Effects of XAI on Perception, Trust, and Acceptance

appliedAI Seminar — Further Methods and Issues in XAI

Maternus Herold

05.10.2023

APPLIEDAI INSTITUTE FOR EUROPE

Agenda

1. Introduction
2. Effect of XAI on Cognitive Load
3. Effect of XAI on Trust
4. Conclusion

Introduction

Effects of XAI on Perception, Trust and Acceptance

XAI is the ability to explain the way in which an algorithm works in order to understand how and why it has delivered particular outcomes [4].

Effects of XAI on Perception, Trust and Acceptance

XAI is the ability to explain the way in which an algorithm works in order to understand how and why it has delivered particular outcomes [4].

BUT

Recent XAI approaches have mainly been designed by developers for developers, as opposed to addressing the end-user [5].

Important Factors for establishing TRUST

Honesty & Transparency

Competence

Integrity

Clear Communication

Influential Factors fo Acceptance and Perception

Ease of use

Compatibility with Goals

Effort and Time Savings

Feedback Loop

Comprehensibility

Problem Setting and Motivation

Problem Setting

- More complex systems
- Understanding requires expertise $\Rightarrow$ black-boxes
- Challenging explanations have a negative effect on perception
- Transparency is fundamental to trust and acceptance

Problem Setting and Motivation

Problem Setting

- More complex systems
- Understanding requires expertise $\Rightarrow$ black-boxes
- Challenging explanations have a negative effect on perception
- Transparency is fundamental to trust and acceptance

Motivation

Use XAI to provide

- users with an understanding on how the algorithm generates its results
- assurance and build confidence that AI systems works well
- an indication of the *right amount / appropriate level* of trust into the system

Problem Setting and Motivation

Problem Setting

- More complex systems
- Understanding requires expertise ⇒ black-boxes
- Challenging explanations have a negative effect on perception
- Transparency is fundamental to trust and acceptance

Motivation

Use XAI to provide

- users with an understanding on how the algorithm generates its results
- assurance and build confidence that AI systems works well
- an indication of the *right amount / appropriate level* of trust into the system

→ XAI should be perceived as mentally efficient [1].

Expectations on the Talk

Exemplify effects of XAI **vs.** using XAI w.r.t. certain attributes

Expectations on the Talk

Exemplify effects of XAI **vs.** using XAI w.r.t. certain attributes

→ Why sociotechnical factors are important

→ Not every type of explanation is appropriate

→ Situations when explanations enhance the performance

Effect of XAI on Cognitive Load

Impact of XAI on Cognitive Load

*"Do XAI **explanation types** affect end-users' cognitive load and what are the ramifications for task performance and task time?" [2]*

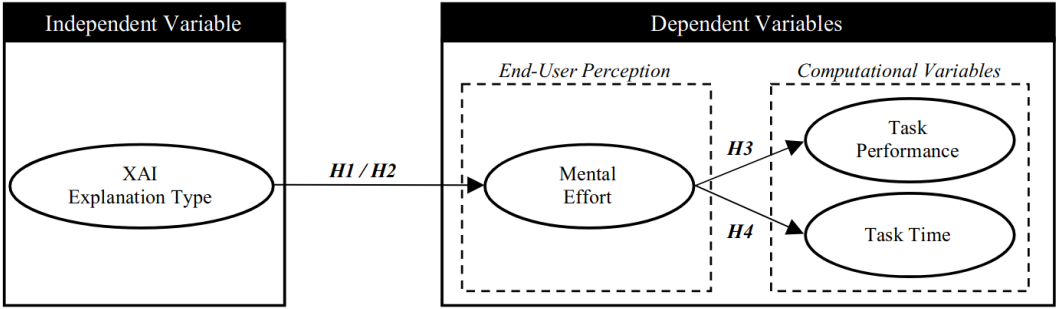
Empirical study, in proceedings of the *European Conference on Information Systems 2023*.

