# The New Regulatory Paradigm: IEEE Std 7003 and Its Impact on Bias Management in Autonomous Intelligent Systems

Warren Huang[1] and Pablo Rivas[2]

[1] Department of Economics, Baylor University, Texas, USA
Warren_Huang1@Baylor.edu
[2] Department of Computer Science, Baylor University, Texas, USA
Pablo_Rivas@Baylor.edu

**Abstract.** This paper critically evaluates the newly introduced IEEE Standard for Algorithmic Bias Considerations (IEEE Std 7003-2024) as a transformative framework for managing bias in autonomous intelligent systems (AIS). Our analysis examines the standard's comprehensive structure—including the development of a bias profile, stakeholder identification, data representation, and risk and impact assessment, complemented by mechanisms for continuous evaluation. The structured approach set forth in the standard establishes a new benchmark for transparency and accountability in AI development, effectively bridging theoretical guidelines with practical implementation. While the standard marks a significant advancement in bias regulation, our evaluation also identifies opportunities for refinement, such as the integration of quantitative metrics and the development of sector-specific operational guidelines. These insights contribute to the broader discourse on responsible AI development, underscoring the promise of systematic bias mitigation and outlining critical directions for future research.

**Keywords:** Algorithmic Bias · Autonomous Intelligent Systems · IEEE Standard

## 1 Introduction

The rapid integration of artificial intelligence (AI) into decision-making processes across various sectors has transformed the landscape of technology and society. As AI systems increasingly influence critical areas such as healthcare, finance, and law enforcement, they bring forth significant ethical challenges, particularly concerning bias. The deployment of AI technologies raises questions about fairness, accountability, and transparency, necessitating a thorough examination of how these systems operate and the implications of their decisions [32,40]. The ethical stakes are high; decisions made by AI can have profound effects on individuals and communities, making it imperative to address the biases that may be embedded within these systems.

In the realm of algorithmic decision-making, distinguishing between desired and unwanted bias is crucial. Desired bias may be intentionally integrated into AI systems to achieve specific ethical or societal goals, such as promoting diversity in hiring practices or prioritizing healthcare resources for underrepresented populations [39]. For instance, algorithms may be designed to favor underrepresented groups in hiring processes to promote diversity and inclusion, thereby reflecting a socially desirable bias [35]. Conversely, unwanted bias arises from flawed data or algorithmic processes, leading to discriminatory outcomes that can perpetuate existing inequalities. Research indicates that biases embedded in training data can significantly affect the performance of AI systems, often disadvantaging marginalized populations [36,12]. The literature also highlights the importance of recognizing bias as a dual-edged sword in AI system design. While some biases are necessary for achieving fairness and equity, others can lead to harmful stereotypes and reinforce systemic discrimination [24,34]. For example, the under-representation of women and minorities in AI development teams can result in products that inadequately address the needs of these groups, thereby perpetuating existing disparities [27]. This dichotomy necessitates a careful examination of the motivations behind bias in AI systems, as well as the implications of these biases for affected stakeholders.

The purpose of this paper is to evaluate how a structured process can help manage bias in AIS. By exploring how IEEE Std 7003-2024 provides a foundation for bias mitigation, this paper assesses how a structured methodology, combined with proposed refinements, can balance the need for bias mitigation with stakeholder requirements. Focusing on key aspects such as stakeholder identification, data representation, risk and impact assessment, and ongoing evaluation, the paper aims to explore how systematic approaches can not only identify potential biases but also ensure that the voices of affected communities are considered in the design and implementation of AI systems [14,5]. As we advance further into this discussion, it is essential to recognize that addressing bias in AI is not merely a technical challenge but a societal imperative. The consequences of unchecked bias can lead to significant harm, reinforcing systemic inequalities and undermining public trust in technology [37,2]. Therefore, a comprehensive understanding of how to manage bias through structured methodologies is vital for fostering responsible AI development and ensuring that these technologies serve the interests of all stakeholders equitably.

