World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Navigating the Complex Landscape of AI Ethics and Privacy

Meghana Bhimavarapu *

Strategic Healthcare Programs, USA.

World Journal of Advanced Research and Reviews, 2025, 26(01), 423-430

Publication history: Received on 25 February 2025; revised on 03 April 2025; accepted on 05 April 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.26.1.1085

Abstract

The rapid integration of Artificial Intelligence systems across diverse sectors of society has generated unprecedented challenges for privacy, ethics, and accountability. This article examines the complex relationship between AI functionality and individual privacy rights, highlighting what researchers term the "privacy paradox"—the disconnect between users' stated privacy concerns and their online behaviors. It explores how sophisticated data collection methods often operate without meaningful user consent, creating pervasive surveillance networks that disproportionately impact marginalized communities. It investigates algorithmic bias and its manifestation across various domains, including criminal justice, healthcare, and financial services, where seemingly objective systems can perpetuate and amplify existing societal inequities. Furthermore, it addresses the fundamental challenge of AI transparency, focusing on the explainability deficit in complex neural networks and the diffusion of responsibility that complicates accountability frameworks. Through analysis of current technical solutions, regulatory approaches, and ethical design principles, this article presents a comprehensive overview of emerging frameworks that aim to balance technological innovation with ethical imperatives and human rights considerations.

Keywords: Algorithmic bias; Privacy paradox; Explainable AI; Ethical frameworks; Accountability mechanisms

1. Introduction

The proliferation of Artificial Intelligence (AI) technologies across various sectors has fundamentally transformed how we live, work, and interact. From recommendation systems that curate our entertainment to predictive algorithms that inform critical decisions in healthcare, finance, and criminal justice, AI's reach continues to expand. However, this technological revolution brings with it profound ethical and privacy challenges that demand urgent attention. This article examines the multifaceted nature of these concerns and explores potential frameworks for addressing them.

2. The Privacy Paradox in AI Systems

AI systems thrive on data—vast quantities of it. This creates an inherent tension between the functionality of AI and the protection of individual privacy. A comprehensive survey conducted by the Pew Research Center reveals that an overwhelming majority of Americans feel they have very little or no control over the data that companies collect about them, and most report being very or somewhat concerned about how companies are using their personal information. Despite these concerns, the same research indicates that only a small fraction of Americans say they always or often read privacy policies before agreeing to them, while most admit they understand very little or nothing at all about current privacy laws and regulations. This fundamental disconnect illustrates what researchers term the "privacy paradox"—the contradiction between users' stated privacy concerns and their actual behavior online [1].

* Corresponding author: Meghana Bhimavarapu

Table 1 Core Components of the Privacy Paradox in AI Systems [1]

Component	Description	Key Concern
User Perception	Most Americans feel they have little control over data collection	Perceived lack of agency
Consent Practices	Complex terms of service agreements discourage thorough reading	Ineffective consent mechanisms
Tracking Technologies	Google tracking is present on the majority of websites; Facebook on approximately a quarter	Hidden surveillance infrastructure
Demographic Disparities	Trust in AI systems varies significantly by demographic group	Disproportionate impact on marginalized communities
Data Security	Biometric data breaches create permanent vulnerabilities	Irrevocable exposure of sensitive identifiers

3. Data Collection and Informed Consent

The effectiveness of AI systems is directly proportional to the volume and quality of data they can access. This has led to increasingly sophisticated methods of data collection, often occurring without meaningful user awareness. The notion of informed consent becomes problematic in contemporary digital environments where users encounter complex terms of service agreements designed to discourage thorough reading, data collection occurs passively through sensors and devices embedded in our environment, and secondary uses of data extend far beyond original collection purposes.

A groundbreaking study by Libert analyzing the privacy policies and tracking technologies of the most popular websites quantified this phenomenon with alarming precision. The research found that Google tracking technologies were present on a substantial majority of websites analyzed, while Facebook trackers appeared on roughly a quarter of these sites. Perhaps more concerning, when examining news websites specifically, Google tracking was detected on virtually all domains. Despite this pervasive tracking, only a small percentage of privacy policies explicitly mentioned these third-party trackers, creating a substantial gap between what users are told and what actually happens with their data. These tracking mechanisms enable tech giants to build comprehensive profiles of user behavior across the internet—profiles that subsequently feed AI systems and algorithmic decision-making processes [2].

