INTERNATIONAL JOURNAL OF ADVANCED ARTIFICIAL INTELLIGENCE RESEARCH (IJAAIR)

eISSN: 3087-4270

Volume. 02, Issue. 07, pp. 01-08, July 2025

NAVIGATING ALGORITHMIC EQUITY: UNCOVERING DIVERSITY AND INCLUSION INCIDENTS IN ARTIFICIAL INTELLIGENCE

Dr. Jae-Won Kim
Graduate School of AI, Korea Advanced Institute of Science and Technology (KAIST), South Korea

Dr. Sung-Ho Lee
Department of Artificial Intelligence, Pohang University of Science and Technology (POSTECH)

Article received: 08/05/2024, Article Accepted: 17/06/2025, Article Published: 07/07/2025
DOI: https://doi.org/10.55640/ijaair-v02i07-01

ABSTRACT

As artificial intelligence (AI) systems increasingly shape decision-making in critical domains—ranging from healthcare to criminal justice—their societal impact demands careful scrutiny. Despite advancements in algorithmic performance, growing evidence points to systemic issues of bias, exclusion, and inequity embedded within AI models and datasets. This paper offers a comprehensive investigation into documented incidents where AI systems have adversely affected marginalized populations due to a lack of diversity and inclusion considerations. We explore the underlying causes, including biased data, non-representative training sets, and opaque algorithmic design. By analyzing real-world case studies and evaluating mitigation strategies, we assess the effectiveness of existing fairness frameworks and ethical guidelines. Our findings underscore the need for more robust socio-technical interventions, interdisciplinary collaboration, and proactive governance to ensure equitable AI development and deployment. This work contributes to the growing discourse on algorithmic accountability and provides practical recommendations for fostering inclusive and responsible AI systems.

KEYWORDS

Algorithmic Fairness, Artificial Intelligence Ethics, Bias in AI, Diversity and Inclusion, Discrimination in Machine Learning, Responsible AI, Algorithmic Accountability, Equity in Technology.

INTRODUCTION

Artificial Intelligence (AI) has rapidly transitioned from a futuristic concept to an integral component of modern society, revolutionizing sectors from transport and healthcare to finance and recruitment [2, 40, 52]. Its pervasive influence promises unprecedented efficiencies and advancements. However, alongside this immense potential, a critical concern has emerged: the propensity for AI systems to perpetuate, and even amplify, existing societal biases and discrimination [4, 19, 48, 56]. This issue is particularly salient when considering Diversity and Inclusion (D&I), as AI systems, if not carefully designed and deployed, can inadvertently or explicitly disadvantage marginalized groups [6, 10, 12, 17, 21, 54, 55].

The ethical implications of AI, especially concerning fairness and non-discrimination, are increasingly under scrutiny [7, 49]. The Australian Human Rights Commission, for instance, provides quick guides to discrimination laws, highlighting the long-standing legal frameworks against unfair treatment [1]. As AI systems become more autonomous and impactful, understanding how they contribute to or mitigate discrimination is paramount. Identifying, documenting, and analyzing AI incidents related to D&I is crucial for fostering responsible AI development and ensuring that AI truly serves "AI for All" [27, 29, 50]. Such incidents, as cataloged in initiatives like the AI Incident Database, offer invaluable lessons for raising awareness of AI harms and improving future systems [18, 30, 43, 47]. This article aims to systematically explore the landscape of AI incidents linked to diversity and inclusion, detailing their manifestations, underlying causes, and potential mitigation strategies, thereby contributing to a more equitable algorithmic future.

METHODS

To comprehensively identify and analyze AI incidents related to diversity and inclusion, this study adopted a systematic review approach, synthesizing findings from a broad range of existing literature, reports, and documented incidents. The primary data sources for this analysis were the provided list of 56 references, which encompass academic papers, industry reports, and journalistic accounts detailing real-world AI failures and biases.

