Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 19 No. 14 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijim.v19i14.56981

SPECIAL FOCUS PAPER

Ethical and Governance Frameworks for Artificial Intelligence: A Systematic Literature Review

Osama Ismail, Naim Ahmad (✉)

College of Computer Science, King Khalid University, Abha, Saudi Arabia

nagqadir@kku.edu.sa

ABSTRACT

Rapid proliferation of artificial intelligence (AI) in key domains such as healthcare, education, public services, and digital economies has heightened global commitment to developing strong ethical and governance standards. This systematic review, based on the PRISMA approach, synthesizes evidence from 22 peer-reviewed journal articles, white papers, and policy documents between 2020 and 2025. It examines governance frameworks, ethical concepts, regulatory approaches, sectoral application, and common challenges in operationalizing AI ethics internationally. The review sees wide confluence on core ethical values of fairness, transparency, accountability, explainability, and sustainability. There remains, however, wide variability in the application and institutionalization of these values across jurisdictions. The European Union's AI Act provides a binding three-tiered risk system with centralized monitoring, whereas global institutions such as the World Health Organization (WHO) and the United Nations advocate high-level ethical frameworks but without statutory implementation. Regional efforts by ASEAN, Hong Kong SAR, and China reflect different states of maturity in the application of lifecycle audit and sector-level policy tools. Five recurring challenges are identified throughout literature: algorithmic bias, privacy risks of the data, patchwork of regulatory environments, corporate and state incumbents, and risks of labor market dislocation. Institutional models by the Alan Turing Institute and corporate players such as NTT DATA stress the value of building ethics from the outset of the life of AI. Educational and civil society voices add further calls for inclusive and collaborative governance frameworks. This paper concludes by advocating for the creation of an internationally coordinated AI regulation agency, cross-border certification schemes, ethics-by-design design methodologies, and adaptive governance mechanisms for the safe, equitable, and transparent application of AI technology.

KEYWORDS

artificial intelligence (AI) governance, AI ethics, algorithmic accountability, risk-based regulation, data privacy, global AI policy, ethics-by-design, AI in healthcare, participatory governance, digital policy, generative AI, cross-border AI regulation, AI frameworks, algorithmic bias, AI transparency, human-centered AI

Ismail, O., Ahmad, N. (2025). Ethical and Governance Frameworks for Artificial Intelligence: A Systematic Literature Review. *International Journal of Interactive Mobile Technologies (iJIM)*, 19(14), pp. 121–136. https://doi.org/10.3991/ijim.v19i14.56981

This article is a revised version of a paper presented at the Annual Conference on Innovating for Impact, held on February 19–20, 2025, at the Manipal Dubai Campus, Dubai, UAE. Article submitted 2025-05-02. Revision uploaded 2025-05-26. Final acceptance 2025-06-03.

1 INTRODUCTION

1.1 Background and context

Artificial intelligence (AI) technologies are revolutionizing world industries, including healthcare, education, finance, and public administration. In the health sector, AI is being applied to assist clinical decision-making, optimize the utilization of resources, and build diagnostic capacity, as highlighted in the World Health Organization's (WHO) ethical AI in health systems guidelines [1]. In finance, AI software assists in its detection against fraud and automation of transactions with attendant risks for bias and undetectable decision-making architectures that need to be regulated through ethics [2]. Learning institutions are adopting generative AI into pedagogy and administration, and there is an immediate need for governance frameworks that maintain academic integrity and ethical alignment [3]. Across jurisdictions, the use of generative models of AI, such as large language models (LLMs) and transformer models, created major ethical issues. They produce biased or discriminatory text based on imbalanced training data and uninterpretable decision-making [4]. Privacy violations and risks to identity on the part of users arising from AI uses are on the rise, particularly with surveillance technologies and unregulated data gathering processes [5]. In response, there has been a call for standardized AI governance frameworks. The Alan Turing Institute introduced the context, accountability, responsibility, explainability (CARE) and assessment, commitment, transparency (ACT) models, emphasizing the integration of ethics in the full AI lifecycle for public sector applications [6]. Industry-driven governance models, such as the AI lifecycle and organizational governance strategy by NTT DATA, underline the necessity for aligning AI development with business ethics and operational standards [7].

1.2 Problem statement and review objective

Despite the emergence of several AI ethics and governance frameworks, practical implementation remains inconsistent. The European Union's AI Act proposes a risk-tier classification and a centralized European AI Office to coordinate oversight, yet enforcement challenges persist across member states [8]. The WHO's ethical guidance outlines six core principles for AI in health—autonomy, transparency, accountability, fairness, non-maleficence, and sustainability—but lacks binding legal power to ensure compliance across countries [1]. Similarly, the UN's model policy on AI governance provides strategic recommendations but relies on voluntary adoption, while regional frameworks such as *ASEAN's Guide on AI Governance and Ethics* emphasize ethical use without strict enforcement mechanisms [9]; [10]. Fragmentation in global AI policy results in duplicated efforts, inconsistent safeguards, and governance gaps, particularly in cross-border applications of AI [11]. Given this context, the aim of this literature review is to systematically analyze and synthesize 22 contemporaries articles and policy reports on ethical and governance frameworks for AI. This review identifies convergences and divergences across governance systems, highlights research gaps, and proposes policy-oriented recommendations to advance a globally harmonized and ethically sound approach to AI governance.

1.3 Research questions

To guide this systematic literature review, the following research questions are proposed: 1. What governance and ethical frameworks currently exist for AI systems?

