Afro-Asian Journal of Scientific Research (AAJSR)
المجلة الأفروآسيوية للبحث العلمي (AAJSR)
E-ISSN: 2959-6505
Volume 3, Issue 1, January - March 2025
Page No: 175-185
Website: https://aajsr.com/index.php/aajsr/index
SJIFactor 2024: 5.028 ISI 2024: 0.580 معامل التأثير العربي (AIF) 2024: 0.74

Ethical Implications of AI in Automated Decision-Making Systems

Elmakzum Kamis Elmakzum Elgedik [1*], Hazem Abdalgader Amer Salem [2]
[1] Department of Computer Science, Faculty of Science, Bani Walid University, Bani Walid, Libya
[2] Department of Computer Science, Higher Institute of Engineering Technologies, Sabha, Libya

الآثار الأخلاقية للذكاء الاصطناعي في أنظمة صنع القرار الآلية

المخزوم خميس المخزوم الجدك [1*]، حازم عبدالقادر عامر سالم [2]
[1] قسم علوم الحاسوب، كلية العلوم، جامعة بني وليد، بني وليد
[2] قسم الكمبيوتر، المعهد العالي للتقنيات الهندسية، سبها، ليبيا

*Corresponding author: mkmkjd85@gmail.com

Received: October 14, 2024	Accepted: November 15, 2024	Published: January 15, 2025

Abstract:
Artificial Intelligence is rapidly embedding in Automated Decision-Making (ADM) systems that transform key sectors like finance, healthcare, or criminal justice with increased decision-making efficiency and scalability. However, the deployment of ADM systems introduces glitches, which are major ethical challenges: bias, transparency, privacy, and accountability. Ethics here review that ADM systems, optimized for performance, often reflect historical biases embedded in their training data which might affect social inequalities. Moreover, the "black-box" nature of many machine learning models is impervious to transparency, further complicating accountability when decisions yield unfavourable results. The analysis shows that ethics schemes should be developed to guide the responsible development and deployment of ADM systems; fairness-aware algorithms, explainable AI techniques, and robust data governance form a suite of basic elements to safeguard individual rights and foster public trust. This paper concludes by recommending ways in which policymakers and practitioners can ensure ADM systems uphold values in society by weighing technological innovation against ethical integrity.

Keywords: Automated Decision-Making, Artificial Intelligence, Ethical Implications, Bias, Transparency, Accountability, Privacy.

الملخص
إن الذكاء الاصطناعي يدمج بسرعة في أنظمة صنع القرار الآلي التي تحول القطاعات الرئيسية مثل التمويل أو الرعاية الصحية أو العدالة الجنائية مع زيادة كفاءة صنع القرار وقابلية التوسع. ومع ذلك، فإن نشر أنظمة صنع القرار الآلي يقدم خللًا، وهي تحديات أخلاقية رئيسية: التحيز والشفافية والخصوصية والمساءلة. تستعرض الأخلاقيات هنا أن أنظمة صنع القرار الآلي، المحسنة للأداء، غالبًا ما تعكس تحيزات تاريخية مضمنة في بيانات التدريب الخاصة بها والتي قد تؤثر على التفاوتات الاجتماعية. وعلاوة على ذلك، فإن طبيعة "الصندوق الأسود" للعديد من نماذج التعلم الآلي لا تتأثر بالشفافية، مما يزيد من تعقيد المساءلة عندما تسفر القرارات عن نتائج غير مواتية. يُظهر التحليل أنه يجب تطوير مخططات الأخلاق لتوجيه التطوير والنشر المسؤول لأنظمة صنع القرار الآلي؛ تشكل الخوارزميات الواعية بالعدالة وتقنيات الذكاء

الاصطناعي القابلة للتفسير وحوكمة البيانات القوية مجموعة من العناصر الأساسية لحماية الحقوق الفردية وتعزيز الثقة العامة. تختتم هذه الورقة بالتوصية بالطرق التي يمكن لصناع السياسات والممارسين من خلالها ضمان دعم أنظمة صنع القرار الآلي للقيم في المجتمع من خلال موازنة الابتكار التكنولوجي بالنزاهة الأخلاقية.

الكلمات المفتاحية: اتخاذ القرارات الآلية، الذكاء الاصطناعي، التداعيات الأخلاقية، التحيز، الشفافية، المساءلة، الخصوصية.

Introduction

As artificial intelligence (AI) is rapidly integrated into Automated Decision Making (ADM) systems, decision-making itself has changed in finance, healthcare, and even criminal justice. Such systems make possible the analysis of large volumes of data in order to come to conclusions with limited human intervention at unprecedented efficiency and scalability. For instance, ADM systems may decide on credit applications in a couple of seconds while supporting diagnoses of medical conditions or informing the police about the allocation of resources based on predictive models. However, with society getting increasingly dependent on such autonomous systems, ethical reflections about their deployment have become a very hot debate.

The underlying concern is rooted in the perception that automated decision-making, or ADM, systems are effective but do not necessarily promote heritable values such as fairness, accountability, and transparency. Whereas human reasoning works on assumptions of reasons, AI systems analyse data and find patterns which are often historically flawed and prejudiced. Illustratively, a recruitment model built on the basis of historical algorithms may be biased towards a woman or an ethnic group based strictly on previously held stereotypes if such stereotypes were present in the initial data. In addition, many state-of-the-art machine learning algorithms are considered “black boxes” and the reasoning behind a choice is not always evident which presents accountability issues for the systems when erroneous and harmful decisions are made.

