



Review

Ethics Without Teeth? Challenges and Opportunities in AI Declarations for Platform Governance

Ahmad Haidar

Université Paris-Saclay, Univ Evry, IMT-BS, LITEM, 91025 Evry-Courcouronnes, France; ahmad.haidar@imt-bs.eu

Abstract

The rapid integration of artificial intelligence (AI) into digital platforms has raised critical questions about how AI's ethical declarations influence this sector. This study adopts a mixed-methods approach. First, a descriptive content analysis examined 54 declarations, including 45 national declarations across Africa, Asia, Europe, and the Americas, and 9 from major global actors (MGAs) such as the OECD, G7, and the EU. Ethical principle frequency was examined, and a benchmarking index was developed to compare "dominant principles" cited in over 50% of regional declarations with those cited in over 50% of MGA declarations. The analysis reveals universal adoption of societal well-being, fairness, accountability, and privacy (100%), while transparency and security show regional variation (75%). Second, a semi-systematic literature review following PRISMA guidelines identified four opportunities (e.g., global participation) and seven limitations (e.g., lack of standard frameworks, definitional ambiguities, implementation challenges, and legal enforcement difficulties). The implications of these limitations for digital platforms are then examined, leading to the identification of two dimensions for responsible platform governance: assessment mechanisms (e.g., UNESCO's Ethical Impact Assessment) and governance implementation structures. The study further distinguishes three tiers of enforceability: declarative, procedural, and institutionalized ethics, bridging normative declarations and operational practice in platform governance.

Keywords: artificial intelligence ethics; ethical declarations; ethical framework; ethical guidelines; policy recommendations; e-commerce; digital platforms; platform governance; responsible AI; PRISMA guidelines

1. Introduction

Artificial intelligence (AI) is progressively transforming diverse sectors, such as business [1] and healthcare [2], with expenditures anticipated to grow at an annual rate of 37.3% from 2023 to 2030 [3]. Among the sectors at the forefront of this transformation are digital commerce and platform-based ecosystems. When thoughtfully integrated into e-commerce platforms, AI holds real potential to guide consumer decision-making and steer business growth [4]. Today's AI-driven e-commerce platforms have become critical economic infrastructure, using sophisticated algorithms to deliver highly personalized marketing, flexible pricing, and smoother transaction experiences [5]. Yet this rapid rise in AI does not come without risks. Rezaei and his colleagues [6] highlight the adverse nature of AI in organizations, including privacy and data protection, bias and fairness, transparency and explainability, which significantly influence decision-making [7]. Additionally, the unexplainable nature of AI recommendations from deep learning-based models in digital business poses a significant obstacle to strategic decision-making [6].



Academic Editor: Albert Y.S. Lam

Received: 16 February 2026

Revised: 22 March 2026

Accepted: 23 March 2026

Published: 26 March 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\) license](#).

These risks become even more pressing in digital service environments, where behind-the-scenes algorithms quietly control access to goods, services, and information—often with little transparency or meaningful user control. Real-world incidents make these risks difficult to ignore. In August 2023, hackers successfully breached Retool, a software company, compromising 27 cloud-based customers through a carefully orchestrated attack that combined phishing tactics with an AI-generated deepfake voice call designed to trick employees into handing over multi-factor authentication codes [8].

In response to these growing concerns, researchers have put forward a range of governance approaches. The field of AI ethics itself emerged largely as a reaction to the misuse of technology and its damaging effects—particularly on consumer trust [9,10]. A digital governance framework spanning analog, augmented, and automated modes has been proposed, emphasizing control, coordination, incentives, and trust mechanisms for managing AI-enabled systems [11]. Taking a different angle, Corporate Digital Responsibility (CDR) has been introduced as a framework tailored to address the full lifecycle risks that AI presents—from algorithmic bias and data privacy violations to labor displacement and threats to human dignity—risks that traditional Environmental, Social and Governance (ESG) framework categories often fail to adequately capture [12]. Similarly, effective AI management requires frameworks that can rein in AI autonomy, improve learning through high-quality data and human oversight, and tackle the problem of inscrutability through understandable AI tools and broader stakeholder education [13]. At the municipal level, 13 distinct roles that local governments can play in AI governance have been identified—ranging from initiators and regulators to solution architects and impact auditors [14].

Recent research on AI-generated content in e-commerce contexts highlights the growing importance of Ethical Labeling Practices (ELPs) as mechanisms to signal platform responsibility and mitigate perceived risks [15]. These practices align closely with the principles of ethical AI declarations, such as transparency, accountability, and fairness. A large body of literature has explored declarations to construct frameworks for ethical AI [16,17] and to compare existing guidelines [18–20]. According to Jobin et al. [18], there was an 88% increase in the number of documents related to ethical principles or guidelines for AI after 2016.

Another part of the literature has criticized these principles, describing them as “not just fruitless but a dangerous distraction, diverting immense financial and human resources away from potentially more effective activity” [21] (p. 869). The same author argues that AI principles are useless for three main reasons. Firstly, many of these principles are deemed ‘meaningless’ due to their contested or incoherent nature, which complicates their application. Secondly, these principles are often ‘isolated,’ emerging within an industry and education system that generally overlooks ethical considerations. Third, ‘toothless’ lacks enforceable consequences and frequently aligns more with corporate interests than ethical accountability. From this perspective, companies choose generative AI open source under normative uncertainty [22]. Responding to this “toothless” characterization, Ressayguier and Rodrigues [23] have called for restoring enforcement mechanisms to AI ethics, arguing that “AI ethics should not remain toothless!” (p. 1).

While critics focus on what ethical declarations lack, there remains limited systematic analysis of how these declarations vary across global contexts, what conditions enable their operationalization, and how their limitations manifest in commercial digital ecosystems. Existing comparative studies have mapped ethical principles across declarations [18,24] but have not systematically examined regional variations or benchmarked national commitments against global governance frameworks. Similarly, critiques of implementation gaps [20,25] have identified challenges but not mapped these to specific business implications for digital platforms.

Consequently, this study aims to bridge these gaps by (1) assessing the opportunities presented by ethical AI declarations for guiding responsible AI development, (2) identifying limitations that hinder their effectiveness, and (3) examining implications for digital platforms and e-commerce firms. Two high-level research questions guiding this research are: *How do global AI declarations vary in their inclusion of ethical principles? What gaps and opportunities emerge in the literature in translating AI ethics into enforceable mechanisms, particularly in commercial and platform contexts such as e-commerce?*

In the context of digital business and e-commerce, this question is timely and consequential, as firms seek to align their AI-driven operations with evolving societal expectations and regulatory standards. This article proceeds as follows: Section 2 defines key concepts and explores the moral philosophies underpinning ethical AI. Section 3 outlines our mixed-methods research design. Section 4 presents findings from the descriptive content analysis and a semi-systematic literature review. Section 5 discusses implications for businesses, policymakers, and scholars, proposing a governance framework and tiered model of ethical enforceability. Section 6 concludes with contributions and directions for future research.

2. Background of the Study

2.1. The Concept of Ethics in Artificial Intelligence

Ethics is a branch of philosophy that deals with moral principles and values that guide human behavior [26], and the term is often called the “moral language” [27] (p. 2). It concerns questions such as “What is the right thing to do?” and “What is the good life?” Numerous scholars have delved into the definition of the term “ethics.” According to Kidder [28], ethics is the “science of the ideal human character” or “the science of moral duty” (p. 63). To further refine the definition, ethics encompasses “a set of concepts and principles that guide us in determining what behavior helps or harms sentient creatures” [29] (p. 2). In the realm of science, ethics involves the contemplation of moral issues that arise during research, publication, and other professional endeavors [30]. Adeusi [26] identified three distinct aspects of ethics: normative ethics, descriptive ethics, and meta-ethics, as shown in Table 1.

Table 1. Aspects of ethics (source: Adeusi, 2020 [26]).

Normative ethics	The primary focus is to make moral judgments about particular kinds of actions and to offer reasons to support those judgments, to show that the judgments are reasonable (for example, prostitution is bad; it can lead to incurable diseases).
Descriptive ethics	It answers factual questions about the moral views of individuals or groups. Basically, statements of fact are made about the moral views held by individuals, groups, or society at large (for example, “Sussan believes prostitution is bad” or “Christians hate corrupt practices”).
Meta-ethics	Concerned with questions about the meaning of ethical terms. It attempts to analyze ethical concepts to determine their actual meanings and logical relations.

2.2. Ethical AI: Definition and Moral Philosophies

The definition of technology shapes our study of it [31]. Accordingly, this study adopts the definition of AI proposed by Kaplan and Haenlein [32], who describe AI as “a system’s ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation” (p. 17). Ethical AI refers to the development and deployment of AI systems that are ethical and aligned with

human values [33]. The ethics of AI can be divided into roboethics and machine ethics [34]. According to these authors, roboethics concerns how people should act when designing, building, using, and interacting with AI, whereas machine ethics concerns how artificial moral agents should act. Gorr et al.'s [35] critique of the performative view of moral status, which calls for a paradigm shift to treat moral status claims as assertions rather than declarations, illustrates the importance of a more deliberative approach to ethical decision-making in AI. This perspective is especially relevant in the context of roboethics and machine ethics.

To navigate the ethical intricacies of machine ethics, it is essential to draw upon established moral philosophies that provide a solid foundation for ethical decision-making in AI. Utilitarianism, defined by Jaquet [36] as a theory of right action that proposes a criterion for rightness, guides the design of AI toward actions that maximize overall well-being. This theory advocates pursuing the maximum achievable happiness, with every individual's happiness equally important [37]. This approach can explain the foundation of the first principle, 'Beneficence,' in the ethical AI framework outlined by Floridi and Cowls [16]. Micewski and Troy [38] differentiate between deontological and teleological ethics theories. According to the authors, deontological ethics prioritizes adherence to moral laws and duties, assessing the rightness of actions independently of their outcomes.

In contrast, teleological ethics, including its extreme form of radical consequentialism, evaluates actions based on the goodness of their consequences, thereby justifying the means by the ends. From a deontological standpoint, AI systems must be constructed and function in alignment with universal moral principles, including respect for autonomy and equity. In this vein, Albrechtslund [39] highlights value-sensitive design (VSD) as a pivotal framework for integrating ethical considerations into technological development. VSD emerges as "a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process." [40] (p. 1). Another theory in the ethical AI discussion is virtue ethics, which focuses on individuals' moral character and virtues, suggesting that variations in character lead to different responses in the same scenarios [41]. Thus, even in identical AI scenarios, the moral character of those involved in design and decision-making can significantly influence ethical direction and outcomes, highlighting the importance of virtue in the development and application of AI technologies.

2.3. Theoretical Framework of the Study

The moral frameworks mentioned in the previous section (utilitarianism, deontology, virtue ethics) do not just define ethics differently; they also differ in whether ethics can be institutionalized at all. Utilitarianism invites measurable outcome-based governance; deontology grounds universal duties that can be codified into law; virtue ethics, by contrast, resists formal institutionalization, depending instead on the character of those who design and implement systems. This variation matters enormously for AI governance: a declaration anchored in virtue ethics may remain aspirational by nature, while one grounded in deontological duties is structurally akin to enforcement. This tension, between what we declare and what we can actually enforce, lies at the heart of the present study.

To interpret AI ethics declarations as dynamic normative governance instruments, this study draws on Winston's [42] concept of the "norm cluster." The author argues that contemporary international norms are not singular, fixed objects but bounded collections of interrelated problems, values, and behaviors—what she calls a tripartite structure. Actors adopting a norm cluster need not perfectly replicate their predecessors' choices; rather, different combinations of the cluster's components are accepted as "close enough" by the adopting community, producing recognizable variation within a shared normative family.

This concept maps directly onto the phenomenon observed in this study: the AI ethics declarations analyzed do not adopt identical ethical principles, yet they are all recognizably part of the same global normative project. The shared problem is the ethical governance of AI; the values—fairness, transparency, accountability, privacy—form the cluster’s core; and the behaviors—the mechanisms through which these values are expressed—vary substantially across regions and actors. Winston’s framework thus provides the theoretical basis for why our frequency analysis finds both convergence and divergence: these are not contradictory findings but the expected signature of norm cluster diffusion.

We acknowledge that Winston’s framework is significant for explaining variation in norm adoption. However, it does not address a critical dimension of AI governance: the degree to which any given combination of norm components can be institutionalized and enforced. To illustrate, “fairness” as a principle may diffuse across all regions and appear in every country’s declaration, fulfilling the conditions for norm cluster adoption, and yet remain entirely without an enforcement mechanism. This study extends Winston’s framework by adding an institutional dimension: the question is not just what principles declarations contain; we must also ask about their degree of enforceability.

3. Methodology

3.1. Research Design

This study adopts a mixed-methods approach, combining two complementary methods. First, it conducts a descriptive comparative content analysis of 54 AI ethics declarations across global regions and actor types. Second, it employs a qualitative semi-systematic literature review to explore scholarly debates and practical challenges in translating ethical principles into enforceable mechanisms, particularly in commercial and platform contexts.

3.2. Descriptive Content Analysis

To address the first research question, a comparative dataset of 54 AI ethics declarations published between 2017 and 2025 was constructed (see Appendix A). The dataset consists of 45 national declarations selected based on their prominence in leading comparative studies of AI ethics declarations [18,24,25], ensuring that the sample reflects the countries most influential in shaping the global AI ethics landscape rather than a random or exhaustive census. In addition, 9 international or multi-stakeholder declarations from major global players (MGA) such as the OECD, UNESCO, the G7, and the EU. This allowed for a regionally balanced mapping of ethical principles across global contexts. Despite its significance, Australia was excluded from the regional comparative analysis because, unlike the other continental groupings—which include multiple national declarations enabling intra-regional comparison—Australia constitutes both a country and a continent, meaning its inclusion would represent an entire continental region through a single national declaration. This would introduce an asymmetry into the regional benchmarking framework, making Oceania structurally incomparable to other regions.

Each declaration was analyzed for its inclusion of AI ethical principles (e.g., transparency, fairness, privacy, accountability, human oversight) using a comparative coding framework. To illustrate, the coding procedure was deliberately straightforward and anchored in the structure of the declarations themselves. Each declaration included a dedicated section explicitly enumerating its AI ethical principles—for example, sections titled “core ai principles,” “ethical guidelines,” or equivalent. Principles were therefore directly extracted from these sections rather than inferred from the broader text, minimizing interpretive ambiguity. A single researcher conducted all coding, and each principle was recorded as either present (1) or absent (0) within a structured SPSS (version 27.0) dataset. Because coding relied on explicit declaration content rather than subjective judgment, in

other words, no inductive or interpretative coding was used, so formal intercoder reliability testing was not applicable to this study design [43]. This approach is consistent with direct extraction methods employed in comparable single-coder content analyses of AI ethics declarations [44].

