A Balancing Act: Designing User-Centered AI-based DSS to Address Overreliance and Distrust in System Outputs

Natalia Echeverry*
nae81@pitt.edu
University of Pittsburgh
Pittsburgh, Pennsylvania, USA

Abstract

Public institutions increasingly rely on AI-based Decision Support Systems (DSS) to transform data into actionable insights. Examples include the Allegheny Family Screening Tool (AFST) in child welfare and Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) in criminal justice. While these tools demonstrate the transformative potential of predictive analytics, their effectiveness is often undermined by practical implementation challenges. In tools such as AFST and COMPAS, a common source of challenges is tool misuse, including overreliance, diffidence, and distrust in tool recommendations. This study uses the AFST as a case study to explore approaches to designing user interfaces for AI-based DSS using predictive risk modeling (PRM). It focuses on how elements such as risk score and error rate presentation influence users' trust and ability to use the tool for effective decision-making, especially among those with low statistical literacy. Drawing on cognitive science, psychology, communication, and HCI literature, it explores ways to enhance cooperation between users and AI-based DSS. First, using the AFST user interface as a case study, the study implements a between-subject experiment with 225 participants to compare two types of risk representations: gauge scales and icon arrays. While risk visualizations do not significantly alter decision outcomes; however, they affect users' perceptions of and confidence in the system. These findings underscore the need for designing user-centered interfaces that help users balance trust and skepticism, preventing overreliance on risk scores while supporting informed decision-making. This research highlights the importance of user interface design and risk representations in shaping users' perceptions of the system. The results also demonstrate the importance of considering socio-cognitive factors —such as beliefs, behaviors, biases, and misconceptions— that affect the user-system interaction.

CCS Concepts

• **Human-centered computing** → **Graphical user interfaces**; • **Information systems** → **Decision support systems**.

Keywords

AI-based Decision Support Systems, Cognitive Bias, Risk Visualization, User Interface Design, Automation Bias, Confirmation Bias, Decision-Making, User Experience, Statistical Literacy

ACM Reference Format:
Natalia Echeverry. 2025. A Balancing Act: Designing User-Centered AI-based DSS to Address Overreliance and Distrust in System Outputs. In *Proceedings of ACM Conference on Intelligent User Interfaces (ACM IUI) 2025 (IUI '25).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

IUI '25, March 24–27, 2025, Cagliari, Italy

ACM ISBN XXXX
https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Background and context

Public institutions adopting data-driven approaches increasingly rely on Decision-Support Systems (DSSs) with predictive analytics to transform vast amounts of data into actionable insights. The adoption of AI-based DSSs by public institutions reflects this shift, with risk models being implemented in areas like social services and the criminal justice system. Prominent examples include the Allegheny Family Screening Tool (AFST) in child welfare and the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) in criminal justice, with additional examples in healthcare and other contexts. These tools demonstrate the transformative potential of predictive analytics, but their effectiveness is often undermined by challenges associated with their implementation in real-world contexts. In healthcare, the lack of HCI consideration for how the tool fits in the user's workflow and the nature of the job independent of the tool has been listed as the main cause of DSS failures [22].

Historically, DSSs with predictive analytics were primarily used by individuals with advanced quantitative and statistical skills. Today, these systems are increasingly used by individuals with varying levels of expertise, requiring careful consideration of how end users engage with AI-based DSSs in decision-making. In the case of AFST and COMPAS, most implementation challenges were associated with a limited understanding of how the AI-based DSS works and low statistical literacy among end-users—social workers and judges—. These challenges stem from end-users lack of entering skills needed to interact with the tool, for example, limited understanding of predictive analytics and statistical concepts, such as uncertainty and probabilistic reasoning, which are non-intuitive for individuals without formal training. Addressing these knowledge gaps is essential to ensure the adequate use of these technologies and to foster user trust and informed decision-making in critical domains like child welfare and criminal justice among others.

