A Qualitative Exploration of User-Perceived Risks of AI to Inform Design and Policy

Lena Recki
lena.recki@h-brs.de
Institut für Verbraucherinformatik,
Bonn-Rhein Sieg University of Applied Sciences
Sankt Augustin, Germany

Dennis Lawo
dennis.lawo@verbraucherinformatik.de
Institut für Verbraucherinformatik,
Bonn-Rhein Sieg University of Applied Sciences
Sankt Augustin, Germany

Veronika Krauß
Verbraucherinformatik Research Group,
University of Siegen
Siegen, Germany

Dominik Pins
Fraunhofer Institute for Applied
Information Technology FIT
Germany

ABSTRACT

AI systems pose unknown challenges for designers, policymakers, and users which aggravates the assessment of potential harms and outcomes. Although understanding risks is a requirement for building trust in technology, users are often excluded from legal assessments and explanations of AI hazards. To address this issue we conducted three focus groups with 18 participants in total and discussed the European proposal for a legal framework for AI. Based on this, we aim to build a (conceptual) model that guides policymakers, designers, and researchers in understanding users' risk perception of AI systems. In this paper, we provide selected examples based on our preliminary results. Moreover, we argue for the benefits of such a perspective.

KEYWORDS

Risk perception, AI-Systems, Explainable AI, Empirical Study

1 INTRODUCTION

With the growing influence of AI-powered systems, automated decision-making systems, and their real-world consequences for people, the call for regulations and policy is increasingly coming into focus by policymakers and society. They aim to "provide AI developers, deployers and users with clear requirements and obligations regarding specific uses of AI" [7].

The need to understand the precarious nature of design and the central role of risk in AI design is also highlighted from a research perspective [13, 19, 20]. Here, research is predominantly influenced by two perspectives — the philosophical and the engineering view [14, 20]. While the first stresses the danger of future super-intelligent systems and the need for regulation, the second perspective stresses a solutionist approach. Our approach, in contrast, is inspired by a realistic stance [20] that focuses on the current capabilities of AI systems as learning algorithms, where risk handling is part of the design process and should be made explicit through user understanding, and user participation.

A vast majority of research, as well as the policy approach [9], are, however, not based on users' understandings of risks, but on experts' opinions. This is, however, problematic as it (1) non-experts perceive risks differently than experts [5, 22], and (2) neglects the role of user-participation in design [20].

To address this gap and inform and qualify designers [24] as well as policymakers [15] we conducted focus group workshops [17, 21] with 18 participants to understand their mental models and explanations about AI and its respective risks. All participants are based in Germany, are in an age range of 23 to 49, and 11 participants self-identified as male, while seven described themselves as female. Based on the 12 use cases presented by the EU in their regulatory framework [7] the sessions consisted of three tasks:

(1) Familiarization with 12 use cases (for instance AI application in robot-assisted surgery) from the EU regulatory framework and group discussion on experiences and opinions to establish a common ground.
(2) Sorting of the use cases into the four risk categories *unacceptable risk*, *high-risk*, *limited risk*, and *minimal risk* proposed by the EU Commission [7].
(3) Disclosure of experts' risk categorization and discussion of similarities and differences.

We transcribed the sessions and analyzed them based on the thematic analysis approach by Braun and Clarke [3]. For this purpose, we started to cluster thoughts and textual excerpts from the transcripts on a virtual whiteboard into initial themes.

In this position paper, we present the first results to support our claim for a stronger involvement of users in the risk assessment of AI. This approach does not only contribute to HCI's understanding of how users perceive AI systems [1] but also attempts to inform policymakers about misconceptions and blind-spot they might have. Overall we aim to develop a (conceptual) model that guides researchers, designers, and policymakers in incorporating users' risk perceptions.

Veröffentlicht durch die Gesellschaft für Informatik e.V.
in P. Fröhlich & V. Cobus (Hrsg.):
Mensch und Computer 2023 – Workshopband, 03.-06. September 2023, Rapperswil (SG)

https://doi.org/10.18420/muc2023-mci-ws16-383

2 PRELIMINARY FINDINGS & DISCUSSION

Our preliminary analysis shows how users perceive and explain the risk of AI. The resulting classification of risks into categories is quite similar between our participants and the experts of the EU, but the explanations differ. In the following, we introduce the findings and provide interesting examples.

2.1 Dimensions of Risk & Example *Designers' Responsibility for the Risk*

The analysis of the focus groups revealed five themes that describe dimensions used to allocate and further explain a risk.

- *When* - (1) risks during design and development, (2) procedural risks, and (3) distributive risks after user receipt.
- *Who* - (4) risks affecting individuals or (5) society as a whole.