Across the 54 declarations, terminological variation was observed in how certain principles were labeled—particularly those relating to well-being and information integrity. Terms such as “do no harm,” “beneficence,” “resilience and sustainability,” “social and environmental benefits,” and “societal and environmental well-being” were encountered across different declarations. These variants were consolidated under two umbrella categories: “societal well-being” and “environmental well-being.” Another typology to mention is “information integrity,” which appears differently in the declarations as “misinformation & disinformation,” “responsible information sharing,” and “deepfakes & criminal law.” Also, terms like “inclusiveness” mentioned in some declarations were grouped as “inclusion.” This consolidation approach follows the precedent established by Jobin et al. [18] and Corrêa et al. [24], both of which grouped terminological variants under shared umbrella categories in their comparative analyses of AI ethics declarations.

As a final step, national-level findings were situated within broader governance trends through a systematic comparison of the 45 national declarations with the 9 global declarations, using major AI governance frameworks as reference points (e.g., the OECD AI Principles and the UNESCO Recommendation on the Ethics of AI).

3.3. Qualitative Analysis

3.3.1. Review Approach

The second research question was addressed through a semi-systematic literature review (SSLR). SSLR was chosen to identify patterns, conceptual debates, and practical challenges associated with the operationalization of AI ethics across scholarly and policy-oriented publications [45]. This SSLR aims to examine, critique, and summarize the literature on a research topic, allowing new theoretical frameworks and viewpoints to emerge [46,47]. Scholars have captured this type of literature across various fields, such as business [48] and the environment [49].

3.3.2. Search Strategy

Following prior scholars who conducted systematic reviews using two databases in their search (e.g., [50]), and in line with the recommendation of Dabić et al. [51] to use two or more databases to avoid potential bias in the results, this study used Google Scholar (GS) and Web of Science (WOS) as the primary search databases. GS was selected for its reliability and consistency in citation counts [52], and for its platform for disseminating scholarship and facilitating networking [53]. GS is also considered the most comprehensive academic database [53,54] covering a diverse array of cross-disciplinary studies and sources that may not be fully indexed in traditional databases [43,55]. Accordingly, GS is particularly appropriate for a cross-disciplinary topic such as AI ethics, which sits at the intersection of computer science, philosophy, law, and governance. The use of the GS search database is consistent with the Systematic Reviews and Meta-Analyses (PRISMA) 2020 reporting guidelines [56], which are intended to guide authors in documenting what was searched and found, while also providing readers with a transparent window into the steps taken throughout the review process [57]. The search goes further by applying the same search terms in WOS, due to its replicability features [55,58] and its selective, high-quality journal coverage [59].

The review, across both databases, focused on publications released between 2017 and 2025, the same timeframe used in the empirical analysis of AI ethics declarations. This

period marks a critical phase in the global development of AI ethics discourse, particularly in relation to governance debates and sector-specific implications.

The following Boolean search string was applied in Google Scholar: (“ethical artificial intelligence” OR “ethical AI declarations” OR “ethical AI guidelines” OR “ethics in AI”). Given Google Scholar’s interface limitations, filters were manually applied to refine the results. The following inclusion parameters were used: (1) publication language restricted to English; (2) time frame; and (3) document type limited to peer-reviewed review articles. In WOS, the search was conducted using the Topic field tag (TS), which searches titles, abstracts, author keywords, and Keywords Plus, applying the following query: TS = (“ethical artificial intelligence” OR “ethical AI declarations” OR “ethical AI guidelines” OR “ethics in AI”). Filters applied in WOS are illustrated in Figure 1, in the screening phase.

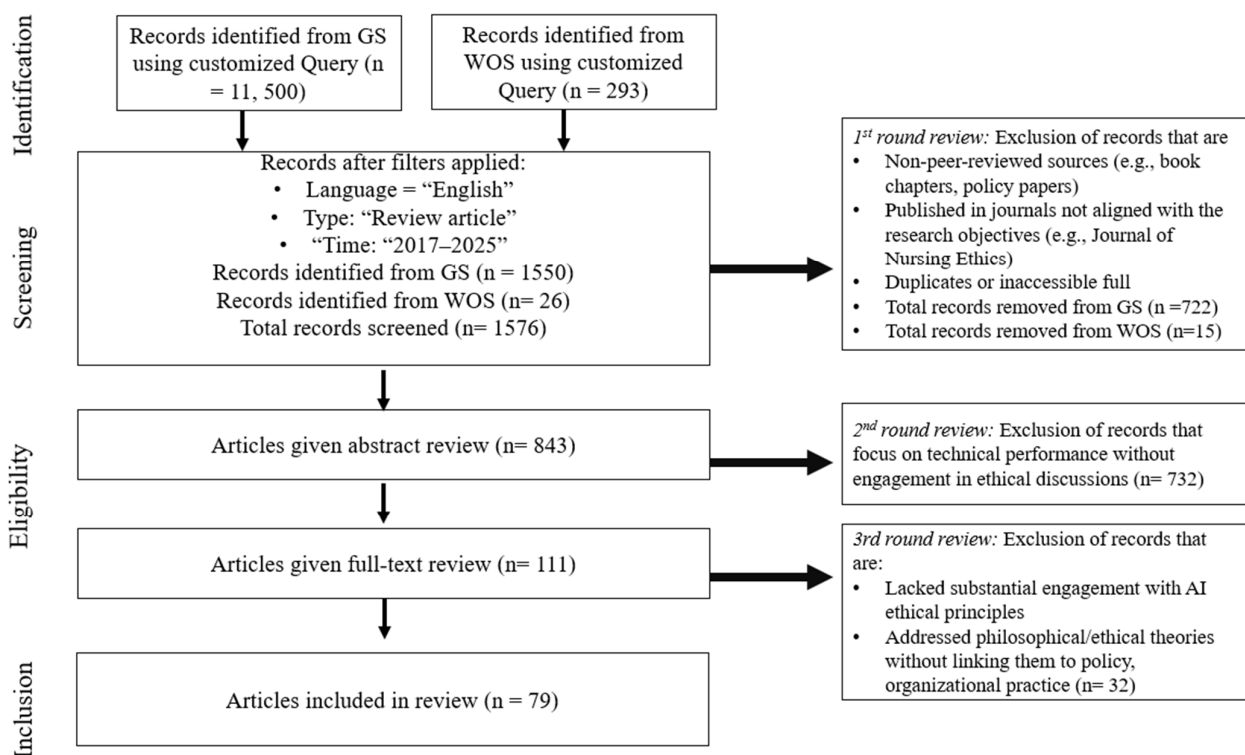


Figure 1. PRISMA flow diagram for the literature review. Note. Source: Authors’ elaboration (adapted from Mazrekaj [60]).

Quality assessment was applied as part of the inclusion criteria. Only peer-reviewed journal articles were considered eligible for inclusion. Gray literature, working papers, blog posts, and non-peer-reviewed policy documents were excluded from the literature review.

3.3.3. Inclusion and Exclusion Criteria

The relevance, quality, and analytical coherence of the reviewed literature were ensured through the application of clear inclusion and exclusion criteria, adopted from Tranfield et al. [61]. Figure 1 presents the Preferred Reporting Items for PRISMA flow diagram, which illustrates the identification, screening, and inclusion process used to select the final set of studies, following best practices outlined by Mazrekaj [60].

Included studies met at least one of the following conditions:

- They explicitly addressed opportunities or challenges related to the implementation or operationalization of AI ethical principles.
- They discussed the governance of AI ethics in policy, legal, or applied organizational contexts.

- They analyzed soft law, ethical frameworks, or emerging tools (e.g., audits, impact assessments) within organizations.

Articles focusing on AI development or performance that did not engage with ethical frameworks, ethical principles, or governance implications were excluded.

4. Results

4.1. Analysis of Ethical Principles in AI Declarations

In recent years, significant advances in AI systems’ capabilities have created an ethical debate about society’s opportunities and concerns. Since 2016, a wide array of global stakeholders, including policymakers, industry leaders, civil society organizations, think tanks, media, and consultancy firms, have been actively engaged in intense discussions [62].

4.1.1. Comparative Presence of AI Ethics Principles in National Declarations

The frequency analysis of the ethical AI principles across national AI declarations. The results, illustrated in Figure 2, reveal both convergences and notable divergences in the prioritization of principles.

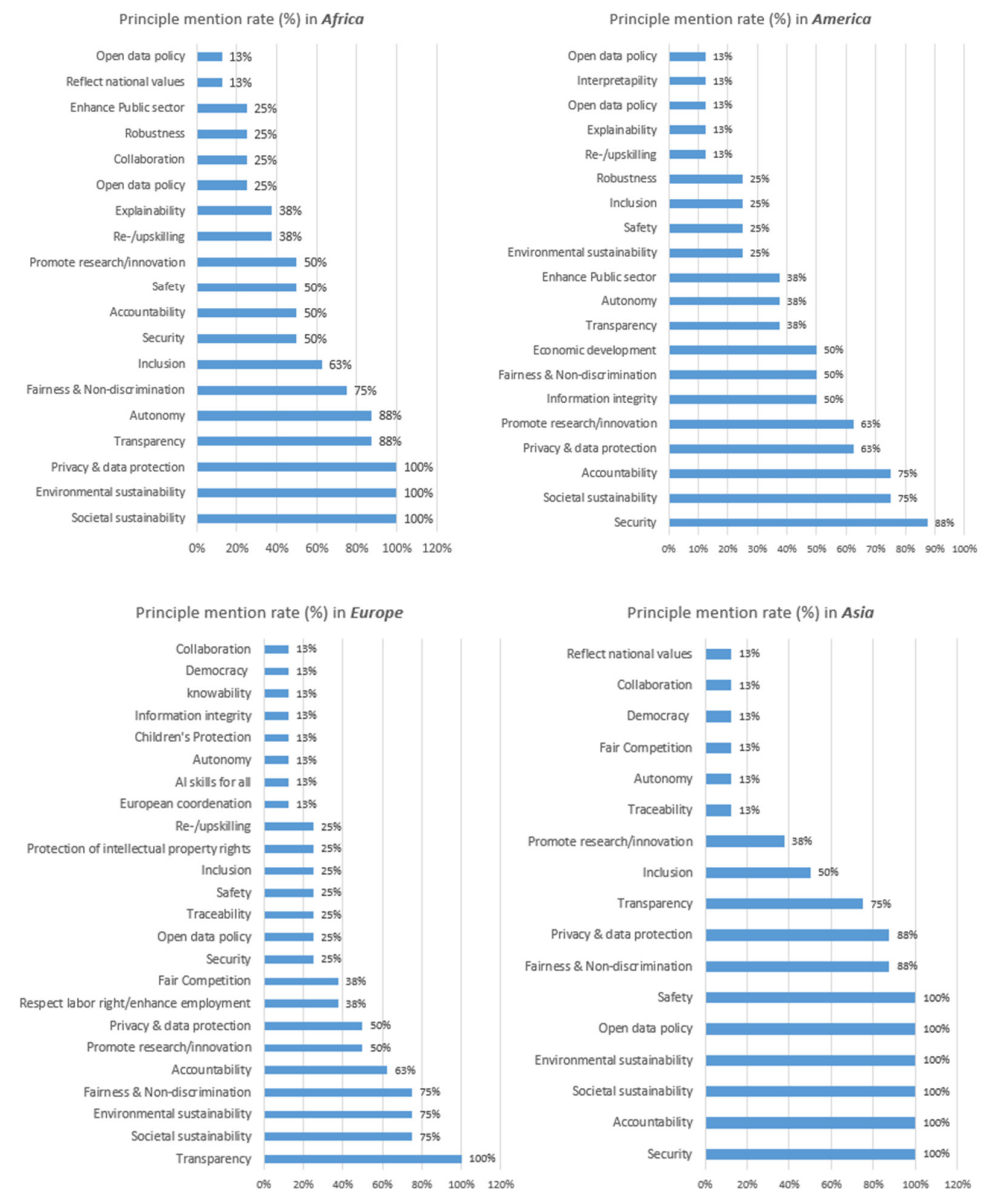


Figure 2. Comparative overview of AI ethical principles across global regions.

European declarations most frequently mention transparency (100%), fairness and non-discrimination (75%), environmental sustainability (75%), societal sustainability (75%), accountability (63%), and privacy and data protection (50%). Less frequently mentioned were principles such as democracy, information integrity, and autonomy (13% each), suggesting that Europe’s discourse focuses more on institutional safeguards and technocratic governance than on normative pluralism.

Asia shows universal inclusion (100%) of safety, security, accountability, environmental sustainability, societal sustainability, and open data policy across the examined declarations. Transparency is also highly emphasized (75%), alongside privacy and data protection (88%) and fairness and non-discrimination (88%). This feature reflects a strong focus on control-oriented and collective values, with an emphasis on responsible innovation within national development goals.

Africa also demonstrates strong convergence on environmental sustainability, societal sustainability, privacy and data protection, transparency, inclusion, and autonomy (all above 75%). Explainability, promoting research/innovation, and re-/upskilling appear in 50% of cases, highlighting an emphasis on capacity building and human-centered development. However, mentions of principles such as interpretability, collaboration, and robustness remain limited (<25%).

In the Americas, security (88%) and societal sustainability (75%) rank highest, followed by accountability, privacy and data protection, and fairness and non-discrimination (63–75%). The region shows a broader spread across economic development (50%), autonomy (38%), and enhancing the public sector (38%). However, low-frequency principles like interpretability, open data policy, and explainability (13%) suggest a fragmented approach, potentially shaped by the diversity of national policies across North and Latin America.

