How to Manage Output Uncertainty: Targeting the Actual End User Problem in Interactions with AI

Zelun Tony Zhang[a], Heinrich Hußmann[b]

[a]fortiss GmbH, Research Institute of the Free State of Bavaria, Guerickestraße 25, 80805 Munich, Germany
[b]LMU Munich, Chair of Applied Informatics and Media Informatics, Frauenlobstraße 7a, 80337 Munich, Germany

Abstract

Given the opaqueness and complexity of modern AI algorithms, there is currently a strong focus on developing transparent and explainable AI, especially in high-stakes domains. We claim that opaqueness and complexity are not the core issues for end users when interacting with AI. Instead, we propose that the output uncertainty inherent to AI systems is the actual problem, with opaqueness and complexity as contributing factors. Transparency and explainability should therefore not be the end goals, as such a focus tends to place the human into a passive supervisory role in what is in reality an algorithm-centered system design. To enable effective management of output uncertainty, we believe it is necessary to focus on truly human-centered AI designs that keep the human in an active role of control. We discuss the conceptual implications of such a shift in focus and give examples from literature to illustrate the more holistic, interactive designs that we envision.

Keywords

output uncertainty, human-AI interaction, intelligent systems, transparency, explainability, user control

1. Introduction

The field of artificial intelligence (AI) has witnessed impressive progress in recent years. Yet, in many critical, high-stakes domains such as aviation, medical technology or criminal justice, AI is not yet widely deployable due to challenges like brittleness of the algorithms [1] or algorithmic bias. The results are issues in terms of safety, ethics and social justice. The complexity and opaqueness of most modern AI algorithms are generally seen as the core of the problems, prompting widespread calls for AI transparency and explainability. However, despite the increasingly active research towards transparent and explainable AI, the effectiveness of these efforts on the end user side remains unclear. This paper calls for a more holistic perspective on the issues in end user interactions with AI systems, especially in high-stakes domains. We propose to focus on output uncertainty, i.e. the uncertainty of the user about the case-by-case correctness of the algorithmic output, rather than complexity and opaqueness.

2. Background

2.1. What is AI?

When thinking about human-AI interaction, it is useful to have a clear idea about what AI actually is, something which is often taken for granted by those working with AI, but is in fact not that clear at all. We begin by briefly discussing what AI is from two distinct angles: the definitions and properties of AI, and the conceptualizations of AI usage.

2.1.1. Definitions and properties of AI

AI and what counts as AI are hard to define precisely, as intelligence is itself already a concept with no agreed-upon definition [2]. As a result, experts from different fields or even within the same discipline might have varying understandings of the term. As for the field of interaction design, Völkel et al. analyzed all past IUI proceedings for how intelligence is characterized in the IUI community [3]. They identified a clear trend towards a diversification of characterizations over the years.

Related to our perspective, Yang et al. examined the question from the angle of what makes AI uniquely difficult for designers to work with [4], identifying capability uncertainty and output complexity as the two defining dimensions. In contrast, our focus is on what makes AI uniquely difficult for the end user to interact with. To this end, we suggest that AI is usually applied to complex problems which often cannot be fully specified with hard criteria and are therefore subject to uncertainty. Consider for instance recidivism prediction or making medical diagnoses, where humans at least partially rely on experience and gut feeling. It is in problems like these where AI can achieve what conventional programming cannot. We therefore draw on the definition of intelligence by Albus "as the ability of a system to act appropriately in an uncertain environment" [5]. For the purpose of this paper, AI is thus mainly characterized as *computer systems*

Joint Proceedings of the ACM IUI 2021 Workshops, April 13–17, 2021, College Station, USA
zhang@fortiss.org (Z. T. Zhang); hussmann@ifi.lmu.de (H. Hußmann)
© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

that can appropriately handle problems that are subject to uncertainty. In particular, this implies that the resulting outputs of an AI system are also subject to uncertainty. Note that the purpose of this characterization is to capture what defines AI for the end user. As such, it does not necessarily address every understanding of AI. However, the vast majority of user-facing applications of AI should be covered by our characterization.