4. Surveillance Capabilities

The combination of AI with surveillance technologies creates unprecedented monitoring capabilities. Facial recognition systems deployed in public spaces, gait recognition algorithms that can identify individuals by their walking patterns, and sentiment analysis tools that interpret emotional states all contribute to what scholars term "ambient surveillance"—the continuous, often invisible monitoring of individuals.

These technologies have been deployed globally with varying degrees of regulatory oversight. The increasing sophistication of facial recognition technologies has raised particular concerns about privacy and civil liberties. Research conducted by the Pew Research Center indicates that most Americans have heard about facial recognition technology being used by police, with just over half of Americans trusting law enforcement to use these technologies responsibly. However, trust levels vary significantly across demographic groups, with Black Americans expressing substantially less trust compared to white Americans—a reflection of historical patterns of discriminatory surveillance that modern AI systems risk amplifying [1].

5. Data Security Vulnerabilities

The centralization of vast data repositories creates significant security risks. AI systems trained on sensitive personal information become high-value targets for cyberattacks. When breaches occur, the consequences can be severe and long-lasting for affected individuals. According to data published in the ISACA Journal, the financial impact of data breaches continues to rise each year, with particularly severe costs in the healthcare sector. The typical data breach has a lengthy lifecycle (time to identify and contain), during which unauthorized parties may have continuous access to sensitive information [3].

The breach of Biostar 2, a biometric security platform, exemplifies these risks in the context of AI systems. This breach exposed millions of records containing fingerprint and facial recognition data—biological identifiers that, unlike passwords, cannot be changed if compromised. The breach included substantial amounts of sensitive biometric data along with personally identifiable information. The implications of such breaches for AI systems are particularly severe, as they compromise not only current security but also the integrity of future AI applications that might rely on biometric authentication. Organizations affected by data breaches often pass costs on to consumers, with research indicating that a majority of organizations increased the prices of their products or services following a breach, creating a secondary economic impact beyond the immediate security concerns [3].

6. Algorithmic Bias and Discrimination

Perhaps the most extensively documented ethical concern surrounding AI involves algorithmic bias—the tendency for AI systems to produce outputs that systematically disadvantage certain groups.

6.1. Sources of Bias

Algorithmic bias stems from multiple sources, with quantifiable impacts across various domains. When historical data contains patterns of discrimination, AI systems learn and perpetuate these patterns. For example, recruitment algorithms trained on historical hiring decisions may penalize candidates from underrepresented groups if those groups were historically disadvantaged in hiring. Underrepresentation of certain demographics in training data leads to reduced accuracy for those groups, a phenomenon clearly demonstrated in the influential Gender Shades study conducted by Buolamwini and Gebru. Their research evaluated commercial gender classification systems and found dramatic accuracy disparities across demographic groups. The error rates for darker-skinned females were substantially higher than those for lighter-skinned males—revealing a significant overall accuracy gap. The researchers evaluated individuals from parliaments in European and African countries, creating a balanced dataset that exposed the significant limitations of facial analysis technologies when applied across diverse populations [4].

Even when protected characteristics such as race or gender are explicitly removed from data, AI systems can identify proxy variables that correlate with these characteristics, inadvertently reintroducing bias. The Gender Shades research underscores how seemingly objective systems can produce highly disparate outcomes when deployed in real-world contexts with diverse populations. Without deliberate attention to these bias mechanisms, AI systems risk amplifying rather than mitigating societal inequities [4].

