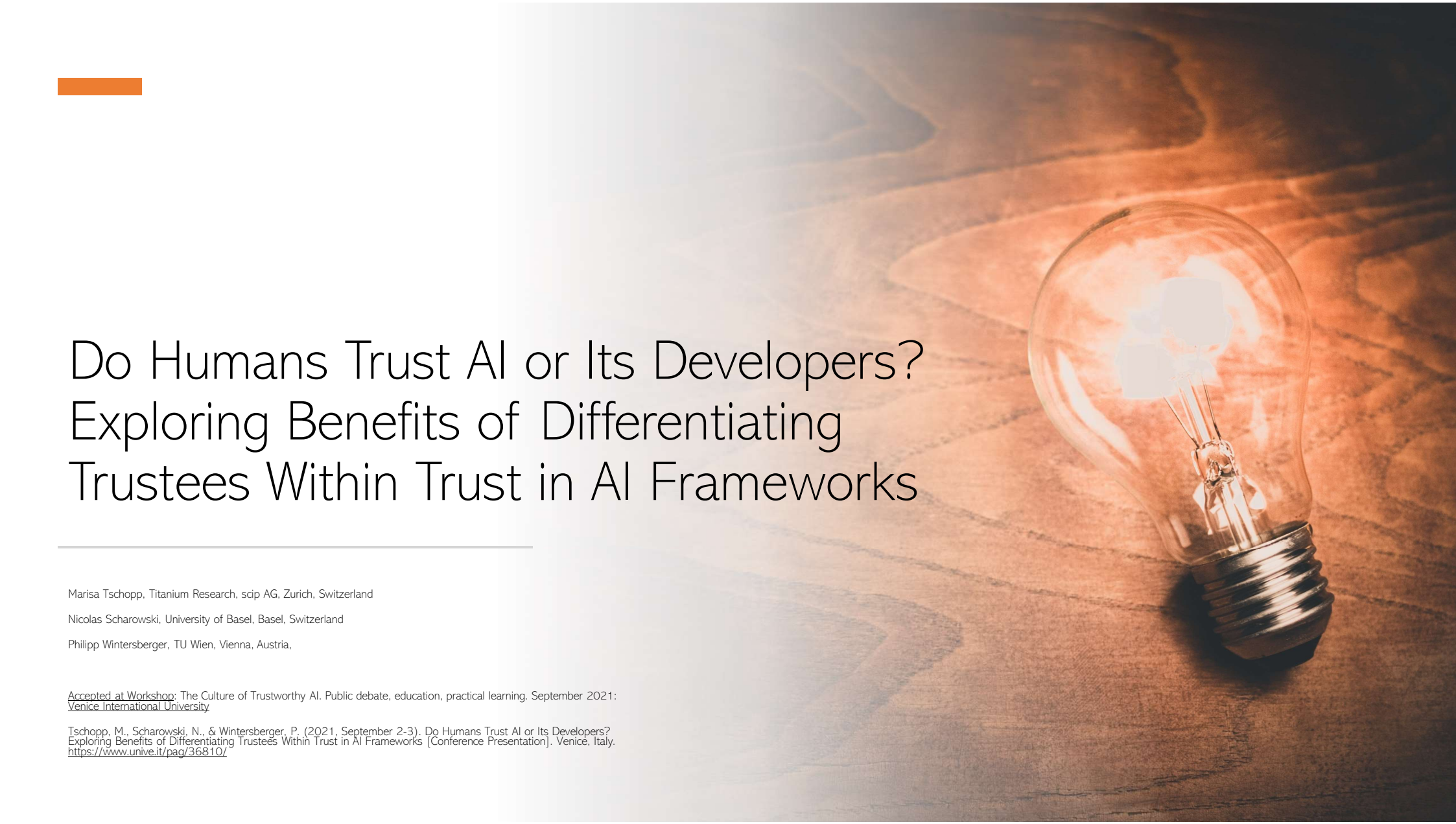
Do Humans Trust AI or Its Developers? Exploring Benefits of Differentiating Trustees Within Trust in AI Frameworks



