Learning how to use a Decision Support System: a simulation of Human-AI interaction frames

Corentin Boidot & Antoine Mourato & Léo Facca — École Navale / IRENav

The integration of AI in Decision Support System faces a hard limit: automation bias (Cummings, 2004). While AI may outperforms human operators in certain tasks, the added value of keeping humans-in-the-loop becomes questionable if they merely validate AI decisions. However, the ideal tool should enable complementary team performance (CTP). CTP describes a situation where the human-AI (HAI) team overcomes both AI capacity and the operator alone in terms of performance, which is incompatible with automation bias. CTP describes a one-time performance, and is not easy to achieve: it means operators stay on their mind mostly when they are correct, and change their mind mostly when the AI is correct, which is optimistic, as shown in a recent meta-analysis (Vaccaro et al., 2024).

However, one observation of CTP doesn't preclude automation bias in the long-run. Beyond a static conception of performance, the dynamics of these interactions have been studied through several theoretical perspectives (Kumar et al., 2021; Williams et al., 2023; Gonzalez et al., 2026). The stress is often put on the evolution of trust in AI through time (Chong et al., 2022), but the issue of learning effects in the task, independent of trust, is poorly understood. What is the impact of AI assistance on human skill acquisition? This question has been mostly answered in educational contexts (Fan et al., 2025; Kosmyna et al., 2025), where the delegation of cognitive tasks to the AI would hinder the learning process of the children. In operational contexts, the fear of deskilling is often expressed by experts like physicists (Natali et al., 2025), but empirical evidence is missing, to our knowledge. This gap in the literature is understandable as the inquiry is challenging: in any domain, future experts are few, their tasks to learn are numerous, and the deployment of measurement settings on long duration might be difficult.

To explore this question without conducting such longitudinal studies, we propose to build a model of human learning of a technical task with an advisor. Our work develop a simulation setting on a classification task to depict possible quantitative effects of advice-taking on performance evolution through time. This modeling scheme allowed us to investigate the impact of the advisor on skill development. While most people expect that a more accurate advisor would lead the learner to higher performances, Vaccaro et al. (2024) suggest that advisors with lesser performance would be more prone to inducing CTP. This claim follows from static experimental contexts: we hereby test it in a dynamical setting. We expect that in the long-run, deskilling could happen, and would affect not-only degraded-mode performance, but also maximal running performance. Is the performance of the advisor the main factor determining performance evolution, or could the assistants have an intrinsic complementarity with the learner, depending on their classification patterns? The type of AI could lead to different results, in terms of performance, reliance, and of dependence.

Identification of Use Cases

The development of AI on high-stake classification tasks is driving towards field applications. Especially in defense, many target recognition tasks —Unmanned Aerial Vehicles, field targets,

mines in seafloor images, targets in acoustic signals— receive significant attention from data scientists, with a surge of automation. In such domain, the achievement of CTP, and its maintaining over the long term can be seen as a sign of effective HAI interactions. In practice, when an AI is deemed imperfect by domain experts, and therefore not prone to CTP in operations, it can still become a tool to help young operators in their training. Design questions arise, as how to promote a reasonable use of this AI in their decision-process: when used naïvely as an advisor, would an AI hinder the development of the future expert's own resources (reliability, autonomy)? In the study of HAI interactions, we conduct a prospective evaluation of HAI capabilities using computational models (Chen et al., 2022; Guo et al., 2024).

Our approach can be applied to any context modeled as classification or regression tasks, where human experience significantly influences performance. It helps identify suitable AI characteristics to achieve complementary in a training context, guiding the selection of variables to prioritize in longitudinal studies. Additionally, by clarifying the dynamic effects of AI assistance over time, it provides insights to mitigate deskilling risks and support controlled skill progression in trainees.