This question explores governmental, institutional, and private-sector initiatives aimed at regulating AI technologies through ethical principles, lifecycle governance, and compliance mechanisms. 2. How do different regions and sectors define and implement ethical AI? The review investigates variations in implementation across sectors such as healthcare, education, and tourism, and compares international approaches in the EU, ASEAN, China, Australia, and BRICS. 3. What are the recurring challenges and limitations in current AI governance systems? This includes technical constraints (e.g., algorithmic bias), institutional limitations (e.g., enforcement gaps), and socio-economic impacts (e.g., job displacement). 4. What future directions and recommendations are proposed to strengthen global AI governance? The review seeks to uncover forward-looking strategies such as cross-border alignment, ethical-by-design approaches, and adaptive regulatory frameworks.

2 METHODOLOGY (PRISMA FRAMEWORK)

2.1 Search strategy

This literature review employed a targeted, manual search strategy using a curated dataset of 21 academic and policy documents provided by the user. These documents encompass peer-reviewed journal articles, institutional white papers, and regional governance guidelines on artificial intelligence. The aim was to ensure a comprehensive and in-depth analysis of global and sectoral AI governance practices. The search strategy prioritized sources published between 2020 and 2025, reflecting recent advancements and evolving governance practices. The relevant documents were located through the occurrence of major keywords in the title, abstract, or thematic focus, and these were AI governance, ethical AI, framework, accountability, and digital policy. The keywords were used to ensure convergence to the review aims with regard to ethical issues, regulatory measures, and implementation loopholes in AI systems. It allowed diversification of sources into health sources [1], education sources [3], and digital economy policy sources [12], which allowed for comprehensive and multi-sector investigation into the availability of frameworks.

2.2 Inclusion and exclusion criteria

This review was based on sources that were investigating ethical practices in AI, models for governance, regulatory environment, or alignment of cross-border policy for thematic coherence reasons. Peer-reviewed articles, systematic reviews, and policy reports, as well as white papers from organizations such as WHO, the EU, the UN, and national regulatory agencies, were used as sources that were qualified for this review. If the source addressed technical matters of AI (architecture planning, deep learning optimization) without issues regarding governance, policy, or ethics, then they were left out. This is because the synthesis did not include articles that analyze AI performance but make no mention of accountability or policy ramifications with respect to the area.

2.3 PRISMA workflow

Adhering to the PRISMA framework (see Figure 1), the review was carried out through four organized phases: identification, screening, eligibility, and inclusion.

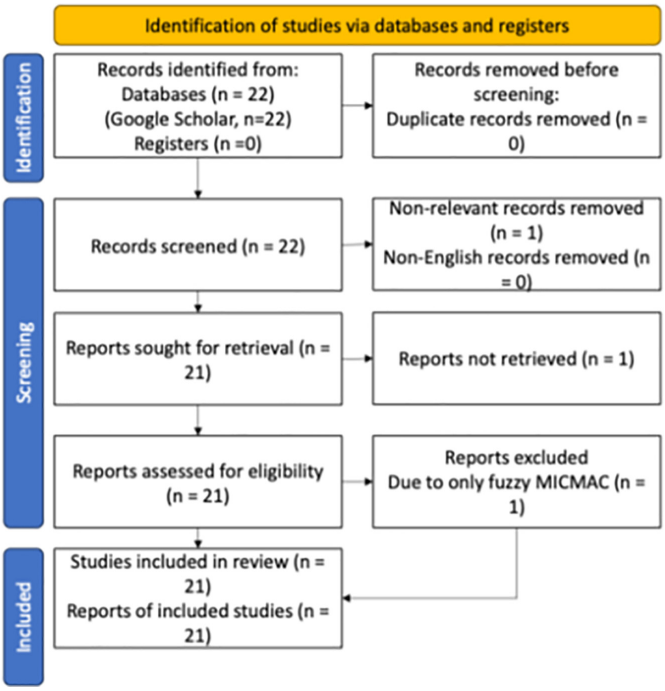

Fig. 1. PRISMA workflow

Identification stage: The initial. The dataset consisted of 21 documents given directly by the user. They were recently peer-reviewed articles, white papers, and strategic frameworks by international institutions and academic researchers.

Screening stage. Every article went through an initial screening on the basis of title and abstract to verify alignment with the study aims. Articles were kept if they discussed ethical principles, policymaking, enforcement issues, or the design of governance. Significantly, articles by [2], [9], and [5] were kept on the basis of theoretical underpinnings and rigorous ethical evaluation.

Eligibility stage. The full-text screening was performed on all 21 documents. The stage made sure that every study had analytical or empirical findings on AI governance and ethics. For instance, [1] was chosen because of its guidance on AI operation in the health sector, and [8] because of its in-depth examination of the EU AI Act enforcement framework.

Inclusion stage. All 21 papers met the inclusion criteria and were integrated into the final synthesis. These papers represent contributions across multiple geographies (EU, ASEAN, BRICS, China, Australia), sectors (health, education, finance), and authorship types (governments, think tanks, private companies, academics). This diverse selection ensured a globally representative and multi-stakeholder analysis.

2.4 Data extraction and categorization

Data from each of the 21 included sources were extracted systematically using a structured matrix. The following metadata were recorded: author(s), publication year, geographic scope, study type (e.g., review, policy paper), AI ethical principles addressed (e.g., transparency, bias, privacy), governance level (national, corporate, intergovernmental), and sector focus (e.g., healthcare, education, public policy, digital economy). Articles were then categorized into five thematic areas: 1) Core Ethical Principles, 2) Framework Types, 3) Key Challenges, 4) Sector-Specific Considerations, and 5) Strategic Solutions, Figure 2.