The rest of the paper is organized as follows. First, we review the background and related work on AI bias and ethical challenges. Next, we present an overview of the standard, highlighting its core components. We then describe our evaluation framework and methodology for assessing bias management strategies. This is followed by an analysis of our findings and a discussion of their implications. Finally, we conclude by summarizing our contributions and outlining directions for future research.

## 2 Background and Related Work

### 2.1 Current Issues

The rapid advancements in AI have brought numerous benefits but also raised critical ethical concerns, particularly regarding bias, fairness, and governance. Recent developments, such as the surge in DeepSeek's popularity and President Trump's rollback of DEI initiatives, have exposed growing gaps in AI ethics enforcement. These challenges emphasize the urgent need for stronger regulatory oversight and accountability mechanisms.

DeepSeek, an advanced AI model designed for content creation, has faced criticism for producing biased outputs and raised concerns about its overall transparency, despite offering some insight into its reasoning processes. Its strict censorship practices, which align with the agenda of the Chinese Communist Party, raise concerns about the ethical implications of AI-driven content moderation, particularly in restricting free expression and shaping public discourse to fit state narratives [20]. These concerns have also resulted in the Congressional push to ban DeepSeek, citing security risks in data collection and storage especially considering the company's close ties to the Chinese military [20,31].

Another critical issue in AI governance is the impact of political and policy changes. President Trump has signed three executive orders titled "Ending Illegal Discrimination and Restoring Merit-Based Opportunity," "Defending Women From Gender Ideology Extremism and Restoring Biological Truth to the Federal Government," and "Ending Radical and Wasteful Government DEI Programs and Preferencing," which may significantly influence how bias is addressed in AI development and deployment. Historically, DEI initiatives have played a crucial role in mitigating bias in AI-driven decision-making, particularly through the hiring of diverse teams and creation of equity programs in higher education [26]. The directive for the Attorney General to identify and potentially investigate private sector companies with "egregious and discriminatory" DEI programs signals a broader impact beyond federal agencies [19]. This move will deter private organizations from maintaining DEI initiatives, further exacerbating disparities and hindering progress toward a more inclusive society.

### 2.2 Societal Impact of Bias

The literature highlights the importance of considering the broader societal implications of algorithmic bias. For instance, biased AI systems can exacerbate existing inequalities in areas such as healthcare, criminal justice, and employment, leading to significant adverse outcomes for marginalized communities [35,24]. Generalizing models can result in suboptimal performance when deployed across varied populations while biased systems risk reinforcing and amplifying existing inequalities [18]. By employing comprehensive risk assessment methodologies, researchers and practitioners can better understand the potential consequences of bias in AI systems and develop strategies to mitigate these risks.

### 2.3 Existing Approaches

Bias evaluation in artificial intelligence has undergone profound transformation over the past decade, driven by interdisciplinary advances in computer science, ethics, and social sciences. Early approaches focused narrowly on statistical parity and outcome imbalances, but modern frameworks now integrate causal reasoning, adversarial training, and dynamic fairness metrics to address systemic inequities. This evolution reflects growing recognition that bias manifests not merely as technical flaws in datasets but as structural phenomena requiring holistic mitigation strategies. [25].

While tools like adversarial debiasing and counterfactual fairness represent significant advances, their efficacy depends on contextual adaptation and alignment with broader ethical frameworks. As AI permeates critical infrastructure, only through rigorous, ethically grounded bias management can we ensure these systems promote equity rather than erode it [1].

## 3 Overview of the Standard

### 3.1 Clause 1: Overview

The exploration of bias in autonomous intelligent systems is a multifaceted endeavor that requires a nuanced understanding of various dimensions, including algorithmic decision-making, stakeholder identification, data representation, risk and impact assessment, and ongoing evaluation. Each of these areas presents unique challenges and opportunities for addressing bias, which can manifest both as a deliberate design choice and as an unintended consequence of system implementation.