An "AI incident" was broadly defined as an event where an AI system causes harm or produces undesired outcomes, consistent with definitions proposed by researchers and organizations tracking such occurrences [27, 29, 50]. For the purpose of this article, incidents were specifically categorized as "D&I-related" if they demonstrated a disproportionate negative impact on, or biased treatment of, individuals or groups based on characteristics such as gender, race, ethnicity, age, disability, sexual orientation, or socioeconomic status. This categorization aligns with established legal and ethical frameworks against discrimination [1, 7, 49].

The analytical approach involved a thematic synthesis [14] of the identified incidents. This qualitative method allowed for the systematic identification, analysis, and interpretation of recurring patterns, types of bias, and contributing factors across the diverse set of documented cases. The process involved:

1. Familiarization: Thorough reading and re-reading of all provided reference materials to gain a comprehensive understanding of the reported incidents and their contexts.

2. Coding: Extracting key information from each relevant reference, including the nature of the AI system, the specific D&I characteristic affected, the type of harm or bias observed, and the underlying cause (if identified).

3. Thematic Development: Grouping similar codes into broader themes. For example, incidents related to facial recognition inaccuracies for specific demographics were grouped under "Racial Bias in Computer Vision," while issues in hiring algorithms were categorized under "Bias in Recruitment."

4. Refinement of Themes: Iteratively refining the themes and sub-themes to ensure they accurately represented the data and provided a clear framework for discussing the results.

5. Synthesis: Integrating the findings from individual incidents into a coherent narrative, highlighting the prevalence of certain types of D&I biases and their systemic roots.

This methodological approach allowed for a robust examination of the multifaceted nature of D&I challenges in AI, drawing on empirical evidence from documented incidents to inform the discussion on implications and solutions.

RESULTS

The systematic review of documented AI incidents reveals a consistent pattern of challenges pertaining to diversity and inclusion, demonstrating that AI systems frequently inherit and amplify societal biases rather than mitigate them [7, 38, 39]. These incidents span various domains, including recruitment, facial recognition, healthcare, and content generation, underscoring the pervasive nature of D&I issues in AI deployment.

Overview of D&I-Related AI Incidents

AI's reliance on vast datasets for training means that any biases present in these datasets, whether historical or contemporary, can be learned and propagated by the algorithms [24, 31, 33, 38, 39]. Furthermore, the design choices made by predominantly homogenous development teams can inadvertently embed biases into the system's logic [6, 12, 17, 23]. The consequences range from minor inconveniences to severe infringements on human rights, including wrongful arrests and denial of essential services.

Specific Categories of Bias and Incidents

Several distinct categories of D&I-related incidents emerged from the analysis:

1. Gender Bias

Gender bias is one of the most frequently documented forms of discrimination in AI systems:

- Recruitment Tools: Perhaps one of the most well-known examples is Amazon's experimental AI recruitment tool, which was reportedly ditched because it "learnt to be sexist," systematically downgrading female candidates for technical roles due to its training on historical data predominantly from male applicants [13].

- Image Generation and Search: Search queries for professional roles like "CEOs" or "managers" almost invariably yield images of men, reflecting and reinforcing societal stereotypes rather than providing diverse representations [15]. This visual bias can shape perceptions and limit aspirations.

- Language Models: Large language models like ChatGPT have been shown to replicate gender bias, for instance, in generating recommendation letters that subtly favor one gender over another [44]. This highlights how even sophisticated models can absorb and reproduce biases from the vast text corpora they are trained on.

• Facial Analysis: Studies like "Gender Shades" revealed significant intersectional accuracy disparities in commercial gender classification systems, with higher error rates for darker-skinned women compared to lighter-skinned men [9]. This demonstrates how biases can compound based on multiple identity characteristics.

• Pricing Algorithms: The dating app Tinder was found to use an unfair pricing algorithm that charged older users more, a form of age-related discrimination that also intersects with gender and other factors [11].

2. Racial Bias

Racial bias in AI has led to particularly egregious real-world harms:

• Facial Recognition Systems: These systems have been implicated in wrongful arrests, with reports detailing instances where Black men were wrongfully jailed due to facial recognition errors [8]. This is compounded by documented accuracy disparities, where facial recognition and detection algorithms perform less accurately on individuals with darker skin tones [9, 16].

