Do Humans Trust AI or Its Developers? Exploring Benefits of Differentiating Trustees Within Trust in AI Frameworks

Marisa Tschopp
Titanium Research, scip AG
Zurich, Switzerland
mats@scip.ch

Nicolas Scharowski
Center for Cognitive Psychology and Methodology, University of Basel
Basel, Switzerland
nicolas.scharowski@unibas.ch

Philipp Wintersberger
TU Wien
Vienna, Austria
philipp.wintersberger@tuwien.ac.at

*corresponding author: Marisa Tschopp

Abstract
"Trustworthy" artificial intelligence (AI) is a globally discussed topic. Trust, distinct but related to trustworthiness, is considered a critical success factor in human-AI-interactions. However, the role of the trusted "partner" remains unclear: Is an AI system or the designers of AI the trusted peer in this relation, and does it even matter? Since trust in technology providers is not frequently evaluated, we conducted an online survey to explore the potential benefit of differentiating the actors of the trust framework in more detail. The results of this early stage research confirms that trust is a relevant construct in the adoption of AI technology, and participants trusted AI in general significantly more than technology providers that develop AI. Preliminary results, limitations and plans for future research will be discussed. We aim to refine the discussion around trustworthy AI through a better understanding of the designers' role in the trust framework with the goal to develop trust-relevant recommendations for companies designing "trustworthy" AI.

Author Keywords
trust; trustworthy AI; AI Trust; provider trust

Introduction
"Trust in technology/automation" has been discussed in different domains such as human factors engineering [6], or human-robot interaction [4], and the psychological con-

struct currently gains additional momentum due to the current "trustworthy AI" discussions taking place all over the world. For example, the words "trust" and "trustworthiness" appear over 100 times in a regulatory proposal by the European Commission [2], suggesting that trust and a beneficial use of AI have a close relationship. We argue that making statements about the relevance of trust requires a better understanding of *who* and *what* humans place their trust exactly in. Some researchers have started to question if AI can or should be trusted at all [1, 3], while others claim it being highly relevant [5].

From a human-centered perspective, one can ask: is it relevant in the context of AI applications whether users trust a tool or the developers of that tool? Would users make a difference in this regard? We attempt to contribute to two main issues: Shedding light on (1) the conceptual confusion of who the trusted actors are, accompanied by (2) re-evaluating if trust is a relevant concept for human-AI interactions at all. A more differentiated understanding ideally allows us to develop trust-relevant recommendations for companies designing AI.

Online Survey and Results

To investigate our research questions, we conducted an online survey with 111 participants (between 30 and 50 years old; 62 female, 42 male, 2 diverse, 5 did not answer). Participants are well educated and primarily from western countries. Their self-reported level of expertise in AI ($M = 4.69, SD = 2.24$) was slightly below, their expertise in technology ($M = 6.67, SD = 2.09$) slightly above the midpoint of a 10-point Likert scale, with no significant gender-differences. The convenience sample was distributed using mailing lists and social media channels.

We wanted to find out if trust is relevant and if *trust in AI* and *trust in providers of AI* are distinct dimensions (see Table 1). Further, we assessed how these two factors influence participants' willingness to rely on AI products (an automated vehicle and a voice assistant). Comparison of the two scales shows that *trust in AI* ($Mdn = 3, IQR = 1$) is significantly higher than *trust in providers* ($Mdn = 2.75, IQR = .75, z = -6.26, p < .001$) with a medium to large effect (Cohen's $d = .6$). However, a Kendall rank correlation analysis also shows a significant correlation between the two concepts ($r = .53, p < .001$). A Mann-Whitney-U test shows that participants who would use an automated vehicle rated their *general trust in AI* significantly higher than those who would not ($z = -4.703, p < .001, d = .52$, see Table 2). The same pattern follows for the digital assistant ($z = -3.065, p = .002, d = .3$). The dimension *trust in providers of AI* shows similar results. Those who would use automated vehicles ($z = -3.635, p < .001, d = .40$) or use digital assistants ($z = -2.457, p = .013, d = .24$) trust technology companies significantly more than those who claimed the opposite. Limitations include questionnaire design and the low number of participants.

