Am I Being Treated Fairly? A Conceptual Framework for Individuals to Ascertain Fairness

Juliett Suárez-Ferreira[1,*], Marija Slavkovik[2] and Jorge Casillas[1,3]

[1]Data Science and Computational Intelligence Institute (DaSCI), University of Granada, Calle Periodista Daniel Saucedo Aranda, s/n. 18071 Granada, Spain.

[2]Department of Information Science and Media Studies, University of Bergen, Fosswinckels gate 6, 5007 Bergen, Norway.

[3]Department of Computer Science and Artificial Intelligence (DCSAI), University of Granada, Higher Technical School of Computer and Telecommunications Engineering, 18071 Granada, Spain.

Abstract

Current fairness metrics and mitigation techniques provide tools for practitioners to asses how non-discriminatory Algorithmic Decision Making (ADM) systems are. What if I, as an individual facing a decision taken by an ADM system, would like to know: *Am I being treated fairly?*. We explore how to create the affordance for users to be able to ask this question of ADM. In this paper, we argue for the reification of fairness not only as a property of ADM, but also as an epistemic right of an individual to acquire information about the decisions that affect them and use that information to contest and seek effective redress against those decisions, in case they are proven to be discriminatory. We examine key concepts from existing research not only in algorithmic fairness but also in explainable artificial intelligence, accountability, and contestability. Integrating notions from these domains, we propose a conceptual framework to ascertain fairness by combining different tools that empower the end-users of ADM systems. Our framework shifts the focus from technical solutions aimed at practitioners to mechanisms that enable individuals to understand, challenge, and verify the fairness of decisions, and also serves as a blueprint for organizations and policymakers, bridging the gap between technical requirements and practical, user-centered accountability.

Keywords

contestability, fairness, discrimination, procedural fairness

1. Introduction

Artificial intelligence (AI) is increasingly being deployed to automate decisions in the personal, business and public domains [1]. The primary motivation for automation is efficiency and cost reduction, yet AI's societal impact requires ensuring that systems are trustworthy [2]. One of the most sensitive AI applications is algorithmic decision making (ADM), where AI determines outcomes in high-stakes areas like credit, employment, and healthcare [3]. This raises a fundamental question from the end-users point of view: *Am I being treated fairly by this AI system?*

There are different interpretations of fairness; but, the approach towards accomplishing fairness, however interpreted, needs to be both substantive and procedural [4]. Substantive fairness ensures an equitable distribution of benefits and burdens, preventing bias and discrimination. Procedural fairness emphasizes the ability to understand, challenge, and seek redress for AI-driven decisions [5]. However, digitalization and algorithmic opacity complicate the fairness assessment, and the scale of algorithmic decisions vastly exceeds human oversight, raising concerns about systemic discrimination and the accountability of AI-driven processes.

The notion of epistemic rights (the right to access, evaluate, and challenge information that affects one's life) is increasingly recognized as a cornerstone of democratic participation in digitally mediated societies [6]. These rights are especially critical in contexts where decisions are automated and opaque.

Second Multimodal, Affective and Interactive eXplainable AI Workshop (MAI-XAI25 2025) co-located with the 28th European Conference on Artificial Intelligence 25-30 October 2025 (ECAI 2025)

*Corresponding author.

juliettsuarez@correo.ugr.es (J. Suárez-Ferreira); Marija.Slavkovik@uib.no (M. Slavkovik); casillas@decsai.ugr.es (J. Casillas)

0000-0002-1093-040X (J. Suárez-Ferreira); 0000-0003-2548-8623 (M. Slavkovik); 0000-0002-5887-3977 (J. Casillas)

In light of this, we argue that fairness in ADM systems must be reified as an epistemic right, which we term *ascertainable fairness*, where end-users can access, understand, and contest decisions that affect them.

The scope of our proposal focuses on investigating how to enable individual users to actively engage and assess the fairness of decisions made by ADM systems using the current state of the art in substantive and procedural fairness by building upon prior research in algorithmic fairness, explainability, accountability, and contestability.

The remainder of this paper is organized as follows. We begin by establishing the conceptual foundation for fairness in ADM, distinguishing between substantive and procedural dimensions, and examining their relevance to individual experiences of fairness. Next, we analyze the limitations of existing fairness approaches in supporting individuals' ability to ascertain fairness, motivating the need for a more integrated framework. We then outline what it means for individuals to inquire about and be informed of the fairness of ADM decisions, emphasizing the need for personalized explanations, justifications, and mechanisms that go beyond technical metrics to support meaningful contestation. Building on this, we introduce the notion of ascertainable fairness, conceptualizing fairness as an epistemic right, and present a conceptual framework that details its components and their interactions with relevant stakeholders. This is followed by a discussion of our contributions, as well as challenges and open research areas. We conclude by summarizing our findings and reflecting on their implications for policy, organizational practice, and the broader discourse on AI fairness.

