eXplainable AI: Take one Step Back, Move two Steps forward

Investigating Folk Theories and Users' Perception of Artificial Intelligence

Fatemeh Alizadeh
Information systems and new media
University of Siegen
Siegen, Germany
Fatemeh.alizadeh@uni-siegen.de

Margarita Esau
Information systems and new media
University of Siegen
Siegen, Germany
Margarita.esau@uni-siegen.de

Gunnar Stevens
Information systems and new media
University of Siegen
Siegen, Germany
Gunnar.stevens@uni-siegen.de

Lena Cassens
Business sciences
University of Bonn-Rhein-Sieg
Bonn, Germany
Lena.cassens@h-brs.de

ABSTRACT

In 1991 the researchers at the center for the Learning Sciences of Carnegie Mellon University were confronted with the confusing question of "where is AI" from the users, who were interacting with AI but did not realize it. Three decades of research and we are still facing the same issue with the AI-technology users. In the lack of users' awareness and mutual understanding of AI-enabled systems between designers and users, informal theories of the users about how a system works ("Folk theories") become inevitable but can lead to misconceptions and ineffective interactions. To shape appropriate mental models of AI-based systems, explainable AI has been suggested by AI practitioners. However, a profound understanding of the current users' perception of AI is still missing. In this study, we introduce the term "Perceived AI" as "AI defined from the perspective of its users". We then present our preliminary results from deep-interviews with 50 AI-technology users, which provide a framework for our future research approach towards a better understanding of PAI and users' folk theories.

KEYWORDS

Artificial intelligence, Folk theories, Mental models, Perceived AI, Misconception

MuC'20 Workshops, Magdeburg, Deutschland © Proceedings of the Mensch und Computer 2020 Workshop on «Workshop on User-Centered Artificial Intelligence (UCAI 2020) ». Copyright held by the owner/author(s). https://doi.org/10.18420/muc2020-ws111-369

1 Introduction

In 1991, a few artificial intelligence-enabled prototypes of higher education programs were successfully used at the center for the Learning Sciences of Carnegie Mellon University. Thereby, researchers were frequently confronted with the puzzling question of "Where is AI" by the confused users who did not recognize AI in the interaction [1]. Schank considered the mismatch between the prototypes and the then-current users' definition of AI as the main reason for this phenomenon [1]. Thus, he analyzed different viewpoints on AI and distinguished between the four following groups:

1. AI means magic bullets: A machine is intelligent if it builds unanticipated connections.
2. AI means inference engines: A machine is intelligent if it turns experts' knowledge into rules.
3. AI means getting a machine to do something you didn't think a machine could do (the "gee whiz" view): A machine is intelligent if it does a task that no machines have ever done before.
4. AI means machines that can learn: A machine is intelligent if it can learn by itself.

Over the three decades ever since, AI enjoyed an increasingly growing attention from researchers and the leading industry [2]–[5]. Nevertheless, regarding the public perception of AI, several studies have shown that even today, users cannot realize that they are interacting with an AI-enabled technology [6]–[11]. What makes the problem worse is that defining AI is a confusing task, not only for the users with no computational background but even for the researchers and AI practitioners [7], [12]. This is, among others, due to the evolution of the term over time, and that it has never represented one specific technology in one specific period [5], [11]. As a result, the current limited perception, misconceptions, and unrealistic

expectations of AI (e.g. Superhuman fallacy) [7], [14], not only prompt frustration by unmet users' desires but also restrain effective interaction and collaboration with AI [7], [15].

To elaborate further, individuals create informal theories ("Folk theories") to be able to perceive how a system works and interact with it, which are not necessarily accurate and realistic [17]. With the widespread use of recommender systems, collaborative filtering, and personalized services, there is an undeniable need for active interaction of the users with these intelligent agents [16], [17]. However, the misconceptions and limited folk theories lead to active but not effective interactions, which often fail to bring the desired success and cause frustration among users, instead [7].

There is a huge body of research on explainable AI [18] to support users' understanding of an AI-based system and fight against the misconceptions and incomplete mental models [19]–[21]. However, previous scholars have neglected the pre-existing informal theories of the users on how these system work [20], and some of them have admitted that their reported results on users' perception were primarily and fragmented [22].

