Workshop on Trust and Reliance in AI-Human Teams (TRAIT)

Gagan Bansal*
bansalg@cs.washington.edu
University of Washington
Seattle, USA

Alison Smith-Renner*
arenner@dataminr.com
Dataminr
New York, USA

Zana Buçinca
zanabucinca@g.harvard.edu
Harvard University
USA

Tongshuang Wu
wtshuang@cs.washington.edu
University of Washington
Seattle, USA

Kenneth Holstein
kjholste@cs.cmu.edu
Carnegie Mellon University
USA

Jessica Hullman
jhullman@northwestern.edu
Northwestern University
USA

Simone Stumpf
simone.stumpf.1@city.ac.uk
University of Glasgow
UK

ABSTRACT

As humans increasingly interact (and even collaborate) with AI systems during decision-making, creative exercises, and other tasks, *appropriate* trust and reliance are necessary to ensure proper usage and adoption of these systems. Specifically, people should understand when to trust or rely on an algorithm's outputs and when to override them. While significant research focus has aimed to measure and promote trust in human-AI interaction, the field lacks synthesized definitions and understanding of results across contexts. Indeed, conceptualizing trust and reliance, and identifying the best ways to measure these constructs and effectively shape them in human-AI interactions remains a challenge.

This workshop aims to establish building appropriate trust and reliance on (imperfect) AI systems as a vital, yet under-explored research problem. The workshop will provide a venue for exploring three broad aspects related to human-AI trust: (1) How do we clarify definitions and frameworks relevant to human-AI trust and reliance (e.g., what does trust mean in different contexts)? (2) How do we *measure* trust and reliance? And, (3) How do we *shape* trust and reliance? As these problems and solutions involving humans and AI are interdisciplinary in nature, we invite participants with expertise in HCI, AI, ML, psychology, and social science, or other relevant fields to foster closer communications and collaboration between multiple communities.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; *Interactive systems and tools*; • **Computing methodologies →** Machine learning.

KEYWORDS

trust, human-centered artificial intelligence, reliance, uncertainty

ACM Reference Format:
Gagan Bansal, Alison Smith-Renner, Zana Buçinca, Tongshuang Wu, Kenneth Holstein, Jessica Hullman, and Simone Stumpf. 2022. Workshop on Trust and Reliance in AI-Human Teams (TRAIT). In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3491101.3503704

1 INTRODUCTION AND BACKGROUND

AI is increasingly deployed in people-facing systems to assist and collaborate with users (e.g., domain experts) through their outputs, such as for assisted decision making and creative tasks. Consequently, interest in human-AI interaction and collaboration is exploding both within and beyond the HCI community [1, 9, 22, 24]. But when is a human-AI collaboration "successful"—and how do we measure success in these contexts—and what factors (of people, AI, interface, task, etc) impact the success of the human-AI teams? One set of frequently discussed, relevant constructs are human-AI "trust" and "reliance." For example, many studies have started to investigate the impact that system transparency and accuracy have on trust [4, 7, 11, 25]; the role of trust in the user experience of complex algorithmic systems, such as the impact on satisfaction, adoption, and reliance [5, 15]; and the potentially undesirable effects that too much or too little trust might have on a user's (over- or under-) reliance upon a given system [2, 13, 19, 20]. Despite all this interest, the community still lacks a conceptual understanding of definitions of human-AI trust and reliance in the context of successful human-AI collaborations [5, 8, 10].

In practice, to ensure that human-AI collaborations do more good than harm, it is vital that we understand, measure, and shape human-AI trust and reliance; for example, when is user trust and reliance on AI warranted or *appropriate*, how does it evolve over the course of (long-term) human-AI collaborations, which factors (individual, contextual, etc) influence it, and how can we design interventions that guard against harms caused by inappropriate trust or reliance and instead promote more appropriate reliance within human-AI collaborations? Indeed, recent years have seen

*Equal contribution.

Woodstock '18, June 03–05, 2018, Woodstock, NY
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9156-6/22/04.
https://doi.org/10.1145/3491101.3503704

renewed interest in human-AI trust within the AI and HCI community [12, 16, 18, 21, 23] motivated by directions and approaches introduced in prior literature on human-automation interaction and human factors [14]. However, it is not yet well understood what generalizes from these works and what new considerations modern human-AI collaborations introduce [3, 8].