2. Algorithmic Decision-Making and Fairness

Algorithmic decision-making (ADM) refers to computational systems that automate decision processes by analyzing patterns in past data to make predictions and guide decisions, often operating as opaque *black boxes*. This highlights the importance of transparency as a requirement for ADM systems, enabling users to recognize that they are interacting with such a system and to understand not only how decisions are made and why [4], but also how to ensure that these decisions are fair.

While algorithmic fairness research often focuses on formal, statistical definitions [7] foundational theories from philosophy, social science, and law emphasize fairness as both an equitable outcome and a transparent, contestable process. For instance, Rawlsian justice conceptualizes fairness through principles of equality and justified inequality [8], while procedural justice theory highlights legitimacy, voice, and transparency as essential to perceived fairness [9]. Legal traditions emphasize non-discrimination and equal protection, particularly with respect to protected characteristics [10]. Our adoption of the High-Level Expert Group on AI (HLEG) definition of fairness; grounded in substantive (equitable outcomes) and procedural (transparency, contestability and redress) dimensions [4], offers a multidisciplinary bridge between these normative traditions and computational practice. This allows us to frame fairness as a technical property to be engineered and an epistemic right of individuals to understand, challenge, and contest algorithmic decisions that affect them.

2.1. Substantive Dimension of Fairness: Algorithmic Fairness

The substantive dimension of fairness is defined as *a commitment to ensuring equal and just distribution of benefits and costs, ensuring that individuals and groups are free from unfair bias, discrimination, and stigmatization* [4].

The field of *algorithmic fairness* examines what makes ADM decisions fair; representing the operationalization of the substantive dimension of fairness in ADM systems. This field offers *measures* for the quantification of unfairness and *methods* to mitigate discrimination in algorithmic decisions, considering protected attributes such as race, sex, or age [10]. Algorithmic fairness seeks to understand and correct the sources of unfairness [11] identified as discrimination, resulting from human prejudice and stereotyping, and bias, arising from different interactions between humans, data and algorithms [12]. Several literature reviews have attempted to create a taxonomy of different aspects of the field, e.g., [13, 12, 14] typically assessing fairness using two views: *Individual fairness* and *Group fairness*.

Another important area in the field of algorithmic fairness is the process of ameliorating the effect of bias on one or more protected attributes at different stages of the development of the ADM system (pre-, in-, and post-processing), called *bias mitigation* [7, 15].

2.2. Procedural Dimension of Fairness in ADM Systems.

The procedural dimension of fairness includes the ability to contest and seek redress against decisions made by AI systems and their human operators [4]. For this to be effective, the responsible entity must be identifiable and the decision-making processes must be understandable. We examine different parts of this definition that extend the understanding of procedural fairness beyond the fair decision-making process studied by [16] to also include mechanisms for redress and ensuring contestation.

Explicability of decision-making processes. The decision-making processes of AI systems should be explicable, that is, transparent and understandable to the affected parties. This aligns with the transparency requirement for trustworthy AI [4] with the dimensions of traceability, explainability, and communication [17]. Traceability monitors system data, development, and deployment through documentation. Communication covers how decisions and interactions with ADM systems are conveyed to end-users. Explainability (XAI) [18] is crucial to achieve user-perceived fairness by clarifying decision processes to all stakeholders, showing rationale, and offering alternatives. It helps detect ADM biases by revealing decision-making attributes, logic, and organizational rules.

Identifiability of accountable entities and redress. For procedural fairness to be actionable, the entity responsible for the AI decision must be identifiable, and the end-users affected by AI decisions should have mechanisms to obtain remedies if decisions are found to be unjust or erroneous. This allows someone to be responsible for ADM system decisions and ensures mechanisms to correct harm from unfair or incorrect decisions, such as reversing decisions, offering compensation, or implementing preventive measures, which is essential for transparency, trust, and fairness [4]. *Auditability*, the capacity to evaluate algorithms, data and design, is also a key aspect of accountability. Different fairness audit mechanisms are described in [19, 20]. These mechanisms help users verify fairness, but must be tailored to their needs; otherwise, trust in independent auditors is essential [21]. In disputes, auditing offers impartial conflict resolution.

Contestation. The ability to dispute decisions made by ADM systems is crucial to ensuring procedural fairness. Users must have the opportunity to appeal and scrutinize these decisions, granting them the power to challenge the outcomes of such systems, and thus the possibility of an unfair treatment. Although contestability is not considered a principle for trustworthy AI by [4], it is considered a pathway to fairness. Some researchers perceive contestability as a post hoc tool to challenge decisions [22], while others see it as a design feature [23, 24]. We argue that both elements are crucial. ADM systems should be inherently contestable, and a post-deployment mechanism should allow users to contest the results of ADM systems. We consider that contestability should allow users to challenge not only decisions, but also fairness metrics, involved attributes, and their impact on outcomes.