To contribute to closing this gap we need to take one step back by investigating the current public perception of AI. For this aim, we introduce the term "Perceived AI" as an evolving phenomenon and investigate its aspects and dimension in comparison with AI form the point of view of its designers. By acknowledging the fact that XAI does not shape the appropriate mental models, but reshapes those already existing, and analyzing users' pre-perception and theories, we will take two steps forward and will be able to evaluate XAI-system more effectively.

2 Research background

The foundation of this study is grounded in three research areas, namely the previous research on public acceptance and awareness of AI-enabled systems, design of more explainable AI-systems, and the users' theories and mental models of how a specific AI-enabled system works (e.g. intelligent agents or Internet of Things). Therefore, we first outline previous research and related terms in the context of public understanding of AI, namely acceptance, trust, and awareness. We then move to clarify the necessity for more explainable systems to support users' interaction with AI. Finally, we will discuss how public informal theories on AI-systems have been investigated distributedly within the realm of four AI-enabled systems and services.

2.1 Public acceptance and awareness of AI

If we take a look at the several large-scale reports on public acceptance and trust regarding AI-enabled technologies, we recognize an increasing positive tendency towards AI. To begin with, in a survey conducted by the British Science Association with over 2000 responds in 2015, one-third of the respondents considered the rise of AI as a danger for mankind [23]. In 2017 some researchers indicated that users still felt uncomfortable with AI making decisions on their behalf, however, the majority of the 27,901 interviewed EU citizens had a generally positive view about robots and AI [19], [21], [25]. Moving to the more recent surveys, we will see greater support from users for future AI developments [9].

Apart from trust and acceptance, the perception of AI has also been discussed as AI-awareness. In a study from HubSpot with over 1,400 consumers worldwide, the results revealed that 63% of those claiming not to use AI-tools were using them without being aware of it [6]. Pega-systems confirmed these findings by research on 5000 consumers, in which only 34% of the respondent agreed they had used an AI-technology before. However, in reality, 84% of them had already used these technologies [8].

Furthermore, researchers of the Center for the Governance of AI at the University of Oxford found out that the majority of users believed virtual assistants, smart speakers, driverless cars, social robots, and autonomous drones use AI but Facebook photo tagging, Google Search, Netflix or Amazon recommendations, and Google Translate do not use AI [22]. The results from Northstar survey with 3804 consumers in 2020 also indicated a significant difference in recognition of 'invisible AI' (i.e. AI-algorithms behind the scenes) and 'visible AI' (i.e. tangible AI devices). Here, 90% of the respondents knew voice assistants are AI-enabled, however, almost a third of them did not consider social media as an AI-based technology [9]. Davies argued for the dependency of the public awareness of AI on the visibility of its application [9]. Nonetheless, this study shows that users do not define AI as humanoids and know the difference between sci-fi and reality [9].

By distinguishing between public perception of AI and public acceptance, trust, and awareness, we aim at contributing to closing the existing gap in the literature and study AI definition and informal theories on how AI-enabled technologies work, from the perspective of the users.

2.2 Designing for explainable technology

To move towards a more transparent design of technology and address trust concerns [18], "Explainable AI" was introduced as AI systems wherein the activities can be effortlessly comprehended and examined by humans [26]. Van Lent et al. [27] used the term for the first time in 2004 for explaining the behavior of the AI-enabled elements in the simulation applications. As of time, scholarly community and professionals have paid renewed attention to explainable AI [18].

By refining users' mental models of AI-enabled systems and resolving their misconceptions, XAI promises a more effective performance of the users [21]. However, previous researchers have criticized XAI for resulting in less efficient systems and less flexible and capable output [18], [28]. To tackle this issue, XAI researchers have argued that explanations are not always

necessary and introduced specific application domains in which they can bring significant benefits (e.g. healthcare and military) [18], [28], [29]. In this regard, Molnar [28] introduces human-friendly explanations as selected and focused on the abnormal. In other words, he suggests that people do not expect complete explanations that cover a full list of causes, but they rather expect them to explain why a system behaved in the way they did not expect [28].

We need to bear in mind that users were already interacting with AI-enabled systems long before the relatively new trend of XAI was considered and applied to design [18]. And as the operation of these systems was opaque, they have often developed theories of how they work (folk theories) to plan their interaction with them [30]–[33]. Folk theories perform as a frame for forming users' expectations [33]. Therefore, understanding them is not only helpful for designing effective explanations but is also necessary for knowing users' assumptions and expectations of the systems. Hence, we propose to step back and investigate the users' understanding and perception of AI-based technology and their pre-existing expectations to discover more valid explanation use cases.

