Towards a Generalized Scale to Measure Situational Trust in AI Systems

LENA DOLINEK, TU Wien, Austria
PHILIPP WINTERSBERGER, TU Wien, Austria

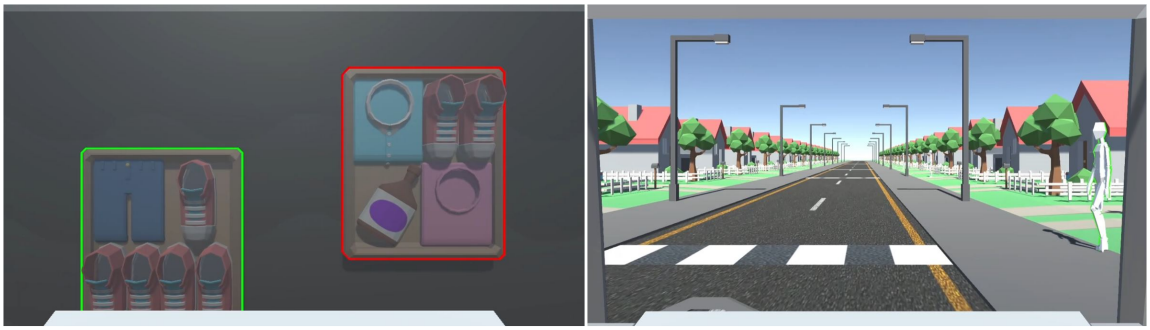

Fig. 1. The updated scale was tested in two human-AI interaction scenarios with equivalent underlying game mechanics, where participants cooperated with a baggage scanner (left) and an automated vehicle (right).

Trust is a highly relevant concept determining how users interact with AI systems. However, while trust is a multi-dimensional construct influenced by various contextual factors, most subjective measurements assess it on a more general level. To better assess the situation- and context-specific nature of trust, we update the situational trust scale for automated driving to allow assessment in other domains. Initial results, based on a lab study with N=23 participants who completed the scale after cooperating with AI systems in two independent scenarios (automated vehicle and AI-supported baggage scanner), confirm that all scale items load onto a single factor. However, additional investigations will be necessary to determine to which degree the scale is sensitive to variations in automation performance. Still, the updated scale can be considered a first step towards measuring situational in various application areas where users interact with automated and AI-driven systems.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**; **Interaction paradigms**; **HCI theory, concepts and models**.

Additional Key Words and Phrases: Human-AI Interaction, Trust in Automation, Scale Development, Trust Measurements

ACM Reference Format:
Lena Dolinek and Philipp Wintersberger. 2022. Towards a Generalized Scale to Measure Situational Trust in AI Systems. In *CHI '22: ACM Conference on Human Factors in Computing Systems, April 30–May 06, 2022, New Orleans, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Trust is a core construct in the interaction between humans and technology. It has a long history in the human factors and automation literature (especially in domains such as aviation, human-robot interaction, or automated driving)

