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)

Challenges of Artificial Intelligence today and future implications for society and the world

Donatien Sakubu *

ICT Independent Researcher.

World Journal of Advanced Research and Reviews, 2025, 26(01), 3045-3054

Publication history: Received on 11 March 2025; revised on 20 April 2025; accepted on 22 April 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.26.1.1380

Abstract

This paper examines the growing challenges artificial intelligence (AI) poses to global society now and in the foreseeable future. Through a systematic review of peer-reviewed literature (2019–2024), analysis of policy discourse, and emerging findings from preprints and recent reports (early 2025), the study identifies critical issues such as job displacement and labor market shifts, algorithmic bias and systemic inequities (including specific biases based on race and gender), privacy erosion and enhanced surveillance capabilities (including threats from deepfakes and cybersecurity vulnerabilities), autonomous weapons and escalating global security dilemmas, existential risks, the significant environmental costs of AI, data scarcity and lack of diverse representation, lack of transparency (opaque decision-making), and challenges posed by centralized AI infrastructure. These challenges are analyzed for their profound societal implications, including exacerbated economic inequality, complex ethical dilemmas, and tangible threats to global security and stability. A holistic framework is proposed to address these multifaceted risks, emphasizing the critical need for multi-stakeholder collaboration, the development and implementation of robust ethical governance frameworks, the widespread adoption of explainable AI (XAI) techniques, and the deliberate alignment of AI development and deployment with the United Nations Sustainable Development Goals (SDGs) to ensure equitable and sustainable global development. The paper concludes with actionable recommendations spanning policy interventions, targeted research directions, and industry best practices aimed at mitigating risks effectively and responsibly harnessing AI's transformative potential for the collective global good.

Keywords: Artificial Intelligence; Societal Impact; Ethical Challenges; Autonomous Systems; AI Governance; Sustainable AI

1. Introduction

The rapid advancement of artificial intelligence (AI) has initiated a profound transformation across diverse sectors, including but not limited to healthcare, finance, education, and public administration. This technological evolution is fundamentally reshaping daily human interactions and presenting unprecedented opportunities for enhancing efficiency, driving innovation, and addressing complex global challenges. However, this progress is not without its complexities, concurrently introducing systemic risks that threaten societal stability, exacerbate existing inequities, and pose new challenges to global security. The burgeoning economic impact of the AI sector, projected to reach an estimated $4.8 trillion in market value by 2033 [1], underscores the scale of this transformation. Alongside the economic shifts, critical ethical dilemmas, such as pervasive algorithmic discrimination [2–4], significant environmental costs stemming from energy-intensive training and infrastructure [5–7], and escalating geopolitical risks, particularly those associated with the militarization of AI and the development of autonomous weapons systems [8–10], urgently demand rigorous interdisciplinary analysis and proactive, globally coordinated governance responses.

* Corresponding author: Donatien Sakubu

This paper addresses two central, interconnected questions critical to navigating the AI era responsibly: What key societal challenges specifically examining economic disruption and job displacement, various forms of bias, the erosion of privacy, the proliferation of autonomous weapons, environmental impact, and potential existential risks that AI technology poses today, and how can these complex issues be effectively mitigated through informed policy, ethical innovation, and robust international cooperation? By synthesizing findings from a systematic review of peer-reviewed studies, comprehensive policy reports from leading international bodies, and industry whitepapers published between 2019 and 2024, alongside emergent data and perspectives from recent preprint repositories and conferences (early 2025), this study provides a comprehensive and up-to-date analysis of AI's multifaceted societal impacts. It proposes a framework for action anchored in the urgent need for ethical governance frameworks that are both adaptable and inclusive, the integration of sustainability principles aligned with the United Nations Sustainable Development Goals (SDGs) as a core tenet of AI development, and the promotion of inclusive socio-technical design processes that prioritize human well-being and rights. These proposed strategies collectively emphasize the foundational importance of transparency, accountability, and fairness in the design, deployment, and governance of AI, aiming to ensure that AI aligns with humanity's collective interests, upholds fundamental societal values, and respects universal human rights through inclusive policymaking and the establishment of international standards [11–14].

2. Methods

This study employed a systematic literature review methodology to synthesize existing knowledge and identify emergent trends concerning the societal challenges and implications of artificial intelligence. The review drew upon a broad range of sources, including peer-reviewed academic literature accessed through prominent databases such as Google Scholar, IEEE Xplore, and Scopus. In addition to academic publications, the review incorporated key grey literature, including comprehensive policy briefs and reports from influential international organizations such as the OECD, IEEE, and the United Nations, as well as reports from respected non-governmental organizations like Human Rights Watch.

The literature search focused on publications released between January 2019 and March 2025 to capture the most recent advancements, discussions, and projected future implications in the rapidly evolving field of AI. This timeframe allowed for the inclusion of late-breaking findings from preprints and conference proceedings, which are critical for identifying emergent trends ahead of formal publication. Search strings were constructed using a combination of relevant keywords (e.g., "AI ethics," "algorithmic bias," "AI governance," "societal impact of AI," "AI policy," "autonomous weapons," "environmental impact of AI," "existential risk AI," "explainable AI") connected with Boolean operators (AND, OR) to ensure comprehensive coverage of the research domain. This initial search yielded a substantial corpus of approximately 1,200 sources.