To systematically manage bias, AI developers must critically examine their systems' potential for discrimination and unfair outcomes. This includes establishing a criteria for dataset selection, defining application boundaries, and managing user expectations [1]. By proactively identifying risks, establishing clear accountability, and implementing iterative monitoring strategies, organizations can work toward minimizing unwanted bias in AI systems [23]. While it sets minimum criteria to reduce unwanted bias in AI systems, it specifies that adherence does not guarantee alignment with mission objectives or prevent adverse consequences.

### 3.2 Clause 4: Requirements for Bias Consideration

The requirements for bias consideration in AIS emphasizes a structured and iterative approach to manage bias throughout the system's life cycle. The process begins with requirements setting, which involves defining the role of bias in achieving the system's functional objectives and distinguishing between wanted and unwanted bias. This stage is essential as it ensures that bias is proactively considered rather than addressed reactively. [3].

The primary outputs of this stage include the bias profile, a structured repository that records bias-related considerations throughout the AIS life cycle, and a values statement that aligns bias considerations with the organization's ethical and operational priorities [1]. To achieve these objectives, specific actions must be taken, including gathering essential documentation to define the AIS's purpose and governance structure, identifying sensitive attributes, and ensuring their representation in the bias consideration process. Organizations must also establish accountability structures that integrate bias considerations into governance frameworks, define specific requirements for AIS development, operation, and decommissioning. The intended outcome of these actions is a clear understanding of the boundaries of acceptability for bias in AIS, allowing developers to anticipate and mitigate risks systematically.

### 3.3 Clause 5: Bias Profile

The bias profile serves as a repository of information created and maintained throughout the life cycle of an AIS to document algorithmic considerations. The purpose of the bias profile is to provide a continuous record of how bias is identified, evaluated, and mitigated within an AIS, recognizing that algorithmic bias can be both a necessary feature and a potential flaw. The standard emphasizes that unwanted bias can stem from multiple sources, including the data used to train the model, the model itself, or the code that builds the model [1]. By maintaining a structured bias profile, organizations can create accountability mechanisms that track bias-related decisions, ultimately fostering transparency and trust [17].

To effectively address this, a five-stage framework is isolated, beginning with bias consideration, which establishes the foundational processes for bias evaluation. Next is stakeholder identification, ensuring that all affected groups are considered throughout the system's development. Data representation follows, assessing whether data accurately reflect diverse perspectives. The fourth stage, risk and impact assessment, identifies potential consequences of bias and develops mitigation strategies. Finally, evaluation involves continuously monitoring the AI system for bias over time, ensuring ongoing accountability and fairness [1].

### 3.4 Clause 6: Stakeholder Identification

Identifying and mapping stakeholders affected by algorithmic decisions is essential for ensuring that diverse perspectives are represented in the design and implementation of AI systems. Various methodologies have been proposed for stakeholder analysis, emphasizing the need for inclusive practices that capture the voices of all affected parties, particularly those from historically marginalized communities [4,16]. Techniques such as participatory design and community engagement have been shown to enhance stakeholder representation, fostering a more equitable approach to AI development [35,12].

Moreover, the identification of stakeholders must consider the intersectionality of various identities, including race, gender, socioeconomic status, and disability. This complexity underscores the necessity for comprehensive frameworks that can adequately capture the diverse experiences and needs of different groups [7,22]. By employing these frameworks, researchers and practitioners can better understand the potential impacts of algorithmic decisions on various stakeholders, ultimately leading to more equitable outcomes.

### 3.5 Clause 7: Data Representation

The representation of data in AI systems is a critical factor influencing the fairness and effectiveness of algorithmic decision-making. Studies have shown that the quality and relevance of data used to train AI models directly affect their performance and the equity of their outcomes [11]. For instance, data that inadequately represents certain demographic groups can lead to biased predictions and reinforce existing disparities [8,29]. Therefore, ensuring that datasets are comprehensive and representative of the populations they serve is paramount for promoting fairness in AI systems.

