Designing and Evaluating Human-Centred AI Systems: Best-Practices from a Multidisciplinary View

Giuseppe Desolda[1], Andrea Esposito[1,*], Rosa Lanzilotti[1] and Antonio Piccinno[1]

[1]Department of Computer Science, University of Bari Aldo Moro, Via E. Orabona 4, 70125 Bari, Italy

Abstract

The rise of Human-Centred Artificial Intelligence (HCAI) highlights the growing need for AI systems that are not only functional but also ethical, transparent, and user-focused. Leveraging the results of a literature review, this paper briefly explores best practices and multidisciplinary methodologies for designing and evaluating HCAI systems. Drawing from fields such as Human-Computer Interaction (HCI), Software Engineering (SE), Machine Learning (ML), and Ethics, we provide a brief synthesis of design techniques, evaluation strategies, and fairness frameworks. We emphasize the importance of iterative, user-centered design processes and propose actionable insights for integrating interpretability, bias mitigation, and ethical considerations into HCAI systems. Despite these advancements, significant challenges persist, including the lack of tailored heuristics for evaluation and the complexity of ethical impact assessments. Addressing these gaps will require scalable evaluation methodologies and enhanced interdisciplinary collaboration.

Keywords

Human-Centered Artificial Intelligence, Design Guidelines, Evaluation Guidelines, Survey

1. Introduction

Nowadays, Artificial Intelligence (AI) is becoming a transformative force across various fields, starting from creative industries (for example, through image generation [1]), up to high-stakes domains like medicine (for example, through medical decision support systems and general diagnostic aids [2, 3]). While AI may benefit users' capabilities, its successful deployment depends on a well-designed human-AI interaction capable of fulfilling the specific users' needs [4]. For instance, medical professionals often struggle to trust black-box AI systems, as their lack of transparency and/or explainability poses risks to both patients and physicians themselves (for example, legal risks), limiting the potential adoption of AI in some real-world scenarios.

To tackle these challenges, the concept of Human-Centered Artificial Intelligence (HCAI) promises to be a beacon for future AI research. HCAI lays at the crossroad of AI and Human-Computer Interaction (HCI). It emphasizes usability, safety, trustworthiness, and ethical considerations to design successful AI systems [5]. Moving beyond "traditional" AI, HCAI for systems that augment human expertise, rather than replace it [6]. Leveraging HCAI techniques, designers can aim towards *proactive* systems that *symbiotically* collaborate with their end-users towards common goals [7, 5].

This shift, however, requires a multidisciplinary approach to AI system design, combining insights from HCI, Software Engineering (SE), Machine Learning (ML), and AI Ethics. While these disciplines contribute distinct methodologies and perspectives, their convergence introduces complexities typical of multidisciplinary work, such as inconsistent terminologies, knowledge gaps, and lack of standardized practices [5]. Leveraging the results of a previous multidisciplinary literature review [5], this article synthesizes the best practices and general methodologies to design and evaluate HCAI systems.

This article is thus structured as follows: Section 2 better introduces the concept of HCAI, Sections 3 and 4 provide an overview of methodologies for designing and evaluating HCAI systems (respectively),

Joint Proceedings of the ACM IUI Workshops 2025, March 24–27, 2025, Cagliari, Italy

*Corresponding author.

giuseppe.desolda@uniba.it (G. Desolda); andrea.esposito@uniba.it (A. Esposito); rosa.lanzilotti@uniba.it (R. Lanzilotti); antonio.piccinno@uniba.it (A. Piccinno)

0000-0001-9894-2116 (G. Desolda); 0000-0002-9536-3087 (A. Esposito); 0000-0002-2039-8162 (R. Lanzilotti); 0000-0003-1561-7073 (A. Piccinno)

Section 5 summarizes the main challenges, while Section 6 ends the article.

