Multi-Level AI Trustworthiness Labels Scale Potential Users' Perceptions and Evaluations of AI Products

Christina. U. Pfeuffer
Department of Psychology – Human-Technology Interaction
Catholic University of Eichstätt-Ingolstadt
Eichstätt, Germany
christina.pfeuffer@ku.de, https://orcid.org/0000-0001-9394-8316

***Abstract*—(Potential) users of artificial intelligence (AI) products (e.g., smart fridges, voice assistants) are most often not fully informed about the data security and data privacy (DSDP) of these AI products as such information is not available in an easily accessible format. Here, I argue that it is essential to provide (potential) users with such information presented in an easily accessible and quick-to-process format as this could help instill appropriate levels of (dis-)trust and thus support (trustworthy) AI acceptance and adoption. Testing the effect of AI trustworthiness labels, I presented participants with hypothetical AI products that were versus were not paired with a graphical or text-based label indicating a low to high DSDP level. Trust, AI anxiety, as well as the monetary value attributed to AI products scaled with an AI trustworthiness label's DSDP level. Importantly, participants' ratings of unlabeled AI products corresponded to their perceptions of AI products labeled with an intermediate DSDP level. This apparent bias towards intermediate DSDP judgements in the absence of information on AI products underscores the relevance of explicitly communicating AI trustworthiness to (potential) users. Interestingly, differences in trust for AI products with a high as compared to intermediate DSDP level further increased with more positive attitudes towards AI technology.**

Keywords—artificial intelligence, data security and data privacy, label, regulation, trust, AI anxiety, technology attitude

I. INTRODUCTION

Artificial intelligence (AI) and corresponding AI products are on the verge of becoming ubiquitous to our everyday lives. This includes both conversational AI user interfaces (e.g., chatbots, AI voice assistants) as well as other AI (e.g., recommender systems, computer vision). AI holds the potential to benefit both individuals, organizations, and society at large by, for instance, optimizing products and services, enhancing productivity and efficiency, or lowering costs [1]. However, this potential can only be realized when human-AI interactions are appropriately shaped [2], [3].

Concerns regarding AI trustworthiness, in particular, data security and data privacy (DSDP) concerns [1], [4], jeopardize a further widespread acceptance and broader adoption of AI products (see e.g., [5], [6], [7], [8], for prominent theories of technology acceptance and adoption). AI products often gain access to users' (sensitive) data, raising concerns regarding data security and data privacy (see e.g., [9], [10], [11] for corresponding considerations regarding algorithm auditing). Recent theorizing emphasizes especially the role of trust (e.g., Intelligent Systems Technology Acceptance Model, ISTAM, [12], linking trust to transparency) as an essential precursors of technology acceptance and adoption. As such, establishing and maintaining the public's trust in AI, derived from a trustworthiness assessment of AI products [13], appears paramount to its further acceptance and adoption.

Users, however, are hardly able to evaluate the trustworthiness of AI accurately [1], [13], as corresponding information is commonly not easily accessible. To access information on the trustworthiness of AI products, for instance, to gain data security and data privacy information, (potential) users need to read long legal statements in an AI product's fine print. (Potential) users therefore currently (dis-)trust AI products mainly based on heuristics [14], [15], [16]. Moreover, strong, often unjustified AI endorsement [1], is coupled with low understanding of AI in the general public [1], [17]. Discrepancies between objective trustworthiness (e.g., adherence to criteria like those proposed by the European Commission and in the EU AI Act [4], [18]) and how trustworthy individuals perceive AI to be call for corresponding affirmative action. At present, (potential) users generally presume that trustworthy AI principles and standards (e.g., data security and data privacy, accuracy, risk and impact mitigation, and AI literacy support) are in place (e.g., for AI used in human resources, healthcare, or security contexts [1]). This notion is not necessarily justified for each AI product. It appears that the public's overall expectations regarding AI safeguards are currently not keeping up with objective protection measures and the speed of AI evolution.