2.1.2. Conceptualizations of AI Usage

Not only is there no agreement on a clear definition of AI, but there are also diverging opinions on how AI should be used. The longstanding debate in HCI whether computer systems should be designed as tools or as agents [6] also translates to how designers and researchers conceptualize the usage of AI. In Shneiderman's terms, the two extremes of possible conceptualizations are the emulation goal and the application goal [7]. The former refers to the intention to emulate human capabilities with AI through autonomous agents, the latter to the usage of AI to create tools that enhance human performance. The emulation goal tends to encourage algorithm-centered thinking, while pursuing the application goal promotes human-centered thinking. Völkel et al. found that elements of both conceptual extremes are represented in IUI publications [3].

An algorithm-centered conceptualization of AI usage is problematic, especially for high-stakes applications. Automation is frequently implemented to supplant the human in various tasks, e.g. with automatic creditworthiness decisions, automatic recidivism predictions, or autopilots in planes. However, due to imperfect algorithms, human operators are appointed as supervisory instances—a task that is long known to be unsuitable for humans [8]. Aircraft pilots need to constantly monitor the automation at their disposal for correct functioning. Likewise, users of automatic decision aids are not involved in the decision-making process. They are confronted with a suggested result and expected to decide whether to accept or to reject it. Such systems are algorithm-centered in that the end users are ordered to compensate for the shortcomings of the algorithm. As a consequence, the automation ironically often needs to be cut back in its abilities to allow the human a chance to perform his or her supervisory task at all, resulting in lower performance than technically possible. Even worse in high-stakes applications are the costly and ethically problematic mistakes that occur when the human supervisor fails to detect, understand, or correct an algorithm error.

Commonly, the fact that the human is out of the loop in these automated systems is identified as the core issue [9]. With the complexity and opacity of modern AI technologies like deep learning, the focus is therefore naturally on enabling AI transparency and explainability to keep the human in the loop. However, outfitting a system like those described above with a transparent and explainable interface does not change the algorithm-centered nature of the design. The human is still supposed to make up for the deficits of the algorithm; the task only gets more user-friendly, but is in its nature still unsuitable for humans. To prevent burdening users with such unsuitable tasks, there needs to be a more fundamental rethinking of how to deploy AI for the benefit of users. Truly human-centered AI designs should not strive for human-in-the-loop, but AI-in-the-loop [7].

2.2. Transparency and Explainability

The current wave of efforts towards more interpretable AI was initiated by the AI and machine learning communities, with an algorithmic focus on how to create transparent models or how to generate explanations. Usability and actual user needs only gained more importance later on as the HCI community shifted its attention towards the topic [10]. Still, the effectiveness of current approaches remains unclear. This already starts with the fact that the understanding of the concepts of AI transparency and explainability is similarly diffuse as the definition of AI itself. Especially explainability and related concepts like interpretability are not well defined [11], are used to describe differing ideas by different authors [12, 13], and lack agreed-upon ways to evaluate them [11].

HCI researchers have resorted to various measures to evaluate the effectiveness of explanations. A range of user studies that evaluate the effect of explanations on trust [14, 15], system understanding [14, 16], or task performance [17] demonstrate the benefits of AI transparency and explainability. However, in contradiction to these results, many studies show no significant effect of providing explanations [18, 19, 20, 21]. On closer inspection, the effects of adding interpretability to AI systems can be counterintuitive or even negative. For instance, Poursabzi-Sangdeh et al. observed in their study that adding transparency can hinder users to detect serious mistakes of the algorithm, likely due to information overload [22]. Furthermore, Bansal et al. showed that users were more likely to follow AI outputs when given explanations, regardless of the correctness of the outputs [23]. The results of Eiband et al. suggest that this overtrust can even occur with placebic explanations, i.e. explanations that contain no actual information [24].

