The Illusion of Understanding: Mitigating Cognitive Bias in Interactive AI Interfaces for Medical Decision-Making

Bachelor/Master project "Information Systems"

Table of Contents

Prof. Dr. Mario Nadj
University of Duisburg-Essen
Faculty of Computer Science

Chair and Website
Information Systems and AI in Marketing
ai.ris.uni-due.de

Address Campus Essen
Universitätsstraße 9
D-45141 Essen

www.uni-due.de

1: Project Overview

Background: Medicine is perhaps the domain with the highest stakes for using artificial intelligence (AI). Responsibility for human lives warrants careful weighing of factors like quality of life, chances of treatment success, and the patient's own wishes. Domains like visual analytics (VA) and explainable AI (XAI) have aimed to increase the transparency of AI-based clinical decision support systems (CDSSs) to make them more understandable and useful. However, in successfully increasing trust and perceived understanding, overreliance on algorithmic inputs has become a major problem.

Project Goal: Develop and implement an interactive interface integrating VA and XAI for explaining AI recommendations in medicine. The interface should integrate techniques that stop clinicians from relying too much on these recommendations, nudging them to engage critically with the AI to detect potential errors.

Highlights: Develop the interface from scratch; engage with fundamental ideas about human reasoning; freedom to pursue individual solutions; close cooperation with the chair team

2: Project Background

AI for medical decision-making has gained traction over the last few years. From initial problems of algorithm aversion and lack of trust that XAI has attempted to solve, to more and more performant models that in some domains like cancer recognition rival human experts (see Sharma et al., 2024), further AI adoption throughout the healthcare sector seems imminent. Two related research streams from the XAI and VA domains however, have highlighted that high trust and high understanding are not causally related (Ha et al., 2024, Bansal et al., 2021). Rather, high trust and high subjective understanding can be signs of overreliance. When users perceive XAI or VA methods as signs of increased transparency, they might be less inclined to properly question AI recommendations. Thus, faulty reasoning by the AI could persist and lead to harm for patients.

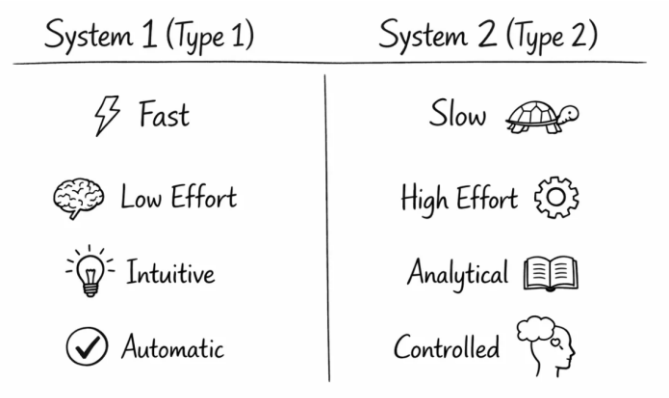

Figure 1: System 1 and System 2 (Teaching and Learning Process | Exploring Dual Process Theory).

This tendency can be explained through the contents of dual-process theory (Kahnemann, 2011). It proposes that humans employ the so-called System 1 in situations where they need to decide quickly and intuitively, while System 2 is used for more analytical, in-depth thinking (see Figure 1). The reliance of System 1 on heuristics opens it up to various cognitive biases with detrimental effects on decision quality.

Literature on how to mitigate such biases is abundant and precedes the current AI field. Proposals include changing the interaction protocol to show AI recommendations only after the human has made an initial decision, implementing a fixed delay before showing the recommendations, or issuing a short warning, when the system finds that a user has not paid attention to the explanations (Bertrand et al., 2022). However, these conflict with goals like timely decision-making and a positive usability experience.

3: Project Objective

The objective of this Bachelor/Master project is to develop and implement an interactive interface for a medical decision-making task in oncology that incorporates strategies for bias mitigation.

At the start of the project, students will conduct a scoping review on interactive interfaces in medicine and bias mitigation techniques. Based on the input from the literature, a series of expert interviews with oncologists will be conducted for requirements elicitation. Goal-directed task analysis (GDTA) will be applied as a framework to structure these interviews (Endsley et al., 2001; Dos Santos & Son, 2024). GDTA was developed to analyze different levels of situational awareness along a range of goals and sub-goals. Through the lens of situational awareness, we hope to gain a better understanding of the factors leading to cognitive biases in medical decision-making.