4.1.2. Comparative Presence of AI Ethics Principles in Major Global Players

The analysis of AI ethical principles across nine major global players’ (MGAs) declarations reveals three distinct clusters of prioritizations, illustrated in Figure 3:

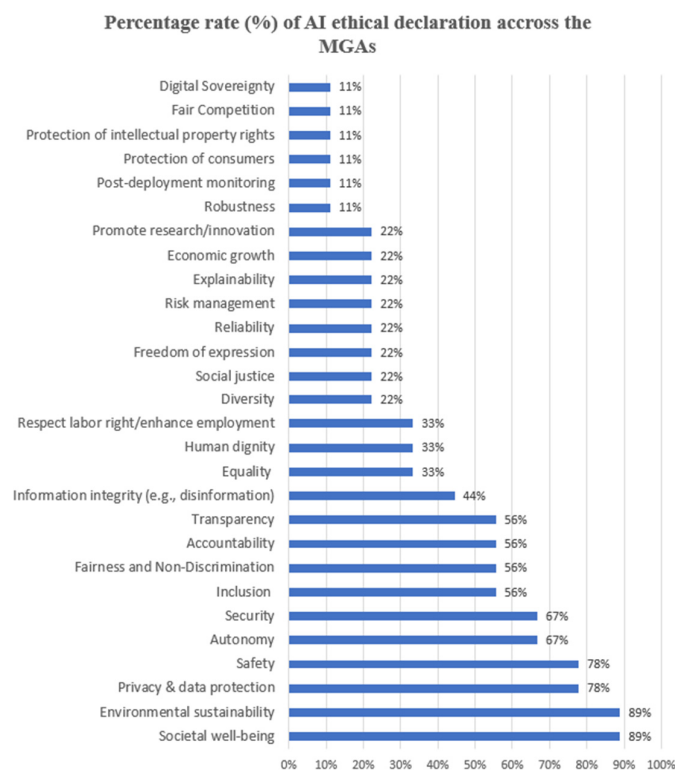


Figure 3. Comparative overview of AI ethical principles across MGAs.

The most consistently cited principles, appearing in nearly 70% of the declarations, are: autonomy, safety, security, privacy and data protection, societal well-being, and environmental sustainability.

A second group of principles emerges as foundational, ranging from 40% to 60% in the declarations. These include transparency, accountability, fairness and non-discrimination, inclusion, security, autonomy, and information integrity. The third group comprises several ethical dimensions that appear more sporadically, indicating either emerging concern or underdevelopment in global declarations. These include values such as digital sovereignty, post-deployment monitoring, robustness, diversity, consumer protection, freedom of expression, and fair competition.

4.2. Integrative Literature Review Findings

4.2.1. Realizing the Potential: Strengths and Opportunities in Ethical AI Declarations

Strengths and opportunities represent internal and external capabilities that can aid or facilitate an organization's development [63]. Echoing ethical AI declarations and pinpointing these strengths and opportunities is crucial for businesses to leverage their assets to navigate the ethical landscape of AI usage and foster innovation responsibly within the AI sector.

Global Participation as a Foundation for Inclusive AI Governance

Global participation in AI declarations is a strength, stressing the crucial inclusion of historically underrepresented communities in AI development, from local engagement in the Global South to significant roles in model development and strategic decision-making, fostering a globally diverse and inclusive AI ecosystem [64]. For example, President Biden's Executive Order on AI on 30 October 2023, incorporates eight guiding principles and priorities that mirror the OECD's [65]. This breadth of AI ethical declarations originates from various sources, including governments, the private sector, and non-governmental organizations.

Based on the government's outlook, Jobin et al. [18] found that a considerable portion of AI ethics principles originated in the United States (25%) and the United Kingdom (15.5%), with EU member states, Canada, Iceland, and Norway collectively contributing 26%. Japan accounted for 4.8%, while the UAE, India, Singapore, South Korea, and Australia each produced one document, totaling 6%. This data suggests that Western values significantly influence the landscape of AI ethics principles, with these regions representing 67% of the total. The 'Data Free Flow with Trust' initiative led by Japan, highlighted by Greenleaf [66], exemplifies such international cooperation, aiming to build trust and cooperation across digital economies during the G20 Leaders' Summit in Osaka. Although there were challenges in forming a unified position due to the broad terms and differing levels of engagement from various countries (24 in total), this initiative represents a significant effort towards creating "global data privacy rules" (p. 1). Fukuda-Parr and Gibbons [67] also observed that over half of the 15 reviewed ethical AI guidelines were grounded in international human rights law, further illustrating how these declarations support ethical governance and standards in AI globally by embracing universal values such as equality, freedom of expression, and media freedom. For instance, the Global Declaration on Information Integrity Online (published in 2023) established high-level international commitments by various participating states [68]. This foundation enhances their credibility and effectiveness. Further reinforcing this global momentum, the European Union reached a landmark milestone in 2024 with the enactment of the EU AI Act—the world's first comprehensive legislative framework governing artificial intelligence—which officially entered into force across all 27 EU Member States on 1 August 2024. The Act

introduces a structured, risk-based classification system alongside mandatory transparency requirements, marking a pivotal step in the formal regulation of AI at an international scale [69].

Through the lens of the private sector and non-governmental organizations. Analysis from the AI Ethics Lab indicates that among the 103 documents surveyed, 34% emanate from private companies. Fjeld et al. [70] also note that, in their examination of 36 documents, private funding accounted for 22%. NGOs are pivotal in shaping AI ethics. For instance, the CEPEJ (European Commission for the Efficiency of Justice) has issued ethical guidelines for AI in judicial systems, highlighting NGOs' critical role in expanding and deepening the understanding of AI ethics across industries. In this vein, Schiff et al. [71] examined 112 ethical AI documents and noted that 26 were issued by the private sector and 32 by NGOs. They also observed differences in priorities across these sectors: the private sector often prioritizes technical challenges like algorithmic bias and transparency, NGOs take a broader approach that includes issues such as accountability and misinformation, and the public sector typically focuses on societal impacts like unemployment and economic growth. In another study, Schiff et al. [72] observed that the private sector was responsible for creating roughly one-quarter of the ethical AI documents, highlighting a pattern of AI leadership within the private sector that mirrors that found in the public sector. Furthermore, Auld et al. [73] emphasized the role of the private sector as delineated in the Montreal Declaration. It asserts the necessity for government entities and private organizations involved in creating and implementing machine learning systems to scrutinize and address associated risks promptly. Moreover, the authors highlight the crucial need to include experts from academia, legal backgrounds, and civil society in these conversations.

Comprehensive Coverage Harmonization of Global Standards

The declarations and efforts span a broad spectrum of ethical considerations in AI, covering a comprehensive strategy towards AI governance. The literature captured this trend, with projections by Corrêa et al. [24] indicating an upward trajectory to over 200 AI ethics guidelines documents by 2023. Also, Smuha [74] delves into the framework set forth by the European Commission's High-Level Expert Group on AI, heralding it as a "comprehensive framework to achieve Trustworthy AI" (p. 97). Similarly, the research conducted by Rothenberger et al. [75] thoroughly explores AI ethics within contemporary discussions, beginning with a detailed review of the literature on classical Western ethical theories, thereby laying the groundwork for crafting essential AI ethical guidelines. Following this, the authors conducted qualitative interviews with experts to secure detailed insights and prioritize these guidelines based on their significance. Their findings show that the arithmetic mean ranking puts "responsibility" at the top (4.51), followed by "protection of data privacy" (4.4), "transparency" (4.28), "robustness" (4.25), "minimization of bias" (3.87), and the need for AI to have a clear purpose (3.58). Also, in his analysis of 22 AI guidelines, Hagendorff [25] found that accountability, privacy, and fairness are present in approximately 80% of them.

Díaz-Rodríguez et al. [76] make a compelling case for a comprehensive AI governance strategy, building on the contributions of Corrêa et al. [24] and Rothenberger et al. [75] by introducing a framework that safeguards the integrity of all processes and stakeholders throughout the AI system's lifecycle. In parallel, Morley et al. [20], identified a compilation of over 70 documents published between 2017 and 2019, put forward a typology that brings together ethical principles and the distinct phases of the AI lifecycle, ensuring that ethical considerations are not confined to a single point but are woven into each stage of development and deployment. Rothenberger et al. [75] contribute to this dialogue by providing an extensive analysis that advocates for a multidisciplinary approach to

AI ethics, encompassing universal principles, philosophical insights, and a risk-based regulatory framework. In this context, Cappelli and Di Marzo Serugendo [77] developed a semi-automated software model designed to assess whether AI systems operate in accordance with established ethical standards. The model is based on thirteen general ethical principles and their sub-principles, which come from a variety of European and international regulatory frameworks. Central to their approach is a two-tiered evaluation structure: the first tier determines whether an AI system satisfies overarching ethical principles, while the second conducts a more granular, sector-specific examination of the relevant sub-principles

Addressing Ethical Issues in AI

One of the main reasons that explains the hype surrounding ethical AI is its ability to address ethical issues. Siau and Wang [34] categorize research into AI ethics into three primary areas: ethical challenges inherent to AI’s characteristics, ethical risks associated with human factors, and strategies for instilling ethical behaviors in AI systems. These authors also distinguish between the ethical considerations specific to narrow AI and weak AI, noting that the ethical implications of narrow AI are particularly concerned with human aspects, with Bostrom and Yudkowsky ([78], p. 5) suggesting, “a different set of ethical issues arises when we contemplate the possibility that some future AI systems might be candidates for having moral status.” A notable strength of AI declarations, summarized in Table 2 by Morley [20] and initially conceptualized by Mittelstadt et al. [19], is their capacity to tackle ethical concerns. In a similar study, Stahl et al. [79] undertook an in-depth analysis of the ethical challenges presented by AI applications, uncovering a broad spectrum of issues. Their research delineates 39 unique ethical dilemmas, encompassing everything from the potential impact on innovation and risks to physical safety to more intricate problems such as power imbalances, job losses, and the possibility of exploitation for military objectives.

Table 2. Ethical concerns related to algorithmic use (Source: Morley et al. [20], adapted from the ‘Map’ created by Mittelstadt et al., 2019 [19]).

Ethical Concerns	Explanation
Inconclusive evidence	Algorithmic conclusions are probabilities and therefore not infallible. This can lead to unjustified actions. For example, an algorithm used to assess credit worthiness could be accurate 99% of the time, but this would still mean that one out of a hundred applicants would be denied credit wrongly.
Inscrutable evidence	A lack of interpretability and transparency can lead to algorithmic systems that are hard to control, monitor, and correct. This is the commonly cited “black-box” issue.
Misguided evidence	Conclusions can only be as reliable (but also as neutral) as the data they are based on, and this can lead to bias. For example, Dressel and Farid [80] found that the COMPAS recidivism algorithm commonly used in pretrial, parole, and sentencing decisions in the United States, is no more accurate at fair predictions made by people with little or no criminal justice expertise.
Unfair outcomes	An action could be found to be discriminatory if it has a disproportionate impact on one group of people. For instance, Selbst [81] articulates how the adoption of predictive policing tools is leading to more people of color being arrested, jailed or physically harmed by police.
Transformative effects	Algorithmic activities, like profiling, can lead to challenges for autonomy and informational privacy. For example, Polykalas and Prezerakos [82] examined the level of access required for personal data by more than 1000 apps listed in the “most popular” free and paid for categories on the Google Play Store. They found that free apps requested significantly more data than paid-for apps, suggesting that the business model of these “free” apps is the exploitation of personal data.
Traceability	It is hard to assign responsibility for algorithmic harms and this can lead to issues with moral responsibility. For example, it may be unclear who (or indeed what) is responsible for autonomous car fatalities. An in-depth ethical analysis of this specific issue is provided by Hevelke and Nida-Rümelin [83].

AI for Social Good and the Preservation of Trust

The main contribution of a regulatory framework for AI is to “build trust among consumers and businesses in AI, and therefore speed up the uptake of the technology” ([84], pp. 9–10). One of the dominant frameworks for ethical AI in the literature proposes five principles: Beneficence, non-maleficence, autonomy, justice, and explicability [16,17]. Table 3 illustrates the five principles.

Table 3. Overview of ethical AI principles (source: author).

Principles	Definitions
Beneficence	This principle has several definitions in declarations, for instance, promoting human well-being and flourishing, peace and happiness, creating socio-economic opportunities, and economic prosperity [18]. As the EU’s High-Level Expert Group on AI ([85], p. 4) emphasizes, “AI is not an end in itself, but rather a promising means to increase human flourishing, thereby enhancing individual and societal well-being and the common good, as well as bringing progress and innovation.”
Non-maleficence	This principle emphasizes the duty not to inflict harm on others, resonating with the <i>primum non nocere</i> guiding principle, or “first, do no harm” [86].
Autonomy	The dimensions of autonomy encompass capacities for self-determination, normative obligations to respect and support it, relational recognition, self-respect, its exercise, and necessary material, economic, legal, cultural, and informational conditions, without these conditions being constitutive of autonomy itself [87].
Justice	It provides a set of standards by which to fairly adjudicate people’s (often competing) claims to liberties, opportunities, resources, and modes of treatment [88].
Explicability	It refers to the ability of AI systems to provide clear, understandable explanations of their decision-making processes, outcomes, and behaviors [89].

4.2.2. Addressing the Challenges: Limitations and Weaknesses in Ethical AI Declarations

Although Namugenyi et al. [90] did not directly focus on ethical AI, their characterization of weaknesses as internal limitations offers a framework for pinpointing distinct challenges within ethical AI declarations. This includes identifying opportunities for enhancement, barriers to expansion, and ingrained problems that hinder the successful execution and compliance with ethical norms in AI development and application.

Lack of a Standard Framework for Ethical AI

A notable weakness in ethical AI declarations is the lack of a standardized ethical AI framework, leading to varied interpretations and implementations of AI ethics across different entities and sectors. A primary reason for this gap is rooted in the disciplinary divide, as ethics experts and AI developers speak different languages [71].

Jobin et al. [18] highlighted this issue and found considerable diversity among 84 ethical AI documents. Their analysis revealed the absence of any universally adopted ethical principle, indicating a fragmented approach to AI ethics. Similarly, Fjeld et al. [70] analyzed documents from diverse organizational backgrounds and identified 8 commonly cited principles. These were fairness/non-discrimination (cited in every document analyzed), followed by privacy and accountability (each at 97%), transparency/explainability (94%), safety/security (81%), professional responsibility (78%), maintaining human control over technology (69%), and upholding human values (also 69%). Despite identifying these commonalities, the organization’s diversity underscores the disparity in ethical prioritization, further illustrating the lack of a standardized ethical framework.

Furthermore, Hagendorff [25] investigated 21 declarations and identified recurring themes: accountability, privacy, justice (each cited in 77% of the declarations), and transparency (68%). Despite identifying these commonalities, the authors illustrate the lack of a standardized ethical framework.