Manuscript submitted to ACM

[4, 6, 19], and recently, it gained additional momentum due to the rise of artificial intelligence (AI) systems and emerging issues regarding AI transparency and explainability [5, 10, 21]. The words "trust" and "trustworthiness" appear over 100 times in a regulatory proposal issued by the European Commission, at it is suggested that trust and beneficial use of AI have a close relationship [3]. Further, "trust issues" have already contributed to various accidents (some even fatal) with driving automation systems. For example, the US National Transport Safety Board concluded in the investigation of a deadly crash of a Tesla driver that the driver *"was overly confident in [Tesla] Autopilot's capabilities"*. Falsely attributed capabilities (i.e., "overtrust") in AI systems can yield dangerous outcomes, while on the other hand, relevant systems may be rejected when they are not trusted by users (i.e., "distrust"). Consequently, trust in technology should be calibrated to appropriate levels so that the benefit of technology can be exploited appropriately [14]. Trust is highly context-specific [15], and a beneficial trust relationship requires users to assess a system concerning its performance and the demands of the dynamic situations where systems are in use. To achieve calibrated trust and best utilize the construct in research and the design of AI systems, the construct must be measured appropriately. Trust measurements can be behavioral, physiological, or subjective. Behavioral measurements assess the outcome of the interaction, for example, in terms of compliance (*"response when an operator acts according to a warning signal"*) and reliance (*"warning system indicates the system is intact and the user does not take precaution"*) rates [12, 17]. However, behavioral measurements are influenced by other psychological constructs (for example, situation awareness, distraction, etc.). Physiological assessment like eye-tracking [7] or stress measurements [18] only allow to infer potential conclusions. Thus, both behavioral and physiological assessment must be considered indirect rather than direct measurements. The only direct ways to measure trust are subjective, for example, by asking users via questionnaires or interview techniques. Many existing measurements, such as the "Automation Trust Scale" by Jian et al. [11] assess trust on a more general level and do not allow investigating the temporal, situation- and context-specific nature of this multi-dimensional construct. Thus, there is a need for additional measurements that can capture the varieties of changing system performance or dynamic interactions. To counter, Holthausen et al. [9] have developed the "situational trust scale". Still, this scale is explicitly tailored to vehicle automation and does not support measuring trust in other domains. However, a more generalized scale would be beneficial to allow investigating the dimension of situational trust in other domains and allow comparison between them.

In this work, we present an adaptation of the scale proposed by Holthausen et al. [9], where we aimed at a more generalizable measurement of situational trust. We have updated the wording to become more neutral and independent of a particular system, added items that represent other essential factors (according to the trust model by Hoff and Bashir [8]), and evaluated the result in a lab study where participants played a game developed in Unity 3D. The game mechanics incorporate relevant aspects of user-trust relationships (such as changing system performance and secondary tasks). We created two different interfaces for the game mechanics to investigate if the scale can be used in different scenarios (in particular, an automated driving system as present in the original publication [9] and an AI-supported baggage scanner, see Figure 1). Study participants were exposed to both settings and completed the updated scale multiple times, where we dynamically varied the performance of the mimicked systems. **Thereby, we aimed at evaluating if (RQ1) the scale items load onto a single factor, (RQ2) if the scale can successfully capture performance variations, and (RQ3) if these systems are different in terms of the trust ratings and user behavior**. Initial results based on 23 participants indicate slight differences between the two scenarios and confirm that the questionnaire is a potentially valid measurement of situational trust, which allows investigating the construct in a broader range of human-AI interaction scenarios.

2 RELATED WORK AND SCALE DEVELOPMENT

Lee and See [14] define trust in automation as *"the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability"*, which is formed as a result of analytic, analogical, and affective processes. A great variety of factors have been revealed to influence trust over the years, and these factors have been integrated into a conceptual model by Hoff and Bashir [8], who distinguish between dispositional, situational, and learned trust. Situational trust is different from trait-based, dispositional factors and refers to contextual differences in trust development. Thereby, Hoff and Bashir [8] distinguish between external (type and complexity of a system, workload, task difficulty, workload, perceived risks and benefit, setting, and framing of a task) and internal (self-confidence, expertise, mood, and attentional capacity) variability. Their conceptual model is frequently cited (over 1100 citations on Google Scholar), yet empirical evaluations of it are scarce [9]. One reason might be that many existing scales address trust more generally. For example, trust scales have been proposed by Chien et al. [1], Körber [13], Schaefer [20], and others, and those scales are (also because of many scale items) typically used at the end of an experimental condition. To measure situational trust multiple times during an experiment (for example, to investigate the effect of changing system performance over time), the scale by Holthausen et al. [9] provides six items referring to situational factors (to be assessed on a 7-point Likert scale from "fully disagree" to "fully agree"). The scale evaluation (conducted in the form of an online survey) confirmed situational trust as independent of more general, trait-based trust via factor analysis. It was shown that the scale can capture performance variations in different situations. Still, as described above, the original scale is tailored to the domain of automated driving systems and does not allow being used in other scenarios. Thus, we have updated the individual scale items and included two additional items that cover other important situational trust factors according to the Hoff Bashir model [8], namely mood, organizational setting, and framing of the task (see Table 1).

