Designing for Technology Transparency

Transparency Cues and User Experience

Ilka Hein
Department of Psychology,
Ludwig-Maximilians-Universität
München
Munich, Germany
ilka.hein@psy.lmu.de

Sarah Diefenbach
Department of Psychology,
Ludwig-Maximilians-Universität
München
Munich, Germany
sarah.diefenbach@psy.lmu.de

Daniel Ullrich
Department of Computer Science,
Ludwig-Maximilians-Universität
München
Munich, Germany
daniel.ullrich@ifi.lmu.de

ABSTRACT

As technologies become more complex, the question of how transparent they should be for users and how transparency cues should be designed comes to the fore. Transparency refers to the extent to which users learn, for example, how the technology works or arrives at certain results. The increased interest in this topic also stems from legal changes such as the debate about a European AI regulation, which demands transparent AI systems and thus necessitates solutions for an optimal design of transparency cues. The paper discusses examples and risks of lacking transparency and approaches and the state of knowledge for improving the user experience by technology-based transparency cues. Finally, we present an outlook on the promising directions for design guidelines and next steps of research.

KEYWORDS

Technology transparency, Explainable AI, User experience, Design approaches

1 INTRODUCTION

Can you tell how your smartphone recognizes your face, how your voice assistant communicates with you, or how you receive news targeted to your interests? The use of intelligent technologies has become part of our daily routines while the exact functioning of such technology often remains opaque to us. In general, with technological advances and the consequent increase in complexity of systems, the transparency of technologies decreases (e.g., [1–3]). Opaque systems progressively fail to provide their users with information about their processes, performance, intent, or plans (e.g., [4–6]), such as what the technology is currently doing, how it arrived at a particular outcome, or why (e.g., [7, 8]). Examples of opaque technologies range from well-established systems (e.g., chatbots, online recommendations) to more recent developments (e.g., smart home, autonomous vehicles). In comparison to such technological progresses, human skills do not increase at the same rate, leading to necessarily less transparent technologies even if transparency cues are not changed. Accordingly, transparency cues have to improve more and more to reverse or at least mitigate this trend.

Veröffentlicht durch die Gesellschaft für Informatik e.V. und die German UPA e.V. 2023 in T. Jackstädt, C. Hauri & J. Hinze (Hrsg.):
Mensch und Computer 2023 - Usability Professionals, 03.-06. September 2023, Rapperswil (SG)
 https://doi.org/10.18420/muc2023-up-448

2 RISKS OF TECHNOLOGY OPACITY

While the reasons for increasing technology opacity are comprehensible, it comes with certain risks and negative effects on users and their experiences. For example, they perceive opaque technologies as less easy to use or useful, trust them to a smaller extent, and are less willing to use them (e.g., [9–11]). In addition, misperceptions or incorrect decisions could arise from a lower level of understanding or incorrect mental models about how the technology works (e.g., [12, 13]). Furthermore, opacity can make it harder to detect system errors, making user intervention less likely [14]. This may be especially dangerous in safety-critical application areas such as autonomous driving. Accidents have been reported in which drivers fully relied on the autopilot and did not consider that they were only driving a level 2 system, that is, a solution that only supports the driver rather than operates completely autonomously [15]. Consequently, opacity may not only negatively impact user experience, but can also have far-reaching, potentially harmful effects. Legal consequences are also becoming increasingly likely due to current legislative developments (e.g., European AI Regulation, General Data Protection Regulation; [16, 17]). These considerations have also been picked up by technology companies such as Microsoft, who have defined transparency as a principle of responsible AI [18]. However, it should not go unmentioned that too much transparency (e.g., forced by endless tutorials or notification boxes) can also be detrimental and exceed the users' capacity for reception [19]. This could increase their mental load, impair the user experience, or lower the acceptance and use of the technology. Therefore, a critical question seems what a well-balanced degree of transparency is and whether there are particular aspects where transparency should be established (and others which can remain "invisible" to the user).

The following section presents approaches to and the state of knowledge on technology-based transparency cues. Both can serve as guidelines for designing transparency in diverse application domains. Afterwards, deficiencies of existing research findings are discussed, and an outlook is given.