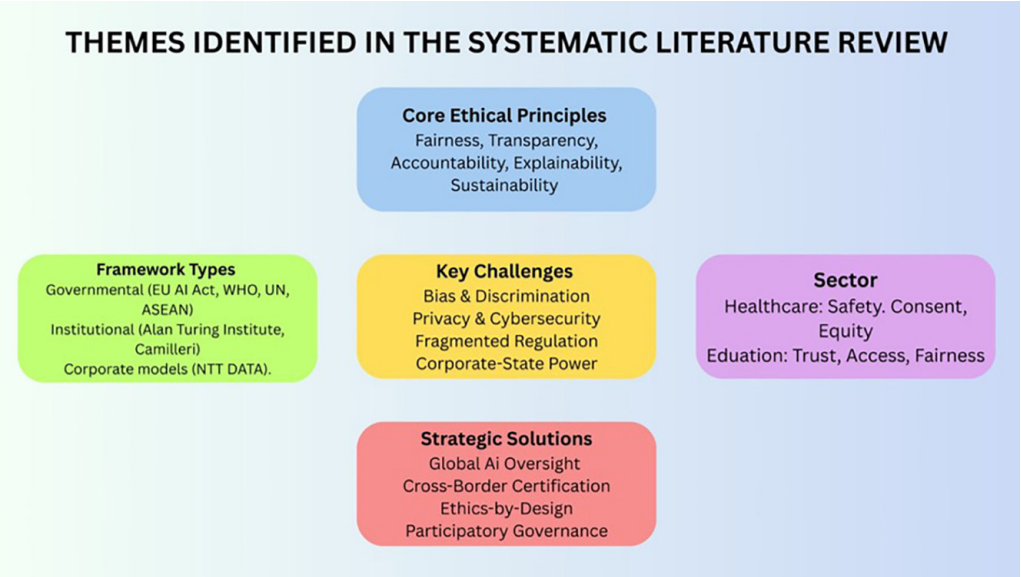

Fig. 2. Themes identified in the systematic literature review on AI governance and ethics

Literature review on AI governance and ethics led to a taxonomy that was critical for drawing comparisons and identifying governance gaps across jurisdictions and sectors.

3 AI MODELS, GOVERNANCE, AND ETHICS FOUNDATIONS

3.1 Types of AI models requiring governance

The rapid evolution of AI has introduced a variety of complex model architectures that demand stringent governance mechanisms. Among the most critical are LLMs and generative models, such as those powering tools such as GPT-4, which have demonstrated the capacity to synthesize human-like language and content but pose risks related to misinformation, data misuse, and bias [4]. Convolutional neural networks (CNNs), commonly used in image recognition, are particularly prevalent in fields like medical diagnostics, surveillance, and autonomous systems [2] Algorithmic decision systems, including AI engines embedded in government welfare systems, hiring software, or banking fraud detection, often operate with limited transparency, making them vulnerable to unfair or unethical outcomes [5]. These models rely on historical data that may carry embedded social biases, which can be amplified when deployed at scale without appropriate governance structures. Real-world deployment examples underscore the breadth of AI impact: facial recognition systems are increasingly used in law enforcement and border control, raising concerns about civil liberties; in healthcare, AI-enabled diagnostic imaging is used to detect tumors or predict disease progression [1]; and in higher education, tools such as ChatGPT are integrated into learning environments for content generation and tutoring assistance [3].

3.2 Need for ethical governance

The use of advanced AI systems without regulation poses important technological and society-related concerns. At the society level, AI systems impinge on the

individual's right to autonomy, particularly when the users do not know or cannot object to automated choices being made about them. AI systems are able to be used to propagate misinformation, widen society's cleavages, and sway political processes through autonomously created algorithmic content delivery on social platforms [13]. The black-box nature or opacity of numerous AI algorithms makes these outputs uninterpretable. This erodes user trust and accountability and makes it harder to regulate [14]. Without explainability, users and developers might not know exactly why an AI model gave a specific outcome. Technologically, AI has the capacity to amplify biases found in existing datasets used to train it. An AI hiring module trained with biased data may invariably prefer candidates from some demographics and perpetuate discrimination under the guise of objectivity [15]. Likewise, public services using AI without equity protection can cause unequal access or discriminatory treatment to marginalized groups [6]. These ethical considerations reinforce the need for regulatory frameworks to provide fairness, auditability, human oversight, and responsible innovation in the creation and implementation of artificial intelligence.

3.3 International calls for AI ethics

Global and regional institutions are already providing broad ethical standards for their use. A model policy for the responsible use of AI within United Nations systems, where the UN reflects the principles of fairness, human rights, safety, and sustainability. To achieve this heterogeneity, risk assessment processes, institutional accountability, and human-centered design need to be incorporated in the AI lifecycle [16]. In recognition of this, the WHO has equally intensified its efforts on the ethical use of AI in healthcare. In its 2021 report, it gives six principles that should guide the use of AI, including autonomy, transparency, accountability, inclusiveness, safety, and sustainability. The aim is for principles to reduce the risk of exposure related to the use of algorithms in clinical environments, protect the privacy of the data, and promote fairness of health outcomes in general and particularly in low- and middle-income nations [1]. At the regional level, ASEAN published a guide to AI governance that supports voluntary uptake of principles such as explainability, human agency, and robustness [10]. The government of Hong Kong SAR created an "Ethical AI Framework" with a focus on lifecycle management, impact assessments, and ethical audits [17]. In turn, China's Digital Policy Office released ethics-focused guidelines with a focus on enterprise-level responsibility and national governance alignment [18]. They are evidence of increasing international agreement on the importance of principled AI development, with differing levels of enforcement.

