PAPER REVIEW NR 1

Paper:

Leichtmann, B., Humer, C., Hinterreiter, A., Streit, M., & Mara, M. (2023). Effects of Explainable Artificial Intelligence on trust and human behavior in a high-risk decision task. Computers in Human Behavior, 139, 107539. https://doi.org/10.1016/j.chb.2022.107539

Reviewer:

Giovanni Filomeno

I confirm that I have read the paper and written the following texts myself ☒

1. Thematic focus

The paper aims to study the effect of different variables (such as AI knowledge, knowledge of mushroom, etc.) on the participant decision about taking or leave a mushroom. The participant where feeded with the information of an AI mushroom classificatory.

2. Foundations

- Explainable AI
- Deep Learning for Image recognition
- Human-robot interaction

3. Method

The paper implemented a classification network based on ResNet50 trained with pre-processed images taken from different sources.

After that, a set of eligibility participant was chosen and informed about the AI-assisted mushroom-picking task as well as about the scenario.

4. Key results

- Educational intervention had no statistically significant effect on any of dependent variable.
- Participants with visual explanations trusted the AI less.
- Domain-specific (i.e., mushroom) did not affect performance in the picking task.

5. Practical implications for AI or robotics

By demonstrating that intuitive, explainable AI features exert the most significant influence, the findings advocate for prioritizing the development of AI and robotics systems that emphasize explainability and user-centric design. This approach is essential to enhancing trust in future applications, highlighting the importance of transparency and accessibility in the interface between users and AI technologies.

6. Strengths of the paper

The strength of the paper is in the methodological rigor for example into the division of the participants into groups or the deliberate introduction of errors in the AI's classification to investigate scenarios of overtrust.

7. Weaknesses of the paper

The paper is limited by its orientation on a virtual, rather than real, scenario which lead to a weak generalizability of the results to an actual high-risk decision-making context. Moreover, the paper focuses visual explanations and educational interventions potentially ignoring other decision-making factors (e.g., previous experience with AI systems). Finally, the paper does not further investigate the long-term impact of the interventions on user trust which may lead to a very short usability of the results.

8. Personal learnings

I personally liked the rigorous methodology and the attentions on details regarding the algorithm/answer. It highlighted the significance of careful planning and execution in experimental design to explore complex interactions between humans and AI systems. I learn about how to structure research in robopsychology.