Towards a Praxis for Intercultural Ethics in Explainable AI

CHINASA T. OKOLO, Cornell University, United States

Explainable AI (XAI) is often promoted with the idea of helping users understand how machine learning models function and produce predictions. Still, most of these benefits are reserved for those with specialized domain knowledge, such as machine learning developers. Recent research has argued that making AI explainable can be a viable way of making AI more useful in real-world contexts, especially within low-resource domains in the Global South. While AI has transcended borders, a limited amount of work focuses on democratizing the concept of explainable AI to the "majority world", leaving much room to explore and develop new approaches within this space that cater to the distinct needs of users within culturally and socially-diverse regions. This article introduces the concept of an intercultural ethics approach to AI explainability. It examines how cultural nuances impact the adoption and use of technology, the factors that impede how technical concepts such as AI are explained, and how integrating an intercultural ethics approach in the development of XAI can improve user understanding and facilitate efficient usage of these methods.

1 INTRODUCTION

Recent research has argued that making AI explainable can make AI more useful in real-world contexts, especially within low-resource domains in the Global South [50]. Other research has also noted the concentration of AI research and development within the Western world [13, 48] and how AI systems primarily embed the cultural values and practices of people within these respective regions, alienating certain groups of users and causing cultural harm [52]. While the development, use, and implementation of AI has transcended borders, a limited amount of work focuses on democratizing the concept of explainable AI to the "majority world" [4], leaving much room to explore and develop new approaches within this space that cater to the distinct needs of users within this region.

This article introduces the concept of an intercultural ethics approach to AI explainability. It examines how cultural nuances impact the adoption and use of technology, the factors that impede how technical concepts such as AI are explained, and how integrating an intercultural ethics approach in the development of XAI can improve user understanding and facilitate efficient usage of these methods. We first discuss what it means to explain, reviewing relevant literature within explainable AI and introducing intercultural ethics. We then highlight barriers that impact explainable AI. Next, we introduce the concept of using intercultural ethics to inform AI explainability, outlining steps for researchers interested in leveraging this approach. We conclude the paper by reflecting on the prospect of an intercultural ethics approach to XAI, the limitations of such an approach, and potential areas to build upon this work.

2 BACKGROUND

2.1 What does it mean to explain?

The Cambridge dictionary defines the verb explain as a way "to make something clear or easy to understand by describing or giving information about it" [16]. Overall, explanations can make things clear and understandable by supplying context, information, and reasoning. Other ways to explain involve simplifying abstract concepts into more interpretable forms. When explaining, additional tools like analogies, diagrams, and examples can be used to convey information simply. To explain effectively, factors such as clear communication skills, understanding the explanation

needs of the target audience, and an ability to adjust communication skills to adapt to target audiences' needs are key. Adequate explanation allows knowledge to transfer clearly and comprehensively. The concept of explaining is present in various fields and has most recently begun to gain traction in the application of technical and human-centered methods in machine learning (ML) and artificial intelligence (AI).

2.1.1 Explainable AI. As a result of the growing need for developers and end users to manage AI systems and understand the decisions produced by these systems, researchers established the field of explainable AI (XAI). XAI aims to improve AI systems' transparency, interoperability, and trustworthiness while enhancing how they are deployed ethically and held accountable for the impact they may cause. XAI consists of methods that enable humans to understand the predictions made by machine learning models [7, 24] and can be present in the form of algorithms, toolkits, and libraries or non-technical forms such as frameworks and design guidelines. Early research in this field can be traced back to the 1980s in work by Buchanan and Shortliffe [11] providing explanations of infection diagnoses to doctors from a rule-based system and work by Mitchell et al. [38] using explanations to provide domain experts with ways to generalize ML systems. XAI methods work across a range of data types and use cases, being leveraged for models trained on images and videos in computer vision [28, 32, 47, 51, 58] to text and tabular data in natural language processing (NLP) [25, 27, 33, 54]. XAI methods can be classified as either pre-hoc explainability, where explanations are produced in tandem with model predictions, and post-hoc explainability, where explanations are produced separately from model predictions, with no effect on the predictive accuracy of the model as a whole [42, 63].

