Shadows of The Past: The Effects of User's Past Experiences and Expectations on Human-AI Partnership

Mahsan Nourani*
University of Florida

ABSTRACT

With Artificial Intelligence (AI) models being broadly integrated in decision-support and analytical systems, designing effective tools wary of their target users and stakeholders is essential. People from different backgrounds encounter interactive intelligent systems on a daily basis, in situations that necessitates human-AI partnership for the analysis and decision-making. In my dissertation, I will investigate people's prior experiences and how they affect human-AI collaborations. I will start by investigating *which* past experiences affect human-AI collaborations. Through an initial analysis of empirical studies, I organized a preliminary conceptual model of users' past experiences and factors that influence human-AI collaborations, organized as *long-term* past, *short-term* past, and *present*. Using *anchoring bias* as a form of short-term and *domain expertise* to represent long-term past experiences, I demonstrate two empirical user studies, contributing significant evidence that past experiences spawn differences in usage behaviors and perceptions. To further explore the effects of individual experiences on usage differences, I plan to create (and later design a user study to test the effectiveness of) a proof-of-concept questionnaire to capture how previously formed expectations of AI affect behaviors towards AI systems. I also propose a systematic literature review in the fields of human-centred AI to support and improve the preliminary conceptual model of user's past behaviors.

Index Terms: Human-centered computing—Visualization—Visualization techniques—Treemaps; Human-centered computing—Visualization—Visualization design and evaluation methods

1 INTRODUCTION

Over the last two decades, Machine Learning and Artificial Intelligence (ML/AI) approaches have fundamentally altered users' analytical reasoning and decision-making by making computing systems more intelligent. In fact, a large number of existing ML/AI systems have achieved relative autonomy in that they can decide and act on their own with minimal human intervention. While many AI-powered technology automates outcomes for people's needs, many require people as agents to monitor, interact, analyze the outcomes, and *collaboratively* make decisions together with the AI. In many systems, *partial autonomy* is preferred over full autonomy for cases when the final decision is significant and/or the user's expertise is required to evaluate the predictions and reason analytically [48]. This motivates the need for mixed-initiative systems [15] and collaborative human-AI systems, where the goal is to exploit the merits of both human and the machine for better decision making (also known as AI-assisted decision making [48] or data-driven decision-making for visual analytics systems [2]).

With such systems, one main challenge is that many types of human differences may lead to distinct usage behaviors, influencing the outcomes of human and AI collaborative efforts. People's unique personal experiences shapes their understanding of technology and their views of the world. This applies to our context of interactive intelligent systems as well, where individual differences are partially rooted in their mental model of AI as a concept and the technologies/devices supported by it.

Studying these differences is crucial given the complex nature of human minds; people think and reason differently based on their unique past experiences, backgrounds, demographic differences, existing biases, and impressions of AI and technology. The implied impressions of AI may lead to divergence in interactions, perceptions, utilization, and understanding of AI-powered tools and models. Ultimately, these assorted behaviors can take place when people of various backgrounds partner up with AI systems to make high-stakes decisions. The body of work in the field of Human-centered AI (HCAI) consists of numerous work investigating *how* certain attributes originated from past experiences affect people's usage behaviors. For instance, researchers have studied how knowledge of AI or domain affects their perceptions of the system and their trust [3, 9, 26]. Investigating the *how* question is critical for improving intelligent systems while keeping their users and stakeholders in mind. However, little is known about *what* these different attributes are in the first place. Identifying *which* prior experiences associate with individual usage differences can not only improve our search for which *how*s to investigate, but can also motivate finding techniques to circumvent imposed challenges or augment behaviors that are caused by people's differences. Ultimately, we can find solutions for leveraging people's differences in building tools that satisfy their needs and facilitate their usage of such AI systems.

