Cognitive Computing and Internet of Things, Vol. 165, 2025, 198–206
https://doi.org/10.54941/ahfe1006103

The Algorithmic Fairness Challenge in Decision-Making

Herwig Zeiner[1], Lucas Paletta[1], and Gustavo Vieira[2]

[1]Joanneum Research Forschungsgesellschaft mbH, 8010 Graz, Austria
[2]Collaborative Laboratory Mountains of Research, Intelligent Technologies, Bragança, Portugal

ABSTRACT

The increasing use of automated decision-making systems has given rise to concerns about fairness. This paper examines the major principles of fairness in decision-making, and it discusses the challenges of implementing fairness principles in practice, such as the trade-offs between different types of fairness and the difficulty of measuring fairness. Finally, the paper puts forward several proposals for promoting fairness in decision-making. These include the use of transparent and explainable algorithms, the involvement of stakeholders in the design of decision-making systems, and the establishment of accountability mechanisms. It is of the utmost importance that fairness is a fundamental principle in decision-making processes. This approach is designed to ensure that individuals facing similar circumstances are treated equally and not subjected to discrimination. Examples of unfair decision-making include situations where individuals are discriminated against based on protected attributes such as race, gender, or age. Decisions that lack transparency in their process may be perceived as biased or unjust. Unfairness can arise in a number of ways, for example when promotions are based on favouritism rather than merit, or when hiring decisions are influenced by personal biases rather than qualifications.

Keywords: Decision making, Fairness, User-centered

INTRODUCTION

First, it is clear that many decisions are made by groups. The theory of social choice (Sen, 1986) deals with this fundamental aspect, namely the aggregation of individual preferences of group members into a collective decision. The question to be answered is: What makes a collective decision a good, i.e. "fair" decision? The goal is to gain a deeper understanding of collective decision making in order to address new technological and social challenges where aspects of decision making and fairness are important. We need to start with the preferences of individuals, machines, or criteria for a set of discrete objects. Then we need to make a "fair" group decision. A key problem of most previous studies is the limited availability of actual preference information. The available information from elections or group decisions is usually limited to the data collected during the voting process. Often, these are only individual alternatives from a relatively large set of alternatives, as in the case of majority voting, where each person can cast

exactly one vote for an alternative. The underlying complete preferences (e.g. the complete ranking of the alternatives) are usually not even recorded. Therefore, it is difficult to understand or even justify whether the collective result really represents some kind of "collective will".

We also explore the field of algorithmic fairness and its goals. To illustrate the importance of this field, we present examples of unfair models and their effects. We discuss the current state and future challenges in meeting the challenges of fair algorithmic decision making. The article explores how biases in the data used to train these algorithms can perpetuate unfairness in real-world decisions (Tolan, 2019). However, the use of automated decision-making algorithms can have unintended effects that lead to discrimination against certain groups (e.g., in the workload example for example as described in a transport logistics application (Zeiner, 2023)). In this context, it is crucial to develop AI services that are not only accurate, but also fair. A study on perceptions of fairness and trust in automated decision-making examines the relationship between people's trust in automated decision-making systems and their understanding of how these systems work. The study shows that a lack of transparency can lead people to question the fairness of such systems.

The research presented in this article is part of the EU-funded FAIRWork project. More details of the concept and approach can be found in (Woitsch, 2023) and (Woitsch, 2024). In this project, we aim to develop the Democratic AI-based Decision Support System to improve decision making in production environments. This system will not only optimize the production process, but also suggest decisions and resource allocations that are "fair" to the users, taking into account their preferences and individual situations. This includes presenting the results in a way that allows workers to understand and trust the proposed decisions. This fosters acceptance in a broader sense.

RELATED WORK

Defining and measuring fairness in resource distribution is a multifaceted challenge with no one-size-fits-all solution (Kim, 2021). The increasing use of AI in various sectors, from healthcare and transportation to loan applications and college admissions, necessitates the development of algorithms that are not only accurate but also objective and fair. However, algorithmic decision-making can be inherently prone to unfairness, even when unintentional, as algorithms may learn and perpetuate historical biases present in the data

The increasing use of algorithms in critical decision-making domains has brought the problem of algorithmic injustice to the fore. This unfairness can manifest itself in discriminatory or biased decisions (Mehrpouyan, 2022), which may stem from various sources, including biased training data, flawed problem formulations, or inherent biases in algorithm designs. There is no universal definition of fairness, but concepts such as group fairness, individual fairness and counterfactual fairness are frequently discussed. A central dilemma is the trade-off between fairness and accuracy of algorithms, since reducing unfairness in some cases can compromise overall accuracy.