Taken together, there is no clear picture of when and how transparent and explainable AI can be achieved, despite the rich and rapidly growing body of research. The widely spread results hint at the complexity of the topic, with numerous contributing factors that might not be obvious and necessitate much more further research.

3. A More Holistic Perspective

3.1. A Provocative Question

Given the complicated matter of making AI interpretable, one should be allowed to ask a provocative question: Why do we actually need transparency and explanations in AI? The common line of thinking is that modern AI systems are too complex and opaque to understand how specific outputs are generated, that they constitute black boxes. For AI engineers, this hinders their development work. For regulators, it complicates the evaluation of compliance with regulations. And also for end users, the black box property is commonly thought to be an issue. However, many systems in our lives are complex black boxes in the eyes of the user, and yet neither the average user nor the designers of these systems care about transparency to open up these black boxes. For instance, a car is sufficiently complex that the average driver has no good understanding of its inner workings. Yet, there is no need to make its complex engineering transparent or explainable, despite the high-stakes, safety-critical nature of the car. So what is the difference for the end user with AI?

3.2. Output Uncertainty is the Actual Problem

In the context of end user interactions, complexity and opaqueness appear not to be the problem per se. Instead, we argue that from an end user interaction perspective, the distinguishing factor of AI as characterized by us in Section 2.1.1 is what we refer to as *output uncertainty*. By output uncertainty we mean the uncertainty of the user about the case-by-case correctness of the algorithmic output. Note that we focus on end users here who directly interact and work with the AI systems, e.g. pilots flying with AI assistance, physicians making diagnoses with the help of AI, or police departments employing AI-enabled predictive policing systems. For other stake holders like developers or regulators, different considerations might apply.

Consider the braking system of a car, which could be arbitrarily complex and opaque, without the average driver ever caring about it. Since the driver can be sure that the car will slow down when stepping on the brake (and that the brake will not apply otherwise), there is no need for him or her to wonder about how the result came to be. On the other hand, even very simple rule-based systems could be problematic for the user, despite a high degree of transparency and explainability. Take for instance an agent that automatically categorizes emails according to a manageable set of simple and explicit rules based on factors like sender, keywords, or time. Since the real criteria for how to categorize the emails are unlikely to be captured fully by these simplistic rules, the user can never be certain that no emails have been misclassified without tedious manual verification and correction. The fact that the user can use the rules to reconstruct how the emails have been categorized does not resolve the issue. These examples certainly by far do not cover the whole range of problems where transparency and explainability are demanded. However, they serve the purpose to illustrate that not complexity and opaqueness are the root of the problem, but rather output uncertainty. This view is supported by studies showing that displaying the confidence of the model in its output is highly effective in calibrating users' trust to appropriate levels[1], while giving explanations has no significant effect [23, 21].

We do regard complexity and opaqueness as highly important issues as they constitute major contributing factors to output uncertainty. However, focusing on them as if they were the root problems can limit our thinking when searching for human-centered ways to deploy AI and reap its benefits in high-stakes scenarios. We propose that designers of AI systems should instead focus on managing output uncertainty, considering complexity and opaqueness as contributing factors. The currently popular practice is to present fully automatically generated outputs to users as a fait accompli. Users need to constantly reconstruct the reasons behind these outputs in order to reject them and to override the algorithm when necessary. The goal should be to come up with more holistic and effective designs than that. Now, this does not preclude current designs and efforts towards fully automatic and hopefully transparent decision aids. But the point is that there needs to be a better understanding of when and why such a design could be appropriate, rather than taking it as a default or a given.

3.3. Output Uncertainty and Related Constructs

As stated in Section 3.2, we define output uncertainty as the uncertainty of the user about the correctness of the model output on a case-by-case basis. We acknowledge that several similar constructs have already been investigated in human-AI interaction research. In order to clarify our perspective, we briefly discuss how output uncertainty differs from these related constructs in this section.

