

# Trust and Reliance in XAI - Distinguishing Between Attitudinal and Behavioral Measures

NICOLAS SCHAROWSKI\*, Center for General Psychology and Methodology, University of Basel, Switzerland

SEBASTIAN A. C. PERRIG, Center for General Psychology and Methodology, University of Basel, Switzerland

NICK VON FELTEN, Center for General Psychology and Methodology, University of Basel, Switzerland

FLORIAN BRÜHLMANN, Center for General Psychology and Methodology, University of Basel, Switzerland

Trust is often cited as an essential criterion for the effective use and real-world deployment of AI. Researchers argue that AI should be more transparent to increase trust, making transparency one of the main goals of XAI. Nevertheless, empirical research on this topic is inconclusive regarding the effect of transparency on trust. An explanation for this ambiguity could be that trust is operationalized differently within XAI. In this position paper, we advocate for a clear distinction between behavioral (objective) measures of reliance and attitudinal (subjective) measures of trust. However, researchers sometimes appear to use behavioral measures when intending to capture *trust*, although attitudinal measures would be more appropriate. Based on past research, we emphasize that there are sound theoretical reasons to keep trust and reliance separate. Properly distinguishing these two concepts provides a more comprehensive understanding of how transparency affects trust and reliance, benefiting future XAI research.

CCS Concepts: • **Human-centered computing** → **HCI theory, concepts and models**.

Additional Key Words and Phrases: AI, XAI, Trust, Reliance, Behavior, Attitude, Measures, Measurement, Operationalization

## ACM Reference Format:

Nicolas Scharowski, Sebastian A. C. Perrig, Nick von Felten, and Florian Brühlmann. 2022. Trust and Reliance in XAI - Distinguishing Between Attitudinal and Behavioral Measures. In *CHI 2022: Workshop on Trust and Reliance in AI-Human Teams, April 30 – May 6, 2022, New Orleans, LA*. ACM, New York, NY, USA, 6 pages. <https://doi.org/XX.XXXX/XXXXXXXX.XXXXXXX>

## 1 INTRODUCTION

Recent advances in artificial intelligence (AI) and its applicability in various everyday applications (e.g., video surveillance, email spam filtering, and product recommendations) have prompted a demand for more transparent AI to mitigate the potentially negative consequences of opaque AI [1, 12]. This call for transparent AI has led to the multidisciplinary research field of explainable artificial intelligence (XAI), which explores methods and models that make the behaviors, predictions, and decisions of AI transparent and understandable to humans [25, 26]. There are different rationales for transparent AI, one of which is the assumption that transparency contributes toward building a relationship of trust between humans and AI [39] and more reliance on AI [15]. Past research has shown that people tend to rely on automation they trust and reject automation they distrust [24]. This makes trust particularly relevant in the misuse (overreliance [35]) and disuse (neglect or underutilization [35]) of automation and AI [15, 39, 41], rendering trust a key property of the interaction between users and AI [19].

---

\*Corresponding Author.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2022 Association for Computing Machinery.

Manuscript submitted to ACM

However, there is mixed empirical evidence about whether transparency is indeed increasing trust [7, 8, 11, 34, 36, 42]. We argue that one possible reason for this ambiguity could be that within empirical XAI research, trust is measured in different ways, and often there does not appear to be a clear distinction between trust and related concepts such as reliance in the XAI community. To provide two examples from empirical work: Lai and Tan [23] conducted an experiment, asking end-users to decide whether a hotel review was genuine or deceptive. They illustrated that heatmaps of relevant instances and example-based explanations improved human performance and increased the trust humans place in the predictions of AI. Lai and Tan defined trust as the percentage of instances for which humans relied on the machine prediction. In contrast, Cheng et al. [7] conducted an experiment where participants used different UI interfaces to comprehend an algorithm’s decision for university admissions. They showed that revealing the inner workings of an algorithm can improve users’ comprehension and found that users’ trust, assessed by a survey scale, was not affected by the explanation interface. These two examples illustrate how trust can be thought of and measured differently in XAI research. Both studies intended to measure *trust*, but their respective operationalization of it varied substantially. Lai and Tan [23] opted for a *behavioral measurement* of trust (percentage of reliance instances) while Cheng et al. [7] chose an *attitudinal measurement* of trust (scale-based measurement).