A primary goal of XAI is to provide human-interpretable reasoning for the predictions and decisions produced by AI systems. Explanations can impact how people use and trust AI systems. Still, issues of bias pervade these systems [8, 19, 45], and many existing XAI methods are only interpretable to a small subset of technically-savvy users. With this in mind, there is a need to broadly study existing use cases of XAI and develop novel methods that ensure end users from various backgrounds (educational, regional, cultural, etc.) can comprehensively understand AI decisions and use explanations to leverage AI in their respective work effectively. Over the past decade, research has emerged that uses human-centered methods to critically examine XAI [5, 18, 50], evaluate XAI with human users [14, 26, 29, 34, 44], and develop novel XAI methods [10, 17, 59]. This growing body of work in human-centered XAI shows much promise in improving current XAI approaches and creating new ones. Improving the interpretability of XAI can possibly enhance user confidence, reduce bias, improve the identification of prediction errors, and allow users to retain their autonomy when deciding whether or not to act on AI predictions. Additionally, leveraging methods from the humanities can help make progress toward more interpretable XAI, improving its utility for a diverse set of users in global contexts.

2.2 Intercultural ethics

Researchers should incorporate an intercultural ethics approach into AI development to understand the challenges posed by the widespread use of AI and shape novel approaches to AI explainability that account for a globally diverse user base. Evanoff introduces the concept of intercultural ethics and defines it as "the process by which people from different cultures negotiate the norms that will govern relations between them at a variety of levels, including the interpersonal, intergroup and international" [20–22]. More generally, intercultural ethics help guide human behavior when interacting with individuals or groups from cultures different from their own. In these respective interactions, cultural beliefs, practices, and perspectives are understood and respected, and ethical standards that are sensitive to cultural differences are applied.

In the case of explainable AI, applying intercultural ethics can improve the efficacy of explanations and make AI systems, in general, more useful for end users with low levels of AI knowledge. This paper introduces and examines the concept of an intercultural ethics approach to AI explainability to unpack the existing challenges of XAI and to promote human-centered approaches to model interpretability. This work aims to foster new ground for explainability approaches that cater to a wide range of users and improve how these techniques convey the complexity behind model decision-making.

3   BARRIERS TO EXPLANATIONS

In the case of technology development, use, and adoption, cultural nuances can have a substantial impact. Cultural nuances are demonstrated in how people in different cultures express their beliefs, customs, values, and behaviors. They affect many aspects of our lives, including how we communicate and interact with others in verbal and non-verbal ways, such as language, social cues, and gestures. If cultural nuances are understood, cross-cultural communication is more effective, improving the respect of cultural perspectives and practices, and preventing misunderstandings and conflict. Cultural nuances also affect the design and development of technology due to the expectations, preferences, and values of groups of people. These factors can also influence how people adopt and use technology due to social norms around trusting and actively relying on technical systems. Research within HCI and ICTD has demonstrated that there exist cultural differences regarding privacy [43], technology use [31], interface design [6], and algorithmic fairness [55], which could impact how ethical values are embedded into technical products. For example, in cultures where collective interests are prioritized over individual privacy and autonomy, device sharing amongst multiple users is common [3, 12, 57], but technical interfaces that account for this type of usage in such contexts are not the norm despite the many challenges noted by researchers [2, 56]. With this in mind, as XAI becomes a popular choice for technology developers and user interface designers to integrate within the systems they develop, they should make more effort to be aware of the cultural nuances that could affect the use of these methods. By actively working to understand these nuances, technology will become accessible, effective, and socially responsible for all users, regardless of their respective backgrounds.