2.3 Making sense of technology

To make sense of technology, users often generate folk theories through direct experiences and social interactions [30], [33]. These informal theories explain how systems operate and support users in responding to them [33]. Therefore, previous studies have argued for the folk theories acting as a window to capture users' perceptions and assumptions [33]. However, the existing studies on mental models and folk theories in the field of AI are highly distributed [34]. Hence, by dividing earlier work into different AI domains, we provide an overview of prior research on folks' theories and their role in understanding users' perceptions, misconceptions, and expectations.

2.2.1 Content curation algorithms and recommender systems Although curating, filtering, and clustering the information form users' perception of the world, previous studies suggest that more than half of the users are not aware of content curation [35]. To address this gap Eslami et al. have conducted several studies on users' perception and folk theories of data curation algorithms [30], [31], [35]. In researching 40 Facebook users to understand their perception of the Facebook News Feed curation algorithm, they found that 65% of the users were completely unaware of it. They continued their study by investigating the perceptions of Facebook users and indicated very specific folk theories, ranging from curation based on users' engagement with other accounts or content to balancing their friends or content by the algorithm [30]. DeVito et al. set the same aim by analyzing 102,827 tweets from a hashtag related to rumors on algorithmic curation on a Twitter timeline (#RIPTwitter) [33]. Their research revealed the abstract and functional folk algorithmic theories that define algorithms as concepts or processes respectively [33]. They also argued that user's folk theories influence their resistance to the algorithmic change and that the more explicit folk theories will cause more explicit user reactions [33].

For their in-the-field learning, personalized agents are also highly dependent on active users' participation. This is due to the significant role users play in adjusting the outputs to their expectations and achieving the desired results [36], [37]. In this respect, Kuhl et al. have investigated the interplay between users' mental models and users' performances [37]. Among others, they have found that there were no consistent understandings of learning algorithms and users' mental models that varied based on their background and experiences [37]. In their study on a music recommender system, Kulesza et al. [36] also showed the impact of mental model soundness on users' interaction with the system. Their study suggests that providing users with structural knowledge on their recommender systems' reasoning will improve their mental model soundness and increase their active participation in the interaction for receiving the desired results [36].

2.2.2 Internet of Things (IoT) Regarding sensor-enabled systems like the Internet of Things (IoT), users develop folk theories on how this data can be used rather than where the data comes from and how it is collected in the first place [38], [39]. Rader and Slaker [39] showed in their study that users explain their activity based on their perception of their performed activities and the processed data provided by the interface. This data can be categorized by the input. Although users entered data (e.g. age and weight) themselves (so that distance and calories were calculated based on steps or location), they did not have a full understanding of their raw data [39]. Hence, users could not make informed decisions for the collection of their personal data and were not aware of further data processing and the underlying operating principles of the device.

Misconceptions may lead to unawareness of privacy risks. This is particularly true when information is collected without consent like in the case of Bluetooth Beacon Systems [40]. Beacons were part of an invisible IoT infrastructure where no direct user interactions were needed for entering personal data (e.g. location). In their study, Yao et al. [40] used drawing as a methodology to reveal user's folk theories on how beacon-based systems work.

The most common misconception was that these sensors collect and store user information, which they actually did not. Further, they assumed they initially needed to actively consent to receive location-based information. Finally, most of the users thought that the location like the store or mall which collected the data also owned the data. As a result, users who consciously or unconsciously decided to be a part of the sensor-enabled system had little understanding of the actual mechanisms of raw data collection and processing. Folk theories focused more on the reasoning of the visible and embodied data collection.

2.2.3 Robots Humans tend to anthropomorphize computational artifacts to rationalize their actions and

behavior that they cannot reasonably explain to themselves. Inaccurate mental models frequently deceive people who as consequence credit autonomous systems with more capability and knowledge than they hold [41]. Those mental models are influenced by the appearance and physical attributes [42], dialogue, personality traits, language, and origin [41] and miss a clear understanding of mechanical and conceptual functionality. These attributes affect the credibility of robots [43]. Powers and Kiesler [43] showed that the robots' facial features like the dimension of forehead and chin impact the perception of intelligence. Hence, people developed high expectations that were not fulfilled and harmed their collaboration and relationship building which require their trust in autonomous systems. The anthropomorphism can support the users in their approach and reactions to the robot and prevent initial rejection. But it might also lead to misconceptions which have detrimental consequences for the outcome of the task that even endanger human lives, like in military settings or critical workspaces [41]. Therefore, humans need a clear and accurate understanding of how robots collect and process data and make decisions beyond anthropomorphism.