Taking into account the capability of ADM systems to affect lives and outcomes, it becomes imperative to reflect on their ethical aspects. This paper discusses the ethical challenges encountered in the course of AI applied in automated decision-making systems, which include aspects like bias, transparency, privacy, and accountability. It then discusses some of the frameworks and guidelines that may assist in making sure that ADM systems are used in a safe and ethical manner. To put it simply, in a time when AI is increasingly getting involved in decision-making processes, it is no longer optional to respond to these issues, but it is crucial for the promotion of trust and fairness in outcomes for all AI applications.

The Role of AI in Automated Decision-Making

Automated Decision Making (ADM); systems are the sets of algorithms or, say, ways of deciding an action on data via computation by itself- or performing an automated process devoid of any influence from human attention. Systems use artificial intelligence (AI) and, increasingly, machine learning- permeating all fields with complex information analysis and pattern identification to make decisions or recommendations from audiovisual, spatial, real-time, or internal databases of processed information. The application is not limited to this. It covers different sectors, such as human resources, financial services, healthcare, and public safety. In the recruitment sector, for example, ADM systems help in the selection of applicants by using AI tools to analyse elaborate resumes, match job requirements with the corresponding skill sets of an applicant, and filter out applicants who do not meet the requisite qualifications; thus, the system hugely increases the efficiency of speeded selection and potentially reduces bias (Raghavan et al., 2020).

Similarly, ADM systems can be used to automate loan approval in the financial sector by determining the expected credit risk associated with a potential customer in an informed, data-driven way to expedite the approval process (Cowgill et al., 2020; Fuster et al., 2019). This is usually faster and considered more objective than decision-making by human judgment. Similarly, health applications also demonstrate the transformational power of ADM through AI-powered diagnosis, which is done using medical images and patient histories for early detection and personalized treatment for better patient outcomes (Topol, 2019; Esteva et al., 2019). Predictive policing is another area where ADM is employed; this area checks historical crimes to utilize the available resources better, although this application is highly scrutinized due to ethical and biased points of view (Brantingham et al., 2018).

ADM systems have several advantages in terms of efficiency, objectivity, and scalability. Thus, ADM enhances processing speed through the automation of repetitive and data-intensive tasks, which is essential in high-decision-volume operations in the financial and retail industries (Mehrabi et al., 2021). Moreover, when the data used in designing ADM systems is clear and unbiased, a degree of standardization, through which subjective biases on the part of humans can be minimized, is availed to arrive at more objective decision-making processes (Barocas et al., 2023). Other advantages include

scalability because ADM systems can handle increasing workloads without necessarily having to increase resources proportionally, thereby rendering it particularly beneficial for organizations with resource optimization needs (O'Neil, 2017). Despite these gains, there are also challenges in that ADM systems can perpetuate prejudices integrated into their training data; hence, again, responsible design and implementation are vital to avoid such unfair outcomes (Noble, 2018). In this regard, this present paper discusses the positioning of ADM in the making of decisions in all sectors, with a focus on an efficient-equitable balance as technologies of artificial intelligence continue to develop.

Ethical Challenges in AI-Driven ADM Systems

Bias and Fairness

Biases can be transmitted onto AI-driven ADM systems through biases in the training data that then produce discriminatory or unfair outcomes. Training data reflect the patterns of history and society alike, including all human prejudices, and when objectified into AI systems, these can be furthered or even magnified. For example, an ADM system used for hiring, if it is trained with historical data from male-dominated industries, may inadvertently show a predisposition toward males because of the developed patterns in the information that have been fed into it. It will make biased decisions on hires based on gender (Raghavan et al., 2020). Such biases can seep into other aspects of decision-making and result in disparate prejudicial treatment against persons or groups who are already disadvantaged or underrepresented.

A very good example of this is the COMPAS algorithm, which has been used in predictive policing in the United States and also has its own set of racial biases in the risk assessment scores. Indeed, studies have shown that it defines Black defendants as a higher risk than white ones with similar criminal records, leading to Black people being assessed for tougher legal sentences and contributing to systemic racial disparities (Angwin et al., 2016). Bias appeared in hiring when the system, in Amazon's experimental recruitment algorithm, was found to downgrade women's resumes by using language associated with women, based on historical data biased towards male-dominated language in resumes (Dastin, 2022). These cases illustrate the potential of ADM to lock in the existing inequities if such data are uncritically accepted and no correction mechanisms are in place.

Given this, the work goes toward applying fairness-aware algorithms that consciously reduce dependence on biased features or incorporate fairness within the constraints while training models (Mehrabi et al., 2021). True fairness, however, is difficult to achieve, and it has to be a conscious balancing act between accuracy and ethics, with continuous monitoring, transparent auditing, and diverse datasets to reflect variation in societal realities to a better extent in order to minimize discriminatory outcomes (Noble, 2018; Barocas et al., 2023).