Researchers’ choice of measures has implications for the variable of interest, and we will demonstrate that there are compelling theoretical reasons to distinguish between trust and reliance. We argue that a lack of distinction between behavioral and attitudinal measures leads to misunderstandings regarding what is actually being investigated in a study - trust or reliance. This misunderstanding could be a reason for the inconclusive findings in current XAI literature regarding the effect of transparency. For this reason, a rigorous distinction between trust and related concepts like reliance seem warranted. Although the two mentioned studies [7, 23] are merely illustrative, they indicate conceptual inconsistencies in current XAI research that needs to be addressed and provide a research opportunity to better define and formalize key concepts of XAI such as trust and reliance. We encourage a comprehensive literature review of past empirical work to better understand the extent to which such conceptual inconsistencies exist in current XAI literature. The purpose of this position paper is to emphasize the importance of adequately distinguishing between trust and reliance in measurement and provide theoretical reasons to make this distinction. We will further demonstrate that this measurement indifference has implications for the conclusions derived from XAI research. If research properly considers and distinguishes different ways of measuring trust and related concepts, this can bring great potential for a more comprehensive understanding of XAI.

## 2 THE DISTINCTION BETWEEN TRUST AS AN ATTITUDE AND RELIANCE AS A BEHAVIOR

Lee and See [24] defined trust in automation as "the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability." [24, p. 6] This definition encapsulates the notion of uncertainty, vulnerability, and risk, a necessary prerequisite for human trust in AI [19] and is consistent with the notion of vulnerability emphasized in the most widely used and accepted definition of trust [37] by Mayer et al. [29].

Defining trust as an *attitude* suggests that it should be regarded as a psychological construct. In this context, the term construct refers to unobservable features, such as psychological traits or abilities of people [17]. Psychological constructs are typically measured by survey scales, sometimes referred to as questionnaires or rating scales. The data collected using survey scales is inherently subjective, given that it reflects participants’ own perspectives. At their core, survey scales are about using a series of items to measure a construct of interest to a researcher [9]. Related to XAI research, a researcher could be interested in a user’s *subjective* trust in an automated system (i.e., the construct "trust in automation") and have users respond to the 12 items of the survey scale by Jian et al. [20]. Combining the responses

to the items of the scale (e.g., by forming a total score or calculating the mean value) provides the researcher with a number reflecting users' *subjective level of trust* toward the system. This number can then be used for statistical analysis during the research process and compared across research if the specific scale is used more than once.

Lee and See [24] thus defined trust as an *attitude* and distinguished it from reliance as a *behavior*. In contrast to trust as a psychological construct, *behavior* (e.g., reliance) is directly observable and therefore open to more objective approaches of measurement. Based on Lee and See we define *reliance* on a system as *a user's behavior that follows from the advice of the system*. In their work on metrics for XAI, Hoffman et al. [16] make a similar distinction when they differentiate between trusting a machine's output and following its advice. Given this definition, one possibility to measure reliance is "weight of advice" (WOA), stemming from the literature on advice taking [14]. WOA measures the degree to which people change their behavior and move their initial estimate (e.g., a price estimate) towards an (AI) advice to form a final estimate. While WOA has been used in the past by researchers in XAI as a more objective trust measurement [27, 33, 36], it has never been explicitly referred to as reliance and thus clearly differentiated from trust to the best of our knowledge.

Given these arguments, a proposed distinction between trust and reliance relates to subjective and objective measurement approaches. This differentiation between subjective and objective measures of trust in XAI was addressed by Mohseni et al. [32]. They pointed out that subjective trust measures in XAI include interviews, survey scales, and self-reports via questionnaires. More recently, Bućinca et al. [5] pointed out that subjective measures of trust have been a focal point in evaluating AI. For objective measures of trust, Mohseni et al. [32] proposed user's system competence, understanding, and users' compliance (i.e., reliance) with a system. Given these clarifications, it is evident that Lai and Tan [23] and Cheng et al. [7] made different choices of method (objective behavioral measurement vs. subjective scale-based measurement) while intending to capture the same variable of interest (i.e., trust).