Following the initial search, a rigorous screening process was undertaken. Titles and abstracts were systematically reviewed against predefined inclusion criteria, which prioritized sources with an empirical focus, clear policy relevance, or significant technical novelty related to AI's societal impacts. Duplicates and texts not available in English were excluded at this stage. This screening process resulted in a refined set of approximately 200 articles that proceeded to full-text review.

A final selection of 120 sources was chosen for in-depth thematic analysis. Inductive coding was applied iteratively to the selected texts to identify recurring themes and patterns related to AI's challenges and societal implications. This coding process was guided by established qualitative data analysis techniques, specifically drawing upon principles of open coding as described in foundational texts such as [15,16]. To ensure the trustworthiness and rigor of the coding process, two independent researchers conducted the inductive coding, achieving a strong level of intercoder reliability ($\kappa = 0.85$), indicating a high degree of consistency in the application of codes and interpretation of themes.

The thematic analysis converged on five key areas of significant societal challenge: economic disruption and labor market shifts, algorithmic bias and systemic inequities, privacy erosion and surveillance capabilities, autonomous weapons and global security dilemmas, and the intertwined issues of environmental costs and existential risks. AI governance and explainability (XAI) emerged as crucial cross-cutting concerns relevant to mitigating risks across all identified challenge areas [17].

During the review, a significant regional disparity in the geographic origin of studies was observed, with a notable 82% of the analyzed sources originating from institutions in North America and Europe [18]. Recognizing the potential for bias in a review dominated by perspectives from these regions, the methodology was consciously adapted to incorporate reports and frameworks from a wider global scope. This included integrating recent UNCTAD reports

focusing on the global economic implications of AI, frameworks developed by the African Union addressing AI development and governance in Africa, technical assessments from the Asia-Pacific region, and relevant initiatives and research originating from the Global South. This deliberate effort aimed to enhance the global representativeness and analytical depth of the study's findings and recommendations, ensuring a more balanced understanding of AI's worldwide impacts.

3. Results and Discussion

3.1. Economic Disruption and Labor Market Shifts

Generative AI's rapid adoption is fundamentally reshaping labor markets, a process characterized by both significant disruption and the potential for transformation, albeit unevenly distributed across sectors and geographies. Projections regarding the scale of job displacement and creation vary across studies and timelines, reflecting the inherent uncertainty in forecasting the impact of a rapidly evolving technology. Some earlier estimates suggested AI could displace 75 million jobs globally by 2025 while simultaneously creating 133 million new ones [19,20]. More recent analyses, however, offer different perspectives; for instance, some estimate that the potential for automation could eventually impact up to 300 million full-time jobs [21,22]. Research from sources like the IMF indicates that nearly 40% of global employment is exposed to AI's impact, with a higher propensity for disruption observed in advanced economies compared to low-income countries [23,24]. Within advanced economies, a substantial proportion of jobs, estimated around 60%, may see at least 10% of their tasks affected by the capabilities of large language models [24,25], suggesting a significant reshaping of existing roles rather than outright elimination in many cases.

The impact is particularly acute in low-income nations, which often confront systemic vulnerabilities due to weaker social safety nets, limited resources for workforce adaptation, and a potential lack of infrastructure to fully harness AI ‘driven job creation opportunities [1]. Conversely, AI is also expected to create new jobs, particularly in fields related to AI development, maintenance, and ethical oversight, as well as roles that leverage AI tools to augment human capabilities. The World Economic Forum's 2023 report highlights the emergence of new job roles and the changing skill demands driven by technological adoption [26].

Tangible impacts are already being observed in various economies. Reports from the US, for example, indicate instances of firms replacing certain roles with AI tools, including generative models like ChatGPT [21]. Data entry, administrative support, and certain clerical roles are frequently cited as facing significant projected losses due to automation potential [21,26]. However, it is crucial to consider that the perception of job displacement by AI and robots can sometimes be inflated compared to the actual reported experience [27], suggesting a need for careful analysis of both qualitative and quantitative data.

These significant labor market shifts risk exacerbating existing economic inequalities if not proactively and thoughtfully addressed through targeted interventions. To ensure that AI's integration into the workforce translates into broadly shared prosperity and enhances overall employment quality, policymakers must prioritize the development and implementation of sector-specific reskilling and upskilling initiatives. These programs should be designed to equip workers with the digital literacy, AI fluency, and adaptable skills necessary to thrive in an AI-augmented economy. Examples include targeted training for administrative professionals to leverage AI tools or programs focused on developing skills in emerging AI related fields. Furthermore, exploring and strengthening adaptive social safety nets and vocational training programs is essential to support workers in transition. Drawing inspiration from successful models, such as Brazil's initiatives leveraging AI for workforce development, can provide valuable insights. Crucially, international collaboration is paramount to channeling resources, sharing best practices, and building capacity toward equitable AI education and workforce adaptation strategies globally, ensuring that the benefits of AI driven economic transformation are accessible to all nations and populations [1].