Different Explanation Types

Type[1]	Description[1]	Exemplary Implementations[2]
How	Holistic representation of how the ML model's inner decision logic operates – global explanation type.	ProfWeight, SHAP, DALEX, Saliency
How-To	Hypothetical adjustment of the ML model's input yielding a different output (counterfactual explanation) – local explanation type.	DiCE, KNIME, PDP
What-Else	Representation of similar instances of inputs that result in similar ML model outputs (explanation by example) – global explanation type.	SMILY, Alibi
Why	Description of why a prediction was made by informing which input features are relevant to the ML model – local explanation type.	SHAP, LIME, ELI5, Anchor
Why-Not	Description of why an input was not predicted to be a specific output (contrastive explanations) – local explanation type.	CEM, ProtoDash

Legend: *1)* Types and definitions adapted from Mohseni et al. (2021); *2)* exemplary classification of frequently mentioned XAI implementation packages based on Das and Rad (2020), Dwivedi et al. (2022), Liao and Varshney (2022), and Mohseni et al. (2021).

Effects of Explanations on Performance, Time, and Mental Effort



$$\text{Mental Efficiency} = \frac{Z_{\text{perf.}} \cdot Z_{\text{time}} - Z_{\text{effort}}}{\sqrt{2}}$$

Study Design: Medical Decision Support System

Input Image:

Explanation:

Description of Explanation:

In the center section, the system's decision-making process is explained. Here, the light gray area with black border represents the area that the system considers relevant to the overall classification of Covid-19 or no Covid-19. The rest of the image is not considered as relevant.

Rate your perceived level of mental effort during this task.

Extremely low | Low | Somewhat low | Neutral | Somewhat low | High | Extremely high

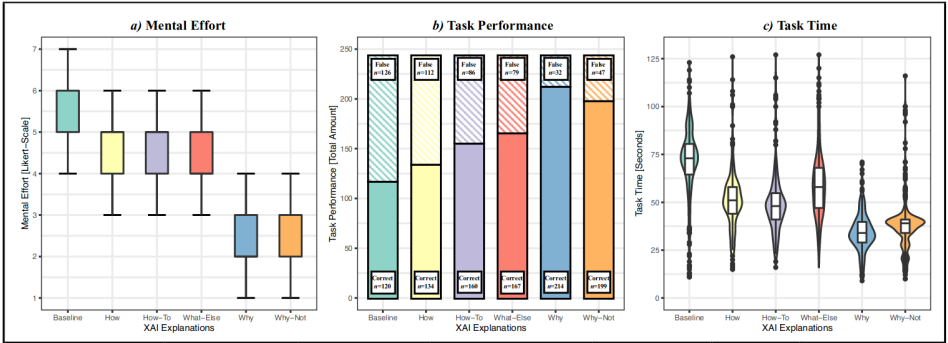
Is this chest diseased with Covid-19?

Please use the information provided above to solve this task.

Yes | No

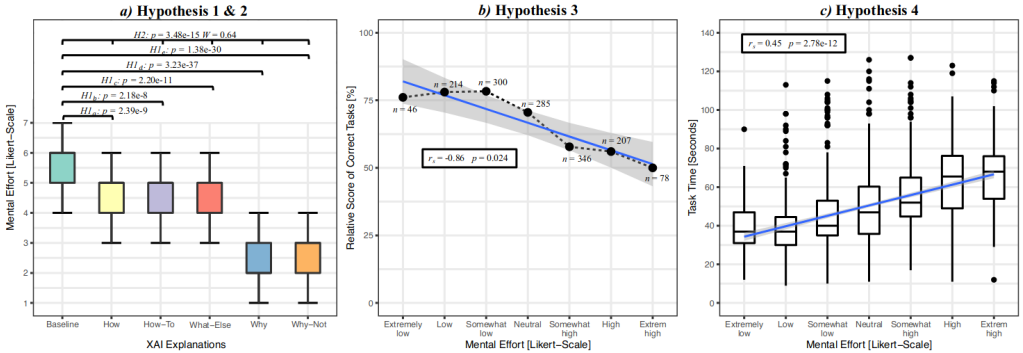
Study: $n = 271$ of novice AI users, all enrolled as medical students

Results 1/2



Type	Median Mental Effort[1]	SD Mental Effort[2]	AVG Task Performance[3]	AVG Task Time[3]	SD Task Time[2]	AVG Mental Efficiency[3,4]
Baseline	**6**	**1.34**	0.49	**72.59**	**26.15**	−0.34
How	5	1.15	0.55	51.68	17.49	−0.15
How-To	5	1.05	0.65	49.84	16.71	−0.11
What-Else	4	1.20	0.68	60.10	18.49	−0.08
Why	2	0.92	**0.87**	34.50	10.25	**0.34**
Why-Not	3	0.90	0.81	38.92	15.40	0.23