Marisa Tschopp, Titanium Research, scip AG, Zurich, Switzerland

Nicolas Scharowski, University of Basel, Basel, Switzerland

Philipp Wintersberger, TU Wien, Vienna, Austria,

Accepted at Workshop: The Culture of Trustworthy AI. Public debate, education, practical learning. September 2021: Venice International University

Tschopp, M., Scharowski, N., & Wintersberger, P. (2021, September 2-3). Do Humans Trust AI or Its Developers? Exploring Benefits of Differentiating Trustees Within Trust in AI Frameworks [Conference Presentation]. Venice, Italy. https://www.unive.it/pag/36810/

Introduction

AI systems should be *"understandable to non-technical audiences and providing them with meaningful information, which is necessary to evaluate fairness and gain trust"*

(Regulatory proposal 2021, European Commission)

- The words "trust" and "trustworthiness" appear over 100 times in a regulatory proposal by the European Commission → Suggesting, that trust and beneficial use have a close relationship
- Do they? "No one should Trust AI!" (Joanna Bryson, 2018)

References

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

*European Commission. 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

Research questions

As AI systems in various contexts become part of human lives, and increasing awareness by the public, a broad understanding of how people make sense of AI as an actor is necessary [1] which cannot exclude the topic of trust. To better understand the relevance of trust, a better understanding of who and what humans place their trust "exactly" in is needed [2,3,4].

RQ 1: Is trust even relevant when it comes to user behavior? (No trust, no use?)

RQ 2: Are trust in AI and trust in tech companies of AI different dimensions?

References

[1] Andrea L. Guzman and Seth C. Lewis. 2020. Artificial intelligence and communication: A Human–Machine Communication research agenda. New Media & Society 22, 1 (2020), 70–86. https://doi.org/10.1177/1461444819858691

[2] Ella Glikson and Anita Williams Woolley. 2020. Human Trust in Artificial Intelligence: Review of Empirical Research. Academy of Management Annals 14, 2 (2020), 627–660. https://doi.org/10.5465/annals.2018.0057

[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. https://doi.org/10.1016/j.ibmed.2020.100001

[4] Kristin E Schaefer, Jessie YC Chen, James L Szalma, and Peter A Hancock. 2016. A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. Human factors 58, 3 (2016), 377–400.

Method 1 / 2: Participants

- Online survey with various questions around trust/use ratings and demographic variables (date of data collection, 2019).
- Participants:
  - N = 111 between 30 and 50 (62 female, 42 male, 2 diverse, 5 did not answer) . 76% have at last a Bachelor, Master Degree or higher, and participants are primarily from western countries (72% European, 15% American, 13% other).
  - Their self-reported level of expertise in AI (M = 4.69, SD = 2.24) was slightly below, and 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.

Method 2/2: Questionnaire

Scale items
General Trust in Artificial Intelligence (Cronbach's α = .71)
1 In general do you trust AI?
2 In general, are you sceptical about AI? (reverse coded)
Trust in Tech Companies (providers of AI, Cronbach's α = .81)
3 In general, do you trust the big tech companies who develop AI?
4 Tech-companies (providers of AI) are trustworthy and keep up to ethical standards
5 Tech-companies (providers of AI) care about humanity rather than their own benefit
6 Tech-companies are taking much care in building safey and high-quality AI products

- Two self developed items to measure trust in AI (5-point Likert scale)
- Four self developed items to measure trust in tech companies (5-point Likert scale)
- Acceptable Cronbach's Alpha to calculate scale means
- Intention to use was measured on a nominal scale (yes, no, depends) to stimulate a «behavioural response»:

