Per Aspera ad Astra, or Flourishing via Friction: Stimulating Cognitive Activation by Design through Frictional Decision Support Systems

Chiara Natali

University of Milano-Bicocca, viale Sarca 336, Milan, 20126, Italy

Abstract

My research proposal explores the *frictional design* approach to stimulate cognitive engagement in Decision Support Systems, especially so for those that adopt eXplainable AI techniques (XAI). The aim of these *programmed inefficiencies* is the mitigation of human overreliance, while promoting thoughtfulness and cognitive enhancement, and ultimately improving the effectiveness of decision-making. Specifically, I will explore how frictional design principles can be applied to the development of AI decision support systems in order to create interfaces that encourage users to engage with the system actively and thoughtfully, rather than passively accepting its recommendations, in the form of *cautious*, *comparative*, *judicial* or *adjunct* support. By increasing transparency and promoting cognitive engagement, frictional design can help ensure that XAI decision support systems remain valuable tools for decision-makers, rather than becoming a crutch or a source of manipulation.

Keywords

Decision Support System (DSS), eXplainable AI (XAI), Cognitive friction, Human-AI Interaction

1. Introduction

In these frantic months of 2023, we are witnessing the proliferation of generative and conversational systems that have the capability to produce complex content and engage in human-like interactions, fostering the illusion of high competence, reliability, and trustworthiness. Although these systems may not possess sentience or awareness, they are becoming increasingly persuasive, particularly in sensitive domains. The optimization of these systems aims to achieve higher agreement with humans, reduce aversion, and foster trust by employing increasingly effective persuasion strategies studied and classified by psychological and design sciences. As a result, decision makers may be more inclined to rely on machine advice than ever before, prompted also by the increasing intuitiveness and user-friendliness of AI system interfaces that minimize cognitive friction [1], following the UX design maxim "Don't Make Me Think!" that was popularized by a book of the same title [2]. While these principles have largely remained relevant, it is evident that some suggestions inadvertently led to designs that elicited immediate and thoughtless reactions from users, as in the case of dark patterns [3] in design. Drawing on the insights of Frischmann and Selinger [4], according to whom "Tolerating some congestion,

CHItaly 2023: Crossing HCI and AI, Doctoral Consortium, 20–22 September 2023, Turin, Italy

chiara.natali@unimib.it (C. Natali)

0000-0002-5171-5239 (C. Natali)

CEUR Workshop Proceedings (CEUR-WS.org)

some friction, some inefficiency, even some transaction costs may be necessary to sustain an underdetermined environment conducive to human flourishing", my research as a first-year PhD student under the supervision of Prof. Federico Cabitza has focused on the role of AI Decision Support Systems (DSS) in the short-term decision-making process within the medical domain. Specifically, I investigate the phenomena of automation bias and deskilling. My objective is to explore whether the cognitive-forcing functions [5] embedded in 'programmed inefficiencies' [6] or 'frictional protocols', that make the interpretation of the AI output less immediate and promote the cognitive activation of users, can yield benefits and mitigate potential negative impacts in terms of decision accuracy, usability and skill retention. Recognizing the significant risks of overreliance and deskilling associated with the development of DSS, it is essential to acknowledge that AI interventions encompass various design choices that become evident in the presentation of the system's output: it is the case of choosing between clear-cut categories, probabilities, prioritized lists of alternatives, similar cases, or explanations, just to name a few options. Additionally, lesser-explored design decisions include the order and level of automation (e.g., AI as a first-opinion or second-opinion giver [7], AI as an autonomous agent or case-mining tool [8]), optimization criteria (e.g., accuracy or utility), and adaptation to the target population (e.g., novices or experts). These configurations collectively form a protocol of human-AI interaction, which defines the "process schema that stipulates the use of AI tools by competent practitioners to perform a certain task or job" [7]. Each specific protocol possesses a distinct potential for influence and user reliance, leading to varying levels of dominance (i.e. "the influence that an AI system can exert on human judgment and decisions" [9]) after proper evaluation. Within the scope of my investigation, I propose and aim to evaluate empirically four specific protocols aimed at safeguarding human agency: cautious protocols, comparative protocols, judicial protocols, and adjunct protocols.

2. The protocols

Cautious protocol. This protocol involves presenting a set of potential answers (possibly also with varying individual confidence scores), accompanied by a predetermined probability level that the proposed set encompasses the correct answer, such as in conformal prediction [10]. Consequentially, the system may offer the option to abstain from providing an answer [11]. In cautious protocols, the limitations of the AI system and of its predictions are transparently communicated to the user, adopting a similar approach to 'seamful design' [12], where knowledge gaps, uncertainties, and discrepancies between the training data and the application domain are deliberately revealed as to empower human agency [13].