Importantly, both misplaced distrust [13], [19] and misplaced trust (due to expectancy violations, [20], [21]) prevent the further acceptance and adoption of new (and trustworthy) AI technologies and obstruct corresponding benefits of AI usage. I propose that informative, multi-level labels (e.g., similar to the Nutri-Score indicating the nutritional value of food, e.g., [22]; for prior studies on technology/AI certification labels see e.g., [23], [24], [25], [26]) constitute the best-suited means of achieving accurate assessments of AI trustworthiness with very limited (potential) user effort across varying levels of AI literacy. At present, empirically-validated product labels like the Nutri-Score do not yet exit for AI products (existing labels typically do not indicate the degree and/or criteria of trustworthiness, but see [27] for a notable exception).

In this experiment, I communicated the data security and data privacy (DSDP; i.e., a specific AI trustworthiness criterion) level of hypothetical AI products using three-level labels (low vs. intermediate vs. high) of two label types (graphical vs. text-based label). My hypotheses were clear-cut: I expected trust and attributed monetary value to increase and AI anxiety to decrease for AI products with higher DSDP levels communicated by a corresponding DSDP/trust-

worthiness label. Furthermore, I expected to observe differences between the two label types.

II. Experimental Methods

An extended preprint (https://osf.io/preprints/psyarxiv/q25nr_v1), the preregistration of this study (https://osf.io/vbxqy) and all study materials (https://doi.org/10.17605/OSF.IO/HD3NA) are available online.

A priori, I planned Bayesian linear mixed model analyses and thus used a Bayesian approach to sample size estimation [28]. The Bayesian stopping criterion for ending data collection was a Bayes factor BF_{10} larger than 3 (or smaller than 1/3) for the effects of label type and DSDP level on the dependent variables trust and AI anxiety. 102 participants (35 male, 64 female, 3 diverse; age: M = 26.7 years, SD = 8.9; attitude towards technology [19]: M = 14.4, SD = 2.86, [4;20]; recruited via open online advertisements) took part after providing informed consent. The study was conducted in line with the regulations of the local ethics committee.

First, participants were informed about the features and functions of two hypothetical AI product types (smart fridge, voice assistant). Subsequently, they rated their trust (4-item trust questionnaire by Choung et al. [19]; 1 = strongly disagree to 5 = strongly agree) and (state) anxiety (4-item sociotechnical blindness subscale of the AI anxiety scale by Wang & Wang [29]; 1 = strongly disagree to 5 = strongly agree) regarding each AI product (represented by a corresponding name text and icon). First (baseline; 2 trials: 1 per AI product; order randomized), participants were presented only with the AI product (name text and icon) without further information. Then, second (after an introduction of the DSDP labels; DSDP criteria adapted from [30]; compare [4], [18]), participants were again presented with the AI products now paired with a DSDP label (label type: graphical vs. text-based; within) indicating a low, intermediate, or high level of trustworthiness (DSDP level; within; 12 trials: 2 AI products x 2 label types x 3 DSDP levels; order randomized) and indicated their trust and AI anxiety (see Fig. 1 for experimental design, procedure, and label appearance).

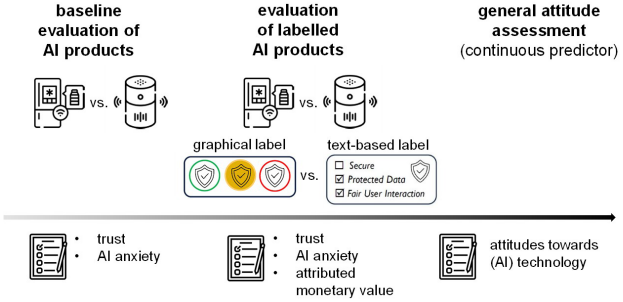

Fig. 1. Study design and time course.

In a final phase, the monetary value participants attributed to the respective labeled AI products was assessed by showing two different levels of the same label type per AI product and trial (level comparison: low-intermediate vs. intermediate-high vs. low-high; within; 12 trials: 2 AI products x 2 label types x 3 DSDP level comparisons; order randomized). Here, participants' task was to indicate how much more (percent acceptable price increase) they were willing to pay for the AI product with the respective higher DSDP level. Finally, participants rated their attitude towards (AI) technology (4-item attitude toward (AI) technologies scale by Choung et al. [19]; 1 = strongly disagree to 5 = strongly agree) and were debriefed.

