Natali, Chiara (2024): Frictional AI. Designing Desirable Inefficiencies in Decision Support Systems for Knowledge Work. In: Proceedings of the 22nd European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centered Computing on the Design of Cooperation Technologies - Doctoral Colloquium Contributions, Reports of the European Society for Socially Embedded Technologies (ISSN 2510-2591), DOI: 10.48340/ecscw2024_dc07

Frictional AI. Designing Desirable Inefficiencies in Decision Support Systems for Knowledge Work

Chiara Natali
University of Milano Bicocca, Italy
chiara.natali@unimib.it

Abstract. My research involves the conceptualization of 'Frictional AI' as a novel approach for enhancing cognitive engagement in Decision Support Systems (DSS) through intentional design inefficiencies. Through empirical studies and theoretical analysis, I explore the balance between human intuition and automation-induced enhancements to decision-making in medical diagnostics. With the introduction and assessment of four distinct frictional protocols (cautious, comparative, judicial, and adjunct), this design framework prioritizes the efficacy and integrity of human knowledge work, ensuring that professionals are engaged, critical thinkers, capable of counteracting automation bias and deskilling–even at a slight cost in efficiency and comfort.

1 Introduction

The CSCW community is wary of "underrating the skills and competencies that are required in even the most routine of tasks," (Hartswood et al., 2003) instead acknowledging the skills and expertise demanded by tasks often perceived as mundane. This ethos stems from an appreciation of the complexities inherent in collaborative efforts and the view of technology as a means to empower, not replace, human participation.

In this light, my doctoral research is driven by the goal of informing the debate around "what to automate and what to leave to human skill and ingenuity" (Hartswood et al., 2003) by proposing design strategies that promote AI-supported decision-making processes.

I propose a series of design solutions informed by the concept of Frictional AI: an AI design principle that leverages programmed inefficiencies (Cabitza et al., 2019) to stimulate human cognitive activation and mitigate overreliance on automated systems (Cabitza et al., 2024a).

The development of Decision Support Systems (DSS) brings to light the critical issues of dependency and the erosion of skills, highlighting the necessity of considering the multitude of design options available in AI deployments. These options manifest in how a system's outputs are displayed, ranging from oracular advice, clear-cut classifications, probabilities or ranked alternatives, among other possibilities. Equally important, yet less frequently addressed, are design choices regarding the sequence and level of automation (Parasuraman et al., 2000), the benchmarks for optimization (such as precision or practicality), and the customization for specific user groups (like beginners or specialists).

Together, these elements constitute a Human-AI Collaboration Protocol (HAI-CP), outlining a structured approach for the application of AI instruments by skilled professionals in executing particular tasks or roles, as described by Cabitza et al. (2023b). Each protocol offers a unique scope for affecting and shaping user reliance, thereby determining the extent of 'dominance' - the level of impact an AI entity may have on human decision-making (Cabitza et al., 2023a).

Although reducing friction is often sought to improve user experience and system usability, I question the wisdom of favoring ease and efficiency at the expense of the quality and integrity of knowledge work (Kelloway and Barling, 2000; Bossen and Pine, 2023). Echoing the thoughts of Frischmann and Selinger (2018), who argue that embracing a degree of inefficiency and friction might be vital for maintaining an environment that fosters human flourishing, my research aims to investigate the potential advantages of 'frictional protocols'.

These protocols are designed to make the interpretation of AI output less immediate, thereby stimulating user cognitive engagement, with the goal of enhancing decision-making accuracy, sense of agency and the preservation of professional skills. Such an approach is particularly pertinent for professionals,

like those in the medical field, who are involved in high-stakes processes, ensuring they remain actively involved, critically minded participants rather than mere operators. My work focuses on designing a conceptual framework for AI systems that accurately mirror the nuances of diagnostic tasks and are implemented in a manner that honors the irreplaceable contribution of human expertise.

My research encompasses the development and empirical assessment of four distinct protocols devised to protect human autonomy: cautious protocols, which encourage prudence; comparative protocols, which facilitate the evaluation of multiple outcomes; judicial protocols, which involve agonistic decision-making processes; and adjunct protocols, which move the AI support to an ancillary role to human judgment.