Thereby, the *When* and the *Who* are mutually related, and in most cases, more than one theme is used to explain and understand the risk. Interestingly, users perceive most risks occurring before usage and during system development. In contrast, the legal framework [7] focuses more on regulating high-risk systems during use. Participants for example discussed the involved people in the design process:

> "If you think about, who feeds the algorithms and who develops the system? You have to check, are the people diverse?" (P13)

Therefore, a user-centered policy should also prioritize regulating and guiding system development and deployment. This provides an interesting starting point for HCI to disseminate results, as we made contributions to accessibility policies earlier [16].

Moreover, our participants paid little or no attention to economic, military, or legal risks [9, 12]. This could be problematic if systems are banned or restricted based on concerns, as users may feel patronized and unable to comprehend such regulations. Thus, a democratization of the discourse is crucial, as argued by Garvey [9]. However, participants had a limited understanding of AI capabilities, primarily focusing on questions about what AI is allowed to do and what it can do. This perspective can lead to underestimating risks or overemphasizing legal frameworks due to future use cases [14]. Therefore, promoting literacy on AI capabilities and risks is essential for informed decision-making and risk management.

2.2 Assessment of Risk & Example *Impact of the Risk in Relation to Human Decisions*

From the participant discussions, we identified six criteria for sorting use cases:

(1) Risk Occurrence Probability
(2) Impact of the Risk
(3) Impact of the Risk in Relation to Human Decisions
(4) Voluntary Use
(5) Abusive Use Probability
(6) Design Intention

The first three categories focus on assessing specific key risks, while the other three focus on overall use case assessment. These criteria overlap and represent different perspectives used for risk assessment. Participants made an interesting comparison between the impact of risks and human decisions. They recognized risks associated with both humans and AI systems, determining the higher probability of risk occurrence for either humans or machines. For example, in robot-assisted surgery, some participants preferred AI systems due to the computer model's extensive learning experience compared to a surgeon's limited professional years.

> "I see zero risk and would always prefer the AI system because of the computer model's learning experience; a surgeon can not collect in his life. We are only limited beings; he has a maximum of 40 professional years and can also have a bad day"(P6)

However, other studies [2] suggest that the mere fact that a machine or algorithm is making decisions and evaluating human behavior can trigger negative feelings. It is essential to consider whether mechanical or human requirements are more suitable for solving a decision, and who is more likely to experience the risk [18]. In practice, it might be challenging to define humans' and machines' exact risk occurrence probabilities concerning a given risk for each system. Nevertheless, we should try to find out whether users have fewer negative feelings about AI-powered systems if they know the probability of risk relative to humans. In doing so, we would like to point out that the goal is not to avoid risk, but provide means to engage users and allow them to more actively manage risks.

2.3 Towards a (Conceptual) Model of User-Centered Risk-Understanding

Many studies and surveys focus on specific applications or privacy risks [8, 11, 12, 19]. However, our approach addresses the general risks of AI-powered systems based on use cases and the EU framework [6]. We aim to develop a conceptual model or framework that considers end-user perspectives and differences in risk perception (see e.g., [22]), which is crucial for building trust in AI systems [10]. Designers and developers play a vital role in mitigating these risks by adopting user-centered practices and emphasizing all aspects of the AI life cycle [4]. Providing information about the design process, considerations, data used, and organizational aspects helps users make informed decisions. While previous approaches have primarily focused on technical perspectives [23] to address bias and harm along the AI life cycle [4], our user-oriented approach complements this by highlighting users' perceptions of risks and their consequences. By defining risks along the AI life cycle, users gain a better understanding of when specific risks may occur, enhancing explanations of AI.

3 CONCLUSION

The increasing prevalence of AI systems has led to calls for regulations, but there is a lack of understanding regarding how users perceive the risks associated with AI. Our research goal is to serve as an initial approach to bridge the gap between different perspectives, enabling designers, companies, and regulatory authorities to address user concerns. Our findings highlight the importance of the user perspective. The explanations of AI differ between users and experts, and there is a need to discuss uncertainties and potential negative consequences more openly. We aim to initiate a nuanced discourse about the risks and consequences of AI systems that enhances trust and highlights their benefits.

