The Role of User Control in Enhancing Human-AI Collaboration Effectiveness: Insights from a Pilot Study

Burak Oz, Alexander Karran, Jared Boasen, Constantinos Coursaris, and Pierre-Majorique Léger

Abstract In this research program proposal, we aim to investigate why experts override AI suggestions and identify design principles for more effective human-AI teams. Specifically, we propose testing whether increasing the perceived locus of control of human decision-makers over AI functions will lead to fewer overrides and improved performance. We present a mixed-factorial, multi-trial experimental design in which participants receive AI recommendations regarding demand forecasting decisions in a business simulation. Prior to each trial, one group specifies how they want the AI to function (experimental), and the other group does not (control). We use electroencephalography and oculometry to capture attention to recommendations and user interface elements. Behavioral data from a preliminary pilot study with four participants align with our hypotheses. We observed that participants in the experimental condition applied smaller adjustments to AI suggestions and had higher decision performance than the control group. The experiment's results will contribute to our understanding of AI aversion and inform the design of human-AI interactions to improve performance.

B. Oz (✉) · A. Karran · J. Boasen · C. Coursaris · P.-M. Léger
Tech3Lab, HEC Montréal, Montréal, Québec, Canada
e-mail: burak.oz@hec.ca

A. Karran
e-mail: alexander.karran@hec.ca

J. Boasen
e-mail: jared.boasen@hec.ca

C. Coursaris
e-mail: constantinos.coursaris@hec.ca

P.-M. Léger
e-mail: pml@hec.ca

J. Boasen
Faculty of Health Sciences, Hokkaido University, Sapporo, Japan

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_15

Keywords AI aversion · Artificial intelligence · Decision-making · Attention · Perceived locus of control · Human-AI interaction · NeuroIS · EEG · ERP · P300 · Eye-tracking

1 Introduction

Data-driven intelligent systems empowered by artificial intelligence (AI)[1] methods are increasingly being implemented in various fields [1–4]. However, the implementation potential of AI in complex decision-making tasks is far from full automation, primarily due to challenges in applying algorithmic solutions in critical contexts where there are concerns regarding ethical, social, or life-critical aspects of the final decision [5–9]. These concerns regarding full automation have given rise to semi-autonomous human-AI teams, where labor is divided between humans and AI. These systems involve collaboration between human experts and AI systems to perform tasks that require both human intuition and judgment abilities for dealing with unstructured problems and machine capabilities of big data analytics for dealing with structured problems [10, 11]. However, reports of AI implementation in knowledge work have shown that expert decision-makers often override AI suggestions and end up with worse decision performance [12–15].

Experts' reluctance to accept AI suggestions, even when AI performs well, can be referred to as AI aversion, akin to algorithm aversion [13]. AI aversion has been attributed to numerous factors such as lack of transparency, fear of bias, lack of accountability, preference for human judgment, and influence of personal experiences leading to different user expectations [16]. In essence, AI aversion occurs because of a lack of cognitive compatibility between human experts and intelligent systems, especially when these systems lack transparency and configurability [16–18].

Increasing compatibility between humans and AI systems can be possible through two primary methods: increasing human control over the system and improving system explanations [16, 19]. Research on AI explainability is vast and has produced valuable insights. On the other hand, although there is growing interest in studying user control, there is still a need for a better understanding of the impact of varying types and degrees of users' perceived locus of control on the effectiveness and acceptance of AI systems [19]. In this study, we define perceived locus of control as a person's perception of their ability to influence the human-AI semi-autonomous decision-making process [19, 20]. As part of a new research program investigating experts' AI aversion, we propose research investigating the impact of perceived locus of control on users' AI acceptance. More specifically, the proposed study aims to address the following research question:

[1] We employ the term "AI" to encompass a broad range of automation and autonomy capabilities, irrespective of the underlying technology, which could include learning agents like machine learning, deep learning, or artificial intelligence, as well as more static reasoning systems.

RQ1: To what extent does a user's perceived locus of control improve human-AI team performance?

While reducing AI aversion is critical, there is an equally important and relevant concept on the other end of the spectrum: algorithm overreliance [21]. Some experts may rely too heavily on algorithmic suggestions without considering contextual nuances and outliers, defying the purpose of implementing a human-AI team. Indeed, researchers express concerns that overreliance on AI can decrease system performance and error detection [22]. However, these are yet to be empirically tested:

RQ2: To what extent does overreliance on AI suggestions impact a user's ability to detect AI errors?