Most notably, managing output uncertainty appears to be very similar to calibrating *trust* in AI systems. However, output uncertainty management recognizes the inevitability of AI errors and is concerned with designs to manage these errors. As such, it is a wider problem than trust calibration, which relies mostly on explanations

[1] The user has appropriate trust in the model if the user follows the model output when it is correct and rejects it when it is wrong.

to help users recognize when to trust or to dismiss an algorithmic output. While this is a possible approach, there are more ways to manage output uncertainty, as we describe in Section 3.4.2. Furthermore, trust in AI is a highly convoluted construct with many different meanings. [25]. Its conceptualization is also influenced by our human intuition about interpersonal trust, which can cause misleading conclusions [25]. In contrast, output uncertainty is a much simpler and more focused notion, which could be seen as one influencing factor in the complex of trust and trustworthiness.

We also differentiate output uncertainty from *unpredictability*, as a system can behave predictably on some level while still inducing output uncertainty. This could for instance be the case for the exemplary email agent mentioned in Section 3.2. Due to its simple and explicit rules, the system behavior can be considered predictable. Yet still, the agent might unexpectedly miscategorize some emails due to the flexible nature of language, creating output uncertainty on the end user side. Another way unpredictability can differ from output uncertainty is when the latter is precisely quantifiable, i.e. the user knows the likelihood that the system is correct in any given situation. Hoping for a specific number while throwing a dice would be the simplest example. In such a case, the global behavior of the system is predictable to the user, but the uncertainty about the case-by-case correctness remains.

In the same vein, output uncertainty is not the same as the *confidence* of the model in the correctness of its outputs—or rather the lack thereof: Confidence scores are supposed to reveal how (un)certain the *model* is about its outputs, whereas output uncertainty is the uncertainty of the *user* about the model outputs. While confidence scores might possibly be a viable method to manage output uncertainty in specific situations, both concepts can also be detached from each other. An algorithm can be correct despite low confidence and vice versa. Hence, an uncertainty on the user's side about the case-by-case correctness of the model output can persist, even for very high or very low system confidence.

3.4. Conceptual Implications

A more holistic perspective on human-centered AI deployment with a focus on output uncertainty has conceptual implications in two distinct but closely related ways: in terms of how we conceptualize AI and its usage and in terms of the design solutions we consider. We discuss both briefly in the following.

3.4.1. Recalibrating conceptualizations of AI towards human-centered, user-empowering tools

As discussed in Section 2.1.2, the currently predominant algorithm-centered conceptualization of AI is problematic. The emphasis on transparency and explainability tends to reinforce such a conceptualization. This is not because demands for these properties would be wrong, but because the isolated focus on them is based on the assumption that AI is necessarily implemented as fully automatic systems that need human supervision. Looking solely at transparency and explainability does not question whether this assumption is compatible with human cognitive limits; it is merely concerned with making this inherently algorithm-centered paradigm more user-friendly at best (see Fig. 1, left).

Focusing on output uncertainty on the other hand has the potential to recalibrate the conceptualization of AI towards a more human-centered direction. As shown by decades of research in human-automation interaction, humans are inherently not well suited to deal with output uncertainty in a passive supervisory role [9]. Therefore, when considering effective ways to manage output uncertainty, it is necessary to consider how to actively engage the user in performing the task. In this way, addressing the problem of output uncertainty encourages a shift towards giving the user an active role at the center. AI systems would be designed around the user, as tools that enhance the user's ability to perform the task.

3.4.2. Recalibrating design thinking with regards to AI towards more holistic, interactive designs

Solely focusing on transparency and explainability is an example of jumping straight to the solution without proper search for the root problem first. The result is that a large part of the solution space is not considered at all. The proposed focus on output uncertainty would no longer regard complexity and opaqueness as the root problems, but as contributing factors to output uncertainty. This means that transparency and explainability are not seen as final goals, but as possible building blocks for effective human-AI interactions. By this reframing of the problem, the focus on output uncertainty can open up the ideation activities of a design thinking process for a much wider range of possible solutions. For instance, ideation does not always need to focus on how to make fully automatic decision aids more transparent and explainable. A solution could instead revolve more around how to allow users to steer the algorithm so that it enhances the user's abilities while actively performing the task him- or herself (see Fig. 1, right). This could involve more exploration of effective input techniques and how

