

The Challenge of Bias Mitigation in Clinical AI Decision Support: A Balance Between Decision Efficiency and Quality

NIELS VAN BERKEL, SANDER DE JONG, JOEL WESTER, and NAJA KATHRINE KOLLERUP ALS, Aalborg University, Denmark

An increasing number of intelligent data-driven health systems seek to support patients and clinicians in decision making tasks. However, the recommendations provided by such systems can negatively impact the reasoning abilities of its users, giving rise to cognitive biases. Such mental processes can subsequently harm the quality of the user’s decision. While decision support systems are typically designed to increase user efficiency, known approaches to mitigate such biases primarily rely on slowing down the decision making process—offsetting any efficiency benefits. This position paper calls attention to the efficiency–quality trade-off in bias mitigation and outlines a future research direction for bias mitigation in AI decision support.

1 BIAS MITIGATION IN AI DECISION SUPPORT

Bertrand et al. reviewed the relation between AI-assisted decision making and cognitive biases [1]. While some biases can in fact be mitigated by decision support (e.g., reliance on representativeness heuristic), other biases are likely to be exacerbated as the result of AI suggestions. This includes well-known biases such as automation bias and confirmation bias, as explained by the reliance that users—either consciously or subconsciously—put in these recommendations.

Recent work in HCI has begun to systematically study the effect of cognitive biases, as well as explore new ways to mitigate end-user bias. Healthcare is a common focus in these studies due to the potential impact of clinical care decisions. Solomon ran a deception study in which participants were led to believe they could customise the recommendation algorithm in a decision support system [4]. Regardless of the accuracy of the algorithm, those who believed to have customised the algorithm were biased in their decisions and therefore more likely to follow its recommendations. Buçinca et al. showed how cognitive forcing, in which people push themselves to put in additional reflection, can reduce reliance on AI [2]. In their implementation, users were faced with either a thirty-second timeout prior to the AI recommendation or had to actively request the recommendation, thereby necessitating cognitive forcing. Bach et al. highlight the challenges encountered when implementing bias mitigation techniques [3]. In an explorative study on ophthalmologists’ perceptions of bias mitigation in their AI-assisted decision support tool, Bach et al. encountered substantial concern over a decrease in their efficiency. This study evaluated three established and distinct bias mitigation techniques, showcasing their limited applicability in real-world scenarios.

2 BALANCING THE EFFICIENCY–QUALITY TRADE-OFF

Healthcare systems around the globe are under increasing pressure due to growing patient numbers and rising costs. Meanwhile, AI-enabled decision support is introduced to aid both patients and clinicians in making life-changing decisions and increase overall efficiency. While recent work in the HCI community shows a growing awareness of the potential for cognitive biases to arise as a result, the current literature is short on solutions that have real-world potential to mitigate these biases. In complex time-pressured and multi-stakeholder contexts, as encountered in healthcare settings, currently evaluated bias mitigation techniques will result in frustration and abandonment [3]. Following this, we call on the research community to explore bias mitigation techniques as integrated into AI-enabled decision support systems in ecologically valid settings.

Authors’ address: Niels van Berkel, nielsvanberkel@cs.aau.dk; Sander de Jong, sanderdj@cs.aau.dk; Joel Wester, joelw@cs.aau.dk; Naja Kathrine Kollerup Als, nkka@cs.aau.dk, Aalborg University, Aalborg, Denmark.

REFERENCES

- [1] Astrid Bertrand, Rafik Belloum, James R. Eagan, and Winston Maxwell. 2022. How Cognitive Biases Affect XAI-Assisted Decision-Making: A Systematic Review. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (Oxford, United Kingdom) (AIES '22). Association for Computing Machinery, New York, NY, USA, 78–91. <https://doi.org/10.1145/3514094.3534164>
- [2] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 188 (2021), 21 pages. <https://doi.org/10.1145/3449287>
- [3] Anne Kathrine Petersen Bach, Trine Munch Nørgaard, Jens Christian Brok, and Niels van Berkel. 2023. “If I Had All the Time in the World”: Ophthalmologists’ Perceptions of Anchoring Bias Mitigation in Clinical AI Support. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems* (CHI’23). <https://doi.org/10.1145/3544548.3581513>
- [4] Jacob Solomon. 2014. Customization Bias in Decision Support Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ’14). Association for Computing Machinery, New York, NY, USA, 3065–3074. <https://doi.org/10.1145/2556288.2557211>