Transparency and Explainability

The important ones are deep learning and complex neural network models, somewhat black box-like in nature, that form significant barriers towards understanding how decisions are obtained from these systems (Samek, Wiegand, & Müller, 2017). This is because the processes of reaching a decision in these kinds of models involve so many layers and parameters it becomes, at best, extremely hard for human reasoning to trace the exact pathways through which given inputs lead to a certain output (Doshi-Velez & Kim, 2017). Accountability is a significant issue in AI, where it is required that developers, users, and regulatory bodies should have the ability to investigate the decision-making process of this type of model for their compliance with ethical, legal, and social standards (Guidotti et al., 2018). Transparency in this respect is absolutely critical for accountability. In the absence of such an ability to probe, users would find little or no reason to believe AI systems in areas of greater importance, like healthcare or even criminal justice, since unjustified or biased verdicts lead to disastrous results. In this regard, Explainable AI techniques are being developed that either provide models that turn out to be intrinsically more interpretable or build surrogate models approximating complex systems in a way that attempts for transparency and accuracy without ever completely compromising model accuracy (Arrieta et al., 2020). By making the AI models more understandable, transparency also brings in accountability whereby systems can be audited for errors or biases, and intervention is made where necessary, hence enabling more responsible AI deployment (Adadi & Berrada, 2018).

Privacy Concerns

In general, an ADM system needs a great amount of personal information for good predictions and decisions. This is naturally related to considerable privacy risks (Zarsky, 2016). Since these systems are operated on sensitive information like financial records, health data, or behavioural patterns, the criticality of risk becomes very high due to the possibility of data misuse or unauthorized access. This dependency on personal data is increasing not only the probability of data breaches; it also raises questions as to how safely and with respect to ethical considerations this information is treated within

the ADM system (Mantelero, 2018). In the absence of proper privacy protection, the person may be caught up in some unpleasant aftermath of data disclosure, such as identity theft, profiling, and even discrimination drawing upon unauthorized or incorrect inferences from that data (Van der Sloot, 2017). Moreover, data in ADM systems can be disclosed to unauthorized parties, sometimes weakening public confidence in those sectors, which always deal with sensitive information, such as financial, healthcare, and law enforcement institutions. Legal frameworks, such as the General Data Protection Regulation of the EU, generally attempt to minimize these risks by implementing strict guidelines concerning data collection, processing, and storage. Still, given the rapidity of the evolution of ADM technology, many times these regulations cannot keep pace with such growth (Kaminski, 2021). Privacy in ADM systems would therefore need to be underpinned by robust data governance, transparency in the use of data, and state-of-the-art cybersecurity measures that have checks against misuse and unauthorized access. This will balance innovation with respect for individual privacy.

Autonomy and Control

Greater machine autonomy in making decisions that have traditionally required human judgment raises serious ethical issues in terms of displacement of human oversight in such areas as healthcare, criminal justice, and finance (Coeckelbergh, 2020). As machines increasingly make decisions, responsibility, within the realm of ethics, takes precedence when those decisions bear adverse impacts on the affected individuals. It follows that increased reliance on machine rather than human decision-making threatens to undermine accountability since it increasingly becomes unclear who, if anyone, can be held responsible for decisions reached by algorithms or other AI-driven systems (Matthias, 2004). For instance, an autonomous system may reach a decision that is harmful, and often one would not be in a position to point out any developers, operators, or the AI itself who could be responsible; thus, possible accountability gaps appear (Floridi & Cowls, 2022; Khaleel et al., 2023). This ambiguity in accountability does raise serious concerns not only on the aspect of fairness but also on ethical lapses that have a bearing on the well-being and rights of individuals. Seeing as AI technologies are growing, much more needs to be done in accountability frameworks to make quite sure human oversight and responsibility remain central, even in cases where increased autonomy is granted to machines. Setting clear lines demarcated on the assignment of responsibility, along with the development of AI systems capable of delivering explainable and auditable decisions, is important to maintain trust and ethical standards within AI applications (Mittelstadt, 2019; Khaleel et al., 2024).

Economic and Social Impact

While improving efficiency and productivity, ADM systems are able to bring in major economic and social consequences, like job loss, widening of inequalities, and deepening of social gaps. Integration of ADM in such industries as manufacturing, retail, and even white-collar sectors may displace certain jobs traditionally held by humans, thus causing significant shifts in the labour market, such as rises in unemployment, especially among the low-skilled workers (Susskind & Susskind, 2022). Hence, this displacement affects the populations that are already vulnerable and contributes to further deepening income and opportunity inequalities (Acemoglu & Restrepo, 2018). ADM systems can sometimes inadvertently further stratify society, especially when biases are baked into algorithms that extend discriminatory practices against demographic subgroups in lending, hiring, and law enforcement (Noble, 2018). That is to say, algorithmic bias often bears down particularly hard on racial minorities, exacerbating old social inequalities rather than overcoming them (Eubanks, 2018). This means that, as ADM evolves, policy interventions at the comprehensive and multi-faceted levels are needed to mitigate the negative social consequences of job replacement by upskilling programs, guidelines on ethics to ward off bias, and regulatory mechanisms for fairness across all sections of society (Brynjolfsson & McAfee, 2014).