Comparative protocol. In this protocol, the AI system provides users with the most similar cases associated with either a ground truth or all available classes in relation to the case being considered. By leveraging this approach, the system facilitates analogical reasoning, resembling a transactive memory that taps into a repository of past cases and their corresponding correct decisions [8]. An illustrative example of a comparative protocol is the implementation of *pro-hoc* explanations, which replace the AI's decision support rather than being provided alongside it.

Judicial protocol. In a judicial protocol, the AI system provides arguments and explanations that support multiple, conflicting decisions or interpretations. It is the example of perorative

explanations generated by opposing conversational agents [14] or even the incorporation of two or more antagonistic machine learning models, belonging to different families, trained on distinct representations, ground truths, and parametrizations [15]. Similarly, the introduction of "conflicting rules/knowledge" has been explored as a debiasing technique against *Overconfidence* and *Underconfidence* [16]. While this approach does not entirely eliminate the risk of manipulation by the more persuasive agent, it places the human decision maker at the forefront. It fosters an appreciation for maintaining an impartial, focused, and responsible stance, while unveiling the persuasive tactics employed by the AI systems.

Adjunct protocol. First introduced in [17], the adjunct protocol differs from the previous three protocols as it pertains not to the presentation of 'desirably inefficient' output modalities but instead encompasses the design of the decision-making process itself through specialized Human-AI Collaboration Protocols [7]. The typical configuration is that of AI as a second-opinion giver, as exemplified in the Human-first HAI-CP discussed in [7]. Other process-based cognitive forcing functions [5], such as checklists, diagnostic time-outs, withholding the AI suggestion or incorporating longer waiting times (previous studies have indicated that slower algorithms enhance user accuracy [18]) also contribute to this protocol. During the waiting period for the AI's suggestion, the user may formulate their own hypothesis regarding the correct solution and subsequently evaluate the AI's explanation to assess its alignment with their own conception, possibly alleviating anchoring bias [19] and confirmation bias [16].

3. Research questions and methodology

The research questions to be addressed in my doctoral thesis are the following:

- **RQ1.** Are frictional design principles conducive to more effective, or at least equally effective, decision-making compared to traditional protocols (cf. *non-inferiority* trials)?
- **RQ2a.** Do frictional DSS mitigate the risk of automation bias and dominance (as operationalized in [9])?
- **RQ2b.** Do frictional DSS mitigate the risk of deskilling as well as upskilling inhibition?
- **RQ3.** What is the usability of frictional design patterns and does it change according to the user expertise and task complexity?

To address these research questions, a series of user experiments involving medical professionals were conducted [20, 9, 8, 6, 7] and more are in the works to employ and assess related Human-AI Collaboration Protocols (HAI-CP). For the first and second research questions, I will consider the concepts of technology benefit and reliance patterns presented and operationalized in [9] for the assessment of the impact of DSS on decision-making in terms of decision effectiveness and cognitive biases, using the rates of decision change following the exposure to AI advice as a proxy for the (positive or negative) influence of the DSS. The most ambitious and complex research question focuses on the consequences and dynamics of long-term use of frictional and non-fritcional HAI-CP in terms of the deskilling or lack of acquisition of new competences, to be addressed by consulting the relevant literature as it develops and designing feasible empirical investigations. Finally, I will investigate the fourth research question by referring to the state-of-the-art in usability questionnaires and possibly conducing qualitative interviews following the grounded theory approach [**?**].

4. Preliminary results and future work

The laboratory I am affiliated to (MUDI Lab, University of Milano-Bicocca) conducted an empirical user study in the field of radiology to investigate the use of pro-hoc explanations within a human-first, comparative protocol. The study employed elements of an adjunct protocol, utilizing a human-first interaction approach, and a comparative protocol via pro-hoc explanations, as the AI intervention involved presenting the three most similar cases (to the case at hand) retrieved from the available dataset. The study's findings are currently being prepared for submission to a medical informatics journal. Although the AI intervention led to a non-significant improvement in accuracy (approximately 2%), physicians perceived the AI support as significantly useful, especially so those with less clinical experience.

Building upon these preliminary results, future studies will mainly focus on evaluating the effects of cautious and judicial protocols. A collaborative experimentation with the University of Pavia is currently underway to investigate the application of the judicial protocol in the radiological domain.

Additionally, a conceptual emphasis will be placed on developing a taxonomy of frictional design patterns. The objective is to create a repository of reusable design patterns that incorporate frictional elements in several human-AI interaction protocols across diverse contexts.