3. How current work help individuals to be informed about fairness

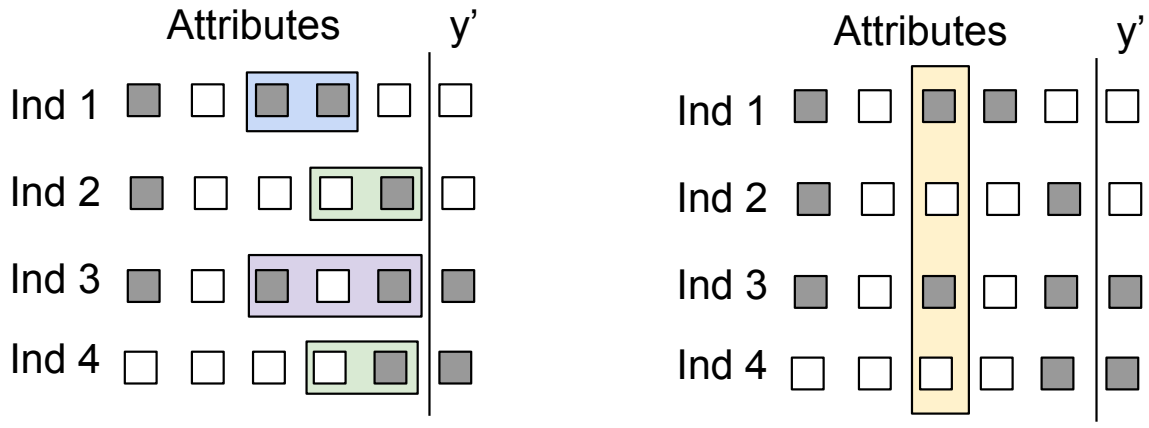
In this section, we discuss the extent to which ascertainable fairness can be achieved by existing approaches to substantive and procedural fairness.

3.1. Algorithmic Fairness and Ascertainable Fairness

For the purposes of ascertainable fairness, unfairness should be evaluated against a subset of personal characteristics that encapsulates the group (collective) identity of the end-user, and this subset of characteristics can vary between individuals. A *collective identity* is one that is shared with a group of

others who have (or are believed to have) some characteristic(s) in common [25]. Beyond the legally defined set of protected attributes, individuals can identify with different groups at the same time and identify more strongly with some of these groups over others. Individuals try to find balance in their need to belong and in their need to be different [26].

An organization may prioritize fairness for one protected group over others. However, the attributes they use to demonstrate system unfairness may differ from those reflecting the collective identity of individuals using the ADM system. This need for balance is evident in aligning with the protected group each individual identifies with. Figure 1 illustrates differences in individual perceptions of fairness versus fairness implementation in ADM systems.



(a) Individual perceptions of fairness can vary, as each person may possess a unique set of attributes that contribute to their collective identity.

(b) The implementation of fairness within ADM systems typically involves organizations adopting a particular concept of fairness that considers specific protected attributes.

Figure 1: Variations in the perception of fairness by individuals and organizations.

Mitigation techniques and fairness metrics help practitioners create models for ADM systems that align with non-discrimination standards. However, these are not directly available tools for individuals to evaluate fairness, as they often lack expertise and access to necessary data. Many algorithms are either non-transparent or proprietary, restricting public scrutiny. This limits individuals' ability to assess the fairness of algorithms affecting their lives. Fairness metrics can inform users, but understanding them is required for individual and societal benefits [6].

3.2. Procedural Fairness and Ascertainable Fairness

We argue that the procedural dimension of fairness gives more resources to end-users, which allows them to ascertain fairness providing not just a metric, but also the understanding of how decisions were taken (explainability), the possibility to challenge them (contestability) and obtain effective compensation (accountability and redress), which supports the epistemic right of ascertainable fairness. However, these fields are not put together in the service of procedural fairness. Furthermore, although comparatively much work has been done in the field of explainable AI, the same effort is not matched in contestability, the technical aspects of accountability such as auditability and redress. The existing work in these fields studied in the previous section provide the building blocks for our ascertainable fairness framework.

4. What It Means to Ask and Inform About Fairness

Consider a scenario in which an ADM system S, deployed by some institution, takes input from a data point, a set of known attributes $A = a_1, a_2, ..a_n$, whose values describe a particular problem instance

associated with an individual. S produces on output y', which is the decision taken in our case. The individual should be able to inquire about and determine how fair the decision made by S is in their case. We say that **the end user can ascertain the fairness of the decisions that affect them**.

Fairness in ADM systems can be assessed at both individual and collective levels. When we compare groups against groups, we evaluate fairness on a collective level. In this case, fairness is evaluated against a subset of attributes denoted as $P_A \in A$, whose values encapsulate the group (or collective) identity of an individual. However, unfairness is experienced individually. The collective identity of a person is unique; the attributes that are considered important to an individual may not be the attributes considered by the organizations that provide the ADM systems. Moreover, the perception of fairness is individual and can be manifested differently in different end-users. In other words, different things may matter to different people. A user may attempt to find out if particular attribute values are causing the different treatment they are facing, irrespective of their group affiliation.