Table 2 Sources of Algorithmic Bias and Their Manifestations [4]

Bias Source	Description	Example	Impact Domain
Training Data Bias	Historical discrimination patterns learned by AI systems	Recruitment algorithms penalizing candidates from underrepresented groups	Employment
Representation Disparities	Underrepresentation of certain demographics in training data	Facial recognition systems perform worse for darker-skinned females	Public Security
Proxy Discrimination	AI systems identifying variables that correlate with protected characteristics	Zip codes serving as proxies for race in lending algorithms	Financial Services
Feedback Loops	Biased predictions leading to actions that reinforce initial bias	Predictive policing concentrating resources in already over-policed areas	Criminal Justice
Interpretability Gaps	Inability to audit decision-making processes for bias	Healthcare algorithms prioritizing care based on opaque criteria	Healthcare

6.2. Domains of Impact

The consequences of algorithmic bias manifest across numerous domains, with measurable effects that impact individuals' life opportunities and outcomes. In criminal justice settings, predictive policing algorithms and recidivism risk assessment tools have been shown to produce disparate outcomes for different racial groups. The ProPublica investigation into COMPAS, a widely used recidivism prediction algorithm in the US criminal justice system, found that Black defendants who did not re-offend were misclassified as higher risk at a significantly higher rate compared to white defendants who did not re-offend. Conversely, white defendants were labeled as lower risk but reoffended at a higher

rate, demonstrating a systematic pattern of bias that could have profound consequences for sentencing and parole decisions [4].

In healthcare contexts, diagnostic algorithms trained primarily on data from certain demographic groups may perform poorly for others. This pattern mirrors the findings from the Gender Shades study, suggesting that across domains, AI systems often perform worse for already marginalized populations. The Pew Research Center found that a majority of Americans believe computer programs will always reflect the human biases of their designers, while fewer believe it's possible to create computer programs that make decisions free from human bias—indicating public awareness of these challenges even as technical solutions remain elusive [1].

Financial services represent another domain where algorithmic bias can perpetuate historical patterns of exclusion. When lending algorithms incorporate variables that correlate with protected characteristics, they risk reproducing patterns of discrimination that have historically limited access to capital in marginalized communities. Even in employment contexts, automated screening tools can systematically disadvantage qualified candidates from underrepresented groups, particularly when trained on historical hiring data that reflects past discriminatory practices [2].

7. Transparency Challenges and Governance Frameworks

The opacity of many AI systems presents significant challenges for accountability and oversight. According to Pew Research, a large majority of Americans say it is very or somewhat important to them to understand who has access to data collected about them, yet most acknowledge having a limited understanding of what companies actually do with this data [1]. This transparency gap is mirrored in the findings from Libert's research on website tracking, which revealed that while third-party tracking is nearly ubiquitous, only a small fraction of privacy policies explicitly mention these trackers—creating a substantial information asymmetry that limits meaningful user consent [2].

Researchers have proposed various frameworks for addressing these challenges, including enhanced disclosure requirements, algorithmic impact assessments, and technical approaches to fairness and accountability. The Gender Shades project demonstrates the value of targeted evaluation of AI systems across demographic groups, providing a model for identifying and addressing disparities before deployment [4]. Similarly, the regulatory approaches emerging in different jurisdictions reflect varying perspectives on the balance between innovation and protection, with the European GDPR model emphasizing user rights and the more fragmented U.S. approach focusing on sector-specific protections.

8. The Transparency Challenge in Artificial Intelligence: Explainability, Accountability, and Ethical Solutions

The "black box" nature of many advanced AI systems—particularly deep learning models—presents significant ethical challenges. As these systems increasingly make or influence decisions with profound impacts on individuals, the inability to fully understand their internal reasoning processes raises fundamental questions about fairness, accountability, and trust. This article explores the transparency challenges inherent in modern AI systems and examines potential frameworks for addressing them.

8.1. Explainability Deficit

Complex AI systems, especially neural networks, often operate in ways that are difficult or impossible to interpret, even for their creators. This lack of explainability raises concerns when these systems influence consequential decisions about individuals. Without understanding how a system reached a particular conclusion, it becomes difficult to verify the system's reasoning, identify potential errors or biases, or contest unfavorable decisions.

This challenge is particularly evident in AI healthcare applications. Research published in The Lancet Oncology examined the performance of machine learning algorithms compared to human experts in dermatological diagnosis. While the study demonstrated that AI systems could achieve diagnostic accuracy comparable to board-certified dermatologists when identifying skin lesions, it highlighted a critical transparency issue: the algorithms provided no explanatory rationale for their classifications. This absence of interpretability meant that clinicians had no insight into which visual features the algorithm had identified as significant, potentially limiting clinical adoption despite the systems' technical performance. The study authors emphasized that in clinical contexts, knowing why a diagnosis was reached can be as important as the diagnosis itself, highlighting the tension between performance and explainability in high-stakes domains [5].