2 Related work

Increasing our understanding of how users interact with and interpret information from predictive analytics is essential for "complementary computing" [14], which considers combining human and machine intelligence to be more effective than relying on humans or AI alone. This aligns with theories of collaborative decision-making and human-computer interaction (HCI), which aims to design systems that support and enhance human judgment, rather than replacing it entirely. While ensuring the fairness, reliability, and accuracy of the predictive risk models is critical, successful integration of AI-based DSSs also requires consideration of human factors that influence how users interact with these systems [8], [9], including cognitive and behavioral patterns identified in risk and decision-making analysis and HCI. AI-assisted decision-making introduces cognitive biases, such as automation, confirmation, and anchoring biases, which are well-documented in cognitive psychology and have significant implications for how users interpret and act on risk scores.

2.0.1 ***Cognitive Biases.*** Predictive analytics in AI-based DSSs introduces new challenges to how information is presented to users. More research is needed to find design patterns that (i) assist users in decision-making effectively and (ii) mitigate cognitive biases. Achieving this may also involve additional measures such as coupling the interface design with sufficient user training and help.

[13] found that users incur misconceptions and biases when reading generic risk representations they do not understand. [8], found that AI explainability did not improve human performance and that providing user feedback for human decision calibration decreased their accuracy. [15], [1], [22]] provide additional background information for this research work. They all demonstrate that risk communication in the context of predictive analytics and AI-based DSS needs to be further studied.

Cognitive biases are distorted perceptions and interpretations of objective information that can result in distortions or inaccuracies of judgment [15]. Cognitive biases identified in the context of decision-making include anchoring bias [4], [17], confirmation bias [18], and automation bias [19], [16]. These biases are usually correlated with a lack of understanding of the tool which in turn results in interpretation errors such as treating risk probability percentages on a scale of zero to one hundred percent as a binary score [8], among others.

While these biases and interpretation errors have been identified and documented in the literature, this research focuses on automation bias which is defined as overreliance on automated systems when dealing with conflicting information [19], [16]. Drawing on the trust theory [3], the cognitive load theory [3], and the risk communication theory [12] this research explores interface design and risk representation patterns that mitigate overreliance on the decision support system's outputs.

2.0.2 ***Automation bias.*** [10], [11] contextual inquiries with users of the AFST showed that users struggled with integrating statistical predictions in their decision workflows resulting in overreliance on the tool's risk classification. [7] found that in early implementations of AFST users were "(...) less likely to adhere to the machine's recommendation when the score displayed is an incorrect estimate of risk (...)". These contrasting results motivate an exploration of the possible effects that the user interface design and the risk representation of AI-based DSS have on users' perceptions of the system.

2.0.3 ***Mistrust.*** As biases and interpretation errors, negative perceptions such as mistrust can be detrimental to the successful implementation of the AI-based DSS. Explanatory approaches, communicating uncertainty and model limitations are forms of transparency that may encourage users to trust the tool [11], [20], [2], [5]. Designing AI-based DSS that enhances user-system cooperation in decision-making demands a better understanding of users.

[22] noted that "the interaction design of most clinical decision support tools instead assumes that individual clinicians will recognize when they need help, walk up and use a system that is separate from the electronic health record and that they want and will trust the system's output." Considering human factors such as users' awareness of the tool, misconceptions, beliefs, and trust brings an opportunity to integrate interdisciplinary approaches from cognitive, psychological, and HCI literature.

As noted by [3] "without establishing trustworthy relationships, these new infrastructures and services, these new artificial agents, these new robots, these new pervasive technologies, do not impact with sufficient strength and in fact of these do not really integrate with the real society" [3]. They also pointed out that "Technology should not only be reliable, safe, secure, but it should be also perceived as such, the user must believe that it is reliable, and must feel confident while using it and depending on it" [3].

2.1 Research Questions and Variables

This study explores the effect that interface design and risk representation patterns of AI-based DSS have on users with low statistical skills decision outputs and perceptions of the tool. The research questions are as follows:

- **RQ1:** Does the interface design and risk representation in AI-based DSS affect users' decisions?
- **RQ2:** Does the interface design and risk representation in AI-based DSS influence users' trust in the system?

The decisions are the dependent variable while the independent variables are two different risk representations tested in this study.