Individuals must be able to determine whether their characteristics have influenced biased decisions and whether there are feasible actions to rectify unfair outcomes. Fairness metrics and explanations are useful in this regard; different tools have been designed to verify the fairness of predictions made by the ADM systems [27, 28, 29] and fairness of recourses interpreted as the actions that the end-users of the ADM systems need to do to change the decision [30, 31]. Nevertheless, the result of these verification processes is a value of a fairness measure, but is a numerical value from a measure or a simple true/false derived from them enough? We consider it insufficient and hard to understand but helpful as grounds for establishing a contestation dialog (as suggested by [32]) with the ADM that should justify the relevance/suitability of the measures, attributes and processes used to make the decision.

Justifications for decisions are essential to evaluate fairness. In some cases, differential treatment can be justified, such as in medical risk assessments based on empirical evidence [33], while other disparities, such as biased recidivism predictions, may indicate unjustified discrimination. Transparency in decision making allows users to distinguish between fair differentiation and unlawful bias.

Ultimately, assessing and informing fairness involves more than technical fairness metrics. It requires enabling individuals to engage in explanations, challenge discriminatory practices, and receive justified responses from ADM systems. These procedural elements ensure that fairness remains an enforceable right, rather than a theoretical concept.

4.1. Contesting Dialogues for Ascertaining Fairness

We now consider what allowing for a contesting dialogue can mean. [32] discusses how to build a computationally plausible contestation process based on argumentation. We consider that this general process of contestations based on the exchange of arguments operationalizes contestability as an essential mechanism of the procedural dimension of fairness to allow users to ascertain fairness (see Figure 2).

End-users or other stakeholders can challenge the suitability of the ADM system for a task. Apart from contesting the system itself concerning its adequacy and appropriateness for addressing the problem at hand, we provide a list of fairness-related contestations they can raise:

- The output of the ADM system. The user could disagree with the decision received, this is the primary contestation.
- The use (or non-use) of attributes. This should include the particular combination of attributes that the user is identifying with.
- The importance of an attribute or the correlation of an attribute with the output received.
- The fairness measure used. This will challenge the concept of discrimination utilized by the organization that supplies the ADM system.
- The fairness of the predictions, explanations, and recourses received by the end-user as well as their validity.
- The validity (legitimacy) of the justifications given in the contestation process.
- The variation of the outcome for a different case similar to the end-user's case.

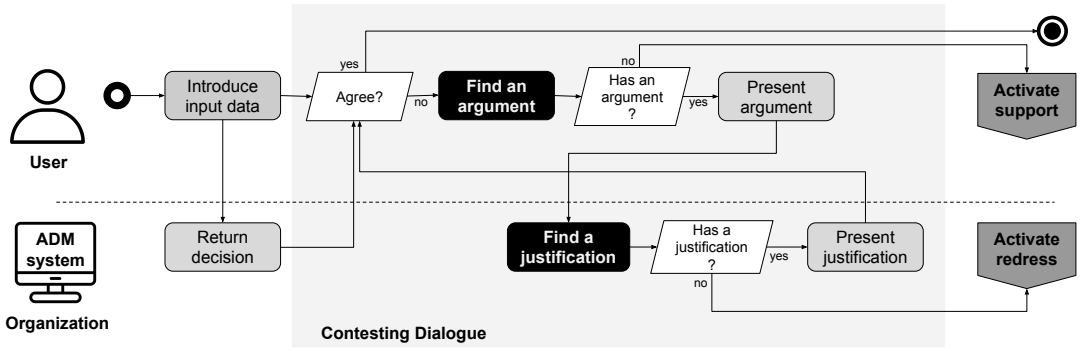


Figure 2: Contestation dialogue between the user and the ADM system. The process models a structured exchange where the user presents an argument challenging a decision, and the system responds with a justification. If the justification is unsatisfactory or absent, support or redress mechanisms are activated, enabling the user to escalate the issue. This dialogue operationalizes contestability as a procedural mechanism for ascertaining fairness.

- An error in the applied norms/rules specific to the solution .
- An internal rule of the organization revealed in a justification.

The contestation dialogue is an iterative and structured interaction through which users can challenge the fairness of decisions made by ADM systems. As illustrated in Figure 2, the process involves the user presenting an argument and the system responding with a justification. If the user finds the justification inadequate, the process escalates. Ground generation tools as explained by [32] can help users formulate arguments, especially when they lack the expertise to articulate their concerns.