The two DSDP label types compared were a graphical (horizontal traffic light: green, yellow, red) and a text-based label (vertical tick boxes with bullet point category descriptions: secure, data protected, fair user interaction; see Fig. 1 for label appearances). Label types differed in appearance, but conveyed the same information. That is, the degree to which the respective hypothetical AI product fulfilled the three main DSDP criteria: Secure, data protection, and fair user interaction (label content adapted from the Swiss Digital Initiative [30]). Participants were instructed about the meaning of the labels and DSDP criteria, that label types conveyed the same information in different visual ways, and the criteria for the respective DSDP levels (i.e., how many criteria were fulfilled to a satisfactory degree) before they began to evaluate the hypothetical AI products. They were explicitly told that both labels represented the same information. Importantly, however, the graphical label summarized the three criteria in one traffic light-like label, whereas the text-based label listed each criterion separately. The selected criteria themselves were used as exemplary material. The thresholds per criterion and for the three DSDP levels in this study were intentionally kept very simplistic and hypothetical (criterion: degree/percentage to which a corresponding ideal was fulfilled; DSDP level: number of criteria reaching this threshold). The focus of this study was not a dissemination of the best threshold criteria, but an assessment of (potential) users' perception of labels conveying trustworthiness criteria like DSDP information to determine whether pursuing further research on such labels and threshold criteria appears promising. Future studies will have to revisit the set of criteria as well as criterion and level thresholds for such labels if DSDP labels or broader AI trustworthiness labels (e.g., building on the criteria detailed in the EU AI Act [4], [18]) are to be implemented. The purpose of the present study was only to confirm whether such DSDP labels (irrespective of their exact level criteria) successfully instilled appropriate levels of (dis-)trust in participants. That is, this study assessed whether participants' perceived trust, AI anxiety, and the monetary value they attribute to DSDP-labeled AI products indeed scaled with the labels' DSDP level as well as how graphical and text-based labels generally differed in this regard.

III. Results

A Bayesian linear mixed model analysis approach (criterion: $BF_{10} > 3$ or $< 1/3$) was used. To account for differences between a person's ratings of the respective AI product type at baseline (i.e., without a DSDP label) and when presented with a DSDP label, I analyzed corresponding difference scores (i.e., dependent variable in respective DSDP label condition - dependent variable at baseline).

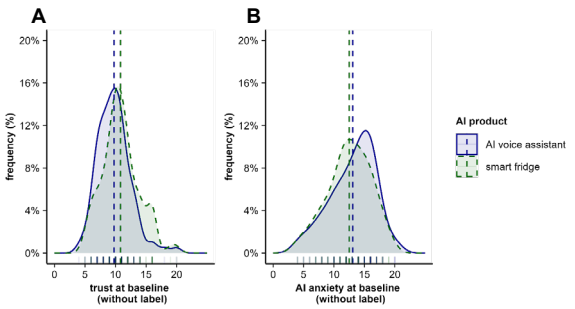

Fig. 2. Distribution of trust and AI anxiety ratings at baseline.