Conceptualizing trust and reliance, formalizing methods to measure them, and studying and shaping their impact on human-AI performance remains challenging for many reasons. For example, defining and measuring human "trust" in an AI system is subject to questions that have long plagued other behavioral sciences, such as the extent to which researchers should focus on *realized behaviors* when interacting with an AI system (e.g, adoption, reliance, or adherence) versus the *beliefs* that give rise to those behaviors [5, 8]. Shaping human-AI trust and reliance may be even cumbersome—Even in the context of simple, ubiquitous human-AI teams, such as AI-assisted decision making, trust may depend on many aspects of the human's mental model: how they conceive the task, their domain expertise, how they conceive of the AI's objectives and decision strategies, their understanding of the AI's capabilities and limitations, and (their understanding of) their own capabilities and limitations, to name just a few. While (post-deployment) AI systems are often conceived as independent of its developers, in real-world organizational settings, system trust may in fact depend on the user's trust in those developers [6], and humans may choose to rely on an AI system for reasons unrelated to their trust in those systems (e.g., organizational pressures or incentives). Finally, "trust" is relevant to human-AI collaboration in large part because such collaborations often involve decision-making under uncertainty. And, behavior under uncertainty is complex and humans decisions under uncertainty exhibit many biases that depend heavily on context. For example, people can be either under- and over-sensitive to important features of a decision task like sample size or variance, depending on the specific situation. As a result, recent results indicating under- or over-reliance in human-AI collaboration may be more context-dependent than is commonly appreciated.

As research in the areas of human-AI interaction and collaboration grows, it is increasingly important to synthesize empirical results across contexts and conceptualize and formalize key concepts. In the absence of a theoretical basis that grounds our understanding, findings from research (e.g., focused on socio-technical approaches) on improving trust and reliance in human-AI collaborations can seem ad-hoc and at times contradictory. For example, prior research has found that exposing system confidence (or uncertainty) can both help [26] or hinder [17] trust calibration. Without a synthesized understanding of research space around trust, reliance and the effectiveness of human-AI collaboration, the risk is that this body of research will continue to expand in size, but not necessarily in insight and impact on the real-world.

2 WORKSHOP GOALS

Our goal is to **establish appropriate trust and reliance on (imperfect, people-facing) AI systems as a vital yet under-explored research problem**, especially for developing human-AI interactions that effectively *augment* users. We wish to highlight incentives and challenges from pursuing this problem and help spur impactful future research.

We aim to **synthesize (and conceptualize) various definitions and frameworks relevant to human-AI trust and reliance.** Many prior works across research sub-communities (such as, HCI, AI, ML, NLP, psychology, etc) motivate their research using concepts and terms similar to "trust" and "reliance" in AI; however, we notice a lack of consensus on many dimensions: (1) What does "trust" mean in different contexts? (2) When should users trust AI and when should trust in AI be defined as warranted or appropriate? (3) What potential harms would inappropriate trust cause in different domains? Thus through this workshop, we wish to synthesize domain-agnostic or -specific themes, definitions, known observations and results, and key challenges to inform future research. We plan to include a broad perspective on human-AI trust by including topics on (and discussions around) (1) people interacting with complex algorithmic systems (e.g., in various domains and social contexts, with varying levels of prior experiences and biases), (2) human-human interaction and group dynamics, and (3) trust from the perspective of machine learning systems (e.g., calibrating AI's confidence, explainable AI, etc).

In accordance with these key themes and challenges, we also would like spur research on *measuring* and *shaping* trust and reliance. **Measuring trust and reliance** is a pre-requisite for developing mechanisms that lead to appropriate trust and reliance. Such research may include categorizing objective and subjective trust (e.g., through quantitative human-AI team performance and qualitative user reflection) and eliciting humans' mental models on AI systems, as well as identifying factors (e.g., system uncertainty, mental model, decision risk, etc.) that affect humans' trust and reliance . We also aim to further **promote research on developing mechanisms to shape trust and reliance with respect to the factors**, e.g., by calibrating user expectations on model capabilities through careful user training and on-boarding, by designing human-AI team architectures, visualizations and interfaces that supports more effective human-AI communications in a team setting, etc.