3.2. Algorithmic Bias and Systemic Inequities

A significant and pressing challenge in the deployment of AI systems is their inherent risk of perpetuating and amplifying societal inequities. This often occurs when AI models are trained on historical data that reflects existing societal biases or when they are developed without adequate consideration for diverse perspectives and potential disparate impacts [4]. Empirical audits and research consistently reveal instances of systemic inequities embedded within AI-driven decision-making processes. For example, a compelling 2024 study from the University of Washington demonstrated how résumé-ranking algorithms exhibited significant bias, favoring names perceived as white by a considerable margin (24%), while Black male applicants faced compounded discrimination [2,3,28]. These

discriminatory outcomes often stem from the use of skewed or unrepresentative training data and a lack of diversity within the AI development teams themselves, a problem further evidenced by the minimal contribution of datasets from the Global South in many widely used AI training corpora [29].

Beyond hiring, similar biases have been identified in healthcare algorithms, potentially leading to disparities in diagnosis, treatment recommendations, and patient care, as highlighted by research on algorithms used to manage the health of populations that systematically disadvantaged Black patients [30–32]. Such biases can also manifest with detrimental effects in other critical sectors, including finance (e.g., loan applications) and insurance, if not rigorously identified and addressed throughout the AI lifecycle [33]. The issue of data scarcity and the lack of adequate representation of diverse populations, particularly from the Global South, further exacerbate these problems, resulting in AI models that may perform poorly or unfairly when applied to underrepresented groups.

Mitigating algorithmic bias requires a multifaceted approach. This includes implementing rigorous fairness audits and bias detection mechanisms not only before initial deployment but also continuously throughout the operational life of AI systems. Crucially, fostering diverse and inclusive development teams is essential to bring varied perspectives to the design and evaluation process [4]. Investing in dedicated research aimed at developing advanced bias detection techniques and fairness-aware machine learning algorithms is also critical. Furthermore, supporting grassroots initiatives, such as those exemplified by the African AI Research Network's efforts to build locally relevant and representative datasets, is vital for addressing data imbalances. Ensuring strict compliance with data protection regulations, such as the GDPR, and adhering to provisions within emerging AI governance frameworks like the EU AI Act regarding the lawful processing of sensitive data for the explicit purpose of detecting and mitigating bias, are also paramount legal and ethical requirements [33]. Navigating bias and fairness in digital AI systems necessitates a proactive and ongoing commitment to fairness-aware design and deployment practices [32].

3.3. Privacy Erosion and Surveillance Capabilities

The increasing proliferation and sophistication of AI technologies significantly enhance capabilities for widespread data collection, analysis, and cross-referencing, thereby intensifying concerns regarding individual privacy and the potential for pervasive surveillance. AI systems are instrumental in powering advanced surveillance infrastructure, raising profound ethical questions about the balance between state security interests and the potential for excessive monitoring and control over citizens. Simultaneously, corporations extensively leverage AI to analyze vast quantities of personal data, frequently harvested through online services and connected devices, further entrenching the 'surveillance capitalism' model [35,36]. This model, driven by the economic imperative to collect and monetize personal information, poses substantial risks to individual autonomy and informational self-determination.

Moreover, the advent of highly realistic and easily generated deepfakes presents novel and evolving threats to individual reputation, public trust, and the integrity of democratic processes. The capabilities for creating synthetic media are improving rapidly, constantly challenging the ability of detection methods to keep pace [35,36]. Recent reports underscore the alarming surge in the volume and sophistication of deepfake content, highlighting its potential for malicious applications such as sophisticated phishing attacks, social engineering campaigns, and the spread of disinformation [35,36].

While AI undoubtedly contributes to these privacy challenges, it also offers potential avenues for enhancing data protection. Techniques such as differential privacy, which adds noise to data to obscure individual data points while allowing for aggregate analysis, and federated learning, which enables model training on decentralized data without the need to centralize sensitive information, represent promising AI driven approaches to enhance privacy [37]. Despite these technological solutions, the significant potential for misuse of AI necessitates the implementation and rigorous enforcement of robust data protection regulations. Frameworks such as the General Data Protection Regulation (GDPR) in Europe provide strong legal safeguards for personal data. Furthermore, emerging regulations like the EU AI Act specifically classify certain AI applications with surveillance capabilities as high-risk, subjecting them to stringent requirements and limitations [33,38].

Implementing comprehensive transparency frameworks and promoting the adoption of explainable AI (XAI) methods are also crucial steps to provide individuals with meaningful insight into how their data is being used and how automated decisions affecting them are being made [39,40]. Beyond privacy, the increasing integration of AI into systems and infrastructure introduces new and complex cybersecurity challenges. These include novel attack vectors such as data poisoning, where malicious data is introduced to corrupt AI models, and the exploitation of vulnerabilities within AI models and their underlying libraries [41]. Addressing privacy erosion and surveillance requires a multi-

pronged approach combining strong legal frameworks, ethical development practices, the strategic deployment of privacy-enhancing technologies, and continuous vigilance against evolving cybersecurity threats.