3.4 Institutional and corporate responsibility

Parallel to governmental efforts, organizational and institutional stakeholders have created in-house governance structures intended to balance AI innovation with ethics and performance agendas. NTT DATA's white paper defines one such three-pillar approach—AI Strategy, Organization, and Operational Lifecycle—with an emphasis on ethics in designing, deploying, and ongoing monitoring. Their model encompasses technical checks and balances, fairness checks, and AI literacy training to foster an ethics culture in enterprises [7]. [2] highlights the "ethics-by-design" principle by calling for the incorporation of fairness, transparency, and accountability

into the very creation of AI. His framework consists of corporate social responsibility (CSR) objectives and calls on tech firms to develop internal checks and balances and practice open algorithmic audits [2]. It highlights the "ethics-by-design" principle by calling for the incorporation of fairness, transparency, and accountability into the very creation of AI. His framework consists of CSR objectives and calls on tech firms to develop internal checks and balances and practice open algorithmic audits [6]. Student voices are being recognized in the education system too. Participatory governance in learning environments has been underscored by Barus et al., who recommended that there should be learner participation in the shaping of AI policies to achieve legitimacy and local relevance [3].

4 COMPARATIVE ANALYSIS OF AI GOVERNANCE FRAMEWORKS

4.1 Governmental and supranational policies

Artificial intelligence regulation has become an increasingly important global policy priority. Supranational and governmental institutions are taking the lead to create enforceable ethical guidelines and legal frameworks to mitigate AI risks. Though there are differences in strategies being adopted in approaches to regulation, there are some threads that run through many, and these include transparency, accountability, and human rights [19].

European Union AI act. The EU released the artificial intelligence Act (AIA) in 2024, the most ambitious effort to regulate AI to date. The AIA uses the risk-tier system to categorize AI systems into four classes: unacceptable risk, high risk, limited risk, and minimal risk. AI systems that control human behavior or are system risks—social scoring and real-time biometric monitoring are mentioned among these—are prohibited under the category of unacceptable risk [8]. To enforce the law, the AIA institutes an EU Artificial Intelligence Board with members from member states and an EU Central Artificial Intelligence Office with responsibilities to oversee implementation, including compliance monitoring and coordinating inspections. The board encourages cooperative regulation by making conformity assessments compulsory for developers of risky systems, keeping records, and having to register systems in an EU database [8].

World Health Organization. Within the health sphere, the WHO released an extensive set of AI ethics principles to use in health-related uses. Six main ethical principles are presented in the framework to govern the creation and implementation of AI into health systems: well-being, autonomy, transparency, accountability, equity, and sustainability [1]. WHO's advice is that AI systems must support and supplement human decision-making and that proper safeguards are installed to provide clinical validity, privacy protection, and equitable access. It emphasizes the need for external technical inspections of AI systems on a regular basis, particularly those with implications for vulnerable people or used in resource-poor environments [11].

United Nations AI policy model. The United Nations has also put forward a model framework for responsible AI use within the UN system. This policy promotes multi-level risk governance, where risk evaluation is conducted at both institutional and operational levels. The framework aligns with broader UN values by stressing sustainability, human rights, transparency, and fairness [16]. The UN policy notably includes mechanism design strategies to ensure that AI systems align incentives among stakeholders and encourages all agencies to develop internal ethical review procedures. Although non-binding, this framework is envisioned as a blueprint

for international organizations and developing countries aiming to establish AI oversight [19].

ASEAN and Hong Kong SAR. The *ASEAN Guide on AI Governance and Ethics* serves as a voluntary framework intended for commercial and non-military AI deployments across Southeast Asia. It promotes principles of robustness, safety, explainability, and proportionality, and recommends lifecycle audits and human-in-the-loop mechanisms to ensure trust and safety in AI-driven decision-making [10]. In Hong Kong, the Ethical AI Framework published by the Digital Policy Office builds upon sector-specific implementation strategies. It emphasizes AI Application Impact Assessments (AI-AIA) and integrates ethical considerations throughout the development lifecycle. The framework follows a three-lines-of-defense accountability model—spanning developers, managers, and auditors—to promote comprehensive risk management [17].

China's digital governance model. China's Digital Policy Office advances an AI policymaking approach that is closely aligned with national goals in surveillance, public safety, and industrial competitiveness. National principles invoke corporate responsibility in ethical design, data protection, and fairness in algorithms, albeit with considerable room given to state-directed uses like facial recognition and internet filtering [18]. The strategy exemplifies a top-down model of governance with national agencies overseeing private sector uptake and ramping up the application of AI to public administration based on an expanded concept of technological sovereignty.

