Reliability and Efficiency of Human - Automation Interaction in Automated Decision Support Systems

Jinyu Li *

Department of Mathematics, University of Washington, Seattle, United State

* Corresponding Author Email: jli35@uw.edu

Abstract. The goal is to analyze the interactions between humans and automated decision support systems (DSS) in different working environments. The review emphasizes their results and factors affecting efficiency and reliability including the correctness of decisions and how people rely on automation under limited time constraints. By reviewing these research papers, the investigation shows how time constraints and operator reliance on DSS influence the performance of decision making. DSS have some potential to mitigate negative effects of consistent work and enhance performance, but problems remain This suggests that while automation can help improve decision accuracy, the optimal integration of human intuition and automation needs more development.

Keywords: Human–automation interaction, decision making, human operator engagement, operational efficiency.

1. Introduction

With the tremendous advancements in Artificial Intelligence (AI), the revolution fueled by it has led to more automated innovations in various fields. Not only to create more automatable systems, but also to improve the human-computer interaction environment based on what is currently available. Scenarios where human operators work with the assistance of automated decision aids in both regular and professional work situations are already commonplace. For example, airport security personnel inspect passenger luggage with the aid of X-rays. Similarly, in some places where the identity of people needs to be verified, the inspectors will assist in screening through scanners or face recognition.

Automated decision support systems (DSS) are largely involved in these efforts. In general, DSS provides users with information about the real state of the world based on the automatic processing and evaluation of environmental information [1]. Such systems can vary depending on different purposes, but all are designed to provide straightforward and referable results so that people can make decisions based on this information.

DSS are ideally thought to make work faster, more accurate, and reduce human workload. However, considering the environment in which assistance systems are applied, the factors that need to be considered for the work between human and assistance systems increase. As it presents complex challenges, the dynamics of human-automation interaction need a deeper examination of environmental and operational factors that may affect the overall efficiency of the work, such as time constraints and the use of operators. This exploration is essential not only to maximize the potential benefits of AI-driven automation, but also to identify and mitigate any adverse effects on workflows and human operators.

Table 1. Summary.

Author	Keywords	Main Theme	Results
T. Rieger and D. Manzey	Temporal stress, DSS, decision-making	Impact of temporal stress on DSS interaction	Time pressure reduces decision quality, but providing DSS reduces this effect if participants made their own decision before DSS's suggestion.
M. L. Bartlett and J. S. McCarley	Automated assistance, decision-making strategy	Strategy in automated decision assistance	Participants with automatic assistance showed little difference in the results of using automatic assistance, indicating that the performance of automatic assistance was poor.
D. Buser, Y. Sterchi and A. Schwaninger	Work duration, performance, security screening	Effect of long working hours on performance in security screening	Screeners were able to maintain consistent sensitivity and accuracy in detecting threats in baggage during continuous work
E. Alberdi, A. Povyakalo, L. Strigini, and P. Ayton	CAD, medical decision-making, reliability	Impact of incorrect CAD outputs on decision-making	Users use CAD as a detection aid, and a classification or diagnostic aid. The incorrect CAD output will affect the user's decision making
E. Alberdi, A. Povyakalo, L. Strigini, and P. Ayton	Vigilance, attention, task performance	Sensitivity and attention in vigilance tasks	Decline in vigilance and sensitivity over time is confirmed, and attention lapses is a key factor leading to decrement in performance.

2. Case Overview

2.1. Temporal Stress and Decision Support Systems

Rieger, Tobias, and Dietrich primarily investigated the effects of temporal stress on human interaction with an automated decision support system (DSS) and the associated performance consequences. Since the results of previous studies may be a product of extreme temporal stress conditions (i.e., 2 s for complex visual stimuli). With reference to previous relevant studies, the authors of this paper took the step of revisiting this issue using less extreme and more realistic temporal stress variations, giving participants a realistic opportunity to examine the temporal stimuli themselves. In addition, based on ideas derived from previous research, Rieger, Tobias, and Dietrich explored and validated these results, considering operator compliance and reliance on DSS within the experimental results. Then, two experiments that only differed in the order in which images were displayed and automatic suggestions yielded. Participants were randomly assigned to use one of manual decision making, 75% reliable DSS, and 95% reliable DSS [1]. Through two sets of experiments, comparing data from percent correct (PC), Signal detection (d'), Compliance (C), and Reliance (R), the researchers found that time pressure reduces the quality of decision-making [1]. Based on the results from Experiment 2 (participants made their own decision before DSS's suggestion), providing DSS (75% and 95% reliable DSS) can reduce this effect, but DSS should be provided after human decision making in advance. This is because under high time constraints, participants would rely more on the DSS for decision making. Therefore, the percentage of correctness would not be affected significantly. Finally, suggestions for future experiments are made. The authors suggest that this may be since if humans make decisions after seeing the results of the DSS they would prefer to have to make their own decisions rather than relying on the results of the DSS.