Tomašev et al. [91] listed 10 principles crucial to AI collaborations for sustainable development, emphasizing realistic expectations, simplicity, inclusivity, ethical compliance, clear goals, and durable partnerships to tackle significant challenges. These principles emphasize the value of aligning incentives, maintaining trust, pursuing cost-effective AI development, ensuring data readiness, and ensuring secure data processing, all in line with respect for human rights and privacy.

Franzke [44] examines 70 international ethics guidelines spanning academia, NGOs, and corporations, highlighting frequent mentions of transparency, security, and privacy. Similarly, Royakkers et al. [92] conducted a systematic literature review on ethical technology, identifying key themes such as privacy, security, autonomy, justice, human dignity, control over technology, and the balance of powers, further underscoring the range of ethical considerations in AI discussions.

Expanding on this research, Corrêa et al. [24] analyzed 200 documents from 37 countries across six continents and five languages, uncovering 17 ethical principles commonly cited, each appearing in more than 10% of the documents. Their findings indicate consistency in the top five ethical principles across North America and Europe, while Asia distinctly prioritized beneficence/non-maleficence (74%) and accountability/liability (70%). The study also found that there were differences between regions. For example, freedom/autonomy and beneficence/non-maleficence were more important in the US than cooperation/fair competition and diversity/inclusion were in the UK. Differences by institution type were also observed, with governmental entities prioritizing transparency (89.5%), private corporations prioritizing reliability (87.5%), and CSOs/NGOs prioritizing fairness (88.2%).

The variation in the frequency of these themes' mentions suggests inconsistencies in ethical prioritization, reflecting once more the absence of a unified framework for ethical considerations in AI.

Divergence in Definitions and Terminology in AI Ethics

Before addressing the lack of consensus around key terms in the AI ethics context, scholars such as Hagedorff [25] noted that the ethics guidelines examined refer exclusively to the term 'AI'. They rarely use more specific terminology. However, 'AI' is just a collective term for a wide range of technologies or an abstract, large-scale phenomenon.

With this in mind, Theodorou and Dignum [33] highlight the challenge of defining the ethical principles guiding AI's design and use, as discrepancies often stem from the diverse cultural, legal, and societal contexts from which these principles emerge. Similarly, Mittelstadt [19] stated that the work on AI ethics has primarily produced vague, high-level principles. In this regard, Whittlestone et al. [93] note that ambiguity and differing interpretations of critical ethical terms complicate their universal application. For instance, the Montreal Declaration [94] interprets 'well-being' as enhancing users' mental and physical health and living conditions. In contrast, the Australian AI Ethics Framework [95] introduces 'net benefit' to denote AI systems generating more benefits than costs.

Among the scholars working in this area, Gree Binns [96] discusses the broad consensus on 'fairness' yet acknowledges the intense political debates surrounding its definition. Cultural differences also affect the prioritization of values, such as privacy. Also, Green et al. [66] claim that values such as "fairness" and "accountability" are "so vague and contested" that organizations can interpret them to align with existing corporate incentives.

In a broader scope, Franzke [44] argues that most existing guidelines do not adequately define "ethics" and related terms. Ryan and Stahl [79] highlight how the tone of AI ethical declarations ranges from mandatory "must do" principles to more advisory "if possible" suggestions. Furthermore, the variability in document length and accessibility, ranging

from summaries to detailed manuals exceeding 266 pages, is designed to appeal to a range of stakeholders, from everyday users to developers, corporations, policymakers, and the general public. This variation emphasizes the difficulty in formulating a unified set of principles for AI ethics.

The use of terminology also shows divergence, with terms like “responsible AI” used by the Montreal Declaration, “beneficial AI” by the Future of Life Institute, and “Trustworthy AI” by UNESCO indicating a lack of consensus in conceptualizing AI ethics.

Challenges in Translating Principles into Practices

A significant shortcoming of existing ethical AI declarations is that they include principles that are considered value statements, promising to be action-guiding, but in practice offer few specific recommendations and fail to address fundamental normative and political tensions embedded in key concepts [19]. Scholars like Schiff et al. [71] consider the primary reason for this challenge to be the mismatch between organizational incentives and those of ethicists, as ethical principles might conflict with profit motives. From an organizational perspective, seven key components that could impact the effective adoption of AI principles: (1) communication, which raises the question of whether awareness should be limited to managers or extended organization-wide; (2) management support, where endorsement exists but behavioral modeling is lacking; (3) training, which faces resistance when mandatory; (4) ethics oversight, where accountability remains diffuse in the absence of a dedicated officer; (5) reporting mechanisms, with non-malicious breaches often going unaddressed; (6) enforcement, where penalties apply only to malicious violations; and (7) measurement, which remains underdeveloped in most organizations [97].

Recent systematic reviews confirm insufficient operationalization as a central knowledge gap, noting that principles focus on the ‘what’ rather than the ‘how’ of AI ethics [92]. Toreini and colleagues [98] identify a critical missing element: mid-level norms that bridge high-level principles (e.g., ‘fairness’) with low-level technical requirements. Empirical studies reveal additional barriers, including role ambiguity (who is responsible for implementing ethics?), competing priorities (transparency vs. operational efficiency), and conflicting principles that resist simultaneous optimization [99,100]. Building on Tang et al.’s [101] finding that only 1.8% of empirical evidence addresses AI’s ethical issues, Arbelaez Ossa and colleagues [102] argue that this gap between high-level ethical principles and practice reveals a deeper problem: “the real work of AI ethics research must focus on translating and implementing principles to practice not in a formulaic way but as a path to understanding the real ethical challenges of AI that may go unnoticed in theoretical discussions” (p. 3). In their review of the UK’s AI Assurance Roadmap, Barrance et al. [103] contend that a developed ecosystem comprising regulation, technical standards, independent auditing, and professional accreditation is necessary to translate ethical AI principles into trustworthy governance. They also identify “ethics washing” as a significant risk when assurance is selective rather than systematic. In this vein, according to Dara et al. [47], even within a single applied sector, like agriculture, it is still challenging to operationalize the principles of fairness, transparency, accountability, privacy, and robustness. According to their analysis, there are ongoing accountability gaps in digital agriculture, where technology companies deny accountability for AI-driven decisions and privacy violations take place in spite of declared ethical commitments.

One of the key reasons for the challenge of translating principles into practice is the tension among these principles during implementation. Following this line of thought, Whittlestone et al. [93] highlight four principal conflicts inherent in AI deployments. The initial conflict involves the trade-off between enhancing service quality and efficiency through data use and safeguarding personal privacy and autonomy, underscoring the

challenges of privacy protection and the importance of obtaining meaningful consent. The second conflict lies in balancing the precision of algorithmic decisions and forecasts with the imperative of ensuring fairness and equal treatment for all. The third conflict is between the benefits of more personalization in digital spaces and the need to promote citizenship and solidarity. The fourth and final conflict juxtaposes the convenience and empowerment that automation brings against the necessity of fostering self-actualization and dignity. Specifically, while automation has the potential to make arts, science, and languages more accessible to groups previously marginalized, it concurrently poses the risk of widespread deskilling, the solidification of practices, homogenization, and the erosion of cultural diversity. Building on this observation, principles of AI ethics predominantly concentrate on system improvements but “rarely question the business culture, revenue models, or incentive mechanisms that continuously push these products into the markets” ([104], p. 43).

Moreover, Attard-Frost et al. [105] point out a fundamental weakness in AI ethics guidelines: a concentrated focus on algorithmic decision-making, such as fairness and transparency, while often overlooking the ethical implications of AI’s business models and the broader political and economic contexts in which AI operates. Ray [106] offers an in-depth analysis of conversational AI system evaluations, including ChatGPT, by reviewing various benchmarks and standards from OpenAI, IEEE, the Montreal Declaration, and the Partnership on AI, indicating a need for a more comprehensive approach to AI ethics that includes business and political considerations to ensure holistic AI governance. This suggests that for AI ethics guidelines to be fully practical, they must encompass both the technical aspects and the socio-economic impacts of AI technologies.

Integration with Existing Legal Frameworks

The European Commission [107] asserts that both activity in and the trustworthiness of AI necessitate “a regulatory framework that is flexible enough to promote innovation while ensuring high levels of protection and safety” (p. 8). For this aim, the General Data Protection Regulation (GDPR) stands as one of Europe’s most notable pieces of legislation, setting a benchmark for the importance of legal foundations in technology governance, encompassing the realm of XAI [108]. Yet aligning AI ethical principles with existing laws and regulations can be complex [109]. According to Pagallo [110], a critical obstacle to incorporating AI ethics declarations into current legal structures is the complex task of implementing the GDPR’s forward-looking data protection strategies amid the rapidly evolving realms of Big Data and AI. Furthermore, the authors underscore the GDPR’s focus on early risk assessments and highlight the difficulty of maintaining legal adaptability. Similarly, Vedder and Spajić [111] found that prioritizing informed consent, the GDPR, and European recommendations on health data may inhibit the moral duty to share data to enhance healthcare, despite various ethical justifications supporting that duty.

Additionally, Owczarczuk [112] identifies three main challenges in formulating AI regulations: defining AI precisely for regulatory clarity, addressing ethical threats to uphold principles and human rights, and managing the competitive nature of AI regulation across key global regions like the EU, the U.S., and China, each with distinct governance approaches.

The challenge of effectively integrating AI ethical principles into law, noting that public sector organizations often aim to establish regulatory groundwork, whereas private entities focus on compliance and marketing their ethical practices, underscores the complexity of aligning ethics with legal standards [71]. This trend in the private sector has raised concerns about “ethics washing,” “ethics shopping,” or “ethics shirking,” in which the outward appearance of ethical commitment masks a lack of substantive action [23].

Erosion of Trust Due to Limited Focus on Impact

Recent literature points out significant deficiencies in ethical AI declarations. Bolte et al. [113] criticize the lack of discourse on AI/ML's environmental sustainability, indicating a gap in addressing ecological impacts. Morley et al. [20] articulate apprehensions regarding the inadequacy of AI systems in acknowledging user autonomy, as machine learning frequently merges prediction with decision-making without examining the ramifications for actionable decisions. Hagendorff [25] identifies additional neglected areas, including labor rights, the potential for technological unemployment, the militarization of AI, the proliferation of disinformation, and the risk that AI will be used for dual purposes. The study also underscores the absence of dialogue on long-term risks, such as safety and existential threats posed by artificial general intelligence. Furthermore, Corrêa et al. [24] reviewed 200 AI governance guidelines and found that only 1.5% addressed long-term AI risks. Also, Lokshina et al. [114] identify factors influencing trust in AI, including perceived benefits and assurance of technology's reliability.

Several examples highlight the narrow attention given to the repercussions of AI technologies. For instance, AI applications have inferred sensitive personal details, such as political leanings or sexual orientation, from Facebook interactions [115] and have evaluated mental health states by analyzing Facebook statuses [116]. An early 2020 incident involved Clearview, a company in New York City, creating an AI tool that could build detailed profiles of individuals using data found online, raising significant alarm over privacy degradation [117]. More recently, Italy's Data Protection Authority imposed a ban on ChatGPT, citing OpenAI's inadequate justification for data gathering and processing and the absence of measures to prevent minors from accessing unsuitable content [118]. Despite AI declarations such as the EU Ethics Guidelines for Trustworthy AI highlighting the significance of privacy, real-world applications show that privacy breaches still occur, as illustrated by AI's ability to derive personal information from users' social media behavior.

Conflicting Objectives in Ethical AI Declarations

The literature highlights two main types of conflicts. First, the conflict stems from the objective perspective. In this vein, the balance between innovation and regulation is a crucial concern. Claessens et al. [119] raise concerns that strict regulations might stifle AI innovation, suggesting that regulatory efforts to control AI could inadvertently curtail its potential and global market competitiveness. This concern is echoed in the White House's draft guidance for AI regulation [120], which advocates a regulatory approach that impedes AI innovation, reflecting the competitive imperative to keep advancing and innovating within the industry. Truby et al. [121] argue that principles such as transparency, accountability, and fairness can foster AI growth and innovation, indicating a trend toward responsible self-regulation and a balance between ethical integrity and competitive progress. Arvan [122] presents the 'moral-semantic trilemma' in ethical AI programming, emphasizing three primary challenges: the formulation of guidelines that are excessively rigid or overly flexible, and the unpredictability of outcomes. This trilemma illustrates the difficulty of crafting AI ethics declarations that balance specificity and flexibility to ensure consistent, practical application across diverse contexts. Furthermore, resolving this trilemma risks fostering skepticism among developers and stakeholders about the viability and consistency of ethical AI, potentially undermining wider adoption and trust in AI systems governed by these ethics declarations. Thus, pursuing ethical AI, which is viewed as a potential competitive disadvantage in the short term, poses a significant challenge to advancing ethically aligned AI technologies.

Second, the conflict arises from the principle's objective, which is also considered a challenge for practitioners. In this context, Sanderson et al. [123] identify four main conflicts

among principles. Privacy protection conflicts with accuracy (more data improves model performance but violates privacy principles), fairness conflicts with accuracy (restricting data scope to address bias reduces model effectiveness), and transparency conflicts with security (explaining models may expose vulnerabilities). Another study by Song et al. [124] demonstrates the apparent trade-off between robustness and privacy, noting that while enhancing robustness, adversarial defense mechanisms may inadvertently heighten the model's susceptibility to membership inference attacks, thereby elevating the risk of training data exposure. As Zhu and Lu [125] argue, effective AI governance requires "managing tensions among different goals and logics" rather than assuming principles can be universally satisfied. Critically, Bleher and Braun [126] identify that ethically aligned approaches lack justification theories to adjudicate between competing principles, leaving practitioners without systematic methods to resolve conflicts.