3 APPROACHES TO AND STATE OF KNOWLEDGE ON TRANSPARENCY CUES

Transparency cues can be implemented and differentiated in many ways. Table 1 shows exemplary aspects that designers could consider when creating transparent technologies derived from the research literature, broadly structured along different levels. A first fundamental distinction is whether transparency aspects refer to the *what level* or the *how level* [20]: On the one hand, different

Table 1: Design Aspects for Transparency Cues

Aspect	Possible considerations
	What aspects
Content	Global or local information (e.g., [5, 27])
	Descriptive or explanatory information [27]
	What, how, or why information (e.g., [7, 8])
	Past-, present-, or future-related information (e.g., [23, 29])
	Input- or output-related information [23]
	Certainty, performance, or accuracy (e.g., [4, 23])
Level of detail	Amount or completeness of information (e.g., [20, 30])
	Abstract or technically detailed information (e.g., [31, 32])
Temporal change	Same or adapted content within time of use [7]
	How aspects
Format of presentation	Text-based (e.g., phrase, keyword) or multimedia cues (e.g., visualization, table; e.g., [21, 33])
Modality of presentation	Visual (e.g., text, illustrations), auditory (e.g., sounds, spoken text), or use of multiple modalities (e.g., [33, 34])
Timing	Before, during, or after use [34]
Way of provision	Automatic, adaptive, or invoked by user (e.g., [21, 23])
	Always available or event-based availability (e.g., [21, 23])
Interactivity	Presentation of information or dialogue with system (e.g., [7, 35])
Location of provision	Technology-based or external location [34]
	Location on display/ technology [20]
Salience of provision	Inobtrusive or prominent cues (e.g., [14, 36])
Wording	Complexity (e.g., [20, 37])
	Length (e.g., [33, 37])
Personalization	Availability of adjustable settings [33]

aspects of the technology's way of functioning can be made transparent (i.e., *what* does the technology disclose to the user?) and on the other hand, the same transparency information can be presented differently (i.e., *how* does the technology convey the disclosed information to the user?).

Referring to the what level, the content of the transparency cue has to be specified at the beginning of the design process [20]. Many researchers here referred to the distinction of global information about how the whole technology works and local information describing why the technology produced a certain result in a particular case. Whereas some argued that case-specific information leads to more effective learning and should be strived for [21], others stated that more general information serves as a better explanation [7] or that users prefer having both global and local cues [22]. Furthermore, scholars claimed, for instance, that it should not only be explained why an event happened but also why this event occurred instead of a different one (i.e., *contrastive* or *why-not explanation*; [7]). For instance, Miller [7] discussed the example of an animal classification system and posited that contrastive explanations ("Why is a particular image labelled as a beetle instead of a spider?") are also easier to derive than plain why explanations ("Why is a particular image labelled as a beetle?") because a smaller number of causes has to be indicated. While this is also controversial and may be context-specific [23], regarding the level of detail, most researchers agreed that simple explanations with less information should be aimed at (e.g., [7, 11, 20]). For example, Silva et al. [11] studied decision-making tools and found that users rated explanations with probability ratings for each decision option as less explainable than explanations consisting of one sentence with removed information. In this context, Miller [7] referred to the dilution effect expressing that additional information may not only increase the cognitive capacity for processing but also dilute the effects of more important information. As these examples illustrate, although many propositions for distinguishing and structuring transparency contents exist, concrete recommendations are largely not feasible because little and fragmentary research has been done on users' experiences of such transparency cues.

Once a decision has been made on the content of the transparency cue, how aspects move into focus [20]. Regarding the format of presentation, various studies underlined that users preferred visualizations over text-based cues (e.g., [24, 25]). As an example, Wastensteiner et al. [25] compared energy consumption feedback in the form of a bar or line diagram to that in the form of sentences. Contrary to the users' preference ratings, the authors also found that the text explanation was better understood, and Silva et al. [11] even showed that various presentation formats (e.g., text, decision trees, probabilities) did not affect trust perceptions or performances differently when using a decision-making system. While most empirical studies focused on comparing presentation formats, other how aspects of transparency cues, such as the presentation modality, are mainly discussed at a conceptual level. For instance, auditory cues are less explored than visual cues. This could be because visual hints are easier to design and more intuitively understandable than auditory ones. However, users may prefer the latter when they are engaged in other activities and therefore have less capacity to process written or illustrated information. Increasingly and probably due to technological advances, research is being conducted on how transparency cues should be provided (e.g., adaptive, automatic, or user-invoked) and on how interactive they should be (e.g., [7, 26]). For example, imagine receiving transparency cues only upon request and being able to pose questions about given cues rather than getting non-tailored and automatic information. In this regard, dialectical theories of explanation were mentioned assuming that explanations are a form of conversation (e.g., [27, 28]). However, many how questions have not yet been satisfactorily answered or even investigated.