Furthermore, the literature emphasizes the importance of ongoing assessments of data quality and representativeness. Techniques such as data auditing and bias detection algorithms can help identify and mitigate representation bias in datasets, thereby enhancing the overall fairness of AI systems [34,8]. By prioritizing diverse and representative data, AI developers can create systems that are more likely to serve the needs of all users, rather than perpetuating systemic inequalities.

### 3.6 Clause 8: Risk and Impact Assessment

Assessing the risks and impacts of bias in algorithmic systems is a critical component of responsible AI development. Various frameworks and methodologies have been proposed to evaluate potential adverse outcomes stemming from both intended and unintended biases [4,15]. For example, risk assessment frameworks can help identify the likelihood and severity of negative impacts associated with biased algorithmic decisions, enabling stakeholders to make informed choices about system design and implementation [28,12].

A key aspect is the recognition that bias-related risks are not static but evolve alongside changes in system design, data inputs, and societal contexts [33]. As a result, the risk and impact assessment process allows for a continuous evaluation that ensures algorithmic systems remain aligned with ethical, legal, and operational expectations while minimizing unintended consequences [1]. By fostering transparency through rigorous review, this approach helps organizations preemptively address potential harm, reducing the likelihood of biased outcomes that could result in reputational damage, regulatory penalties, or broader societal harm.

### 3.7 Evaluation and Monitoring

Ongoing evaluation and monitoring of bias in AI systems are essential for ensuring that these systems remain fair and equitable over time. Research indicates that biases can shift as societal norms and values evolve, necessitating regular assessments of algorithmic performance [21,8]. Techniques such as continuous monitoring, feedback loops, and adaptive algorithms can help identify and address emerging biases, thereby promoting long-term fairness in AI systems [21,8].

Additionally, the literature emphasizes the importance of transparency and accountability in the evaluation process. By making evaluation methodologies and results publicly accessible, stakeholders can hold AI developers accountable for the performance of their systems and advocate for necessary changes [4,21]. This transparency fosters trust in AI technologies and encourages a collaborative approach to addressing bias and promoting equity.

Hopefully, it has become clear to the reader that addressing bias in autonomous intelligent systems requires a comprehensive understanding of various dimensions, including algorithmic decision-making, stakeholder identification, data representation, risk and impact assessment, and ongoing evaluation. By synthesizing insights from the literature, it becomes clear that a multifaceted approach is necessary to mitigate bias and promote fairness in AI systems. This approach must prioritize diverse representation, rigorous data quality assessments, and transparent evaluation methodologies to ensure that AI technologies serve the needs of all stakeholders equitably.

## 4 Evaluation Framework and Methodology

### 4.1 Assessing Guideline Effectiveness

The effectiveness of IEEE Std 7003-2024 as a framework for bias mitigation in AIS requires systematic evaluation. This section outlines the evaluation framework and methodology used to assess the standard's practical applicability, feasibility, and effectiveness in addressing bias. To systematically evaluate IEEE Std 7003-2024, we employed a multi-faceted approach incorporating qualitative and quantitative methods. The assessment focused on three key areas.

1. The first is clarity and comprehensiveness, or whether the standard provides explicit, actionable guidance that is sufficiently detailed to be applied in both technical and managerial situations [10].
2. Next is focusing on practical implementation, or the feasibility of directly integrating the standard into existing AI development workflows [6].
3. Lastly, is measuring impact on bias mitigation, or the extent to which adherence to the standard can reduce unwanted bias and ultimately enhance fairness in AI systems [38].

### 4.2 Bias Mitigation and Stakeholder Goals

Given that bias in AI systems affects multiple stakeholders including developers, regulators, businesses, and end-users, our assessment also considered the alignment of IEEE Std 7003-2024 with the goals of these groups. Our examination focused on two key areas.

1. The first is whether the bias profile and risk assessment methodologies account for ethical concerns while simultaneously incorporating stakeholder perspectives [1].
2. Next is the degree to which the standard facilitates transparency and accountability in AI decision-making processes [1].