Original Item	**Neutral Item**	**Trust Factor**	**Factor Loading**
I trust the automation in this situation	I trust the system in this situation	Type of system System complexity	.812
*I would have performed better than the AV in this situation	*I would have performed better than the system in this situation	Self-confidence Subject matter expertise	.752
In this situation, the AV performs well enough for me to engage in other activities	Given the system's performance, there was no need to monitor it continuously	Perceived benefits Workload Task difficulty Attentional capacity	.685
*The situation was risky	*The situation was risky	Perceived risks	.796
*The AV made an unsafe judgement in this situation	*The system made an usafe judgement in this situation	Perceived risks	.907
The AV reacted appropriately to the environment	The system reacted appropriately in this situation	Perceived risks Perceived benefits	.893
-	The system's use is appropriate in this setting if it behaves like it did in this situation	Organizational setting Framing of task	.898
-	I felt positive about working with the system in the experienced situations	Mood Type of system	.885

Table 1. Situational Trust Scale items, based on the original formulation provided by Holthausen et al. [9]. *reverse scored.

3 SCALE EVALUATION

We developed a computer game in Unity3D to evaluate the scale in different scenarios in a dual-task setting. In the following, we will explain the details of the game on hand of the "baggage scanner" scenario (see Figure 1 left).

3.1 Game Mechanics and Interfaces

Study participants engaged in the game have to perform two tasks simultaneously. The primary task represents cooperation with the AI system, which in the presented example, identifies problematic items (in particular, bottles with liquid) in baggage parcels. Those are placed on a conveyor belt and are scanned by the AI in the screen center. In case the AI detects a problematic item, it is removed. However, sometimes the AI makes an error, and a problematic baggage parcel slips through as it is not detected correctly. This represents a "miss" by the AI, and we solely focused on misses and did not include false positives (i.e., non-problematic item removed). The players' task is to observe the system and intervene manually by clicking on the baggage item if it is not detected. Simultaneously, the player is engaged in a secondary number sorting task. A list of random numbers is presented on a second virtual display, which must be sorted in ascending or descending order (see Figure 2). Players can navigate between the two screens using the mouse wheel, and in case the secondary task is brought to the foreground, the view on the conveyor belt and the AI system is obstructed and blurred. These mechanics allow investigating the monitoring behavior of the user, who is told to perform as many secondary number sorting tasks as possible without missing items that are wrongly classified as safe. In each trial, five congregations of baggage items are passing the conveyor belt, where we can configure the number of errors the AI system makes in detecting them to vary the system's performance. Consequently, the mechanics allow investigating the effect of the scenario and varying performance on:

- **Reliance Behavior** , which can be assessed by quantifying the number correct interventions by the user vs. the number of problematic baggage parcels that "slipped through" when the simulated AI system fails.
- **Monitoring Behavior** , which can be assessed by quantifying number and frequency of the user inspecting the AI's behavior (i.e., switching back from the secondary number ordering task to monitor the system).
- **Subjective Situational Trust** , which is assessed by letting the player complete the situation trust scale as described above (see Table 1) after each trail (which consists of 5 subsequent scans).