2.2 Experimental Setup and Participants

This study uses the Allegheny Family Screening Tool (AFST) as a case study to test the research questions. The AFST is a predictive risk model (PMR), the user interface displays a risk score over a gauge scale with "Lower Risk", "Medium Risk", and "Higher Risk" labels (Figure 1). I built four high-fidelity wireframes of an AI-based DSS modeled after the AFST [2]. The gauge scale is the control group while an icon array is used in the experiment group.

AI-based DSS relies on risk predictive models which are integrated into a user interface that includes the risk representation in numerical or verbal form. As a probability percentage, ratio, single- or double-digit number, or as a text label. Whether the risk is communicated as a number or a text label, it could be supplemented by a graphical representation such as charts, linear and round gauge scales, and icon arrays among others [11]. Additional information

about the tool. Examples of this include the tool's error rate and a disclaimer with the tool's limitations.

Participants Call screeners of child protection hotlines usually have completed a bachelor's degree in social work or related fields and have some experience working with children and families. Because of their educational and professional background, it is uncommon for them to have substantial training in statistics. Given that the purpose of the study is to explore effective ways of communicating risk predictions to users with low and medium statistical skills, paired with the difficulty of working with call screeners directly, individuals older than 18 years old, who speak English, and have completed some years of college or above are accepted as surrogate participants. A background in social work or a related field is preferred.

2.2.1 Online survey design and questionnaires. For this study, I administered an anonymous online survey with three main parts: (i) demographic, educational, and professional background questions, (ii) a simulation of a call referral and case classification modeled after Allegheny County's child protection program, and (iii) a set of system usability Likert-scale questions.

Four possible risk communication strategies were created for this study: gauge scale with a text label, gauge scale with a percentage, icon array with a text label, and icon array with a percentage.

Simulation of a call referral and case classification with FST After participants complete the first part of the survey, they are prompted with a short text describing a child protection program that requires call screeners to utilize an AI-based DSS when deciding how to classify a referral. The prediction target and error rate of the FST are communicated to participants. Participants are instructed to situate themselves in the role of a call screener.

"The Department of Human Services manages a child protection hotline where people can report alleged cases of child abuse and neglect, also called referrals. The hotline is operated by call screeners who are required to collect information over the phone and utilize a risk assessment tool called the Family Screening Tool (FST). The FST is an AI-based tool that predicts the risk of an unfavorable outcome occurring to the child in the next two years. Call screeners are required to use FST and consider the risk level provided by the tool when deciding if there is a need for an intervention. The error rate measures the percentage of times the FST makes an incorrect risk prediction. The error rate of the Family Screening Tool is 5.5%-symbol. This survey will ask you to assume the role of a call screener in the following case scenario." Then, they were prompted with a case scenario.

After participants had read the scenario, they were randomly assigned a wireframe of FSTs with one of the risk communication strategies (Figure 1). All risk communication strategies used a 57.5 risk score on a scale of 0 to 100. The case scenario does not present a clear-cut situation and the risk score is slightly above the midpoint. The error rate given was 5.5%-symbol These conditions aimed at encouraging participants to consider both the case and the risk score in their decision-making process. Then, they are asked to classify the case as low, medium, or high risk. This portion of the survey was presented as a Likert question. After participants selected a classification for the case, the survey asked them to type down the reason for the classification they chose.

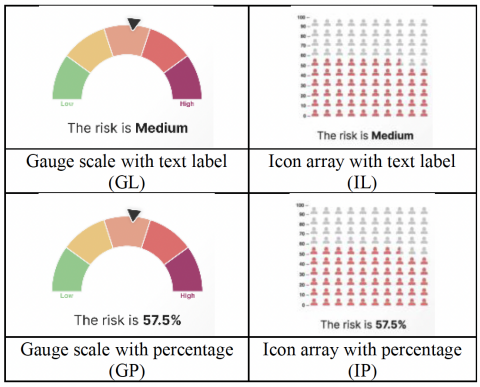

Figure 1: Risk representation strategies/widgets

Risk communication strategies: verbal versus numeric risk descriptions and graphical representations. [6] noted the limitation of relying upon a verbal description exclusively because of the difficulty of mapping it on a numeric scale. Providing a graphical representation of the risk output in addition to a verbal or numeric risk description is considered a good practice [6]. However, graphical representations can be counterproductive if they are not selected according to the users' familiarity with them [9]. Gauge scale and icon array are the two graphical representations selected for this study. Gauge scales are pervasive in visual analytics, when they are used to communicate risk, they usually follow the speedometer metaphor that starts in low (green) and ends in high (red). The icon array is also used to represent risk graphically. Icon arrays are considered the best option for risk communication because they allow for a precise "discrete representation of risk" [21].