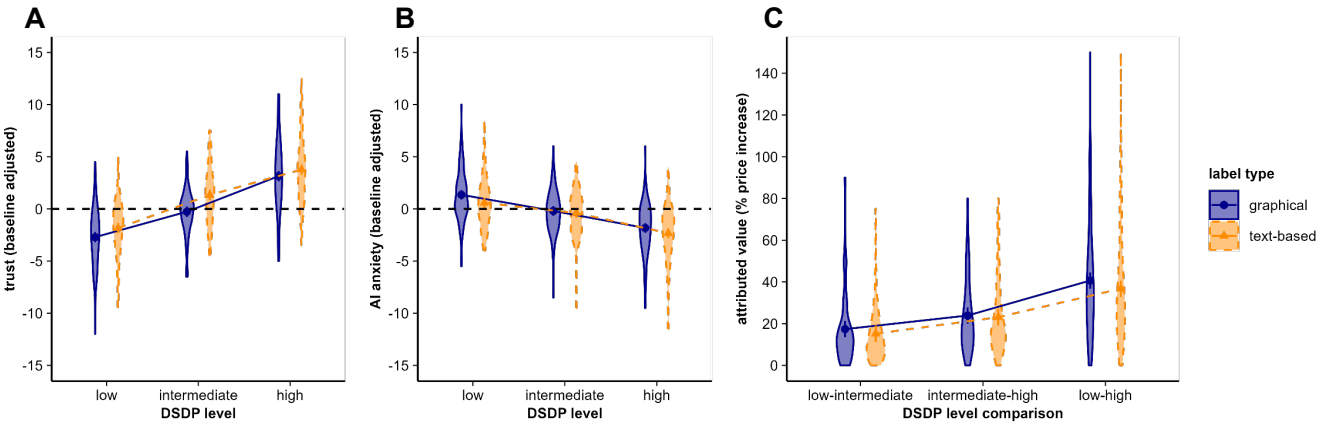

Fig. 3. Effects of data security and data privacy (DSDP) level/level comparison and label type on A) trust, B) AI anxiety, and C) attributed monetary value. Trust and AI anxiety scores are displayed relative to a participant's respective baseline rating of the corresponding AI product (0 = rating equivalent to baseline). Violins around the respective mean depict the corresponding rating distribution per condition.

Baseline. Trust at baseline was 9.8 (SD = 2.7; [0;20])/10.8 (SD = 3.0) for the AI voice assistant/smart fridge and AI anxiety at baseline was 13.1 (SD = 3.7; [0;20])/12.5 (SD = 3.5) for the AI voice assistant/smart fridge (see Fig. 2).

Trust. Trust ratings increased with increasing DSDP levels, BF_{10} = 1.36 x 10^{44} ±1.16%, showing a strong impact of the DSDP labels on participants' trust both between low and intermediate, BF_{10} = 1.23 x 10^{18} ±5.25%, and between intermediate and high DSDP levels, BF_{10} = 1.72 x 10^{19} ±1.46% (see Fig. 3A). Moreover, trust ratings were higher for text-based as compared to graphical labels, BF_{10} = 7.18 x 10^{7} ±0.88%. Label type and DSDP level interacted, BF_{10} = 4.03 ±1.61%.

Trust – Attitudes towards Technology. A post-hoc, exploratory analysis including attitudes towards technology showed evidence against an effect of attitudes towards technology, BF_{10} = 0.16 ±1.58%, as well as against an interaction between label type and attitudes towards technology, BF_{10} = 0.20 ±2.88%, and against a three-way interaction, BF_{10} = 0.13 ±2.14%. There was, however, evidence in favour of an interaction of DSDP level and attitude towards technology, BF_{10} = 6.73 ±3.57%. Specifically, there was evidence against differences in the slopes of individual attitudes towards technology scores between low and intermediate DSDP levels, BF_{10} = 0.27 ±2.06%, but evidence in favour of differences in the slopes of individual attitudes towards technology scores between intermediate and high DSDP levels, BF_{10} = 5.47 ±4.48%. Differences in trust scores between intermediate and high DSDP levels increased with more positive attitudes towards technology (see Fig. 4). Whereas, the evidence in favour of the effects of label type, BF_{10} = 4.6 x 10^{5} ±1.51%, and DSDP level, BF_{10} = 1.12 x 10^{147} ±1.88%, on trust ratings remained substantial, there was inconclusive evidence against an interaction of label type and DSDP level, BF_{10} = 0.70 ±2.07%, when additionally accounting for attitudes towards technology.

AI Anxiety. AI anxiety ratings decreased with increasing DSDP levels, BF_{10} = 8.4 x 10^{25} ±1.76%, showing a strong impact of the DSDP labels on participants' AI anxiety both between low and intermediate, BF_{10} = 7.5 x 10^{7} ±2.29%, and between intermediate and high DSDP levels, BF_{10} = 4.9 x 10^{10} ±1.45% (see Fig. 3B). AI anxiety ratings were lower for text-based as compared to graphical labels, BF_{10} = 9.5 ±1.75%. There was evidence against an interaction of label type and DSDP level, BF_{10} = 0.1 ±2.23%.

AI anxiety – Attitudes towards Technology. A post-hoc, exploratory analysis including attitudes towards technology showed evidence against an effect of attitudes towards technology, BF_{10} = 0.20 ±3.47%, as well as against an interaction between label type and attitudes towards technology, BF_{10} = 0.18 ±1.67%, against an interaction between DSDP level and attitudes towards technology, BF_{10} = 0.05 ±9.07%, and against the three-way interaction, BF_{10} = 0.04 ±1.67%. The evidence in favour of the effects of label type, BF_{10} = 9.5 ±2.93%, and DSDP level, BF_{10} = 8.3 x 10^{25} ±3.05%, on AI anxiety ratings remained substantial, when additionally accounting for attitudes towards technology. Similarly, there was still evidence against an interaction of label type and DSDP level, BF_{10} = 0.09 ±2.30%.