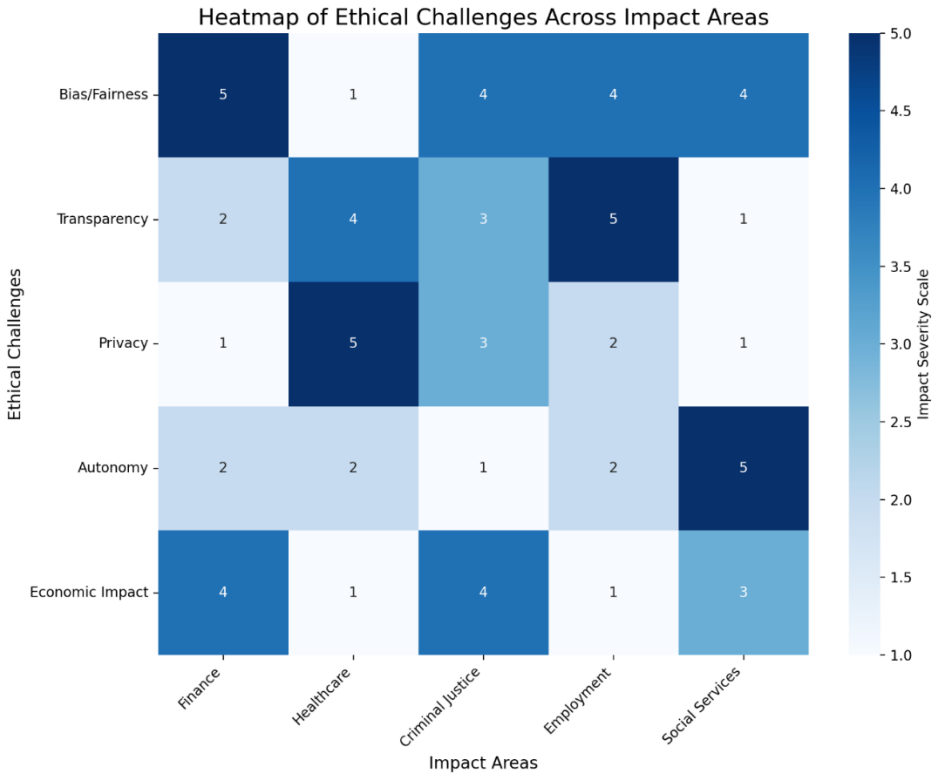

Figure 1. Comparative Analysis of Ethical Challenges in ADM Systems.

Frameworks for Ethical AI in ADM Systems

Ethics by Design

Ethics should underpin the design of automated systems so that considerations with respect to fairness, accountability, and transparency remain foundational, rather than retroactive.

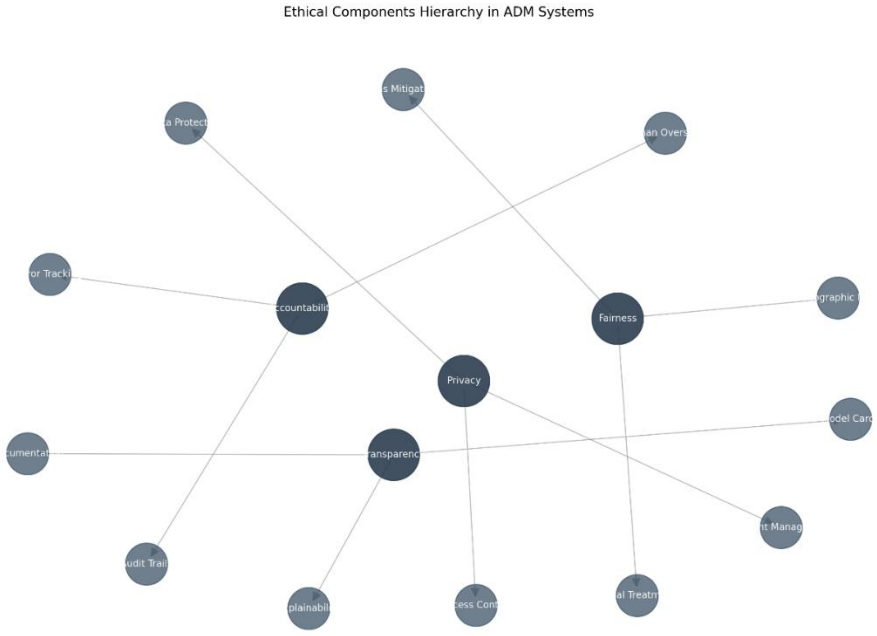

Figure 2. Conceptual Framework of Ethical Components in AI-Driven Automated Decision-Making Systems.

Ethics by design proactively calls for embedding ethical standards within the development process, calling for designers and developers to account for potential biases from the outset, ensuring accountability, and maintaining transparency (Dignum, 2019). It involves the realization of frameworks,

such as Fairness, Accountability, and Transparency, which guide a developer right from the very beginning in proactive identification and mitigation of ethical risks, which may create harmful biases, leading to discrimination or other forms of unfair treatment (Floridi & Cowls, 2022). Ethics at an early stage of development allows developers to create systems that have room for human oversight, giving way to transparency in automated decision-making processes, and introducing mechanisms for traceability and explainability (Jobin, Ienca, & Vayena, 2019). Ethics-by-design allows the developer to warrant accountability by making the systems designed in a way that enables audits and tracking of the decision paths such that once there is an error or bias, they can be fixed over time. Ethical principles, in this approach, are integrated a priori in order to ensure automated systems fully conform not only with existing regulations but also further trust and equity in their interaction with users (Mittelstadt 2019).