Geopolitical Tension

Geopolitical tensions delineate how global power dynamics, particularly between the United States and China, complicate collaborative efforts in AI regulation and ethical governance [127]. In a related aspect, the European Union seeks to set itself apart from the United States and China by focusing on an approach that is ethical, human-centric, and grounded in values, as exemplified by the European Commission [128]: "There is strong global competition on AI among the USA, China, and Europe. The USA leads for now, but China is catching up fast and aims to lead by 2030. For the EU, it is not so much a question of winning or losing a race but of finding a way to embrace the opportunities offered by AI in a way that is human-centered, ethical, secure, and true to our core values" (pp. 12–13). Bächle and Bareis's [129] study examines the complex geopolitical factors that influence Autonomous Weapon Systems (AWS), highlighting how these technologies are intertwined with national power ambitions. The study emphasizes that the challenges associated with AWS—ranging from vague definitions to varied interpretations—are deeply intertwined with geopolitical strategies. Nations leverage AWS in their political narratives, reflecting differing national interests and strategic applications of AI technologies. In this vein, Lingevicius [130] underscores a contradiction in the European Commission's approach: it excludes military AI from its forthcoming AI policies while supporting military and defense AI initiatives at the EU level. Consequently, he introduces the "Military Power Europe" definition, which incorporates four categories derived from the Europe-as-power debate. This geopolitical landscape complicates the formulation of global standards and regulations for AWS, affecting the efficacy and perceived impartiality of AI ethical declarations in international settings.

5. Discussion

A growing body of scholarship has examined AI ethical declarations to map common principles across documents [18,24,25], while a separate stream of research has explored the ethical challenges that arise within digital marketing and e-commerce environments. Yet these two arguments have largely remained in parallel. To our knowledge, no study has yet crossed this divide, systematically examining the weaknesses of global AI ethical declarations and asking what those weaknesses actually mean for the firms and platforms operating in commercial digital spaces. Bridging this gap is what the present study sets out to do.

5.1. Implications of AI Ethical Declarations for Digital Business and E-Commerce Firms

The findings of this study offer significant insights into how the opportunities and challenges of AI ethical declarations manifest within digital platforms and e-commerce

ecosystems. While these declarations tend to be written at a fairly abstract level, they are increasingly shaping how companies communicate their responsibilities around user data governance, algorithmic transparency, and automated decision-making.

On the opportunities side, global involvement in drafting AI ethics guidelines drives ethics adoption across platforms, as they are recognized universally, thereby instilling trust in consumers. Such an effort is vital in e-commerce, as trust is paramount for consumer engagement with platforms. The global alignment of ethics fosters compatibility and compliance among digital enterprises operating internationally, thereby guaranteeing stability in the governance landscape. Furthermore, these principles guide digital platforms to the critical ethical concerns raised by Mittelstadt et al. [19], thus enhancing algorithmic transparency—facilitating traceability.

However, the limitations identified in this study present substantial challenges for digital platforms attempting to translate high-level ethical commitments into operational practice. Table 4 introduces an explanatory framework that maps policy-level ethical gaps to real commercial frictions.

Table 4. The relationship between the limitations in AI ethical declarations and digital business (source: author’s elaboration).

Limitation in Ethical AI Declarations	How It Affects Digital Business (Mechanism of Influence)	Implications for E-Commerce Precisely
Lack of a standard framework	(1) Fragmented compliance costs as they must simultaneously navigate conflicting regional regulations; (2) complicated cross-border operations and platform governance; (3) competitive disadvantages when competitors exploit lax jurisdictions through regulatory arbitrage.	(1) Suffer massive inefficiencies as they must build and maintain separate AI systems for each market (EU version, US version, etc.); (2) disadvantage smaller platforms that cannot afford multi-jurisdictional compliance teams, thereby entrenching monopolies of large players like Amazon.
Divergence in definitions and terminology	(1) Ambiguity in key terms (e.g., “fairness”, “transparency”) makes it hard to align AI systems with compliance or customer expectations; (2) increase in the compliance cost as companies must design adaptive governance frameworks, not one-size-fits-all AI systems (for instance, fairness might be defined as equal opportunity in one company /country and demographic parity in another), which also pose a challenge to juridical department.	(1) E-commerce platforms face operational paralysis when identical AI features require contradictory implementations across markets. Suppose we consider a product recommendation system that aims to promote “fairness.” Under demographic parity, the system shows the same products to men and women; under equal opportunity, it shows products based on browsing history; and under national interest, it prioritizes domestic brands.
Challenges in translating principles into Practice	(1) Lack of implementation guidance, ethical KPIs. For instance, data scientists cannot translate “be fair” into code without concrete metrics. (2) invest heavily in trial-and-error experimentation that risks regulatory penalties; (3) adopt superficial compliance measures (ethics washing) that satisfy stated principles on paper while failing to address actual harms in practice.	(1) Ethical AI becomes a checkbox exercise with limited efficient/practical safeguards, especially in recommender systems or personalization engines. For instance, suppose the principle is “maintain transparent pricing”. It is unclear, explain to whom? (customers? regulators? researchers?); in what format? (technical documentation vs. plain language?); is it transparency, explainability, or interpretability principle?
Integration with legal frameworks	(1) Ethical declarations might not match actual laws, creating confusion between legal compliance and ethical responsibility. (2) AI Ethics declarations that ignore misinformation, disinformation, leave digital businesses legally and ethically unprotected; thus, platforms face reputational damage and regulatory backlash for spreading election misinformation or health disinformation.	(1) Businesses either over-invest in compliance or risk underperforming in new markets. (2) Silence on information disorder allows e-commerce platforms to weaponize AI for deception, like generating fake reviews at scale, creating synthetic “influencer” endorsements, producing misleading product comparisons, and using deepfake demonstrations.
Erosion of trust due to limited inclusion	(1) Ethical principles rarely involve input from SMEs, civil society, or diverse global stakeholders, making them seem top-down and biased.	(1) Platforms may struggle to build user trust or face boycotts over perceived bias or unfairness.
Conflicting objectives within declarations	(1) Some declarations promote both innovation and strict control, leaving firms uncertain how to balance speed vs. safety; (2) Wasted resources training employees on multiple conflicting versions of the same principle without clear guidance on which interpretation governs when conflicts arise.	(1) It provides perfect cover for exploitation where e-commerce platforms justify invasive behavioral tracking by claiming “personalization accuracy requires comprehensive data collection,” defend discriminatory dynamic pricing as “fairness unfortunately conflicts with revenue optimization,” rationalize opaque recommendation algorithms through “transparency risks competitive intelligence leakage or security”.
Geopolitical tensions	It forces digital businesses into incompatible regulatory trilemmas where operating globally requires simultaneously satisfying US innovation-first requirements, EU human-centric mandates, and Chinese sovereignty imperatives: (1) creating technical impossibilities where a single AI system cannot comply with all three jurisdictions,(2) forcing costly market-specific versions, (3) strategic market exits.	(1) It might mandate e-commerce platforms to perform geopolitical loyalty by adopting the ethics narrative of their chosen bloc. For instance, the US bans Chinese platforms like Temu and SHEIN, citing “surveillance threats,” while American platforms deploy equivalent tracking, the EU uses strict AI regulation to handicap US tech giants under the guise of “ethical superiority.”

These frictions are observable in documented real-world cases across digital platform contexts. Amazon’s algorithmic pricing systems have repeatedly demonstrated the tension between commercial efficiency and fairness principles [4]. The system’s dynamic pricing

engine has been shown to adjust prices based on user location and browsing behavior, raising concerns about discriminatory pricing that existing ethical declarations identify as a priority but provide no mechanism to prevent [105]. A comparable pattern is documented in Yee et al. [131], who audited Twitter’s English-language content moderation AI and found that it systematically assigned higher false-positive rates to content from marginalized communities, including Black communities, despite Twitter’s published ethical commitments to fairness and non-discrimination.

The large number of failures at the platform level extends beyond fairness violations to encompass other core AI ethical principles, most notably privacy. IBM’s 2024 Cost of a Data Breach Report reveals that 46% of data breaches involve personally identifiable information, with the global average breach cost reaching \$4.88 million [132]. These figures highlight the insufficiency of relying solely on declarative ethics in the implementation and design phases and underscore the need for enforcement mechanisms and a framework that includes additional dimensions.

5.2. Global vs. National Declarations in the Context of Digital Business

To compare ethical priorities across regions and globally, we built a benchmark matrix based on the results of the descriptive content analysis. This matrix identifies principles that appear in more than 50% of declarations within each region and compares them against the ethical principles found in MGA declarations that appear in more than 50%, to calculate a matching percentage, as illustrated in Table 5. This approach follows several studies that have employed frequency analysis in large-scale analyses of AI ethical guidelines, where principles appearing in more than 50% of documents were interpreted as dominant normative categories (e.g., [20,24,25,70,92]). Consistent with this tradition, and as a methodological extension of their results, the present benchmarking index is descriptive and comparative in nature, aimed at mapping patterns of the “dominant principles” and comprehending where national declarations converge with or diverge from global MGA frameworks.

Table 5. Benchmarking regional AI ethics declarations for digital governance: Alignment with MGA principles (source: author’s elaboration).

Benchmark: MGA		Ethical AI Principles with Frequency Above 50% in Different Regional Declarations				Matching Percentage
Ethical AI Principles with Frequency Above 50% in MGAs	Europe	America	Africa	Asia		
Societal well-being	1	1	1	1	100%	
Accountability	1	1	1	1	100%	
Environmental well-being	1	0	1	1	75%	
Transparency	1	0	1	1	75%	
Fairness and non-discrimination	1	1	1	1	100%	
Inclusion	0	0	1	1	50%	
Security	0	1	1	1	75%	
Autonomy	0	0	1	0	25%	
Safety	0	0	1	1	50%	
Privacy and data protection	1	1	1	1	100%	
AI Ethical Principles Mentioned in More Than 50% of Regional Declarations but Not in MGAs	Promote research and innovation	Economic development; Promote research and innovation; Information integrity	Promote research and innovation	Open data policy		

This matching percentage essentially measures how closely regional priorities align with global frameworks. The results reveal an interesting pattern: core principles such as societal well-being, fairness, and privacy enjoy universal adoption across all regions (100%), while others like accountability, transparency, and security show regional variation (75%).

Principles such as inclusion and autonomy, meanwhile, remain notably fragmented at the regional level (25%).

The benchmarking matrix developed in this study complements existing AI governance indices, such as the Responsible AI Index and national AI governance scoring frameworks, by introducing a principle-level normative diffusion metric. Whereas most global indices assess institutional readiness, regulatory maturity, or the presence of national strategies, the present benchmark captures the internal ethical architecture of those strategies. By identifying which principles surpass majority-level adoption within regions, the index adds analytical depth to governance comparisons and provides a diagnostic layer that can strengthen cross-national AI governance assessment tools.

5.2.1. Implications for Policymakers and International AI Governance

From a policy and business standpoint, the matrix has several implications for digital commerce, e-governance, and platform regulation:

First, the results reveal a strong cross-regional convergence around a core set of ethical principles, notably societal well-being, environmental sustainability, fairness and non-discrimination, and privacy and data protection. These principles appear consistently across all regions and align closely with the MGA benchmark, suggesting the emergence of a shared ethical baseline in global AI governance. For policymakers, this convergence provides a foundation for harmonizing AI governance approaches across jurisdictions, particularly in the context of cross-border digital trade and platform regulation.

Second, Table 5 reveals significant regional disparities in the prioritization of certain ethical principles. Europe demonstrates strong alignment with MGAs on transparency and sustainability but shows weaker emphasis on safety, security, and autonomy at the national declaration level. Asia and Africa, by contrast, display broader alignment with MGA principles related to security, safety, accountability, and inclusion, reflecting governance concerns associated with systemic risk management and societal protection. The Americas exhibit a mixed profile, aligning with core MGA principles while simultaneously emphasizing economic development-related considerations. These differences suggest that while ethical convergence exists at a high level, regional governance logics continue to shape how AI ethics are interpreted and institutionalized.

Third, the benchmark exposes a set of emerging priorities at the regional level, which MGA frameworks underrepresent. These include open data policy in Asia, economic development in the Americas, and the promotion of research and innovation in Africa and Europe. These patterns show that AI governance is not only top-down but also shaped by bottom-up normative experimentation. National and regional strategies reflect developmental, institutional, and informational needs. For international governance bodies, this observation is critical: global frameworks must remain flexible enough to integrate regionally relevant principles if they aim to foster inclusive and effective global governance.

From the point of view of digital business and e-commerce, this benchmark is useful because it serves as an ethical alignment map for platform governance. Digital firms operating across multiple regions face increasing pressure to comply not only with legal requirements but also with ethical expectations embedded in national AI strategies. The table makes visible where ethical expectations are broadly shared and where they diverge, enabling firms to anticipate governance risks, reputational challenges, and future regulatory tightening. In areas where MGA-aligned principles such as transparency, accountability, or security are weakly emphasized at the national level, digital platforms may need to adopt stronger internal governance mechanisms to maintain trust and legitimacy. Conversely, where regions emphasize principles not yet embedded in global frameworks, such as

information integrity or open data, businesses may encounter emerging ethical demands that are not yet standardized internationally.

A preliminary observation from the dataset suggests that declarations published following the EU AI Act [50] introduce principles not present in earlier documents, indicating an early pattern of norm evolution. For instance, Italy's national AI law (Law No. 132/2025) [133] explicitly incorporates principles of children's protection and deepfake regulation under criminal law, while the updated OECD AI Principles similarly reflect emerging concerns around information integrity and intellectual property protection—signaling a shift toward more specific and enforceable normative commitments that future longitudinal research should systematically track.

5.2.2. Implications for Digital Platforms Strategy

A closer examination of national-level declarations reveals additional nuances with profound consequences for digital business actors. Among the countries included in this study, only the United States, through its National AI R&D Strategic Plan (2025), explicitly lists the need to “combat synthetic media” as an ethical priority. This principle closely aligns with the broader commitment to information integrity across four major global AI frameworks: the OECD, the G7 Hiroshima Leaders' Statement, the Hamburg Declaration, and the BRICS AI Declaration. These multilateral declarations—especially the OECD's 2024 [134] update—directly address the rise in misinformation and disinformation as pressing ethical and economic challenges. In digital business contexts, the absence of robust ethical commitments to information integrity can leave platforms vulnerable to content manipulation, erosion of public trust, and disruptions to e-commerce operations, including fraud, reputational risk, and reduced user engagement.

Second, cultural and political value systems embedded in national declarations can shape the operational environment for digital platforms. For example, Indonesia anchors its ethical AI governance within the Pancasila philosophical framework, which includes principles such as belief in one God, just and civilized humanity, democratic consultation, and social justice for all. While these values are not standard in international AI ethics declarations, they frame how technology—including digital services and e-commerce—must operate in line with cultural and societal norms. Platforms operating in such contexts must integrate local ethical expectations into service delivery, moderation practices, and platform design to avoid ethical frictions and ensure compliance with evolving norms of digital responsibility.