To address algorithmic unfairness, various techniques are used that can be grouped into three main categories: pre-processing techniques modify the data before training to reduce bias. In-processing techniques integrate fairness constraints directly into the training process. Post-processing methods adjust the output of a trained model to improve fairness. Selecting the appropriate technique and evaluating success depends heavily on the specific requirements of the use case and the underlying societal values. Various metrics are available for evaluating fairness, including demographic parity, equal opportunities and equity, although the choice of the appropriate metric is context-dependent.

Current research in the field of algorithmic fairness is very active and focuses on more advanced concepts such as causality and intersectionality in order to better understand the causes of injustice and develop more effective solutions. Other important research areas include dynamic fairness in sequential decision-making processes, the development of explainable and interpretable fairness models, and the application of fairness principles in complex AI systems (Corbett-Davies, 2023) or (Suárez Ferreira, 2025). Challenges remain in operationalising fairness in real-world applications, balancing fairness with other performance goals, and considering ethical and societal implications.

The field of algorithmic fairness is developing rapidly (Amanatidis, 2023). Although significant progress has been made, many questions remain. Future research efforts will focus on developing more robust, interpretable and context-sensitive fairness techniques that are not only technically sound but also ethically sound and socially relevant. The transfer of fairness research into practice and the examination of the broader social implications of algorithmic decisions will be central tasks in this regard.

METHODOLOGY

Our design process, as a general feature, adopts a service-based approach. **A significant innovation lies in the integration of Multi-Agent Systems (MAS) that incorporate fairness aspects** into decision-making for workload balance, surpassing the current state-of-the-art. Unlike traditional systems focused solely on productivity, MAS enables a human-centered paradigm by modeling diverse stakeholders and their interactions, aligning resource allocation decisions with both production goals and worker well-being. By leveraging autonomous, decentralized processes, MAS simulates complex social interactions and integrates a broad spectrum of inputs, including human conditions, preferences, and well-being metrics. This fosters decision-making that is not only efficient but also fair, adaptive, and inclusive, aligning with Industry 5.0's vision of socially responsible and technologically advanced industrial environments.

This **novel application of MAS represents a paradigm shift,** moving beyond operational optimization to prioritize equitable workload distribution and increased worker satisfaction. MAS enhances human participation in governance processes, creating democratic decision frameworks even in constrained industrial settings by balancing relevant human considerations

with production demands. It demonstrates the potential to redefine industrial decision-making by harmonizing technological efficiency with human-centric values, paving the way for sustainable and socially responsible advancements in resource allocation and workforce management.

A critical aspect of implementing MAS, particularly in the context of human-centered decision-making, is ensuring algorithmic fairness. Algorithms, by their nature, can perpetuate or even amplify existing societal biases if not carefully designed. Fairness in algorithms addresses the challenge of mitigating these biases to ensure equitable outcomes for all stakeholders. This involves considering various definitions of fairness, such as statistical parity, equality of opportunity, and predictive value parity, and selecting the most appropriate criteria for the specific application. For instance, in workload distribution, fairness might mean ensuring that no particular group of workers consistently bears a disproportionate burden, regardless of their demographic characteristics.

To achieve algorithmic fairness, it is essential to incorporate fairness-aware techniques (Paletta, 2024). These techniques focus on developing algorithms that explicitly account for fairness considerations during the training process. This can involve pre-processing data to remove biases, modifying the learning objective to include fairness constraints, or post-processing the algorithm's outputs to correct for unfair outcomes. Furthermore, transparency and explain ability are crucial. By making the decision-making process of the algorithm more transparent, it becomes possible to identify and address potential sources of bias, fostering trust and accountability. It also allows for human oversight and intervention, ensuring that the algorithm's decisions align with ethical and societal values.