Motivated by these challenges, my PhD dissertation research will focus on studying the impact of user past differences in interactive human-AI systems to inform their improved design and implementation. Furthermore, this research will contribute to and is critical in building responsible and ethical AI systems. In particular, this work is motivated by the following Guiding Research Questions (GRQs):

- **GRQ1**: What past experiences lead to differences with current usage behaviors with intelligent systems that rely on human-AI partnerships?
- **GRQ2**: How do these individual past experiences affect human-AI collaborations?
- **GRQ3**: How does identifying differences spawned from past experiences help with predicting usage behaviors and the effectiveness of collaborative decision-making between humans and AI?

I will explore these questions with studies and literature review analysis through a human-centered perspective. In this manuscript, I will provide an overview of my proposed work (completed and future work) to cover these guiding research questions.

2 RELATED WORK

A variety of different research communities have studied the design of user-centered AI and eXplainable AI (XAI) systems. The goal is to design while being mindful of users and their needs. In the visualization community, researchers have focused on building and

*Mahsan Nourani is a PhD Candidate advised by Dr. Eric D. Ragan at the University of Florida. She can be contacted via mahsannourani@ufl.edu.

designing tools for XAI systems in an attempt to improve understanding [34], analytical process [30, 46], debugging [35, 46], and fairness [2]. Work in the HCI community has resulted in user-centered design guidelines and frameworks for XAI systems [10, 41, 42], while others have compared and evaluated design variations for explanations to understand how they can improve user experience in an XAI system [17, 20]. Some HCI researchers focused on building empirical knowledge around explainable intelligent systems by studying user behaviors; for instance, how users form and calibrate mental models of intelligent systems.

In the context of intelligent systems, mental models refer to a user's map of how a system works based on their interactions. Some researchers focused on studying the effects of transparency on user mental models. For instance, Kulesza et al. [18] studied how explanation soundness and completeness affect user mental model. Empowering users to build more accurate mental models is critical for various reasons, one of the common goals being to improve user-machine trust [12, 28, 43], as humans might find a hard time trusting what they do not understand. While transparency is known as another approach to improve understanding and the fidelity of formed mental models [23], previous research has shown other factors, such as user's cognitive biases, can affect the quality of these formed mental models, despite the presence of explanations [28].

Semi-automated intelligent systems heavily rely on human analysis and sense-making for the final decision or analysis. Human judgments and decisions come from their complex brains, structure and inner-workings of which is yet to be fully understood. Human reasoning can be influenced by various conscious and subconscious factors, some of which are rooted in prior experiences and formed expectations of AI. Preexisting factors, such as prior knowledge, background culture, experiences, cognitive abilities, age, and expertise can influence human perceptions and cognition abilities [13, 21]. Researchers in HCI, HCAI, and CSCW communities have evaluated how factors of these natures affect people's usage behaviors and understanding of AI. For instance, Zhang et al. [47] studied how people perceive and form expectations of their AI teammates.

People's biases and cognitive heuristics are one of the more studied preexisting factors. Cognitive biases are a subcategory of these biases that are caused by erroneous decision-making heuristic [39] and may influence usage and human behaviours of intelligent systems. While such biases have been studied in various areas, HCI research attempts to acknowledge their existence and impacts when humans work with computing systems. Today, many known variations of cognitive biases exist [1]. In 2018, Dimara et al. [7] organized a taxonomy of 154 known cognitive biases in seven categories to make visualization designers aware when designing tools. In the context of human-centered AI, some of the most common studied biases include automation bias [26, 28, 41], confirmation bias [24], anchoring effect [26, 38], and availability heuristic [33]. Cognitive biases have been studied in various human-centered domains of interest to this review, including visualization and visual analytics [7, 40], decision-making [7, 39], and XAI [26, 28, 41]. Research on first impressions has shown that human's early observations and judgements can bias and affect their behaviours towards people [11, 44], systems [25], and/or agents [6, 31]. In the machine learning community, some researchers have focused on the effects of order of training data on model's performance accuracy [4, 45]. In visual analytics, Wall et al. [40] argue that heuristics such as anchoring effect can occur during the analytical process, causing the bias to propagate to the findings or reaching false conclusions.