However, the choice of measurement has implications on the variable of interest. Assuming that trust as an attitude is a psychological construct, behavioral measures are less adequate to capture its subjective nature, and scale-based approaches (e.g., Jian et al. [20]) are more appropriate. Conversely, attitudinal measures may be inadequate to reflect the objective constituents of reliance as a behavior. This indicates that XAI research could be at risk of confusion between the variable of interest and their respective attitudinal and behavioral measures. This lack of clarity regarding what is measured has implications not only for the data gathered with a particular method but also for the conclusions derived from said data, and thus the validity of a study as a whole [28]. We argue that this ambiguity could hinder the comparison of empirical work in XAI since researchers may have the same variable of interest in mind (e.g., trust) but miss each other's work or implications because of different measurement choices. Rigorously distinguishing between trust and reliance also offers a more comprehensive understanding of the effects of transparency and is, therefore, an opportunity for the XAI community.

### 3 DISCUSSION

We argue that the above-mentioned conceptual considerations have been insufficiently considered in past XAI research and emphasize that researchers need to differentiate between trust and reliance, further accentuating the importance of this distinction proposed by Lee and See [24]. The conceptual distinction between trust and reliance carries significant implications for XAI evaluation and uncovers two potential challenges. First, if researchers only assess trust as an attitude via questionnaires, they could falsely assume that people will (or will not) rely on an AI system. Second, if only behavior (i.e., reliance) is measured, researchers might incorrectly deduce that people necessarily trust (or not trust) the system in question. Consequently, researchers and practitioners have to decide which variables of interest to

consider and investigate when evaluating AI. We think it is useful to conceptualize trust as a psychological construct that is associated with reliance but does not fully determine it [24]. While trust is considered the main factor for human reliance on automation, other factors (e.g., workload, self-confidence, and task-associated risks) are likewise important [18, 24] and whether trust translates into reliance is more nuanced than often assumed and depends on an interaction between the operator, the automation, and the situation [22, 24].

We propose that it is sensible to consider trust and reliance to be conceptually distinct, because attitudes and behavior do not share a deterministic but a probabilistic relationship [2]. While researchers have shown that attitudinal judgments have an impact on people’s reliance on automated systems [8, 30], trust does not necessarily lead to more reliance. Even if an AI system is trusted, reliance must not follow [21, 22], and people may claim to trust an AI system, yet behave in a way that suggests they do not [31]. The empirical findings of Dzindolet et al. [10] support this argument: although some automated decision aids were rated as more trustworthy than others, all were equally likely to be relied upon. Multiple models of the relationship between trust and reliance have been proposed [6, 15, 24, 38, 40] and it has been hypothesized that trust mediates between human reliance and reliability of automation [24, 35]. This relationship has recently been empirically supported [18]. Research on XAI should take such models as a theoretical basis for their empirical work and refine them whenever necessary in the context of AI to advance the field with a rigorous science underlying it.

While we have provided a theoretical justification for a distinction between trust and reliance, others have provided more philosophical arguments, suggesting that the use of the term trust is not appropriate and sensible in the context of technology [4, 13, 19]. Because trust requires the acceptance of vulnerability, but objects have no understanding about this vulnerability, they cannot intentionally betray our trust, we can only rely on them. Therefore, so the argument, it might be unreasonable to talk about trust when referring to interactions between humans and AI, and instead, it is a matter of reliance. Jacovi et al. [19] argued that using the term trust is appropriate when attributing AI with intent and that in the case of non-anthropomorphized AI, the term reliance should be considered. Given this philosophical perspective, there appears to be potential for future research distinguishing between trust and reliance, both with anthropomorphized and non-anthropomorphized AI. Such research could then provide empirical evidence, adding to the philosophical discussion raised by Baier [4] and others [13, 19].