2.2.4 Conversational Agents Besides the anthropomorphism of Conversational Agents (CA) [44], extensive research in HCI explores Voice Interaction and the adoption of domestic CA. In particular, some studies have investigated the perception and understanding of the attributed intelligence [39], [44]–[47]. Researchers used methods of drawing [39], [44], [45], interactive tasks [44], [46], [47] and interviews [46], [47] to explore the reasoning and explanations of systems' behavior that children of different ages create. Xu and Warschauer [45] showed that children used behavioral references like listening and talking to justify cognitive properties, reciprocity as an indicator for psychological properties, and biological references like mechanical causality or fantasy reasoning to explain behavioral properties. Thus, children's conceptions of CA, with their mixed animated and unanimated elements, seemed to be more multifaceted. They tended to allocate CA rather to a continuum between humans and artifacts than to a distinct category. Especially younger children who were less aware of the underlying concepts attempted to make sense of computational artifacts by personifying them. When children acknowledged the devices to be more intelligent than themselves [45], [47], they were more likely to trust and believe the information provided which contributed to improved learning [47]. Additionally, understanding the user's emotion and showing affection in reciprocity promotes learning as well [47]. To examine the state of intelligence or consciousness, the study of Druga et al. [46] indicated that younger children asked personal questions like "what is your favorite color?" and older children asked the device to perform actions that they knew humans were capable of. With prior technology experience or an engineering background more thoughtful reasoning was applied [39]. Nonetheless, voice and tone affected the perceived friendliness [46] and available input modalities communicated the expected level of intelligence [46], [47]. Further, the system's output allows us to judge and interpret the actual intelligence and behavior [47]. To enable a meaningful interaction and collaboration, design decisions for the influencing factors [39] should aim at meeting users' expectations or communicating the actual capabilities of CA transparently.

3 Research setting and method

In this section we will explain our research questions and method, then we will describe the preliminary results of a pre-study and how they shape the foundation for our future research and approach.

3.1 Understanding users' perception of AI

When it comes to folk theories and users' assumptions on how a complex system works, in-depth interviews as well as participatory and co-design workshops together with creative methods (e.g. sketching and storytelling) have indicated great potential previously [44]-[47]. Hence, we applied the semi-structured interview method to form an initial understanding of PAI and established a research framework for our future work towards understanding AI as it is perceived. Our aim was to address the following question:

RQ1) How is AI perceived by users?
RQ2) What are the common folk theories about AI?
RQ3) Which factors influence users' perceptions of AI?
RQ4) What are the AI-related misconceptions and what causes them?

3.2 Pre-study and preliminary findings

We conducted in-depth interviews with 50 AI-technology users (29 M, 21 F), including users of voice assistants, fitness trackers, social media, Roboadvisors, and Recommender systems (e.g. in YouTube or Netflix). The majority of the participants were between 20-30 years old and were familiar with the daily use of technology. They were recruited by convenient sampling and were not compensated. The interviews were conducted remotely (via either video or audio call) and took between 20-30 minutes. Apart from demographic information we mainly collected users' answers to the five following questions:

1. What are three examples of artificial intelligence, you can think of?
2. How would you explain to a friend what AI is?
3. Have you ever talked about AI with others (e.g. family, friends, etc.), if yes, what did you talk about?
4. If we tell you that AI is embedded into these objects what would you expect? a) a door b) a bank account
5. If we tell you AI will disappear tomorrow what do you think the result will be?

With the consent of the participants, the interviews were audio-recorded and transcribed afterward. By taking a thematic analysis approach and application of MAXQDA each interview was in-vivo coded and analyzed by two different researchers. In the following, our first findings will be described.