For the other scenario, we modeled an automated driving task similar to the situations evaluated in [9]. Instead of observing the baggage scanner, the user is placed in an automated vehicle that drives down a straight road with multiple subsequent zebra crossings (see Figure 1 right). Analogous to the baggage scanner scenario, a person crosses the street on five of these zebra crossings. The automated vehicle either detects the pedestrian and stops, or it continues driving, requiring the pedestrian to step back (representing the failure situation). If such a failure occurs, the player can manually stop the car. All other aspects (secondary number ordering task, obstructed/blurred view when the secondary task is open to investigating the monitoring behavior, etc.) are precisely the same as the other scenario. Additionally, the developed framework allows configuring the exact game logic (i.e., how many trials with a particular scenario, sequences of performance/error rates), and also additional Likert scale questions to be displayed while playing the game can be included using an XML configuration file.

3.2 Study Design

For the presented evaluation, study participants were exposed to a series of five trials (i.e., 25 situations as one trial consists of 5 situations, i.e., scans or zebra crossings) for each of the two scenarios (baggage scanner and automated

Fig. 2. Dual-task setting: In both the baggage scanner (left) and the automated vehicle scenario (right), participants had to complete as many number sorting tasks as possible while intervening in case the AI system fails. When the secondary task was in foreground, the view on the primary task was partly obstructed and blurred. The assignment of ascending/descending number sorting tasks was randomized.

vehicle; within-subjects design in counterbalanced order). After each trial, they completed the situational trust scale as outlined in Table 1. The performance of the AI systems was modified so that each participant experienced two trials without any failure (i.e., 100% performance), while the remaining three trials had a random error rate between 1 and 3 failures, which yielded an overall performance of 76-96% over all trials. The order of the errors within an individual trial and the order of faulty trials in the overall sequence was randomized.

3.3 Participants and Procedure

The present publication is based on the evaluation of the first N=23 participants (primarily students and University staff; 13 male and 10 female, M = 35.65; SD = 14.61 years).
After expressing consent and completing a short demographic survey, we issued pre-test questionnaires addressing perceived risks of AI technologies and propensity to trust [16]. Then, participants saw a short video explaining their tasks for the first scenario (how to control the game, focusing on performing well in the secondary task while intervening in case a failure occurs). Then, they experienced the five trials in the respective first scenario. The condition concluded with another survey that included the trust in automation scale by Jian et al. [11] and the technology acceptance model [2]. The publication at hand solely focuses on the ratings of the situational trust scale as described in Table 1. The whole procedure was then repeated for the other scenario. The experiment lasted 45-60 minutes per participant, and the study was conducted under consideration of the Universities' precautions regarding the COVID-19 pandemic.

3.4 Results

The evaluation was carried out using IBM SPSS Version 27, and results are reported as statistically significant at $p<.05$. In the following, we present the statistical analyses regarding the proposed research questions.

3.4.1 Scale Reliability and Factor Analysis. After reversing the respective items (see Table 1), we calculated Cronbach's alpha to assess the inter-item reliability of the scale. The analysis was performed by averaging the ratings for each scale item over all trials independently for the two investigated scenarios. The results show a Cronbachs' alpha of .900 for the baggage scanner and .925 for the automated vehicle scenario, which indicates a high internal consistency.
Then, we calculated the average of the individual scale items across all investigated trials and over both investigated scenarios (baggage scanner and automated vehicle) to conduct a factor analysis. The evaluation confirms that all 8 items load onto a single factor, which suggests that the underlying construct of "situational trust" is successfully measured as single trust dimension. The factor loading for each item can be seen in Table 1 on page 3.