2. Background and Definitions

Since AI systems are increasingly integrated within critical domains, their design must consider technical performances and their alignment with human needs, values, and ethical standards. Although HCAI promises to guide towards such a goal, various difficulties arise. In fact, HCAI is not a monolithic research effort but results from a convergence of methodologies and insights from several disciplines [5]. Such disciplines, each with its own priorities and terminologies, contribute to the design of systems that can truly augment humans, ensuring safety, reliability, and transparency. To better guide the following discussion, we report a possible definition of HCAI systems by Desolda et al. [5]:

> HCAI systems are AI systems that are designed, developed, and evaluated by involving users in the process (i.e., through a Human-Centered Design process) to increase the performance and satisfaction of humans in specified tasks. HCAI systems, therefore, aim to be usable and useful for specified users, who might be described through a formal model, to achieve their goals in their context of use while being reliable, safe and trustworthy.

Since the multidisciplinary nature of HCAI is one of its defining features, creating HCAI systems requires the integration of methodologies and insights from various disciplines. The main disciplines that contribute to HCAI are HCI and AI, which provide the main cornerstone of the entire discipline. The former focuses on improving the usability and user experience of interactive systems. It contributes principles and methodologies for designing AI systems that align with human values by involving end-users throughout the whole lifecycle of AI (from its design to its evaluation and monitoring); similarly, AI provides the technical backbone of HCAI as it enables systems to perform tasks that require intelligence.

Although nowadays, the term AI is often employed as a synonym of ML and Deep Learning, as these are the most successful subfields of AI, we prefer to avoid generating additional confusion in the way these terms are employed. Therefore, throughout the rest of this article, we will refer to the general high-level concept of “intelligent” machines through the term “AI.” In contrast, we will refer to the statistical techniques to implement such intelligent behavior through learning with the term “ML.”

Strongly related to AI is the field of eXplainable Artificial Intelligence (XAI). Generally, it aims to ensure transparency in AI systems by creating models and methods that are either interpretable and understandable by humans, or that can be explained through post-hoc techniques. In the context of HCAI, it enhances trust and accountability, which are extremely valued, especially in high-risk domains.

Two additional disciplines involved in HCAI are SE and Ethics. The former provides the methodologies for building robust, scalable, and maintainable systems and guidelines to integrate AI into larger software ecosystems while upholding quality and reliability. A sub-field of SE, at the intersection with HCI, is End-User Development (EUD). EUD aims to empower users to customize or even create software artifacts without advanced programming knowledge. It is highly relevant in HCAI as it may democratize AI creation and may provide effective techniques to intervene in AI systems’ decision processes to deal with errors or to reconfigure them [8].

Finally, Ethics guides the design and development of AI to create fair, transparent, and accountable systems while ensuring alignment with societal values and norms and preventing biases and harmful consequences.

3. Design Techniques

Designing HCAI systems requires methodologies prioritizing usability, ethical alignment, and adaptability. As illustrated above, the multidisciplinary nature of HCAI draws from fields like HCI, AI, SE, and Ethics. This section outlines key design techniques identified in the literature, with practical examples and approaches for HCAI systems.

3.1. HCI Solutions

Human-Centred Design (HCD) serves as the foundation of HCAI system development, emphasizing iterative processes that integrate user feedback at every stage [5]. By actively involving users throughout the design lifecycle, HCD ensures that systems are not only functional but also aligned with user needs, preferences, and contexts.

A key technique in HCD is rapid prototyping, which allows designers to draft and evaluate user interface prototypes quickly, testing ideas efficiently and refining designs based on user interactions and observations.

Participatory design is a key HCD strategy that actively involves end-users as collaborators in the design process. This approach ensures systems accurately reflect user needs and preferences by incorporating their perspectives and expertise. Beyond participatory design, HCD offers specialized methodologies tailored to the unique challenges of HCAI. Value Sensitive Design, for instance, integrates ethical considerations and human values directly into design choices, ensuring that systems uphold principles such as fairness, transparency, and equity.

Other notable approaches include activity-centered design and contextual design. Activity-centered design prioritizes understanding and optimizing user tasks, creating systems that align with specific workflows or activities. This method is particularly useful in professional domains where task efficiency and accuracy are paramount. In contrast, contextual design employs observational techniques to study the social, environmental, and cultural factors influencing user interactions with technology. By focusing on the broader context of use, contextual design ensures that systems are adaptable and relevant to real-world scenarios.