4 DEFICIENCIES OF THE STATE OF KNOWLEDGE AND OUTLOOK

As outlined above, there is a wide variety of approaches describing possible design aspects to enhance technology transparency on different levels. At the same time, there are still many gaps and the current state of knowledge has major deficiencies that are to be summarized in the following.

First, under the heading of Explainable AI, existing research was mostly concerned with the technical implementation of transparent systems rather than the experience of users (e.g., [31, 33]). For decision support and recommender systems, Nunes and Jannach [33] found that only 21.5% of the existing studies evaluated their proposed techniques and tools. Thus, the psychological perspective on transparency cues has been given little attention. Yet, the intended effects of the designers and developers may be different from the users' actual experience of the transparency cue. For example, developers may mistakenly assume that users are able to understand a specific explanation because it is easily comprehensible to them as experts. Also, they might have a misconception of what may be relevant for the user to know or how users prefer the transparency cue to be presented. In this context, Miller et al. [38] used Cooper's [39] phrase "inmates running the asylum" meaning that developers may design software for themselves. Therefore, user studies should become an essential part of developing transparency cues. However, we rather propose that studies on user preferences should be conducted before as well as during and not only after designing transparency cues. Thereby, resources (e.g., time, work resources, money) could be saved and invested more effectively.

Second, consistent with the mentioned scarcity of user studies, measures for evaluating transparency cues are lacking (e.g., [11, 40]). Yet, standardized measurements would allow for dependable insights and reliable comparisons between different transparency cues. Silva et al. [11] also highlighted that proposed scales are not appropriate for lay users indicating that it is crucial to include potential users when constructing scales. In addition, various constructs were investigated to assess the quality of transparency cues (e.g., satisfaction, trust, performance; e.g., [5, 33]), further impeding the comparability of evaluation study results. Before concrete measures are developed, important criteria for transparency cues should therefore be defined and agreed upon.

Third, transparency cues were mostly developed for applications that are targeted to experts rather than lay users (e.g., [22, 41]). For instance, recommender systems for health professionals or admission specialists were featured with transparency information (e.g., [42, 43]). Although such domains are crucial to investigate, end users should also be taken into account. On the one hand, decisions of experts who used a recommender system might be more accepted by patients or applicants if they understand its reasoning. On the other hand, the focus on expert systems signals that other technologies are rather neglected in research. Yet, findings on one user group should not be transferred to other user groups because they may have different needs (e.g., [5, 44]). As such, experts and novices may require different transparency information because of their divergent level of knowledge (e.g., [44, 45]). Research findings also showed that interindividual differences influence the perception of transparency information. For example, users' extraversion, openness to experience as well as their technical affinity affected their willingness to spend time with transparency cues [26]. Thus, the experience of transparency cues should be examined with and compared between different user groups using a broader range of technologies.

Fourth, most research was not theory- or hypothesis-driven but rather exploratory (e.g., [32, 38]). Thus, no comprehensive theories or models exist that could guide the design of transparency cues. However, understanding the reasons why certain transparency notices lead to more favorable user perceptions and developing theories based on these reasons would enable more generalizable statements about the design of transparency notices. Current publications already suggested theories from diverse research areas – such as explanation, sensemaking, social attribution, and cognitive processing research – that could be applied to designing transparency cues (e.g., [7, 27]). Future studies should aim at empirically investigating and further developing these propositions for the new application domain or developing alternative theoretical approaches if needed.

5 CONCLUSION

Technology transparency is an issue of increasing importance in HCI research and practice. Yet, the current state of research shows a lack of consensus on how transparency cues should be designed [20, 33] and provides only limited foundations for practitioners [4, 46]. The present article can be seen as first step to address this shortcoming. On the one hand, the overview on the state of knowledge given in section 3 can serve as an initial guidance for usability professionals on which aspects could be considered when designing transparency cues. On the other hand, the analysis of deficiencies and next research steps in section 4 offers a basis for future studies by showing which issues need to be systematically investigated and which research gaps should be filled. In all of these steps, it is important to put the user at the center. Ultimately, transparency aims at happy, self-determined users – at least this is our vision of positive user experience.