2.2. Human Strategy in Automated Assistance

Bartlett and McCarley examine the strategy of human operators interacting with highly through three sets of experiments. Participants performed a binary signal detection task that required them to

classify random dot images as blue or orange as the dominant color. Experiments 1 and 3 were performed with 93% reliable automated decision assistance, while Experiment 2 provided only binary cues [2]. Experiment 3 was a close replication of Experiment 1, but with the addition of a scoring system that allowed participants to track their performance. Finally, the researchers compared automated-assisted performance with seven statistical models for predictive collaborative decision making.

The authors found that Automation-aided participants in Experiment1 had almost no difference in results using automated assistance, and this result was demonstrated by Experiment 2 [2]. The authors concluded that the performance of automated assistance was very inefficient. Since there was little difference in the results of Experiments 1, 2, and 3, the performance of the automated assistance was confirmed to be very inefficient.

2.3. Work Duration and Performance in Security Screening

Buser, Sterchi, and Schwaninger study examines whether the performance of security screeners working with the help of X-ray to check passengers' luggage is affected by long working hours. The experiment divided the participants into two groups, with one group scanning continuously for one hour, while the other group took a 10-minute break every 20 minutes [3]. The objective was to evaluate whether these breaks would counteract any decline in performance. Since the experimental results were similar for both sets of experiments, this unexpected result showed that screeners were able to maintain consistent sensitivity and accuracy in detecting threats in baggage during continuous work. Such findings challenge the common assumption that performance necessarily declines with prolonged periods of uninterrupted continuous mental work.

2.4. Accuracy and Reliability of Computer-Aided Detection

Alberdi, Povyakalo, Strigini, and Ayton investigated the impact of incorrect computer output on the reliability of human user decisions and wanted to find possible ways to improve human computer CAD (computer-aided detection) interactions. The experiment is divided into two groups, the first group of experiments will consist of 20 film-readers, and they would provide "recall/no recall" decisions for each case. They will observe 60 sets of mammograms (30 of which contained cancer) which contained incorrect output generated by CAD (50%) [4]. the second group consisted of 19 different readers and watched the same set of cases under similar conditions, but only without the computer output.

As indicated above, the average sensitivity of the human readers in Study 1 was very low: 52% (minimum: 27%; maximum: 70%) [4]. In Study 2 the average reader sensitivity was still relatively low: 68% (minimum: 50%; maximum: 87%). Both sets of experiments confirm that the proportion of correct human decisions depends heavily on whether a case is normal or cancerous, and whether the CAD tool correctly prompts it or misses it. Thus, it suggests that the CAD output does have an impact on readers' decisions. The authors suggest that the CAD output (especially the absence of prompts) may be used by readers as a reassurance that they will make a "no recall" decision under normal circumstances. This means that the reader may use CAD not only as an inspection aid, but also as a classification or diagnostic aid, which is not what the tool is designed for. More importantly, these tools may undermine the reliability of human decision making in ways, and this may lead to potential damaging consequences for patients.

2.5. Attention and Vigilance in Monitoring Tasks

McCarley and Yamani’s study was designed to verify whether the loss of sensitivity in the vigilance task was spurious, and whether the lack of attention led to a failure of vigilance. The experiment consisted of 99 participants [5]. If the gap between a pair of probes embedded with dynamic punctate visual noise exceeds the standard value of 2 cm, ask the participant to press the space bar on each trial, otherwise stop responding [5]. Besides, Participants would indicate whether their judgment was correct or incorrect based on not receiving post-test feedback, and they would be

immediately asked to complete the subjective Workload Assessment Technique (SWAT). The results suggest that altered prejudice, loss of sensitivity, and attention deficit may lead to a decrease in vigilance. The loss of sensitivity in the vigilance task did exist and decreased in a fairly short period of time. Furthermore, they confirm that attention deficit is a source of decreased vigilance.

3. Discussion

The series of research reviewed has one common limitation across these studies. These studies are still mainly in the somewhat artificial setting of laboratory experiments, which may not capture the nuances of real-world scenarios where environmental and psychological factors vary widely. For example, ever though Riegel, Tobias, and Dietrich's study is rigorously designed, the environment cannot fully simulate the stress and unpredictability of real-life decision-making situations. Airport security or medical professionals will be very cautious and conscientious in detecting danger, whereas participants may not have a similar awareness. Such disconnect may affect the generality of the conclusions from those papers, especially when professionals have decision fatigue and cognitive overload under more variable and prolonged stress. Similarly, Buser, Sterchi, and Schwaninger's study challenges the assumption that performance declines over time, but this required further validation in real-world settings. Despite those limitations, Rieger, Tobias, and Dietrich's study reveal the interactions between human operators and automated systems under time pressure. Bartlett and McCarley's experiment results suggest that human performance is a bit better with DSS aided, but participants used automated aids very inefficiently. Their findings coincide in that the combination of humans and automation tends to be less effective than automation alone. In addition, the results of their experiments are consistent and may reflect that participants may have less reliability on the results of the assistance system than if they make their own decisions first and then receive the results from the assistance system. Under time pressure, the best strategy is for a human to make the decision first and then make a judgment based on the results given by DSS. Together, these studies have enlightened us that humans can make use of DSS in decision-making and have more accurate and convincing decisions.