Table 3 Transparency Challenges in AI Systems [5]

Challenge	Description	Implication	Domain Example
Black Box Decision-Making	Neural networks operate in ways difficult to interpret	Cannot verify the reasoning	Dermatological diagnosis
Technical Complexity	Increasing computational resources and parameters	Widening interpretability gap	Large language models
Distributed Responsibility	Decision-making spread across humans and algorithms	Difficulty assigning accountability	High-stakes AI systems
Implementation Gaps	Organizations updating general practices without specific explainability measures	Compliance without substance	GDPR implementation
Accountability Diversion	"Algorithmic passing the buck" between operators and developers	Evasion of responsibility	Public sector AI

The technical complexity of modern AI systems exacerbates this challenge. The computational resources required for training state-of-the-art AI models have increased dramatically in recent years, with models like GPT-3 requiring significant computational resources for a single training run. This exponential increase in model complexity has outpaced advancements in explainability techniques, creating what researchers term an "interpretability gap" that widens as models become more sophisticated. As models incorporate more parameters and training data, traditional approaches to understanding their decision-making processes become increasingly inadequate [5].

9. Accountability Mechanisms

The opacity of AI systems complicates traditional accountability frameworks. When negative outcomes occur, determining responsibility becomes challenging when decision-making is distributed across human operators, organizational processes, and algorithmic systems. Research published in the Journal of Information Technology examines this problem in depth, noting that among organizations deploying high-stakes AI systems, only a minority had formal audit procedures in place to verify algorithmic outcomes, and even fewer maintained comprehensive documentation of model development decisions that would facilitate accountability [6].

The General Data Protection Regulation (GDPR) in Europe attempts to address this through provisions granting individuals the "right to explanation" for automated decisions, though the practical implementation of this right remains contested. The Journal of Information Technology study identifies that many organizations struggle to implement meaningful accountability mechanisms due to a combination of technical complexity, organizational barriers, and regulatory uncertainty. The research indicates that companies subject to GDPR have updated their data processing practices in general terms, but a significantly smaller proportion have implemented specific measures to enhance the explainability of their algorithmic systems. This implementation gap suggests that legal requirements alone may be insufficient to ensure meaningful algorithmic accountability without corresponding technical solutions and organizational processes [6].

The accountability challenge extends beyond legal compliance to include broader societal considerations. Research on algorithmic impact assessments reveals a fragmented approach to accountability across technical and organizational dimensions. Many frameworks focus primarily on either technical validation or procedural compliance, failing to address the multifaceted nature of algorithmic accountability. The study on algorithmic impact assessments identifies a phenomenon termed "accountability gaps," where organizations implement partial measures that satisfy formal requirements without addressing the substantive transparency needs of affected individuals. These gaps are particularly pronounced in public sector applications, where resource constraints may limit the implementation of comprehensive accountability frameworks [7].

The research on algorithmic impact assessments also identifies the phenomenon of "algorithmic passing the buck," where human operators blame algorithms for poor outcomes while developers attribute failures to improper use. This diffusion of responsibility creates significant challenges for establishing meaningful accountability in socio-technical systems. The effectiveness of algorithmic impact assessments varies considerably based on implementation details, with the most successful approaches incorporating both proactive evaluation during development and ongoing monitoring throughout deployment [7].

10. Frameworks for Ethical AI Development

Addressing these ethical and privacy concerns requires multi-faceted approaches spanning technical innovation, regulatory frameworks, and organizational practices.

10.1. Technical Solutions

Various technical approaches aim to mitigate specific ethical challenges in AI development and deployment. Differential privacy techniques, which add calibrated noise to data or results to protect individual privacy while preserving aggregate insights, have been implemented by organizations, including government statistical agencies. Research in healthcare contexts shows that privacy guarantees can be maintained while limiting accuracy loss for most statistical queries. These techniques enable organizations to derive valuable population-level insights while providing mathematical guarantees against individual re-identification [5].