ACKNOWLEDGMENTS

This research was funded by the German Research Foundation (DFG), projects PerforM and TransforM (425412993) as part of the Priority Program SPP2199 Scalable Interaction Paradigms for Pervasive Computing Environments.

REFERENCES

[1] Timo Jakobi, Gunnar Stevens, Nico Castelli, Corinna Ogonowski, Florian Schaub, Nils Vindice, Dave Randall, Peter Tolmie, and Volker Wulf. 2018. Evolving needs in IoT control and accountability: A longitudinal study on smart home intelligibility. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4, Article 171, 1–28. https://doi.org/10.1145/3287049

[2] Wojciech Samek, Thomas Wiegand, and Klaus-Robert Müller. 2017. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv: 1708.08296. Retrieved from https://doi.org/10.48550/arXiv.1708.08296

[3] Feiyu Xu, Hans Uszkoreit, Yangzhou Du, Wei Fan, Dongyan Zhao, and Jun Zhu. 2019. Explainable AI: A brief survey on history, research areas, approaches and challenges. In Tang, J., Kan, MY., Zhao, D., Li, S., & Zan, H. (Eds.), *Natural Language Processing and Chinese Computing* (pp. 563–574). Springer. https://doi.org/10.1007/978-3-030-32236-6_51

[4] Jessie Y. C. Chen, Katelyn Procci, Michael Boyce, Julia Wright, Andre Garcia, and Michael Barnes. 2014. *Situation awareness-based agent transparency*. Army Research Laboratory Report. Retrieved August 3, 2023 from https://apps.dtic.mil/sti/citations/AD1143367

[5] Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. 2021. A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Transactions on Interactive Intelligent Systems* 11, 3-4, Article 24, 1-45. https://doi.org/10.1145/3387166

[6] Eric S. Vorm and David J. Y. Combs. 2022. Integrating transparency, trust, and acceptance: The Intelligent Systems Technology Acceptance Model (ISTAM). *International Journal of Human–Computer Interaction* 38, 18–20, 1828–1845. http://dx.doi.org/10.1080/10447318.2022.2070107

[7] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

[8] Ribana Roscher, Bastian Bohn, Marco F. Duarte, and Jochen Garcke. 2020. Explainable machine learning for scientific insights and discoveries. *IEEE Access* 8, 42200–42216. http://dx.doi.org/10.1109/access.2020.2976199

[9] Donghee Shin. 2020. User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability. *Journal of Broadcasting & Electronic Media* 64, 4, 541-565. https://doi.org/10.1080/08838151.2020.1843357

[10] Donghee Shin, Bu Zhong, and Frank A. Biocca, 2020. Beyond user experience: What constitutes algorithmic experiences?. *International Journal of Information Management* 52, Article 102061. https://doi.org/10.1016/j.ijinfomgt.2019.102061

[11] Andrew Silva, Mariah Schrum, Erin Hedlund-Botti, Nakul Gopalan, and Matthew Gombolay. 2022. Explainable artificial intelligence: Evaluating the objective and subjective impacts of XAI on human-agent interaction. *International Journal of Human–Computer Interaction* 39, 7, 1390–1404. http://dx.doi.org/10.1080/10447318.2022.2101698

[12] Victoria Alonso and Paloma de la Puente. 2018. System transparency in shared autonomy: A mini review. *Frontiers in Neurorobotics* 12. http://dx.doi.org/10.3389/fnbot.2018.00083

[13] Jack Muramatsu and Wanda Pratt. 2001. Transparent queries: Investigation users' mental models of search engines. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '01)*, September 9-12, New Orleans, LA, USA. ACM, New York, NY, USA, 217–224. https://doi.org/10.1145/383952.383991

[14] Mica R. Endsley. 2017. From here to autonomy. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 59, 1, 5–27. http://dx.doi.org/10.1177/0018720816681350

[15] Faiz Siddiqui and Jeremy B. Merrill. 2023. 17 fatalities, 736 crashes: The shocking toll of Tesla's autopilot. Washington Post. Retrieved August 3, 2023 from https://www.washingtonpost.com/technology/2023/06/10/tesla-autopilot-crashes-elon-musk/