In their recent book chapter, Vaughan and Wallach [37] discuss ML *intelligibility*, defined as the stakeholders' ability to monitor AI systems enough to achieve their tasks or goals. They argue that, since humans (be it those who implement or build such systems, or those who use it or are affected by their predictions) are the center of intelligent systems, the strategy to *intelligibility* should start from user needs. Here, I will also focus on the group of stakeholders who are directly using the system, i.e., the people who are among the target users. There are many ways of studying and categorizing the differences between stakeholders. Two examples include studying stakeholders based on their demographics and level of education/expertise; especially, knowledge in the task domain. As Hoff and Bashir [13] maintain, prior knowledge of the domain can influence user reliance and trust in decision-aids. Preexisting knowledge can also impact users' understanding of and ability to detect errors and understand the uncertainties of model predictions [26, 37].

In HCI, researchers have explored how domain expertise motivates effective system designs. For example, Schaffer et al. [32] found that those with less task familiarity are more likely to rely on explanations, often leading to cases of automation bias

Doshi-Velez and Kim [8] also discuss the importance of user domain expertise on model transparency. They maintain that the amount and purpose of explanations should differ based on user's level of familiarity of the task and domain and motivations for using the system. Additionally, some researchers studied the perceptual comparison of novice and expert users on how they analyze outputs of intelligent systems [36].

3 Proposed Work

People's individual differences are rooted in prior experiences and circumstances, including, but not limited to: their exposures to and perceptions of AI, their expertise and knowledge of AI and domain, and their cognitive and societal biases. These prior experiences affect human-AI collaborations and the outcomes and decisions achieved through this partnership.

3.1 Conceptual Model of Users' Past Experiences (WiP)

Before studying *how* these past experiences affect human-AI teaming, it is critical to understand *what* these past experiences[2] are, i.e., which elements of the past can impact this relationship.

While there has been prior work in the field to organize our understanding of human-AI teaming (e.g., [1, 14, 16, 19, 22]), there is no concise organization that characterizes human's prior experiences that influence people's behaviours with AI system. Such a taxonomy can serve as a road map for intelligent system engineers and designers to build AI systems mindful of people's backgrounds and differences (the 'who'). Furthermore, it helps raise their awareness and attention towards people's experiential differences and elevating those to improve the efficiency of human-AI teaming.

For these purposes, I have organized a preliminary conceptual model of users' prior experiences that describes such experiences through three lenses of *long-term* past, *short-term* past, and *present* and depicts the potential influences by outcomes and factors. The preliminary version of this conceptual model (as shown in Figure 1) is based on an initial literature analysis of empirical studies (including my own past work and that of others). The current organization of the conceptual model is based on the inner workings of human brain in capturing and maintaining memory over time; i.e., long-term, short-term, and working memory [5]. While a user's formed experiences of an intelligent system can be influenced by their current encounters with the AI outcomes (i.e., via both working and short-term memories), their long-term experiences of their past encounters with this AI systems or others can significantly impact their usage behaviors and the outcomes of collaborative decision-making (as noted by prior research [8, 26, 37]).

[1] List of Available Biases on Wikipedia–Accessed: [May 1, 2021]: https://en.wikipedia.org/w/index.php?title=List_of_cognitive_biases&oldid=791032058

[2] Note that we are looking for prior experiences in the context of human-AI collaborations (or HCAI at a higher level), while acknowledging that the term *experiences* can be too broad. In fact, I am seeking suggestions for better terminology from VIS DC.