Third, the U.S. Executive Order 14110 (“Safe, Secure & Trustworthy Development and Use of AI”), though recently revoked, is notable for being among the few national-level initiatives to explicitly emphasize consumer and worker protection as ethical principles. These principles are critical in digital commerce, where AI-driven automation, recommendation systems, and content personalization increasingly affect labor markets, consumer rights, and workplace fairness. Their absence from many national declarations and multilateral frameworks suggests a gap in translating socio-economic rights into enforceable ethical standards within digital ecosystems.

In short, digital platforms must do more than acknowledge AI ethics in the abstract; they need to understand how those ethics shift across regions and what happens in practice when they fall short. Issues such as information integrity, worker rights, and culturally embedded values may not yet be fully mainstreamed into global ethical benchmarks like the MGAs, but their growing prominence at the national level makes them *de facto* expectations in several markets. Firms that proactively engage with these emerging norms can gain a competitive advantage by fostering user trust, preempting regulatory backlash, and differentiating their platforms in ethically contested spaces.

5.3. Ethical Declarations for Responsible Platform Governance

The lack of institutional capacity to enforce these principles is referred to as “ethical inaction” [135]. The ethical declarations and frameworks analyzed in this study identify two critical dimensions of responsible platform governance that guide their enforcement.

5.3.1. Dimension 1: Assessment Mechanism

The ethical declarations and frameworks analyzed in this study identify two critical dimensions of responsible platform governance. The first dimension focuses on systematic *assessment and auditing mechanisms*. One such mechanism is the ethics impact assessment model. A critical example is UNESCO’s Ethical Impact Assessment (EIA) tool, which helped Germany reduce complaints of algorithmic discrimination in the public sector by 32%. The EIA provides a structured framework requiring platforms to implement transparent content-curation policies with formally institutionalized checks and balances accessible to all stakeholders, including marginalized groups. The second mechanism is algorithmic auditing, which provides a systematic evaluation of AI systems to detect bias and ensure compliance with ethical standards. Under the EU Digital Services Act, the European Commission’s Joint Research Centre identifies four audit study designs: risk-uncovering studies that identify potential algorithmic harms; reverse-engineering studies that reveal algorithmic decision-making parameters; interface design studies that assess how platform features amplify risks; and risk-measuring studies that quantify disproportionate exposure to harmful content. Through regular algorithmic audits, digital platforms can proactively identify and mitigate discriminatory patterns in content moderation and recommendation systems before they cause widespread harm.

5.3.2. Dimension 2: Governance Implementation Structures

First, this dimension involves establishing clear *governance architectures* from both national and industry perspectives to ensure responsible platform operation. The first type of governance architecture is law enforcement from the state’s perspective. In this system, public authorities are in charge of making sure that AI standards are followed and enforced. Italy’s AI law illustrates how ethical commitments can be paired with supervisory infrastructure by designating enforcement and oversight responsibilities to public authorities, including the Agency for Digital Italy and the National Cybersecurity Agency, reinforcing the organizational dimension of ethics-through-governance rather than merely aspirational principles. The second governance architecture is the establishment of independent ethics boards. For instance, Meta’s Oversight Board operates as an independent body with binding authority to review and overturn content moderation decisions, like ensuring freedom of expression, while issuing policy recommendations, demonstrating how ethics boards can shape platform policies.

Second, this dimension leverages *design requirements in practice*. For instance, this dimension facilitates digital inclusion for marginalized users. The OECD identifies several disability-centered AI-powered solutions that demonstrate how platforms can operationalize ethical commitments to non-discrimination and human dignity. Examples include RogerVoice, a French application that enables deaf users to read phone calls through automated speech-to-text captioning. Envision Glasses is a Dutch platform that uses image recognition to give blind people real-time descriptions of text, people, and their surroundings. These applications illustrate how AI systems, when designed with accessibility principles, can transform digital platforms from potential sources of exclusion into enablers of equal participation, demonstrating that ethical AI governance extends beyond harm mitigation to actively advancing social inclusion.

5.4. A Tiered Framework for Evaluating AI Ethics Enforcement

Building on the findings of this study, the pattern of high and low principle citations, the literature review findings on enforcement gaps, and the benchmarking matrix converge on a single diagnostic insight: AI ethical declarations differ fundamentally in their degree of enforceability. Accordingly, we propose a three-tiered framework for assessing the enforceability of AI ethical declarations.

Tier 1: Declarative ethics (toothless)

This tier encompasses declarations that articulate ethical principles without accompanying implementation mechanisms, monitoring systems, or enforcement consequences. These declarations typically list high-level values such as transparency, fairness, and accountability, but provide neither operational definitions nor guidance for translating them into organizational practice. The majority of the 54 declarations analyzed in this study fall within this tier, offering normative guidance that remains largely symbolic. While such declarations may serve important legitimization functions and foster public discourse, they lack the institutional infrastructure necessary to ensure compliance or measure impact. A clear example of this tier is the lower adoption of accountability principles in African declarations, reflecting the absence of independent regulatory bodies with AI mandates, which creates a structural barrier to procedural commitment independent of normative intent.

Observable indicators that a specific declaration is considered toothless include standalone ethical principles with no accompanying legislation, no designated oversight body, and no compliance reporting requirement; verification is limited to document analysis confirming the absence of any procedural commitment.

Tier 2: Procedural ethics (emerging teeth)

At this intermediate level, ethical declarations are supplemented with procedural mechanisms such as assessment tools, industry self-regulation initiatives, or voluntary compliance frameworks. The limited legal mandate, inconsistent enforcement, and reliance on organizational willingness to implement recommendations constrain the effectiveness of these accountability mechanisms. This tier represents a transitional phase in which ethics begin to influence practice but are not yet systematically enforced.

In sum, to label a specific declaration as an emerging teeth it should exhibit two characteristics: it introduces procedural governance mechanisms, and no law enforcement. Several practical examples fall into this tier: the EU's Ethics Guidelines for Trustworthy AI; UNESCO's AIA; the OECD AI Impact Assessment (AIIA); and Canada's Algorithmic Impact Assessment (AIA). To illustrate, the Canadian AIA provides a structured assessment tool comprising 51 risk and 34 mitigation questions. The responses are used to generate a score that determines the AI system's impact level and guides evaluation of system design, data sources, decision impacts, and risks to individuals. Each score requires specific governance measures such as peer review, public notice, or human oversight.

Tier 3: Institutionalized ethics (enforceable)

The most advanced tier integrates ethical principles with legally binding frameworks, mandatory assessment protocols, designated enforcement agencies, and measurable accountability mechanisms, such as mandatory incident reporting or financial sanctions. For instance, the EU AIA specifies fines of up to €35 million or 7% of global annual turnover for violations of prohibited AI practices (Article 5). Also, the act requires providers of high-risk AI systems to report serious incidents to national authorities within defined timeframes (Article 99).

The convergence of normative principles, institutional oversight, and regulatory power characterizes this tier. Evidence from this study demonstrates that such integration produces tangible outcomes: Germany's adoption of UNESCO's Ethical Impact Assess-

ment tool led to a 32% reduction in algorithmic discrimination complaints in public-sector applications, while the EU Digital Services Act mandates algorithmic auditing with legal consequences for non-compliance. At this tier, ethical declarations transition from aspirational guidelines to enforceable governance instruments embedded within regulatory ecosystems.

Within this interpretation, we emphasize that the transition from Tier 1 to Tier 3 is not merely theoretical. The trajectory of principles established in the OECD AI Principles [136] and the UNESCO Recommendation on AI Ethics [137], which articulated fairness, transparency, and accountability at the declarative and procedural levels, was codified as legal obligations under the EU AI Act [69]. Three conditions enabled this transition from Tier 1 to 3: first, political consensus across EU member states that created the legislative instrument; second, the prior existence of procedural frameworks and impact assessment methodologies providing the technical architecture for enforcement; and third, the designation of national market surveillance authorities with explicit sanctioning powers, transforming normative commitments into enforceable governance.

6. Conclusions

This study offers three contributions to the discourse on AI governance, particularly ethics of AI, and responsible platform governance. First, it provides an empirical analysis of how 54 AI ethics declarations diverge in their articulation of core ethical principles. Second, it develops a benchmarking index that compares ethical principles adopted in national declarations against those endorsed by MGAs. The matching percentages generated by this index provide policymakers, governance bodies, and platform operators with a structured diagnostic tool: they can identify where cross-regional convergence is strong enough to support harmonized regulation and where local priorities diverge, requiring tailored governance approaches. For digital firms operating across multiple jurisdictions, the index also serves as an ethical alignment map, enabling proactive compliance strategies and anticipating regulatory trajectories. Third, the three-tiered framework for evaluating AI ethics enforceability, distinguishing declarative, procedural, and institutionalized ethics, constitutes another original contribution of this study. The framework moves beyond prior comparative studies that mapped principle frequency without explaining governance outcomes. Accordingly, this study bridges three research areas: normative AI ethics, which asks what principles matter and why; platform governance, which examines how digital firms navigate ethical and regulatory obligations in algorithmic environments; and regulatory governance, which asks under what institutional conditions ethical commitments become enforceable.

Like all research, this study has limitations that should be acknowledged. The literature search was conducted across two databases; future studies would benefit from expanding coverage to additional databases, such as Scopus. Additionally, the empirical analysis is based on 54 declarations. Future studies could expand the corpus to include additional national declarations, particularly from underrepresented regions, to test the generalizability of the proposed framework.

Future research should utilize methodologies that are based on surveys or interviews and target AI governance practitioners, product managers, legal compliance teams, and ethics officers in e-commerce and digital platform firms. Such research could explore which declarations, frameworks, or principles organizations actually use in their processes, what challenges prevent these from being put into practice, and how companies manage the balance between ethical values like fairness and business goals like personalization, efficiency, and revenue growth.

A helpful next step would be to study specific platforms like Amazon, Meta, or Alibaba, as case studies, to see how they deal with the gap between their public ethical statements and how their algorithms actually operate. Researchers could use UNESCO’s Ethical Impact Assessment to check whether these companies are following their own ethical rules, linking the macro analysis from this study with the micro-level real challenges faced at the business level to assess different ways in which ethics work in practice.

Lastly, and most significantly, future research should investigate the conditions, including specific tools and practices at both macro and micro levels, under which ethical declarations successfully transition from Tier 1 to Tier 3 in the enforceability framework proposed in this study.

Funding: The APC was funded by Good in Tech research chair.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data analyzed in this article are available in Appendix A.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A. AI Declarations Used in This Study

Table A1. Overview of AI Ethical Declarations Across the *American* Continent.

Declaration/Policy Name	Year Issued/Latest Update	Country	AI Principles Mentioned
America’s AI Action Plan/National AI R&D Strategic Plan	2025 (draft, published 10 July 2025)	United States	Deregulation; Promoting open source; Re-/upskilling; Promote AI research; Interpretability; Robustness; Control; Promote AI in government; Promote commercial AI innovation; Combat synthetic media; Security
Executive Order 14110—“Safe, Secure & Trustworthy Development and Use of AI” (revoked)	2023	United States	Safety; Transparency; Accountability; Human oversight; Risk assessment; Interagency coordination; Safety & Security; Privacy Protection; Prevent discrimination and unlawful bias; Consumer & Worker Protection; Detecting synthetic content
Executive Order 14179—“Removing Barriers to American Leadership in AI”	2025	United States	Human flourishing; Economic competitiveness; National security; Freedom from bias; National Security; Public trust; Accountability; Risk mitigation; Fairness; Transparency; The need to protect civil rights and privacy
Montreal Declaration for a Responsible Development of Artificial Intelligence	2018	Canada	Promote well-being; Autonomy; Justice; Privacy; Knowledge; Democracy; Responsibility; Caution principle; Sustainable Development; Solidarity
Plano Brasileiro de Inteligência Artificial (PBlA)	2024–2028	Brazil	Inclusive growth & well-being; Human-centric & equitable values; Transparency & explainability; Robustness & security; Accountability (aligned with OECD AI Principles)
National AI Policy (Chile)	2021	Chile	AI for well-being; Respect for human rights & security; Sustainable development; Inclusiveness; Global cooperation
Chile—Draft AI Regulation Bill (Boletín 16 821-19)	2024 (Reviewed in 2025)	Chile	Transparency; Promote innovation; Address the issue of deepfakes; Individual dignity, democratic integrity, or public security; Responsibility
Iniciativa con proyecto de decreto por el que se reforma la fracción XVII del artículo 73 de la Constitución Política de los Estados Unidos Mexicanos, en materia de inteligencia artificial.	2025	Mexico	Personal data protection; Responsible AI use; Fostering responsible innovation: protecting human rights, privacy; National security; Address societal and economic challenges
2019 Plan Nacional de Inteligencia Artificial	2019	Argentina	Promote social and economic development; Respect human rights; Minimize social risks; Personal data and privacy; Ethics by design

Table A2. Overview of AI Ethical Declarations Across the *Africa* Continent.

Declaration/Policy	Year Issued/Latest Update	Country/Region	Security
African Union Continental AI Strategy	2024	African Union (Continental)	Risk mitigation; Data protection; Cybersecurity (Safety); Consumer protection; Inclusion; Reviewing labor protections; Aligning social media regulations; Non-discrimination; Job creation, health, and education
Kenya’s National AI Strategy 2025–2030	2025	Kenya	Inclusivity; Non-discrimination; Transparency; Accountability; Explainability; Human-oversight; Auditing, Cultural preservation and contextualization; Environmental sustainability; Achieve national priorities; Enhance public services; Promote inclusive economic growth
African Declaration on Artificial Intelligence	2025	African Union + 49 nations	Sovereignty; Inclusivity; Diversity; Privacy, ethics, transparency, and explainability while prioritizing human dignity, rights, freedoms, and environmental sustainability; Regional and international collaboration; Re-/upskilling; open data; Promote innovation
Ghana’s National Artificial Intelligence Strategy (2023–2033)	2024	Ghana	Transparency; Security; Robustness; Human rights; Inclusiveness; Re-/upskilling; Facilitate data access and governance; Collaboration; Promote AI research; Promote AI adoption in the public sector
Rwanda’s National AI Strategy	2025	Rwanda	Transparency; Privacy; Fairness; Accountability; Human oversight; Fairness & Non-Discrimination; Safety & Robustness; Accountability; Societal & Environmental Well-being
Republic of Ghana National Artificial Intelligence Strategy 2023–2033	Ongoing (2024 draft)	Nigeria	Transparency; Fairness; Accountability; Respect for human rights; Inclusivity; Non-discrimination; Innovation and adaptation; Long-term economic, social, and environmental goals; Risk management and resilience; Human-centric; Global leadership
SA National AI Policy Framework	2024	South Africa	Re-/upskilling; Safety; Sectoral oriented; Transparency; Explainability; Data protection; Promote research; Human oversight
Arab Republic of Egypt (2025).	2025	Egypt/Senegal/Tunisia	Fairness; Transparency; Accountability; Privacy protection; Human rights; Social justice; Safety; Non-discrimination

Table A3. Overview of AI Ethical Declarations Across the *Asia* Continent.