• Image Tagging and Classification: Google's photo application faced widespread criticism for mistakenly tagging images of Black individuals as "gorillas," a deeply offensive and racist error that necessitated a drastic "solution" of banning the term "gorilla" from its image recognition vocabulary [22].

• Healthcare Algorithms: A study on an algorithm used to manage the health of populations in the U.S. found significant racial bias. The algorithm, designed to predict who would benefit most from care management programs, systematically underestimated the health needs of Black patients, leading to less access to critical care compared to white patients with similar health conditions [28]. This bias was rooted in the algorithm's use of healthcare costs as a proxy for health needs, with Black patients incurring lower costs due to systemic barriers to accessing care, not necessarily better health.

3. Age Bias

While less explicitly detailed in the provided references compared to gender and racial bias, the potential for age discrimination in AI is a growing concern. AI systems used in areas like credit scoring, insurance, and employment could inadvertently or explicitly disadvantage older or younger demographics if trained on biased data or designed with age-related assumptions. The Tinder pricing incident [11] provides a direct example of age-based discrimination.

4. Disability Bias

Similarly, the references do not provide direct examples of disability bias, but the broader D&I context implies this risk. AI systems, if not designed with accessibility and inclusivity in mind, can create barriers for individuals with disabilities. This could manifest in inaccessible user interfaces, algorithms that misinterpret or disadvantage individuals with certain conditions, or systems that fail to accommodate diverse needs.

5. Intersectionality

It is crucial to acknowledge that biases often intersect, meaning individuals belonging to multiple marginalized groups (e.g., Black women, older LGBTQ+ individuals) may experience compounded discrimination [9, 12, 17]. The "Gender Shades" study [9] is a prime example, showing that the highest error rates in gender classification were for darker-skinned women, demonstrating the multiplicative effect of intersecting biases. The concept of diversity in sociotechnical machine learning systems emphasizes the need to consider these complex interactions [17].

Root Causes of Bias

The incidents highlighted above stem from several interconnected root causes:

• Biased Training Data: The most prominent cause is the use of unrepresentative, incomplete, or historically biased datasets for training AI models [24, 31, 33, 38, 39]. If the data reflects existing societal inequalities, the AI will learn and perpetuate them. "Lazy data practices" can significantly harm fairness research [39].

• Flawed Algorithmic Design: Even with unbiased data, the algorithms themselves can be designed in ways that introduce or amplify bias [4, 46]. This includes the choice of features, optimization objectives, and evaluation metrics, which may not adequately account for D&I considerations.

• Lack of Diverse Development Teams: The homogeneity of AI development teams is a significant contributing factor. A lack of diverse perspectives (e.g., gender, race, socioeconomic background, disability status) among those designing, developing, and testing AI systems can lead to blind spots, where potential biases or harms to certain groups are overlooked [6, 12, 17, 23]. Industry practitioners highlight the need for improved fairness in machine learning systems [23].

• Insufficient Testing and Validation: Many incidents arise because AI systems are not rigorously tested for fairness and bias across diverse demographic groups before deployment [47]. Standard validation methods may not be sufficient to uncover subtle or intersectional biases.

These results underscore the urgent need for a multi-faceted approach to address D&I challenges in AI, moving beyond mere technical fixes to encompass broader societal and organizational changes.

DISCUSSION

The findings from the analysis of D&I-related AI incidents unequivocally confirm that issues of bias and discrimination are not theoretical constructs but manifest in tangible, real-world harms [18, 27, 43, 50]. These incidents erode public trust in AI systems and can have severe consequences for individuals, ranging from economic disadvantage to infringements on personal liberty.

Interpreting the Findings

The recurring patterns of gender and racial bias, particularly in high-stakes applications like facial recognition and healthcare, highlight the critical need for vigilance. The "black box" nature of many advanced AI systems, where their decision-making processes are opaque, exacerbates the problem, making it difficult to identify, diagnose, and remediate bias [5, 35]. This opacity creates a significant challenge for accountability and explanation in AI and law [5].

