.013.025

Jongepier, F., & Keymolen, E. (2022). Explanation and Agency: Exploring the normative-epistemic landscape of the "Right to Explanation." *Ethics and Information Technology*, *24*(4), 49. https://doi.org/10.1007/s10676-022-09654-x

Kaminski, M. E., & Urban, J. M. (2021). The Right to Contest AI. *Columbia Law Review*, *121*(7), 1957–2021.

Kaplow, L., & Shavell, S. (1994). Accuracy in the Determination of Liability. *The Journal of Law and Economics*, *37*(1), 1–15. https://doi.org/10.1086/467304

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, *20*(1), Article 1. https://doi.org/10.1080/1369118X.2016.1154087

Liao, Q. V., & Vaughan, J. W. (2023). *AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap* (No. arXiv:2306.01941). arXiv. http://arxiv.org/abs/2306.01941

Lind, E. A., & Tyler, T. R. (1988). *The Social Psychology of Procedural Justice*. Springer. https://doi.org/10.1007/978-1-4899-2115-4

MacCoun, R. J. (2005). Voice, Control, and Belonging: The Double-Edged Sword of Procedural Fairness. https://escholarship.org/uc/item/011185w5

Mackenzie, C. (2020). Procedural justice, relational equality, and self-respect. In Meyerson, D., Mackenzie, C. MacDermott, T. (Eds.), *Procedural Justice and Relational Theory*. Routledge.

Marchiori, T., *Framework for Measuring Access to Justice Including Specific Challenges Facing Women: Guidance Note* (UN Women and Council of Europe, July 2016). https://rm.coe.int/framework-for-measuring-access-to-justice-including-spesific-challenge/1680a876b9

Meers, J., Halliday, S., & Tomlinson, J. (2023). Chapter 28*:* Why we need to rethink procedural fairness for the digital age and how we should do it, in B. Brożek, O. Kanevskaia, & P. Pałka (Eds.), Research handbook on law and technology. Edward Elgar Publishing. https://www.elgaronline.com/edcollchap/book/9781803921327/chapter28.xml

Merken, S., & Merken, S. (2023, June 26). New York lawyers sanctioned for using fake ChatGPT cases in legal brief. *Reuters*. https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/

Metikoš, L. (2024). Explaining and Contesting Judicial Profiling Systems. *Technology and Regulation*, *2024*, 188–208. https://doi.org/10.26116/techreg.2024.017

Metikoš, L., & Ausloos, J. (2025). The right to an explanation in practice: Insights from case law for the GDPR and the AI Act. *Law, Innovation and Technology* 17(1), 205-140. https://doi.org/10.1080/17579961.2025.2469349

Metikoš, L. (2024b), The AI Act: Weak, Weaker, Weakest (June 13, 2024), Mediaforum 2024-3, p. 73-74. http://dx.doi.org/10.2139/ssrn.4873496

Metikoš, L. (2025b), A Procedural Sedative: The GDPR's Right to an Explanation (September 26, 2025). *Data, Cybesecurity & Privacy*, 18 & 19, September 2025. pp. 24-27.

Meyerson, D. (2020). The inadequacy of instrumentalist theories of procedural justice. In Meyerson D., Mackenzie C. MacDermott T. (Eds.) *Procedural Justice and Relational Theory; Empirical, Philosophical, and Legal Perspectives*. Routledge.

Meyerson, D., & Mackenzie, C. (2018). Procedural justice and the law. *Philosophy Compass*, *13*(12), e12548. https://doi.org/10.1111/phc3.12548

Meyerson, D., Mackenzie, C., & MacDermott, T. (Eds.). (2021). *Procedural Justice and Relational Theory: Empirical, Philosophical, and Legal Perspectives*. Routledge. https://directory.doabooks.org/handle/20.500.12854/72677

Miao, M. (2022). Debating the Right to Explanation: An Autonomy-Based Analytical Framework. *Singapore Academy of Law Journal*, 34, 736-762. https://doi.org/10.2139/ssrn.4225103