To address recent calls for research to leverage NeuroIS methods to develop more effective human-AI teams [22] and generate prescriptive IS design knowledge [23], we started a pilot study involving a mixed factorial, multi-trial experiment consisting of a demand forecasting task. We use electroencephalography (EEG) and oculometry to measure attention, cognitive processing, and gaze behaviors. We manipulate AI configurability across participants for the first research question and AI performance across trials for the second research question. We hypothesize that subjects permitted to specify AI function parameters will have a higher perceived locus of control, override the AI less often, exhibit stronger attentional processing and gaze behaviors, and perform the task better compared to the control group, even if the AI system provides the same suggestions regardless of user input. The findings of this study are expected to provide a better understanding of human-AI interactions that can be useful in informing developers on the design of effective and efficient human-AI teams.

In this submission, we report on our study design, briefly discuss preliminary results from a pilot study comprised of four participants and conclude with the expected contributions of the complete study currently underway.

2 Methods and Materials

To address the aims of this research plan, we developed an experimental task in which participants are asked to perform a series of demand forecasting tasks. This section explains the experimental task, followed by the experimental conditions, procedure, and data recording tools.

Participants

We will use convenience sampling to recruit 100 participants, primarily from our university's student participant panel. Participants will be screened based on their experience with SAP and the courses they have taken to ensure a basic knowledge of the experimental task and experience with stimuli. In addition to their task-related

knowledge, participants will be screened for having a normal vision and not being diagnosed with a neuropathological condition.

Material and Procedure

Building upon previous studies using the ERPsim business simulation to investigate human–computer interactions [24–28], we used this platform to generate a realistic demand dataset for the demand forecasting task. The experimental task uses the newsvendor problem, a well-established experimental paradigm in which participants play the role of a demand management agent [29, 30]. They decide how many units of a perishable product to order for the next period, knowing there is uncertain demand and different costs for having excess demand or losing sales.

Participants will be presented with a time series chart displaying a fictitious distribution company's weekly ice cream demand over the last 20 weeks. In each trial, they will evaluate the weekly demand data, receive a forecast estimation from the AI, and decide whether they want to edit the AI suggestion or continue directly. After they submit their decision, the simulation will advance by one week, and they will be presented with their performance in terms of lost and saved money. Following the performance report, participants continue doing this demand forecasting task in the same market, one virtual week per trial. With the addition of AI functionality to the newsvendor paradigm, we extend and demonstrate the use of this well-established decision-making task in an information systems study.

To ensure the ecological validity of the task, participants will use an industry-standard organizational resource planning software, SAP, to complete their experimental tasks. The SAP screen designed for this experiment shows pricing information, past demand data, and AI-supported demand forecasting functions. At the beginning of a trial, participants will not be provided with any AI suggestions or explanations, but the AI suggestion field will be visible without any number. Participants will be asked to scan and evaluate all demand information first, then fixate on the suggestion area while clicking the "Calculate AI Suggestions" button. One second after they click the button, a three or four-digit number will appear as an AI suggestion in its designated area, marking the onset of each trial's stimulus. Two seconds later, an explanation of the factors contributing to each suggestion will appear, together with buttons to accept or edit suggestions.

The designed SAP interface will be shown in full screen mode on a 22″ screen at 1920 × 1080 resolution at a 60 Hz refresh rate. Participants will mainly use a mouse to interact with the system and use the keyboard only when they adjust AI suggestions.

Experiment design. To test the impact of perceived locus of control on user behaviors in an AI-supported decision-making scenario, we designed a mixed factorial, multi-trial experiment with a between-subjects, two-level factor of AI configurability, and a within-subjects two-level factor of AI error.

Participants receive AI suggestions in all between-subjects conditions before completing their tasks. To investigate the impact of perceived user control on acceptance of AI suggestions and human-AI interactions, we aim to create an increased locus of control for participants in one of the between-subjects conditions. In the configurable AI (experimental) condition, the system asks for the user's configuration input at the beginning of each trial. In contrast, the unconfigurable AI (control) condition uses an AI system that does not ask for user input. Regardless of user inputs, both systems perform identically to isolate the psychological impact of the locus of control over performance. Participants in both groups can override AI suggestions.

The three AI parameters we ask participants to configure are smoothing, trend, and seasonality. These parameters are crucial in the Holt-Winters method, a popular time series forecasting model used in supply chain management [31]. We hypothesize that ability to provide input regarding relevant task parameters will increase their locus of control, which then will lead to a greater acceptance and utilization of AI suggestions in trials with high AI performance. Moreover, by making participants think about these parameters in each trial, we aim to increase their AI error detection abilities through increased task engagement and accumulated task knowledge.