3.4. Autonomous Weapons and Global Security Dilemmas

The integration of AI into military systems represents one of the most urgent and ethically complex challenges today, leading to the accelerated development and potential deployment of lethal autonomous weapons systems (LAWS). These systems, defined by their capability to select and engage targets without direct human intervention, fundamentally alter the nature of warfare. Their proliferation lowers the threshold for engaging in armed conflict and introduces significant risks of accidental escalation, miscalculation due to unforeseen AI behaviors, and a profound erosion of meaningful human control over decisions concerning the use of force, particularly those involving the taking of human life [8,10]. The documented use of AI guided systems in recent and ongoing conflicts tragically underscores these dangers and brings the theoretical concerns into stark reality [10].

Growing international concern over LAWS is increasingly evident in multilateral forums. This concern is perhaps most clearly reflected in recent resolutions adopted by the United Nations General Assembly. Notably, a resolution in December 2024 calling for urgent action and the potential negotiation of a new international treaty to regulate or prohibit LAWS received overwhelming support, with 166 nations voting in favor [8,10]. Despite this strong signal for action, multilateral discussions, particularly within the framework of the Convention on Certain Conventional Weapons (CCW), continue to struggle with complex legal and technical questions. Key among these is the interpretation and application of fundamental principles of International Humanitarian Law (IHL) such as distinction (distinguishing combatants from civilians), proportionality (ensuring civilian harm is not excessive compared to military advantage), and precaution (taking feasible precautions to avoid civilian harm) to the context of LAWS [42]

Central elements under intense discussion in these international dialogues include defining and ensuring the maintenance of context appropriate human control over autonomous weapons, especially for critical life and death decisions; establishing clear restrictions on the operational parameters and environments in which LAWS can be deployed; and ensuring robust mechanisms for accountability in cases of violations of international law [10,42]. While a legally binding global regulatory consensus specifically on LAWS remains elusive, the landscape of governance is further complicated by the development of national policies, such as the US Department of Defense Directive 3000.09 on autonomy in weapons systems, and the approach taken by regional frameworks like the EU AI Act, which, while generally excluding military applications of AI from its scope, emphasizes the obligation of Member States to adhere to existing International Humanitarian Law [9]. Establishing clear international norms, developing legally binding regulations where necessary, and implementing effective oversight mechanisms through dedicated multilateral forums like the UN are absolutely essential steps to prevent an uncontrolled AI arms race, mitigate the tangible threats LAWS pose to global stability and security, and uphold humanitarian principles in the age of artificial intelligence.

3.5. Environmental Costs and Existential Risks

Beyond the immediate societal and security challenges, the significant environmental footprint of AI and the potential for long-term existential risks represent increasingly recognized, albeit distinct, areas of concern. The training and operation of large-scale AI models, particularly deep learning architectures, require substantial computational resources that translate into considerable energy consumption and associated carbon emissions. For instance, the training of a single large model like GPT-4 has been estimated to generate a significant carbon footprint, potentially equivalent to hundreds of tons of CO_2 equivalent [43]. Furthermore, the data centers that power AI systems are voracious consumers of electricity and require vast amounts of water for cooling, contributing significantly to global electricity demand, which is projected to grow rapidly with increasing AI adoption [5–7]. This substantial energy and resource requirement for AI development and deployment presents a direct conflict with global efforts towards climate action, particularly challenging the achievement of Sustainable Development Goal (SDG) 13 (Climate Action).

While AI's current environmental trajectory is concerning, the technology also holds considerable potential to contribute positively to climate change mitigation. AI can be leveraged to optimize energy grids, improve the efficiency of industrial processes, develop more sustainable materials, and enhance climate modeling and prediction capabilities [5]. Addressing the environmental costs of AI thus requires a focused and deliberate effort towards developing energy-efficient AI architectures, often referred to as "Green AI". This involves research into more efficient algorithms, hardware, and data management techniques. Additionally, promoting transparency regarding the environmental impact of AI models and infrastructure is crucial, potentially through standardized reporting mechanisms. The development of regulatory standards, such as those being pursued by organizations like ISO [6], and the emergence of

initiatives promoting sustainable AI practices, such as the Coalition for Environmentally Sustainable AI [44], are promising steps in this direction.

Alongside these immediate environmental considerations, the long-term potential for artificial general intelligence (AGI) or superintelligence gives rise to complex and debated existential risks. These concerns primarily revolve around the ''alignment problem,'' which questions our ability to ensure that highly intelligent AI systems, once developed, will remain aligned with complex and often nuanced human values and intentions [45]. There is a tangible risk of unintended consequences or the potential loss of human control over systems far surpassing human cognitive abilities [11,46]. Expert opinions on the probability and potential timeline of these existential risks vary widely [47], reflecting the speculative nature of forecasting such advanced AI development. Nevertheless, the potentially catastrophic severity of these risks necessitates dedicated and serious research into AI safety, robust alignment techniques, and reliable control mechanisms. Recent studies suggest that public and academic discourse around existential risks does not necessarily detract from addressing the more immediate and tangible harms of AI, indicating that it is possible and necessary to consider both near-term and long-term challenges simultaneously [48]. Navigating the future of AI responsibly requires acknowledging and actively working to mitigate both its environmental footprint and the complex, albeit debated, existential risks it might pose.