## 5 Analysis and Results

### 5.1 Existing Strengths

A critical strength of IEEE Std 7003-2024 lies in Clause 4 through its structured approach to documenting bias considerations. The requirement to maintain a bias profile establishes a clear framework for ensuring transparency and accountability throughout an AI system's life cycle [1]. By enforcing consistent documentation, the standard allows auditors and regulators to trace how bias risks were identified and addressed at each stage of system development. This enhances the reproducibility and reliability of bias mitigation efforts, particularly in high-stakes domains like healthcare [7].

The standard also offers clear steps for recording bias in Clause 5, providing organizations with a structured methodology for tracking and addressing algorithmic bias [1]. By requiring AI developers to document bias-related decisions, the standard fosters an evidence-based approach to mitigating unfair outcomes. This emphasis on record-keeping ensures that bias mitigation is not merely a theoretical exercise but a practical, actionable process that can be evaluated, revised, and audited over time [9].

Another notable strength is the standard's emphasis on stakeholder mapping in Clause 6, which mandates early identification of both system influencers and impacted groups. This requirement compels developers to consider diverse perspectives in the early stages of system design [1]. Without sufficient representation of diverse engineers and stakeholders, society risks ceding control to biased AI systems, reinforcing existing inequities and diminishing human agency in critical decision-making processes [6]. By explicitly defining stakeholder engagement as an essential step, the standard mitigates the risk of biased assumptions driving system design choices.

Lastly, the standard provides robust guidelines for risk assessment, particularly through Clause 8, which outlines structured procedures for assessing the likelihood and severity of bias-related harms. The standard's approach aligns with best practices in AI ethics by incorporating quantitative and qualitative risk analysis to evaluate the impact of AI decisions on different demographic groups. For instance, IEEE Std 7003-2024 requires developers to define application boundaries, the specific contexts in which an AI system is intended to operate, thus preventing the model from being deployed in scenarios where it has not been validated [1].

### 5.2 Proposed Improvements

Despite these strengths, the standard exhibits notable gaps that require further refinement. One critical limitation is the lack of specific, quantifiable metrics for determining whether dataset representation is sufficient to mitigate bias. While the standard advocates for datasets that reflect all relevant stakeholder groups, it does not specify industry-specific thresholds for data diversity. This ambiguity has prompted calls for sector-specific annexes that tailor bias mitigation requirements to fields such as finance, healthcare, and criminal justice. In industries like recruitment, biased models have historically disproportionately impacted marginalized communities, yet the standard does not provide clear statistical benchmarks for evaluating fairness [10].

Another key challenge is the conflict between competing stakeholder priorities. Businesses may prioritize efficiency and accuracy, regulators focus on compliance, and marginalized communities demand greater representation. The current framework does not provide guidance on resolving these tensions, leaving developers without a clear decision-making framework when faced with competing ethical and business objectives [38]. With recent policy shifts away from DEI mandates, companies may be less incentivized to prioritize fairness, opting instead for performance-driven AI models. The lack of safeguards within the standard to address bias mitigation in environments where DEI policies are weakened or repealed poses a significant risk. Without explicit enforcement mechanisms, companies may deprioritize fairness considerations, leading to regressions in AI equity [30].

Furthermore, existing AI governance structures often place significant control in the hands of corporate entities, which may introduce biases through content moderation practices. Recent challenges with AI-generated content, such as DeepSeek's censorship issues, illustrate how regulatory compliance efforts can themselves create new forms of bias [20]. When AI developers adjust models to avoid controversy, they risk overcorrecting, suppressing valid perspectives, and reinforcing dominant narratives. The standard does not explicitly address the balance between fairness and ideological neutrality, leaving a gap in the guidance for navigating AI governance in politically charged environments.