We will conduct manipulation checks using previously validated self-report measurement item sets for perceived autonomy in human-AI interactions [19]. Moreover, to control the experiment duration across conditions, participants assigned to the control group will see a loading page for a duration that is based on our pretest observations.

In both between-subjects conditions, we follow an oddball paradigm design and randomly assign 15 of the 70 trials as oddballs in which the AI misbehaves, i.e., provides suggestions having more than three times the standard error of its average performance.

Experimental procedure. The experiment starts with a practice round in which participants are asked to complete five trials of demand forecasting tasks without AI support. After the practice block, there will be a training block with AI functionality, which is then followed by 70 trials of AI-supported demand forecasting tasks.

Data Recording and Analysis

Neurophysiological measures of attention to AI recommendations and the interface. EEG signals will be recorded from 32 scalp sites at a sampling rate of 1000 Hz. From this recorded data, event-related potentials (ERP) will be derived from the visual onset of the AI recommendation (the three to four-digit number at the bottom of Fig. 1). In this study, we will focus on the P300 ERP component, which reflects the cognitive processing of decision-relevant information and the allocation of attention to the stimulus [32].

The recorded EEG data will be filtered using 30 Hz low-pass and 1 Hz high-pass filters. The filtered data will then be divided into segments of 1 s per trial, starting 0.1 s

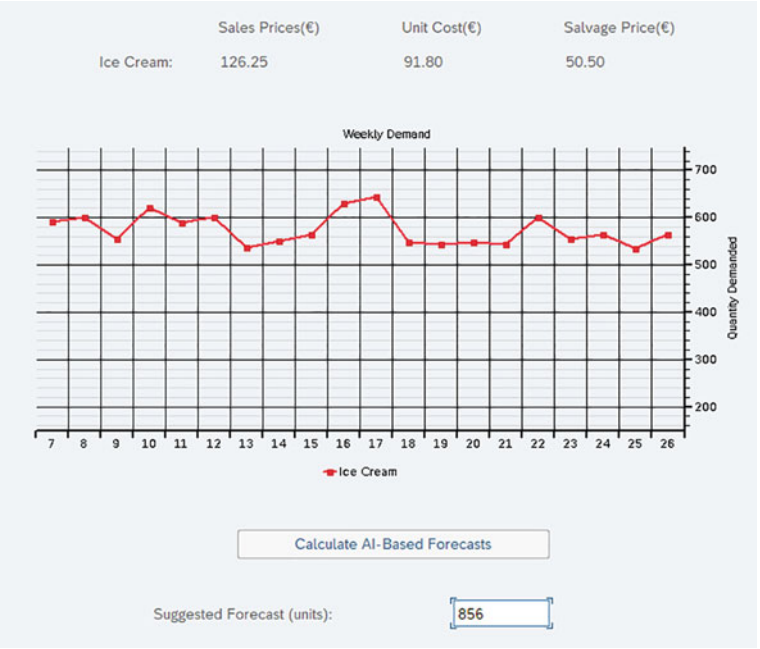

Fig. 1 Experimental stimuli. The relevant input information for the decision is on the top, whereas the AI-supported decision-making interface elements are on the bottom half of the screen. Participants are asked to analyze the data before clicking the "Calculate AI-Based Forecasts" button. They are also asked to focus on the suggestion field at the bottom. Once they receive suggestions, participants can accept or edit the suggested forecasts

before and 0.9 s after the stimulus onset of AI suggestion numbers. Following the epoching of the data, artifact rejection will be applied using automatic and manual procedures, assuming peak-to-peak amplitude ranges greater than 100 μV to be artifacts.

After filtering, epoching, and artifact rejection, the ERPs will be averaged across trials. These averaged ERPs will be used in statistical tests to determine whether our manipulation leads to a higher P300 peak amplitude, indicating higher attention to AI suggestions and improved ability to detect AI errors.

Eye-tracking measures. Gaze data will be recorded using a Tobii Pro Nano (Tobii Technology AB, Sweden) at a sampling rate of 60 Hz. Research has demonstrated that gaze transition entropy (GTE), a measure of the randomness of eye movements during visual processing, is related to attentional processing [33, 34]. Using GTE, we aim to gain insight into how individuals allocate attention when making forecasting decisions and detecting errors, and how this process is impacted by their perceived control over the AI system. Additionally, eye-tracking data will complement the ERP measurements by verifying participants' fixation on the stimulus at the onset.