The result of a contestation process to ascertain fairness should be a legitimate justification that convinces the individual of the fairness of the treatment received; otherwise, if discrimination is exposed, this process could potentially lead to a change in the decision received using a redress mechanism. If contestation shows discrimination or the user doubts the justification's legitimacy, they should be able to request an audit from a relevant regulatory authority. This process makes contestability a concrete procedural safeguard, enabling users to interrogate and rectify perceived unfairness rather than passively accept automated outcomes.

5. Ascertainable Fairness

After analyzing different aspects of substantive and procedural dimensions of fairness as well as the interoperability of different fields in response to the epistemic right of an individual to ascertain the fairness of the ADM systems decisions that affect them; we can describe ascertainable fairness as *the ability of end-users of ADM systems to authenticate the concept of fairness with which they resonate, considering their shared identity and individual traits, potentially identifying sources of unfairness and acquiring justifications via a contestability mechanism, leading to either verification of fairness or availability of redress and / or audit results.*

5.1. Ascertainable Fairness: a Conceptual Framework

In this section, we present a conceptual framework that will support individuals in the process of ascertaining fairness using different tools. Figure 3 illustrates the interaction between the components of the framework and the stakeholders involved in the process.

Component 1 (**C1**) is a tool to check the fairness of the predictions. The output output y' received by the user giving the attributes provided. The tool for assessing the fairness of the predictions will have access to query the system S and build a parallel model that can simulate the behavior of the system and verify bias in the algorithm decision.

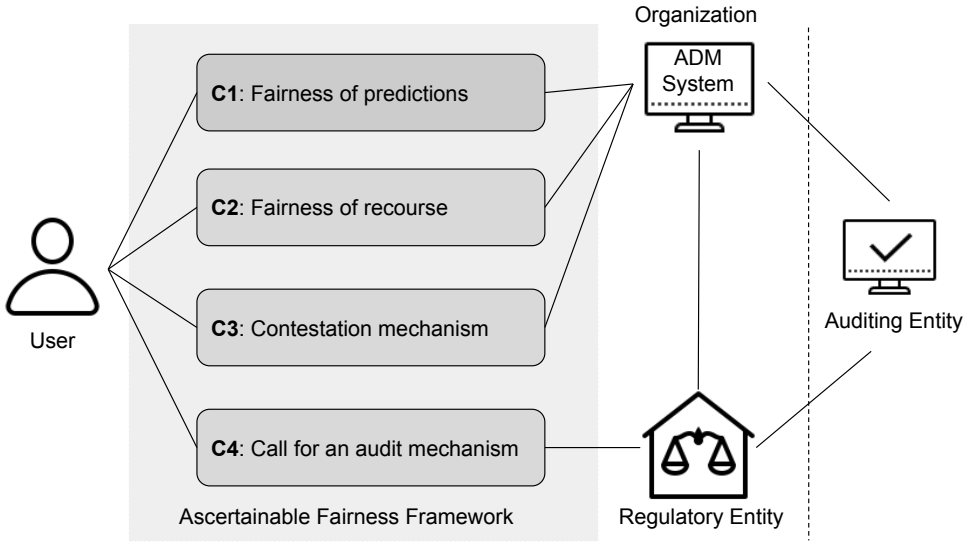


Figure 3: Ascertainable Fairness Framework. The end-user interacts with different tools to ascertain the fairness of the decision received.

Component 2 (**C2**) is a tool to check the fairness of recourses. The explanations E given to the user may point to changes that the user can apply to change the final result. Are those actions i.e. recourses fair for the user? The tool to check the fairness of the recourses will use the system S and the explanations E given to the user to check if the changes the user needs to make to receive a positive decision are fair.

The fairness measures adopted by components 1 and 2 should be able to verify the metrics reported by the organization and other metrics, as well as various combinations of attributes, thus allowing end-users to confirm their self-identity.

Component 3 (**C3**) is a contestation mechanism that allows the user to challenge the ADM. This tool will use the results of components 1 and 2 as well as the explanations E provided by the ADM and possibly a ground generation tool to establish an exchange of arguments with the system that need to provide justification to the individuals not just for the decision made but also for the process to obtain it, as well as the different elements that can be subject to a contestation.

Component 4 (**C4**) is a mechanism to report the organization to a regulatory entity and request an audit. If there is a conflict between the organization and the end-user, this channel serves as the end-user's final option, not to ascertain fairness, as it is already confirmed for the user, but to request validation of the decision from a regulatory entity and ultimately seek redress.

Figure 3 illustrates how these components interact with the stakeholders we have identified. *End-users* are the individuals affected by the decision of the ADM system. The end-users are responsible for questioning the decision to which they were subject and taking the necessary steps to challenge the systems and possibly reverse the decision. They should define the set of attributes $P_A \in A$ that correspond to the group with which they are identified and use a framework to verify whether it is discriminated against, taking into account this identity.