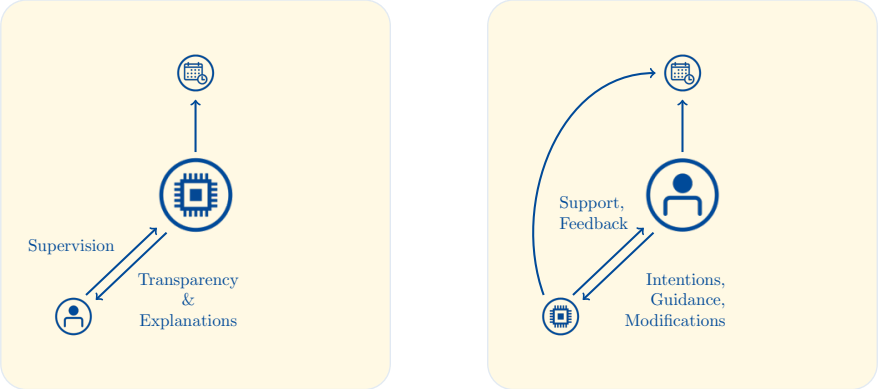

Figure 1: User-friendly, but algorithm-centered versus truly human-centered AI systems. **Left**: The algorithm is designed to handle the task fully automatically. The human is not involved, but is placed into a supervisory role to make up for the shortcomings of the AI, an inherently algorithm-centered design where the user's task is not suitable to humans. Focusing on transparency and explainability does not change the nature of the user's task and can merely make the task more user-friendly. **Right**: A truly human-centered design frees the user from the supervisory role. Instead, the AI supports the user in his or her task and is steered by the user according to his or her goals.

transparency could be integrated with those to enable feedback in interactions with the system.

The literature provides several promising examples for how such more holistic, interactive designs could look like. Cai et al. developed a deep learning-based image retrieval system for medical decision making with three different tools to help physicians refine the retrieved results [26]: cropping to indicate important regions of an image, pinning of examples that contain the searched-for concept, and sliders to (de-)emphasize certain clinical concepts. Weber et al. proposed an image restoration tool where the user can iteratively control and guide the inpainting algorithm by manually painting directly onto the image to be restored [27]. Heer presented three case studies in the domains of data cleansing and formatting, data exploration, and natural language translation [28]. In all of his case studies, the predictive models work on a task representation shared with the user, and are integrated into interactive systems such that they provide helpful assistance to the user.

All these exemplary designs allow users to manage the output uncertainty of the underlying AI algorithms. Instead of being an all-or-nothing affair depending on whether the algorithms are right or wrong, these systems provide helpful assistance to their users even in cases where their output is not entirely correct. Users can still work with imperfect outputs by manipulating the results. Furthermore, users can work forwards towards their goals, instead of being forced to work backwards from an automatic AI output. Transparency in these systems is therefore not achieved by providing explicit explanations, but by actively engaging the user and giving the user control in performing the task.

Note that the presented examples bear a strong resemblance to techniques of *interactive machine learning* (iML) [29], where interactive user feedback is a key concern. However, our focus is on managing output uncertainty, while iML has the specific purpose to make machine learning more accessible to users that are not machine learning practitioners. The ultimate goal of iML is therefore to improve the performance of the algorithm through a well designed, usable training process. We regard output uncertainty as a more fundamental issue that plagues end user interactions with AI in general. Techniques from iML can be important contributions to designs that effectively manage output uncertainty, though.

We reiterate that our point is not to rule out fully automatic systems with transparent and explainable interfaces. Instead, we call for a more complete view of the solution space by focusing on output uncertainty. We see two pillars to this: (1) We need a framework for when fully automatic systems are appropriate, and when more interactive solutions are necessary to manage output uncertainty. (2) There is currently little understanding on how to design more interactive AI systems like those mentioned above. We therefore see a need for more research into pertinent guidelines and techniques.