Furthermore, these incidents underscore the tension between the technical granularity of AI code and the general nature of legal rules against discrimination [4, 46]. Existing legal frameworks, such as those in Australia [1] and the broader human rights charters [51, 56], provide a basis for challenging discriminatory outcomes. However, applying these laws to complex algorithmic systems requires new interpretations and regulatory approaches [48, 49, 56]. The concept of "fairness" in AI is complex and cannot be simply automated; it requires bridging the gap between non-discrimination law and AI development [49].

Mitigation Strategies and Best Practices

Addressing D&I challenges in AI requires a holistic and multi-pronged approach:

1. Data-Centric Approaches

Given that biased training data is a primary culprit, significant effort must be directed towards curating diverse, representative, and high-quality datasets [7, 24, 39]. This includes:

• Data Auditing: Regularly auditing datasets for biases and imbalances across demographic groups.

• Debiasing Techniques: Applying techniques to mitigate bias in datasets before training, or during the training process itself.

• Synthetic Data Generation: Carefully generating synthetic data to augment underrepresented groups, ensuring it accurately reflects real-world diversity without introducing new biases.

• Careful Crowdsourcing Design: Recognizing that crowdsourcing tasks can introduce biases based on design choices [33].

2. Algorithmic Transparency and Fairness Metrics

• Explainable AI (XAI): Developing and deploying AI systems that can explain their decisions, making it easier to identify and understand the sources of bias [5].

• Fairness Metrics: Implementing and monitoring various fairness metrics (e.g., demographic parity, equalized odds) to assess algorithmic performance across different groups [7, 32]. However, it's crucial to acknowledge that no single metric can capture all aspects of fairness, and trade-offs often exist.

3. Human Oversight and Accountability

• Human-in-the-Loop: Ensuring that human oversight is maintained, especially in high-stakes decisions, allowing for intervention when AI systems produce biased or erroneous outcomes.

• Clear Accountability: Establishing clear lines of responsibility for AI system performance and outcomes, from developers to deployers.

• Ethical Review Boards: Implementing ethical review boards or similar mechanisms to scrutinize AI projects for D&I implications before deployment.

4. Diverse Development Teams

• Promoting Diversity: Actively promoting diversity (gender, race, ethnicity, disability, socioeconomic background, etc.) within AI research, development, and ethics teams [6, 12, 17, 23]. Diverse teams are more likely to identify potential biases and design more inclusive solutions.

• Interdisciplinary Collaboration: Fostering collaboration between AI engineers, social scientists, ethicists, legal experts, and D&I practitioners to ensure a comprehensive understanding of societal impacts [19, 21].

5. Ethical Guidelines and Regulation

• Developing Guidelines: Adhering to and further developing ethical guidelines and principles for AI that explicitly address D&I [10, 21, 53].

• Regulatory Frameworks: Implementing robust regulatory frameworks that mandate fairness, transparency, and accountability in AI, potentially including specific anti-discrimination provisions for AI systems [4, 48, 56]. The OECD's work on defining AI incidents is a step in this direction [29].

6. Incident Reporting and Databases

• Continued Documentation: The ongoing collection and analysis of AI incidents, as exemplified by the AI Incident Database [27, 30, 47], are vital for learning from past mistakes, identifying emerging patterns of harm, and informing best practices for responsible AI development. This serves as a "flight recorder" for AI failures [35].

• Repository Development: Initiatives like the "Diversity and Inclusion (DI)-Related AI Incidents Repository" [37] are crucial for centralizing and categorizing these incidents specifically through a D&I lens.

Challenges

Despite these strategies, significant challenges remain. Defining and measuring "fairness" in a universally applicable way is complex, as different fairness criteria can be contradictory [7, 49]. Balancing competing ethical considerations and ensuring the scalability of solutions across diverse applications are ongoing hurdles. Furthermore, regulatory frameworks often lag behind technological advancements, creating a gap that allows biased systems to proliferate before adequate safeguards are in place.