To summarize, the workshop will focus on three broad aspects:

(1) How do we *clarify* definitions and frameworks relevant to human-AI trust and reliance?
(2) How do we *measure* trust and reliance and identify factors that affect these constructs?
(3) How to we *shape* trust and reliance toward effective human-AI collaboration?

As the problems and solutions involving AI and people are interdisciplinary in nature, we will invite people with expertise in HCI, AI, ML, psychology, social sciences, or other relevant fields to foster closer communication and collaborations between multiple communities.

3 ORGANIZERS

Gagan Bansal Gagan Bansal is a Ph.D. candidate in the Allen School of Computer Science and Engineering at the University of Washington, Seattle. He is part of the UW Lab for Human-AI Interaction and conducts interdisciplinary research on Artificial Intelligence and Human-Computer Interaction with focus on developing human-centered AI systems for augmenting people.

Alison Smith-Renner is a Senior Research Scientist at Dataminr. Her research interests lie at the intersection of AI and HCI, focusing on transparency and control for human-in-the-loop systems to engender appropriate trust and improve human performance. Alison received her Ph.D. from the University of Maryland, College Park. She has organized various workshops on explainable AI and human-centered ML, including at IUI, CHI, and TEI, and she has held senior committee roles at IUI.

Zana Buçinca is a Ph.D. Candidate at Harvard University. Her research lies at the intersection of Human-Computer Interaction and Artificial Intelligence. Informed by cognitive science theories, Zana designs, builds, and evaluates AI for decision-making support.

Tongshuang (Sherry) Wu is a Ph.D. Candidate at the University of Washington, Seattle. Her research lies at the intersection of Human-Computer Interaction and Natural Language Processing, aiming to support humans interacting with imperfect AIs, by debugging and correcting AIs interactively. Her work improves system transparency and controllability in human-AI collaborations, and encourages global understanding and refinement in model analysis.

Kenneth Holstein is an Assistant Professor in Human-Computer Interaction at Carnegie Mellon University. His research interests lie at the intersection of HCI, AI, design, and cognitive science, focusing on the design, development, and evaluation of human-AI collaborative systems in complex social contexts.

Jessica Hullman is an Associate Professor of Computer Science at Northwestern University. Her research looks at how to design, evaluate, coordinate, and think about visual representations of data and model predictions for inference, decision making, and communication, including the effects of visualizing uncertainty on belief updating and potential for behaviorally induced feedback loops in visualizing model predictions in strategic settings.

Simone Stumpf is a Reader in Responsible and Interactive AI at University of Glasgow, UK. She has a long-standing research focus on user interactions with machine learning systems. Her work has contributed to shaping the field of Explainable AI (XAI) through the Explanatory Debugging approach to interactive machine learning, providing design principles for crafting explanations. She is a member of the organising committee of the ExSS workshop at IUI, and has held senior committee roles at CHI, IUI and EICS conferences.

4 WEBSITE

We will set up a website[1] to advertise and disseminate the workshop's information and call for proposals. We will also use this website to share workshop contributions, including accepted papers, and support future engagement.

5 PRE-WORKSHOP PLAN

In addition to the website, we will advertise the workshop through email distribution lists at relevant conferences and research institutions (including, but not limited to, FAccT, IUI, CHI, ACL, and CSCW mailing lists), direct communication with colleagues in the field, and social media.

We will have a program committee (PC) with experts from diverse research organizations and backgrounds who will help us to curate the workshop by disseminating the call for papers and reviewing submissions. So far, we have commitments from 21 PC members with expertise in human-AI interaction and related topics, including Ben Shneiderman (University of Maryland), Elena Glassman (Harvard), Jenn Wortman Vaughan (MSR), Krzystof Gajos (Harvard), Matthew Kay (Northwestern), Tim Miller (University of Melbourne), Alon Jacovi (Bar Ilan University), Maria De-Arteaga (UT Austin), Vera Liao (MSR), Hal Daume (University of Maryland), Gonzalo Ramos (MSR), Michael Terry (Google), Ming Yin (Purdue), Maia Jacobs (Northwestern), Erin Chiou (ASU Adapt Lab), Ella Glikson (Bar Ilan University), Tom Williams (Colorado School of Mines), Shi Feng (University of Maryland), Zahra Ashkortab (IBM), Brian Lim (National University of Singapore), and Michael Bernstein (Stanford HAI).