When explaining technical concepts, there exists a variety of barriers that can prevent users from interpreting such information. Technical concepts may also be abstract, containing convoluted theories or ideas that may be challenging for individuals without sufficient background knowledge. Such concepts may also involve technical jargon unfamiliar to users, exacerbating this gap. While the goal of AI explainability has generally been to improve transparency regarding the function of and predictions from AI systems, these benefits are often limited to those with technical backgrounds. The effectiveness of AI explanations is often dependent on factors, including the complexity of the system, the quality of the explanations, and the technical expertise of end users. There are also cases where explanations may be of little or no benefit to users, particularly in situations where interpreting an explanation takes more mental effort than receiving a decision or prediction from an AI system. Miller [37] finds that explanations help improve user understanding of AI but can be counterproductive if they are too complex for users to understand. Explanations are not always complete and can be misleading, limiting their effectiveness for end users. Mittelstadt et al. find that in cases where explanations are technically correct, their respective lack of interpretability by end users can mislead users in understanding functions about AI behavior [39]. Furthermore, if users do not trust AI systems or the predictions produced by them, the utility of XAI is also reduced. Some work [14, 35, 60, 64] finds that while explanations help improve user understanding of AI, the presence of explanations does not always result in enhanced trust in AI systems. Trust is often influenced by users' perceptions of the accuracy and consistency of AI systems and their prior experience interacting with AI systems.

Currently, there exist many barriers to effective XAI. While much technical progress has been made in developing XAI solutions, these systems often do not incorporate findings of human-centered XAI research that advocate for easier interpretability, lower complexity, and improved transparency. Additionally, many of these systems do not account for technology literacy and cross-cultural differences, especially when it comes to nuances that occur when explaining concepts to end users from locales outside of the West and those with varying levels of technical knowledge. However, prioritizing participatory design and introducing concepts such as intercultural ethics can help address some of these concerns. In the next section, we'll introduce the concept of using an intercultural ethics approach to develop better XAI methods while leveraging participatory methods commonly practiced in fields such as human-computer interaction (HCI) and information and communication technologies for development (ICTD).

4 USING INTERCULTURAL ETHICS TO INFORM XAI

To initiate an intercultural ethics approach to XAI, researchers should consider how varying cultural perspectives and values shape the design, development, and implementation of AI systems. There should also be considerations focusing on such perspectives and values that can inform the transparency and accountability of these systems to serve the needs of a diverse set of stakeholders. When we consider the development of XAI from an intercultural ethics approach, researchers should recognize that different cultures have varying expectations for factors such as accountability, transparency, and privacy and that other factors such as trust, power dynamics, and historical experiences shape these expectations. As researchers understand users' expectations, they should also consider how end users in particular contexts value certain aspects of AI systems, such as accuracy and efficiency, and what oversight and control users expect to have over AI decision-making. With this in mind, we outline steps for researchers interested in incorporating intercultural ethics into the development of XAI.

Engaging with diverse stakeholders through participatory design First, participatory methods should be leveraged by engaging with diverse stakeholders. By engaging end users from different cultural, religious, educational, and professional backgrounds, researchers can expect to gain a broad understanding of the expectations and concerns these users have regarding XAI. Researchers will also gain an understanding of user needs, perspectives, and values. These engagement methods can include co-creation workshops, focus groups, interviews, and user feedback sessions. All of these methods are useful when aiming to tailor XAI to the needs and contexts of specific users. Co-creation workshops bring together stakeholders such as end users, designers, researchers, and developers to generate ideas and design concepts collaboratively. In these workshops, stakeholders can also participate in interactive activities to provide deeper feedback on prototypes developed in these sessions. As these prototypes evolve, hosting focus groups and conducting interviews [1, 23, 30] can be helpful for XAI practitioners to get insights on early iterations of the intended product. User feedback sessions involving members of local communities, especially those from historically marginalized backgrounds, can be further used to solicit feedback on the design and functionality of more advanced prototypes to help progress toward a final product. These sessions can also leverage methods such as interviews, surveys, and usability testing to help practitioners identify areas of improvement. Researchers can create XAI systems that account for user needs and incorporate diverse cultural perspectives by incorporating these participatory methods within their development practices.