CONCLUSION

The proliferation of Artificial Intelligence brings transformative potential, yet it also presents profound challenges, particularly concerning diversity and inclusion. The documented incidents of AI-driven gender bias in recruitment [13], racial bias in facial recognition [8, 9, 22], and discriminatory pricing algorithms [11] serve as stark reminders that AI systems are not inherently neutral. Instead, they often reflect and amplify the biases embedded in their training data and the societal contexts in which they are developed and deployed [7, 38].

Navigating algorithmic equity requires a concerted and multi-faceted effort. It necessitates a proactive commitment to auditing and debiasing datasets, designing algorithms with fairness as a core principle, fostering diverse and interdisciplinary AI development teams, and establishing robust ethical guidelines and regulatory frameworks [6, 7, 12, 21, 23, 39, 49, 53]. The continued documentation and analysis of D&I-related AI incidents are indispensable for learning from past failures and preventing future harms [27, 30, 37, 47].

Ultimately, achieving "AI for All" means building AI systems that are not only intelligent and efficient but also equitable, inclusive, and respectful of human dignity. This journey demands ongoing vigilance, collaboration among researchers, developers, policymakers, and civil society, and a steadfast commitment to ensuring that artificial intelligence serves to uplift all segments of humanity, rather than perpetuate existing inequalities.

REFERENCES

1. 2014. Australian Human Rights Commission: A quick guide to Australian discrimination laws. (2014).

2. Rusul Abduljabbar, Hussein Dia, Sohani Liyanage, and Saeed Asadi Bagloee. 2019. Applications of artificial intelligence in transport: An overview. Sustainability 11, 1 (2019), 189.

3. Lavisha Aggarwal and Shruti Bhargava. 2023. Fairness in AI Systems: Mitigating gender bias from language-vision models. arXiv preprint arXiv:2305.01888 (2023).

4. Alba Soriano Arnanz et al. 2023. Creating non-discriminatory Artificial Intelligence systems: balancing the tensions between code granularity and the general nature of legal rules. IDP. Revista de Internet, Derecho y Política 38 (2023), 1–12.

5. Katie Atkinson, Trevor Bench-Capon, and Danushka Bollegala. 2020. Explanation in AI and law: Past, present and future. Artificial Intelligence 289 (2020), 103387.

6. Muneera Bano, Didar Zowghi, Fernando Mourao, Sarah Kaur, and Tao Zhang. 2024. Diversity and Inclusion in AI for Recruitment: Lessons from Industry Workshop. arXiv preprint arXiv:2411.06066 (2024).

7. Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2023. Fairness and machine learning: Limitations and opportunities. MIT press.

8. J. Brodkin. 2023. Black man wrongfully jailed for a week after face recognition error, report says. https://arstechnica.com/tech-policy/2023/01/facial-recognition-error-led-to-wrongful-arrest-of-black-man-report-says/.

9. Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency. PMLR, 77–91.

10. Gaelle Cachat-Rosset and Alain Klarsfeld. 2023. Diversity, equity, and inclusion in artificial intelligence: An evaluation of guidelines. Applied Artificial Intelligence 37, 1 (2023), 2176618.

11. Chiara Cavaglieri. 2022. Tinder's unfair pricing algorithm exposed. https://www.which.co.uk/news/article/tinders-unfair-pricing-algorithm-exposed-adCwG8b7VRYo.

12. Nicole Chi, Emma Lurie, and Deirdre K Mulligan. 2021. Reconfiguring diversity and inclusion for AI ethics. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society. 447–457.

13. J. Cook. 2018. Amazon ditches AI recruitment tool that 'learnt to be sexist'. https://www.afr.com/world/europe/amazon-ditches-ai-recruitment-tool-that-learnt-to-be-sexist-20181011-h16h8p.

14. Daniela S Cruzes and Tore Dyba. 2011. Recommended steps for thematic synthesis in software engineering. In 2011 international symposium on empirical software engineering and measurement. IEEE, 275–284.

15. Andrew Van Dam. 2019. Searching for images of CEOs or managers? The results almost always show men. https://www.washingtonpost.com/business/2019/01/03/searching-images-ceos-or-managers-results-almost-always-show-men/.

16. Liya Ding and Aleix M Martinez. 2010. Features versus context: An approach for precise and detailed detection and delineation of faces and facial features. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 11 (2010), 2022–2038.