Participants interested in giving a presentation at the workshop will need to submit a short paper (2-6 pages). Submission types will include, but are not limited to, position papers summarizing authors' existing research in the area and how it relates to the workshop theme, papers offering an industrial perspective or real-world approach to the workshop theme, papers that review the related literature and offer a new perspective, and papers that describe work-in-progress research projects. We will encourage submissions that present diverse viewpoints on the workshop topics, and encourage participation across relevant fields, such as AI, HCI, and cognitive psychology. We will use Easychair to collect and review these submissions. Each submission will be reviewed by at least two PC members and one organizing committee member. Accepted papers will be presented either during paper sessions, or with posters during coffee breaks (see the next section for details on the workshop schedule.)

As workshop organizers, we value and are committed to diversity, equity, and inclusion. We welcome and encourage the participation of people who identify with any historically marginalized or underrepresented group. Further, the listed platforms and technologies should not be a barrier to the participation of anyone interested in this workshop. If the technologies listed do not accommodate participants needs, we will work with participants to find alternative solutions.

6 WORKSHOP STRUCTURE

Based on prior similar workshops and the interest shown by colleagues, we expect between 40 and 50 participants. As such, the workshop is currently designed for 50 participants. If interested participants exceed this number after the initial advertisement of the workshop, we may adjust the workshop structure to accommodate up to 100 participants. We plan to organize our proposal as a single-day workshop, from 9:00 AM to 5:00 PM local time (including breaks), in a hybrid format. We hope most participants join the in-person workshop but will also plan for synchronous online access to the workshop. For the virtual participation experience, we intend to to use a combination of Zoom (for synchronized talks) and Discord (for virtual and asynchronous question answering and online discussions). However, the organizers will work with the technical team at CHI 2022 to utilize provided streaming methods (with captioning for accessibility), so as to minimize the jump across platforms.

[1] https://chi-trait.github.io

Slot	Theme
09:00 – 09:15 (15min)	Welcome
09:15 – 10:15 (60min)	Keynote talk by Prof. John D. Lee
10:15 – 10:45 (30min)	Paper sessions 1
10:15 – 10:45 (30min)	Coffee break (concurrent with poster presentations)
10:45 – 11:30 (45min)	Panel with experts that have diverse and well-balanced expertise
11:30 – 12:00 (30min)	Paper sessions 2
12:00 – 13:00 (60min)	Lunch break
13:00 – 14:30 (90min)	Group activity 1 *(60 min discussion + 30 min group result sharing)*
14:30 – 15:00 (30min)	Coffee break (concurrent with poster presentations)
15:00 – 16:30 (90min)	Group activity 2 *(60 min discussion + 30 min group result sharing)*
16:30 – 16:45 (15min)	Closing remarks

Table 1: Tentative schedule for the proposed single day workshop. The workshop will dedicate sufficient time for group discussions and activities (afternoon session) in addition to a knowledge-sharing and discussion in the form of a keynote, paper presentations, and an expert panel discussion (morning session).

The tentative workshop schedule is detailed in Table 1. Since one of our goals is to synthesize knowledge and expertise from various communities and spur impactful future research, the workshop will dedicate sufficient time for group discussions and activities (afternoon session) in addition to a knowledge-sharing and discussion in the form of a keynote, paper presentations, and an expert panel discussion (morning session). This will help connect participants that share similar interests and provide them with a chance to contribute and learn.

The morning session will begin with a keynote by Professor John Lee from the University of Wisconsin-Madison.[2] His immense expertise in appropriate trust from the psychology and human factors perspective will attract multi-disciplinary participants and spark interest and motivation in later workshop events. Participants will have the opportunity to share their accepted work with either paper or poster presentations. The morning session will include two paper sessions for authors to share their accepted work; presentations will consist of 1-2 minute lightening talks (or 3-5 minutes, depending on the numbers of accepted papers), followed by a joint (around 10 minute) Q&A session. Concurrent with the planned coffee breaks (in the morning and afternoon sessions), we also plan to hold poster presentations to support more interactions between authors and participants. We will also host a discussion panel of experts from the organizing and program committees, to form the discussion around our diverse research interests—trust calibration, human-AI teaming, understanding AI uncertainty, etc.