3 Data Collection and Analysis Methods

3.0.1 Survey distribution and participant demographics. Participants were offered a $5 Amazon e-gift card as compensation for finishing the online survey. Of the 493 responses received, 225 passed the minimum acceptance criteria. The age range mode is 25-34 years old followed by 35-44 years old. More than 60%-symbol of participants selected Graduated 4-year College as their highest level of education followed by Postgraduate with 20%-symbol and Graduated 2-year College with 15%-symbol. All participants in this sample self-reported that they have formal training or experience in social work or a related field.

3.0.2 Data sampling. Social work programs typically include statistical training in their curriculum. The extent of statistical training can vary depending on the program. College-level programs may include introductory courses on statistics and research methods, while post-graduate programs typically have a stronger emphasis on statistics, including courses on descriptive and inferential statistics. Given that training in predictive statistics is not typically part of the curriculum even at the post-graduate level programs, participants in all education groups will be included in this study.

Participants were randomly assigned to one of the four risk communication strategies (Figure 1). The risk prediction in the FST was the same for all four groups, 57.5 on a scale from 0 to 100.

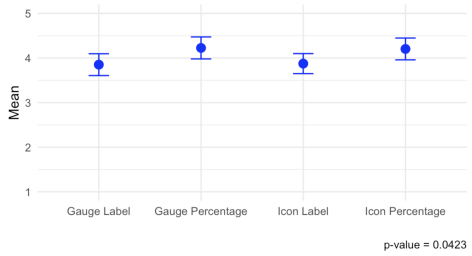

Figure 2: "Overall, I consider the FST to be useful for the evaluation of child referrals." (Strongly agree = 5, Agree = 4, Neutral = 3, Disagree = 2, Strongly disagree 1)

The only difference between groups was the risk communication strategy used in the FST.

3.0.3 Summary of Risk Classifications. The distribution of the risk classification in all groups was the same or similar for all groups. The mean for gauge with label, gauge with percentage and icon array with percentage groups is 2.4 with a standard deviation of 0.5. The icon array with text label group had a slightly different mean of 2.3 and a standard deviation of 0.6 presumably because of a bigger sample size.

4 Results

4.0.1 Impact of Risk Representation on Decision Outcomes. I applied one-way ANOVA to test the null hypothesis that the mean risk classification of all four groups is the same. The test indicates that there is no statistical evidence to reject the null hypothesis (p-value=0.5). The risk communication strategy does not seem to influence the decision the user makes. The mean risk classification across groups is statistically the same. However, this may be due to a small sample size (n=225) and data quality issues. Participants of the online anonymous survey may have had the incentive to self-report characteristics that are not accurate because of the offer to receive a reward.

4.0.2 Impact of Risk Representation on User's Perceptions of the Tool. The survey prompted participants with Likert-scale questions about their agreement with statements about the FST after they had classified the case and typed their responses. The one-way ANOVA tests for each statement did not find any statistically significant difference between groups except for the final statement: "Overall, I consider the FST to be useful for the evaluation of child referrals".

Participants in the gauge with text label and icon array with text label rated the FST lower in the Likert-scale with means of 3.8 and 3.9 while the gauge with percentage and icon array with percentage rated the FST higher both with a mean of 4.2. The p-value of the one-way ANOVA test is 0.04 which shows that there is statistical evidence to infer that the risk communication strategy may influence users' trust in the usefulness of the tool. A possible explanation for the differences between groups is that users who were administered the risk communication strategies with a percentage perceived the tool, as more specialized and precise. A similar grouping appears in the one-way ANOVA test for the statement "All the information provided by the FST is useful and relevant." where groups gauge scale with percentage and icon array with percentage perceived the information in the FST as more useful and relevant compared to the other two groups. With a p-value of 0.1, there is no statistical evidence to consider that the means are different. However, these results show similitudes between groups based on how the risk score is communicated. The graphical representation does not seem to have as much weight as the format of the risk score communication, whether verbal or numeric.