4.2 Institutional frameworks from academia and NGOs

Scholarly and non-state institutions have contributed to AI decision-making by creating independent, evidence-informed ethical principles. They are particularly focused on public interest, inclusiveness, and transparency and are filling the gaps in policies created by governments and industry. The Alan Turing Institute has formulated two foundational public sector AI governance models: the CARE and ACT models. Both models focus on ongoing ethical considerations beginning with project initiation and leading through deployment. CARE encourages decision makers to think about the socio-political context of the use of AI systems, to clarify who bears the responsibility of using them, and to make systems explainable to the public. Instead, ACT proposes such a governance action model based on these values applied to existing projects as an assessment and measures of transparency [6]. In his CSR-oriented ethical framework [2], manages to merge four corporate AI system high-level constructs, fairness, transparency, accountability, and explainability. The research pinpointed how multinationals need to draw up explicit rules and internal regulations in order to forestall risks like algorithmic bias, breach of privacy, discriminatory treatment among others. The reasons why auditability matters, why there should be third-party algorithmic audits, and how regulatory collaboration could help build public trust are given by Camilleri [2]. According to [14], an analytical approach to AI governance can be developed under an integrated framework based on the life cycle continuum. Instead of their framework having just these four stages, their stages include governance requirements during design, development, deployment, and the monitoring of these after deployment. The reason this framework stands out is that it is based on anticipatory governance, in which ethical risks are assessed well in advance during the design stage by performing scenario planning and including contributions from stakeholders. And their study of self-driving cars is a good

example of the value of dynamic adaptation in processes of governance and how this has become increasingly important as an emergent risk plays out with other AI innovations [14].

4.3 Corporate and industry-focused governance models

Companies that bring AI into the private sector have developed intricate organizational structures that map AI development with business objectives and risk management requirements. Based on the model, including strategy organization as well as lifecycle management, NTT DATA provides AI governance in three tiers. The strategy tier aligns business priorities and ethical considerations with AI development so that AI deployments reflect overarching values like fairness and accountability. The organization tier fosters AI literacy and culture transformation as well as specialized governance teams like "AI hubs" or ethics boards. The lifecycle management tier provides monitoring and audit mechanisms throughout the AI project life cycle to ensure compliance and risk reduction and long-term value delivery [7]. For emerging economies, Saba and Ngepah consider the impact of AI governance on employment and the economy in BRICS economies. Employing cross-sectional and time-series model estimations, they establish and confirm that the interaction among AI investment and the quality of governance has significant implications for the economy. Their work advises that governments adopt supportive regulatory environments that strike an equilibrium between innovation and labor market protection and accountability [20]. A theoretical system is proposed by Ashok et al., establishing a multilayered framework of ontology in digital ethics. The structure classifies ethical issues in relation to seven types of digital technology archetypes (for instance, cognitive, governance, and physical technologies) and plots them on dimensions such as autonomy, privacy, fairness, and accountability. It suggests this framework be used by organizations to evaluate ethical risks on several levels—from user interaction to organizational conduct—to make more context-sensitive and scalable systems of governance [15]. These models reflect the heterogeneity of ethical concerns within and among sectors and regions, with some featuring systematic processes and others targeting socio-economic consequences.

4.4 Education and societal involvement

Public involvement and education are paramount to the establishment of inclusive and reliable AI control systems. Barus et al. explored students' attitudes towards generative AI regulation in Indonesian universities through a mixed-method study. They found that students strongly supported introducing the use of AI tools such as ChatGPT in the learning space, with concerns regarding academic integrity problems, uses, and context of institutional policies. Students demanded models of co-governance where the students are involved in deriving guidelines for the use of AI. For the study, transparency, and ethical education, collective responsibility is key in building trust through their use and in developing responsible AI use among learners [3]. The ASEAN Guide to AI governance and ethics adds fuel to the fire by suggesting construction of frameworks appropriate to commercial, non-military use that serve to provide substantial benefit to ASEAN. It focuses on human-in-the-loop systems, inclusive stakeholder consultations, and ethics-by-design cycles of development. From a formal legal perspective, however, ASEAN's approach is not binding;

however, it is regionally compatible and culturally conscious AI that developers and regulators can adopt [10].

5 KEY CHALLENGES IN IMPLEMENTING ETHICAL AI GOVERNANCE

5.1 Algorithmic bias and discrimination

One of the most pressing and multifaceted issues in AI ethical regulation is bias in algorithms. Artificial intelligence systems are trained on past data and may unknowingly replicate or extrapolate existing social biases on the basis of race, gender, and income level. Such biases are usually concealed in impenetrable algorithms and thus more far-reaching and harder to correct [6]. For instance, the AI systems employed in hiring, lending scoring, or disease diagnosis could be based on training data that undercounts some demographics or embodies biases in society. Therefore, these systems can disproportionately harm women, ethnic minorities, or poor individuals. Face recognition software that does not work well on non-white individuals is an extensively documented instance of this effect [4]. The problem is exacerbated by the lack of diversity in datasets, and this is more pronounced in emerging economies with limited digitized data. Digital economy growth in China is found by Dong et al. to be related to regional imbalances in green innovation by way of one-sided digital infrastructure and under-representation in decision-supporting datasets applied to optimize policymaking and industry decision-making. Most digital systems are found by Ashok et al. to exhibit Western and urban-biased viewpoints, entrenching global imbalances regarding AI harms and benefits. Existing governance regimes fail to sufficiently enforce requirements on representativeness when data is being audited or on explainability when algorithms are required to prevent discriminatory outcomes. Algorithmic bias will continue to be an unresolved ethical risk when equity audits and diversity requirements are not enforceable regulatory standards.