Organizations are the responsible o developing the ADM systems. Within organizations, the *practitioners* are members of organizations with different roles in the development and deployment of ADM systems. They are are not represented in the figure but are worth to mention for their role in the developing and deploying process. Organizations and practitioners are responsible for the implementation of all the mechanisms to avoid unfairness in the development of ADM systems and to disclose the fairness definitions used to evaluate the proposed solution. Organizations provide a mechanism to challenge their system; this mechanism should be different from the redress mechanism or may include it.

Regulatory Entities are competent authorities responsible for ensuring nondiscriminatory ADM systems. Regulatory entities should create mechanisms for appealing a decision to provide the user

with tools, external to the organizations, that can lead to an audit of the process by an auditing entity.

In addition to the components that will help the user verify the treatment received, Figure 3 shows an additional entity that could act on behalf of regulatory entities in two main forms: (1) an audit process made by an independent entity in the form of a certification that checks a specific requirement, fairness in this case; (2) an audit process triggered by a reclamation related to unfairness originated by the user. *Auditing Entities* are independent entities that perform conformity assessments in ADM systems. The auditing entities should be designated by these regulatory entities.

To ensure ascertainable fairness, an ADM system should meet certain requirements. It must allow unlimited petitions with new data input, providing decisions to facilitate fairness verification. The system should disclose all attributes used in decision making, ensuring transparency between user input and model features. It must provide users with explanations and recourses (minimal changes needed to alter decisions) enabling contestation. A redress mechanism must be in place to compensate affected users and implement corrective actions to prevent recurring unfair outcomes. Finally, the system must disclose its fairness criteria and report corresponding values, helping stakeholders to assess its fairness approach. These requirements collectively empower users to verify, contest and seek redress in ADM decisions.

6. Discussion

In this section, we examine our theoretical and practical contributions as well as the open areas and limitations of our approach. The proposed framework integrates fairness, explainability, contestability, and accountability into four components that enable users to ascertain fairness.

6.1. Theoretical and Practical Contributions

This paper advances the theoretical understanding and practical implementation of fairness in ADM systems. We distinguish between theoretical contributions that expand current concepts and practical contributions that provide implementable solutions.

As theoretical contributions, we introduce ascertainable fairness, treating fairness as an individual's right to access and verify decision-related information. We integrate fields such as algorithmic fairness, explainable AI, contestability, and accountability into a conceptual framework that expands the understanding of procedural fairness and provides a foundation for fairness verification mechanisms. Finally, we propose a user-centered fairness perspective, linking individual perceptions to collective identity, enabling fairness authentication based on personal context.

Our research offers two practical contributions. We propose four components combined in a framework for ascertainable fairness: a tool to verify prediction fairness, a mechanism to assess fairness of recourse, a contestation method, and an audit mechanism. These components are accompanied by specific requirements that ADM systems must meet to enable ascertainable fairness in practice and the necessary interactions between different stakeholders in the system. Moreover, we provide customized guidance for stakeholders: processes for end-users to verify and contest decisions, strategies for organizations to implement fairness and contestability, requirements for practitioners to build fairer systems, and guidance for regulators on auditing fairness claims.

6.2. Open Areas and Limitations

While the proposed conceptual framework provides a structured approach to ascertainable fairness, several challenges remain:

- **Fairness Metrics Standardization:** Current fairness measures lack a unified standard, leading to discrepancies between ADM providers and end-user expectations. Further research is needed to harmonize metrics in different applications.

- **User Literacy:** Fairness assessments and contestability mechanisms should be accessible to non-experts. Future work should focus on developing user-friendly tools that facilitate fairness verification without requiring deep technical knowledge.
- **Implementation of the components**: Components 1 and 2 of the framework can be implemented with current tools but need significant optimization to benefit end-users; ideally, they should be implemented by regulatory entities that develop standardized tools for fairness verification, similar to sandboxes for ADM system testing and validation. Component 3 is still unexplored to support users in formulating contestation requests and receiving appropriate justifications. With respect to component 4, effective auditing mechanisms must be standardized and integrated into the governance of ADM. Further efforts should define clear protocols for auditing fairness claims and handling disputes.
- **Human-In-The-Loop Considerations:** This framework primarily addresses fully automated decision-making. Future research should adapt it to hybrid systems where human oversight interacts with ADM systems, ensuring fairness in both automated and human-influenced decisions.

Despite these challenges, the proposed framework aligns with regulatory initiatives, such as the AI Act of the European Union [34], reinforcing transparency and accountability.

7. Conclusions

This work introduces the concept of ascertainable fairness, positioning fairness as an individual's right to access and verify decision-related information in ADM systems. Unlike traditional fairness approaches aimed at practitioners, this framework shifts the focus to empower end-users, allowing them to understand, contest, and verify AI-driven decisions.