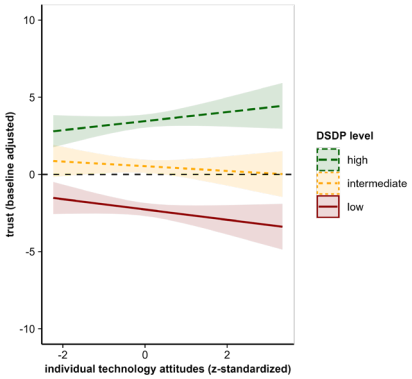

Fig. 4. Influence of individual attitudes towards (AI) technology on trust ratings by data security and data privacy (DSDP) level.

Attributed Value. Attributed monetary value (acceptable percent price increase for a higher DSDP level) increased across DSDP level comparisons, BF_{10} = 5.8 x 10^{29} ±1.26% (see Fig. 3C). The increase in attributed value was substantially larger between the intermediate and high DSDP level than between the low and intermediate DSDP level, BF_{10} = 1.2 x 10^{7} ±8.78%. Higher monetary value was attributed to AI products labelled with graphical as compared to text-based labels, BF_{10} = 8.1 ±1.15%. There was inconclusive evidence against an interaction of label type and DSDP level comparison, BF_{10} = 0.44 ±1.81%.

Attributed Value – Attitudes towards Technology. A post-hoc, exploratory analysis including attitudes towards technology showed evidence against an effect of attitudes

towards technology, BF_{10} = 0.35 ±3.31%, as well as against an interaction between label type and attitudes towards technology, BF_{10} = 0.19 ±9.27%, against an interaction between DSDP level comparison and attitudes towards technology, BF_{10} = 0.06 ±9.04%, and against the three-way interaction, BF_{10} = 0.05 ±14.96%. The evidence in favour of the effects of label type, BF_{10} = 9.9 ±3.09%, and DSDP level comparison, BF_{10} = 8.0 x 10^{111} ±2.66%, remained substantial, when additionally accounting for individual attitudes towards technology. The evidence against an interaction of label type and DSDP level comparison was conclusive when additionally accounting for individual attitudes towards technology, BF_{10} = 0.07 ±7.92%.

IV. Discussion

In this study, I either paired AI products (smart fridge, voice assistant) with an AI trustworthiness label indicating a low to high DSDP level or presented them without such a label (baseline). As expected, DSDP labels effectively communicated AI trustworthiness and correspondingly scaled participants' perceptions and evaluations of the AI products. That is, participants' trust increased and AI anxiety decreased from low to high DSDP levels. Moreover, participants attributed a higher monetary value to AI products with a higher DSDP level as indicated by the corresponding AI trustworthiness label.

Importantly, for the conducted Bayesian analyses, difference scores were used which subtracted a participant's baseline rating of the respective AI product from its rating in the DSDP level and label type conditions. Due to this baseline adjustment of trust and AI anxiety ratings, values of 0 equated a participant's perception of the respective AI product at baseline in the absence of DSDP information. The analyses revealed that baseline assessments of AI products corresponded to participants' assessment of AI products labeled with an intermediate DSDP level. This means that participants unjustifiedly attributed an intermediate DSDP level to AI products in the absence of DSDP information, suggesting a bias. The present findings thus underscore the importance of introducing corresponding DSDP labels for AI products to prevent both unjustified trust and unjustified distrust in (potential) users.

Furthermore, text-based DSDP labels indicating AI trustworthiness were associated with higher trust and lower AI anxiety than graphical labels. Conversely, however, participants attributed higher monetary value to AI products labeled with a graphical rather than text-based label. This suggests that text-based labels might be better suited to increase trust [12], [19], [31] in (potential) users and thereby the acceptance and adoption of AI. However, graphical labels might appear more appealing to companies selling AI products as products with graphical labels were attributed a higher monetary value. Furthermore, graphical labels can be processed even faster and more easily by (potential) users than text-based labels. Therefore, one might also prefer graphical labels to potentially increase AI companies' acceptance of regulatory AI trustworthiness labels and further optimize user effort.