Arrieta et al. [3] emphasized that only agreed-upon metrics and their respective measurements allow for meaningful comparisons in XAI research and that without such consistency, empirical claims are ambiguous and do not provide a solid foundation for future XAI research. Therefore, it seems particularly important that researchers carefully differentiate *attitude* from *behavior* and choose appropriate measures for human-centred AI accordingly. Future research could conduct a literature review of how measures are used in the current XAI literature to understand better the extent to which conceptual inconsistencies exist between subjective attitudinal and objective behavioral measures. Such work could further develop a framework to guide the selection of measures of trust and reliance, depending on different research purposes. A way forward could be measuring both, trust *and* reliance and recognizing theoretical models on their relationship that have been proposed by past research [6, 15, 24, 38, 40]. Given that AI is increasingly applied to make critical decisions with far-reaching consequences, selecting appropriate measurements for clearly defined concepts is of paramount importance in XAI.

## REFERENCES

- [1] ACM FAccT Conference. 2022. *ACM Conference on Fairness, Accountability, and Transparency 2022 (ACM FAccT 2022) Call For Papers*. Retrieved 2022-01-10 from <https://facctconference.org/2022/cfp.html>

- [2] Icek Ajzen and Martin Fishbein. 1980. *Understanding attitudes and predicting social behavior*. Prentice-Hall, Englewood Cliffs, NJ, USA.
- [3] Alejandro B. Arrieta, Natalia Diaz-Rodríguez, Javier D. Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58 (June 2020), 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- [4] Annette Baier. 1986. Trust and antitrust. *ethics* 96, 2 (1986), 231–260. <https://doi.org/10.1086/292745>
- [5] Zana Bućinca, Phoebe Lin, Krzysztof Z. Gajos, and Elena L. Glassman. 2020. Proxy Tasks and Subjective Measures Can Be Misleading in Evaluating Explainable AI Systems. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari, Italy) (IUI '20). ACM, New York, NY, USA, 454–464. <https://doi.org/10.1145/3377325.3377498>
- [6] Eric T Chancey, James P Bliss, Yusuke Yamani, and Holly AH Handley. 2017. Trust and the compliance–reliance paradigm: The effects of risk, error bias, and reliability on trust and dependence. *Human factors* 59, 3 (2017), 333–345. <https://doi.org/10.1177/0018720816682648>
- [7] Hao-Fei Cheng, Ruotong Wang, Zheng Zhang, Fiona O’Connell, Terrance Gray, F. Maxwell Harper, and Haiyi Zhu. 2019. Explaining Decision-Making Algorithms through UI: Strategies to Help Non-Expert Stakeholders. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, UK, GBR) (CHI '19). ACM, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300789>
- [8] Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. 2008. The Effects of Transparency on Trust in and Acceptance of a Content-Based Art Recommender. *User Modeling and User-Adapted Interaction* 18, 5 (Nov. 2008), 455–496. <https://doi.org/10.1007/s11257-008-9051-3>
- [9] Robert F. DeVellis. 2017. *Scale development: Theory and applications (Fourth Edition)*. Vol. 26. SAGE publications, Inc., 2455 Teller Road, Thousand Oaks, California 91320.
- [10] Mary T. Dzindolet, Scott A. Peterson, Regina A. Pomranky, Linda G. Pierce, and Hall P. Beck. 2003. The role of trust in automation reliance. *International Journal of Human-Computer Studies* 58, 6 (June 2003), 697–718. [https://doi.org/10.1016/S1071-5819\(03\)00038-7](https://doi.org/10.1016/S1071-5819(03)00038-7)