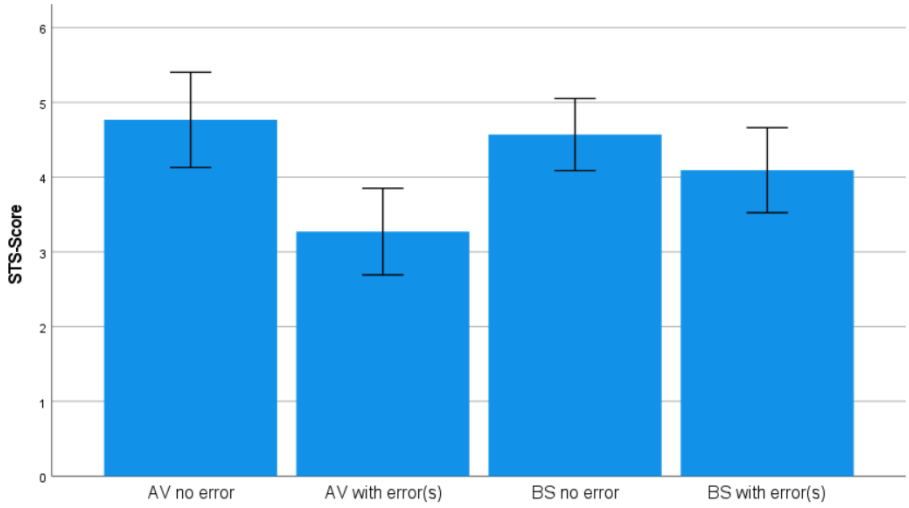

Fig. 3. Diagrams showing the mean situational trust scores for the automated vehicle (AV, left) and the baggage scanner (BS) scenarios without and with one or more errors (error bars: 95% CI).

3.4.2 Scale Sensitivity. Given the high internal consistency according to Cronbach's alpha, we calculated the scale values for situational trust by averaging the individual scale items (see Figure 3). To investigate if the scale can capture variances in automation performance, we calculated the average situational trust ratings for trials with and without failures per user and conducted Wilcoxon signed rank tests for comparison. Regarding the automated driving scenario, the difference was significant ($Z = -3.587, p < .001$). Participants trusted the vehicle significantly more ($M = 4.77, SD = 1.47$) in error-free trials compared to situations with one or more errors ($M = 3.27, SD = 1.34$). However, there was not significant difference visible for the baggage scanner scenario ($Z = -1.542, p = .123$), although descriptive statistics indicate that participants attributed the baggage scanner higher situational trust when it operated error-free ($M = 4.57, SD = 1.12$) than in situations with one or more failures ($M = 4.09, SD = 1.31$).

3.4.3 Differences between the Scenarios. We also compared the trials with and without failures between the two scenarios using Wilcoxon signed rank tests. Results confirm what can be seen in Figure 3: While the situational trust scale ratings were similar in trials without errors ($Z = -.909, p = .363$), the ratings for trials with one or more failures differs between the automated vehicle and the baggage scanner scenario ($Z = -2.662, p = .008$), where the former showed a higher trust decline.

Further, we compared the monitoring ratio and the number of completed secondary tasks between the two scenarios for all trials with and without failures. Regarding the monitoring ratio, participants monitored the behavior of the automated vehicle significantly more often than the baggage scanner. This is visible for both the non failure (automated vehicle monitoring $M = .24, SD = .12$; baggage scanner monitoring $M = .18, SD = .10$; $Z = -2.357, p = .018$) and failure groups (automated vehicle monitoring $M = .29, SD = .12$; baggage scanner monitoring $M = .17, SD = .09$; $Z = -3.057, p = .002$). A logical consequence of this behavior is the flipped number of completed number sorting tasks, which are significantly higher for the baggage scanner scenario. In both the non-failure (number of secondary tasks completed in the automated vehicle $M = 6.86, SD = 1.98$; number of tasks completed in the baggage scanner

scenario $M = 8.11, SD = 2.24$; $Z = -3.242, p = .001$) and failure groups (number of secondary tasks completed in the automated vehicle $M = 6.09, SD = 1.71$; number of tasks completed in the baggage scanner scenario $M = 8.33, SD = 2.49$; $Z = -3.726, p < .001$), participants completed more secondary tasks than in the automated vehicle scenario.

