

Association for Information Systems

AIS Electronic Library (AISeL)

ICIS 2024 TREOS

AIS TREO Papers

12-15-2024

Are AI practitioners ready for AI Fairness? The need for Institutional Work for Early Prioritization of Fairness in AI Practices

Pouria Akbari Ghatar

University of Agder, pouria.akbarighatar@uia.no

Ilias Pappas

Norwegian University of Science and Technology, ilias.pappas@uia.no

Polyxeni Vassilakopoulou

University of Agder, polyxenv@uia.no

Follow this and additional works at: https://aisel.aisnet.org/treos_icis2024

Recommended Citation

Akbari Ghatar, Pouria; Pappas, Ilias; and Vassilakopoulou, Polyxeni, "Are AI practitioners ready for AI Fairness? The need for Institutional Work for Early Prioritization of Fairness in AI Practices" (2024). *ICIS 2024 TREOS*. 61.

https://aisel.aisnet.org/treos_icis2024/61

This material is brought to you by the AIS TREO Papers at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2024 TREOS by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Are AI practitioners ready for AI Fairness?

The need for Institutional Work for Early Prioritization of Fairness in AI Practices

Pouria Akbarighatar (pouriaa@uia.no); Ilias Pappas (ilias.pappas@uia.no);

Polyxeni Vassilakopoulou (polyxenv@uia.no)

Artificial Intelligence (AI) systems can reproduce or introduce biases, raising fairness concerns (Barocas et al. 2023). Data used for AI training often reflect societal biases perpetuating unfairness. Mislabeling, data gaps, and erroneous entries may introduce new biases. Algorithm choices can introduce biases due to feature selection or optimization criteria. Even more importantly, fairness is a societal issue as much as a technical one; an archetypical sociotechnical problem (Dolata et al. 2022). AI models are used in different contexts where what is fair depends on multiple factors and requires normative judgments. Different definitions of fairness can conflict with each other and result in different outcomes (Teodorescu, et al. 2021). Addressing fairness should start early to allow for defining fairness goals and aligning data and algorithms. We present findings from surveying 130 AI professionals from Europe and the US with significant AI experience (68% more than 5 years) and education (66% at least a master's). We asked about different Responsible AI principles and found that many view fairness as an afterthought rather than a foundational part of responsible AI development (Table 1). This is striking as fairness is difficult and costly to address retroactively. Organizations need to transition to new AI practices

	C-level managers			AI developers		
	address first	address after initial work	address in later stages	address first	address after initial work	address in later stages
Benevolence	82%	14%	4%	71%	27%	2%
Transparency	37%	61%	2%	25%	70%	5%
Fairness	37%	32%	31%	22%	25%	53%

Table 1. Perspectives of AI practitioners on prioritizing different Responsible AI principles

Building on the concept of “institutional work” (Lawrence and Suddaby 2006) we explore the following Research Question: What forms of institutional work are needed to disrupt current AI practices and prioritize fairness early in AI development processes? We suggest that the required change in AI practices is not just a matter of adjusting Information Systems development lifecycle models but also requires a mindset shift requiring work to reconfigure belief systems, alter abstract categorizations, and enact new rules.

References

- Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. MIT press.
- Dolata, M., Feuerriegel, S., & Schwabe, G. (2022). A sociotechnical view of algorithmic fairness. *Information Systems Journal*, 32(4), 754-818.
- Lawrence, T. B. and Suddaby, R. (2006), 'Institutions and institutional work. ', in S. R. Clegg, et al. (eds.), *Handbook of organization studies.*; London: Sage, 215-54.
- Teodorescu, M. H., Morse, L., Awwad, Y., & Kane, G. C. (2021). Failures of Fairness in Automation Require a Deeper Understanding of Human-ML Augmentation. *MIS Quarterly*, 45(3), 1483-1500.