Studying cultural values and perspectives As researchers engage in these participatory design methods, there will be a need to separately study how cultural concepts such as accountability, fairness, privacy, and trust are encoded in the respective context. Early work in this area by Sambasivan et al. [55] has examined algorithmic fairness in the Indian context, but there has not been much work focused on other regions of the Global South. By extensively studying

these factors, researchers can better understand how these cultural nuances can impact the design and development of XAI. Additionally, incorporating local knowledge can provide additional context to XAI systems, making them much more relevant and valuable in places where they are used. There is much work needed to democratize access to AI, and efforts that focus on translating technical jargon into local languages, incorporating cultural nuances, customs, and practices into AI development, increasing training opportunities for AI practitioners, and developing the capacity to support local data collection, data sharing, and cloud storage will be essential to broadening inclusion within AI, eventually impacting how underrepresented users and regions are represented within XAI [49]. Such work cannot be conducted in silos, and AI practitioners will have to intentionally form interdisciplinary collaborations with researchers based in local contexts, researchers with experiences studying cultural nuances, and researchers with domain expertise in the unique interplay between culture and technology. These practices can help avoid "cultural incongruencies", as coined by Prabhakaran et al. [52], where the development and usage of AI systems cause cultural barriers, impose hegemonic classifications, cause safety gaps, violate cultural values, and lead to cultural erasure.

Addressing power dynamics to increase transparency and accountability As researchers design and develop XAI systems, they should be aware of the power dynamics between them and local communities and work meaningfully to address them. Additionally, researchers must be mindful of power dynamics between certain groups of people in local communities, understanding how social factors such as age, caste, gender, and religion impact communities. To help address power dynamics between themselves and local communities, researchers should implement mechanisms for stakeholder feedback and active input into the design, development, and evaluation processes. Additionally, as explainability becomes incorporated into AI systems, researchers should consider how existing power dynamics may impact the interpretation and use of XAI by local communities. This section has previously highlighted various participatory design methods that practitioners can leverage, which can help address issues of power dynamics. Using these methods, researchers can work towards inclusive XAI systems that prioritize transparency and accountability in their decision-making processes. To ensure these systems are responsive to stakeholder concerns and expectations, practitioners can also implement mechanisms to audit and monitor XAI systems. XAI can augment standard auditing practices in AI [15, 53] by ensuring that the outcomes of audits are interpretable to stakeholders across a spectrum of technical knowledge. AI has been shown to exacerbate disability-based discrimination [41], and XAI can further amplify such discrimination if researchers fail to incorporate inclusive measures within AI development. By addressing power dynamics, prioritizing transparency and accountability, and leveraging accessibility measures in HCI and AI [46, 61, 62], practitioners can make XAI more accessible to populations within the Global South who have limited experience with digital technologies and low exposure to AI.

Conducting human-centered evaluations Prior work within XAI has advocated for robust evaluations of XAI in low-resource contexts, especially those situated within the Global South [50]. Low-resource contexts, regardless of location, provide a variety of challenges that can hamper the effective integration and use of XAI in these domains. Extenuating factors include a lack of data sources, limited internet connectivity, low AI/digital literacy, smaller concentrations of AI practitioners, and a shortage of computing resources needed to train and evaluate these systems for real-world use. In the Global South, these factors can be compounded by cultural nuances and societal structures that set different expectations for who engages with, uses, and governs technology. To proactively understand and address these challenges, researchers should rigorously test XAI in real-world settings to ensure they are effective, fair, and reliable. Our advocacy for leveraging participatory methods within the design process also translates to the evaluation process. To create effective environments for evaluating XAI, practitioners should also intentionally partner with local organizations and institutions to conduct user testing and evaluations. Collaborations with such partners leverages

their respective domain expertise and experience interacting with local communities and can also help quickly identify issues for researchers to address.