In the afternoon, we plan to allocate adequate time for two sessions of in-depth group activities, each session containing an one-hour within-group discussion, and a half-hour between-grouping insight sharing. The groups will be in the form of "birds-of-feather" discussions and practices around several topics, including measures, challenges, and mechanisms and interactions for shaping trust. We will finalize the group activities based on the number of participants and their interests, but some initial ideas include,

(1) a *debating format*, where two groups are paired to represent the claims and counterclaims relating to themes within human-AI trust, so to motivate people to play devil's advocates to each others ideas. The organizers would provide inspirational questions, as well as imaginary use scenarios that can ground these discussion (e.g., in high-stake domains like education, medical, etc.)
(2) *concept mapping* around definitions, measures, and factors for appropriate trust and reliance. Groups' could collaborate to enumerate and discuss relevant concepts—predictability, uncertainty, trust, reliance, adaptability, etc. and identify their overlaps and relationships.
(3) *ideation* for solutions for shaping trust for particular use cases, such as news recommender systems or autonomous vehicles; here groups can brainstorm possible solutions, including system interactions, explanations, or visualizations, and iterate on these ideas with input from other groups. Outputs of this activity might consist of a set of solution ideas or low-fidelity mockups for system designs.
(4) *on-the-spot paper writing and reviewing*: where participants come up with one research idea, or an imaginary paper they would like to write around the topic of trust and reliance. Groups' output might be an abstract, certain teaser figures illustrating the core idea, or compelling use cases. Then, participants will review these deliveries, hopefully to help better articulate what aspects people would care about around a particular research idea.

Each group will be moderated by at least one organizer; we will also encourage paper authors to join groups related to their paper topics and share their posters within the group, so they can have more dedicated discussions around the broader theme, but in the context of their own work. For hybrid participation, activities will make use of collaborative virtual environments like Google Documents and Miro boards.

We will host the paper lightening talks, posters, and group sharings in Google Slides and on the website, and we will later convert them into medium posts to share with the broader audience.

[2] https://csl.engr.wisc.edu/labgroupresearch.html

7 POST-WORKSHOP PLAN

The workshop outcomes will be synthesized as a poster to be presented at CHI 2022. In order to reach a larger audience, the recorded sessions will be uploaded on YouTube and the group activity outcomes will be published as Medium posts, all of which will be shared via social media.

8 CALL FOR PARTICIPATION: WORKSHOP ON TRUST AND RELIANCE IN AI-HUMAN TEAMS (TRAIT)

As humans increasingly interact with AI systems during decision-making, creative tasks, and other workflows, appropriate trust and reliance are necessary to ensure proper usage and adoption of these systems. For example, people should understand when to trust or rely on an algorithm's outputs and when to override them. While significant research focus has aimed to measure and promote trust in human-AI interaction, the field lacks synthesized results across contexts, formalized key concepts, and definitions. The workshop will provide a venue to explore three broad aspects related to human-AI trust: (1) How do we clarify definitions and frameworks relevant to human-AI trust and reliance (e.g., what does trust mean in different contexts)? (2) How do we measure trust and reliance? And, (3) How do we shape trust and reliance? Themes include, but are not limited to:

- Definitions of trust and reliance.
- Human-human trust and lessons from social sciences.
- Qualitative (e.g., user reflection) and quantitative methods (e.g., usage, adoption, team performance, etc.) for evaluating trust and reliance.
- Tradeoffs with other objectives (e.g., team performance, creativity, etc)
- Solutions (and their limitations) for promoting appropriate trust (e.g., XAI, control mechanisms, human agency, communicating uncertainty etc).
- Safety mechanisms for when trust is broken.