Theoretically, the proposed framework integrates algorithmic fairness, explainability, contestability, and accountability into a user-centered fairness approach. Practically, it provides structured components: fairness of predictions, fairness of recourse, contestability mechanisms, and auditing tools, to operationalize fairness in ADM systems. However, while the framework offers actionable tools, additional research is needed to refine contestability mechanisms, standardize fairness metrics, and adapt solutions for human-in-the-loop decision systems.

Our proposal also benefits organizations by providing guidelines to ensure fairness, transparency, and accountability in ADM systems. For policymakers, it offers an approach to regulating ADM systems by defining fairness verification mechanisms, contestability processes, and audit requirements.

This work bridges the procedural and substantive dimensions of fairness and ensures that fairness is not only a technical property of ADM systems but a right that individuals can ascertain and uphold in practice. Allowing end-users to find an answer for *Am I being treated fairly?* employing a systematically organized framework of tools and processes improves transparency, thus increasing the trustworthiness of ADM systems.

Declaration on Generative AI

During the preparation of this work, the authors used Writefull to improve grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and assumed full responsibility for the content of the publication.

References

[1] L. Tangi, C. Van Noordt, M. Combetto, D. Gattwinkel, F. Pignatelli, AI Watch. European landscape on the use of Artificial Intelligence by the Public Sector, Technical Report, Publications Office of the European Union, Luxembourg, 2022. doi:10.2760/39336.

[2] C. Lahusen, M. Maggetti, M. Slavkovik, Trust, trustworthiness and AI governance, Scientific Reports 14 (2024) 20752. doi:10.1038/s41598-024-71761-0.
[3] C. Castelluccia, D. L. Métayer, Understanding algorithmic decision-making: Opportunities and challenges, Technical Report, European Parliament Study, 2019. URL: https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2019)624261.
[4] U. Bergmann, C. Bonefeld-Dahl, V. Dignum, J.-F. Gagné, T. Metzinger, N. Petit, S. Steinacker, A. V. Wynsberghe, K. Yeung, Ethics guidelines for trustworthy AI, Technical Report, High-Level Expert Group on Artificial Iintelligence (AI HLEG) set up by the European Commission, 2019. URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
[5] L. P. Robert, C. Pierce, L. Marquis, S. Kim, R. Alahmad, Designing fair ai for managing employees in organizations: a review, critique, and design agenda, Human–Computer Interaction 35 (2020) 545–575. doi:10.1080/07370024.2020.1735391.
[6] H. Nieminen, Why We Need Epistemic Rights, Springer International Publishing, Cham, 2024, pp. 11–28. doi:10.1007/978-3-031-45976-4_2.
[7] S. Barocas, M. Hardt, A. Narayanan, Fairness and Machine Learning: Limitations and Opportunities, fairmlbook.org, 2019. http://www.fairmlbook.org.
[8] J. Rawls, A Theory of Justice, Harvard University Press, Cambridge, MA, 1971.
[9] T. R. Tyler, Why People Obey the Law, Yale University Press, New Haven, CT, 1990.
[10] European Parliament, Charter of Fundamental Rights of the European Union, 2012. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A12012P%2FTXT.
[11] D. Pessach, E. Shmueli, A review on fairness in machine learning, ACM Computing Surveys 55 (2022) 1–44. doi:10.1145/3494672.
[12] K. Makhlouf, S. Zhioua, C. Palamidessi, Machine learning fairness notions: Bridging the gap with real-world applications, Information Processing and Management 58 (2021). doi:10.1016/j.ipm.2021.102642.
[13] S. Feuerriegel, M. Dolata, G. Schwabe, Fair AI: Challenges and Opportunities, Business and Information Systems Engineering 62 (2020) 379–384. doi:10.1007/s12599-020-00650-3.
[14] A. Castelnovo, R. Crupi, G. Greco, D. Regoli, I. G. Penco, A. C. Cosentini, A clarification of the nuances in the fairness metrics landscape, Scientific Reports 12 (2022) 4209. doi:10.1038/s41598-022-07939-1.
[15] R. K. E. Bellamy, K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilovic, S. Nagar, K. N. Ramamurthy, J. Richards, D. Saha, P. Sattigeri, M. Singh, K. R. Varshney, Y. Zhang, Ai fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias, 2018. doi:10.48550/ARXIV.1810.01943.
[16] M. C. Decker, L. Wegner, C. Leicht-Scholten, Procedural fairness in algorithmic decision-making: the role of public engagement, Ethics and Information Technology 27 (2024) 1. doi:10.1007/s10676-024-09811-4.
[17] N. Díaz-Rodríguez, J. D. Ser, M. Coeckelbergh, M. L. de Prado, E. Herrera-Viedma, F. Herrera, Connecting the dots in trustworthy artificial intelligence: From ai principles, ethics, and key requirements to responsible ai systems and regulation, Information Fusion 99 (2023). doi:10.1016/j.inffus.2023.101896.
[18] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, Information Fusion 58 (2020) 82–115. doi:10.1016/j.inffus.2019.12.012.
[19] G. Tang, W. Tan, M. Cai, Privacy-preserving and trustless verifiable fairness audit of machine learning models, International Journal of Advanced Computer Science and Applications 14 (2023). doi:10.14569/IJACSA.2023.0140294.
[20] E. Toreini, M. Mehrnezhad, A. van Moorsel, Fairness as a service (faas): verifiable and privacy-preserving fairness auditing of machine learning systems, International Journal of Information Security 23 (2024) 981–997. doi:10.1007/s10207-023-00774-z.
[21] K. Dowding, B. R. Taylor, Algorithmic decision-making, agency costs, and institution-based trust,