Additionally, the development of ethical watchdogs within MAS represents an innovation that introduces mechanisms for embedding ethical oversight directly into decentralized decision-making processes. These watchdogs act as autonomous agents designed to monitor and alert to the ethical implications of decisions made within the system, ensuring alignment with predefined human-centric values and societal norms. Ethical watchdogs help mitigate biases and safeguard against unintended consequences in resource allocation or workload distribution, where computer agents and humans (Gal, 2022) share decision-making in order to address ethical conflicts (Belloni, 2015). This capability allows MAS to uphold ethical standards in dynamic and complex industrial settings. The integration of the watchdog advances the field by operationalizing ethical principles in decision-making, offering a novel approach to fostering accountability and trust (Woodgate, 2022) in technology-driven environments, and reinforcing the alignment of industrial processes with broader societal and human values.

USE CASE SCENARIO FOR FAIRNESS ASPECTS

Before the start of each work period, it provides a proposal for the deployment of personnel, taking into account relevant information on the available team, the precise requirements of the activities to be carried out and the characteristics of the products and assembly lines in

question. A compliance monitor, designed to monitor key fairness principles parameters in the selection process, signals any breach of these parameters, enabling the operator to initiate a re-assignment of the staff.

This information is processed by the engine in order to identify a sound deployment scheme that complies with the defined limits to meet the established ethical standards, taking into account the wide range of potential choices, subject to constraints and objectives. The agent-based resource allocation function demonstrated its ability to support selection processes in industrial workforce management by flexibly matching team member profiles with production line requirements.

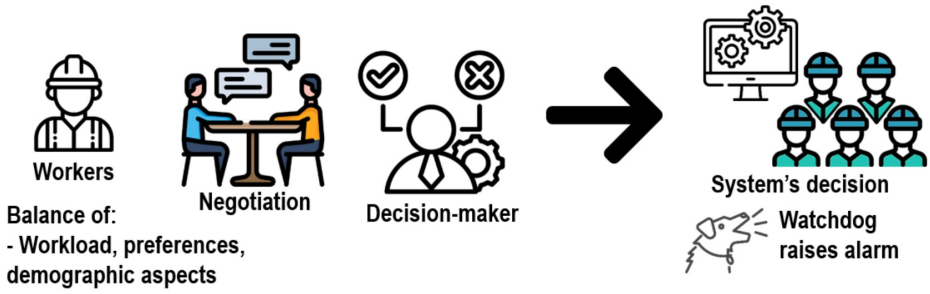

Figure 1: The watchdog raises arlam.

The system consistently generated deployment proposals by ordering team members based on flexibility and preference metrics. This method not only facilitated resource allocation through agent interaction and negotiation, but also ensured that human-centred values were maintained throughout the selection process, validating the system's utility in complex industrial environments. The creation of the agent-driven distribution process in an industrial manufacturing context has demonstrated potential improvements in workload balancing and selection process support.

The method incorporates principles that emphasise fairness, consideration of team members, and digital representation of relevant factors in the distribution process. Prior to assignments, factors such as team member availability and physical condition where operational requirements are assessed to determine team distribution. Monitoring systems are in place to track and modify team assignments based on ethical considerations. An agent-based method takes into account factors that influence task suitability, involving the representation of human participants in an interactive selection process. If a team member does not meet the criteria required for a given task, alternative assignments are explored. The process can be dynamically modified based on significant parameters. The selection process culminates in an integrated approach that combines different metrics to promote fairness and balance. Team members' preferences for different tasks are incorporated into the allocation process, ensuring that individual needs are considered alongside task compatibility. In addition, the inclusion of negotiation protocols increases accountability and strengthens a reliable distribution framework.

Figure 2: Renegotiation to comply with the watchdog fairness principle.

Overall, the incorporation of human-centred considerations into an agent-based selection methodology has provided support for human resource management. The method provides insights into balancing operational needs with team member factors, contributing to more informed selection in dynamic environments.