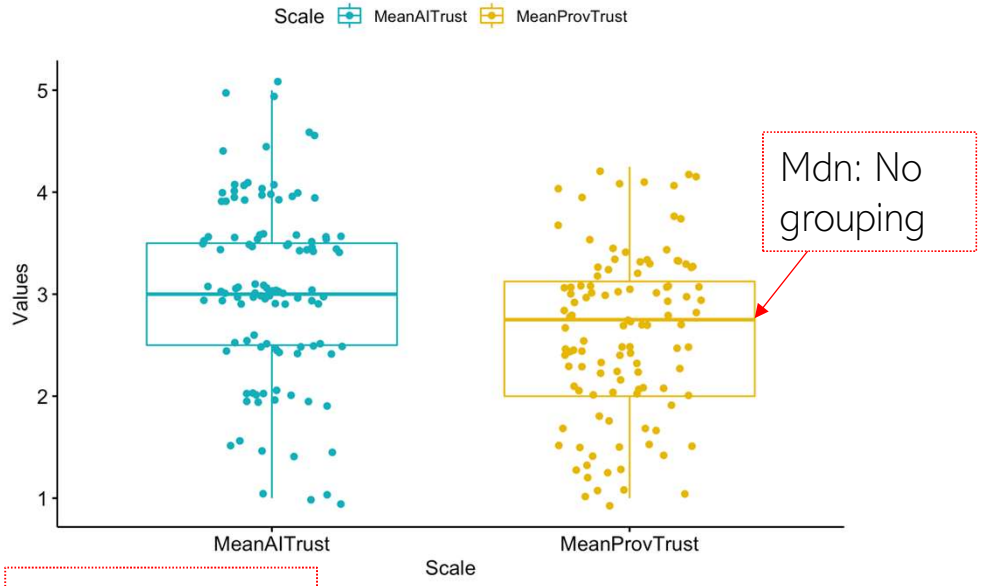
1. Would you use a self-driving car? (high risk scenario)
2. Are you using a conversational AI? (low risk scenario)

Results 1 / 4 RQ 1: No trust, no use?

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 automated vehicle	yes	50	3.5 (1)	3.48 (.82)	3 (.94)	2.86 (.82)
	no	31	2.5 (1)	2.53 (.75)	2 (1.25)	2.16 (.74)
Use digital assistant	yes	43	3.5 (1)	3.38 (.84)	3 (.75)	2.88 (.80)
	no	60	3 (1)	2.87 (.81)	2.5 (1.06)	2.45 (.77)

- Normality assumptions violated (Shapiro-Wilk's test's p-values < 0.05)
- 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 ($z = -4.703$, $p < .001$, $d = .52$)
- 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
- No difference in trust ratings between using a car or using a digital assistant

Results 2/ 4 RQ 2: Trust in AI vs. Trust in Tech companies?