Regulation and Governance

The rapid growth of AI and ADM systems demands the establishment of robust legal frameworks combined with regulatory oversight of ethical and responsible use. Existing regulations, such as those under the GDPR in the European Union, constitute foundational guidelines for the protection of individual privacy as a means of ensuring data security in AI applications, especially through provisions like the right to explanation and the right to be forgotten (Kaminski, 2021). These frameworks also fail to deal with the emerging challenges regarding AI development in terms of raising new ethical issues and addressing concerns about fairness, accountability, and transparency within a wide range of domains. According to Wachter, Mittelstadt, and Floridi (2017), the creation of independent AI oversight bodies solidifies these regulations through regular monitoring by way of bias reviewing and compliance checking with ethical guidelines (Gasser & Almeida, 2017). These units would also be able to evaluate the impact AI systems have upon society in high-stakes fields such as finance, healthcare, and law enforcement, where ADM has a great deal of influence on the lives of people living in the group (Calo, 2017). What is needed are modern regulatory schemes that match the development and growth of AI capabilities, plus independent oversight bodies that create a solid foundation, where innovation is balanced with ethics to help build trust among the general public and minimize the risk of misuse or other unintended consequences.

Stakeholder Engagement

The stakeholders mentioned above need to be actively involved in the design and deployment process of AI systems, speaking to ethical issues and assurance that these systems answer societal needs. Afflicted communities, ethicists, and technical experts allow a three-dimensional insight into making design decisions-specifically in areas recognized by high-sounding or even self-deluding slogans such as health and law enforcement. In the ethics of AI applications in law enforcement, for example, a modified participatory design model active in stakeholder role-playing, especially by ethicists and community representatives, shapes ethics at the design point. The inclusions of this type raise the bar for transparency and public confidence in design practices and policies (Watson et al., 2009). In biomedical areas, the review helps appraise stakeholder involvement for its inclusiveness, earliness, and dynamism toward better and more sensitive interventions (Slack et al., 2018). Similarly, stakeholder involvement helps in addressing the sustainability issues arising from AI design. It has also been observed in several works that the involvement of a wide range of stakeholders, including environmental interest groups and local communities, helps the AI system to get more closely aligned with environmental goals and reduce resource utilization, leading towards sustainability (Kunkel et al., 2023). These examples further demonstrate very clearly that stakeholder involvement forms the bedrock on which any AI system must be based if, from a technical point of view, the system is to be sound, but also from the ethical perspective robust, consistent with greater values at the level of society. To quantify and visualize the multifaceted influence of different stakeholders in ADM systems, Figure 3 presents a comprehensive radar chart analysis of stakeholder impact across key dimensions.

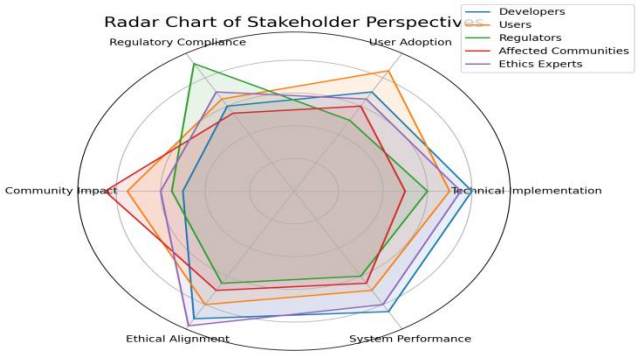

Figure 3. Stakeholder Impact Analysis Across Key Dimensions of ADM Systems.

Bias Mitigation Techniques

Bias mitigation in AI systems is multi-dimensional: because of the need to identify, and correct biases continuously throughout the data life cycle and model development.

Figure 4. Detection and Mitigation Framework for Bias in Automated Decision-Making Systems.

One useful approach to this problem is the diversification of sources, which ensures that the training datasets represent a wide range of demographic and cultural backgrounds, reducing the overrepresentation of certain groups at the expense of others. Data collection should, for that matter, be representative to avoid biases for some unfair outcomes- favouring a particular demographic within hiring algorithms or loan approval mechanisms. Fair Data Generation uses various causality models to investigate and then adjust the bias in datasets by re-weighting sensitive feature contributions to best construct balanced datasets, retaining critical context for accurate prediction (González-Sendino et al., 2024).

Another approach, bias auditing, employs a systematic evaluation methodology to assess the model for fairness through multiple bias metrics on measurable disparities across protected attributes. As an illustration, the AI Fairness 360 toolkit comes with strategies for the different stages of model development in algorithmic fairness evaluation and offers pre-processing, in-processing, and post-processing bias correction methods (Malhotra & Thulal, 2024). Bias auditing also includes post-hoc evaluations, which include debiasing procedures where algorithms are “re-weighted” or modified through adversarial methods to correct extraneous variables, as well as methods which adaptively change over time and correct existing biases in the algorithm (Gupta et al., 2024).