4 DISCUSSION

The results of the presented study suggest that a more neutral and less scenario-specific formulation of the "situational trust scale" (initially developed for the use case of automated driving [9]) could yield to a general measurement for the construct of situational trust, which is an important trust dimension according to the conceptual trust model proposed by Hoff and Bashir [8]. We extended the scale by two additional items, which represent the organizational setting, the framing of a task, and the user's mood and type of system, to allow application to and comparison of other human-AI interaction scenarios. We could obtain similar results as in the original publication of the scale [9]: The 8 scale items show a high internal consistency according to Cronbach's alpha, and a factor analysis confirms that they load onto a single factor (**RQ1**).

However, regarding the presented evaluation, the scale showed higher sensitivity to performance variations in the automated driving domain than the scenario of an AI-supported baggage scanner. In the automated driving scenario, the situational trust ratings for trials with and without failures significantly differed, while such a difference was not present in the baggage scanner scenario. There are multiple possible explanations for this behavior. The number of study participants (N=23) in the presented evaluation was comparably low, and a larger sample might account for the missing statistically significant effects in all performance variations. A higher sample size will also allow to investigate the groups according to the number of (one, two, or three) failures individually, which was not performed in this initial evaluation due to a lack of statistical power: a power analysis resulted in 36 participants needed to detect differences in a 2 (scenarios) by 4 (failure groups) design (with an expected effect size of .25). Another explanation could be a different perception of the risks associated with the two scenarios. Perceived risks are said to be an important factor in trust formation [8, 15], and while failures of automated driving systems can quickly lead to safety-critical situations, overlooking a bottle with liquid in a baggage scanner could be perceived as less risky from the perspective of study participants. Given that a comparison of the trust ratings between the two scenarios showed differences only for the failure groups (while no differences could be obtained for non-failure groups), maybe less risky scenarios require a larger performance drop to show a significant decline in subjective situational trust (**RQ2**).

This explanation would fit other differences that have been obtained in this study. Participants completed a significantly higher number of the (secondary) number sorting tasks and, in turn, monitored the automation less frequently when operating the AI-controlled baggage scanner, compared to the automated vehicle scenario (**RQ3**). Given that there are, to our best knowledge, no other dedicated measurements for situational trust, we believe that the modified and extended questionnaire still can provide a valuable tool for the community. Since trust in AI and explanations, robotic, as well as automated systems is frequently addressed in HCI studies, an extended and generalized situational trust scale will be important to compare results and identify trust-related differences and similarities between various trust dimensions, context factors, and user interfaces in different application domains. With only 8 items, the scale is also relatively small compared to other trust measurements [1, 13, 20], and it can be administered multiple times during a study to evaluate different UI designs (for example, considering explanations or other HMI interventions), experimental conditions, performance variations, or to research the temporal development of trust formation.

5 LIMITATIONS AND FUTURE WORK

Although we put much effort into the game mechanics, intending to make the two investigated scenarios comparable, we cannot completely rule out that the source of some observed differences (like the monitoring behavior or the number of completed secondary tasks) emerge from other factors than participants' perception of the scenarios. There exist slight differences regarding the exact timings of the simulated systems' actions and failures, and also the obstruction due to the secondary task is slightly different. Still, we claim that comparing different results gathered in independent studies would be even less valid. Another limitation in most trust-related lab studies is the absence of real risk, which emerged solely from participants' imaginations. In this sense, the visuals are also relatively schematic, and future implementations should provide a more realistic visual appearance and interaction methods, for example, by putting the scenarios in virtual reality.

Further, the evaluations presented in this work are based on a relatively small sample with only 23 participants. We will include more participants, add additional scenarios, re-evaluate our initial results, and investigate other issues not addressed in this work, like correlation analyses between error rates, monitoring behavior, and interventions to gain better insights into the complex interplay between subjective situational trust and users' behavioral responses. Finally, the scale addresses a particular set of human-AI interaction scenarios, namely supervisory control situations. Those are characterized by the continuous nature of the AI application in operation (i.e., a time critical setting). Still, a large collection of AI systems are non-continuous decision aid systems, i.e., they allow users to evaluate and judge their output without inherent time pressure. To allow the scale assessing such systems, some questions (such as *"Given the system's performance, there was no need to monitor it continuously"*, or one could speak of particular "decisions" instead of "situations") need to be adapted.