Federated learning techniques enable AI model training across multiple devices while keeping raw data local, addressing privacy concerns by minimizing data centralization. This approach has been implemented for applications like keyboard prediction on mobile devices, demonstrating the feasibility of developing effective models without transmitting sensitive user data to central servers. The growth in federated learning implementations reflects increasing awareness of privacy concerns among both users and developers, with particular adoption in sectors handling sensitive personal information, such as healthcare and finance [6].

Explainable AI (XAI) methods enhance the interpretability of complex models through various approaches, including local explanations, counterfactual explanations, and model-agnostic interpretation techniques. Research on clinical decision support systems indicates that implementations with structured explainability features achieve better adoption rates among healthcare professionals compared to black-box alternatives. The ability to understand and interrogate model recommendations appears particularly important in domains where professional judgment and expertise remain central to decision-making processes [5].

10.2. Regulatory Approaches

Table 4 Regional Approaches to AI Governance [6]

Region	Primary Regulatory Framework	Key Features	Strengths	Limitations
European Union	General Data Protection Regulation (GDPR) & AI Act	"Right to explanation"; risk-based categorization of AI systems	Comprehensive protection; rights-based approach	Implementation challenges; interpretation ambiguity
United States	Sectoral laws (CCPA, CDPA, etc.)	State-by-state regulation focused on data rights	Innovation-friendly; adaptable to contexts	Fragmented landscape; inconsistent protection
Canada	Directive on Automated Decision-Making	Impact assessments; human oversight requirements	Public sector focus; accountability emphasis	Limited to government applications
China	Personal Information Protection Law & AI governance frameworks	National security emphasis; strategic AI development	Coordinated strategy; rapid implementation	Limited individual rights; surveillance concerns
Global Standards	IEEE, ISO, NIST frameworks	Technical standards for AI ethics and safety	Cross-border compatibility; industry expertise	Non-binding; limited enforcement mechanisms

Regulatory frameworks are evolving globally to address AI ethics and privacy concerns. The European Union's proposed AI Act seeks to categorize AI applications by risk level and impose appropriate requirements, with more stringent obligations for systems operating in high-risk domains like healthcare, education, and law enforcement. The regulatory

approach reflects a risk-based philosophy that recognizes the varying potential for harm across different applications of similar technologies [6].

In the United States, California's Consumer Privacy Act (CCPA) and Virginia's Consumer Data Protection Act introduce new data rights, including provisions relevant to automated decision systems. Research on CCPA implementation indicates that compliance required significant organizational adjustments for affected companies, with surveyed organizations reporting substantial operational changes to meet requirements. These state-level initiatives represent a fragmented approach to AI governance in the US context, creating a complex regulatory landscape for organizations operating across multiple jurisdictions [6].

Canada's Directive on Automated Decision-Making, which establishes requirements for impact assessments and human oversight, represents a public sector-focused approach to algorithmic governance. Research on algorithmic impact assessments indicates that agencies have conducted evaluations for numerous automated decision systems, with those classified as high-impact requiring enhanced transparency and human oversight provisions. This approach emphasizes the importance of tailoring governance requirements to the potential impact of automated systems, with more stringent measures applied to systems with greater potential for affecting individual rights and opportunities [7].

10.3. Ethical Design Principles

Organizations and researchers have proposed various principles for ethical AI development. Research published in the Advances in Artificial Intelligence journal surveying AI ethics frameworks found substantial convergence around core values, with transparency, fairness, and privacy appearing consistently across frameworks. However, the research also identified significant variations in how these principles are operationalized across organizations. While many entities have publicly committed to ethical AI principles, a much smaller proportion have translated these principles into specific technical requirements, and an even smaller group has implemented metrics to evaluate adherence [8].

Value alignment approaches, which ensure AI systems are designed to align with human values and societal norms, show promising results in experimental settings. Research at academic centers focused on human-compatible AI has demonstrated that systems trained with value alignment techniques can make decisions more consistent with surveyed human preferences compared to conventionally trained systems. These findings suggest that explicit attention to value alignment during development can lead to systems that better reflect human ethical intuitions, though challenges remain in defining and measuring alignment for diverse user populations [8].