Proposed Technical Approach

We proof-tested our simulation method on a low-dimensional data classification, with class unbalance, as it is a frequent difficulty of applied ML. We use artificial neural networks (NN) as proxies for human learning: they are flexible enough in terms of complexity, and their learning curve can be easily observed. Psychologically-grounded models of learning often fall within the scope of Reinforcement Learning (Gonzalez & Dutt, 2011) : here, NNs are a simple attempt to picture learning in a supervised setting. We aim to model goal-oriented behaviors, where humans adapt their actions in an interaction frame—reliance is seen as a result of this task-driven dynamic. We consider two baseline interaction frames: solo decision-making (S) and assisted decision-making (A). In addition to these baselines, we implemented the ideas of trust quotas (Q) and AI mentalization (M) — see (Boidot et al., 2025). In conjunction to these independent variables, several assistants have been tested : a logistic regression, two Support Vector Machine and two Decision Trees, with varied degrees of performance (Lin, SVM+, SVM-, DT+, DT-). Their performance levels are on par with those of our neural learner, an MLP with two hidden layers of 4 neurons. These AI are simple, yet realistic compared to the Wizard-of-Oz AIs used in most human experiments. This assistant prediction score is presented as a supplementary variable in input of the NN.

The simulation results consists in two learning curves: one for performance (measured by the f1-score), one for reliance (learner-assistant agreement), over 30 training epochs. The task is a new synthetic task with three variables, as a traditional dataset, like moon, has yielded near-perfect performance with most algorithms. Our simulation also allows us to derive “degraded mode” performances, ie. performances of an operator trained to work with an assistant, in a situation where it is unavailable (emulated by a constant 50% signal).

Our results illustrate that transitory CTP can be observed on learners performance, using less accurate AIs, while on the long term CTP is not achieve — ie. the operator learning without an assistant reaches better performances.

Bibliography

Boidot, C., Mourato, A., Facca, L., & Lemaître, D. (2025). Simuler les interactions humain-IA : Plans de recherche autour de l'émergence d'une décision hybride. *Conférence Nationale en Intelligence Artificielle*. https://hal.science/hal-05113982/

Chen, V., Johnson, N., Topin, N., Plumb, G., & Talwalkar, A. (2022, août 20). *Use-Case-Grounded Simulations for Explanation Evaluation*. NeurIPS-2022. https://doi.org/10.48550/arXiv.2206.02256

Chong, L., Zhang, G., Goucher-Lambert, K., Kotovsky, K., & Cagan, J. (2022). Human confidence in artificial intelligence and in themselves : The evolution and impact of confidence on adoption of AI advice. *Computers in Human Behavior*, *127*, 107018.

Cummings, M. (2004, septembre 20). Automation Bias in Intelligent Time Critical Decision Support Systems. *AIAA 1st Intelligent Systems Technical Conference*. AIAA 1st Intelligent Systems Technical Conference. https://doi.org/10.2514/6.2004-6313

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness : Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*(2), 489-530. https://doi.org/10.1111/bjet.13544

Gonzalez, C., Donahue, K., Goldstein, D. G., Heidari, H., Jalali, M. S., Schelble, B., Singh, A., & Woolley, A. W. (2026). *Toward a science of human–AI teaming for decision making : A complementarity framework*.

Gonzalez, C., & Dutt, V. (2011). Instance-based learning : Integrating sampling and repeated decisions from experience. *Psychological review*, *118*(4), 523.

Guo, Z., Wu, Y., Hartline, J. D., & Hullman, J. (2024). A Statistical Framework for Measuring AI Reliance. *CoRR*. https://openreview.net/forum?id=vLi9WdImDy

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your brain on ChatGPT : Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv preprint arXiv:2506.08872*, *4*. https://collimateur.uqam.ca/wp-content/uploads/sites/11/2025/12/2506.08872v1_comp.pdf

Kumar, A., Patel, T., Benjamin, A. S., & Steyvers, M. (2021). Explaining algorithm aversion with metacognitive bandits. *Proceedings of the annual meeting of the cognitive science society*, *43*(43). https://escholarship.org/uc/item/7xc470dt

Natali, C., Marconi, L., Dias Duran, L. D., & Cabitza, F. (2025). AI-induced Deskilling in Medicine : A Mixed-Method Review and Research Agenda for Healthcare and Beyond. *Artificial Intelligence Review*, *58*(11), 356. https://doi.org/10.1007/s10462-025-11352-1

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful : A systematic review and meta-analysis. *Nature Human Behaviour*, 1-11.

Williams, K. J., Yuh, M. S., & Jain, N. (2023). A Computational Model of Coupled Human Trust and Self-confidence Dynamics. *ACM Transactions on Human-Robot Interaction*, *12*(3), 1-29. https://doi.org/10.1145/3594715