Discussion and Future Research Agenda

Our results indicate that people can differentiate between *general trust in AI* and their *trust in providers of AI*. A qualitative review of the data supports this hypothesis, i.e., one participant stated: *"It is not the AI [I am skeptical about]. It is who is designing it and how it is being used behind the scene that deserves scrutiny and caution."* This suggests that human users explicitly take such considerations into account. However, despite the general difference, we could not see any effect of the provider dimension on the actual usage/intention to use a particular technology. Furthermore, we can confirm that the "no trust, no use" hypothesis holds in the context of AI. This might feel obvious and

Table 1: Survey items used to assess general trust in AI, as well as trust in the tech companies developing AI applications.

Construct
General Trust in Artificial Intelligence (Cronbach's $\alpha = .71$)
In general do you trust AI?
In general, are you sceptical about AI? (reverse scored)
Trust in Tech Companies (providers of AI, Cronbach's $\alpha = .81$)
In general, do you trust the big tech companies who develop AI?
Tech-companies (providers of AI) are trustworthy and keep up to ethical standards
Tech-companies (providers of AI) care about humanity rather than their own benefit
Tech-companies are taking much care in building safey and high-quality AI products

Table 2: Descriptive statistics of the dimensions general *trust in AI* and *trust in providers of AI*, split along the corresponding intention to use a particular product.

Concept & Artifact	**Value**	**N**	**trust in AI mdn (IQR)**	**trust in AI M (SD)**	**trust in provider mdn (IQR)**	**trust in provider M (S)**
Use	yes	50	3.5 (1)	3.48 (.82)	3 (.94)	2.86 (.82)
automated vehicle	no	31	2.5 (1)	2.53 (.75)	2 (1.25)	2.16 (.74)
Use	yes	43	3.5 (1)	3.38 (.84)	3 (.75)	2.88 (.80)
digital assistant	no	60	3 (1)	2.87 (.81)	2.5 (1.06)	2.45 (.77)

was previously confirmed for a wide range of technologies. Still, given the sharply criticised relevance of trust in AI, we believe it must remain in current debates. Finally, given the significance of trust in AI, we recommend to include "trust calibration", a dynamic phenomenon in the human-AI interaction, in further studies [6]. It is expected that trust changes with respect to time and performance of AI-based systems. These could present critical factors which we could not explore within the scope of this pilot study. Despite methodological limitations, we consider the results of this piloting study important and meaningful enough for us to plan more extensive experiments in the future.

Conclusion

Within the current discussions of trustworthy AI, the question of who represents the "trustee" in human-AI interactions, and if it matters, is not fully answered. We conducted a pilot study to investigate potential users' general *trust in AI* and their *trust in providers of AI*. We could confirm that trust matters in human-AI interaction and that users may trust AI, in general, more than technology providers developing AI products. It seems that users differentiate between the two aspects. By addressing the methodological limitations, we believe it is beneficial to create a future research agenda expanding current trust in AI frameworks with a new actor.

REFERENCES

[1] Joanna Bryson. 2018. AI & Global Governance: No One Should Trust AI. (2018). https://cpr.unu.edu/publications/articles/ai-global-governance-no-one-should-trust-ai.html

[2] European Commision. 2021. Proposal for a Regulation laying down harmonised rules on artificial intelligence. (2021). https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-//harmonised-rules-artificial-intelligence

[3] Felix Gille, Anna Jobin, and Marcello Ienca. 2020. What we talk about when we talk about trust: Theory of trust for AI in healthcare. *Intelligence-Based Medicine* 1-2, 4 (2020), 100001. DOI: http://dx.doi.org/10.1016/j.ibmed.2020.100001

[4] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human factors* 53, 5 (2011), 517–527. DOI: http://dx.doi.org/10.1177/0018720811417254

[5] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: integrating empirical evidence on factors that influence trust. *Human factors* 57, 3 (2015), 407–434. DOI: http://dx.doi.org/10.1177/0018720814547570

[6] John D. Lee and Katrina A. See. 2004. Trust in automation: designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80. DOI: http://dx.doi.org/10.1518/hfes.46.1.50{_}30392