[16] European Parliament. 2023. MEPs ready to negotiate first-ever rules for safe and transparent AI. Retrieved August 3, 2023 from https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai

[17] Bryce Goodman and Seth Flaxman. 2017. European Union regulations on algorithmic decision making and a "right to explanation". *AI Magazine* 38, 3, 50–57. http://dx.doi.org/10.1609/aimag.v38i3.2741

[18] Microsoft. 2022. Microsoft responsible AI standard, v2: General requirements. Retrieved August 3, 2023 from https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf

[19] Sarah Diefenbach, Lara Christoforakos, Daniel Ullrich, and Andreas Butz. 2022. Invisible but understandable: In search of the sweet spot between technology invisibility and transparency in smart spaces and beyond. *Multimodal Technologies and Interaction* 6, 10, Article 95. http://dx.doi.org/10.3390/mti6100095

[20] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing transparency design into practice. In *23rd International Conference on Intelligent User Interfaces*, March 7-11, Tokyo, Japan. ACM, New York, NY, USA, 211–223. http://dx.doi.org/10.1145/3172944.3172961

[21] Shirley Gregor and Izak Benbasat. 1999. Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS Quarterly* 23, 4, 497-530. http://dx.doi.org/10.2307/249487

[22] Jiaxin Dai. 2022. *Scenario-based design: Exploring the effect of explainability on user's adoption intention of smart home lighting systems*. Ph.D. Dissertation. Eindhoven University of Technology. Retrieved from https://research.tue.nl/en/studentTheses/scenario-based-design

[23] Brian Y. Lim and Anind K. Dey. 2009. Assessing demand for intelligibility in context-aware applications. In *Proceedings of the 11th international conference on Ubiquitous computing*, September 30-October 3, Orlando, FL, USA. ACM, New York, NY, USA, 195–204. http://dx.doi.org/10.1145/1620545.1620576

[24] Jo Vermeulen, Kris Luyten, and Karin Coninx. 2012. Understanding complex environments with the feedforward torch. In Paternò, F., de Ruyter, B., Markopoulos, P., Santoro, C., van Loenen, E., & Luyten, K. (Eds.), *Ambient Intelligence* (pp. 312–319). Springer. http://dx.doi.org/10.1007/978-3-642-34898-3_22

[25] Jacqueline Wastensteiner, Tobias M. Weiss, Felix Haag, and Konstantin Hopf. 2022. Explainable AI for tailored electricity consumption feedback – An experimental evaluation of visualizations. arXiv: 2208.11408. Retrieved from https://arxiv.org/abs/2208.11408

[26] Katharina Weitz, Alexander Zellner, and Elisabeth André. 2022. What do end-users really want? Investigation of human-centered XAI for mobile health apps. arXiv: 2210.03506. Retrieved from https://arxiv.org/abs/2210.03506

[27] Shane T. Mueller, Robert R. Hoffman, William Clancey, Abigail Emrey, and Gary Klein. 2019. Explanation in human-AI systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI. arXiv: 1902.01876. Retrieved from https://arxiv.org/abs/1902.01876

[28] Douglas Walton. 2011. A dialogue system specification for explanation. *Synthese* 182, 3, 349–374. https://doi.org/10.1007/s11229-010-9745-z

[29] Victoria Bellotti and Keith Edwards. 2001. Intelligibility and accountability: Human considerations in context-aware systems. *Human–Computer Interaction* 16, 2–4, 193–212. http://dx.doi.org/10.1207/s15327051hci16234_05

[30] Todd Kulesza, Simone Stumpf, Margaret Burnett, Sherry Yang, Irwin Kwan, and Weng-Keen Wong. 2013. Too much, too little, or just right? Ways explanations impact end users' mental models. In *Proceedings of Symposium on Visual Languages and Human Centric Computing*, September 15-19, San Jose, CA, USA. IEEE, Piscataway, NJ, USA, 3-10. http://dx.doi.org/10.1109/vlhcc.2013.6645235

[31] Wojciech Samek and Klaus-Robert Müller. 2019. Towards explainable artificial intelligence. In Samek, W., Montavon, G., Vedaldi, A., Hansen, L., & Müller, KR. (Eds.), *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (pp. 5–22). Springer. https://doi.org/10.1007/978-3-030-28954-6_1