4. Policy Recommendations and Future Directions

Effectively addressing the complex and multifaceted societal challenges posed by artificial intelligence necessitates the urgent development and implementation of adaptive, inclusive, and globally coordinated governance frameworks. A risk-tiered regulatory approach, similar to that adopted by the European Union in its AI Act, offers a promising model. This involves implementing stringent bans on AI applications deemed to pose unacceptable risks, such as pervasive social scoring or real-time, indiscriminate biometric surveillance in public spaces. For high-risk AI systems, including those used in critical areas like employment screening, loan applications, and criminal justice, the EU AI Act's requirements for conformity assessments, risk management systems, and data governance should be considered as potential global standards. Furthermore, expanding these regulations to mandate independent third-party audits for critical AI applications could significantly enhance accountability and public trust [38,49].

Addressing the economic disruption and ensuring equitable distribution of AI's benefits requires proactive policy interventions focused on resource allocation. Funding accessible and comprehensive reskilling and upskilling programs is paramount, with a deliberate focus on reaching vulnerable populations and communities in the Global South who are at higher risk of displacement and have fewer resources for adaptation [1]. Alongside workforce development, initiatives aimed at diversifying AI development teams and leadership structures are crucial to mitigate bias and ensure that AI systems are designed with a broader range of societal needs and perspectives in mind; this could potentially include exploring mechanisms like diversity quotas for publicly funded AI projects.

Countering the increasing centralization of AI power among a few dominant technology companies is also a critical policy objective. Governments and international bodies should explore strategies to decentralize AI infrastructure, perhaps through supporting the development of government-backed or publicly accessible open-source AI platforms and models. Promoting data-sharing initiatives in non-competitive or public-interest domains, with appropriate privacy safeguards, can also help democratize access to the data necessary for AI development and counter the data monopolies held by large corporations.

Enhancing transparency and trust in AI systems is fundamental. This can be enforced through mandating the implementation of explainable AI (XAI) frameworks, particularly for high-stakes applications where algorithmic decisions have significant impacts on individuals' lives. Integrating well-established XAI methods such as LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) should be standard practice, guided by principles outlined by organizations like the OECD [39,40]. These measures allow for greater scrutiny of AI decision-making processes, which is essential for identifying and mitigating bias and ensuring accountability.

Future research in AI and its societal implications should actively adopt participatory and community-based models. Engaging directly with marginalized communities and populations disproportionately affected by AI's negative consequences, such as those facing algorithmic bias or job displacement, through co-design workshops and collaborative research projects is essential. This ensures that policy and technical solutions are grounded in real-world experiences and address the specific needs and concerns of those most vulnerable to AI's harms. Such participatory approaches can inform adaptive strategies to address the socio-economic impacts of automation.

Aligning AI development with sustainability goals is also a critical future direction. Policies should incentivize the development of energy-efficient AI architectures ("Green AI") through research funding, tax breaks, or regulatory requirements. Promoting transparency in reporting AI's environmental footprint is also necessary to enable informed decision-making and drive innovation in sustainable AI practices. Supporting initiatives like the Coalition for Environmentally Sustainable AI can accelerate the adoption of greener AI technologies and infrastructure [44].

Finally, multilateral cooperation is indispensable for addressing AI challenges that transcend national borders, particularly in the realm of global security. Advancing towards a potential legally binding treaty on lethal autonomous weapons systems (LAWS) under the auspices of the United Nations remains a priority. Such a treaty should aim to establish clear prohibitions on unacceptable LAWS, require meaningful human control over the use of force, and establish robust international oversight and accountability mechanisms [8–10]. Navigating the evolving and varied global AI governance landscape, which includes different approaches in jurisdictions like the EU, UK, US, and China, requires sustained international dialogue and cooperation to establish shared norms and best practices [38]. These interconnected policy, research, and international cooperation efforts are crucial to ensure that AI's trajectory is steered towards fostering equitable, sustainable, and human-centric advancement globally.

5. Conclusion

The foregoing analysis underscores that the rapid advancement and pervasive integration of artificial intelligence into global society present a complex web of interconnected challenges. These challenges span critical domains, including significant economic disruption and the reshaping of labor markets, the perpetuation and amplification of algorithmic bias leading to systemic inequities, the erosion of individual privacy alongside the expansion of surveillance capabilities, the profound security dilemmas posed by autonomous weapons systems, the substantial environmental footprint of AI technologies, and the debated but potentially transformative long-term existential risks. Effectively navigating these challenges demands an urgent, coordinated, and multilateral approach to governance and a commitment to proactive, inclusive policymaking that centers human well-being and global sustainability.