The interviews for requirements elicitation will be conducted in two phases. The first phase serves to build a fundamental understanding of decision-making in oncology, the technologies used, their evaluation of the CDSSs currently in use and the potential for improvement. In the second phase, students prepare more specific questions and the interviewees are asked to propose prototypical designs themselves and evaluate mock-ups previously created by the students.

Following these two formative steps, the students will develop and implement an interactive interface that includes elements of VA and XAI. This interface will incorporate visualizations of the AI's predictions that allow users to explore and drill deeper for more detailed information. From the interviews, the students will elicit a set of situations where overreliance would be potentially harmful and that are non-obvious to detect in practice. During the development of the interface, students will also develop interview guidelines for a summative evaluation of the interface with the aforementioned situations. The final evaluation will be

conducted with a different group of oncologists, with the aim of investigating which design features lead them to detect faulty reasoning, and where they fail to do so. These interviews will apply the think-aloud method (Eccles & Arsal, 2017), where participants interact with the proposed artefact and verbally express their thoughts during the interaction.

The objectives of this specific project are the following:

- **Developing** an interactive interface for medical decision-making
- **Implementing** a treatment response prediction model for the interactive interface
- **Investigating** when clinicians exhibit low situational awareness, and how the interface can be designed to increase awareness and thus improve reliance

Depending on group size, adjustments to the project scope can be made. This will be announced at the kickoff and in subsequent milestone meetings.

Each finished artifact has to be properly documented to facilitate understanding and further development by students in follow-up projects.

4: Project Registration and further Information

All Bachelor- and Master students interested in this IS-project may apply via our online application form. The link to the application will be **available here** (Link hinterlegt) as of the start of the application period on **Wednesday, 25.03.2026.** The deadline for the application is **Wednesday, 08.04.2026, 23:59.**

For successful completion of the project, Master students receive 12 ECTS (corresponding to 360 working hours) and Bachelor students receive 6 ECTS (corresponding to 180 working hours)

- Master students will invest 18 hours per week into the project.
- Bachelor students will invest 9 hours per week into the project.

In addition to project-specific deliverables during the process, we also require some fixed deliverables from each group. On the 05.06.2026, the project team needs to submit a one-page management summary on the current state of the project. A template will be provided by the Chair. As final deliverables, the project team has to submit three items until 24.07.2026. These are (1) a 5-minute video presenting the results (we will provide an example video for orientation), (2) a poster summarizing the results (whether the poster has to be presented will be determined through a centralized vote among the Information Systems Chairs), and (3) a project report along with the project code. The project report will contain the final results on 10-20 pages. We will provide a template for the final project report. The code needs to be accessible as Open Source via GitHub with appropriate documentation. We want to put up your posters (2) at the Chair premises to highlight your achievements. The final video (1) and a link to your code (3) will be shown on the Chair page.

5: Project Schedule (Dates are to be announced)

Deliverables for each milestone are marked in blue.

6: Grading

The final grade will be calculated as follows based on the sum of the milestones:

- Milestone 1: 5%
- Milestone 2: 20%
- Milestone 3: 5%
- Milestone 4: 15%
- Milestone 5: 15%
- Milestone 6: 40%

7: Sources

- Sharma, P., Nayak, D. R., Balabantaray, B. K., Tanveer, M., & Nayak, R. (2024). A survey on cancer detection via convolutional neural networks: Current challenges and future directions. Neural Networks, 169, 637-659.
- Ha, S., Monadjemi, S., & Ottley, A. (2024). Guided By AI: Navigating Trust, Bias, and Data Exploration in AI-Guided Visual Analytics. In Computer Graphics Forum, 43(3), 15108-15120.
- Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T. & Weld, D. (2021). Does the whole exceed its parts? the effect of ai explanations on complementary team performance. In Proceedings of the 2021 CHI conference on human factors in computing systems (pp. 1-16).
- Bertrand, A., Belloum, R., Eagan, J. R., & Maxwell, W. (2022). How cognitive biases affect XAI-assisted decision-making: A systematic review. In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (pp. 78-91).
- Dos Santos, V., & Son, C. (2024). Identifying firefighters' situation awareness requirements for fire and non-fire emergencies using a goal-directed task analysis. Applied ergonomics, 114, 104136.
- Endsley, M. R. (2001, November). Designing for situation awareness in complex systems. In Proceedings of the Second International Workshop on symbiosis of humans, artifacts and environment (pp. 1-14).
- Eccles, D. W., & Arsal, G. (2017). The think aloud method: what is it and how do I use it?. Qualitative Research in Sport, Exercise and Health, 9(4), 514-531.
- Kahneman, D. (2011). Thinking, fast and slow. macmillan.