- [11] Upol Ehsan, Pradyumna Tambwekar, Larry Chan, Brent Harrison, and Mark O. Riedl. 2019. Automated Rationale Generation: A Technique for Explainable AI and Its Effects on Human Perceptions. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) (IUI '19). ACM, New York, NY, USA, 263–274. <https://doi.org/10.1145/3301275.3302316>
- [12] Upol Ehsan, Philipp Wintersberger, Q. Vera Liao, Martina Mara, Marc Streit, Sandra Wachter, Andreas Riener, and Mark O. Riedl. 2021. *Operationalizing Human-Centered Perspectives in Explainable AI*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3411763.3441342>
- [13] Martin Hartmann. 2020. *Vertrauen - Die unsichtbare Macht*. Fischer Verlag, Berlin, Germany.
- [14] Nigel Harvey and Ilan Fischer. 1997. Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational behavior and human decision processes* 70, 2 (May 1997), 117–133. <https://doi.org/10.1006/obhd.1997.2697>
- [15] Kevin A. Hoff and Masooda Bashir. 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors* 57, 3 (May 2015), 407–434. <https://doi.org/10.1177/0018720814547570>
- [16] Robert R. Hoffman, Shane T. Mueller, Gary Klein, and Jordan Litman. 2018. Metrics for Explainable AI: Challenges and Prospects. *CoRR abs/1812.04608* (2018). arXiv:1812.04608 <http://arxiv.org/abs/1812.04608>
- [17] Kenneth D Hopkins. 1998. *Educational and psychological measurement and evaluation*. Pearson, London, UK.
- [18] Aya Hussein, Sondoss Elsayah, and Hussein A Abbass. 2020. Trust mediating reliability–reliance relationship in supervisory control of human–swarm interactions. *Human Factors* 62, 8 (2020), 1237–1248. <https://doi.org/10.1177/0018720819879273>
- [19] 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* (Virtual Event, Canada) (FAccT '21). ACM, New York, NY, USA, 624–635. <https://doi.org/10.1145/3442188.3445923>
- [20] 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 (March 2000), 53–71. [https://doi.org/10.1207/S15327566IJCE0401\\_04](https://doi.org/10.1207/S15327566IJCE0401_04)
- [21] Alex Kirlik. 1993. Modeling strategic behavior in human-automation interaction: Why an "aid" can (and should) go unused. *Human Factors* 35, 2 (June 1993), 221–242. <https://doi.org/10.1177/001872089303500203>
- [22] Moritz Körber. 2018. Theoretical considerations and development of a questionnaire to measure trust in automation. In *Proceedings of the 20th Congress of the International Ergonomics Association* (Florence, Italy) (IEA '18). Springer International Publishing, Cham, CH, 13–30. [https://doi.org/10.1007/978-3-319-96074-6\\_2](https://doi.org/10.1007/978-3-319-96074-6_2)
- [23] Vivian Lai and Chenhao Tan. 2019. On Human Predictions with Explanations and Predictions of Machine Learning Models: A Case Study on Deception Detection. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) (FAT\* '19). ACM, New York, NY, USA, 29–38. <https://doi.org/10.1145/3287560.3287590>
- [24] John D. Lee and Katrina A. See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (March 2004), 50–80. [https://doi.org/10.1518/hfes.46.1.50\\_30392](https://doi.org/10.1518/hfes.46.1.50_30392)
- [25] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). ACM, New York, NY, USA, 1–15. <https://doi.org/10.1145/3313831.3376590>
- [26] Zachary C. Lipton. 2018. The Mythos of Model Interpretability. *Commun. ACM* 61, 10 (Sept. 2018), 36–43. <https://doi.org/10.1145/3233231>
- [27] Jennifer M. Logg, Julia A. Minson, and Don A. Moore. 2019. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes* 151 (March 2019), 90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>