6 CONCLUSION

Many existing trust scales assess trust on a general level and do not consider the context-and situation-specific nature of trust formation. Thus, new standardized measurements are needed to capture the concept of "situational trust" [8]. In this work, we have updated and extended the "situational trust scale for automated driving" (STS-AD) towards a more neutral version (STS) that can be applied in multiple human-AI interaction scenarios. We evaluated the scale with 23 participants, who cooperated with (simulated) AI-controlled automated systems in two different supervisory control scenarios (automated driving and an AI-supported baggage scanner) and completed the scale after experiencing different performance variations of these systems. Our results confirm the existence of situational trust as a relevant trust dimension, where all items loaded onto the single factor. However, at this point we cannot confirm that the scale is sensitive to performance variations in all investigated scenarios – while the situational trust ratings significantly dropped when errors were present in the automated driving scenario, ratings were still comparably high when the system performance dropped (to 76-96%) in the baggage scanner scenario. Further, we have revealed some differences and similarities of the investigated scenarios, such as situational trust ratings concerning automation performance, monitoring, and behavior in secondary tasks.

We will conduct additional evaluations with a larger sample to evaluate if researchers and designers can use the updated situational trust scale in a wide range of human-AI interaction scenarios in the future.

REFERENCES

[1] Shih-Yi Chien, Zhaleh Semnani-Azad, Michael Lewis, and Katia Sycara. 2014. Towards the Development of an Inter-cultural Scale to Measure Trust in Automation. In *Cross-Cultural Design*, P. L. Patrick Rau (Ed.). Springer International Publishing, Cham, 35–46.

[2] Fred D Davis. 1993. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International journal of man-machine studies* 38, 3 (1993), 475–487.

[3] 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

[4] 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.

[5] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. 2019. XAI—Explainable artificial intelligence. *Science Robotics* 4, 37 (2019).

[6] 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

[7] Sebastian Hergeth, Lutz Lorenz, Roman Vilimek, and Josef F Krems. 2016. Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving. *Human factors* 58, 3 (2016), 509–519.

[8] 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.

[9] 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.

[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. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 624–635.

[11] 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.

[12] David Keller and Stephen Rice. 2009. System-wide versus component-specific trust using multiple aids. *The Journal of General Psychology: Experimental, Psychological, and Comparative Psychology* 137, 1 (2009), 114–128.

[13] Moritz Körber. 2019. Theoretical Considerations and Development of a Questionnaire to Measure Trust in Automation. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*, Sebastiano Bagnara, Riccardo Tartaglia, Sara Albolino, Thomas Alexander, and Yushi Fujita (Eds.). Springer International Publishing, Cham, 13–30.

[14] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.

[15] 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* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 177–185. https://doi.org/10.1145/3342197.3344525

[16] Stephanie M Merritt, Heather Heimbaugh, Jennifer LaChapell, and Deborah Lee. 2013. I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human factors* 55, 3 (2013), 520–534.

[17] Joachim Meyer. 2004. Conceptual issues in the study of dynamic hazard warnings. *Human factors* 46, 2 (2004), 196–204.

[18] Drew M. Morris, Jason M. Erno, and June J. Pilcher. 2017. Electrodermal Response and Automation Trust during Simulated Self-Driving Car Use. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61, 1 (2017), 1759–1762. https://doi.org/10.1177/1541931213601921

[19] Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. 2008. Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of cognitive engineering and decision making* 2, 2 (2008), 140–160.

[20] Kristin E. Schaefer. 2016. *Measuring Trust in Human Robot Interactions: Development of the "Trust Perception Scale-HRI"*. Springer US, Boston, MA, 191–218. https://doi.org/10.1007/978-1-4899-7668-0_10

[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.