Finally, I will devote time to the study of ethical considerations connected to frictional support, developing impact assessment checklists and work on governance issues: my aim will be to develop guidelines and recommendations to ensure the responsible and ethical implementation of programmed inefficiencies in various contexts.

5. Conclusion

Instead of attributing the phenomenon of human over-reliance on AI systems to cognitive biases and human limitations, my research proposes a paradigm shift by focusing on how technology is designed and deployed, with designers and programmers responsible for the promotion of user agency, responsibility and skill. While recognizing that increased friction may negatively affect how users perceive system usability, I criticize the prioritization of efficiency and comfort over the efficacy and integrity of our knowledge work, putting at the forefront the promotion of responsible and thoughtful decision-making as well as the mitigation of risks of over-reliance, technology dominance and deskilling. This approach aligns with the principles of *slow design* [21], which does not advocate for slowness in itself in a technophobic fashion, but rather aims to facilitate users in engaging in the right actions at the appropriate time and pace. By embracing frictional design of DSS, we can encourage users to better understand and reflect on their actions, fostering more intentional and meaningful interactions with technology.

References

[1] A. Cooper, The inmates are running the asylum, Springer, 1999.
[2] S. Krug, Don't make me think!, Pearson, 2000.

[3] C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, A. L. Toombs, The dark (patterns) side of ux design, in: Proceedings of the 2018 CHI conference on human factors in computing systems, 2018, pp. 1–14.
[4] B. Frischmann, E. Selinger, Re-engineering humanity, Cambridge University Press, 2018.
[5] Z. Buçinca, M. B. Malaya, K. Z. Gajos, To trust or to think: cognitive forcing functions can reduce overreliance on ai in ai-assisted decision-making, Proceedings of the ACM on Human-Computer Interaction 5 (2021) 1–21.
[6] F. Cabitza, A. Campagner, D. Ciucci, A. Seveso, Programmed inefficiencies in dss-supported human decision making, in: Modeling Decisions for Artificial Intelligence: 16th International Conference, Milan, Italy, September 4–6, 2019, Springer, 2019, pp. 201–212.
[7] F. Cabitza, A. Campagner, et al., Rams, hounds and white boxes: Investigating human-ai collaboration protocols in medical diagnosis, Artificial Intelligence in Medicine (2023).
[8] F. Cabitza, A. Campagner, C. Simone, The need to move away from agential-ai, International Journal of Human-Computer Studies 155 (2021) 102696.
[9] F. Cabitza, A. Campagner, R. Angius, C. Natali, C. Reverberi, Ai shall have no dominion: on how to measure technology dominance in ai-supported human decision-making, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 2023.
[10] G. Shafer, V. Vovk, A tutorial on conformal prediction., Journal of Machine Learning Research 9 (2008).
[11] A. Campagner, F. Cabitza, D. Ciucci, Three-way decision for handling uncertainty in machine learning: A narrative review, in: IJCRS 2020, Havana, Cuba, June 29–July 3, 2020, Proceedings, Springer, 2020, pp. 137–152.
[12] M. Chalmers, I. MacColl, M. Bell, Seamful design: Showing the seams in wearable computing, in: 2003 IEE Eurowearable, IET, 2003, pp. 11–16.
[13] S. Inman, D. Ribes, Beautiful seams: Strategic revelations and concealments, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019.
[14] T. Miller, Explainable ai is dead, long live explainable ai! hypothesis-driven decision support, arXiv preprint arXiv:2302.12389 (2023).
[15] M. Hildebrandt, Algorithmic regulation and the rule of law, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 376 (2018) 20170355.
[16] T. Kliegr, Š. Bahník, J. Fürnkranz, A review of possible effects of cognitive biases on interpretation of rule-based machine learning models, Artificial Intelligence 295 (2021).
[17] F. Cabitza, C. Natali, Open, multiple, adjunct. decision support at the time of relational ai, in: HHAI2022: Augmenting Human Intellect, IOS Press, 2022, pp. 243–245.
[18] J. S. Park, R. Barber, A. Kirlik, K. Karahalios, A slow algorithm improves users' assessments of the algorithm's accuracy, Proceedings of the ACM on Human-Computer Interaction 3 (2019) 1–15.
[19] C. Rastogi, et al., Deciding fast and slow: The role of cognitive biases in ai-assisted decision-making, Proceedings of the ACM on Human-Computer Interaction 6 (2022) 1–22.
[20] F. Cabitza, A. Campagner, L. Famiglini, et al., Color shadows (part i): Exploratory usability evaluation of activation maps in radiological machine learning, in: Machine Learning and Knowledge Extraction, Springer, 2022, pp. 31–50.
[21] B. Grosse-Hering, et al., Slow design for meaningful interactions, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2013, pp. 3431–3440.