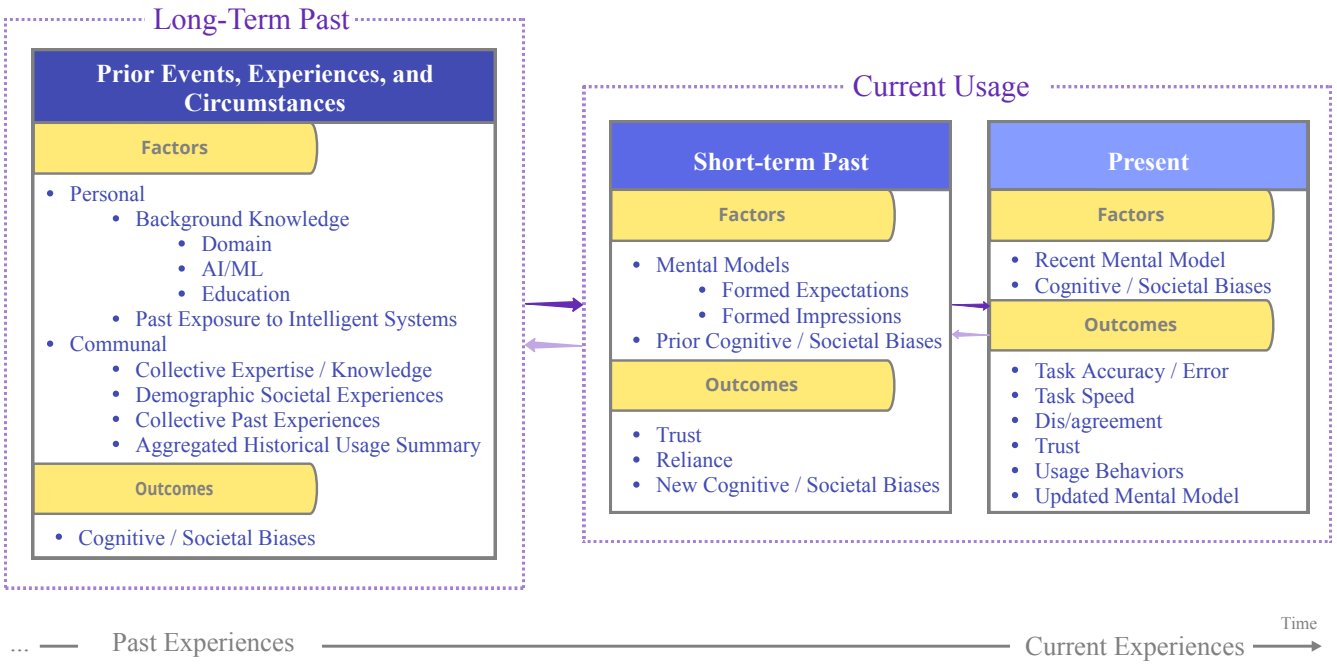

Figure 1: An overview of the preliminary conceptual model of user's past experiences [29]. The model works based on time, where each box represents time of experience: long-term past experiences, short-term past experiences, and present usage. Each of the boxes includes a set of *factors* that influence user actions that lead to certain *outcomes*. The dark purple arrows (from left to right) demonstrate how the past experiences affect those taking place more recently. As the time goes by (i.e., present turning to past), the state of each of the factors update based on more recent interactions. This constant back-and-forth between present, short- and long-term pasts is denoted by light purple arrows (from right to left).

Through working memory, during the current usage stage, people may not only develop biases that are only relevant in the short-term, but could also form new biases that have lasting effects. For instance, anchoring bias is a heuristic that is formed based on early impressions of information, but can impose challenges on usage behaviors. Moreover, continued usage of intelligent systems affects current usage behaviors and how different human factors calibrate over time. For instance, people's trust can change with any given interaction, and when examined over time, one can understand their overall trends in changes of trust over time. This continued back-and-forth between long-term, short-term, and current usage is depicted by arrows on the conceptual model in Figure 1, with new experiences contributing to the past experiences, and the past experiences affecting people's usage behaviors. The preliminary conceptual model of users' past experiences has been recently accepted to appear in a journal paper at ACM TiiS [29]. However, this is still a **Work-in-Progress**. Moving forward, I plan to update and refine the model by conducting a semi-systematic literature review of the broad field of human-centered AI (based on various venues, from ACM CHI/IUI to IEEE VIS/TVCG) to formalize the current research efforts in this community, identify the existing gaps of knowledge, and propose future research directions.