3.2.1 Why PAI? Although most of our participants were similar regarding their age and educational background, they did not have an equal understanding of artificial intelligence. Therefore, we could divide their perceptions of AI into seven main groups:

1) AI means that machines can learn and develop (machine learning).
2) AI means machines that work independently to complete a task and make a decision (automation).
3) AI means machines that know everything (expert systems).
4) AI means machines that simulate human intelligence (neural network).
5) AI means improving human intelligence (intelligent machines with no emotions).
6) AI means machines that can predict.
7) AI means machines that recognize patterns (pattern recognition).

This variety in users' understanding of AI emphasizes the importance and relevance of our research question and the need for a deeper investigation with a larger and representative sample.

Furthermore, our first results confirmed the findings from the Northstar study in 2019 [9], since the majority of the participants mentioned tangible AI devices (e.g. voice assistants and self-driving cars) as prevailing examples of AI. Therefore, future research should cover the differences emerging from the perception of the "invisible" and "visible" AI.

3.2.2 Where is the gap? Almost one-third of the participants defined AI as machines that can learn and develop themselves. However, as they were meant to imagine AI being used in a product like a door, many of them (regardless of their understanding of AI) mentioned a door that closes automatically. Based on this result we assume that many still see AI as a synonym for automation. Some of the respondents also mentioned they see AI as machines that take over the difficult tasks. The terms like "big computers" and "a large amount of data" were also often mentioned in the responses. Therefore, the border between large industrial machines in manufactories and AI as an everyday experience seems to be vague for the users.

Furthermore, some of our respondents had high expectations of the AI-enabled door like "a door that recognizes me and communicates with me based on its knowledge about me" or "a door that can receive the packages itself". This phenomenon suggests that the users do not have the appropriate knowledge when it comes to the implementation, which can lead to unrealistic expectations from the technology and disappointment when these expectations are not met.

Therefore, in the next step, we will aim at identifying this gap by using more profound research techniques to grasp folk theories and understanding of how AI can be embedded in the systems and what we can expect from them.

3.2.3 PAI as the premise of XAI As previous studies have suggested [29], explanations should not only address users' needs but should be adjusted to their understanding. Hence, we cannot design for explainable AI when we do not know how people perceive AI and their theories of how it works in the first place. Our pre-study with a homogenous sample suggests that categorizing users based on their computational background is not enough since this is not the only influencing criteria on users' perception and expectations of AI.

Therefore, we strive for more profound research on the influencing criteria and their effects on users' perception and understanding as a foundation for explainable AI technologies.

4 Conclusion

In this work, we emphasized the importance of understanding AI according to users' perceptions to design efficient and productive explainable AI-enabled systems. By outlining the fragmented previous work on users' mental models and folk theories, reviewing the structure of user-friendly interpretations, and presenting the preliminary results of a pre-study with AI-technology users, we established a framework for our future research questions and further approaches. As our results suggest, users' perception of AI is not purely dependent on their computational background, which indicates the need for future investigation on the influencing factors and the dimensions of PAI and users' folk theories.

REFERENCES

[1] R. C. Schank, „Where's the AI?", *AI Magazine*, Bd. 12, Nr. 4, Art. Nr. 4, Dez. 1991, doi: 10.1609/aimag.v12i4.917.

[2] "Artificial intelligence and life in 2030", one-hundred-year study on artificial intelligence, report of the 2015 study panel, https://ai100.stanford.edu/sites/g/files

[3] „The 6 Stages in the Evolution of AI and Customer Experience". https://zoovu.com/blog/stages-evolution-ai-customer-experience/

[4] Y. Gil und B. Selman, „A 20-Year Community Roadmap for Artificial Intelligence Research in the US", *arXiv:1908.02624 [cs]*, Aug. 2019

[5] GoodAI, „Understanding the public perception of AI", *Medium*, Feb. 18, 2019. https://medium.com/goodai-news/understanding-the-public-perception-of-ai-a14b0e6b6154

[6] M. An, „Artificial Intelligence Is Here - People Just Don't Realize It". https://blog.hubspot.com/marketing/artificial-intelligence-is-here

[7] D. Long und B. Magerko, „What is AI Literacy? Competencies and Design Considerations", in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA, Apr. 2020, S. 1–16, doi: 10.1145/3313831.3376727.

[8] „What Consumers Really think about AI", Pegasystems, https://www.pega.com/system/files/resources/2019-01/what-consumers-really-think-about-ai-study-de.pdf.