5 DISCUSSION

Comprehensively, an intercultural ethics approach to XAI promotes diverse cultural perspectives and values while actively working to incorporate these considerations into the design, development, and deployment of AI systems. As research within the field of human-centered XAI continues to expand, practitioners can take advantage of the recommendations provided within this paper to develop effective and equitable systems that meet the needs and expectations of users. However, these recommendations are not to be taken without examining the risks that XAI and AI, in general, pose to communities. Although AI has the potential to provide many benefits to communities, the harms of these technologies have rightly overshadowed this potential. Introducing XAI into flawed AI systems can exacerbate these ills, resulting in a detrimental user experience that harms user agency, violates cultural and personal values, and decreases trust in AI systems. In cases where (X)AI systems are used for high-stakes decisions in domains like healthcare or lawmaking, negative impacts made by AI-assisted decision-making tools could be irrevocable, permanently harming people subjected to these systems. With this in mind, AI practitioners should critically reflect on their role as researchers, designers, and developers to understand their motivations for developing XAI and to what ends it serves their interests over those of populations who will be affected by the systems as they reach the real world.

Limitations We posit an intercultural ethics approach to XAI as only part of a solution for AI practitioners to improve XAI development. However, we acknowledge the existing challenges and barriers in AI research and the implications of using (X)AI to address societal problems. AI should not be seen as the solution to all issues, and feasible approaches that don't leverage AI should also be considered as practitioners outline approaches to their research problems. While this paper primarily focuses on XAI, intercultural ethics can be applied to other domains of artificial intelligence, machine learning, and even technology as a whole. We encourage researchers interested in building upon this work, which we characterize as *intercultural XAI*, to apply our guidelines in various contexts and experiment with augmentations to improve the robustness of this approach and its applicability to domains outside of AI.

6 THE FUTURE OF INTERCULTURAL XAI

An intercultural ethics approach to XAI is only one approach to improving the cultural competency of XAI methods. As researchers continue building and evaluating XAI methods, this research may inspire other avenues of exploration. Identifying complementary approaches to intercultural ethics is a promising area of further research, as Prabhakaran et al. state: "culture is not a static variable that can be easily encoded into a technology, rather a complex and dynamic system that is constantly being transmitted and transformed" [52]. We encourage parallel and tangential approaches to our work and look forward to this subdomain growing even further. We end this article outlining a set of questions for researchers to contemplate as they assess the prospect of integrating intercultural ethics:

- **How can we leverage intercultural XAI to examine best practices for incorporating cultural values and perspectives into XAI development?** To understand what intercultural XAI looks like in practice, interdisciplinary collaborations that leverage the work presented in this paper could produce novel findings, providing rich contributions to various fields, including computing, digital humanities, sociology, and anthropology.
- **What lessons can (X)AI practitioners learn from practices of colonial exploitation on modern-day AI to improve their research and development practices?** To complement the literature on decolonial AI

[9, 36, 40], researchers can examine these perspectives in tandem with intercultural XAI to understand better practices to foster inclusion in the development pipelines of AI systems and develop more user-interpretable methods.

- **What impact does intercultural XAI have on the efficacy and adoption of XAI?** As XAI continues to be developed and reconstructed, evaluation is essential to this process. To understand the value of intercultural XAI, researchers may consider developing novel evaluation metrics that provide sufficient context to model performance and user experience.

7 CONCLUSION

This paper introduces the concept of examining AI explainability through the lens of intercultural ethics. We highlight current work in XAI, noting the significant progress made towards improving interpretation methods, discuss current barriers to explaining AI, and outline ways to leverage intercultural ethics to inform the design, development, and implementation of XAI. We acknowledge that XAI is not the panacea to completely address the entirety of issues plaguing AI and understand that there are use cases where explanations are less helpful or not needed for a respective context. By considering and implementing an intercultural ethics approach to XAI development, we hope that researchers can improve the efficiency and inclusivity of XAI while ensuring that their tools serve the needs of a diverse user base. As AI continues to evolve as a field and transcend borders, novel approaches such as these will be needed to help mitigate negative impacts and broaden the potential of AI as a force for good.