3.2 Short-Term Past on Current Usage (Completed)

Following the framing of the described conceptual model, a first step was to verify and better understand how people' *short-term* experiences with intelligent systems affect their perceptions of the model and current usage; specifically, when forming mental models of systems that they have not used before, as mental models play an important role on usage behaviours. By using anchoring bias (cognitive heuristic describing people's tendency to rely on information based on the order they are received) as an example of short-term past experience, I designed and conducted a user study to understand how people's mental models and confidence in an interactive XAI system's outputs is affected by their negative or positive impressions of the intelligent system. To investigate the interactions between *short-term* past experience and the development of mental models, the study investigated the interplay between system explainability and anchoring bias in a cooking video activity recognition scenario where participants were asked to verify if kitchen policies are being followed by the employees. By altering the order of the policies to control *when* users encounter errors, we observed people's formed mental models and confidence were anchored and varied based on whether their formed impressions were positive or negative. This study, which contributes new empirical knowledge of potential effects of *short-term* recent experiences and provides insights into what to consider when designing XAI tools, has been **completed** and previously published and presented in ACM IUI and was a paper award winner [27].

3.3 Short-Term vs. Long-Term Past (Completed)

To better understand the interplay between *short-term* (via anchoring bias) and *long-term* (via domain experiences) past experiences as described in the conceptual model, I designed another study to compare elements from each of these stages. The goal was to understand how present usage behaviours, such as trust, is anchored by people's first impressions of a domain-specific AI system and how prior task domain knowledge affects these behaviours and impressions. We recruited novice and experienced participants from entomology and *only* controlled the order of observing misclassifications. This study found evidence to support that prior experiences of recent and long-term nature each play a different role in how people understand and trust an intelligent system, highlighting the importance of investigating human-AI partnership by accounting for people's prior differences, as well as how such differences can be augmented and accounted for when designing AI systems. This study is **com-**

pleted and was previously published in the 2020 AAAI Conference in Human Computation and Crowdsourcing (HCOMP) [26].

3.4 Expectations of AI via AI-Related Content (Proposed)

The final chapter of my proposal is allocated to explore an existing gap identified from the conceptual model of past experiences. Due to their biases, people form expectations of their encounters, which can ultimately affect their experiences. With AI, aside from their first-hand daily experiences with tools and systems powered by this technology, people may be exposed to content about AI which may or may not be true, but could influence their judgment of real-world AI-powered systems. This can be from creative content about AI, such as fictional and entertainment content (as seen in movies and series, magazines, comic books, and novels), or content seen on the news, social media, and social interactions.

Motivated by this observation, I aim to investigate whether people's exposure to such AI-related content leads to expectation formations towards AI in real world; whether it is possible to categorize people's usage behaviours based on their formed expectations of AI based on exposure to this content; and if it is possible to predict (and potentially, mitigate) people's behaviours based on these categories. To answer these questions, through a literature review of prior work in *digital literacy* and *human-centered AI*, I will first identify and refine a potential questionnaire that can help categorizing people's differences in their formed expectations towards AI and computing systems based on their exposures to AI-related content.

With the questionnaire formalized, I will design and conduct a user study with two goals: 1) to test the hypothesis that the questionnaire can help identify people's different behaviours based on the expectation category the questionnaire helped identify, and 2) to test whether we can capture and predict people's behaviours based on their expectation category. To achieve this, I will use a AI system in the context of human-AI teaming where the user and the AI algorithm are jointly collaborating to make a decision in order to test these detected categories against human factors such as mental model formations, trust, reliance, and task performance. The proposed study will be designed to adhere to all the three phases that are proposed in the conceptual model in Section 3.1. Similar to the completed work, The *short-term* past experiences are exemplified with anchoring bias, while the information on the *long-term* past experiences will be captured and collected through the questionnaire. Through a collaborative human-AI decision-making task in such setup, the outcomes of *current* experiences will be measured to understand usage behaviors.

My work contributes empirical evidence to demonstrate the importance of studying AI systems from the perspectives of human users and their differences, and advocates for designing tools that elevate these prior differences to improve human-AI collaborations and partnership. As a final step, once I am finished with all the proposed work, I will update the conceptual model of prior experiences to better reflect the structured literature review and my contributions.