Philosophy & Technology 37 (2024) 68. doi:10.1007/s13347-024-00757-5.
[22] H. Lyons, E. Velloso, T. Miller, Conceptualising contestability: Perspectives on contesting algorithmic decisions, Proceedings of the ACM on Human-Computer Interaction 5 (2021). doi:10.1145/3449180.
[23] H. Lyons, E. Velloso, T. Miller, Fair and responsible ai: A focus on the ability to contest, 2021. Preprint at https://doi.org/10.48550/arXiv.2102.10787.
[24] M. Almada, Human intervention in automated decision-making: Toward the construction of contestable systems, in: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, ICAIL '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 2–11. doi:10.1145/3322640.3326699.
[25] R. D. Ashmore, K. Deaux, T. McLaughlin-Volpe, An organizing framework for collective identity: Articulation and significance of multidimensionality., Psychological Bulletin 130(1) (2004) 80–114. doi:10.1037/0033-2909.130.1.80.
[26] M. J. Hornsey, J. Jetten, The individual within the group: balancing the need to belong with the need to be different., Personality and social psychology review : an official journal of the Society for Personality and Social Psychology 8(3) (2004) 248–264. doi:10.1207/s15327957pspr0803_2.
[27] S. Sharma, J. Henderson, J. Ghosh, Certifai: A common framework to provide explanations and analyse the fairness and robustness of black-box models, in: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, AIES '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 166–172. doi:10.1145/3375627.3375812.
[28] S. Goethals, D. Martens, T. Calders, Precof: counterfactual explanations for fairness, Machine Learning (2023). doi:10.1007/s10994-023-06319-8.
[29] A. Kuratomi, E. Pitoura, P. Papapetrou, T. Lindgren, P. Tsaparas, Measuring the burden of (un)fairness using counterfactuals, in: I. Koprinska, P. Mignone, R. Guidotti, S. Jaroszewicz, H. Fröning, F. Gullo, P. M. Ferreira, D. Roqueiro, G. Ceddia, S. Nowaczyk, J. Gama, R. Ribeiro, R. Gavaldà, E. Masciari, Z. Ras, E. Ritacco, F. Naretto, A. Theissler, P. Biecek, W. Verbeke, G. Schiele, F. Pernkopf, M. Blott, I. Bordino, I. L. Danesi, G. Ponti, L. Severini, A. Appice, G. Andresini, I. Medeiros, G. Graça, L. Cooper, N. Ghazaleh, J. Richiardi, D. Saldana, K. Sechidis, A. Canakoglu, S. Pido, P. Pinoli, A. Bifet, S. Pashami (Eds.), Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Springer Nature Switzerland, Cham, 2023, pp. 402–417.
[30] V. Gupta, P. Nokhiz, C. D. Roy, S. Venkatasubramanian, Equalizing recourse across groups, 2019. Preprint at http://arxiv.org/abs/1909.03166.
[31] J. v. Kügelgen, A.-H. Karimi, U. Bhatt, I. Valera, A. Weller, B. Schölkopf, On the fairness of causal algorithmic recourse, Proceedings of the AAAI Conference on Artificial Intelligence 36 (2022) 9584–9594. doi:10.1609/aaai.v36i9.21192.
[32] F. Leofante, H. Ayoobi, A. Dejl, G. Freedman, D. Gorur, J. Jiang, G. Paulino-Passos, A. Rago, A. Rapberger, F. Russo, X. Yin, D. Zhang, F. Toni, Contestable ai needs computational argumentation, 2024. Preprint at https://arxiv.org/abs/2405.10729.
[33] J. E. Rodríguez, K. M. Campbell, Racial and Ethnic Disparities in Prevalence and Care of Patients With Type 2 Diabetes, Clinical Diabetes 35 (2017) 66–70. doi:10.2337/cd15-0048.
[34] European Parliament, Regulation (eu) 2024/1689 of the european parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence and amending regulations (ec) no 300/2008, (eu) no 167/2013, (eu) no 168/2013, (eu) 2018/858, (eu) 2018/1139 and (eu) 2019/2144 and directives 2014/90/eu, (eu) 2016/797 and (eu) 2020/1828 (artificial intelligence act) (text with eea relevance), 2024. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.