Despite the lack of statistical significance, the gauge scale with percentage group had the highest mean rate compared to the other groups consistently in most Likert-scale statements such as: "The FST provided sufficient information", "The information in the FST is straightforward", and "I would not need the support of a supervisor or colleague to understand FST results". The "Overall, I consider the FST to be useful for the evaluation of child referrals" shared the mean of 4.2 with the icon array with a percentage group. These statements have to do with users' trust in the statistical prediction given by the system and users' confidence in their capability of using FST to make decisions.

4.0.3 Participant's written responses. Participants in the icon array percentage share similarities with the gauge percentage group in terms of risk classification and explanations. P22 from the icon array percentage group and P69 from the gauge percentage group classified the case as high risk and wrote:

- P22: "The risk rate is above average"
- P69: "It's more than half"

5 Conclusion and Future Work

The risk communication strategies did not influence participants' decision outcomes but they influenced their perceptions of the usefulness of the system. Participants in the gauge scale with percentage and icon array with percentage rated higher on the statement referring to the usefulness of the tool. However, while the study demonstrated that the user design interface and risk representation affect users' perceptions of the tool, it is futile to propose user interface and risk representations without modeling users' beliefs and misunderstandings. The findings are not conclusive about the type of risk representation or user interface that contributes to better user-system cooperation therefore more research is needed to model users' trust, beliefs, and misconceptions in multiple dimensions, considering the complexity of their work, collaborations, job workflow and the AI-based DSS itself in the particular context. The *cognitive trustor model* [3] and *mental models approach* [12] are promising frameworks to aid designers in these explorations.

Conducting online or in-person interviews and simulations with think-aloud protocols would help understand better how users interact and interpret statistical predictions. Interviews would also allow an in-depth understanding of users' quantitative and statistical skills, and their perceptions of such systems. Moreover, models from trust theory and risk communication, such as the *cognitive trustor model* [3] and *mental models approach* [12], can help clarify users' needs within a specific implementation context.

6 Citations and Bibliographies

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* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 81, 16 pages. https://doi.org/10.1145/3411764.3445717

[2] Umang Bhatt, Javier Antorán, Yunfeng Zhang, Q. Vera Liao, Prasanna Sattigeri, Riccardo Fogliato, Gabrielle Melançon, Ranganath Krishnan, Jason Stanley, Omesh Tickoo, Lama Nachman, Rumi Chunara, Madhulika Srikumar, Adrian Weller, and Alice Xiang. 2021. Uncertainty as a Form of Transparency: Measuring, Communicating, and Using Uncertainty. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (Virtual Event, USA) *(AIES '21)*. Association for Computing Machinery, New York, NY, USA, 401–413. https://doi.org/10.1145/3461702.3462571

[3] Christiano Castelfranchi and Rino Falcone. 2010. *Trust Theory*. John Wiley Sons Ltd.

[4] Isaac Cho, Ryan Wesslen, Alireza Karduni, Sashank Santhanam, Samira Shaikh, and Wenwen Dou. 2017. The Anchoring Effect in Decision-Making with Visual Analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*. 116–126. https://doi.org/10.1109/VAST.2017.8585665

[5] Valdemar Danry, Pat Pataranutaporn, Yaoli Mao, and Pattie Maes. 2023. Don't Just Tell Me, Ask Me: AI Systems that Intelligently Frame Explanations as Questions Improve Human Logical Discernment Accuracy over Causal AI explanations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 352, 13 pages. https://doi.org/10.1145/3544548.3580672

[6] Mohammad Daradkeh, Clare Churcher, and Alan McKinnon. 2013. Supporting informed decision-making under uncertainty and risk through interactive visualisation. In *Proceedings of the Fourteenth Australasian User Interface Conference - Volume 139* (Melbourne, Australia) *(AUIC '13)*. Australian Computer Society, Inc., AUS, 23–32.