Miró-Nicolau, M., Jaume-i-Capó, A., & Moyà-Alcover, G. (2024). Assessing fidelity in XAI post-hoc techniques: A comparative study with ground truth explanations datasets. *Artificial Intelligence*, *335*, 104179. https://doi.org/10.1016/j.artint.2024.104179

Mohseni, S., Zarei, N., & Ragan, E. D. (2021). A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *ACM Transactions on Interactive Intelligent Systems*, *11*(3–4), 1–45. https://doi.org/10.1145/3387166

Molnar, C., König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C. A., Casalicchio, G., Grosse-Wentrup, M., & Bischl, B. (2022). General Pitfalls of Model-Agnostic Interpretation Methods for Machine Learning Models. In A. Holzinger, R. Goebel, R. Fong, T. Moon, K.-R. Müller, & W. Samek (Eds.), *xxAI - Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers* (pp. 39–68). Springer. https://doi.org/10.1007/978-3-031-04083-2_4

Munch, L. A., Bjerring, J. C., & Mainz, J. T. (2024). Algorithmic decision-making: The right to explanation and the significance of stakes. *Big Data & Society*, *11*(1). https://doi.org/10.1177/20539517231222872

Naudts, L. (2024). The Digital Faces of Oppression and Domination: A Relational and Egalitarian Perspective on the Data-driven Society and its Regulation. *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 701–712. https://doi.org/10.1145/3630106.3658934

Naudts, L., & Vedder, A. (2025). Fairness and Artificial Intelligence. In N. A. Smuha (Ed.), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence* (pp. 79–100). Cambridge University Press. https://doi.org/10.1017/9781009367783.006

Nišević, M., Cuypers, A., & De Bruyne, J. (2024). Explainable AI: Can the AI Act and the GDPR go out for a date? *2024 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN60899.2024.10649994

Ogonjo, F., Gitonga, J., Wairegi, A., & Rutenberg, I. (2021). *Utilizing AI to Improve Efficiency of the Environment and Land Court in the Kenyan Judiciary Leveraging AI Capabilities in Land Dispute Cases in the Kenyan Environmental and Land Court System*.

Osa, D. U. S. de la, & Remolina, N. (2024). Artificial intelligence at the bench: Legal and ethical challenges of informing—or misinforming—judicial decision-making through generative AI. *Data & Policy*, *6*, e59. https://doi.org/10.1017/dap.2024.53

Palmiotto, F. (2021). The Black Box on Trial: The Impact of Algorithmic Opacity on Fair Trial Rights in Criminal Proceedings. In M. Ebers & M. Cantero Gamito (Eds.), *Algorithmic Governance and Governance of Algorithms: Legal and Ethical Challenges* (pp. 49–70). Springer. https://doi.org/10.1007/978-3-030-50559-2_3

Pasquale, F., & Malgieri, G. (2024). Generative AI, Explainability, and Score-Based Natural Language Processing in Benefits Administration. *Journal of Cross-Disciplinary Research in Computational Law*, *2*(2), Article 2. https://journalcrcl.org/crcl/article/view/59

Pereira, J., Assumpcao, A., Trecenti, J., Airosa, L., Lente, C., Cléto, J., Dobins, G., Nogueira, R., Mitchell, L., & Lotufo, R. (2024). INACIA: Integrating Large Language Models in Brazilian Audit Courts: Opportunities and Challenges. *Digit. Gov.: Res. Pract.* https://doi.org/10.1145/3652951

Peters, T. M., & Visser, R. W. (2023). The Importance of Distrust in AI. In L. Longo (Ed.), *Explainable Artificial Intelligence* (pp. 301–317). Springer. https://doi.org/10.1007/978-3-031-44070-0_15

Porębski, A. (2024). Institutional Black Boxes Pose an Even Greater Risk than Algorithmic Ones in a Legal Context. In J. Mańdziuk, A. Żychowski, & M. Małkiński (Eds.), *Progress in Polish Artificial Intelligence Research* (pp. 562–570). https://papers.ssrn.com/abstract=4971723

Posner, R. A. (1973). An Economic Approach to Legal Procedure and Judicial Administration. *The Journal of Legal Studies*, *2*(2), 399–458.