REFERENCES

[1] Fatemeh Alizadeh, Gunnar Stevens, and Margarita Esau. 2021. I don't know, is AI also used in airbags? *I-com* 20, 1 (2021), 3–17. https://doi.org/10.1515/icom-2021-0009

[2] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's Reducing a Human Being to a Percentage'; Perceptions of Justice in Algorithmic Decisions. *Conference on Human Factors in Computing Systems - Proceedings* 2018-April (1 2018). https://doi.org/10.1145/3173574.3173951

[3] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological.* (3 2012), 57–71. https://doi.org/10.1037/13620-004

[4] Shipi Dhanorkar, Christine T. Wolf, Kun Qian, Anbang Xu, Lucian Popa, and Yunyao Li. 2021. Who needs to know what, when?: Broadening the Explainable AI (XAI) Design Space by Looking at Explanations across the AI Lifecycle. *DIS 2021 - Proceedings of the 2021 ACM Designing Interactive Systems Conference: Nowhere and Everywhere* (6 2021), 1591–1602. https://doi.org/10.1145/3461778.3462131

[5] David Etkin. 2016. 3 - Disaster Risk. In *Disaster Theory*, David Etkin (Ed.). Butterworth-Heinemann, Boston, 53–101. https://doi.org/10.1016/B978-0-12-800227-8.00003-X

[6] European Commission. 2021. Proposal for a Regulation laying down harmonised rules on artificial intelligence | Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

[7] European Commission. 2022. Regulatory framework proposal on artificial intelligence | Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[8] Isaac J Gabriel and Easwar Nyshadham. 2008. A Cognitive Map of People's Online Risk Perceptions and Attitudes: An Empirical Study. (2008). https://doi.org/10.1109/HICSS.2008.6

[9] Colin Garvey. 2018. AI Risk Mitigation Through Democratic Governance: Introducing the 7-Dimensional AI Risk Horizon. *AIES 2018 - Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (12 2018), 366–367. https://doi.org/10.1145/3278721.3278801

[10] Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. 2021. Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. *FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (3 2021), 624–635. https://doi.org/10.1145/3442188.3445923

[11] Timo Jakobi, Maximilian von Grafenstein, Patrick Smieskol, and Gunnar Stevens. 2022. A Taxonomy of user-perceived privacy risks to foster accountability of data-based services. *Journal of Responsible Technology* 10 (7 2022). https://doi.org/10.1016/J.JRT.2022.100029

[12] Sabrina Karwatzki, Manuel Trenz, Virpi Kristiina Tuunainen, and Daniel Veit. 2018. Adverse consequences of access to individuals' information: an analysis of perceptions and the scope of organisational influence. *European Journal of Information Systems* 26, 6 (11 2018), 688–715. https://doi.org/10.1057/S41303-017-0064-Z

[13] Scott R Klemmer, Björn Hartmann, and Leila Takayama. 2006. How Bodies Matter: Five Themes for Interaction Design. *Proceedings of the 6th ACM conference on Designing Interactive systems - DIS '06* (2006), 140–149. https://doi.org/10.1145/1142405

[14] P. M. Krafft, Meg Young, Michael Katell, Karen Huang, and Ghislain Bugingo. 2020. Defining AI in Policy versus Practice. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. ACM, New York, NY, USA, 72–78. https://doi.org/10.1145/3375627.3375835

[15] Dennis Lawo, Thomas Neifer, Margarita Esau-Held, and Gunnar Stevens. 2023. Digital Sovereignty: What it is and why it matters for HCI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7.

[16] Jonathan Lazar. 2015. Public policy and HCI: making an impact in the future. *Interactions* 22, 5 (2015), 69–71.

[17] Moon-Hwan Lee, Da-Hoon Kim, Hyun-Jeong Kim, and Tek-Jin Nam. 2012. Understanding Impacts of Hidden Interfaces on Mobile Phone User Experience. *In Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design (NordiCHI '12)* (2012), 45–48. https://doi.org/10.1145/2399016.2399024

[18] Min Kyung Lee. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5, 1 (3 2018). https://doi.org/10.1177/2053951718756684

[19] Mengyao Li, Brittany E Holthausen, Rachel E Stuck, and Bruce N Walker. 2019. No Risk No Trust: Investigating Perceived Risk in Highly Automated Driving. (2019). https://doi.org/10.1145/3342197.3344525

[20] Aale Luusua and Johanna Ylipulli. 2020. Artificial Intelligence and Risk in Design. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. ACM, New York, NY, USA, 1235–1244. https://doi.org/10.1145/3357236.3395491

[21] Stephanie Rosenbaum, Gilbert Cockton, Kara Coyne, Michael Muller, and Thyra Rauch. 2002. Focus Groups in HCI: Wealth of Information or Waste of Resources? *CHI '02 extended abstracts on Human factors in computing systems - CHI '02* (2002). https://doi.org/10.1145/506443

[22] Paul Slovic. 1990. Perceptions of Risk: Reflections on the Psychometric Paradigm. (1990).

[23] Harini Suresh and John Guttag. 2021. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. *ACM International Conference Proceeding Series* (10 2021). https://doi.org/10.1145/3465416.3483305

[24] Allison Woodruff, Sarah E Fox, Steven Rousso-Schindler, and Jeff Warshaw. 2018. A Qualitative Exploration of Perceptions of Algorithmic Fairness. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018). https://doi.org/10.1145/3173574