Moreover, while the standard provides broad recommendations for risk assessment methodologies, it lacks detailed operational guidance on implementation. Many organizations struggle with embedding fairness audits into their existing AI development workflows, particularly in cases where bias mitigation may conflict with performance optimization goals [38]. Practical steps, such as requiring external fairness evaluations, implementing anonymous bias audits, or mandating independent third-party oversight, would strengthen compliance mechanisms and ensure that bias mitigation efforts are not merely self-reported but subject to rigorous verification.

To address these gaps, the standard should consider enhancing its sector-specific guidance, providing clearer conflict resolution mechanisms, and introducing stronger safeguards for fairness enforcement. Anonymous data audits, external fairness evaluations, and clearer dispute resolution protocols would provide organizations with practical tools to balance competing stakeholder priorities while ensuring adherence to bias mitigation principles [13]. By refining these areas, IEEE Std 7003-2024 could more effectively support researchers, practitioners, and regulators in the development of equitable and accountable AI systems.

## 6 Conclusion

This paper has evaluated the recently introduced IEEE Standard for Algorithmic Bias Considerations (IEEE Std 7003-2024) as a structured framework for mitigating bias in autonomous intelligent systems. Our analysis reveals that the standard represents a significant advancement in the field, offering a comprehensive blueprint that integrates multiple dimensions of bias management—from the creation of a bias profile and rigorous stakeholder identification to careful data representation and dynamic risk and impact assessment.

By mandating systematic documentation of bias-related decisions throughout an AIS's lifecycle, the standard sets a new benchmark for transparency and accountability in AI development. Its structured approach provides clear guidance for both technical and managerial practices, thereby bridging a critical gap in the current landscape of AI ethics and regulation.

At the same time, our evaluation identifies opportunities for further enhancement. The standard would benefit from the inclusion of specific, quantifiable metrics to assess data representativeness and bias-related risks, which are essential for its practical application in complex, real-world scenarios. Additionally, while the emphasis on stakeholder engagement is commendable, more detailed mechanisms for reconciling competing stakeholder priorities could further strengthen the framework. Sector-specific annexes and operational guidelines are also recommended to tailor the standard to the diverse challenges encountered across different industries.

The IEEE Std 7003-2024 marks a transformative step toward the regulation of bias in AI systems. Its structured methodology not only advances academic discourse but also provides practical tools for developers, regulators, and other stakeholders. With continued refinement, empirical validation, and collaborative efforts among industry and academia, this standard holds great promise for fostering the development of autonomous intelligent systems that are both fair and accountable.

**Acknowledgments.** The authors thank the Rivas.AI Lab (`https://lab.rivas.ai`) for the support and helpful feedback throughout this project. This research was, in part, funded by the National Science Foundation under grant CNS-2136961.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. IEEE standard for algorithmic bias considerations. IEEE Std 7003-2024 pp. 1–59 (2025). `https://doi.org/10.1109/IEEESTD.2025.10851955`