4 Discussion and Potential Feedback Areas

This dissertation thesis is proposing novel work in the area of human-centered AI/XAI that can help improve the design of intelligent systems more ethically and responsibly, while shedding light into potential avenues of research not yet examined. Currently, I am working on performing a systematic literature review draw from existing body of work in this domain to support and improve the organized conceptual model from Section3.1. The next steps involve creating a proof-of-concept questionnaire from Section 3.4.

Given that many interactive intelligent systems take advantage of visualization and visual analytics techniques, I am certain the experts (organizers and panelists) of Doctoral Colloquium organized at IEEE VIS 2022 conference can provide invaluable insight into how I can improve my dissertation research. To this end, I am seeking feedback regarding the general framing of my dissertation (as summarized in this manuscript), as well as some of the more specifics of the proposed, incomplete work. In particular, regarding the latter, I hope the expert panelists and DC participants to help me in this research with the following areas/questions.

Regarding the conceptual model of users' prior experiences in Section 3.1:

1. It has been challenging to find the appropriate terminology to use for the conceptual model. For example, during my proposal defense, I have received comments on how the term 'experiences' is broad, complex, and can be referred to different concepts and impossible to capture. I was hoping to get suggestions on terminology that can better capture what I try to study.

2. Which relevant experiences exist in HCAI which is not reflected through the model? Are there anything missing? How can the model be improved?

3. The current model is organized, resembling people's short-/long-term, and working memory. Are there any other intuitive ways that may be relevant/interesting to organize the model based on literature from psychology/social sciences?

Regarding the questionnaire and study in Section 3.4:

1. Are there any questionnaires of this sort that already exist in other fields? If so, from which fields and what are they?

2. The questionnaire will be designed with the goal to capture different aspects of human-AI expectations and perceptions, such as people's anxieties, emotions, and expectations of competency towards AI technology. Are there any other categories here that would make sense/are relevant?

5 Conclusion

Designing interactive, collaborative AI-powered tools that support users with their decision-making and analytical reasoning is a timely topic, researched in various fields and communities. Meanwhile, the field of responsible AI is focused on developing practices and guidelines to build this technology wary of their users and stakeholders. The challenge to understanding human behaviors with intelligent systems is that many types of human differences may lead to distinct usage behaviors, influencing the outcomes of the human and AI collaborative efforts. People's unique personal experiences shapes their understanding of technology and their views of the world. In my dissertation project, I focus on formulating *which* prior experiences influence people's usage behaviors by organizing a conceptual model of users' prior experiences and their influences on human-AI collaborations, and propose studies that contribute to raising researchers and developers' awareness towards users and their differences, as a first step to building tools that augment these differences and are able to mitigate potential challenges imposed by them.

Acknowledgments

I would like to thank my PhD advisory committee, Eric Ragan, Juan Gilbert, Peter Kvam, Jenn Wortman Vaughan, and Vincent Binschaedler for their endless supports on this journey.

REFERENCES

[1] A. Adadi and M. Berrada. Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, 6:52138–52160, 2018.

[2] Y. Ahn and Y.-R. Lin. Fairsight: Visual analytics for fairness in decision making. *IEEE transactions on visualization and computer graphics*, 2019.

[3] A. Bussone, S. Stumpf, and D. O'Sullivan. The role of explanations on trust and reliance in clinical decision support systems. In *2015 International Conference on Healthcare Informatics*, pp. 160–169. IEEE, 2015.

[4] E. F. Can and A. Ezen-Can. The effect of data ordering in image classification. *arXiv preprint arXiv:2001.05857*, 2020.

[5] N. Cowan. What are the differences between long-term, short-term, and working memory? *Progress in brain research*, 169:323–338, 2008.

[6] M. Desai, P. Kaniarasu, M. Medvedev, A. Steinfeld, and H. Yanco. Impact of robot failures and feedback on real-time trust. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 251–258. IEEE, 2013.

[7] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic. A task-based taxonomy of cognitive biases for information visualization. *IEEE transactions on visualization and computer graphics*, 26(2):1413–1432, 2018.

[8] F. Doshi-Velez and B. Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.