4. Conclusion

In our view, complexity and opaqueness are not the root problems for end users when interacting with AI, as is commonly assumed. Instead, these properties contribute to what we see as the actual problem that needs to be ad-

dressed: output uncertainty. We believe that effectively addressing output uncertainty requires more holistic, interactive designs than merely transparent and explainable interfaces. Such designs are not all-or-nothing affairs depending on the correctness of the algorithm output; allow users to work forwards towards their goal instead of backwards from the AI output; and allow the AI system to effectively support the user. Overall, such designs would be much more human-centered. However, we still need a much better understanding of how to achieve these designs and when it is appropriate to choose fully automatic system designs instead.

We believe that such a human-centered approach that goes beyond transparency and explainability is necessary to overcome the barriers to AI deployment concerning safety, ethics and social justice. Therefore, we initially plan to develop our line of thinking concretely into a concept for assessing human factors in the certification of AI systems in the aviation domain. Our long-term goal is to extend the expected results of this project to other high-stakes domains as well.

Acknowledgments

This work was supported by the German Federal Ministry for Economic Affairs and Energy (BMWi) under the LuFo VI-1 program, project KIEZ4-0.

References

[1] D. Heaven, Deep trouble for deep learning, Nature 574 (2019) 163–166. doi:10.1038/d41586-019-03013-5.

[2] S. Legg, M. Hutter, A collection of definitions of intelligence, arXiv:0706.3639 [cs] (2007). arXiv:0706.3639.

[3] S. T. Völkel, C. Schneegass, M. Eiband, D. Buschek, What is "intelligent" in intelligent user interfaces?: A meta-analysis of 25 years of IUI, in: Proceedings of the 25th International Conference on Intelligent User Interfaces, IUI '20, ACM, 2020, pp. 477–487. doi:10.1145/3377325.3377500.

[4] Q. Yang, A. Steinfeld, C. Rosé, J. Zimmerman, Re-examining whether, why, and how human-AI interaction is uniquely difficult to design, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, ACM, 2020, pp. 174:1–174:13. doi:10.1145/3313831.3376301.

[5] J. S. Albus, Outline for a theory of intelligence, IEEE Transactions on Systems, Man, and Cybernetics 21 (1991) 473–509. doi:10.1109/21.97471.

[6] B. Shneiderman, P. Maes, Direct manipulation vs. interface agents, Interactions 4 (1997) 20. doi:10.1145/267505.267514.

[7] B. Shneiderman, Human-centered artificial intelligence: Three fresh ideas, AIS Transactions on Human-Computer Interaction 12 (2020) 109–124. doi:10.17705/1thci.00131.

[8] L. Bainbridge, Ironies of automation, Automatica 19 (1983) 775–779. doi:10.1016/0005-1098(83)90046-8.

[9] M. R. Endsley, From here to autonomy: Lessons learned from Human–Automation research, Human Factors: The Journal of the Human Factors and Ergonomics Society 59 (2017) 5–27. doi:10.1177/0018720816681350.

[10] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, M. Kankanhalli, Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, ACM, 2018, pp. 582:1–582:18. doi:10.1145/3173574.3174156.

[11] F. Doshi-Velez, B. Kim, Towards a rigorous science of interpretable machine learning, arXiv:1702.08608 [cs, stat] (2017). arXiv:1702.08608.

[12] Z. C. Lipton, The mythos of model interpretability, Queue 16 (2018) 31–57.

[13] A. Adadi, M. Berrada, Peeking inside the black-box: A survey on explainable artificial intelligence (XAI), IEEE Access 6 (2018) 52138–52160. doi:10.1109/ACCESS.2018.2870052.

[14] C. J. Cai, J. Jongejan, J. Holbrook, The effects of example-based explanations in a machine learning interface, in: Proceedings of the 24th International Conference on Intelligent User Interfaces, IUI '19, ACM, 2019, pp. 258–262. doi:10.1145/3301275.3302289.