2. Agarwal, A., Agarwal, H.: A seven-layer model with checklists for standardising fairness assessment throughout the ai lifecycle. Ai and Ethics **4**, 299–314 (2023). https://doi.org/10.1007/s43681-023-00266-9
3. Agbese, M., Mohanani, R., Khan, A., Abrahamsson, P.: Implementing ai ethics: Making sense of the ethical requirements. Association for Computing Machinery (2023). https://doi.org/10.1145/3593434.3593453
4. Ananny, M., Crawford, K.: Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society **20**, 973–989 (2016). https://doi.org/10.1177/1461444816676645
5. Asante, K., Sarpong, D., Boakye, D.: On the consequences of ai bias: when moral values supersede algorithm bias. Journal of Managerial Psychology (2024). https://doi.org/10.1108/jmp-05-2024-0379
6. Ashok, M., Madan, R., Joha, A., Sivarajah, U.: Ethical framework for artificial intelligence and digital technologies. International Journal of Information Management **62** (2022). https://doi.org/10.1016/j.ijinfomgt.2021.102433
7. co author, C., co author, J., Alhuwail, D., Peltonen, L., Topaz, M., Block, L.: The untapped potential of nursing and allied health data for improved representation of social determinants of health and intersectionality in artificial intelligence applications: a rapid review. Yearbook of Medical Informatics **31**, 094–099 (2022). https://doi.org/10.1055/s-0042-1742504
8. Bhattacharya, A., Stumpf, S., Verbert, K.: Representation debiasing of generated data involving domain experts. In: Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization. pp. 516–522 (2024). https://doi.org/10.1145/3631700.3664910
9. Bunn, J.: Working in contexts for which transparency is important. Records Management Journal **30**, 143–153 (2020). https://doi.org/10.1108/RMJ-08-2019-0038
10. Chen, Z.: Ethics and discrimination in artificial intelligence-enabled recruitment practices. Humanities and Social Sciences Communications **10** (2023). https://doi.org/10.1057/s41599-023-02079-x
11. Cook, L., Sachs, J., Weiskopf, N.: The quality of social determinants data in the electronic health record: a systematic review. Journal of the American Medical Informatics Association **29**, 187–196 (2021). https://doi.org/10.1093/jamia/ocab199
12. Dancy, C., Saucier, P.: Ai and blackness: toward moving beyond bias and representation. Ieee Transactions on Technology and Society **3**, 31–40 (2022). https://doi.org/10.1109/tts.2021.3125998
13. Deshpande, A.: Regulatory compliance and ai: Navigating the legal and regulatory challenges of ai in finance. In: 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS). vol. 1, pp. 1–5. IEEE (2024)
14. Durach, C., Kembro, J., Wieland, A.: A new paradigm for systematic literature reviews in supply chain management. Journal of Supply Chain Management **53**, 67–85 (2017). https://doi.org/10.1111/jscm.12145
15. Franklin, M.: S. voeneky, p. kellmeyer, o. mueller and w. burgard (eds) cambridge handbook of responsible artificial intelligence: interdisciplinary perspectives. Prometheus **39** (2023). https://doi.org/10.13169/prometheus.39.1.0066
16. Guo, A., Kamar, E., Vaughan, J., Wallach, H., Morris, M.: Toward fairness in ai for people with disabilities sbg@a research roadmap. Acm Sigaccess Accessibility and Computing pp. 1–1 (2020). https://doi.org/10.1145/3386296.3386298