Results 2/2



H.	Description	Test	*p*-Value[1,2]	Dec.[3]
H1	Mental effort of every XAI explanation is lower than baseline.	Kruskal-Wallis	Cf. Figure *a)****	Acc.
H2	Mental effort of every XAI explanation differs.	Friedman	3.48e-15***	Acc.
H3	Decreased mental effort results in increased task performance.	Spearman	0.024**	Acc.
H4	Decreased mental effort results in decreased task time.	Spearman	2.78e-12***	Acc.

Effects of Explainability

- **Mental Effort:** *Why* / *Why-Not* $\gg$ *How* / *How-To*
- **Task Performance:** *Why* / *Why-Not* $\gg$ *How*
- **Task Time:** *Why* / *Why-Not* outperformed others
- **Mental Efficiency:** only local explanations with a positive score

Effects of Explainability

- **Mental Effort:** *Why* / *Why-Not* $\gg$ *How* / *How-To*
- **Task Performance:** *Why* / *Why-Not* $\gg$ *How*
- **Task Time:** *Why* / *Why-Not* outperformed others
- **Mental Efficiency:** only local explanations with a positive score

→ Adapt Explanations to Users and Use Case

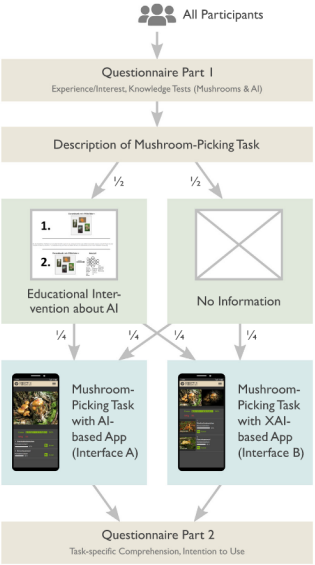
Effect of XAI on Trust

Impact of XAI on Trust

There are two routes to user comprehension of AI-based decisions to achieve improved performance and trust: improving users' general AI knowledge and enabling the AI system to explain its decisions [3].

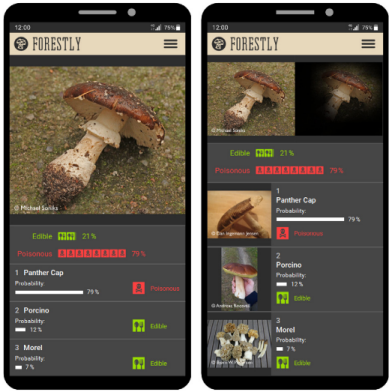
Empirical study, published in *Computers in Human Behavior 139, 2023*.

Study Design: Mushroom Picking

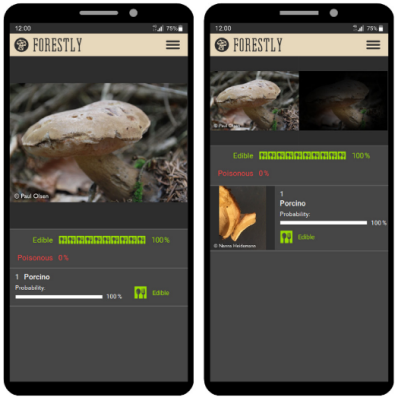


- Prior education on AI
- Decide whether or not to pick a mushroom
- Decide whether or not to eat a mushroom
- UX questionnaire