2 Frictional protocols

- **Cautious protocol.** This approach offers a range of possible answers, each with its own confidence level, and includes a probability that the correct answer is among them, similar to conformal prediction (Shafer and Vovk, 2008). It allows the system to choose not to give an answer if uncertainty is too high (Campagner et al., 2020), making users aware of the AI's limitations in a clear way, similar to the concept of 'seamful design' (Chalmers et al., 2003).
- **Comparative protocol.** Here, the AI system presents users with cases similar to the one being examined, along with their outcomes. This method is akin to a Transactive Memory System (Lewis and Herndon, 2011), utilizing a database of previous cases and their outcomes to aid decision-making.
- **Judicial protocol.** This protocol offers multiple, sometimes conflicting, viewpoints or decisions, supported by explanations. It can involve using AI to generate arguments for different sides (Miller, 2023), similar to having a debate between opposing conversational AI agents, or using antagonistic machine learning models optimized according to different parameters (Hildebrandt, 2018). Kliegr et al. (2021) has similarly proposed using "conflicting rules/knowledge" as a way to reduce biases like overconfidence or underconfidence
- **Adjunct protocol.** Differently from the other protocols, which focus on output presentation, this one focuses on the decision-making process itself through specialized HAI-CP. Introduced in previous work as a method where AI serves more as a consultant providing a second opinion (Cabitza and Natali, 2022; Cabitza et al., 2023b), it involves designing the decision-making process to encourage critical thinking. Cognitive forcing functions (Buçinca et al., 2021) such as checklists, delaying AI recommendations, or

longer wait times are used to counteract common biases like anchoring bias (Rastogi and et al., 2022) or confirmation bias (Kliegr et al., 2021).

3 Research questions

Main question. How can the principles of Frictional AI inform the design of AI systems in a way that aligns with CSCW tenets, fostering environments where professionals are engaged, critical thinkers and decision-makers?

Investigating this wide-reaching research question requires addressing four sub-questions:

- **Q1.** Are frictional design principles conducive to more effective, or at least equally effective, decision-making compared to traditional AI-supported protocols?
- **Q2a.** Can frictional DSS mitigate the risks of automation bias, as operationalized by Cabitza et al. (2023a)? Similarly, **Q2b.** Do frictional DSS mitigate the risk of deskilling as well as upskilling inhibition?
- **Q3.** How does the usability of frictional design patterns change according to user expertise and task complexity?
- **Q4.** What is the ethical and conceptual basis for the development and acceptance of Frictional AI systems in (work) practice?

4 Methodological approach

To address these research questions, I participated to a series of empirical user studies involving medical professionals (Natali et al., 2023; Cabitza et al., 2023a,b,d,c, 2024b), with additional studies planned to further explore and evaluate other frictional HAI-CP.

For the first and second research questions, which tackle the effectiveness of frictional AI and its potential to reduce automation bias, I will leverage the strategies for evaluating technology benefit and reliance patterns developed by Cabitza et al. (2023a), specifically aimed at examining the influence of DSS on decision-making effectiveness and susceptibility to cognitive biases.

The most challenging investigation delves into the possible deskilling effects and upskilling prevention stemming from prolonged engagement with non-frictional HAI-CPs. Given the time-intensive nature of such an experiment, as well as its ethical implications, I will address this research question by consulting the relevant literature as it develops and designing feasible qualitative investigations.

As for the third research question, on the usability of Frictional AI according to user experience and task complexity, I will tackle it via user studies as well state-of-the-art usability questionnaires. The findings could be strengthened via

ethnomethodological observations and qualitative interviews analyzed following the grounded theory approach (Muller and Kogan, 2010).

I will tackle the fourth research question by employing a robust theoretical examination to situate the concept of Frictional AI within CSCW literature, inspired by Bossen and Pine's (2023) observations on factors for successful human-AI collaboration, such as AI supporting "what motivates many knowledge workers: reflecting and learning through experimentation". This will be followed by principle elicitation initiatives through workshops and conference special tracks (see Section *Future Work*).