REFERENCES

[1] Anne Adams, Peter Lunt, and Paul Cairns. 2008. *A qualititative approach to HCI research.* Cambridge University Press.

[2] Syed Ishtiaque Ahmed, Md Romael Haque, Jay Chen, and Nicola Dell. 2017. Digital privacy challenges with shared mobile phone use in Bangladesh. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–20.

[3] Syed Ishtiaque Ahmed, Steven J Jackson, Maruf Zaber, Mehrab Bin Morshed, Md Habibullah Bin Ismail, and Sharmin Afrose. 2013. Ecologies of use and design: individual and social practices of mobile phone use within low-literate rickshawpuller communities in urban Bangladesh. In *Proceedings of the 4th Annual Symposium on Computing for Development.* 1–10.

[4] Shahidul Alam. 2008. Majority world: Challenging the West's rhetoric of democracy. *Amerasia Journal* 34, 1 (2008), 88–98.

[5] Kiana Alikhademi, Brianna Richardson, Emma Drobina, and Juan E Gilbert. 2021. Can explainable AI explain unfairness? A framework for evaluating explainable AI. *arXiv preprint arXiv:2106.07483* (2021).

[6] Jasem M Alostath, Sanaa Almoumen, and Ahmad B Alostath. 2009. Identifying and measuring cultural differences in cross-cultural user-interface design. In *Internationalization, Design and Global Development: Third International Conference, IDGD 2009, Held as Part of HCI International 2009, San Diego, CA, USA, July 19-24, 2009. Proceedings 3.* Springer, 3–12.

[7] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58 (2020), 82–115.

[8] Ruha Benjamin. 2019. *Race after technology: Abolitionist tools for the new jim code.* Polity.

[9] Abeba Birhane. 2020. Algorithmic colonization of Africa. *SCRIPTed* 17 (2020), 389.

[10] Joost Broekens, Maaike Harbers, Koen Hindriks, Karel Van Den Bosch, Catholijn Jonker, and John-Jules Meyer. 2010. Do you get it? User-evaluated explainable BDI agents. In *Multiagent System Technologies: 8th German Conference, MATES 2010, Leipzig, Germany, September 27-29, 2010. Proceedings 8.* Springer, 28–39.

[11] Bruce G Buchanan and Edward H Shortliffe. 1984. Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project. (1984).

[12] Jenna Burrell. 2010. Evaluating Shared Access: social equality and the circulation of mobile phones in rural Uganda. *Journal of computer-mediated communication* 15, 2 (2010), 230–250.

[13] Alan Chan, Chinasa T Okolo, Zachary Terner, and Angelina Wang. 2021. The limits of global inclusion in AI development. *arXiv preprint arXiv:2102.01265* (2021).

[14] 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.*

1–12.

[15] Jennifer Cobbe, Michelle Seng Ah Lee, and Jatinder Singh. 2021. Reviewable automated decision-making: A framework for accountable algorithmic systems. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 598–609.

[16] Cambridge Dictionary. 2023. Retrieved April 8, 2023 from https://dictionary.cambridge.org/us/dictionary/english/explain

[17] Upol Ehsan, Pradyumna Tambwekar, Larry Chan, Brent Harrison, and Mark O Riedl. 2019. Automated rationale generation: a technique for explainable AI and its effects on human perceptions. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 263–274.

[18] Upol Ehsan, Philipp Wintersberger, Q Vera Liao, Elizabeth Anne Watkins, Carina Manger, Hal Daumé III, Andreas Riener, and Mark O Riedl. 2022. Human-Centered Explainable AI (HCXAI): beyond opening the black-box of AI. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7.

[19] Virginia Eubanks. 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

[20] Richard Evanoff. 1996. Intercultural Ethics: New Ways of Learning to Get Along with Each Other. *Language and Culture in International Communication* (1996), 145–225.