[9] U. Ehsan, S. Passi, Q. V. Liao, L. Chan, I. Lee, M. Muller, M. O. Riedl, et al. The who in explainable ai: How ai background shapes perceptions of ai explanations. *arXiv preprint arXiv:2107.13509*, 2021.

[10] M. Eiband, H. Schneider, M. Bilandzic, J. Fazekas-Con, M. Haug, and H. Hussmann. Bringing transparency design into practice. In *23rd International Conference on Intelligent User Interfaces*, pp. 211–223. ACM, 2018.

[11] E. Fourakis and J. Cone. Matters order: The role of information order on implicit impression formation. *Social Psychological and Personality Science*, 11(1):56–63, 2020.

[12] F. J. C. Garcia, D. A. Robb, X. Liu, A. Laskov, P. Patron, and H. Hastie. Explainable autonomy: A study of explanation styles for building clear mental models. In *Proceedings of the 11th International Conference on Natural Language Generation*, pp. 99–108, 2018.

[13] K. A. Hoff and M. Bashir. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, 57(3):407–434, 2015.

[14] F. Hohman, M. Kahng, R. Pienta, and D. H. Chau. Visual analytics in deep learning: An interrogative survey for the next frontiers. *IEEE transactions on visualization and computer graphics*, 25(8):2674–2693, 2018.

[15] E. Horvitz. Mixed-initiative interaction. *IEEE Intelligent Systems*, pp. 14–24, September 1999.

[16] T. Kaluarachchi, A. Reis, and S. Nanayakkara. A review of recent deep learning approaches in human-centered machine learning. *Sensors*, 21(7):2514, 2021.

[17] F. C. Keil. Explanation and understanding. *Annu. Rev. Psychol.*, 57:227–254, 2006.

[18] T. Kulesza, S. Stumpf, M. Burnett, S. Yang, I. Kwan, and W.-K. Wong. Too much, too little, or just right? ways explanations impact end users' mental models. In *2013 IEEE Symposium on Visual Languages and Human Centric Computing*, pp. 3–10. IEEE, 2013.

[19] V. Lai, C. Chen, Q. V. Liao, A. Smith-Renner, and C. Tan. Towards a science of human-ai decision making: A survey of empirical studies. *arXiv preprint arXiv:2112.11471*, 2021.

[20] T. Lombrozo. Simplicity and probability in causal explanation. *Cognitive psychology*, 55(3):232–257, 2007.

[21] T. Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, 2019.

[22] S. Mohseni, N. Zarei, and E. D. Ragan. A survey of evaluation methods and measures for interpretable machine learning. *ACM Transactions on Interactive Intelligent Systems*, 2018.

[23] S. Mohseni, N. Zarei, and E. D. Ragan. A multidisciplinary survey and framework for design and evaluation of explainable ai systems. *arXiv preprint arXiv:1811.11839*, 2019.

[24] R. S. Nickerson. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2):175–220, 1998.

[25] M. Nourani, D. R. Honeycutt, J. E. Block, C. Roy, T. Rahman, E. D. Ragan, and V. Gogate. Investigating the importance of first impressions and explainable ai with interactive video analysis. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp. 1–8, 2020.

[26] M. Nourani, J. T. King, and E. D. Ragan. The role of domain expertise in user trust and the impact of first impressions with intelligent systems. In *Eighth AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*, 2020.

[27] M. Nourani, C. Roy, J. E. Block, D. R. Honeycutt, T. Rahman, E. Ragan, and V. Gogate. Anchoring bias affects mental model formation and user reliance in explainable ai systems. In *26th International Conference on Intelligent User Interfaces*, pp. 340–350, 2021.

[28] M. Nourani, C. Roy, J. E. Block, D. R. Honeycutt, T. Rahman, E. D. Ragan, and V. Gogate. Anchoring bias affects mental models and user reliance in explainable ai systems. In *Proceedings of the 26th International Conference on Intelligent User Interfaces (IUI)*, 2021.