[7] 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* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376638

[8] Ben Green and Yiling Chen. 2019. The Principles and Limits of Algorithm-in-the-Loop Decision Making. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 50 (Nov. 2019), 24 pages. https://doi.org/10.1145/3359152

[9] Naja Holten Møller, Irina Shklovski, and Thomas T. Hildebrandt. 2020. Shifting Concepts of Value: Designing Algorithmic Decision-Support Systems for Public Services. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (Tallinn, Estonia) *(NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, Article 70, 12 pages. https://doi.org/10.1145/3419249.3420149

[10] Anna Kawakami, Venkatesh Sivaraman, Hao-Fei Cheng, Logan Stapleton, Yanghuidi Cheng, Diana Qing, Adam Perer, Zhiwei Steven Wu, Haiyi Zhu, and Kenneth Holstein. 2022. Improving Human-AI Partnerships in Child Welfare: Understanding Worker Practices, Challenges, and Desires for Algorithmic Decision Support. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 52, 18 pages. https://doi.org/10.1145/3491102.3517439

[11] Anna Kawakami, Venkatesh Sivaraman, Logan Stapleton, Hao-Fei Cheng, Adam Perer, Zhiwei Steven Wu, Haiyi Zhu, and Kenneth Holstein. 2022. "Why Do I Care What's Similar?" Probing Challenges in AI-Assisted Child Welfare Decision-Making through Worker-AI Interface Design Concepts. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 454–470. https://doi.org/10.1145/3532106.3533556

[12] M. Granger Morgan, Baruch Fischhoff, Ann Bostrom, and Cynthia J. Atman. 2002. *Risk Communication: A Mental Models Approach*. Cambridge University Press.

[13] Mahsan Nourani, Chiradeep Roy, Jeremy E Block, Donald R Honeycutt, Tahrima Rahman, Eric Ragan, and Vibhav Gogate. 2021. Anchoring Bias Affects Mental Model Formation and User Reliance in Explainable AI Systems. In *Proceedings of the 26th International Conference on Intelligent User Interfaces* (College Station, TX, USA) *(IUI '21)*. Association for Computing Machinery, New York, NY, USA, 340–350. https://doi.org/10.1145/3397481.3450639

[14] Joon Sung Park, Rick Barber, Alex Kirlik, and Karrie Karahalios. 2019. A Slow Algorithm Improves Users' Assessments of the Algorithm's Accuracy. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 102 (Nov. 2019), 15 pages. https://doi.org/10.1145/3359204

[15] Charvi Rastogi, Yunfeng Zhang, Dennis Wei, Kush R. Varshney, Amit Dhurandhar, and Richard Tomsett. 2022. Deciding Fast and Slow: The Role of Cognitive Biases in AI-assisted Decision-making. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 83 (April 2022), 22 pages. https://doi.org/10.1145/3512930

[16] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M. Howard, and Alan R. Wagner. 2016. Overtrust of robots in emergency evacuation scenarios. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 101–108. https://doi.org/10.1109/HRI.2016.7451740

[17] Spencer Soper. 2021. Fired by Bot at Amazon: 'It's You Against the Machine'. (2021). https://www.bloomberg.com/news/features/2021-06-28/fired-by-bot-amazon-turns-to-machine-managers-and-workers-are-losing-out

[18] Megha Srivastava, Hoda Heidari, and Andreas Krause. 2019. Mathematical Notions vs. Human Perception of Fairness: A Descriptive Approach to Fairness for Machine Learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Anchorage, AK, USA) *(KDD '19)*. Association for Computing Machinery, New York, NY, USA, 2459–2468. https://doi.org/10.1145/3292500.3330664

[19] Alan R. Wagner, Jason Borenstein, and Ayanna Howard. 2018. Overtrust in the robotic age. *Commun. ACM* 61, 9 (Aug. 2018), 22–24. https://doi.org/10.1145/3241365

[20] Ruotong Wang, F. Maxwell Harper, and Haiyi Zhu. 2020. Factors Influencing Perceived Fairness in Algorithmic Decision-Making: Algorithm Outcomes, Development Procedures, and Individual Differences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376813

[21] Allison Woodruff, Sarah E. Fox, Steven Rousso-Schindler, and Jeffrey Warshaw. 2018. A Qualitative Exploration of Perceptions of Algorithmic Fairness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174230

[22] 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300468

Received December 2024; revised TBD; accepted TBD