Declaration Name	Year Issued/Latest Update)	Country	AI Principles Mentioned
Beijing AI Principles	2019 (reaffirmed in regulatory drafts 2023–2024)	China	AI R&D should serve human interests; Fairness, reduced discrimination/bias; Transparency; Explainability; Predictability; Traceability; Auditability; Accountability; Diversity; Inclusion
Social Principles of Human-Centric AI	2019 (reinforced via AI Promotion Act 2025)	Japan	Human-centricity; Fairness; Transparency; Accountability; Privacy; Innovation; Fair competition; Safety & security; Education & literacy
India AI Governance Guidelines (Sutras)	2025	India	Trust is the foundation; Do no harm; Innovation over restraint; Fairness & equity; Accountability; Understandable by Design; Safety, Resilience & sustainability
Model AI Governance Framework	2019 (updated 2020)	Singapore	Transparency; Explainability; Repeatability / reproducibility; Safety; Security; Robustness; Fairness; Data governance; Accountability; Human agency and oversight; Inclusive growth; Societal and environmental well-being
Stranas KA—National AI Strategy (2020–2045)	2020	Indonesia	Pancasila values ((1) Belief in the one true God, (2) A fair-minded and civilized humanity, (3) Unity of Indonesia, (4) Democracy (from the people) led by wisdom of consultation (of the) representatives (of the people), and (5) Social justice for every person in Indonesia); OECD principles
UAE National AI Strategy (2031)	2017 (vision extended to 2031)	United Arab Emirates	Reflect national values; Privacy & data protection; Sustainability; Human-centered development; Safety; Fairness; Transparency; Explainability; Accountability; Peaceful coexistence with AI; Inclusive access and equity; Legal compliance; societal benefit
Saudi Arabia’s AI Ethics Principles 2.0	2023	Saudi Arabia	Fairness; Privacy & security; Humanity; Social & environmental benefits; Reliability & safety; Transparency & explainability; Accountability & responsibility
AI Seoul Summit Declaration	2024	South Korea	Safety; Innovation; Inclusivity; International cooperation; Human-centered governance

Table A4. Overview of AI Ethical Declarations Across the *European* Continent.

Declaration Name	Year Issued (Latest Update)	Country	AI Principles Mentioned
Stratégie nationale pour l’intelligence artificielle (SNIA)	2018 (updated acceleration phase 2023–2025)	France	Transparency; Fairness; Security; Accountability; Societal & environmental well-being; Open data policy; Promote research/innovation; European coordination; Respect labor right; Privacy & data protection
Germany’s National AI Strategy	2018 (updated 2020 & 2023)	Germany	Transparency; Traceability; Safety; Inclusion; Security; Robustness; Sustainability (society & environment); Non-discrimination; AI skills for all; Promote research/innovation

Table A4. *Cont.*

Declaration Name	Year Issued (Latest Update)	Country	AI Principles Mentioned
Italy’s national AI law (Law No. 132/2025)	2025 (law passed)	Italy	Human-centric (enhance human decisions); Children’s protection; Workplace transparency; Deepfakes & criminal law; Protection of intellectual property rights; Privacy & data protection; Knowability; Traceability; Respect labor right; Transparency; accountability
Spanish AI regulatory framework (AESIA)	2023 (agency); 2025 bills	Spain	Transparency; Promote innovation; Non-discrimination; Privacy & data protection; Protection of intellectual property rights; Fairness & non-discrimination; Fair Competition
Sweden National AI Strategy	2018 (updated 2026)	Sweden	Fair competition; AI ecosystem that has a positive global impact on people and the planet; Democracy, Inclusion; Diversity; Re-/upskilling; Transparency and availability of data; Accountability; Collaboration.
Policy for the Development of Artificial Intelligence in Poland 2025–2030 (and	2024	Poland	Re-/upskilling; Promote innovation; Fair competition; Sovereignty; Transparency; Accountability; Robustness; Safety; Environmental and societal well-being;
Recommendations on AI in education, teaching and training/Digivisio 2030 program	2024	Finland	Fair competition; Provide best public services; Enhance employment; Investment in research; Promote green economic growth; Interoperability; Privacy; Security; Good usability; Portability
UK National AI Strategy (National Data Strategy (2020), a Plan for Digital Regulation (2021), and the UK Innovation Strategy (2021)	2024–2025	United Kingdom	Public participation; Safety; Privacy & data protection; Transparency; Promote research; Risk management; Accountability; Public trust & transparency; Fairness; Beneficence; Collaboration

Table A5. Overview of AI Ethical Declarations Across the *Major Global Actors (MGAs)*.

Declaration Name	Year Issued	Latest Update	Ethical Principles Mentioned
OECD AI Principles (OECD Recommendation on AI)	2019	2024	Inclusive growth; Well-being; Sustainable development; Environmental sustainability; Non-discrimination; Equality; Human dignity; Autonomy; Privacy & data protection; Diversity; Fairness and non-discrimination; Social justice; Labor rights; Misinformation & disinformation; Freedom of expression; Robustness; Security; Safety; Risk management; Accountability; Transparency; Explainability; International collaboration
EU Ethics Guidelines for Trustworthy AI	2019	–	Human agency and oversight; Technical robustness and safety; Privacy and data governance, Transparency, diversity, non-discrimination, and fairness; Societal and environmental well-being; Accountability
UNESCO Recommendation on the Ethics of AI	2021	–	Do no harm; Human agency and oversight; Safety; Security; Fairness and non-discrimination; Privacy & data protection; Transparency; Explainability; Accountability; Inclusion; Well-being; Sustainable development; Environmental sustainability; Economic growth; International collaboration
G7 Hiroshima AI Process Leaders’ Statement	2023	2025	Risk management; Post-deployment monitoring; Transparency; Responsible information sharing; Security; Information integrity; Well-being; Privacy & data protection; Interoperability; International collaboration; Intellectual property protection
Statement on Inclusive and Sustainable AI for People and the Planet (Paris AI Action Summit)	2025	–	Human rights; Human dignity; Autonomy; Privacy & data protection; Fairness and non-discrimination; Social justice; AI accessibility; Sustainable for people and the planet; Safety; Reliability; Security; Accountability; Transparency; Explainability; Inclusion; Labor rights; Protection of consumers; Protection of intellectual property rights
Framework Convention on Artificial Intelligence	2022	2024	Human dignity; Autonomy, Equality & non-discrimination; Privacy and data protection; Transparency; Accountability and responsibility; Reliability; Safe innovation
BRICS AI Governance Declaration	2025	–	Inclusion; Human supervision; Transparency; Accountability; Maximizing societal benefits; Information integrity; Privacy; Safety; Security; Sustainable development; Environmental sustainability; AI must ensure decent work; Promote innovation; Fair competition; Digital sovereignty; International collaboration
Hamburg Declaration on Responsible AI for the SDGs	2025	–	Equality; Inclusive participation; Resource-efficient and climate-friendly; Economic growth; Combat disinformation; Security; International collaboration
Malta Declaration on the Use of AI	2025	–	Primacy and sanctity of human life; Security; Safety; International collaboration; Privacy & data protection; Promote Sustainability; Promote innovation; Respect for the environment; Promote innovation

References

- Lazazzara, A.; Za, S.; Georgiadou, A. A taxonomy framework and process model to explore AI-enabled workplace inclusion. *J. Bus. Res.* **2025**, *201*, 115697. [CrossRef]
- Hassan, R.; Nguyen, N.; Finserås, S.R.; Adde, L.; Strümke, I.; Støen, R. Unlocking the black box: Enhancing human-AI collaboration in high-stakes healthcare scenarios through explainable AI. *Technol. Forecast. Soc. Change* **2025**, *219*, 124265. [CrossRef]
- Grand View Research. Artificial Intelligence Market Size, Share & Trends Analysis Report by Solution (Hardware, Software, Services), by Technology (Deep Learning, Machine Learning, NLP), by Function, by End-Use, By Region, and Segment Forecasts, 2023–2030 (Report No. GVR-1-68038-955-5). 2023. Available online: <https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market> (accessed on 8 October 2025).

4. Van Loo, R.; Aggarwal, N. Amazon's Pricing Paradox. *Harv. J. Tech.* **2023**, *37*, 1.
5. Bian, Z.; Che, C. How AI Overview of Customer Reviews Influences Consumer Perceptions in E-Commerce? *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 315. [[CrossRef](#)]
6. Rezaei, M.; Pironti, M.; Quaglia, R. AI in knowledge sharing, which ethical challenges are raised in decision-making processes for organisations? *Manag. Decis.* **2025**, *63*, 3369–3388. [[CrossRef](#)]
7. Teng, Z.; Xia, H.; He, Y. Algorithmic Fairness and Digital Financial Stress: Evidence from AI-Driven E-Commerce Platforms in OECD Economies. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 213. [[CrossRef](#)]
8. Chaudhary, S.; Khalil, A.; Attri, R.; Ractham, P. Deploying explainable AI in entrepreneurial organizations: Role of the human-AI interface. *Technol. Forecast. Soc. Change* **2025**, *220*, 124324. [[CrossRef](#)]
9. Kan, M. Hacker Deepfakes Employee's Voice in Phone Call to Breach IT Company. *PCMag*. 2023. Available online: <https://www.pcmag.com/news/hacker-deepfakes-employees-voice-in-phone-call-to-breach-it-company> (accessed on 8 October 2025).
10. Wang, S.; Peng, K.L.; Huang, Z.; Ma, L. AI-generated videos: Influencing trustworthiness, awe, and behavioral intention in space tourism e-commerce. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 307. [[CrossRef](#)]
11. Kazim, E.; Koshiyama, A.S. A high-level overview of AI ethics. *Patterns* **2021**, *2*, 100314. [[CrossRef](#)]
12. Hanisch, M.; Goldsby, C.M.; Fabian, N.E.; Oehmichen, J. Digital governance: A conceptual framework and research agenda. *J. Bus. Res.* **2023**, *162*, 113777. [[CrossRef](#)]
13. Laviola, F.; Cucari, N. From promise to concern: Public perceptions of AI in ESG frameworks over time. *Technol. Soc.* **2026**, *85*, 103219. [[CrossRef](#)]
14. Berente, N.; Gu, B.; Recker, J.; Santhanam, R. Managing artificial intelligence. *MIS Q.* **2021**, *45*, 1433–1450. [[CrossRef](#)]
15. Yu, T.; Pan, Y.; Jang, W. Modeling Consumer Reactions to AI-Generated Content on E-Commerce Platforms: A Trust–Risk Dual Pathway Framework with Ethical and Platform Responsibility Moderators. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 257. [[CrossRef](#)]
16. Floridi, L.; Cows, J. A unified framework of five principles for AI in society. In *Machine Learning and the City: Applications in Architecture and Urban Design*; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2022; pp. 535–545. [[CrossRef](#)]
17. Floridi, L.; Cows, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds Mach.* **2018**, *28*, 689–707. [[CrossRef](#)] [[PubMed](#)]
18. Jobin, A.; Ienca, M.; Vayena, E. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* **2019**, *1*, 389–399. [[CrossRef](#)]
19. Mittelstadt, B. Principles alone cannot guarantee ethical AI. *Nat. Mach. Intell.* **2019**, *1*, 501–507. [[CrossRef](#)]
20. Morley, J.; Floridi, L.; Kinsey, L.; Elhalal, A. From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Sci. Eng. Ethics* **2020**, *26*, 2141–2168. [[CrossRef](#)]
21. Munn, L. The uselessness of AI ethics. *AI Ethics* **2023**, *3*, 869–877. [[CrossRef](#)]
22. Hong, S.; Ryee, H.; Jin, X.; Yang, D. How Organizations Choose Open-Source Generative AI Under Normative Uncertainty: The Moderating Role of Exploitative and Exploratory Behaviors. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 250. [[CrossRef](#)]
23. Resselguier, A.; Rodrigues, R. AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data Soc.* **2020**, *7*, 2053951720942541. [[CrossRef](#)]
24. Corrêa, N.K.; Galvão, C.; Santos, J.W.; Del Pino, C.; Pinto, E.P.; Barbosa, C.; Massmann, D.; Mambrini, R.; Galvão, L.; Terem, E.; et al. Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns* **2023**, *4*, 100857. [[CrossRef](#)]
25. Hagendorff, T. The ethics of AI ethics: An evaluation of guidelines. *Minds Mach.* **2020**, *30*, 99–120. [[CrossRef](#)]
26. Adeusi, S.O. Roles and Importance of Ethics. In *Research Anthology on Rehabilitation Practices and Therapy*; IGI Global: Hershey, PA, USA, 2020; pp. 1–11.
27. Fisher, A. *Meta-Ethics: An Introduction*; Routledge: Abingdon, UK, 2014.
28. Kidder, R.M. *How Good People Make Tough Choices: Resolving the Dilemmas of Everyday Living*; Harper Perennial: New York, NY, USA, 2003.
29. Paul, R.; Elder, L. *The Thinker's Guide to Understanding the Foundations of Ethical Reasoning: Based on Critical Thinking Concepts & Tools*; Foundation for Critical Thinking: Dillon Beach, CA, USA, 2006.
30. Proctor, J.D. Ethics in geography: Giving moral form to the geographical imagination. *Area* **1998**, *30*, 8–18. [[CrossRef](#)]
31. Martin, K.E.; Freeman, R.E. The separation of technology and ethics in business ethics. *J. Bus. Ethics* **2004**, *53*, 353–364. [[CrossRef](#)]
32. Kaplan, A.; Haenlein, M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Bus. Horiz.* **2019**, *62*, 15–25. [[CrossRef](#)]
33. Theodorou, A.; Dignum, V. Towards ethical and socio-legal governance in AI. *Nat. Mach. Intell.* **2020**, *2*, 10–12. [[CrossRef](#)]
34. Siau, K.; Wang, W. Artificial intelligence (AI) ethics: Ethics of AI and ethical AI. *J. Database Manag. (JDM)* **2020**, *31*, 74–87. [[CrossRef](#)]