5 Preliminary findings

As a first step, I introduced the concept of Frictional AI within the CSCW context through a study on AI system perceptions by oncological radiologists in two French hospitals, detailed in "Invisible to machines: Designing AI that supports vision work in radiology" (Anichini et al., 2024).

As for user studies, I explored the impact of pro-hoc explanations in AI systems as an example of comparative, human-first protocols in fracture detection tasks (Cabitza et al., 2024b). This approach involved comparing similar past cases to assist in diagnosis, akin to a Transactive Memory System (Lewis and Herndon, 2011) leading to a decrease in diagnostic errors by 10%. The study highlighted differences in how specialists and residents adjusted their diagnoses when using the system, particularly in complex cases. Notably, decision changes were generally beneficial, though some instances of automation bias were observed, more so among residents than specialists. My contribution included a section exploring in depth the topic of Frictional AI.

My latest research, currently in peer review, examines the use of AI in group decision-making, finding that groups show less reliance on AI compared to users in individual Human-AI decision-making settings. The study suggests that group discussions before AI consultation can lead to less dependence on the AI's advice, although potentially leading to the phenomenon of groupthink. However, when AI advice is considered first, after a first private decision (Bahrami et al., 2012), it appears to foster more critical discussion and analysis initiated by reflections on AI's contributions (Chiang et al., 2023), reducing groupthink and automation bias.

This work proposes a model where AI serves as a catalyst for group deliberation, playing an ancillary or adjunct role, ensuring collective decisions are driven by human intelligence.

6 Future work

Building upon these preliminary results, future studies will focus on cautious and judicial protocols. Additionally, an ethnomethodological investigation would shed light on the receptiveness of medical professionals towards the principles of Frictional AI, providing valuable context and understanding of its practical implications in real-world settings.

Empirical evidence and literature analysis will provide the basis of a comprehensive framework dedicated frictional design principles for AI systems.

The goal is to curate a collection of frictional design patterns for the development of HAI-CP suitable for a broad spectrum of scenarios. This would include the development of guidelines and recommendations to ensure responsible deployment, which will be conducted via principle elicitation activities.

A first example is my organization of the workshop “Stimulating Cognitive Engagement in Hybrid Decision-Making: Friction, Reliance and Biases” to be held at HHAI24 on June 11th 2024, intended to be the first of its kind in its discussion of Frictional AI.

As for further investigation of strategies to decrease overreliance, I will be chair of the Special Track “Calibrating Trust in XAI” at the 2nd World Conference on eXplainable AI in July 2024 (date TBA).

7 Expected contributions

The expected contributions of my research to the CSCW research community includes introducing and elaborating on the concept of Frictional AI as a novel approach to designing AI systems that enhance, rather than replace, human decision-making capabilities.

I will explore both the potential benefits and the inherent challenges of incorporating frictional design into AI systems. This exploration will involve both empirical studies and theoretical analysis to understand whether, and how, frictional design can foster greater cognitive engagement and effectively reduce overreliance on automation, thus guiding the creation of AI systems that prioritize human-centered design principles.

8 Conclusion

My doctoral research introduces design principles rooted in the concept of Frictional AI, presenting a paradigm that may seem to deviate from the dominant market trajectory focused on enhancing specificity and certainty in AI technologies.

The industry's current trajectory gravitates towards AI solutions that are marketed on the promise of increased efficiency, unambiguous outcomes, and, purportedly, superior returns on investment. Yet, it is within this juxtaposition that Frictional AI finds its most significant justification and potential for impact.

This strategy advocates for a more sustainable model of technology adoption, where the value of AI is measured not just in terms of efficiency gains but also in its ability to maintain and enhance the professional skills of its users.

By steering AI development to better reflect the uncertain, intuitive, and cooperative aspects of professional work environments, this research aims to pave the way for a more nuanced and impactful integration of technology in the realm of diagnostic work.