Qiao, C. and Metikoš, L (2025), Judicial Automation: Balancing Rights Protection and Capacity-Building (Preprint published February 05, 2025). In U. Schultz, H. Hyden, P. Scharff Smith (Eds.) *Elgar Encyclopedia of the Sociology of Law, Forthcoming* 2026 (forthcoming), ISBN 9781035333158 papers.ssrn.com/sol3/papers.cfm?abstract_id=5125645

Rahnama, A., & Hossein, A. (2025). *Evaluating the Faithfulness of Local Feature Attribution Explanations: Can We Trust Explainable AI?* (Doctoral thesis), KTH Royal Institute of Technology https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-360228

Ranchordas, S. (2021). Empathy in the Digital Administrative State Automating the Administrative State. *Duke Law Journal*, *71*(6), 1341–1390.

Rawls, J. (1999). *A theory of justice* (Rev. ed). Belknap Press of Harvard University Press.

Re, R. M., & Solow-Niederman, A. (2019). Developing Artificially Intelligent Justice. *UCLA School of Law, Public Law Research Paper*, *16–19*, 48.

Rebera, A. P., Lauwaert, L., & Oimann, A.-K. (2025). Hidden Risks: Artificial Intelligence and Hermeneutic Harm. *Minds and Machines*, *35*(3), 1–18. https://doi.org/10.1007/s11023-025-09733-0

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), Article 5. https://doi.org/10.1038/s42256-019-0048-x

Rueda, J., Ausín, T., Coeckelbergh, M., Valle, J. I. del, Lara, F., Liedo, B., Albareda, J. L., Mertes, H., Ranisch, R., Raposo, V. L., Stahl, B. C., Vilaça, M., & Miguel, Í. de. (2025). Why dignity is a troubling concept for AI ethics. *Patterns*, *6*(3). https://doi.org/10.1016/j.patter.2025.101207

Ruiz-Mateos v. Spain, No. 12952/87 (ECtHR June 23, 1993). https://hudoc.echr.coe.int/eng?i=001-57838

Sarra, C. (2020). Put Dialectics into the Machine: Protection against Automatic-decision-making through a Deeper Understanding of Contestability by Design. *Global Jurist*, *20*(3). https://doi.org/10.1515/gj-2020-0003

Selbst, A. D., & Barocas, S. (2018). The Intuitive Appeal of Explainable Machines. *Fordham Law Review*, *87*(3), 1085–1139.

Shi, J. (2022). Artificial Intelligence, Algorithms and Sentencing in Chinese Criminal Justice: Problems and Solutions. *Criminal Law Forum*, *33*(2), 121–148. https://doi.org/10.1007/s10609-022-09437-5

Smuha, N. A. (2021a). Beyond the individual: Governing AI's societal harm. *Internet Policy Review*, *10*(3). https://policyreview.info/articles/analysis/beyond-individual-governing-ais-societal-harm

Smuha, N. A. (2021b). The Human Condition in An Algorithmized World: A Critique through the Lens of 20th-Century Jewish Thinkers and the Concepts of Rationality, Alterity and History (SSRN Scholarly Paper No. 4093683). Social Science Research Network. https://doi.org/10.2139/ssrn.4093683

Solon Barocas & Andrew D Selbst. (2016). Big Data's Disparate Impact. *California Law Review*, *104*(3), Article 3.

Solum, L. B. (2004). Procedural Justice Articles & Commentary. *Southern California Law Review*, *78*(1), 181–322.

Sourdin, T. (2018). Judge v. Robot: Artificial Intelligence and Judicial Decision-Making. *University of New South Wales Law Journal*, *41*(4), Article 4.

Sourdin, T. (2021a). *Chapter 1: Judges and technology*. In *Judges, Technology and Artificial Intelligence* (pp. 1–31). Edward Elgar Publishing. https://www.elgaronline.com/monochap/9781788978255.00005.xml

Sourdin, T. (2021b). The role and function of a judge: The adoption and adaptation of technology by judges. In *Judges, Technology and Artificial Intelligence* (pp. 32–63). Edward Elgar Publishing. https://www.elgaronline.com/view/9781788978255.00006.xml

Stern, R. E., Liebman, B. L., Roberts, M. E., & Wang, A. Z. (2021). Automating Fairness? Artificial Intelligence in the Chinese Courts. *The Columbia Journal of Transnational Law*, *59*(3), 515–553.