[9] „Arm 2020 Global AI Survey". https://pages.arm.com/artificial-intelligence-survey.html

[10] M. Eslami, K. Vaccaro, M. K. Lee, A. Elazari Bar On, E. Gilbert, und K. Karahalios, „User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms", in *Proceedings of the 2019*

CHI Conference on Human Factors in Computing Systems - CHI '19, Glasgow, Scotland Uk, 2019, S. 1–14, doi: 10.1145/3290605.3300724.

[11] M. Eslami *u. a.*, „‚I always assumed that I wasn't really that close to [her]': Reasoning about Invisible Algorithms in News Feeds", in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul, Republic of Korea, Apr. 2015, S. 153–162, doi: 10.1145/2702123.2702556.

[12] R. C. Schank, „What Is AI, Anyway?", *AI Magazine*, Bd. 8, Nr. 4, Art. Nr. 4, Dez. 1987, doi: 10.1609/aimag.v8i4.623.

[13] „What is AI Called?! | SAP Blogs". https://blogs.sap.com/2017/06/20/what-is-ai-called/

[14] M. A. Boden und R. P. of C. S. M. A. Boden, *The Creative Mind: Myths and Mechanisms*. Psychology Press, 2004.

[15] E. Fast und E. Horvitz, „Long-term trends in the public perception of artificial intelligence", in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, San Francisco, California, USA, Feb. 2017, S. 963–969

[16] T. Kulesza, S. Stumpf, M. Burnett, und I. Kwan, „Tell me more? the effects of mental model soundness on personalizing an intelligent agent", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Austin, Texas, USA, Mai 2012, S. 1–10, doi: 10.1145/2207676.2207678.

[17] M. Eslami *u. a.*, „First I ‚like' it, then I hide it: Folk Theories of Social Feeds", in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose, California, USA, Mai 2016, S. 2371–2382, doi: 10.1145/2858036.2858494.

[18] A. Adadi und M. Berrada, „Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)", *IEEE Access*, Bd. 6, S. 52138–52160, Sep. 2018, doi: 10.1109/ACCESS.2018.2870052.

[19] R. R. Hoffman, S. T. Mueller, G. Klein, und J. Litman, „Metrics for Explainable AI: Challenges and Prospects", *arXiv:1812.04608 [cs]*, Feb. 2019, Zugegriffen: Juni 03, 2020. [Online]. Verfügbar unter: http://arxiv.org/abs/1812.04608.

[20] „Why and why not explanations improve the intelligibility of context-aware intelligent systems | Proceedings of the SIGCHI Conference on Human Factors in Computing Systems". https://dl.acm.org/doi/abs/10.1145/1518701.1519023

[21] M. Ribera und À. Lapedriza, „Can we do better explanations? A proposal of user-centered explainable AI", 2019.

[22] B. Z. and A. Dafoe und C. for the G. of A. Oxford Future of Humanity Institute, University of, *2 General attitudes toward AI | Artificial Intelligence: American Attitudes and Trends*. .

[23] „One in three believe that the rise of artificial intelligence is a threat to humanity", *British Science Association*. https://www.britishscienceassociation.org/news/rise-of-artificial-intelligence-is-a-threat-to-humanity

[24] „AI, Automation, and Corporate Reputation", *Ipsos*. https://www.ipsos.com/en/ai-automation-and-corporate-reputation.

[25] „ec.europa.eu". https://perma.cc/9FRT-ADST (zugegriffen Mai 30, 2020).

[26] H. Hagras, „Toward Human-Understandable, Explainable AI", *Computer*, Bd. 51, Nr. 9, S. 28–36, Sep. 2018, doi: 10.1109/MC.2018.3620965.

[27] M. van Lent, W. Fisher, und M. Mancuso, „An Explainable Artificial Intelligence System for Small-unit Tactical Behavior", S. 8.

[28] C. Molnar, *Interpretable Machine Learning*. .

[29] M. Ribera und A. Lapedriza, „Can we do better explanations? A proposal of User-Centered Explainable AI", 2019.

[30] M. Eslami *u. a.*, „First I ‚like' it, then I hide it: Folk theories of social feeds", *Conference on Human Factors in Computing Systems - Proceedings*, S. 2371–2382, 2016, doi: 10.1145/2858036.2858494.