The research by Alberdi, Povyakalo, Strigini, and Ayton implies that there is a need for highly accurate and reliable systems, since people could be misled by decision making systems. Machines of the class of assistance system as DSS can help the accuracy of work. This coincides with Bartlett and McCarley's finding again. Moreover, McCarley and Yamani's study shows that human sensitivity and vigilance decline over time when working alone without an auxiliary system, and that these same factors can have a negative impact on productivity and correctness. This concern may be solved by Buser, Sterchi, and Schwaninger's findings. They showed a work scenario in which humans interact with an assistive system, and there is no difference in final performance between one-hour and other intermittent breaks. Combining the two results hypothetically suggests that humans can consistently perform well with the help of an assistive system and that the assistive system can reduce the negative effects of continuous work on performance.

These findings can guide fields such as healthcare and transportation, for which accuracy and effectiveness in decision-making are important. However, there is a need to be mentioned that over-reliance on automated processes can lead to a degrade in skill over time. In addition, the question remains whether humans should be excluded from the automation cycle. Highly accurate assistive systems will perform better on their own than human-automation interaction, but so far it is still unknown whether computer-assisted systems with 100% accuracy will ever emerge. In addition, decision makers tend to disuse DSS that are not perfectly reliable is well established [2]. For future research, people can extend these findings by exploring more realistic environmental conditions and different levels of automated reliability. Then, we can get more accurately identified factors that may influence human performance. With more of these factors, further research can be made to investigate the long-term impacts of DSS on human cognitive accuracy, vigilance, and decision fatigue. As DSS is integrated into key decision-making processes nowadays, it is important to address ethical issues

like transparency and accountability. Automated systems can be biased and require more rigorous testing and evaluation to ensure fairness and prevent errors that could have serious consequences.

4. Conclusion

In summary, the interaction between humans and these automation-based computer aided systems is currently indispensable, because, according to the results of these studies, the overall performance and efficiency deteriorates over time when humans work alone. The research findings indicate that human decision-making performance is more stable and consistent in longer time with automation system aided. At the same time, with the development of AI, interaction between humans and automation is more prevalent. This requires researchers to investigate a more reliable and accurate decision-making system to help people have better performance on their works. Although automated aided systems enhance work performance, people need to consider the time constraint factor which has a negative impact on daily work. Therefore, people need proper strategy in interacting with machines. The research findings suggest such negative influence can be reduced if the operator's decision sequence is changed with the help of a highly accurate assistance system. Together, these conclusions suggest that humans in performing decision-making tasks such as security screener and doctors need the help of reliable automated decision support system, and that it is better to make their own decision before getting information from the machine. Such strategies maximize the strength of interaction between humans and machines and improve efficiency and accuracy in critical decision-making work.

Reference:

[1] T. Rieger and D. Manzey, “Human Performance Consequences of Automated Decision Aids: The Impact of Time Pressure,” Human *Factors: The Journal of the Human Factors and Ergonomics Society*, p. 001872082096501, Oct. 2020, doi: https://doi.org/10.1177/0018720820965019.

[2] M. L. Bartlett and J. S. McCarley, “Benchmarking Aided Decision Making in a Signal Detection Task,” *Human* Factors*: The Journal of the Human Factors and Ergonomics Society*, vol. 59, no. 6, pp. 881–900, Mar. 2017, doi: https://doi.org/10.1177/0018720817700258.

[3] D. Buser, Y. Sterchi and A. Schwaninger, "Effects of Time on Task, Breaks, and Target Prevalence on Screener Performance in an X-ray Image Inspection Task," 2019 International Carnahan Conference on Security Technology (ICCST), Chennai, India, 2019, pp. 1-6, doi: 10.1109/CCST.2019.8888408.

[4] E. Alberdi, A. Povyakalo, L. Strigini, and P. Ayton, “Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography,” Academic Radiology, vol. 11, no. 8, pp. 909–918, Aug. 2004, doi: https://doi.org/10.1016/j.acra.2004.05.012.

[5] J. S. McCarley and Y. Yamani, “Psychometric Curves Reveal Three Mechanisms of Vigilance Decrement,” Psychological *Science*, vol. 32, no. 10, pp. 1675–1683, Sep. 2021, doi: https://doi.org/10.1177/09567976211007559.