As this paper has argued, mitigating AI's potential harms while harnessing its transformative potential requires deliberate and strategic interventions. Key among these are measures to counter the centralization of AI development and power, fostering a more distributed and equitable technological landscape. Implementing and rigorously enforcing risk-tiered regulations, as exemplified by approaches that ban unacceptable high-risk applications and impose strict requirements on others, is essential for establishing clear boundaries and ensuring accountability. Promoting transparency through the widespread adoption of explainable AI (XAI) techniques is crucial for building trust and enabling scrutiny of algorithmic decision-making, particularly in critical sectors. Furthermore, prioritizing sustainability through the development and incentivization of Green AI standards and practices is necessary to align AI's growth with urgent climate action goals.

Crucially, addressing these challenges equitably necessitates dedicated resource allocation for global reskilling and upskilling initiatives, with a particular focus on empowering vulnerable populations and bridging the digital divide in the Global South. Adopting participatory policymaking processes that genuinely involve marginalized communities is vital to ensure that governance frameworks are informed by diverse experiences and effectively address disparate impacts. While binding international frameworks may be necessary for issues with global ramifications, such as the regulation of autonomous weapons, all governance efforts must ensure that AI development and deployment align fundamentally with universal human rights and contribute positively to the achievement of the United Nations Sustainable Development Goals (SDGs), notably promoting decent work and economic growth (SDG 8), reducing inequalities (SDG 10), enabling climate action (SDG 13), and fostering peace, justice, and strong institutions (SDG 16).

In conclusion, as AI continues its rapid evolution and its influence on every facet of life deepens, its trajectory must be deliberately anchored in robust ethical guardrails, unwavering transparency, clear accountability mechanisms, and a steadfast commitment to collective well-being. Only through concerted global effort, informed by interdisciplinary research and inclusive dialogue, can we ensure that AI serves truly as a catalyst for equitable, sustainable, and human-centric progress worldwide.

Compliance with ethical standards

Disclosure of Conflict of interest

The author declares that there are no conflicts of interest regarding the publication of this paper. This research did not receive any specific funding.

References

[1] UNCTAD. AI's $4.8 trillion future: UN Trade and Development alerts on divides, urges action | UN Trade and Development (UNCTAD) [Internet]. 2025 [cited 2025 Apr 19]. Available from: https://unctad.org/news/ais-48-trillion-future-un-trade-and-development-alerts-divides-urges-action

[2] Çırtlık B. Algorithmic Gender Bias, AI Design, Social Inequality, Artificial Intelligence Gender Bias in AI feministasylum. Feminist Asylum: A Journal of Critical Interventions feministasylum.pitt.edu [Internet]. 2024 [cited 2025 Apr 19];2. Available from: https://feministasylum.pitt.edu/faci/article/view/124/176

[3] Pulivarthy P, Whig P, Pulivarthy P, Whig P. Bias and Fairness Addressing Discrimination in AI Systems. https://services.igi-global.com/resolvedoi/resolve.aspx?doi=104018/979-8-3693-4147-6.ch005 [Internet]. 2024 Jan 1 [cited 2025 Apr 19];103–26. Available from: https://www.researchgate.net/publication/385029381_Bias_and_Fairness_Addressing_Discrimination_in_AI_Systems

[4] UN Women. How AI reinforces gender bias—and what we can do about it | UN Women – Headquarters [Internet]. 2025 [cited 2025 Apr 19]. Available from: https://www.unwomen.org/en/news-stories/interview/2025/02/how-ai-reinforces-gender-bias-and-what-we-can-do-about-it

[5] CG. Accelerating Climate Action with AI [Internet]. 2023 [cited 2025 Apr 19]. Available from: https://web-assets.bcg.com/72/cf/b609ac3d4ac6829bae6fa88b8329/bcg-accelerating-climate-action-with-ai-nov-2023-rev.pdf

[6] Roma D. Environmental Impact of Generative AI | Stats & Facts for 2025 [Internet]. 2024 [cited 2025 Apr 19]. Available from: https://thesustainableagency.com/blog/environmental-impact-of-generative-ai/

[7] Zewe A. Explained: Generative AI's environmental impact | MIT News | Massachusetts Institute of Technology [Internet]. 2025 [cited 2025 Apr 19]. Available from: https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117

[8] Human Rights Watch. Killer Robots: UN Vote Should Spur Treaty Negotiations | Human Rights Watch [Internet]. 2024 [cited 2025 Apr 19]. Available from: https://www.hrw.org/news/2024/12/05/killer-robots-un-vote-should-spur-treaty-negotiations

[9] akub D. Legal aspects of the development of weapon systems with artificial intelligence in 2025 | ARROWS, law firm. Legal services quickly and globally [Internet]. 2025 [cited 2025 Apr 19]. Available from: https://arws.cz/news-at-arrows/legal-aspects-of-the-development-of-weapon-systems-with-artificial-intelligence-in-2025

[10] errin B. Volume: 29 Issue: 1 Lethal Autonomous Weapons Systems & International Law: Growing Momentum Towards a New International Treaty [Internet]. 2025 [cited 2025 Apr 19]. Available from: https://www.asil.org/insights/volume/29/issue/1