5.2 Data privacy and cybersecurity

Artificial intelligence governance also faces major challenges in the realm of data privacy and cybersecurity, especially with the growing use of surveillance technologies and data-driven behavioral analytics. Facial recognition systems, predictive policing tools, and emotion detection software are being deployed in both democratic and authoritarian contexts, often with limited public consent or regulatory oversight [21]. These tools are particularly prevalent in tourism-heavy regions where public monitoring is increasing under the guise of safety and efficiency. Su warns that data commodification and lack of standardization in cybersecurity protocols across AI systems leave users vulnerable to surveillance and manipulation [5]. Many systems collect granular user data—location, preferences, biometric information—without clear boundaries for use, retention, or third-party access. In the absence of ethical guardrails, AI becomes a vector for mass data collection and behavioral profiling. Governments like China have promoted centralized control over AI data management, but this has led to conflicts between surveillance-driven state objectives and privacy rights [18]. Ligot emphasizes that many generative AI models, such as LLMs and diffusion models, absorb large amounts of public and private data, often without consent or oversight. This introduces systemic risks of data breaches, identity theft, and unauthorized surveillance. Therefore, robust cyber resilience

policies, secure AI pipelines, and enforceable data governance frameworks are essential to ensure ethical AI deployment and protect public trust [4].

5.3 Enforcement gaps and fragmentation

Despite a growing number of ethical AI principles and national frameworks, there remains a significant enforcement gap across jurisdictions. The European Union AI Act, though pioneering in scope, lacks clear operational mechanisms to coordinate with non-EU jurisdictions, raising concerns about inconsistent application and legal fragmentation across borders [8]. Even within the EU, enforcement is decentralized across national supervisory authorities, which vary in capacity and political will. This can result in delayed or uneven implementation of high-risk AI safeguards, particularly for complex systems like autonomous weapons or algorithmic public services [9]. In low- and middle-income countries, enforcement is even more difficult due to a lack of technical expertise, weak institutions, and limited financial resources. [13] critiques the global governance vacuum in AI, describing how current discussions often treat AI as a homogenous technology, while in reality it is a distributed computing system embedded in varied social, legal, and cultural contexts. The lack of a universal, enforceable global standard means that AI developers can "forum shop" for the least restrictive regulatory environments—a problem known as "AI regulatory arbitrage." The ASEAN framework, while encouraging voluntary compliance, does not impose binding requirements or cross-national harmonization tools. This leads to fragmented governance landscapes where businesses operating in multiple countries face unclear obligations and inconsistent oversight [10]. Addressing this gap requires the creation of transnational regulatory agreements, cross-border audit systems, and mutual recognition of AI certifications to prevent systemic governance failure and promote global interoperability.

5.4 Governance in the face of corporate and state power

A critical challenge in AI governance lies in the concentration of power within private corporations and state institutions that design, deploy, and profit from AI systems. Tech companies often own data pipelines, control the algorithmic architectures, and define ethical boundaries—all without sufficient external accountability [1]; [7]. WHO warns that corporate-led AI deployment in health settings can lead to profit-driven decision-making that undermines patient safety, equity, and transparency. NTT DATA highlights how internal AI governance strategies—though well-structured, may prioritize business goals over ethical mandates unless monitored by independent bodies. Simultaneously, state-driven AI programs, especially in authoritarian regimes, often leverage AI for surveillance, censorship, and social control. These applications blur the line between governance and coercion, creating ethical concerns around state overreach and lack of democratic oversight [18]. Civil society plays a limited role in the regulatory ecosystem. The Alan Turing Institute advocates for inclusive governance models that involve communities, domain experts, and the general public in shaping AI norms, yet few countries have institutionalized such participatory frameworks [6]. Barus et al. shows that even in academia, students and educators often have minimal input in policy design despite being key stakeholders in educational AI deployments. Effective AI governance should enable multi-stakeholder governing bodies to be more empowered, broaden algorithmic

transparency responsibilities, and promote equal participation by non-state actors to counter unchecked power in institutions [3].

5.5 Economic displacement and labor shifts

Adoption of AI poses huge risks of job displacement, far more so in the Global South. Saba and Ngepah, in examining BRICS economies, reported that investments in AI were associated with productivity growth but depressed unemployment in those sectors that were not adaptable to digitization. They infer that poor institutions don't act to shield labor markets from automation shocks. They advocate for AI models that include employment prospects forecasting, reskilling initiatives, and safeguarding institutions so that AI-fostered innovation does not worsen economic inequality or undermine local employment markets [20].

6 DISCUSSION AND FUTURE DIRECTIONS

6.1 Global harmonization of standards

There is still an urgent need amid this speed in regional AI regulation for globally unified standards to provide ethical uniformity, legal harmonization, and just innovation. The pluralist landscape, in which nations follow differential frameworks, regimes of enforcement, and risk definitions, poses hurdles to multinationals and trans-border systems of AI. Marwala highlights the need to move towards unified global regulation based on the Universal Declaration of Human Rights and calls for collective standards on AI ethics, data regulation, and institutional responsibility [13]. The United Nations' model policy upholds this stance by advocating an internationally applicable framework for UN agencies and governments nationwide to adopt. The model policy advises nations to adopt risk-tiered assessment strategies, sustainability goals, and human rights impact assessments into AI technologies development and implementation practices [16]. To put global governance into practice, Mueller suggests institutions like an international AI watchdog modeled on the International Atomic Energy Agency (IAEA). Such an organization would promote monitoring for international compliance, perform technical and ethical inspections, and regulate convergence around the world. Mueller suggests that if this authority does not exist, powerful private players and national regimes will be able to take advantage of differences in regulation and engage in "AI regulatory arbitrage" and widen disparities in the development and regulation of AI. Globally standardized certification systems—like algorithmic transparency badges or conformity seals on ethics—would promote global cross-border trust and set up accountability standards. Mutually recognized agreements and intergovernmental agreements on digital ethics will be pivotal in allowing AI systems to work responsibly beyond national and cultural borders, especially with rising technologies crossing borders [13].