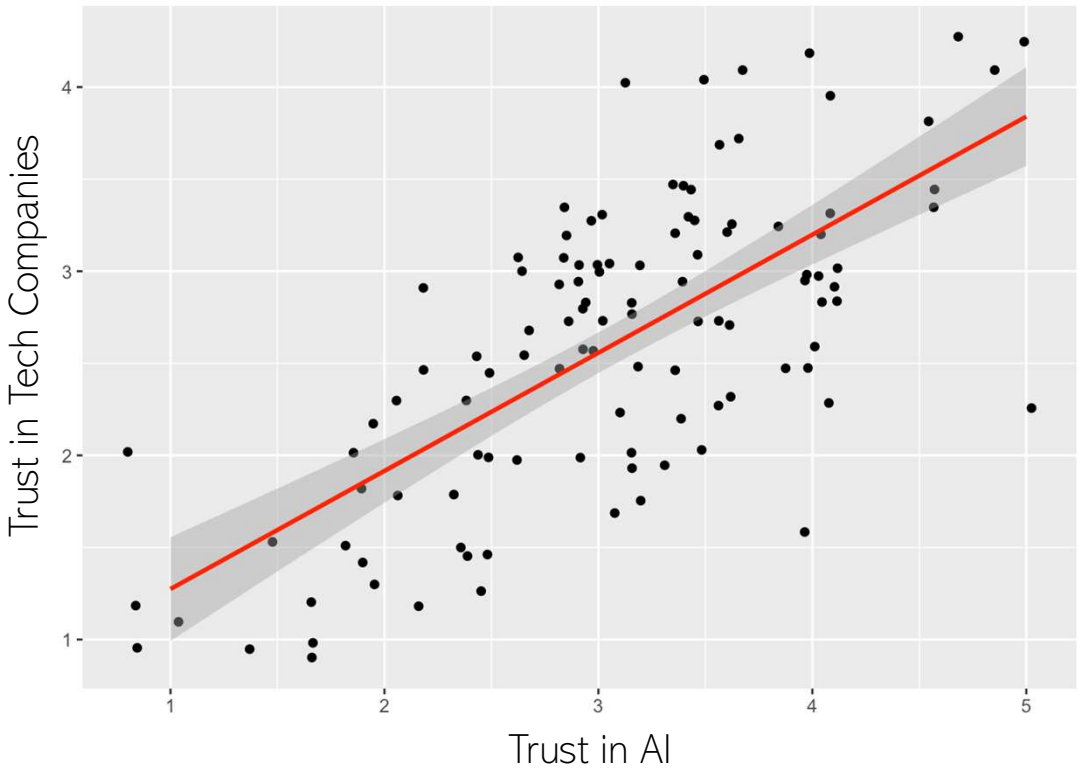


Mdn/M: grouped

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)

- Normality assumptions violated (Shapiro-Wilk's test's p-values < 0.05)
- The boxplots show lower ratings for the provider trust and higher ratings for trust in AI (more scattered/ higher SD)
- The median of the different scales is significantly different with an estimated difference of 0.62. (Wilcoxon z = -6.26, p = 3.899e-10, n = 111; effect size is Cohen's d = .6, medium to large)
- Similar patterns for comparing users – and not-users of AI artifacts

Results 3/3 RQ 2: Trust in AI vs. Trust in Tech companies?



- While AI trust and provider trust ratings are significantly different (as shown by the Wilcoxon test), there is a moderate correlation between the two dimensions.
- People with higher AI trust also show higher Provider trust ($r_K = .53$)
- Principle component analysis (varimax) indicates two different factors but cannot be interpreted due to low sample size.

Discussion

The pilot study suggests two main findings:

1. People are able to differentiate between trust in AI and trust in the provider. But is this influencing their behavior? So, does it even matter?

Statement of one participant: "*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.*"

2. The "no trust, no use" hypothesis also applies in the context of AI. People who are already using or would use a conversational AI or an automated vehicle have a **higher level of trust**. But why is there no difference between the risk scenarios?

Limitations and Future Work

Limitations

- Small, convenience sample
- Tech-Savvy, well educated, culturally biased
- No investigation of particulary providers
- Self-developed items
- Use presented as self-report and not differentiated between actual use and intention to use

→Results cannot be generalized

Future work

- Diverse, validated methods and larger sample size (re-evaluate existing trust measures)
- Real product experiences with corresponing providers (e.g. Amazon and Alexa)
- Can we create a trust experiment including all trust dimensions comparable to the «moral machine experiment»?
- «Tech-Convergence»: Is trust in Alexa different than trust in an autmated vehicle or any other automated/AI technology?

Conclusion

- Debates on «who is the trustee» are still a hot topic, but impact is unclear
- Normative caveat every «trust researcher» should address: Differentiate between trust in the context of responsibility/accountability, and trust calibration in direct human-machine interactions
- Trust believers vs. Trust skeptics: We hope that thinking about the similarities between the two will allow us to develop better methodologies and products in the future

Authors

Marisa Tschopp*
Titanium Research, scip AG
Marisa researches human-AI interaction, focusing on relationship patterns, trust, and anthropomorphism. She is Chief Research Officer and Ambassador for Women in AI (WAI) and co-chair of the IEEE Trust and Agency AIS Committee.

Nicolas Scharowski
Nicolas is a doctoral student at the Center for Cognitive Psychologyand Methodology at the University of Basel focusing on the role of trust and explainability in AI.