Behavioral measures. We will also collect behavioral and self-report measures to gain a more comprehensive understanding of the phenomenon and triangulate our neurophysiological measurements. Behavioral measures include forecasting accuracy, response times, AI suggestion acceptance, and the magnitude of adjustments to AI suggestions. Self-report measures include the perceived level of control [19, 35, 36], perceived use and ease of use [37], trust in AI [38, 39], explanation satisfaction and understandability [39–41] and confirmation of expectations [40].

3 Preliminary Results and Future Outlook

We have conducted four pilot study sessions and analyzed behavioral and self-reported data. These results provide promising initial figures. First, our measurement of perceived level of control [19] supports the effectiveness of our manipulation. Second, our pilot study participants with a higher sense of control (i) made smaller adjustments to the AI suggestions, and (ii) obtained a higher forecasting performance. More specifically, in a time series with a seasonality ranging from 400 to 1,000 units, participants in the experimental group had an average adjustment of 128 units, whereas the control group participants had an AI adjustment average of 149 units across 70 trials. Also, participants in the experimental condition generated an average profit of 699 Euros (based on the game scenario), compared to 396 Euros by the control (140 observations per condition). Analysis of the ERP data from these pilot sessions is currently in progress.

This study is expected to provide empirical evidence for the relevance and importance of users' perceived sense of control over an AI in the success of human-AI collaborative system implementation in the context of decision-making. Moreover, the insights gained from the results of this study will have practical implications for the design of human-AI interactions. We believe that utilizing neurophysiological measures will enhance our understanding of the cognitive processes underlying human-AI interactions, leading to the design of more effective systems that optimize the contributions of both AI and human experts in their collaborative efforts.

References

1. Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization, 28*, 62–70. https://doi.org/10.1016/j.infoandorg.2018.02.005
2. Currie, G., Hawk, K. E., Rohren, E., Vial, A., & Klein, R. (2019). Machine learning and deep learning in medical imaging: Intelligent imaging. *Journal of Medical Imaging and Radiation Sciences, 50*, 477–487.
3. Fourcade, A., & Khonsari, R. (2019). Deep learning in medical image analysis: A third eye for doctors. *Journal of Stomatology, Oral and Maxillofacial Surgery, 120*, 279–288.

4. Giger, M. L. (2018). Machine learning in medical imaging. *Journal of the American College of Radiology, 15*, 512–520.
5. Anderson, P. L. (2019). *Damages caused by AI errors and omissions: Management complicity, malware, and misuse of data.* Anderson Economic Group.
6. Cheatham, B., Javanmardian, K., & Samandari, H. (2019). Confronting the risks of artificial intelligence. *McKinsey Quarterly, 2*, 38.
7. Akhtar, N., & Mian, A. (2018). Threat of adversarial attacks on deep learning in computer vision: A survey. *IEEE Access, 6*, 14410–14430.
8. Ferrer, X., van Nuenen, T., Such, J. M., Coté, M., & Criado, N. (2021). Bias and discrimination in AI: A cross-disciplinary perspective. *IEEE Technology and Society Magazine, 40*, 72–80.
9. Littman, M. L., Ajunwa, I., Berger, G., Boutilier, C., Currie, M., Doshi-Velez, F., Hadfield, G., Horowitz, M. C., Isbell, C., & Kitano, H. (2022). *Gathering strength, gathering storms: The one hundred year study on artificial intelligence (AI100) 2021 study panel report*. arXiv:2210.15767
10. Endsley, M. R. (2022). Supporting human-AI teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior,* 107574. https://doi.org/10.1016/j.chb.2022.107574
11. Rai, A., Constantinides, P., & Sarker, S. (2019). Next generation digital platforms: Toward human-AI hybrids. *MIS Quarterly, 43*, iii–ix.
12. Allen, R., & Choudhury, P. (Raj). (2022). Algorithm-augmented work and domain experience: The countervailing forces of ability and aversion. *Organization Science, 33*, 149–169. https://doi.org/10.1287/orsc.2021.1554
13. Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*, 114–126. https://doi.org/10.1037/xge0000033
14. Khosrowabadi, N., Hoberg, K., & Imdahl, C. (2022). Evaluating human behaviour in response to AI recommendations for judgemental forecasting. *European Journal of Operational Research, 303*, 1151–1167. https://doi.org/10.1016/j.ejor.2022.03.017
15. Kesavan, S., & Kushwaha, T. (2020). Field experiment on the profit implications of merchants' discretionary power to override data-driven decision-making tools. *Management Science, 66*, 5182–5190. https://doi.org/10.1287/mnsc.2020.3743
16. Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making, 33*, 220–239. https://doi.org/10.1002/bdm.2155
17. Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002
18. European Commission. (2021). *Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts.*
19. Westphal, M., Vössing, M., Satzger, G., Yom-Tov, G. B., & Rafaeli, A. (2023). Decision control and explanations in human-AI collaboration: Improving user perceptions and compliance. *Computers in Human Behavior,* 107714. https://doi.org/10.1016/j.chb.2023.107714
20. Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80*, 1.
21. Grissinger, M. (2019). Understanding human over-reliance on technology. *Pharmacy and Therapeutics, 44*, 320.
22. Rubin, D. L. (2019). Artificial intelligence in imaging: The radiologist's role. *Journal of the American College of Radiology, 16*, 1309–1317. https://doi.org/10.1016/j.jacr.2019.05.036
23. vom Brocke, J., Hevner, A., Léger, P. M., Walla, P., & Riedl, R. (2020). Advancing a NeuroIS research agenda with four areas of societal contributions. *European Journal of Information Systems, 29,* 9–24. https://doi.org/10.1080/0960085X.2019.1708218