[11] Hikkasani DC. Navigating Artificial General Intelligence (AGI): Societal Implications, Ethical Considerations, and Governance Strategies [Internet]. 2024. Available from: https://www.preprints.org/manuscript/202407.1573/v3

[12] Nadella GS, Meduri SS, Maturi MH, Whig P, Nadella GS, Meduri SS, et al. Societal Impact and Governance: Shaping the Future of AI Ethics. https://services.igi-global.com/resolvedoi/resolve.aspx?doi=104018/979-8-3693-4147-6.ch012 [Internet]. 2025 Jan 1 [cited 2025 Apr 19];261–82. Available from: https://www.igi-global.com/gateway/chapter/359647

[13] Parson EA, Fyshe A, Lizotte D. Artificial Intelligence's Societal Impacts, Governance, and Ethics: Introduction [Internet]. 2019. Available from: https://ssrn.com/abstract=3476399

[14] Patnaik R. Artificial Intelligence: Transforming Industries, Enhancing Daily Life and Tackling Ethical Challenges. Journal of Educational Research and Policies [Internet]. 2024 Dec 30 [cited 2025 Apr 19];6(12):55–6. Available from: https://bryanhousepub.com/index.php/jerp/article/view/1163

[15] Lungu M. The Coding Manual for Qualitative Researchers. American Journal of Qualitative Research [Internet]. 2021 May 26 [cited 2025 Apr 19];6(1):232–7. Available from: https://www.ajqr.org/download/the-coding-manual-for-qualitative-researchers-12085.pdf

[16] Saldaña Johnny. Johnny Saldaña - The Coding Manual for Qualitative Researchers. 2021;

[17] Batool A, Didar C, Csiro Z, Csiro MB, Author F, Author T. AI Governance: A Systematic Literature Review. 2024 Jul 24 [cited 2025 Apr 19]; Available from: https://www.researchsquare.com

[18] Abanga EA, Acquah T. A Bibliometric Analysis of Global Research Trends in Artificial Intelligence from 2019 to 2023. Asian Journal of Research in Computer Science [Internet]. 2024 Dec 26;17(12):220–33. Available from: https://journalajrcos.com/index.php/AJRCOS/article/view/540

[19] McKinsey &Cpmany. Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages [Internet]. 2017 [cited 2025 Apr 19]. Available from: https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages

[20] WEF. Machines Will Do More Tasks Than Humans by 2025 but Robot Revolution Will Still Create 58 Million Net New Jobs in Next Five Years > Press releases | World Economic Forum [Internet]. 2018 [cited 2025 Apr 19]. Available from: https://www.weforum.org/press/2018/09/machines-will-do-more-tasks-than-humans-by-2025-but-robot-revolution-will-still-create-58-million-net-new-jobs-in-next-five-years/

[21] Josh H. 2025. 2025 [cited 2025 Apr 19]. 60+ Stats On AI Replacing Jobs (2025). Available from: https://explodingtopics.com/blog/ai-replacing-jobs#conclusion

[22] Nexford University. Nexford University. 2024 [cited 2025 Apr 19]. How Will Artificial Intelligence Affect Jobs 2024-2030 | Nexford University. Available from: https://www.nexford.edu/insights/how-will-ai-affect-jobs

[23] IMF. Gen-AI : Artificial Intelligence and the Future of Work [Internet]. International Monetary Fund; 2024 [cited 2025 Apr 19]. Available from: https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2024/01/14/Gen-AI-Artificial-Intelligence-and-the-Future-of-Work-542379

[24] IMF. AI Will Transform the Global Economy. Let's Make Sure It Benefits Humanity. [Internet]. 2024 [cited 2025 Apr 19]. Available from: https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity

[25] Eloundou T, Manning S, Mishkin P, Rock D. GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. 2023 Mar 17 [cited 2025 Apr 19]; Available from: https://arxiv.org/abs/2303.10130v5

[26] WEF. Shaping the Future of Learning : The role of AI in Education 4.0 : Insight Report [Internet]. 2023 [cited 2025 Apr 19]. Available from: https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf

[27] Dahlin E. Are Robots Stealing Our Jobs? Socius [Internet]. 2022 [cited 2025 Apr 19];5. Available from: https://explodingtopics.com/blog/ai-replacing-jobs#conclusion

[28] University of Washington. AI tools show biases in ranking job applicants' names according to perceived race and gender | UW News [Internet]. 2024 [cited 2025 Apr 19]. Available from: https://www.washington.edu/news/2024/10/31/ai-bias-resume-screening-race-gender/

[29] Petrušić R. BIASES IN THE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE. Nauka i tehnologija [Internet]. 2024 Dec 30 [cited 2025 Apr 19];12(2):43–9. Available from: https://naukaitehnologija.iu-travnik.com/article/333

[30] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations [Internet]. 2019 [cited 2025 Apr 19]. Available from: https://www.ehidc.org/sites/default/files/resources/files/Dissecting%20racial%20bias%20in%20an%20algorithm%20used%20to%20manage%20the%20health%20of%20populations.pdf