6.2 Public participation and ethical design

Equitable and responsible AI policymaking needs to include public engagement and human-centered design. AI systems usually lack the lived experiences of

diversified populations because policymaking frameworks typically are crafted by technocratic elites rather than including civil society members, local communities, or end users to any significant extent. Barus et al. makes the argument in support of the implementation of co-design processes, particularly in learning environments, in which students and academics are given the right to influence AI policies that will govern learning environments. Their work shows that participatory models strengthen legitimacy and acceptance of AI policymaking structures. The Alan Turing Institute too highlights the need for inclusive decision-making through its CARE and ACT frameworks, which focus on accountability and explainability and promote active engagement by stakeholders throughout the AI lifecycle [6]. A further obstacle to equal participation is the universal digital divide. WHO observes that populations in middle- and lower-income nations frequently have poor access to digital networks and equipment, which restricts them from accessing, benefitting from, or critically examining AI technologies [1]. Su supplements that these gaps in digital skills and competency, particularly in underrepresented groups, decrease the practicability of bottom-up ethics control. Overcoming these issues demands bringing together digital inclusion policies and making participatory control genuine rather than symbolic [5].

6.3 Sector-specific ethical adjustments

Ethical leadership should also be adapted to the distinctive needs of particular sectors based on differences in risks, stakeholders, and the regulatory environments in play. In health, the consequences are especially dire since AI systems directly impact diagnostic performance, treatment plans, and patient well-being. The WHO presents six fundamental principles to inform ethical AI in health: well-being, autonomy, transparency, accountability, equity, and sustainability. The advice requires informed consent processes, especially if AI is applied to assist clinical decision-making and requires external checks to ascertain diagnostic tools' levels of safety and accuracy. Ethical regulation in education and public services will need to touch on issues of access, equity, and human oversight [1]. Putra Barus et al. identifies student concerns regarding generative AI tools used in higher education based on issues related to academic integrity and participatory democracy. As such, the ACT framework of the Alan Turing Institute is particularly useful to public administration and education in terms of acting as an actionable model for incorporating ethics in the design and delivery of digital services [6]. Both sectors will have to go deeper than abstract principles and move to specific operational instruments such as bias audits through real public feedback and an explicit hierarchy of accountability for each individual case.

6.4 Emerging technologies and ethics-by-design

Going forward, governing structures of developing AI innovations will have to be looking forward while considering risks associated with this emerging innovation, and to do this from the onset, the principle of ethics by design. [4] points out that to integrate risk mitigation into generative AI frameworks, federated learning can be used to improve data privacy, and prompt checks are needed to prevent abuse. Instead, C amendment Eri advocates for mechanisms design practices that establish incentives for stakeholders to align the technological goals and principles of

transparency and fairness from the outset. Initially these are incorporated to reduce ethical debt, increasing the long-term accountability and trust [4].

7 CONCLUSION

The 21 scholarly and policy-based contributions surveyed to AI governance and ethics in this systematic literature review are from various sectors, geographies, and institutional lenses. Broadly, fairness, accountability, transparency, and inclusivity were converged as core ethical principles, however, implementation, enforcement, and coordination were found to differ widely. It is one of the most urgent needs: globally harmonized regulation. Fragmented legal regimes leave regulatory gaps as powerful actors are able to exploit them [16]. A third way of securing compliance and interoperability beyond national boundaries is the proposal of an international AI watchdog that is inspired by the IAEA [13]. The review also underscored the need for human-centric AI ethics. Models like the WHO's principles on health [1], the Alan Turing Institute's CARE and ACT frameworks [6], and participatory education models proposed by Barus et al. [3] all stress that people-focused ethical AI needs to serve humans rather than optimize processes. Ethical design, inclusive policymaking, and public participation are necessary conditions to achieve long-term legitimacy. Ethical regulation calls for constant monitoring and dynamic regulation. Technological innovations—such as generative AI and federated learning—require evolving policies to cope with innovation [2]. Future efforts need to be directed towards enforcement instruments, global accountability frameworks, and sectoral governance systems. Research needs to investigate workable models of AI audit operations, cross-border data governance agreements, and resilience models in environments with poor governance. AI can be implemented in ways that maintain human rights prerequisites, support innovation, and maintain public trust only through multi-stakeholder action, anticipatory regulation, and inclusive oversight.