IMPLEMENTATION MECHANISMS FOR PROMOTING FAIRNESS

The relevance of Multi-Agent Systems (MAS) has significantly influenced resource allocation strategies in the manufacturing sector. The integration of MAS into Decision Support Systems (DSS) holds substantial promise for fostering more informed, robust, and democratic decision-making in industrial contexts. MAS can address challenges related to coordination, conflict resolution, and human-centric considerations in production environments, encapsulating complex interactions and stakeholder perspectives. In contrast to traditional approaches that focus solely on fulfilling production requirements, MAS introduces a broader perspective by incorporating stakeholders' inputs and negotiation mechanisms in the decision process. Among these inputs are the worker resilience and preferences to be allocated in production lines. It dynamically recommends workers to production lines while considering parameters such as availability, medical constraints, experience, and physical resilience. Integrating worker preferences and human factors into agents' decision logic elevates the role of workers from passive recipients to active participants.

In this framework, worker agents store individual attributes (e.g., availability, medical constraints, and preferences for specific tasks), and production line agents specify requirements for staffing levels, deadlines, and skill sets. Communication occurs via the Contract Net Protocol, wherein production line agents issue calls for proposals, and worker agents respond with bids calculated from weighted metrics of resilience and preference. Upon receiving bids, production line agents negotiate optimal allocations through iterative exchanges, thereby accommodating varying levels of production demand and worker preferences.

In any system where multiple stakeholders interact, conflicts are inevitable. MAS architecture inherently supports negotiation and conflict resolution through algorithms that address competing goals. The decision-making

platform can systematically mediate among multiple stakeholders, leading to outcomes that are both effective and equitable.

In the event of identical scores among workers (e.g., equally suitable in terms of resilience and preference), the system resolves conflicts by referencing auxiliary attributes such as job rotation frequency and line-specific experience. This ensures that no single individual is repeatedly tasked with the same role in ways that might undermine fairness. Furthermore, the MAS architecture supports scalability; additional production line agents or worker agents can enter or exit as needed, allowing the system to maintain robust performance under changing operational conditions.

Democratic decision-making hinges on mechanisms that adequately represent each stakeholder's perspective. In this strategy, agent-based modeling ensures that workers, managers, and other relevant entities each operate as individual agents with distinct goals. This distributed control structure not only promotes fairness but also enables the negotiation of outcomes that reflect diverse interests and constraints.

DISCUSSION AND MONITORING ETHICAL ASPECTS

MAS-based methods account for human-centric criteria, but they do not inherently guarantee compliance with broader ethical or demographic goals. A watchdog agent must continuously monitor allocation outcomes against predefined ethical criteria (e.g. minimum gender balance or maximum average age thresholds). If any criterion is breached, the watchdog raises an alarm to signal potential ethical infractions. The decision-maker must then initiate an outcome renegotiation. During this phase, the relevant ethical constraints, initially disregarded or insufficiently addressed, are incorporated into the allocation process. The system then recalculates bids and proposals to fulfil both operational requirements and the ethical considerations.

For example, the initial allocation recommended only male workers for a particular production line, violating a rule stipulating that at least 50% of the assigned workforce must be female. The watchdog flagged this outcome, allowing the decision-maker to prompt a renegotiation for worker and line agents in order to fulfil the ethical requirements for the use case. The revised allocation complied with the demographic requirement without sacrificing worker resilience or preference. The same applies in scenarios where age-based constraints are in place. The watchdog makes sure that allocations remain within acceptable ethical boundaries. The watchdog agent provides feedback on allocation decisions to ensure greater transparency in workforce management. Decision-makers retain control over whether to accept the initial recommendation or act on the watchdog's alerts. This process, in which human oversight complements automated negotiation, fosters accountability and trust in the system's recommendations.

CONCLUSION

The approach to identifying decision-making problems, including fairness principles and their implementation using this overall approach, is then presented. The use of fairness principles is key to allocating the necessary

resources differently and ensuring compliance with some principles. The approach is flexible, allowing for efficient adaptation of resource allocation decisions. Our approach increases user trust and understanding, leading to greater acceptance of the decisions made, all while incorporating the fairness principles. The combined use of MAS-based resource allocation and the watchdog agent offers a structured approach to promoting fairness. The MAS framework tailors resource assignments by incorporating both operational demands and human-centered metrics, while the watchdog agent enforces compliance with higher-level ethical or demographic guidelines. These mechanisms enhance transparency, encourage inclusive decision-making, and underscore the importance of aligning industrial processes with equitable and socially responsible principles.