3.2. AI and ML Solutions

AI and ML algorithms form the backbone of HCAI systems, offering diverse approaches to meet specific design requirements. These algorithms enable systems to learn, adapt, and operate effectively in dynamic and complex environments, making them indispensable for creating user-centered, intelligent solutions.

A useful technique is reinforcement learning, particularly suited for HCAI systems operating in unpredictable contexts [9, 10]. Reinforcement learning enables systems to learn optimal behaviors through trial-and-error interactions with their environment, making it ideal for applications requiring flexibility and adaptability.

Transfer learning provides a practical solution when labeled data is scarce by leveraging pre-trained models from related domains [11]. This approach allows HCAI systems to adapt existing knowledge to new tasks with minimal additional data, reducing development time and resource requirements. Transfer learning is particularly beneficial in specialized domains, such as healthcare, where labeled datasets are often small and expensive to produce [5].

Federated learning represents another useful approach for HCAI systems, especially those requiring personalization while preserving user privacy. Federated learning allows the development of a global model that can be fine-tuned by local client models, enabling personalized interactions without centralizing sensitive user data [12]. This capability is critical for applications like personalized healthcare monitoring or collaborative tools, where individual user preferences and privacy must be respected.

3.3. XAI Solutions

For HCAI systems, ensuring that end-users can interpret AI models and understand their decisions is essential to establish trust [5]. Trust in these systems is particularly critical in high-stakes domains, such as healthcare or finance, where decisions must be transparent and justifiable [13]. However, the current state of AI, mainly ML, often necessitates the use of "black-box" models, such as deep neural networks, whose internal decision-making processes are opaque to users [14].

In situations where black-box models are employed, post-hoc model-agnostic explanation techniques offer a versatile solution for making predictions understandable. These methods are applicable across a wide range of ML models, regardless of their complexity or application domain [14]. Two prominent examples of such techniques are LIME and SHAP. LIME provides local explanations by approximating the model's behavior around a specific prediction [15], while SHAP quantifies the importance of individual input features to the model's output [16]. These tools are widely used to make complex ML systems more accessible and interpretable to non-technical stakeholders.

In some cases, HCAI systems may leverage inherently interpretable models, such as decision trees or rule-based systems [14]. These models are inherently understandable by humans, as they provide clear, logical paths from inputs to outputs. When the use of interpretable models is feasible, they are generally preferred over black-box approaches, as they eliminate the need for post-hoc explanations, which can sometimes be incorrect or misleading. Moreover, relying on interpretable models mitigates risks associated with faithfulness and trustworthiness, which are limitations of many post-hoc explanation techniques [17].

3.4. Software Engineering Solutions

SE provides the principles, techniques, and methods necessary to develop robust, scalable, and maintainable HCAI systems. One of the main contributions of SE to HCAI development is Agile methodologies. Agile methodologies emphasize iterative development cycles and allow frequent user feedback, allowing designers to incorporate end-users' insights at every stage of the development process [18]. This iterative approach is especially valuable for HCAI, as it enables the continuous refinement of both usability and technical performance, ensuring that the system evolves in alignment with user needs and expectations [5].

DevOps practices can be adapted to further support the development of HCAI systems to enhance collaboration between development and operations teams [19]. A specialization of DevOps tailored for ML, called MLOps, allows for continuously integrating, delivering, and testing ML models [20]. MLOps facilitates the deployment of updated models into production environments while ensuring that changes are rigorously tested for consistency and reliability. This is particularly important for HCAI systems, where frequent updates are often required to improve accuracy, address biases, or adapt to changing user contexts [5].

3.5. Ethics Solutions

A fundamental principle of HCAI systems is developing and deploying systems that are ethical and uphold human values. Achieving this goal requires carefully integrating ethical principles throughout the system's lifecycle. Ethical frameworks and guidelines provide a valuable foundation for embedding fairness, transparency, and accountability into HCAI systems. These frameworks not only help align system behavior with societal norms but also aim to ensure that the systems operate responsibly in diverse and sensitive contexts [5].