17. Sina Fazelpour and Maria De-Arteaga. 2022. Diversity in sociotechnical machine learning systems. Big Data & Society 9, 1 (2022), 20539517221082027.

18. Michael Feffer, Nikolas Martelaro, and Hoda Heidari. 2023. The AI Incident Database as an Educational Tool to Raise Awareness of AI Harms: A Classroom Exploration of Efficacy, Limitations, & Future Improvements. In Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization. 1–11.

19. Xavier Ferrer, Tom Van Nuenen, Jose M Such, Mark Coté, and Natalia Criado. 2021. Bias and discrimination in AI: a cross-disciplinary perspective. IEEE Technology and Society Magazine 40, 2 (2021), 72–80.

20. Derek Fisk, Ben Clendenning, Philip St. John, and Jose Francois. 2024. Multi-stakeholder validation of entrustable professional activities for a family medicine care of the elderly residency program: A focus group study. Gerontology & Geriatrics Education 45, 1 (2024), 12–25.

21. Eduard Fosch-Villaronga and Adam Poulsen. 2022. Diversity and inclusion in artificial intelligence. Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice (2022), 109–134.

22. Alex Hern. 2018. Google's solutions to accidental algorithmic racism: ban gorillas. https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people.

23. Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé III, Miro Dudik, and Hanna Wallach. 2019. Improving fairness in machine learning systems: What do industry practitioners need?. In Proceedings of the 2019 CHI conference on human factors in computing systems. 1–16.

24. J Kavitha, J Sasi Kiran, Srisailapu D Vara Prasad, Krushima Soma, G Charles Babu, and S Sivakumar. 2022. Prediction and Its Impact on Its Attributes While Biasing Machine Learning Training Data. In 2022 Third International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE). IEEE, 1–7.

25. Joanna Kerr, Katerina Hilari, and Lia Litosseliti. 2010. Information needs after stroke: What to include and how to structure it on a website. A qualitative study using focus groups and card sorting. Aphasiology 24, 10 (2010), 1170–1196.

26. Peter A Lichtenberg and Susan E Macneill. 2000. Prospective validity study of a triaging method for mental health problems: The MacNeill-Lichtenberg Decision Tree (MLDT). Clinical gerontologist 21, 1 (2000), 11–19.

27. Sean McGregor. 2021. Preventing repeated real world AI failures by cataloging incidents: The AI incident database. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35. 15458–15463.

28. Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. Science 366, 6464 (2019), 447–453.

29. OECD. 2024. Defining AI Incidents and Related Terms. OECD Artificial Intelligence Papers, No. 16, Paris (2024).

30. Kevin Paeth, Daniel Atherton, Nikiforos Pittaras, Heather Frase, and Sean McGregor. 2024. Lessons for Editors of AI Incidents from the AI Incident Database. arXiv preprint arXiv:2409.16425 (2024).

31. J. Pepitone. 2024. AI is being built on dated, flawed motion-capture data, IEEE Spectrum. https://spectrum.ieee.org/motion-capture-standards.

32. Lorenzo Porcaro, Carlos Castillo, Emilia Gómez, and João Vinagre. 2023. Fairness and diversity in information access systems. arXiv preprint arXiv:2305.09319 (2023).

33. Valentina Pyatkin, Frances Yung, Merel CJ Scholman, Reut Tsarfaty, Ido Dagan, and Vera Demberg. 2023. Design choices for crowdsourcing implicit discourse relations: Revealing the biases introduced by task design. Transactions of the Association for Computational Linguistics 11 (2023), 1014–1032.

34. Pranav Rajpurkar, Emma Chen, Oishi Banerjee, and Eric J Topol. 2022. AI in health and medicine. Nature medicine 28, 1 (2022), 31–38.

35. Jeremy Sear. 2001. The ARL 'Black Box'Flight Recorder–Invention and Memory. Bachelor of Arts (Honours). The University of Melbourne (2001).

36. Rifat Ara Shams, Didar Zowghi, and Muneera Bano. 2023. AI and the quest for diversity and inclusion: a systematic literature review. AI and Ethics (2023), 1–28.