Future work will go beyond generic definitions to consider the specific needs and circumstances of different application domains. This will include considering the distinction between fairness as a property of the algorithm and justice as a property of the allocation principle and developing methods that address both aspects in the decision-making process. This will require the development of methods for detecting and preventing manipulation of fairness metrics and carefully evaluating the potential impacts of fairness interventions.

ACKNOWLEDGMENT

This work has been supported by the *FAIRWork* project (www.fairwork-project.eu) and has been funded within the European Commission's Horizon Europe Programme under contract number 101069499. This paper expresses the opinions of the authors and not necessarily those of the European Commission. The European Commission is not liable for any use that may be made of the information contained in this paper.

REFERENCES

Amanatidis, G., Aziz, H., Birmpas, G., Filos-Ratsikas, A., Li, B., Moulin, H.,... & Wu, X. (2023). Fair division of indivisible goods: Recent progress and open questions. *Artificial Intelligence*, *322*, 103965.

Belloni, A., Berger, A., Boissier, O., Bonnet, G., Bourgne, G., Chardel, P.-A., … Others. (2015). Dealing with ethical conflicts in autonomous agents and multi-agent systems. Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence.

Corbett-Davies, S., Gaebler, J. D., Nilforoshan, H., Shroff, R., & Goel, S. (2023). The measure and mismeasure of fairness. *Journal of Machine Learning Research*, *24*(312), 1–117.

Gal, K., & Grosz, B. J. (05 2022). Multi-Agent Systems: Technical & Ethical Challenges of Functioning in a Mixed Group. Daedalus, 151(2), 114–126. doi: 10.1162/daed_a_01904.

Kim, P. T., Gummadi, K. P., & Loiseau, P. (2021). Intersectional fairness in algorithmic decision making: A comparative study. In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT) (pp. 429–440).

Liu, S., Lu, X., Suzuki, M., & Walsh, T. (2024). Mixed fair division: A survey. *Journal of Artificial Intelligence Research*, *80*, 1373-1406.

Mehrpouyan, P., De Cremer, D., & Rahwan, I. (2022). Algorithmic fairness: from group to individual and beyond. Nature Machine Intelligence, 4(5), 377–389

Noriega-Campero, A., Bakker, M. A., Garcia-Bulle, B., & Pentland, A. S. (2019, January). Active fairness in algorithmic decision making. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (pp. 77–83).

Paletta, L., Zeiner, H., Schneeberger, M., Pszeida, M., Mosbacher, J. A., & Tschuden, J. (2024, September). Resilience Scores for Decision Support Using Wearable Biosignal Data with Requirements on Fair and Transparent AI. In *2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA)* (pp. 1–4). IEEE.

Sen, A. (1986). Social choice theory. Handbook of mathematical economics, 3, 1073–1181.

Tolan, S. (2019). Fair and unbiased algorithmic decision making: Current state and future challenges. arXiv preprint arXiv:1901.04730.

Suárez Ferreira, J., Slavkovik, M. & Casillas, J. General procedure to measure fairness in regression problems. *Int J Data Sci Anal* (2025). https://doi.org/10.1007/s41060-025-00721-2

Woitsch, R., Muck, C., Utz, W., & Zeiner, H. (2023, September). Towards a democratic AI-based decision support system to improve decision making in complex ecosystems. Joint Proceedings of the BIR 2023 Workshops and Doctoral Consortium co-located with 22nd International Conference on Perspectives in Business Informatics Research (BIR 2023). CEUR Workshop Proceedings. Retrieved from https://ceur-ws.org/Vol-3514/paper94.pdf

Woitsch, R., Muck, C., Utz, W., & Zeiner, H. (2024). Enable Flexibilisation in FAIRWork's Democratic AI-based Decision Support System by Applying Conceptual Models Using ADOxx. *Complex Systems Informatics and Modeling Quarterly*, (38), 27–53.

Woodgate, J., & Ajmeri, N. (2022). Macro Ethics for Governing Equitable Sociotechnical Systems. Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems, 1824–1828. Presented at the Virtual Event, New Zealand. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Zeiner, H., Unterberger, R., Tschuden, J., & Quadri, M. Y. (2023, May). Time-aware optimisation models for hospital logistics. In *International Conference on Decision Support System Technology* (pp. 45–55). Cham: Springer Nature Switzerland.