To further support the ethical integrity of HCAI systems, a-priori techniques can address fairness and mitigate bias during the design and development stages. These proactive approaches, such as rebalancing datasets, introducing fairness constraints, and debiasing algorithms, enhance the reliability of AI systems by reducing the risk of producing skewed or discriminatory outputs [21]. Addressing potential ethical concerns early in the process, these techniques help build trustworthy and equitable systems, fostering greater acceptance and trust among end-users [5].

4. Evaluation Techniques

Once HCAI systems are designed and developed, rigorous evaluation techniques are essential to ensure the final system meets quality standards and adheres to the requirements of its intended domain.

4.1. HCI Solutions

HCI offers several methods for evaluating systems, focusing primarily on user-centered qualities such as usability and user experience. These methods can be broadly categorized into three main approaches: analytical evaluation (which does not involve direct user participation), usability testing, and in-the-wild studies [22].

Analytical evaluation relies on formal models or heuristics to assess the quality of a system without involving end-users. While this approach is cost-effective and can be applied early in the design process, it is limited by the availability of domain-specific heuristics. Unfortunately, few established heuristics exist for evaluating HCAI systems [5]. However, existing design guidelines, such as those proposed by Amershi et al. [23] and the Google PAIR team [24], can serve as starting points for evaluation purposes.

4.2. Software Engineering Solutions

Techniques such as SE verification and validation, model checking, and general software testing play a crucial role in assessing the technical robustness and functionality of the system, providing confidence in the reliability of the final product.

However, the multidisciplinary nature of HCAI introduces additional layers of complexity to the evaluation process. Quality measures for HCAI systems are inherently multifaceted, encompassing not only technical performance but also usability, transparency, fairness, and ethical alignment. Addressing these diverse dimensions requires employing methodologies drawn from fields such as HCI, AI, and ethics, ensuring a holistic approach to evaluation that aligns with the system's multidisciplinary foundations.

4.3. AI, ML, and XAI Solutions

In the context of HCAI systems, traditional AI and ML metrics, such as accuracy, precision, and recall, remain highly valued as they provide critical insights into the quality and reliability of the intelligent decision-making process. These metrics are essential for assessing how well the system performs its core tasks, ensuring that the underlying models meet the functional requirements of the application [5].

Beyond performance metrics, fairness and bias detection techniques are useful to evaluate HCAI systems, complementing traditional performance metrics as they help ensure equitable and non-discriminatory outcomes.

Since transparency is a fundamental pillar of HCAI, evaluating the interpretability and explainability of AI systems is imperative. Metrics for assessing interpretability focus on how easily users can understand the internal workings of a model [25]. In contrast, post-hoc explainability metrics measure the quality and faithfulness of explanations provided for specific decisions [14].

4.4. Ethics Solutions

One of the significant challenges in evaluating HCAI systems lies in their ethical assessment. Despite the critical importance of Ethics in HCAI, there is a lack of comprehensive and standardized guidelines to evaluate the ethical implications of these systems systematically. The most widely recognized approach to ethical evaluation is the use of Ethical Impact Assessments, a framework designed to analyze the ethical consequences of systems before, during, and after deployment [26].

Ethical Impact Assessments provide a thorough and structured process for identifying and addressing ethical concerns, such as fairness, transparency, and accountability. However, conducting ethical impact assessments can be labor- and resource-intensive, requiring significant expertise in ethics, law, and technology. These demands make ethical impact assessments particularly challenging to implement for organizations with limited resources or expertise [5].

An alternative approach is algorithmic auditing, which involves investigating the algorithms used in an AI system to identify potential ethical issues, such as bias or discrimination [27]. While less systematic than ethical impact assessments, algorithmic auditing is more accessible and easier to adopt.

It allows for targeted evaluations of specific components within a system, making it a practical solution for organizations seeking to address ethical concerns without undertaking the full scope of an ethical impact assessment [5].