Interestingly, participants' attitudes towards AI technology did not yield a general influence on their trust, AI anxiety, and the value attributed to the AI products. These findings for DSDP labeled AI products appear to be in stark contrast to prior findings regarding perceptions and evaluations of AI in the literature. Prior studies found evidence for positive correlations between attitudes towards technology and trust in AI-supported technologies [32], [33], [34] as well as for negative correlations between attitudes towards technology and AI anxiety [33]. Additionally, there was conclusive evidence against an impact of a person's attitude towards technology on the influence of DSDP labels on AI anxiety and attributed monetary value. Trust in the corresponding AI products, however, differed more strongly between the three DSDP levels for persons with more positive general attitudes towards technology and AI. Participants with more positive attitudes towards technology and AI appeared to increase more strongly in their trust for AI products labeled with a high DSDP level in comparison to AI products labeled with an intermediate DSDP level. That is, the DSDP information contained in the labels had a stronger positive impact on the trust of participants with more positive attitudes towards AI. Further systematic comparison studies will be required to fully address what underlies the observed discrepancies between findings regarding attitude towards technology and AI observed here as compared to reported in prior studies (see also e.g., [35] for a prior study showing that information on disruptive technologies, there videos on cryptocurrency, can both increase and decrease trust). Regarding trust, however, the present findings already indicate a potential explanation that merits further investigation. That is, the relationship between attitudes towards technology and trust appears to crucially depend on the information participants have received on an AI product's DSDP level. Put differently, persons with more positive attitudes towards technology and AI appear to be more susceptible to positive DSDP information on AI products.

It is important to note that the best-suited AI trustworthiness criteria and the best-suited level thresholds for AI trustworthiness labels were never the focus of this study. The simplistic, exemplary trustworthiness criteria (here 3 DSDP criteria) and simplistic threshold descriptions used in this study can be considered as placeholders for future, better-suited criteria derived from corresponding research. The aim of this psychological study was to determine whether multi-level AI trustworthiness labels (here focused on DSDP information as an example) are able to instill suitable degrees of (dis-)trust in (potential) users. The present findings clearly confirm that multi-level AI trustworthiness labels can fulfill this purpose. Separate future studies are, however, required to determine the exact, appropriate criteria for such an AI trustworthiness labels as well as their thresholds.

Future research will, for instance, need to incorporate additional trustworthiness criteria (e.g., [18]), select more informed thresholds for AI trustworthiness levels, and account for label effects at different AI literacy levels (e.g., [36]) as well as further investigate the impact of individual attitudes towards technology and other individual characteristics of (potential) users.

Acknowledgment

I thank Anja Ruff for her help with material preparation as well as setting up and advertising the online study.

References

[1] N. Gillespie, S. Lockey, C. Curtis, J. Pool, and Ali Akbari, "Trust in Artificial Intelligence: A global study," The University of Queensland; KPMG Australia, Brisbane, Australia, Feb. 2023. doi: 10.14264/00d3c94.

[2] S. Amershi et al., “Guidelines for Human-AI Interaction,” in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow Scotland Uk: ACM, May 2019, pp. 1–13. doi: 10.1145/3290605.3300233.

[3] G. Fragiadakis, C. Diou, G. Kousiouris, and M. Nikolaidou, “Evaluating Human-AI Collaboration: A Review and Methodological Framework,” July 09, 2024, arXiv: arXiv:2407.19098. doi: 10.48550/arXiv.2407.19098.

[4] European Parliament and Council of the European Union, Regulation (EU) 2022/206 of 21 April 2022 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0206

[5] F. D. Davis, “Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology,” MIS Quarterly, vol. 13, no. 3, p. 319, Sept. 1989, doi: 10.2307/249008.

[6] N. Marangunić and A. Granić, “Technology acceptance model: a literature review from 1986 to 2013,” Univ Access Inf Soc, vol. 14, no. 1, pp. 81–95, Mar. 2015, doi: 10.1007/s10209-014-0348-1.

[7] Venkatesh, Morris, Davis, and Davis, “User Acceptance of Information Technology: Toward a Unified View,” MIS Quarterly, vol. 27, no. 3, p. 425, 2003, doi: 10.2307/30036540.