24. Léger, P.-M. (2006). Using a simulation game approach to teach enterprise resource planning concepts. *Journal of Information Systems Education, 17*, 441–447.
25. Léger, P., Robert, J., Babin, G., Pellerin, R., & Wagner, B. (2007). *ERPsim*. ERPsim Lab (erpsim.hec.ca), HEC Montreal.
26. Oz, B., Tran-Nguyen, K., Coursaris, C. K., Robert, J., & Léger, P.-M. (2020). Using digital nudges on analytics dashboards to reduce anchoring bias. In *SIGHCI 2020 Proceedings* (p. 3).
27. Léger, P.-M., Davis, F. D., Cronan, T. P., & Perret, J. (2014). Neurophysiological correlates of cognitive absorption in an enactive training context. *Computers in Human Behavior, 34,* 273–283. https://doi.org/10.1016/j.chb.2014.02.011
28. Karran, A., Demazure, T., Léger, P.-M., Labonte-LeMoyne, E., Sénécal, S., Fredette, M., & Babin, G. (2019). Towards a hybrid passive BCI for the modulation of sustained attention using EEG and fNIRS. *Frontiers in Human Neuroscience, 13*. https://doi.org/10.3389/fnhum.2019.00393
29. Whitin, T. M. (1955). Inventory control and price theory. *Management Science, 2*, 61–68. https://doi.org/10.1287/mnsc.2.1.61
30. Benzion, U., Cohen, Y., Peled, R., & Shavit, T. (2008). Decision-making and the newsvendor problem: An experimental study. *Journal of the Operational Research Society, 59*, 1281–1287. https://doi.org/10.1057/palgrave.jors.2602470
31. Supriatna, A., Hertini, E., Saputra, J., Subartini, B., & Robbani, A. A. (2019). The forecasting of foreign tourists arrival in Indonesia based on the supply chain management: An application of artificial neural network and Holt Winters approaches. *International Journal of Supply Chain Management, 8*, 156.
32. Müller-Putz, G. R., Riedl, R., & Wriessnegger, S. C. (2015). Electroencephalography (EEG) as a research tool in the information systems discipline: Foundations, measurement, and applications. *Communications of the Association for Information Systems, 37*. https://doi.org/10/ghtd97
33. Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review, 5*, 3–55.
34. Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience & Biobehavioral Reviews, 96*, 353–366.
35. Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly: Management Information Systems, 24*, 665–694. https://doi.org/10.2307/3250951
36. Tapal, A., Oren, E., Dar, R., & Eitam, B. (2017). The sense of agency scale: A measure of consciously perceived control over one's mind, body, and the immediate environment. *Frontiers in Psychology, 8*, 1552. https://doi.org/10.3389/fpsyg.2017.01552
37. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*, 319–340. https://doi.org/10.2307/249008
38. Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4*, 53–71.
39. Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2023). Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance. *Frontiers in Computer Science, 5*.
40. Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly, 25*, 351–370. https://doi.org/10.2307/3250921
41. Obar, J. A., & Oeldorf-Hirsch, A. (2020). The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society, 23*, 128–147.