5. Open Challenges

Designing and evaluating HCAI systems presents unique and multifaceted challenges. These arise from the complex intersection of usability, ethics, transparency, and the need for interdisciplinary collaboration. This section highlights the main open challenges that increase the difficulty of creating HCAI systems.

One of the main barriers to the effective and efficient evaluation of HCAI systems is the scarcity of standardized heuristics. While traditional usability heuristics exist for general software evaluation and are still helpful for HCAI, they fall short when evaluating qualities like transparency, fairness, and trustworthiness.

Developing a robust set of heuristics specific to HCAI would greatly simplify the evaluation process. Beyond evaluation, these heuristics could serve as a valuable design tool, much like those used in traditional software development. By offering guidance during the design phase, they can help designers anticipate and address potential issues proactively, effectively preventing mistakes before they occur [5].

As already stated, another pain point of HCAI systems evaluation is their ethical assessments. Since ethical impact assessments are resource-intensive and require specific expertise, this framework is often out of reach for smaller organizations or projects with limited resources.

Moreover, ethical impact assessments require extensive collaboration across multiple disciplines, which often exacerbates the inherent challenges of multidisciplinary work. One notable difficulty in HCAI projects is the lack of a shared vocabulary among experts from different fields, such as Computer Science, Ethics, Law, and Human-Computer Interaction [5]. These gaps in terminology can hinder effective communication and alignment, complicating the process of identifying and addressing ethical concerns [5].

6. Conclusions

HCAI systems promise to be effective but also ethical, trustworthy, and aligned with human values. By leveraging HCAI techniques, designers can create systems that *proactively* and *symbiotically* engage with their human end-users. Understanding user behaviors and needs (exploiting HCI techniques), guaranteeing quick update and redeployment strategies (leveraging SE), while defending users' rights and ensuring fairness and transparency (dictated by *ethics*), allows designers to create *acceptable* proactive and symbiotic AI systems.

However, creating such systems is a complex, multidisciplinary endeavor, requiring insights from fields as diverse as AI, HCI, SE, and ethics. This paper has provided an overview of some of the various disciplines' best practices and key methodologies, highlighting the role of human-centered design principles, evaluation techniques, and fairness frameworks in designing and developing HCAI systems.

The lack of tailored heuristics for evaluating HCAI systems hinders systematic and accessible assessments, while the resource-intensive nature of ethical impact assessments limits their adoption. Addressing these challenges will require developing scalable, domain-specific evaluation tools and methodologies alongside frameworks that facilitate interdisciplinary collaboration and shared understanding.

To bridge these gaps, the HCAI community must prioritize research into actionable and adaptable solutions, such as lightweight ethical evaluation methods and domain-specific heuristics. By doing so, we can ensure that HCAI systems not only meet technical and functional goals but also uphold ethical principles, fostering trust and user satisfaction across diverse applications.

Acknowledgments

We are grateful to Prof. Maria Francesca Costabile for her guidance and perspective on the work presented in this paper.

This research is partially supported by:

- the co-funding of the European Union - Next Generation EU: NRRP Initiative, Mission 4, Component 2, Investment 1.3 – Partnerships extended to universities, research centers, companies and research D.D. MUR n. 341 del 15.03.2022 – Next Generation EU (PE0000013 – "Future Artificial Intelligence Research – FAIR" - CUP: H97G22000210007);
- the Italian Ministry of University and Research (MUR) under grant PRIN 2022 "DevProDev: Profiling Software Developers for Developer-Centered Recommender Systems" – CUP: H53D23003620006;
- the Italian Ministry of University and Research (MUR) and by the European Union - NextGenerationEU, under grant PRIN 2022 PNRR "PROTECT: imPROving ciTizEn inClusiveness Through Conversational AI" (Grant P2022JJPBY) – CUP: H53D23008150001;
- the Italian Ministry of University and Research (MUR) and by the European Union - NextGenerationEU, under grant PRIN 2022 PNRR "DAMOCLES: Detection And Mitigation Of Cyber attacks that exploit human vuLnerabilitiES" (Grant P2022FXP5B) – CUP: H53D23008140001.

