Assessing appropriate reliance: a framework for evaluating AI influence on user decision-making

Caterina Fregosi[1,*], Andrea Campagner[2], Chiara Natali[1] and Federico Cabitza[1,2]

[1]Department of Informatics, Systems and Communication, University of Milano-Bicocca, Milan, Italy
[2]IRCCS Ospedale Galeazzi - Sant'Ambrogio, Milan, Italy

Abstract

Human-Computer Interaction (HCI) has traditionally focused on the concept of use, examining how humans interact with and benefit from technological systems. However, this notion alone fails to capture the impact that technology, particularly AI in critical domains, has on human cognition, behavior, and ethical responsibilities. This paper explores the concept of "Appropriate Reliance" (AR) where users accurately assess AI capabilities without over-relying (misusing) or dismissing (disusing) the system: that is, AR refers to the human capability to discern when to trust the machine's decisions and when to override them based on their likely accuracy. Optimizing this dimension is essential, as high machine accuracy is useless if users do not rely on its advice. However, most existing metrics focus on human-AI agreement rather than appropriate reliance and do not account for the complex interaction processes behind decision-making. In this paper, we conduct a comprehensive review of the metrics in the field, assessing their effectiveness in evaluating AR. We identified the most useful metrics and introduced new ones tailored to comprehensively assess the impact of AI on user decision-making beyond the effect of chance on post-hoc agreement. These include, among others, metrics to assess appropriate reliance, automation bias and conservatism bias, as well as metrics that conceptualize and quantify the influence of AI systems. All together these metrics compose a metrics-based framework 1 that shifts the focus from reliance to "influence", assessing the extent AI systems shape user decisions. These metrics were applied in four user studies conducted in the medical field, and we discuss the insights derived from these experiments. The findings emphasize the need for designers and researchers to shift from reliance to influence in AI system evaluation, promoting calibrated trust and preventing automation complacency. Understanding these aspects is critical for selecting the most suitable interaction protocols for specific work settings.

Keywords

Appropriate Reliance, Artificial Intelligence, Decision Support Systems, Human-AI Interaction, Calibrated Trust

Acknowledgments

C. Fregosi and F. Cabitza acknowledge funding support provided by the Italian project PRIN PNRR 2022 InXAID - Interaction with eXplainable Artificial Intelligence in (medical) Decision making. CUP: H53D23008090001 funded by the European Union - Next Generation EU.
C. Natali gratefully acknowledges the PhD grant awarded by the Fondazione Fratelli Confalonieri, which has been instrumental in facilitating her research pursuits.

A. Online Resources

The framework is available at https://mudilab.github.io/dss-quality-assessment/ (last access date: 11.11.2024).

Italian Workshop on Artificial Intelligence for Human Machine Interaction (AIxHMI 2024), November 26, 2024, Bolzano, Italy
*Corresponding author.
†These authors contributed equally.
c.fregosi@campus.unimib.it (C. Fregosi); andrea.campagner@unimib.it (A. Campagner); chiara.natali@unimib.it (C. Natali); federico.cabitza@unimib.it (F. Cabitza)
0009-0004-7626-8131 (C. Fregosi); 0000-0002-0027-5157 (A. Campagner); 0000-0002-5171-5239 (C. Natali); 0000-0002-4065-3415 (F. Cabitza)