[8] V. Venkatesh and F. D. Davis, “A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies,” Management Science, vol. 46, no. 2, pp. 186–204, Feb. 2000, doi: 10.1287/mnsc.46.2.186.11926.

[9] D. Hartmann, J. R. L. De Pereira, C. Streitbörger, and B. Berendt, “Addressing the regulatory gap: moving towards an EU AI audit ecosystem beyond the AI Act by including civil society,” AI Ethics, Nov. 2024, doi: 10.1007/s43681-024-00595-3.

[10] A. Koshiyama et al., “Towards algorithm auditing: managing legal, ethical and technological risks of AI, ML and associated algorithms,” R. Soc. Open Sci., vol. 11, no. 5, p. 230859, May 2024, doi: 10.1098/rsos.230859.

[11] J. Laine, M. Minkkinen, and M. Mäntymäki, “Ethics-based AI auditing: A systematic literature review on conceptualizations of ethical principles and knowledge contributions to stakeholders,” Information & Management, vol. 61, no. 5, p. 103969, July 2024, doi: 10.1016/j.im.2024.103969.

[12] E. S. Vorm and D. J. Y. Combs, “Integrating Transparency, Trust, and Acceptance: The Intelligent Systems Technology Acceptance Model (ISTAM),” International Journal of Human–Computer Interaction, vol. 38, no. 18–20, pp. 1828–1845, Dec. 2022, doi: 10.1080/10447318.2022.2070107.

[13] N. Schlicker, K. Baum, A. Uhde, S. Sterz, M. C. Hirsch, and M. Langer, “How do we assess the trustworthiness of AI? Introducing the trustworthiness assessment model (TrAM),” Computers in Human Behavior, vol. 170, p. 108671, Sept. 2025, doi: 10.1016/j.chb.2025.108671.

[14] Z. Buçinca, M. B. Malaya, and K. Z. Gajos, “To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making,” Proc. ACM Hum.-Comput. Interact., vol. 5, no. CSCW1, pp. 1–21, Apr. 2021, doi: 10.1145/3449287.

[15] Q. V. Liao and S. S. Sundar, “Designing for Responsible Trust in AI Systems: A Communication Perspective,” in 2022 ACM Conference on Fairness, Accountability, and Transparency, June 2022, pp. 1257–1268. doi: 10.1145/3531146.3533182.

[16] Z. Lu and M. Yin, “Human Reliance on Machine Learning Models When Performance Feedback is Limited: Heuristics and Risks,” in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama Japan: ACM, May 2021, pp. 1–16. doi: 10.1145/3411764.3445562.

[17] M. Kasinidou, “Promoting AI Literacy for the Public,” in Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 2, Toronto ON Canada: ACM, Mar. 2023, pp. 1237–1237. doi: 10.1145/3545947.3573292.

[18] European Commission. Directorate General for Communications Networks, Content and Technology. and High Level Expert Group on Artificial Intelligence., Ethics guidelines for trustworthy AI. LU: Publications Office, 2019. Accessed: Feb. 28, 2025. [Online]. Available: https://data.europa.eu/doi/10.2759/346720

[19] H. Choung, P. David, and A. Ross, “Trust in AI and Its Role in the Acceptance of AI Technologies,” International Journal of Human–Computer Interaction, vol. 39, no. 9, pp. 1727–1739, May 2023, doi: 10.1080/10447318.2022.2050543.

[20] M. (Mj) Rheu, Y. (Nancy) Dai, J. Meng, and W. Peng, “When a Chatbot Disappoints You: Expectancy Violation in Human-Chatbot Interaction in a Social Support Context,” Communication Research, vol. 51, no. 7, pp. 782–814, Oct. 2024, doi: 10.1177/00936502231221669.

[21] J. Hong, “Artificial intelligence (AI), don’t surprise me and stay in your lane: An experimental testing of perceiving humanlike performances of AI,” Human Behav and Emerg Tech, vol. 3, no. 5, pp. 1023–1032, Dec. 2021, doi: 10.1002/hbe2.292.