The research of Andrea Esposito is funded by a Ph.D. fellowship within the framework of the Italian "D.M. n. 352, April 9, 2022" - under the National Recovery and Resilience Plan, Mission 4, Component 2, Investment 3.3 - Ph.D. Project "Human-Centered Artificial Intelligence (HCAI) techniques for supporting end users interacting with AI systems", co-supported by "Eusoft S.r.l." (CUP H91I22000410007).

Declaration on Generative AI

During the preparation of this work, the author(s) used Grammarly in order to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, M. Chen, Hierarchical Text-Conditional Image Generation with CLIP Latents, 2022. URL: http://arxiv.org/abs/2204.06125. doi:10.48550/arXiv.2204.06125. arXiv:2204.06125.

[2] S. Cerri, O. Puonti, D. S. Meier, J. Wuerfel, M. Mühlau, H. R. Siebner, K. V. Leemput, A contrast-adaptive method for simultaneous whole-brain and lesion segmentation in multiple sclerosis, Neuroimage 225 (2020) 117471. URL: https://www.ncbi.nlm.nih.gov/pubmed/33099007. doi:10.1016/j.neuroimage.2020.117471. arXiv:33099007.

[3] G. Castellano, A. Esposito, E. Lella, G. Montanaro, G. Vessio, Automated detection of Alzheimer's disease: A multi-modal approach with 3D MRI and amyloid PET, Scientific Reports 14 (2024) 5210. URL: https://www.nature.com/articles/s41598-024-56001-9. doi:10.1038/s41598-024-56001-9, [Q1 Multidisciplinary].

[4] A. Esposito, To Automate or To Augment Humans? That is the Question for AI, and The Answer Depends on Users' Requirements, 2024. URL: https://ai.gopubby.com/to-automate-or-to-augment-humans-c3f805d35d55.

[5] G. Desolda, A. Esposito, R. Lanzilotti, A. Piccinno, M. F. Costabile, From human-centered to symbiotic artificial intelligence: A focus on medical applications, Multimedia Tools and Applications (2024). URL: https://rdcu.be/d1RF4. doi:10.1007/s11042-024-20414-5, [Q1 Media Technology].

[6] B. Shneiderman, Human-Centered AI, 1 ed., Oxford University Press, Oxford, 2022.

[7] S. S. Grigsby, Artificial Intelligence for Advanced Human-Machine Symbiosis, in: D. D. Schmorrow, C. M. Fidopiastis (Eds.), Augmented Cognition: Intelligent Technologies, volume 10915 of *Lecture Notes in Computer Science*, Springer International Publishing, Cham, 2018, pp. 255–266. URL: https://link.springer.com/10.1007/978-3-319-91470-1_22. doi:10.1007/978-3-319-91470-1_22.

[8] A. Schmidt, T. Herrmann, Intervention user interfaces: A new interaction paradigm for automated systems, Interactions 24 (2017) 40–45. URL: https://doi.org/10.1145/3121357. doi:10.1145/3121357.

[9] L. P. Kaelbling, M. L. Littman, A. W. Moore, Reinforcement Learning: A Survey, Journal of Artificial Intelligence Research 4 (1996) 237–285. URL: https://www.jair.org/index.php/jair/article/view/10166. doi:10.1613/jair.301.

[10] Y. Li, Deep Reinforcement Learning, 2018. URL: https://arxiv.org/abs/1810.06339. doi:10.48550/ARXIV.1810.06339. arXiv:1810.06339.

[11] L. Torrey, J. Shavlik, Transfer Learning:, in: E. S. Olivas, J. D. M. Guerrero, M. Martinez-Sober, J. R. Magdalena-Benedito, A. J. Serrano López (Eds.), Handbook of Research on Machine Learning Applications and Trends, IGI Global, 2010, pp. 242–264. URL: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-60566-766-9.ch011. doi:10.4018/978-1-60566-766-9.ch011.

[12] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, Y. Gao, A survey on federated learning, Knowledge-Based Systems 216 (2021) 106775. URL: https://linkinghub.elsevier.com/retrieve/pii/S0950705121000381. doi:10.1016/j.knosys.2021.106775.