[21] Richard Evanoff. 2006. Integration in intercultural ethics. *International Journal of Intercultural Relations* 30, 4 (2006), 421–437.

[22] Richard Evanoff. 2020. Introducing intercultural ethics. *The Cambridge Handbook of Intercultural Communication* (2020), 187–202.

[23] Joel Fredericks, Glenda Amayo Caldwell, and Martin Tomitsch. 2016. Middle-out design: collaborative community engagement in urban HCI. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*. 200–204.

[24] Hani Hagras. 2018. Toward human-understandable, explainable AI. *Computer* 51, 9 (2018), 28–36.

[25] David Harbecke, Robert Schwarzenberg, and Christoph Alt. 2018. Learning Explanations from Language Data. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. 316–318.

[26] Sérgio Jesus, Catarina Belém, Vladimir Balayan, João Bento, Pedro Saleiro, Pedro Bizarro, and João Gama. 2021. How can I choose an explainer? An application-grounded evaluation of post-hoc explanations. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 805–815.

[27] Yichen Jiang, Nitish Joshi, Yen-Chun Chen, and Mohit Bansal. 2019. Explore, Propose, and Assemble: An Interpretable Model for Multi-Hop Reading Comprehension. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2714–2725.

[28] Hyungsik Jung and Youngrock Oh. 2021. Towards better explanations of class activation mapping. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1336–1344.

[29] Sunnie SY Kim, Nicole Meister, Vikram V Ramaswamy, Ruth Fong, and Olga Russakovsky. 2022. Hive: evaluating the human interpretability of visual explanations. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XII*. Springer, 280–298.

[30] Ugur Kuter and Cemal Yilmaz. 2001. Survey methods: Questionnaires and interviews. *Choosing Human-Computer Interaction (HCI) Appropriate Research Methods* (2001).

[31] Leantros Kyriakoullis and Panayiotis Zaphiris. 2016. Culture and HCI: a review of recent cultural studies in HCI and social networks. *Universal Access in the Information Society* 15 (2016), 629–642.

[32] Kwang Hee Lee, Chaewon Park, Junghyun Oh, and Nojun Kwak. 2021. LFI-CAM: learning feature importance for better visual explanation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1355–1363.

[33] Jiwei Li, Xinlei Chen, Eduard Hovy, and Dan Jurafsky. 2016. Visualizing and Understanding Neural Models in NLP. In *Proceedings of NAACL-HLT*. 681–691.

[34] Q Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[35] Brian Y Lim, Anind K Dey, and Daniel Avrahami. 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2119–2128.

[36] Sábëlo Mhlambi and Simona Tiribelli. 2023. Decolonizing AI ethics: relational autonomy as a means to counter AI harms. *Topoi* (2023), 1–14.

[37] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* 267 (2019), 1–38.

[38] Tom M Mitchell, Richard M Keller, and Smadar T Kedar-Cabelli. 1986. Explanation-based generalization: A unifying view. *Machine learning* 1, 1 (1986), 47–80.

[39] Brent Mittelstadt, Chris Russell, and Sandra Wachter. 2019. Explaining explanations in AI. In *Proceedings of the conference on fairness, accountability, and transparency*. 279–288.

[40] Shakir Mohamed, Marie-Therese Png, and William Isaac. 2020. Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology* 33 (2020), 659–684.

[41] Meredith Ringel Morris. 2020. AI and accessibility. *Commun. ACM* 63, 6 (2020), 35–37.

[42] W James Murdoch, Chandan Singh, Karl Kumbier, Reza Abbasi-Asl, and Bin Yu. 2019. Interpretable machine learning: definitions, methods, and applications. *arXiv preprint arXiv:1901.04592* (2019).

[43] Sheza Naveed, Hamza Naveed, Mobin Javed, and Maryam Mustafa. 2022. "Ask this from the person who has private stuff": Privacy Perceptions, Behaviours and Beliefs Beyond WEIRD. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.