[32] Q. Vera Liao and Kush R. Varshney. 2022. Human-centered explainable AI (XAI): From algorithms to user experiences. arXiv: 2110.10790. Retrieved from https://arxiv.org/abs/2110.10790

[33] Ingrid Nunes and Dietmar Jannach. 2017. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27, 393–444. http://dx.doi.org/10.1007/s11257-017-9195-0

[34] Jo Vermeulen, Kris Luyten, and Karin Coninx. 2013. Intelligibility required: How to make us look smart again. In *Proceedings of the 10th Romanian Conference on Human-Computer Interaction*, September 2-3, Cluj-Napoca, Romania. Romanian HCI community. http://hdl.handle.net/1942/15998

[35] Prashan Madumal, Tim Miller, Liz Sonenberg, and Frank Vetere. 2019. A grounded interaction protocol for explainable artificial intelligence. arXiv: 1903.02409. Retrieved from https://doi.org/10.48550/arXiv.1903.02409

[36] Raja Parasuraman and Victor Riley. 1997. Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39, 2, 230–253. http://dx.doi.org/10.1518/001872097778543886

[37] Menaka Narayanan, Emily Chen, Jeffrey He, Been Kim, Sam Gershman, and Finale Doshi-Velez. 2018. How do humans understand explanations from machine learning systems? An evaluation of the human-interpretability of explanation. arXiv: 1802.00682. Retrieved from https://arxiv.org/abs/1802.00682

[38] Tim Miller, Piers Howe, and Liz Sonenberg. 2017. Explainable AI: Beware of inmates running the asylum or: How I learnt to stop worrying and love the social and behavioural sciences. arXiv: 1712.00547. Retrieved from https://arxiv.org/abs/1712.00547

[39] Alan Cooper. 2004. *The inmates are running the asylum: Why high-tech products drive us crazy and how to restore the sanity*. Sams.

[40] Michael Chromik and Martin Schuessler. 2020. A taxonomy for human subject evaluation of black-box explanations in XAI. In *Proceedings of the IUI workshop on Explainable Smart Systems and Algorithmic Transparency in Emerging Technologies (ExSS-ATEC'20)*, Cagliari, Italy. ACM, New York, NY, USA. Retrieved from http://www.mmi.ifi.lmu.de/pubdb/publications/pub/chromik2020iuiworkshop/chromik2020iuiworkshop.pdf

[41] Devleena Das, Yasutaka Nishimura, Rajan P. Vivek, Naoto Takeda, Sean T. Fish, Thomas Plötz, and Sonia Chernova. 2023. Explainable activity recognition for

smart home systems. *ACM Transactions on Interactive Intelligent Systems* 13, 2, 1–39. http://dx.doi.org/10.1145/3561533

[42] Hao-Fei Cheng, Ruotong Wang, Zheng Zhang, Fiona O'Connell, Terrance Gray, F. Maxwell Harper, and Haiyi Zhu. 2019. 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*, May 4-9, Glasgow, Scotland, UK. ACM, New York, NY, USA, 1-12. http://dx.doi.org/10.1145/3290605.3300789

[43] Yiming Zhang, Ying Weng, and Jonathan Lund. 2022. Applications of explainable artificial intelligence in diagnosis and surgery. *Diagnostics* 12, 2, 237. http://dx.doi.org/10.3390/diagnostics12020237

[44] Mireia Ribera and Agata Lapedriza. 2019. Can we do better explanations? A proposal of user-centered explainable AI. In *Joint Proceedings of the ACM IUI 2019 Workshops*, March 20, Los Angeles, CA, USA. ACM, New York, NY, USA, 7 pages. Retrieved from http://hdl.handle.net/10609/99643

[45] Upol Ehsan and Mark O. Riedl. 2020. Human-centered explainable AI: Towards a reflective sociotechnical approach. In Stephanidis, C., Kurosu, M., Degen, H., & Reinerman-Jones, L. (Eds.), *HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence* (pp. 449–466). Springer. http://dx.doi.org/10.1007/978-3-030-60117-1_33

[46] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing theory-driven user-centric explainable AI. In *2019 CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019)*, May 4–9, 2019, Glasgow, Scotland, UK. ACM, New York, NY, USA, 1-15. https://doi.org/10.1145/3290605.3300831