[29] M. Nourani, C. Roy, J. E. Block, D. R. Honeycutt, T. Rahman, E. D. Ragan, and V. Gogate. On the importance of user backgrounds and impressions: Lessons learned from interactive ai applications. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2022.

[30] A. M. Pena, E. H. Nirjhar, A. Pachuilo, T. Chaspari, and E. D. Ragan. Detecting changes in user behavior to understand interaction provenance during visual data analysis. In *IUI Workshops*, 2019.

[31] B. Petrak, K. Weitz, I. Aslan, and E. Andre. Let me show you your new home: studying the effect of proxemic-awareness of robots on users' first impressions. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 1–7. IEEE, 2019.

[32] J. Schaffer, J. O'Donovan, J. Michaelis, A. Raglin, and T. Höllerer. I can do better than your ai: Expertise and explanations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 240–251, 2019.

[33] N. Schwarz, H. Bless, F. Strack, G. Klumpp, H. Rittenauer-Schatka, and A. Simons. Ease of retrieval as information: another look at the availability heuristic. *Journal of Personality and Social psychology*, 61(2):195, 1991.

[34] T. Spinner, U. Schlegel, H. Schäfer, and M. El-Assady. explainer: A visual analytics framework for interactive and explainable machine learning. *IEEE transactions on visualization and computer graphics*, 2019.

[35] H. Strobelt, S. Gehrmann, M. Behrisch, A. Perer, H. Pfister, and A. M. Rush. S eq 2s eq-v is: A visual debugging tool for sequence-to-sequence models. *IEEE transactions on visualization and computer graphics*, 25(1):353–363, 2018.

[36] P. Vaidyanathan, J. Pelz, C. Alm, P. Shi, and A. Haake. Recurrence quantification analysis reveals eye-movement behavior differences between experts and novices. In *Proceedings of the symposium on eye tracking research and applications*, pp. 303–306, 2014.

[37] J. W. Vaughan and H. Wallach. A human-centered agenda for intelligible machine learning. *Machines We Trust: Getting Along with Artificial Intelligence*, 2020.

[38] E. Wall, L. Blaha, C. Paul, and A. Endert. A formative study of interactive bias metrics in visual analytics using anchoring bias. In *IFIP Conference on Human-Computer Interaction*, pp. 555–575. Springer, 2019.

[39] E. Wall, L. M. Blaha, L. Franklin, and A. Endert. Warning, bias may occur: A proposed approach to detecting cognitive bias in interactive visual analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 104–115. IEEE, 2017.

[40] E. Wall, L. M. Blaha, C. L. Paul, K. Cook, and A. Endert. Four perspectives on human bias in visual analytics. In *Cognitive biases in visualizations*, pp. 29–42. Springer, 2018.

[41] D. Wang, Q. Yang, A. Abdul, and B. Y. Lim. Designing theory-driven user-centric explainable ai. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–15. ACM, 2019.

[42] C. T. Wolf. Explainability scenarios: towards scenario-based xai design.

In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 252–257. ACM, 2019.

[43] M. Yin, J. Wortman Vaughan, and H. Wallach. Understanding the effect of accuracy on trust in machine learning models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[44] L. A. Zebrowitz. First impressions from faces. *Current directions in psychological science*, 26(3):237–242, 2017.

[45] J. R. Zech, J. Z. Forde, and M. L. Littman. Individual predictions matter: Assessing the effect of data ordering in training fine-tuned cnns for medical imaging. *arXiv preprint arXiv:1912.03606*, 2019.

[46] J. Zhang, Y. Wang, P. Molino, L. Li, and D. S. Ebert. Manifold: A model-agnostic framework for interpretation and diagnosis of machine learning models. *IEEE transactions on visualization and computer graphics*, 25(1):364–373, 2018.

[47] R. Zhang, N. J. McNeese, G. Freeman, and G. Musick. " an ideal human" expectations of ai teammates in human-ai teaming. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3):1–25, 2021.

[48] Y. Zhang, Q. V. Liao, and R. K. Bellamy. Effect of confidence and explanation on accuracy and trust calibration in ai-assisted decision making. *arXiv preprint arXiv:2001.02114*, 2020.