[44] Giang Nguyen, Mohammad Reza Taesiri, and Anh Nguyen. 2022. Visual correspondence-based explanations improve AI robustness and human-AI team accuracy. *arXiv preprint arXiv:2208.00780* 1 (2022).

[45] Safiya Umoja Noble. 2018. *Algorithms of oppression*. New York University Press.

[46] Zeljko Obrenovic, Julio Abascal, and Dusan Starcevic. 2007. Universal accessibility as a multimodal design issue. *Commun. ACM* 50, 5 (2007), 83–88.

[47] Youngrock Oh, Hyungsik Jung, Jeonghyung Park, and Min Soo Kim. 2021. EVET: enhancing visual explanations of deep neural networks using image transformations. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 3579–3587.

[48] Chinasa T Okolo. 2023. Addressing Global Inequity in AI Development. *Handbook of Critical Studies of Artificial Intelligence*.

[49] Chinasa T Okolo, Dhruv Agarwal, Nicola Dell, and Aditya Vashistha. 2023. "If it is easy to understand then it will have value": Examining Perceptions of Explainable AI with Community Health Workers in Rural India. In *Preprint*.

[50] Chinasa T Okolo, Nicola Dell, and Aditya Vashistha. 2022. Making AI explainable in the global south: A systematic review. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS)*. 439–452.

[51] Badri N Patro, Mayank Lunayach, Shivansh Patel, and Vinay P Namboodiri. 2019. U-cam: Visual explanation using uncertainty based class activation maps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7444–7453.

[52] Vinodkumar Prabhakaran, Rida Qadri, and Ben Hutchinson. 2022. Cultural Incongruencies in Artificial Intelligence. *arXiv preprint arXiv:2211.13069* (2022).

[53] Inioluwa Deborah Raji, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. 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*. 33–44.

[54] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2018. Anchors: High-precision model-agnostic explanations. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.

[55] Nithya Sambasivan, Erin Arnesen, Ben Hutchinson, Tulsee Doshi, and Vinodkumar Prabhakaran. 2021. Re-imagining algorithmic fairness in india and beyond. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 315–328.

[56] Nithya Sambasivan, Garen Checkley, Amna Batool, Nova Ahmed, David Nemer, Laura Sanely Gaytán-Lugo, Tara Matthews, Sunny Consolvo, and Elizabeth F Churchill. 2018. " Privacy is not for me, it's for those rich women": Performative Privacy Practices on Mobile Phones by Women in South Asia.. In *SOUPS@ USENIX Security Symposium*. 127–142.

[57] Nithya Sambasivan, Ed Cutrell, Kentaro Toyama, and Bonnie Nardi. 2010. Intermediated technology use in developing communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2583–2592.

[58] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*. 618–626.

[59] Hong Shen, Haojian Jin, Ángel Alexander Cabrera, Adam Perer, Haiyi Zhu, and Jason I Hong. 2020. Designing alternative representations of confusion matrices to support non-expert public understanding of algorithm performance. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–22.

[60] Donghee Shin. 2021. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies* 146 (2021), 102551.

[61] Constantine Stephanidis, Demosthenes Akoumianakis, Michael Sfyrakis, and Alexandros Paramythis. 1998. Universal accessibility in HCI: Process-oriented design guidelines and tool requirements. In *Proceedings of the 4th ERCIM Workshop on User Interfaces for all, Stockholm, Sweden*. 19–21.

[62] Chathurika S Wickramasinghe, Daniel L Marino, Javier Grandio, and Milos Manic. 2020. Trustworthy AI development guidelines for human system interaction. In *2020 13th International Conference on Human System Interaction (HSI)*. IEEE, 130–136.

[63] Zhichao Xu, Yi Han, Tao Yang, Anh Tran, and Qingyao Ai. 2022. Learning to Rank Rationales for Explainable Recommendation. *arXiv preprint arXiv:2206.05368* (2022).

[64] Yunfeng Zhang, Q Vera Liao, and Rachel KE Bellamy. 2020. 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*. 295–305.