- [28] Scott B MacKenzie. 2003. The Dangers of Poor Construct Conceptualization. *Journal of the Academy of Marketing Science* 31, 3 (2003), 323–326. <https://doi.org/10.1177/0092070303254130>
- [29] Roger C. Mayer, James H. Davis, and F. David Schoorman. 1995. An integrative model of organizational trust. *Academy of Management Review* 20, 3 (July 1995), 709–734. <https://doi.org/10.2307/258792>
- [30] Stephanie M. Merritt. 2011. Affective processes in human–automation interactions. *Human Factors* 53, 4 (July 2011), 356–370. <https://doi.org/10.1177/0018720811411912>
- [31] David Miller, Mishel Johns, Brian Mok, Nikhil Gowda, David Sirkin, Key Lee, and Wendy Ju. 2016. Behavioral measurement of trust in automation: the trust fall. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 60, 1 (Sept. 2016), 1849–1853. <https://doi.org/10.1177/1541931213601422>
- [32] Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. 2021. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *ACM Trans. Interact. Intell. Syst.* 11, 3–4, Article 24 (aug 2021), 45 pages. <https://doi.org/10.1145/3387166>
- [33] Henrik Mucha, Sebastian Robert, Ruediger Breitschwerdt, and Michael Fellmann. 2021. *Interfaces for Explanations in Human-AI Interaction: Proposing a Design Evaluation Approach*. ACM, New York, NY, USA. <https://doi.org/10.1145/3411763.3451759>
- [34] Florian Nothdurft, Tobias Heinroth, and Wolfgang Minker. 2013. The Impact of Explanation Dialogues on Human-Computer Trust. In *Proceedings, Part III, of the 15th International Conference on Human-Computer Interaction. Users and Contexts of Use - Volume 8006*. Springer-Verlag, Berlin, Heidelberg, 59–67. [https://doi.org/10.1007/978-3-642-39265-8\\_7](https://doi.org/10.1007/978-3-642-39265-8_7)
- [35] Raja Parasuraman and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors* 39, 2 (June 1997), 230–253. <https://doi.org/10.1518/001872097778543886>
- [36] Forough Poursabzi-Sangdeh, Daniel G. Goldstein, Jake M. Hofman, Jennifer Wortman Vaughan, and Hanna Wallach. 2021. Manipulating and Measuring Model Interpretability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). ACM, New York, NY, USA, Article 237, 52 pages. <https://doi.org/10.1145/3411764.3445315>
- [37] Denise M. Rousseau, Sim B. Sitkin, Ronald S. Burt, and Colin Camerer. 1998. Not So Different After All: A Cross-Discipline View Of Trust. *Academy of Management Review* 23, 3 (July 1998), 393–404. <https://doi.org/10.5465/AMR.1998.926617>
- [38] 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. <https://doi.org/10.1177/0018720816634228>
- [39] Constantine Stephanidis, Gavriel Salvendy, Margherita Antona, Jessie Y. C. Chen, Jianming Dong, Vincent G. Duffy, Xiaowen Fang, Cali Fidopiastis, Gino Fragoni, Limin Paul Fu, Yinni Guo, Don Harris, Andri Ioannou, Kyeong-ah Jeong, Shin'ichi Konomi, Heidi Krömker, Masaaki Kurosu, James R. Lewis, Aaron Marcus, Gabriele Meiselwitz, Abbas Moallem, Hirohiko Mori, Fiona Fui-Hoon Nah, Stavroula Ntoa, Pei-Luen Patrick Rau, Dylan Schmorow, Keng Siau, Norbert Streitz, Wentao Wang, Sakae Yamamoto, Panayiotis Zaphiris, and Jia Zhou. 2019. Seven HCI Grand Challenges. *International Journal of Human-Computer Interaction* 35, 14 (July 2019), 1229–1269. <https://doi.org/10.1080/10447318.2019.1619259>
- [40] Kees Van Dongen and Peter-Paul Van Maanen. 2013. A framework for explaining reliance on decision aids. *International Journal of Human-Computer Studies* 71, 4 (2013), 410–424. <https://doi.org/10.1016/j.ijhcs.2012.10.018>
- [41] Kun Yu, Shlomo Berkovsky, Ronnie Taib, Dan Conway, Jianlong Zhou, and Fang Chen. 2017. User Trust Dynamics: An Investigation Driven by Differences in System Performance. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (Limassol, Cyprus) (IUI '17). ACM, New York, NY, USA, 307–317. <https://doi.org/10.1145/3025171.3025219>
- [42] Yunfeng Zhang, Q. Vera Liao, and Rachel K. E. Bellamy. 2020. Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) (FAT\* '20). ACM, New York, NY, USA, 295–305. <https://doi.org/10.1145/3351095.3372852>