[13] C. Combi, B. Amico, R. Bellazzi, A. Holzinger, J. H. Moore, M. Zitnik, J. H. Holmes, A Manifesto on Explainability for Artificial Intelligence in Medicine, Artificial Intelligence in Medicine (2022) 102423. URL: https://www.sciencedirect.com/science/article/pii/S0933365722001750. doi:10.1016/j.artmed.2022.102423.

[14] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A Survey of Methods for Explaining Black Box Models, ACM Computing Surveys 51 (2019) 1–42. URL: https://dl.acm.org/doi/10.1145/3236009. doi:10.1145/3236009.

[15] M. T. Ribeiro, S. Singh, C. Guestrin, "Why Should I Trust You?": Explaining the Predictions of Any Classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA, 2016, pp. 1135–1144. URL: https://dl.acm.org/doi/10.1145/2939672.2939778. doi:10.1145/2939672.2939778.

[16] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, pp. 4768–4777.

[17] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nature Machine Intelligence 1 (2019/05/01, 2019) 206–215. URL: https://doi.org/10.1038/s42256-019-0048-x. doi:10.1038/s42256-019-0048-x.

[18] H. Lieberman, F. Paternò, M. Klann, V. Wulf, End-User Development: An Emerging Paradigm, in: H. Lieberman, F. Paternò, V. Wulf (Eds.), End User Development, volume 9, Springer Netherlands, Dordrecht, 2006, pp. 1–8. URL: http://link.springer.com/10.1007/1-4020-5386-X_1. doi:10.1007/1-4020-5386-X_1.

[19] C. Ebert, G. Gallardo, J. Hernantes, N. Serrano, DevOps, IEEE Software 33 (2016) 94–100. URL: https://ieeexplore.ieee.org/document/7458761/. doi:10.1109/MS.2016.68.

[20] D. Kreuzberger, N. Kühl, S. Hirschl, Machine Learning Operations (MLOps): Overview, Definition, and Architecture, IEEE Access 11 (2023) 31866–31879. URL: https://ieeexplore.ieee.org/document/10081336/. doi:10.1109/ACCESS.2023.3262138.

[21] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, A. Galstyan, A Survey on Bias and Fairness in Machine Learning, ACM Computing Surveys 54 (2022) 1–35. URL: https://dl.acm.org/doi/10.1145/3457607. doi:10.1145/3457607.

[22] Y. Rogers, H. Sharp, J. Preece, Interaction Design: Beyond Human-Computer Interaction, 6 ed., John Wiley & Sons, Inc, Hoboken, 2023. Previous edition: 2019. - Includes bibliographical references and index. - Description based on CIP data; resource not viewed.

[23] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, E. Horvitz, Guidelines for human-AI interaction, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 1–13. URL: https://doi.org/10.1145/3290605.3300233. doi:10.1145/3290605.3300233, microsoft Guidelines.

[24] Google PAIR, People + AI guidebook, 2019. URL: https://pair.withgoogle.com/guidebook, updated May 18, 2021.
Updated May 18, 2021.

[25] D. V. Carvalho, E. M. Pereira, J. S. Cardoso, Machine Learning Interpretability: A Survey on Methods and Metrics, Electronics 8 (2019) 832. URL: https://www.mdpi.com/2079-9292/8/8/832. doi:10.3390/electronics8080832.

[26] A. Mantelero, AI and Big Data: A blueprint for a human rights, social and ethical impact assessment, Computer Law & Security Review 34 (2018) 754–772. URL: https://linkinghub.elsevier.com/retrieve/pii/S0267364918302012. doi:10.1016/j.clsr.2018.05.017.

[27] I. D. Raji, A. Smart, R. N. White, M. Mitchell, T. Gebru, B. Hutchinson, J. Smith-Loud, D. Theron, P. Barnes, Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing, in: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, ACM, Barcelona Spain, 2020, pp. 33–44. URL: https://dl.acm.org/doi/10.1145/3351095.3372873. doi:10.1145/3351095.3372873.