8 REFERENCES

[1] World Health Organization, "Ethics and governance of artificial intelligence for health: WHO guidance," 2021. [Online]. Available: https://apps.who.int/iris/handle/10665/341996

[2] M. A. Camilleri, "Artificial intelligence governance: Ethical considerations and implications for social responsibility," *Expert Syst.*, vol. 41, no. 7, p. e13406, 2023. https://doi.org/10.1111/exsy.13406

[3] O. P. Barus, A. N. Hidayanto, E. Y. Handri, D. I. Sensuse, and C. Yaiprasert, "Shaping generative AI governance in higher education: Insights from student perception," *International Journal of Educational Research Open*, vol. 8, p. 100452, 2025. https://doi.org/10.1016/j.ijedro.2025.100452

[4] D. V. Ligot, "AI governance: A framework for responsible AI development," *SSRN*, 2024. [Online]. Available: https://doi.org/10.2139/ssrn.4817726

[5] L. J. Su, "An analysis of AI ethics and governance," ABI Research, 2023. [Online]. Available: https://f.hubspotusercontent10.net/hubfs/6705264/Marketing/Whitepapers/An%20Analysis%20Of%20AI%20Ethics%20and%20Governance/CR-FWEI-102.pdf?hsCtaTracking=f6a879c5-d92d-497b-9bfa-4b8ed352a480%7C236dcf94-4a02-4470-9463-61704a807a8d

[6] D. Leslie *et al.*, "AI ethics and governance in practice: An introduction," The Alan Turing Institute, 2024. [Online]. Available: https://www.turing.ac.uk/ai-ethics-governance

[7] NTT DATA, "AI governance: AI strategy, organization and operational lifecycle for business value generation (White Paper)," 2022. [Online]. Available: https://www.nttdata.com
[8] C. Cancela-Outeda, "The EU's AI Act: A framework for collaborative governance," *Internet of Things*, vol. 27, p. 101291, 2024. https://doi.org/10.1016/j.iot.2024.101291
[9] T. Marwala, "Framework for the governance of artificial intelligence," UNU Technology Brief 5, 2024. [Online]. Available: https://unu.edu/publication/framework-governance-artificial-intelligence
[10] ASEAN, "ASEAN guide on AI governance and ethics," 2024. https://asean.org/book/asean-guide-on-ai-governance-and-ethics/
[11] M. Ogunbukola, "AI governance and ethics: Frameworks, challenges, and case studies," *ResearchGate*, 2024. [Online]. Available: https://www.researchgate.net/publication/383477929
[12] X. Dong, K. Guo, and X. Zhao, "Does digital economy promote regional green innovation?" *Journal of Digital Economy*, vol. 3, pp. 47–61, 2024. https://doi.org/10.1016/j.jdec.2024.07.002
[13] M. L. Mueller, "It's just distributed computing: Rethinking AI governance," *Telecomm. Policy*, vol. 49, p. 102917, 2025. https://doi.org/10.1016/j.telpol.2025.102917
[14] L. Xue and Z. Pang, "Ethical governance of artificial intelligence: An integrated analytical framework," *Journal of Digital Economy*, vol. 1, no. 1, pp. 44–52, 2022. https://doi.org/10.1016/j.jdec.2022.08.003
[15] M. Ashok, R. Madan, A. Joha, and U. Sivarajah, "Ethical framework for artificial intelligence and digital technologies," *Int. J. Inf. Manage.*, vol. 62, p. 102433, 2022. https://doi.org/10.1016/j.ijinfomgt.2021.102433
[16] United Nations, "Framework for a model policy on the responsible use of artificial intelligence in UN system organizations," 2024. [Online]. Available: https://unsceb.org/sites/default/files/2024-11/Framework%20for%20a%20Model%20Policy%20on%20the%20%20Responsible%20Use%20of%20AI%20in%20UN%20System.pdf
[17] Government of Hong Kong Special Administrative Region, "Ethical Artificial Intelligence Framework," 2024. [Online]. Available: https://www.pcpd.org.hk
[18] Digital Policy Office of Government of China, "Ethical Artificial Intelligence Framework," 2024. [Online]. Available: https://www.digitalpolicy.gov.hk/en/our_work/data_governance/policies_standards/ethical_ai_framework/doc/Ethical_AI_Framework.pdf
[19] Governance Institute of Australia, "White Paper on AI Governance: Leadership Insights and the Voluntary AI Safety Standard in Practice," 2024. https://www.governanceinstitute.com.au/app/uploads/2024/09/GovInst-AI-Whitepaper.pdf
[20] C. S. Saba and N. Ngepah, "The impact of artificial intelligence on employment and economic growth in BRICS: Does the moderating role of governance matter?" *Research in Globalization*, vol. 8, p. 100213, 2024. https://doi.org/10.1016/j.resglo.2024.100213
[21] G. Gozgor, C. K. Lau, Z. Lin, and Y. Zeng, "The impact of digital governance on tourism development," *Journal of Digital Economy*, vol. 3, pp. 1–13, 2024. https://doi.org/10.1016/j.jdec.2024.05.003

9 AUTHORS

Osama Ismail is a postgraduate student at King Khalid University at the Department of Informatics and Computer Systems, College of Computer Science. His interest includes decision science and ethical issues in information technology adoption (E-mail: 446810671@kku.edu.sa).

Naim Ahmad received the M.S. degree in computer and information sciences from University of South Alabama, Mobile, USA and PhD in Enterprise Systems from Aligarh Muslim University, India. He worked in the USA as a Research Assistant with U.S. DOE funded project and as a Graduate Assistant. He has taught as an Assistant Professor with the Management and Engineering Colleges, India, for seven years. He has been a Faculty Member of King Khalid University, Saudi Arabia, since 2013. He has publications in refereed international journals, conferences, and book chapters. His current research interests include enterprise systems, E-learning, sustainability, data science, and MCDM. He received the Upsilon Pi Epsilon Award for Computing Excellence, in 2003, at USA. He has developed SmartISM software, globally used for interpretive structural modeling (E-mail: nagqadir@kku.edu.sa).