Stevens, A., & De Smedt, J. (2024). Explainability in process outcome prediction: Guidelines to obtain interpretable and faithful models. *European Journal of Operational Research*, *317*(2), 317–329. https://doi.org/10.1016/j.ejor.2023.09.010

Stohl, C., Stohl, M., & Leonardi, P. M. (2016). Digital Age | Managing Opacity: Information Visibility and the Paradox of Transparency in the Digital Age. *International Journal of Communication*, *10*(0), Article 0.

Summers, R. S. (1974). Evaluating and Improving Legal Processes A Plea for Process Values. *Cornell Law Review*, *60*(1), 1–52.

Surden, H. (2024). ChatGPT, Large Language Models, and Law. *Fordham Law Review*, *92*(5), 1941.

Susskind, R. E. (2019). *Online courts and the future of justice* (First Edition.). Oxford University Press.

Taekema, S., & Burg, W. van der. (2020). Legal Philosophy as an Enrichment of Doctrinal Research Part I: Introducing Three Philosophical Methods. *Law and Method*, *01*. https://doi.org/10.5553/REM/.000046

Tim Wu. (2019). Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems. *Columbia Law Review*, *119*(7), Article 7.

Tyler, T. R. (1988). What is Procedural Justice?: Criteria Used by Citizens to Assess the Fairness of Legal Procedures. *Law & Society Review*, *22*(1), 103–135. https://doi.org/10.2307/3053563

Tyler, T. R., & Blader, S. L. (2000). *Cooperation in groups: Procedural justice, social identity, and behavioral engagement* (pp. ix, 233). Psychology Press.

Tyler, T. R., & Lind, E. A. (1992). A relational model of authority in groups. In *Advances in experimental social psychology, Vol. 25* (pp. 115–191). Academic Press. https://doi.org/10.1016/S0065-2601(08)60283-X

Van Domselaar, I. (2022a). 'Plain' legal language by courts: Mere clarity, an expression of civic friendship or a masquerade of violence? *The Theory and Practice of Legislation*, *10*(1), 93–111. https://doi.org/10.1080/20508840.2022.2033946

Van Domselaar, I. (2022b). Inquiry and Imagination in Adjudication. *Netherlands Journal of Legal Philosophy*, *51*(2), 187–198. https://doi.org/10.5553/NJLP/221307132022051002008

Van Domselaar, I. (2026). The role of legal professionals in large-scale miscarriages of justice: a virtue ethics perspective. *Legal Ethics*, *27*(2), 1–31. https://doi.org/10.1080/1460728x.2025.2590858

Volokh, E. (2019). Chief Justice Robots. *Duke Law Journal*, *68*(6), Article 6.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law*, *7*(2), Article 2. https://doi.org/10.1093/idpl/ipx005

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. *Harvard Journal of Law & Technology (Harvard JOLT)*, *31*(2), 841–888.

Waldron, J. (2011). The Rule of Law and the Importance of Procedure. *Nomos*, *50*, 3–31.

Yeung, K., & Harkens, A. (2023). How do "technical" design choices made when building algorithmic decision-making tools for criminal justice authorities create constitutional dangers? (Part I). *Public Law*, *2023*(Apr), 265–286.

Yurrita, M., Draws, T., Balayn, A., Murray-Rust, D., Tintarev, N., & Bozzon, A. (2023). Disentangling Fairness Perceptions in Algorithmic Decision-Making: The Effects of Explanations, Human Oversight, and Contestability. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–21. https://doi.org/10.1145/3544548.3581161

Zalnieriute, M. (2021). *"Transparency-Washing": The Corporate Agenda of Procedural Fetishism*. *8*(1), 39–53.

Zarsky, T. Z. (2013). Transparent Predictions. *University of Illinois Law Review*, *2013*(4), 1503–1570.

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard? *Philosophy & Technology*, *32*(4), 661–683. https://doi.org/10.1007/s13347-018-0330-6