The use of these techniques enhances transparency in the practice of AI, avoids the encoding of harmful stereotypes into systems, and encourages acceptance from the public. The focus on the need to integrate a technical solution, such as a causal model into an extensive audit, presents a growing toolbox to deal with bias in advance, especially in challenging areas such as healthcare finance and criminal justice systems where algorithmic fairness is paramount for achieving desired objectives.

Case Studies

Undoubtedly, the use of biased recruitment algorithms raises a number of ethical issues, the most prominent of which is the example of Amazon's experimental recruitment tool. This AI system used by the company was trained on ten years’ worth of resumes it received and showed a clear bias against women as the datasets used in their training are historically biased. In particular, this system lowered the score of the applicant if the applicant’s resume contained the word ‘women’ as in ‘women’s chess club’ because it led to unfair discrimination against women applicants and their achievements. The company in the end cancelled this hiring tool after finding these discrimination practices weaknesses, showing the risks of using unbalanced data histories for recruitment. This case highlights the importance of bias audits and the limitation of the datasets for recruiting AI reminding the developers that AIs may reinforce the discriminatory practices in existence if not curbed with policies.

Addressing the issue of ‘big data’ more specifically within the category of policing, research has shown that the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) system places discriminatory factors in its assessments towards particularly black male offenders. It was designed to forecast the probability of an individual reoffending in a specific time frame for example two years yet found that individual black offenders were assessed with a greater level of risk compared to young white males with the same background and crime history. The presence of such bias in the elements of an algorithm that is used within a justice system is a cause for concern especially when issues of fairness arise with the possibility of such biases being institutionalized. The need to address bias in policing algorithms such as COMPAS has led to stakeholders advocating for enhanced transparency, routine bias monitoring, and data diversification practices.

AI systems used in medical diagnosis have shown tendencies of bias that can negatively affect the patient’s health care, especially when the training data used is not diverse enough. For instance, certain melanoma detection systems tend to be less precise for people with dark skin because their skin tone has been underrepresented in the training images. Such disparities may result in some populations

failing to get a correct diagnosis or receiving a late diagnosis and subsequent treatment for that matter. Such risks can be associated with the use of biased training data in health care provision. There are both ethical and pragmatic considerations in support of diverse data in training medical AI systems and such systems need to be thoroughly tested in different settings so as to achieve appropriate care without compromising confidence in AId medicine. All of this demonstrates why such AI systems need to be built with a strong sense of fairness, transparency and inclusivity, particularly because such systems are often used in relation to basic rights and access to resources.

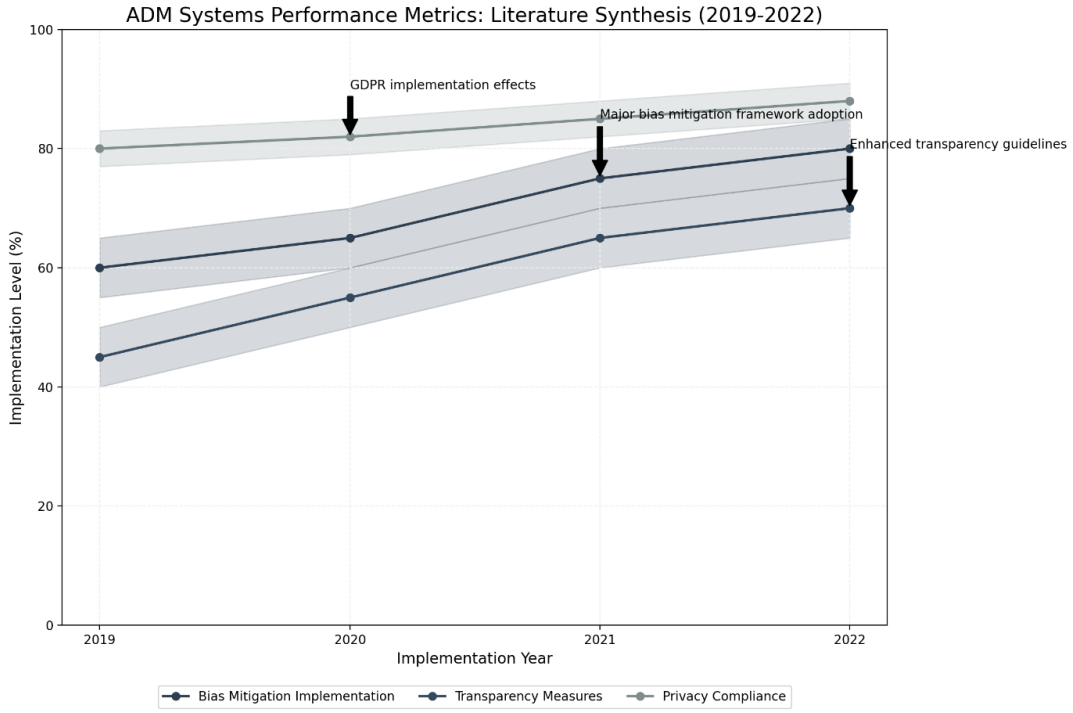

Figure 6: Longitudinal Analysis of ADM System Metrics Across Industry Sectors (2019-2022)

The longitudinal analysis of ADM system metrics presented herein complements the findings in individual case studies through responses from broader industry levels facing ethical challenges. The incremental yet sustained increases in bias mitigation and transparency metrics suggest that organizations learn from documented failures and implement more pragmatic ethical frameworks. However, the variation in improvement rates that emerges in these various metrics also points out that some challenges-especially in areas related to algorithmic transparency demand further concerted effort. Meanwhile, there is evidence that good laws and regulations support real and sustained improvements in the area of privacy protection. The four cases of sustained high levels of privacy compliance, as explained above, have demonstrated that strong regulatory frameworks can be an effective way to ensure systematic improvements. Several key recommendations for both practitioners and policymakers that follow from these patterns across multiple implementations and sectors are discussed in the following section.