Fairness by design practices, incorporating fairness considerations throughout the development lifecycle, are becoming more standardized with the development of open-source toolkits that implement various fairness metrics and bias mitigation algorithms. These tools enable developers to evaluate systems across multiple fairness criteria and implement appropriate interventions based on application requirements. Research indicates the growing adoption of these toolkits in enterprise AI development workflows, suggesting increasing awareness of fairness considerations among practitioners [8].

Privacy by design approaches, building privacy protections into systems from inception rather than as afterthoughts, show significant benefits in both regulatory compliance and breach prevention. Research published in the Journal of Information Technology found that organizations implementing privacy-by-design practices spent considerably less on compliance activities and experienced fewer data breaches compared to organizations implementing privacy measures reactively. These findings suggest that integrating privacy considerations throughout the development process creates both economic and security advantages [6].

Human-centered AI approaches, keeping humans "in the loop" for consequential decisions, are becoming standard practice in high-risk domains. The research on clinical decision support systems found that implementations with structured human oversight achieved substantial reductions in adverse events compared to fully automated systems. This finding highlights the complementary relationship between human and algorithmic capabilities, suggesting that optimal outcomes often result from thoughtfully designed human-AI collaboration rather than complete automation [5].

11. Conclusion

The transparency challenges inherent in advanced AI systems necessitate coordinated responses that span technical, regulatory, and organizational dimensions. As these technologies become increasingly embedded in decision-making processes that affect fundamental rights and opportunities, the ability to understand, verify, and contest algorithmic

decisions emerges as an essential requirement for preserving human autonomy and dignity. While technical solutions such as explainable AI, differential privacy, and federated learning demonstrate promising potential, they must be complemented by robust regulatory frameworks and organizational practices that prioritize transparency and accountability throughout the AI lifecycle. The evidence presented throughout this article underscores the profound gap between ethical principles and their practical implementation, with many organizations publicly committing to responsible AI development while failing to translate these commitments into operational reality. Addressing these challenges requires substantial investment in both developing more interpretable AI systems and creating the governance structures necessary for their responsible deployment. The path forward demands multidisciplinary collaboration between technologists, ethicists, policymakers, and affected communities to ensure that AI systems not only perform efficiently but do so in ways that are transparent, accountable, and aligned with human values and societal well-being.

References

[1] Brooke Auxier et al., "Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information," Pew Research, November 15, 2019, Online, Available: https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/

[2] Timothy Libert, "An Automated Approach to Auditing Disclosure of Third-Party Data Collection in Website Privacy Policies," (International World Wide Web Conference Committee, 2018, Available: https://timlibert.me/pdf/Libert-2018-Automated_Approach_to_Auditing_Website_Privacy_Policies.pdf

[3] Natalie Jorion, Jack Freund, et al., "The True Cost of a Data Breach," 22 February 2023, Isaca, Available: https://www.isaca.org/resources/isaca-journal/issues/2023/volume-1/the-true-cost-of-a-data-breach

[4] Joy Buolamwini, Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," Proceedings of Machine Learning Research 81:1–15, 2018, Available: https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

[5] Amina Adadi, Mohammed Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," September 2018, IEEE Access PP(99):1-1, DOI:10.1109/ACCESS.2018.2870052, Available: https://www.researchgate.net/publication/327709435_Peeking_Inside_the_Black-Box_A_Survey_on_Explainable_Artificial_Intelligence_XAI

[6] Philipp Tschandl et al., "Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study," The Lancet Oncology, Volume 20, Issue 7, July 2019, Pages 938-947, Available: https://www.sciencedirect.com/science/article/abs/pii/S147020451930333X

[7] Jacob Dexe et al., "Explaining automated decision-making: a multinational study of the GDPR right to meaningful information," The Geneva Papers on Risk and Insurance 03 May 2022, Volume 47, pages 669–697, (2022), Available: https://link.springer.com/article/10.1057/s41288-022-00271-9

[8] Jacob Metcalf, et al, "Algorithmic Impact Assessments and Accountability: The Co-construction of Impacts," 2021, ACM, https://doi.org/10.1145/3442188.3445935, Available: https://ranjitsingh.me/wp-content/uploads/2021/04/AIAs_and_Accountability_JM_etal.pdf