Using 2-6 pages and SIGCHI format, authors may present a new position, summarize existing research, provide industry perspective, or in-progress works. At least one author must register and attend the workshop. Submission will be reviewed by program committee and accepted papers will be posted on the workshop website and shared via social media.The workshop will feature keynote by Prof. John Lee (UWisc); moderated panel of diverse experts; short talks or posters by authors; and small group activities to explore workshop themes and promote collaborations

Important Dates:

- Submission: February 11, 2022 (Easychair)
- Notifications: March 11, 2022
- Workshop: TBD (between April 14-15 or April 30-May 6)

REFERENCES

[1] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 2021. Does the whole exceed its parts? the effect of ai explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–16.

[2] Adrian Bussone, Simone Stumpf, and Dympna O'Sullivan. 2015. The role of explanations on trust and reliance in clinical decision support systems. In *2015 international conference on healthcare informatics.* IEEE, 160–169.

[3] Erin K Chiou and John D Lee. 2021. Trusting Automation: Designing for Responsivity and Resilience. *Human Factors* (2021), 00187208211009995.

[4] Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten Van 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 (2008), 455.

[5] Maria De-Arteaga, Riccardo Fogliato, and Alexandra Chouldechova. 2020. A case for humans-in-the-loop: Decisions in the presence of erroneous algorithmic scores. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–12.

[6] Upol Ehsan, Q Vera Liao, Michael Muller, Mark O Riedl, and Justin D Weisz. 2021. Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–19.

[7] Upol Ehsan and Mark O Riedl. 2020. Human-centered explainable ai: Towards a reflective sociotechnical approach. In *International Conference on Human-Computer Interaction.* Springer, 449–466.

[8] Ella Glikson and Anita Williams Woolley. 2020. Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals* 14, 2 (2020), 627–660.

[9] Kenneth Holstein and Vincent Aleven. 2021. Designing for human-AI complementarity in K-12 education. *AI Magazine* (2021).

[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] René F Kizilcec. 2016. How much information? Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* 2390–2395.

[12] Lea Krause and Piek Vossen. 2020. When to explain: Identifying explanation triggers in human-agent interaction. In *2nd Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence.* 55–60.

[13] Himabindu Lakkaraju and Osbert Bastani. 2020. " How do I fool you?" Manipulating User Trust via Misleading Black Box Explanations. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.* 79–85.

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

[15] Min Kyung Lee and Su Baykal. 2017. Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division. In *Proceedings of the 2017 acm conference on computer supported cooperative work and social computing.* 1035–1048.

[16] Ariel Levy, Monica Agrawal, Arvind Satyanarayan, and David Sontag. 2021. Assessing the Impact of Automated Suggestions on Decision Making: Domain Experts Mediate Model Errors but Take Less Initiative. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–13.

[17] Brian Y Lim and Anind K Dey. 2011. Investigating intelligibility for uncertain context-aware applications. In *Proceedings of the 13th international conference on Ubiquitous computing.* 415–424.

[18] Zhuoran Lu and Ming Yin. 2021. 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.* 1–16.

[19] Bonnie M Muir. 1987. Trust between humans and machines, and the design of decision aids. *International journal of man-machine studies* 27, 5-6 (1987), 527–539.

[20] Forough Poursabzi-Sangdeh, Daniel G Goldstein, Jake M Hofman, Jennifer Wortman Wortman Vaughan, and Hanna Wallach. 2021. Manipulating and measuring model interpretability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–52.

[21] Vanessa Sauer, Alexander Mertens, Jens Heitland, and Verena Nitsch. 2021. Designing for Trust and Well-being: Identifying Design Features of Highly Automated Vehicles. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–7.

[22] Dakuo Wang, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. 2020. From human-human collaboration to human-ai collaboration: Designing ai systems that can work together with people. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–6.

[23] David Gray Widder, Laura Dabbish, James D Herbsleb, Alexandra Holloway, and Scott Davidoff. 2021. Trust in Collaborative Automation in High Stakes Software Engineering Work: A Case Study at NASA. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–13.

[24] Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable AI: Fitting intelligent decision support into critical, clinical decision-making processes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* 1–11.

[25] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. 2019. Understanding the effect of accuracy on trust in machine learning models. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–12.

[26] Yunfeng Zhang, Q Vera Liao, and Rachel KE 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*. 295–305.