17. Gutierrez, M.: New feminist studies in audiovisual industries| algorithmic gender bias and audiovisual data: A research agenda. International Journal of Communication **15**(0) (2021), https://ijoc.org/index.php/ijoc/article/view/14906
18. Hanna, M.G., Pantanowitz, L., Jackson, B., Palmer, O., Visweswaran, S., Pantanowitz, J., Deebajah, M., Rashidi, H.H.: Ethical and bias considerations in artificial intelligence/machine learning. Modern Pathology **38** (2025). https://doi.org/10.1016/j.modpat.2024.100686
19. High, T.R., Jordan, J.M., Ostrager, A., LLP, S..C.: President trump acts to roll back dei initiatives (2025), https://corpgov.law.harvard.edu/2025/02/10/president-trump-acts-to-roll-back-dei-initiatives/, accessed: 2025-02-14
20. Jora, R.B., Sodhi, K.K., Mittal, P., Saxena, P.: Role of artificial intelligence (ai) in meeting diversity, equality and inclusion (dei) goals. Conference on Advanced Computing and Communication Systems (ICACCS) pp. 1687–1690 (2022). https://doi.org/10.1109/ICACCS54159.2022.9785266
21. Kale, A., Nguyen, T., Harris, F., Li, C., Zhang, J., Ma, X.: Provenance documentation to enable explainable and trustworthy ai: a literature review. Data Intelligence **5**, 139–162 (2023). https://doi.org/10.1162/dint_a_00119
22. Kamikubo, R., Wang, L., Marte, C., Mahmood, A., Kacorri, H.: Data representativeness in accessibility datasets: A meta-analysis. In: Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility. pp. 1–15 (2022). https://doi.org/10.48550/arxiv.2207.08037
23. Khan, R.S., Sirazy, R.M., Das, R., Rahman, S.: An ai and ml-enabled framework for proactive risk mitigation and resilience optimization in global supply chains during national emergencies. Sage Science Review of Applied Machine Learning **5**, 127–144 (2022), https://journals.sagescience.org/index.php/ssraml/article/view/214
24. Lockwood, A.: Mitigating ai bias in school psychology: Toward equitable and ethical implementation (Nov 2024). https://doi.org/10.31234/osf.io/mh4rj, osf.io/preprints/psyarxiv/mh4rj_v1
25. Manyika, J., Silberg, J., Presten, B.: What do we do about the biases in ai? (2019), https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai, accessed: 2025-02-15
26. Moore, R.: Trump's executive orders rolling back dei and accessibility efforts, explained (2025), https://tinyurl.com/3u333edh, accessed: 2025-02-15
27. Otis, N.: Global evidence on gender gaps and generative ai (2024). https://doi.org/10.31219/osf.io/h6a7c
28. O'Brien, J., Nelson, C.: Assessing the risks posed by the convergence of artificial intelligence and biotechnology. Health Security **18**, 219–227 (2020). https://doi.org/10.1089/hs.2019.0122
29. Park, J.S., Bernstein, M.S., Brewer, R.N., Kamar, E., Morris, M.R.: Understanding the representation and representativeness of age in ai data sets. In: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society. pp. 834–842 (2021). https://doi.org/10.48550/arxiv.2103.09058
30. Press, A.: These u.s. companies are pulling back on diversity initiatives (2025), https://time.com/7209960/companies-rolling-back-dei/, accessed: 2025-02-16
31. Release: Release: Gottheimer, lahood introduce new bipartisan legislation to protect americans from deepseek (2025), https://tinyurl.com/yc4xawad, accessed: 2025-02-13

32. Schiff, D., Borenstein, J., Biddle, J., Laas, K.: Ai ethics in the public, private, and ngo sectors: a review of a global document collection. Ieee Transactions on Technology and Society **2**, 31–42 (2021). https://doi.org/10.1109/tts.2021.3052127
33. Schwartz, R., Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., Hall, P.: Towards a standard for identifying and managing bias in artificial intelligence, vol. 3. US Department of Commerce, National Institute of Standards and Technology (2022). https://doi.org/10.6028/nist.sp.1270
34. Shahbazi, N., Yin, L., Asudeh, A., Jagadish, H.: Representation bias in data: a survey on identification and resolution techniques. Acm Computing Surveys **55**, 1–39 (2023). https://doi.org/10.1145/3588433
35. Sreerama, J.: Ethical considerations in ai addressing bias and fairness in machine learning models. Journal of Knowledge Learning and Science Technology Issn 2959-6386 (Online) **1**, 130–138 (2022). https://doi.org/10.60087/jklst.vol1.n1.p138
36. Tatman, R.: Gender and dialect bias in youtube's automatic captions. In: Proceedings of the first ACL workshop on ethics in natural language processing. pp. 53–59 (2017). https://doi.org/10.18653/v1/w17-1606
37. Varona, D., Suárez, J.: Discrimination, bias, fairness, and trustworthy ai. Applied Sciences **12**, 5826 (2022). https://doi.org/10.3390/app12125826
38. Wan, Y., Wang, W., He, P., Gu, J., Bai, H., Lyu, M.R.: Biasasker: Measuring the bias in conversational ai system. Association for Computing Machinery p. 515–527 (2023). https://doi.org/10.1145/3611643.3616310
39. Wulandari, A., Diko, M.: Hr management transformation in indonesia msmes: the role of ai in sop making and recruitment. Journal of Ecohumanism **3** (2024). https://doi.org/10.62754/joe.v3i7.4641
40. Zhang, J., Zhang, Z.: Ethics and governance of trustworthy medical artificial intelligence. BMC Medical Informatics and Decision Making **23** (2023). https://doi.org/10.1186/s12911-023-02103-9