[22] K. Jürkenbeck, C. Mehlhose, and A. Zühlsdorf, “The influence of the Nutri-Score on the perceived healthiness of foods labelled with a nutrition claim of sugar,” PLoS ONE, vol. 17, no. 8, p. e0272220, Aug. 2022, doi: 10.1371/journal.pone.0272220.

[23] S. Lins, M. Greulich, J. Löbbers, A. Benlian, and A. Sunyaev, “Why so skeptical? Investigating the emergence and consequences of consumer skepticism toward web seals,” Information & Management, vol. 61, no. 2, p. 103920, Mar. 2024, doi: 10.1016/j.im.2024.103920.

[24] D. S. Guamán et al., “TRUESSEC Trustworthiness Label Recommendations,” in Challenges in Cybersecurity and Privacy - the European Research Landscape, 1st ed., New York: River Publishers, 2022, pp. 207–230. doi: 10.1201/9781003337492-10.

[25] N. Scharowski, M. Benk, S. J. Kühne, L. Wettstein, and F. Brühlmann, “Certification Labels for Trustworthy AI: Insights From an Empirical Mixed-Method Study,” in 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago IL USA: ACM, June 2023, pp. 248–260. doi: 10.1145/3593013.3593994.

[26] M. Wischnewski, N. Krämer, C. Janiesch, E. Müller, T. Schnitzler, and C. Newen, “In Seal We Trust? Investigating the Effect of Certifications on Perceived Trustworthiness of AI Systems,” HMC, vol. 8, pp. 141–162, 2024, doi: 10.30658/hmc.8.7.

[27] P. Emami-Naeini, H. Dixon, Y. Agarwal, and L. F. Cranor, “Exploring How Privacy and Security Factor into IoT Device Purchase Behavior,” in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow Scotland Uk: ACM, May 2019, pp. 1–12. doi: 10.1145/3290605.3300764.

[28] F. D. Schönbrodt, E.-J. Wagenmakers, M. Zehetleitner, and M. Perugini, “Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences.,” Psychological Methods, vol. 22, no. 2, pp. 322–339, June 2017, doi: 10.1037/met0000061.

[29] Y.-Y. Wang and Y.-S. Wang, “Development and validation of an artificial intelligence anxiety scale: an initial application in predicting motivated learning behavior,” Interactive Learning Environments, vol. 30, no. 4, pp. 619–634, Apr. 2022, doi: 10.1080/10494820.2019.1674887.

[30] Swiss Digital Initiative, Oct. 2022. Accessed: Feb. 12, 2023. [Online]. Available: https://digitaltrust-label.swiss/criteria/

[31] M. Dekkal, M. Arcand, S. Prom Tep, L. Rajaobelina, and L. Ricard, “Factors affecting user trust and intention in adopting chatbots: the moderating role of technology anxiety in insurtech,” J Financ Serv Mark, vol. 29, no. 3, pp. 699–728, Sept. 2024, doi: 10.1057/s41264-023-00230-y.

[32] A. Y. K. Chua, A. Pal, and S. Banerjee, “AI-enabled investment advice: Will users buy it?,” Computers in Human Behavior, vol. 138, p. 107481, Jan. 2023, doi: 10.1016/j.chb.2022.107481.

[33] C. Montag, J. Kraus, M. Baumann, and D. Rozgonjuk, “The propensity to trust in (automated) technology mediates the links between technology self-efficacy and fear and acceptance of artificial intelligence,” Computers in Human Behavior Reports, vol. 11, p. 100315, Aug. 2023, doi: 10.1016/j.chbr.2023.100315.

[34] L. Schadelbauer, S. Schlögl, and A. Groth, “Linking Personality and Trust in Intelligent Virtual Assistants,” MTI, vol. 7, no. 6, p. 54, May 2023, doi: 10.3390/mti7060054.

[35] H. Treiblmaier and E. Gorbunov, “On the Malleability of Consumer Attitudes toward Disruptive Technologies: A Pilot Study of Cryptocurrencies,” Information, vol. 13, no. 6, p. 295, June 2022, doi: 10.3390/info13060295.

[36] A. Carolus, M. J. Koch, S. Straka, M. E. Latoschik, and C. Wienrich, “MAILS - Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies,” Computers in Human Behavior: Artificial Humans, vol. 1, no. 2, p. 100014, Aug. 2023, doi: 10.1016/j.chbah.2023.100014.