Study Design: Mushroom Picking

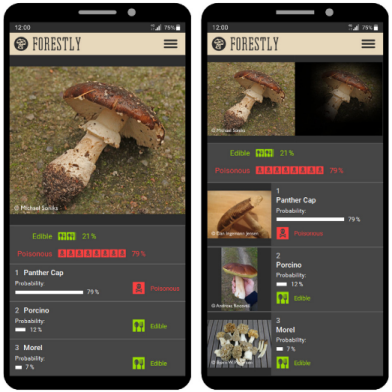


(a) Plain interface (b) XAI interface

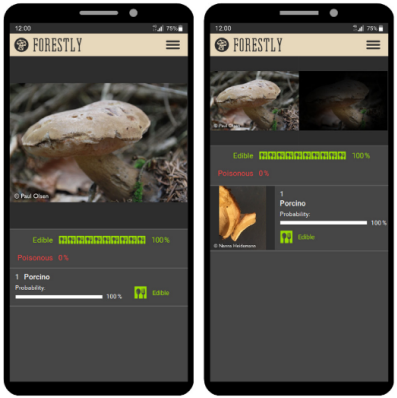


(a) Plain interface (b) XAI interface

Study Design: Mushroom Picking



(a) Plain interface (b) XAI interface



(a) Plain interface (b) XAI interface

Classifier had an accuracy of 71%, which was intended.

Findings

- Educational intervention had not effect
- Positive effect of explanations on performance
- Participants **without explanations** reported higher trust and comprehension

Findings

- Educational intervention had not effect
- Positive effect of explanations on performance
- Participants **without explanations** reported higher trust and comprehension
- Participants with higher trust did worse in mushroom classification

Findings

- Educational intervention had not effect
- Positive effect of explanations on performance
- Participants **without explanations** reported higher trust and comprehension
- Participants with higher trust did worse in mushroom classification

→ Establishing trust via explanation is easier than via knowledge.

Findings

- Educational intervention had not effect
- Positive effect of explanations on performance
- Participants **without explanations** reported higher trust and comprehension
- Participants with higher trust did worse in mushroom classification

→ Establishing trust via explanation is easier than via knowledge.
→ Explanations help to understand the limits of the AI's performance / competencies.

Conclusion

Conclusion

- XAI improves task performance $\rightarrow$ benefits Acceptance & Perception
- Contradicting results have to be explained $\rightarrow$ Trust & Acceptance Issues
- Explanation types depend on the user have an effect on the mental effort
- Acceptance, Perception, & Trust build on transparency
- Trust has to be *calibrated*
- Explanations can help to obtain a realistic estimate of the systems competencies

Effects of XAI on Perception, Trust, and Acceptance

appliedAI Seminar — Further Methods and Issues in XAI

Maternus Herold

05.10.2023

APPLIEDAI INSTITUTE FOR EUROPE

References

[1] Zana Buçinca et al. **“Proxy Tasks and Subjective Measures Can Be Misleading in Evaluating Explainable AI Systems”.** In: *Proceedings of the 25th International Conference on Intelligent User Interfaces*. IUI ’20. New York, NY, USA: Association for Computing Machinery, Mar. 17, 2020, pp. 454–464. ISBN: 978-1-4503-7118-6. DOI: 10.1145/3377325.3377498. URL: https://doi.org/10.1145/3377325.3377498 (visited on 10/05/2023).

References ii

[2] Lukas-Valentin Herm. **“Impact Of Explainable AI On Cognitive Load: Insights From An Empirical Study”.** In: *ECIS 2023 Research Papers* (May 11, 2023). URL: https://aisel.aisnet.org/ecis2023_rp/269.

[3] Benedikt Leichtmann et al. **“Effects of Explainable Artificial Intelligence on Trust and Human Behavior in a High-Risk Decision Task”.** In: *Computers in Human Behavior* 139 (Feb. 2023), p. 107539. ISSN: 07475632. DOI: 10.1016/j.chb.2022.107539. URL: https://linkinghub.elsevier.com/retrieve/pii/S0747563222003594 (visited on 10/02/2023).

References iii

[4] Donghee Shin. **“The Effects of Explainability and Causability on Perception, Trust, and Acceptance: Implications for Explainable AI”.** In: *International Journal of Human-Computer Studies* 146 (Feb. 2021), p. 102551. ISSN: 10715819. DOI: 10.1016/j.ijhcs.2020.102551. URL: https://linkinghub.elsevier.com/retrieve/pii/S1071581920301531 (visited on 08/26/2023).

[5] Jasper van der Waa et al. **“Evaluating XAI: A Comparison of Rule-Based and Example-Based Explanations”.** In: *Artificial Intelligence* 291 (Feb. 1, 2021), p. 103404. ISSN: 0004-3702. DOI: 10.1016/j.artint.2020.103404. URL: https://www.sciencedirect.com/science/article/pii/S0004370220301533 (visited on 10/05/2023).