[31] Singh A. The Manifestation and Implications of Bias in Artificial Intelligence on Global Society. IJFMR - International Journal For Multidisciplinary Research [Internet]. 2024 Oct 8 [cited 2025 Apr 19];6(5). Available from: https://www.ijfmr.com/research-paper.php?id=28477

[32] Tariq MU. Navigating Bias and Fairness in Digital AI Systems. In 2024 [cited 2025 Apr 19]. p. 127–56. Available from: https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/979-8-3693-4147-6.ch006

[33] EPRS. Algorithmic discrimination under the AI Act and the GDPR. EPRS | European Parliamentary Research Service [Internet]. 2025 [cited 2025 Apr 19]. Available from: https://www.europarl.europa.eu/RegData/etudes/ATAG/2025/769509/EPRS_ATA(2025)769509_EN.pdf

[34] DataGuard. https://www.dataguard.com/blog/growing-data-privacy-concerns-ai/. 2024 [cited 2025 Apr 19]. The growing data privacy concerns with AI: What you need to know. Available from: https://www.dataguard.com/blog/growing-data-privacy-concerns-ai/

[35] Sam R. Deepfake Trends to Look Out for in 2025 | Pindrop [Internet]. 2025 [cited 2025 Apr 19]. Available from: https://www.pindrop.com/article/deepfake-trends/

[36] Tuvoc. Deepfake AI Trends 2025: Threats, Opportunities, & Future [Internet]. 2025 [cited 2025 Apr 19]. Available from: https://www.tuvoc.com/blog/deepfake-ai-in-2025-navigating-threats-and-unveiling-opportunities/

[37] Ward S. Anonymization: The imperfect science of using data while preserving privacy. Vol. 10, Science Advances. American Association for the Advancement of Science; 2025.

[38] Nick S. AI Regulations around the World - 2025 [Internet]. 2025 [cited 2025 Apr 19]. Available from: https://www.mindfoundry.ai/blog/ai-regulations-around-the-world

[39] Gaudenz B. Explainable AI (XAI): The Complete Guide (2025) - viso.ai [Internet]. 2024 [cited 2025 Apr 19]. Available from: https://viso.ai/deep-learning/explainable-ai/

[40] DigitalDefynd. Evolution of Explainable AI (XAI) [Include Case Studies] [2025] - DigitalDefynd [Internet]. 2025 [cited 2025 Apr 19]. Available from: https://digitaldefynd.com/IQ/evolution-of-explainable-ai-xai/

[41] HiddenLayer. HiddenLayer.com. 2025 [cited 2025 Apr 19]. AI Security: 2025 Predictions & Recommendations. Available from: https://hiddenlayer.com/innovation-hub/ai-security-2025-predictions-recommendations/

[42] Goussac N, Pacholska M. The Interpretation and Application of International Humanitarian Law in Relation to Lethal Autonomous Weapon Systems [Internet]. 2025 [cited 2025 Apr 19]. Available from: https://unidir.org/publication/the-interpretation-and-application-of-international-humanitarian-law-in-relation-to-lethal-autonomous-weapon-systems/

[43] Patterson D, Gonzalez J, Le Q, Liang C, Munguia LM, Rothchild D, et al. Carbon Emissions and Large Neural Network Training. 2021 Apr 21 [cited 2025 Apr 19]; Available from: https://arxiv.org/abs/2104.10350v3

[44] UNEP. New Coalition aims to put Artificial Intelligence on a more sustainable path [Internet]. 2025 [cited 2025 Apr 19]. Available from: https://www.unep.org/news-and-stories/press-release/new-coalition-aims-put-artificial-intelligence-more-sustainable-path

[45] Lorvo A. Aligning AI with human values | MIT News | Massachusetts Institute of Technology [Internet]. 2025 [cited 2025 Apr 19]. Available from: https://news.mit.edu/2025/audrey-lorvo-aligning-ai-human-values-0204

[46] Wikipedia. Existential risk from artificial intelligence - Wikipedia [Internet]. 2024 [cited 2025 Apr 19]. Available from: https://en.wikipedia.org/wiki/Existential_risk_from_artificial_intelligence

[47] Severin F. Why do Experts Disagree on Existential Risk and P(doom)? A Survey of AI Experts. 2025 Jan 24 [cited 2025 Apr 19]; Available from: http://arxiv.org/abs/2502.14870

[48] Hoes E, Gilardi F. Existential risk narratives about AI do not distract from its immediate harms. Proceedings of the National Academy of Sciences [Internet]. 2025 Apr 22 [cited 2025 Apr 19];122(16):e2419055122. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.2419055122

[49] Atanasovska D. Top 5 AI governance trends for 2025: Compliance, Ethics, and Innovation after the Paris AI Action Summit - GDPR Local [Internet]. 2025 [cited 2025 Apr 19]. Available from: https://gdprlocal.com/top-5-ai-governance-trends-for-2025-compliance-ethics-and-innovation-after-the-paris-ai-action-summit/