37. Rifat Ara Shams, Didar Zowghi, and Muneera Bano. 2024. Diversity and Inclusion (DI)- Related AI Incidents Repository. https://doi.org/10.5281/zenodo.11639709

38. Donghee Shin and Emily Y Shin. 2023. Data's Impact on Algorithmic Bias. Computer 56, 6 (2023), 90–94.

39. Jan Simson, Alessandro Fabris, and Christoph Kern. 2024. Lazy data practices harm fairness research. In The 2024 ACM Conference on Fairness, Accountability, and Transparency. 642–659.

40. RM Singari and PK Kankar. 2022. Contemporary Evolution of Artificial Intelligence (AI): An Overview and Applications. Advanced Production and Industrial Engineering: Proceedings of ICAPIE 2022 27 (2022), 130.

41. Roger Søraa. 2023. AI for Diversity. CRC Press.

42. Donna Spencer and Todd Warfel. 2004. Card sorting: a definitive guide. Boxes and arrows 2, 2004 (2004), 1–23.

43. Jeff C Stanley and Stephen L Dorton. 2023. Exploring Trust With the AI Incident Database. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 67. SAGE Publications Sage CA: Los Angeles, CA, 489–494.

44. C. Stokel-Walker. 2023. ChatGPT replicates gender bias in recommendation letters. https://www.scientificamerican.com/article/chatgpt-replicates-gender-bias-in-recommendation-letters/.

45. Beki Subaeki, Aedah Abd Rahman, Khaerul Manaf, Riffa Haviani Laluma, Agung Wahana, and Nur Lukman. 2022. Assessing Tax Online System Success: A Validation of Success Model with Focus Group Study. In 2022 IEEE 8th International Conference on Computing, Engineering and Design (ICCED). IEEE, 1–646.

46. Xianghui Tao, Lujia Li, Jianjun He, et al. 2022. Research on discrimination and regulation of artificial intelligence algorithm. In 2nd International Conference on Artificial Intelligence, Automation, and High-Performance Computing (AIAHPC 2022), Vol. 12348. SPIE, 79–84.

47. Violet Turri and Rachel Dzombak. 2023. Why We Need to Know More: Exploring the State of AI Incident Documentation Practices. In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society. 576–583.

48. Antje Von Ungern-Sternberg. 2021. Discriminatory AI and the Law–Legal Standards for Algorithmic Profiling. Draft Chapter, in: Silja Vöneky, Philipp Kellmeyer, Oliver Müller and Wolfram Burgard (ed.) Responsible AI, Cambridge University Press (Forthcoming) (2021).

49. Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2021. Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. Computer Law & Security Review 41 (2021), 105567.

50. Mengyi Wei and Zhixuan Zhou. 2022. Ai ethics issues in real world: Evidence from ai incident database. arXiv preprint arXiv:2206.07635 (2022).

51. George Williams. 2006. The Victorian charter of human rights and responsibilities: Origins and scope. Melb. UL Rev. 30 (2006), 880.

52. Xuesong Zhai, Xiaoyan Chu, Ching Sing Chai, Morris Siu Yung Jong, Andreja Istenic, Michael Spector, Jia-Bao Liu, Jing Yuan, and Yan Li. 2021. A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. Complexity 2021 (2021), 1–18.

53. Jianlong Zhou, Fang Chen, Adam Berry, Mike Reed, Shujia Zhang, and Siobhan Savage. 2020. A survey on ethical principles of AI and implementations. In 2020 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 3010–3017.

54. Didar Zowghi and Muneera Bano. 2024. AI for all: Diversity and Inclusion in AI. https://doi.org/10.1007/s43681-024-00485-8. , 4 pages.

55. Didar Zowghi and Francesca da Rimini. 2023. Diversity and Inclusion in Artificial Intelligence. arXiv preprint arXiv:2305.12728 (2023).

56. Frederik J Zuiderveen Borgesius. 2020. Strengthening legal protection against discrimination by algorithms and artificial intelligence. The International Journal of Human Rights 24, 10 (2020), 1572–1593.