Recommendations for Practitioners and Policymakers

Ethics in AI Development: The development process for AI will involve embedding ethics at every stage of the system to ensure that systems are non-discriminatory, accountable, and transparent. Quite literally, ethics involve periodic audits of bias in all phases of a model's life cycle, mostly, if not all of the time, in collection and training before deployment. First, model processes, data sources, and decision logic should be documented to provide insight for stakeholders into the operations and impact the model has. Infusing models with fairness-aware algorithms and developing explainable AI enhances accountability further in that users and regulators will be able to inspect how decisions are derived. Developers should continue testing model performance regularly to identify changes in model behaviour or performance that may impact the fairness or accuracy of said models and support recalibration efforts as the data evolves. This consideration again shall not be limited to the initial deployment itself but also extend to post-hoc reviews and real-time monitoring for unintentional outcomes.

Policy Recommendations: In the context of regulation surrounding AI, policymakers should set the obligation to conduct an ethical impact assessment before the actual deployment of AI systems

especially in Automated Decision-Making (ADM) systems in areas like finance, health and criminal justice. Ethical impact assessments would analyse potential biases, transparency, and risks among other areas before the use of AI, to ensure its use does not contradict societal values. These assessments should include periodic audits by the government agencies in charge, and measures for enforcement and compliance made clear for the companies. Standardization of levels of algorithmic transparency should also be done by the concerned authorities in order to ensure that the developers of an algorithm submit the documentation in a format that can be understood by technical and non-technical audiences. Moreover, user's privacy should be safeguarded and the information processed by the ADM systems should be limited to appropriate and justifiable use as stipulated in the law, with severe consequences for any abuse. In addition, other alternate strategies to these involve setting up an independent advisory board or council whose focus will be on matters of artificial intelligence compliance and ethics that will help in maintaining regulation changes. The purpose of these recommendations is to encourage the development of AI technology in a way that maintains fairness, accountability and public confidence.

Conclusion

The increasing AI capabilities in Automated Decision-Making (ADM) systems have raised ethical issues that need to be addressed in advance solutions. The major issue that arises is Bias and fairness, where training data contains historical biases which cause further and greater inequalities. As in the case of hiring based on one's qualifications using an algorithm or even a crime prediction algorithm which highlights risk groups of individuals, the bias of algorithms has been evident in real-life cases. Therefore, the importance of bias identification and prevention mechanisms cannot be overemphasized. Transparency and explainability are other important challenges because the "black box" nature of complex AI systems makes it a great challenge to provide appropriate accountability. This is where the development and use of XAI techniques are seen to be imperative in establishing a level of mutual trust and guaranteeing a viable degree of scrutiny. Stakeholders require explanations of how better decisions are made to ensure the responsible adoption of the systems and the public's ongoing trust in these systems.

The reliance on vast reaches of personal data by such ADM systems has signalled privacy and data protection concerns to be an area of crucial consideration. Conscientious implementation and operation of robust data governance frameworks have become critical in protecting individual rights while allowing scope for innovation. Potential solutions to this dilemma will be difficult to identify, and seriously so, considering the need to balance technological advancement with privacy protection, requiring quite a depth of thought into the matter of technology as regards data collection, storage, and usage. An increasing level of autonomy for machines within decision-making processes raises important questions of human oversight and responsibility. Clear accountability frameworks arise where machines make consequential decisions affecting people's lives. The challenge is retaining human power where technology must be able to provide such authority.

The rapid advancement of AI technology requires periodic assessment and modification of ethical governance mechanisms. The need for ethical guidelines does not only arise from their initial formulation by organizations; it is necessitated by the changes in technology which require regular review and embossment due to emerging issues. This development calls for the attention of engineers and politicians working hand in hand if at all AI systems will be properly used and developed responsibly. The allocation of responsibilities in ensuring the appropriate application of AI technologies is important. Engineers should design and develop systems with ethics put in place and policymakers should be able to create laws that govern those ethical principles. Effective stakeholder involvement is vital wherever such systems are to be implored to avoid futility and loss of public confidence. Cohesion among the diverse and often competing priorities must be achieved to the possible optimum degree. It is important to encourage the pursuit of innovative practices, but not at the expense of equity and accountability. While gains generated through efficiency might be desirable, they should not be prioritized above ethics. Trust from the public and society at large should only be gained after proper measures on implementation that show, to a reasonable extent, the ethical development of AI has been put in place. The way ahead calls for prompt measures to be taken in many areas at once. It is necessary for organizations to develop and put into use advanced bias testing and mitigation measures, define responsibilities within the organization, improve transparency of the AI procedures, and enhance data protection. These immediate steps should be complemented by long-term initiatives focused on developing comprehensive ethical guidelines, creating sustainable stakeholder engagement mechanisms, enhancing cross-pollination of ideas within and across sectors, and funding bias-free AI systems research. This journey is not without specific responsibilities that different stakeholders must accept. Developers ought to embed ethics-by-design principles in their activities, and organizations