[31] M. Eslami, K. Vaccaro, M. K. Lee, A. E. Bar On, E. Gilbert, und K. Karahalios, „User attitudes towards algorithmic opacity and transparency in online reviewing platforms", *Conference on Human Factors in Computing Systems - Proceedings*, S. 1–14, 2019, doi: 10.1145/3290605.3300724.

[32] M. Eslami *u. a.*, „I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in news feeds", *Conference on Human Factors in Computing Systems - Proceedings*, Bd. 2015-April, S. 153–162, 2015, doi: 10.1145/2702123.2702556.

[33] M. A. De Vito, D. Gergle, und J. Birnholtz, „‚Algorithms ruin everything': #RIPTwitter, folk theories, and resistance to algorithmic change in social media", *Conference on Human Factors in Computing Systems - Proceedings*, Bd. 2017-May, S. 3163–3174, 2017, doi: 10.1145/3025453.3025659.

[34] D. Long und B. Magerko, „What is AI Literacy? Competencies and Design Considerations", in *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, S. 1–16, doi: 10.1145/3313831.3376727.

[35] M. Eslami *u. a.*, „‚I always assumed that I wasn't really that close to [her]': Reasoning about Invisible Algorithms in News Feeds", in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul, Republic of Korea, Apr. 2015, S. 153–162, doi: 10.1145/2702123.2702556.

[36] T. Kulesza, S. Stumpf, M. Burnett, und I. Kwan, „Tell me more? the effects of mental model soundness on personalizing an intelligent agent", in *Conference on Human Factors in Computing Systems - Proceedings*, 2012, S. 1–10, doi: 10.1145/2207676.2207678.

[37] Ni. Kuhl, J. Lobana, und C. Meske, „Do you comply with AI? -- Personalized explanations of learning algorithms and their impact on employees' compliance behavior", *arXiv:2002.08777 [cs]*, Feb. 2020, http://arxiv.org/abs/2002.08777.

[38] E. Rader und J. Slaker, „The importance of visibility for folk theories of sensor data", 2017, S. 257–270, https://www.usenix.org/conference/soups2017/technical-sessions/presentation/rader.

[39] S. Lee, S. Kim, und S. Lee, „‚What does your Agent look like?': A Drawing Study to Understand Users' Perceived Persona of Conversational Agent", in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, Scotland Uk, Mai 2019, S. 1–6, doi: 10.1145/3290607.3312796.

[40] Y. Yao, Y. Huang, und Y. Wang, „Unpacking People's Understandings of Bluetooth Beacon Systems - A Location-Based IoT Technology", *Proceedings of the 52nd Hawaii International Conference on System Sciences*, S. 1638–1647, 2019, doi: 10.24251/HICSS.2019.198.

[41] E. Phillips, S. Ososky, J. Grove, und F. Jentsch, „From tools to teammates: Toward the development of appropriate mental models for intelligent robots", in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2011, Bd. 55, Nr. 1, S. 1491–1495.

[42] S. Kiesler und J. Goetz, „Mental models of robotic assistants", in *CHI'02 extended abstracts on Human Factors in Computing Systems*, 2002, S. 576–577.

[43] A. Powers und S. Kiesler, „The advisor robot: tracing people's mental model from a robot's physical attributes", in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, 2006, S. 218–225.

[44] A. Kuzminykh, J. Sun, N. Govindaraju, J. Avery, und E. Lank, „Genie in the Bottle: Anthropomorphized Perceptions of Conversational Agents", in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA, Apr. 2020, S. 1–13, doi: 10.1145/3313831.3376665.

[45] Y. Xu und M. Warschauer, „What Are You Talking To?: Understanding Children's Perceptions of Conversational Agents", in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, HI, USA, Apr. 2020, S. 1–13, doi: 10.1145/3313831.3376416.

[46] S. Druga, R. Williams, C. Breazeal, und M. Resnick, „‚Hey Google is it OK if I eat you?': Initial Explorations in Child-Agent Interaction", in *Proceedings of the 2017 Conference on Interaction Design and Children*, Stanford, California, USA, Juni 2017, S. 595–600, doi: 10.1145/3078072.3084330.

[47] R. Garg und S. Sengupta, „Conversational Technologies for In-home Learning: Using Co-Design to Understand Children's and Parents' Perspectives", Apr. 2020, doi: 10.1145/3313831.3376631.