Dr. Philipp Wintersberger
Philipp researches and designs systems aiming to strengthen the cooperation between humans and machines with an emphasis on safety-critical systems such as automated vehicles at TU Wien

* Corresponding author: mats@scip.ch

References*

[**1**] Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. 2018. The moral machine experiment. Nature 563, 7729 (2018), 59–64.

[**2**] Joanna Bryson. 2018. AI & Global Governance: No One Should Trust AI. https://cpr.unu.edu/publications/articles/ai-global-governance-no-oneshould-trust-ai.html

[3] Mark Coeckelbergh. 2021. Three Responses to Anthropomorphism in Social Robotics: Towards a Critical, Relational, and Hermeneutic Approach. International Journal of Social Robotics 42, 1 (2021), 143. https://doi.org/10.1007/s12369-021-00770-0

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

[5] Anna-Katharina Frison, Philipp Wintersberger, Andreas Riener, Clemens Schartmüller, Linda Ng Boyle, Erika Miller, and Klemens Weigl. 2019. In UX we trust: Investigation of aesthetics and usability of driver-vehicle interfaces and their impact on the perception of automated driving. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 1–13.

[**6**] 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. https://doi.org/10.1016/j.ibmed.2020.100001

[**7**] Ella Glikson and Anita Williams Woolley. 2020. Human Trust in Artificial Intelligence: Review of Empirical Research. Academy of Management Annals 14, 2 (2020), 627–660. https://doi.org/10.5465/annals.2018.0057

[**8**] Andrea L. Guzman and Seth C. Lewis. 2020. Artificial intelligence and communication: A Human–Machine Communication research agenda. New Media & Society 22, 1 (2020), 70–86. https://doi.org/10.1177/1461444819858691

[9] 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. https://doi.org/10.1177/0018720811417254

[10] 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. https://doi.org/10.1177/0018720814547570

[11] Brittany E Holthausen, Philipp Wintersberger, Bruce N Walker, and Andreas Riener. 2020. Situational Trust Scale for Automated Driving (STS-AD): Development and Initial Validation. In 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. 40–47.

[12] Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. 2021. Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in ai. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 624–635.

[13] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. 2000. Foundations for an empirically determined scale of trust in automated systems. International journal of cognitive ergonomics 4, 1 (2000), 53–71.

[14] Alexandra D Kaplan, Theresa T Kessler, J Christopher Brill, and PA Hancock. 2021. Trust in Artificial Intelligence: Meta-Analytic Findings. Human Factors (2021), 00187208211013988.

[15] Gilles Laurent, Jean-Noel Kapferer, and Francoise Roussel. 1995. The underlying structure of brand awareness scores. Marketing Science 14, 3_supplement (1995), G170–G179.

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

[17] Mengyao Li, Brittany E Holthausen, Rachel E Stuck, and Bruce N Walker. 2019. No risk no trust: Investigating perceived risk in highly automated driving. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. 177–185.

[18] Mark Ryan. 2020. In AI We Trust: Ethics, Artificial Intelligence, and Reliability. Science and engineering ethics 26, 5 (2020), 2749–2767. https: //doi.org/10.1007/s11948-020-00228-y [19] Al F Salam, Lakshmi Iyer, Prashant Palvia, and Rahul Singh. 2005. Trust in e-commerce. Commun. ACM 48, 2 (2005), 72–77.

[20] Kristin E Schaefer, Jessie YC Chen, James L Szalma, and Peter A Hancock. 2016. A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. Human factors 58, 3 (2016), 377–400.

[21] Keng Siau and Weiyu Wang. 2018. Building trust in artificial intelligence, machine learning, and robotics. Cutter Business Technology Journal 31, 2 (2018), 47–53.

[22] S. Shyam Sundar. 2020. Rise of Machine Agency: A Framework for Studying the Psychology of Human–AI Interaction (HAII). Journal of Computer-Mediated Communication 25, 1 (2020), 74–88. https://doi.org/10.1093/jcmc/zmz026

* Bold references were selected for the extended abstract (limited number of references; We used the additional references for our short paper (unpublished). We thought these references might be useful for the readers here.