should observe responsible AI deployment strategies. There should be statutory guidelines put in place by the regulators, while the society should concern itself with the ethics of AI and these issues at all times. The outcome of future ADM systems will be highly determined by our ability or inability as a society to solve all these ethical dilemmas and still make use of AI for the good of society. This will also be possible through the willingness of all the parties to embrace and prioritize the ethical aspect rather than the technological aspect – which renders aspects that are ethical in a secondary position – to avoid this servitude, the technological aspect should be a secondary factor. Otherwise, AIDDM systems will become a tool that is used for their intended purposes and for doing good but tempering the values of justice, transparency and respect for human beings.

References

Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. *The economics of artificial intelligence: An agenda* (pp. 197–236). University of Chicago Press.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica, 23*, 77–91.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115.

Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities.* MIT Press.

Brantingham, P. J., Valasik, M., & Mohler, G. O. (2018). Does predictive policing lead to biased arrests? Results from a randomized controlled trial. *Statistics and Public Policy, 5*(1), 1–6.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* W.W. Norton & Company.

Calo, R. (2017). Artificial intelligence policy: A primer and roadmap. *UCDL Review, 51*, 399.

Dastin, J. (2022). Amazon scraps secret AI recruiting tool that showed bias against women. In *Ethics of data and analytics* (pp. 296–299). Auerbach Publications.

Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way* (Vol. 2156). Cham: Springer.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint* arXiv:1702.08608.

Elliott, R. D., & Musgrove, D. (2019). White resentment in electoral politics from 1964 to 1972: Baltimore, Maryland & Lowndes County, Alabama. *Maryland & Lowndes County, Alabama* (July 10, 2019).

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine, 25*(1), 24–29.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.

Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. *Machine Learning and the City: Applications in Architecture and Urban Design*, 535–545.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? The effects of machine learning on credit markets. *The Journal of Finance, 77*(1), 5–47.

Gasser, U., & Almeida, V. A. (2017). A layered model for AI governance. *IEEE Internet Computing, 21*(6), 58–62.

González-Sendino, R., Serrano, E., & Bajo, J. (2024). Mitigating bias in artificial intelligence: Fair data generation via causal models for transparent and explainable decision-making. *Future Generation Computer Systems, 155*, 384–401.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR), 51*(5), 1–42.

Gupta, S., Kundu, R., Deo, A. K. A., Patnaik, M., Kundu, T., & Dehury, M. K. (2024, June). Enhancing transparency and mitigating bias in large language models' responses with sophistication. In *2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)* (pp. 1–6). IEEE.

Jeong, W., Hadzibeganovic, T., & Yu, U. (2021). Evolution of cooperation in multi-agent systems with time-varying tags, multiple strategies, and heterogeneous invasion dynamics. *arXiv preprint* arXiv:2104.01411.
Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399.
Kaminski, M. E. (2021). The right to explanation, explained. In *Research Handbook on Information Law and Governance* (pp. 278–299). Edward Elgar Publishing.
M. Khaleel, A. Jebrel, and D. M. Shwehdy, "Artificial intelligence in computer science," *Int. J. Electr. Eng. and Sustain.*, pp. 01–21, 2024.
M. Khaleel, A. A. Ahmed, and A. Alsharif, "Artificial Intelligence in Engineering," *Brilliance*, vol. 3, no. 1, pp. 32–42, 2023.
Kunkel, S., Schmelzle, F., Niehoff, S., & Beier, G. (2023). More sustainable artificial intelligence systems through stakeholder involvement? *GAIA - Ecological Perspectives for Science and Society, 32*(1), 64–70.
Malhotra, A., & Thulal, A. N. (2024, March). A comparative analysis of bias mitigation methods in machine learning. In *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)* (pp. 1–8). IEEE.
Mantelero, A. (2018). AI and big data: A blueprint for a human rights, social and ethical impact assessment. *Computer Law & Security Review, 34*(4), 754–772.
Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*(11), 501–507.
Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. In *Algorithms of oppression*. New York University Press.
O'Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown.
Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020, January). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 469–481).
Samek, W. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint* arXiv:1708.08296.
Slack, C., Wilkinson, A., Salzwedel, J., & Ndebele, P. (2018). Strengthening stakeholder engagement through ethics review in biomedical HIV prevention trials: Opportunities and complexities. *Journal of the International AIDS Society, 21*, e25172.
Sloot, B. V. D. (2017). Privacy as virtue: Moving beyond the individual in the age of big data.
Susskind, R., & Susskind, D. (2022). *The future of the professions: How technology will transform the work of human experts.* Oxford University Press.
Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56.
Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law, 7*(2), 76–99.
Watson, P. G., Duquenoy, P., Brennan, M., Jones, M., & Walkerdine, J. (2009, November). Towards an ethical interaction design: The issue of including stakeholders in law-enforcement software development. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7* (pp. 313–316).
Zarsky, T. Z. (2016). Incompatible: The GDPR in the age of big data. *Seton Hall Law Review, 47*, 995.