Acknowledgments

The author gratefully acknowledges the PhD grant awarded by the Fondazione Fratelli Confalonieri, which has been instrumental in facilitating her research pursuits.

References

Anichini, G., C. Natali, and F. Cabitza (2024): 'Invisible to machines: Designing AI that supports vision work in radiology'. Computer supported cooperative work (CSCW), vol. In press, no. tbd, pp. tbd.

Bahrami, B., K. Olsen, D. Bang, A. Roepstorff, G. Rees, and C. Frith (2012): 'What failure in collective decision-making tells us about metacognition'. Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 367, no. 1594, pp. 1350–1365.

Bossen, C. and K. H. Pine (2023): 'Batman and Robin in Healthcare Knowledge Work: Human-AI Collaboration by Clinical Documentation Integrity Specialists'. ACM Transactions onComputer-Human Interaction, vol. 30, no. 2, pp. 1–29.

Buçinca, Z., M. B. Malaya, and K. Z. Gajos (2021): '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, vol. 5, no. CSCW1, pp. 1–21.

Cabitza, F., A. Campagner, R. Angius, C. Natali, and C. Reverberi (2023a): 'AI Shall Have NoDominion: 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.

Cabitza, F., A. Campagner, D. Ciucci, and A. Seveso (2019): 'Programmed inefficiencies in DSS-supported human decision making'. In: Modeling Decisions for Artificial Intelligence: 16th International Conference, Milan, Italy, September 4–6, 2019. pp. 201–212.

Cabitza, F., A. Campagner, and et al. (2023b): 'Rams, hounds and white boxes: Investigating human-AI collaboration protocols in medical diagnosis'. Artificial Intelligence in Medicine.

Cabitza, F., A. Campagner, L. Famiglini, C. Natali, V. Caccavella, and E. Gallazzi (2023c): 'Let Me Think! Investigating the Effect of Explanations Feeding Doubts About the AI Advice'. In: International Cross-Domain Conference for Machine Learning and Knowledge xtraction. pp. 155–169.

Cabitza, F., A. Campagner, C. Natali, E. Parimbelli, L. Ronzio, and M. Cameli (2023d): ‘Painting the black box white: experimental findings from applying XAI to an ECG reading setting’. Machine Learning and Knowledge Extraction, vol. 5, no. 1, pp. 269–286.

Cabitza, F. and C. Natali (2022): ‘Open, multiple, adjunct. Decision support at the time of relational AI’. In: HHAI2022: Augmenting Human Intellect. IOS Press, pp. 243–245.

Cabitza, F., C. Natali, L. Famiglini, A. Campagner, V. Caccavella, and E. Gallazzi (2024a): ‘Never tell me the odds. Investigating pro-hoc explanations, as instances of frictional AI, in medical decision making’. Artificial Intelligence in Medicine.

Cabitza, F., C. Natali, L. Famiglini, A. Campagner, V. Caccavella, and E. Gallazzi (2024b): ‘Never tell me the odds: Investigating pro-hoc explanations in medical decision making’. Artificial Intelligence in Medicine, pp. 102819.

Campagner, A., F. Cabitza, and D. Ciucci (2020): ‘Three-way decision for handling uncertainty in machine learning: A narrative review’. In: IJCRS 2020, Havana, Cuba, June 29–July 3, 2020, Proceedings. pp. 137–152.ù

Chalmers, M., I. MacColl, and M. Bell (2003): ‘Seamful design: Showing the seams in wearable computing’. In: 2003 IEE Eurowearable. pp. 11–16.

Chiang, C.-W., Z. Lu, Z. Li, and M. Yin (2023): ‘Are Two Heads Better Than One in AI-Assisted Decision Making? Comparing the Behavior and Performance of Groups and Individuals in Human-AI Collaborative Recidivism Risk Assessment’. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–18.

Frischmann, B. and E. Selinger (2018): Re-engineering humanity. Cambridge University Press.

Hartswood, M., R. Procter, M. Rouncefield, and R. Slack (2003): ‘Making a case in medical work: implications for the electronic medical record’. Computer Supported Cooperative Work (CSCW), vol. 12, pp. 241–266