[15] F. Yang, Z. Huang, J. Scholtz, D. L. Arendt, How do visual explanations foster end users' appropriate trust in machine learning?, in: Proceedings of the 25th International Conference on Intelligent User Interfaces, IUI '20, ACM, 2020, pp. 189–201. doi:10.1145/3377325.3377480.

[16] H.-F. Cheng, R. Wang, Z. Zhang, F. O'Connell, T. Gray, F. M. Harper, H. Zhu, Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, ACM, 2019, pp. 559:1–559:12. doi:10.1145/3290605.3300789.

[17] V. Lai, C. Tan, On human predictions with explanations and predictions of machine learning models: A case study on deception detection, in: Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19, ACM, 2019, pp. 29–38. doi:10.1145/3287560.3287590.

[18] E. Chu, D. Roy, J. Andreas, Are visual explanations useful? A case study in model-in-the-loop prediction, arXiv:2007.12248 [cs, stat] (2020).

arXiv:2007.12248.

[19] B. Green, Y. Chen, The principles and limits of algorithm-in-the-loop decision making, Proceedings of the ACM on Human-Computer Interaction 3 (2019) 50:1–50:24. doi:10.1145/3359152.

[20] Y. Alufaisan, L. R. Marusich, J. Z. Bakdash, Y. Zhou, M. Kantarcioglu, Does explainable artificial intelligence improve human decision-making?, arXiv:2006.11194 [cs, stat] (2020). arXiv:2006.11194.

[21] Y. Zhang, Q. V. Liao, R. K. E. Bellamy, Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making, in: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT* '20, ACM, 2020, pp. 295–305. doi:10.1145/3351095.3372852.

[22] F. Poursabzi-Sangdeh, D. G. Goldstein, J. M. Hofman, J. W. Vaughan, H. Wallach, Manipulating and measuring model interpretability, arXiv:1802.07810 [cs] (2019). arXiv:1802.07810.

[23] G. Bansal, T. Wu, J. Zhou, R. Fok, B. Nushi, E. Kamar, M. T. Ribeiro, D. S. Weld, Does the whole exceed its parts? The effect of AI explanations on complementary team performance, arXiv:2006.14779 [cs] (2020). arXiv:2006.14779.

[24] M. Eiband, D. Buschek, A. Kremer, H. Hussmann, The impact of placebic explanations on trust in intelligent systems, in: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, CHI EA '19, ACM, 2019, pp. LBW0243:1–LBW0243:6. doi:10.1145/3290607.3312787.

[25] R. R. Hoffman, M. Johnson, J. M. Bradshaw, A. Underbrink, Trust in automation, IEEE Intelligent Systems 28 (2013) 84–88. doi:10.1109/MIS.2013.24.

[26] C. J. Cai, M. C. Stumpe, M. Terry, E. Reif, N. Hegde, J. Hipp, B. Kim, D. Smilkov, M. Wattenberg, F. Viegas, G. S. Corrado, Human-centered tools for coping with imperfect algorithms during medical decision-making, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, ACM, 2019, pp. 4:1–4:14. doi:10.1145/3290605.3300234.

[27] T. Weber, H. Hußmann, Z. Han, S. Matthes, Y. Liu, Draw with me: Human-in-the-loop for image restoration, in: Proceedings of the 25th International Conference on Intelligent User Interfaces, IUI '20, ACM, 2020, pp. 243–253. doi:10.1145/3377325.3377509.

[28] J. Heer, Agency plus automation: Designing artificial intelligence into interactive systems, Proceedings of the National Academy of Sciences 116 (2019) 1844–1850. doi:10.1073/pnas.1807184115.

[29] J. J. Dudley, P. O. Kristensson, A review of user interface design for interactive machine learning, ACM Transactions on Interactive Intelligent Systems 8 (2018) 8:1–8:37. doi:10.1145/3185517.