35. Gorr, M. Is moral status done with words? *Ethics Inf. Technol.* **2024**, *26*, 10. [CrossRef]
36. Jaquet, F. Utilitarianism for the Error Theorist. *J. Ethics* **2021**, *25*, 39–55. [CrossRef]
37. Mill, J.S. Utilitarianism. In *Seven Masterpieces of Philosophy*; Routledge: Abingdon, UK, 2016; pp. 329–375.
38. Micewski, E.R.; Troy, C. Business ethics—deontologically revisited. *J. Bus. Ethics* **2007**, *72*, 17–25. [CrossRef]
39. Albrecht, A. Ethics and technology design. *Ethics Inf. Technol.* **2007**, *9*, 63–72. [CrossRef]
40. Friedman, B.; Kahn, P.; Borning, A. Value sensitive design: Theory and methods. In *University of Washington Computer Science & Engineering Technical Report*; No. 02-12-01; University of Washington: Washington, DC, USA, 2002. Available online: <https://dada.cs.washington.edu/research/tr/2002/12/UW-CSE-02-12-01.pdf> (accessed on 26 October 2025).
41. Trianosky, G. What is virtue ethics all about? *Am. Philos. Q.* **1990**, *27*, 335–344. Available online: <http://www.jstor.org/stable/20014344> (accessed on 26 October 2025).
42. Winston, C. Norm structure, diffusion, and evolution: A conceptual approach. *Eur. J. Int. Relat.* **2018**, *24*, 638–661. [CrossRef]
43. Ibrahim, I.A.; Zaidan, E.; Truby, J.; Hoppe, T. The AI Act and its green blind spots: Hidden environmental risks in the AI lifecycle. *Technol. Soc.* **2026**, *86*, 103284. [CrossRef]
44. Franzke, A.S. An exploratory qualitative analysis of AI ethics guidelines. *J. Inf. Commun. Ethics Soc.* **2022**, *20*, 401–423. [CrossRef]
45. Snyder, H. Literature review as a research methodology: An overview and guidelines. *J. Bus. Res.* **2019**, *104*, 333–339. [CrossRef]
46. Torraco, R.J. Writing integrative literature reviews: Guidelines and examples. *Hum. Resour. Dev. Rev.* **2005**, *4*, 356–367. [CrossRef]
47. Dara, R.; Hazrati Fard, S.M.; Kaur, J. Recommendations for ethical and responsible use of artificial intelligence in digital agriculture. *Front. Artif. Intell.* **2022**, *5*, 884192. [CrossRef]
48. Elsbach, K.D.; van Knippenberg, D. Creating high-impact literature reviews: An argument for ‘integrative reviews’. *J. Manag. Stud.* **2020**, *57*, 1277–1289. [CrossRef]
49. Alcayaga, A.; Wiener, M.; Hansen, E.G. Towards a framework of smart-circular systems: An integrative literature review. *J. Clean. Prod.* **2019**, *221*, 622–634. [CrossRef]
50. Falsafi, A.; Togiani, A.; Colley, A.; Varis, J.; Horttanainen, M. Life cycle assessment in circular design process: A systematic literature review. *J. Clean. Prod.* **2025**, *521*, 146188. [CrossRef]
51. Dabić, M.; Vlačić, B.; Kiessling, T.; Caputo, A.; Pellegrini, M. Serial entrepreneurs: A review of literature and guidance for future research. *J. Small Bus. Manag.* **2023**, *61*, 1107–1142. [CrossRef]
52. Martín-Martín, A.; Thelwall, M.; Orduna-Malea, E.; Delgado López-Cózar, E. Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations’ COCI: A multidisciplinary comparison of coverage via citations. *Scientometrics* **2020**, *126*, 871–906. [CrossRef] [PubMed]
53. Jensenius, F.R.; Htun, M.; Samuels, D.J.; Singer, D.A.; Lawrence, A.; Chwe, M. The benefits and pitfalls of Google Scholar. *PS Political Sci. Politics* **2018**, *51*, 820–824. [CrossRef]
54. Rowe, F. What literature review is not: Diversity, boundaries and recommendations. *Eur. J. Inf. Syst.* **2014**, *23*, 241–255. [CrossRef]
55. Gusenbauer, M.; Haddaway, N.R. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Res. Synth. Methods* **2020**, *11*, 181–217. [CrossRef]
56. Rethlefsen, M.L.; Page, M.J. PRISMA 2020 and PRISMA-S: Common questions on tracking records and the flow diagram. *J. Med. Libr. Assoc.* **2022**, *110*, 253–257. [CrossRef]
57. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, 71. [CrossRef]
58. Kraus, S.; Breier, M.; Lim, W.M.; Dabić, M.; Kumar, S.; Kanbach, D.; Mukherjee, D.; Corvello, V.; Piñeiro-Chousa, J.; Liguori, E.; et al. Literature reviews as independent studies: Guidelines for academic practice. *Rev. Manag. Sci.* **2022**, *16*, 2577–2595. [CrossRef]
59. Singh, V.K.; Singh, P.; Karmakar, M.; Leta, J.; Mayr, P. The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics* **2021**, *126*, 5113–5142. [CrossRef]
60. Mazrekaj, L. Gender Entrepreneurial Behaviour: A SSLR (Semi-Systematic Literature Review) Approach. *South East Eur. J. Econ. Bus.* **2024**, *19*, 77–95. [CrossRef]
61. Tranfield, D.; Denyer, D.; Smart, P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *Br. J. Manag.* **2003**, *14*, 207–222. [CrossRef]
62. Ulnicane, I. Artificial Intelligence in the European Union: Policy, ethics and regulation. In *The Routledge Handbook of European Integrations*; Taylor & Francis: Abingdon, UK, 2022.
63. Phadermrod, B.; Crowder, R.M.; Wills, G.B. Importance-performance analysis based SWOT analysis. *Int. J. Inf. Manag.* **2019**, *44*, 194–203. [CrossRef]
64. Chan, A.; Okolo, C.T.; Terner, Z.; Wang, A. The limits of global inclusion in AI development. *arXiv* **2021**. [CrossRef]
65. The White House. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. 2023. Available online: <https://www.congress.gov/crs-product/R47843> (accessed on 15 January 2026).

66. Greenleaf, G. G20 makes declaration of 'data free flow with trust': Support and dissent. *Priv. Laws Bus. Int. Rep.* **2019**, *160*, 18–19.
67. Fukuda-Parr, S.; Gibbons, E. Emerging consensus on 'ethical AI': Human rights critique of stakeholder guidelines. *Glob. Policy* **2021**, *12*, 32–44. [CrossRef]
68. van Buitenlandse Zaken, M. Global Declaration on Information Integrity Online-Diplomatic Statement-Government. Government of Netherland. 2023. Available online: <https://www.government.nl/documents/diplomatic-statements/2023/09/20/global-declaration-on-information-integrity-online> (accessed on 15 January 2026).
69. European Parliament & Council of the European Union. *Regulation (EU) 2024/1689 of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*; Official Journal of the European Union: Luxembourg, 2024. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689> (accessed on 10 January 2026).
70. Fjeld, J.; Achten, N.; Hilligoss, H.; Nagy, A.; Srikumar, M. Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. In *Berkman Klein Center Research Publication*; Harvard University: Cambridge, MA, USA, 2020.
71. Schiff, D.; Borenstein, J.; Biddle, J.; Laas, K. AI ethics in the public, private, and NGO sectors: A review of a global document collection. *IEEE Trans. Technol. Soc.* **2021**, *2*, 31–42. [CrossRef]
72. Schiff, D.; Biddle, J.; Borenstein, J.; Laas, K. What's next for AI ethics, policy, and governance? A global overview. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES '20), New York, NY, USA, 7–8 February 2020*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 153–158.
73. Auld, G.; Casovan, A.; Clarke, A.; Faveri, B. Governing AI through ethical standards: Learning from the experiences of other private governance initiatives. *J. Eur. Public Policy* **2022**, *29*, 1822–1844. [CrossRef]
74. Smuha, N.A. The EU approach to ethics guidelines for trustworthy artificial intelligence. *Comput. Law Rev. Int.* **2019**, *20*, 97–106. [CrossRef]
75. Rothenberger, L.; Fabian, B.; Arunov, E. Relevance of ethical guidelines for artificial intelligence—A survey and evaluation. In *Proceedings of the 27th European Conference on Information Systems (ECIS 2019), Stockholm-Uppsala, Sweden, 8–14 June 2019*; Association for Information Systems: Atlanta, GA, USA, 2019.
76. Díaz-Rodríguez, N.; Del Ser, J.; Coeckelbergh, M.; de Prado, M.L.; Herrera-Viedma, E.; Herrera, F. Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Inf. Fusion* **2023**, *99*, 101896. [CrossRef]
77. Cappelli, M.A.; Di Marzo Serugendo, G. A semi-automated software model to support AI ethics compliance assessment of an AI system guided by ethical principles of AI. *AI Ethics* **2025**, *5*, 1357–1380. [CrossRef]
78. Bostrom, N.; Yudkowsky, E. The ethics of artificial intelligence. In *Artificial Intelligence Safety and Security*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2018; pp. 57–69.
79. Stahl, B.C. *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies*; Springer Nature: Cham, Switzerland, 2021; p. 124.
80. Dressel, J.; Farid, H. The accuracy, fairness, and limits of predicting recidivism. *Sci. Adv.* **2018**, *4*, eaao5580. [CrossRef]
81. Selbst, A.D. Disparate impact in big data policing. *Ga. Law Rev.* **2017**, *52*, 109–196. [CrossRef]
82. Polykalas, S.E.; Prezerakos, G.N. When the mobile app is free, the product is your personal data. *Digit. Policy Regul. Gov.* **2019**, *21*, 89–101. [CrossRef]
83. Hevelke, A.; Nida-Rümelin, J. Responsibility for crashes of autonomous vehicles: An ethical analysis. *Sci. Eng. Ethics* **2015**, *21*, 619–630. [CrossRef] [PubMed]
84. EU Commission. *White Paper on Artificial Intelligence—A European Approach to Excellence and Trust*; COM(2020) 65 Final; European Commission: Brussels, Belgium, 2020.
85. High-Level Expert Group on Artificial Intelligence. *Ethics Guidelines for Trustworthy AI*; European Commission: Brussels, Belgium, 2019. Available online: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (accessed on 1 March 2026).
86. Farouk, M. Studying Human Robot Interaction and Its Characteristics. *Int. J. Comput. Inf. Manuf. (IJCIM)* **2022**, *2*, 38–49. [CrossRef]
87. Laitinen, A.; Sahlgren, O. AI systems and respect for human autonomy. *Front. Artif. Intell.* **2021**, *4*, 151. [CrossRef] [PubMed]
88. Rafanelli, L.M. Justice, injustice, and artificial intelligence: Lessons from political theory and philosophy. *Big Data Soc.* **2022**, *9*, 20539517221080676. [CrossRef]
89. de Fine Licht, K.; de Fine Licht, J. Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy. *AI Soc.* **2020**, *35*, 917–926. [CrossRef]
90. Namugenyi, C.; Nimmagadda, S.L.; Reiners, T. Design of a SWOT analysis model and its evaluation in diverse digital business ecosystem contexts. *Procedia Comput. Sci.* **2019**, *159*, 1145–1154. [CrossRef]

91. Tomašev, N.; Cornebise, J.; Hutter, F.; Mohamed, S.; Picciariello, A.; Connelly, B.; Clopath, C. AI for social good: Unlocking the opportunity for positive impact. *Nat. Commun.* **2020**, *11*, 2468. [CrossRef] [PubMed]
92. Royakkers, L.; Timmer, J.; Kool, L.; Van Est, R. Societal and ethical issues of digitization. *Ethics Inf. Technol.* **2018**, *20*, 127–142. [CrossRef]
93. Whittlestone, J.; Nyrup, R.; Alexandrova, A.; Cave, S. The role and limits of principles in AI ethics: Towards a focus on tensions. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, Honolulu, HI, USA, 27–28 January 2019; pp. 195–200.
94. Université de Montréal. *Montréal Declaration for a Responsible Development of Artificial Intelligence*; Université de Montréal: Montréal, QC, Canada, 2018. Available online: <https://montrealdeclaration-responsibleai.com/the-declaration/> (accessed on 20 September 2025).
95. Department of Industry; Science; Energy and Resources; Australian Government. *Australia's Artificial Intelligence Ethics Principles*; Australian Government: Canberra, Australia, 2019. Available online: <https://www.industry.gov.au/publications/australias-ai-ethics-principles> (accessed on 20 September 2025).
96. Binns, R. Fairness in machine learning: Lessons from political philosophy. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency, FAT '18, New York, NY, USA, 23–24 February 2018; Volume 81, pp. 149–159.
97. Kelley, S. Employee perceptions of the effective adoption of AI principles. *J. Bus. Ethics* **2022**, *178*, 871–893. [CrossRef] [PubMed]
98. Toreini, E.; Aitken, M.; Coopamootoo, K.; Elliott, K.; Zelaya, C.G.; Van Moorsel, A. The relationship between trust in AI and trustworthy machine learning technologies. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 27–30 January 2020; pp. 272–283.
99. Cools, H. Navigating the responsible AI landscape: Unraveling the principles-to-practices gap of transparency and explainability at the BBC. *Inf. Commun. Soc.* **2025**, 1–21. [CrossRef]
100. Ibáñez, J.C.; Olmeda, M.V. Operationalising AI ethics: How are companies bridging the gap between practice and principles? An exploratory study. *AI Soc.* **2022**, *37*, 1663–1687. [CrossRef]
101. Tang, L.; Li, J.; Fantus, S. Medical artificial intelligence ethics: A systematic review of empirical studies. *Digit. Health* **2023**, *9*, 20552076231186064. [CrossRef]
102. Arbelaez Ossa, L.; Lorenzini, G.; Milford, S.R.; Shaw, D.; Elger, B.S.; Rost, M. Integrating ethics in AI development: A qualitative study. *BMC Med. Ethics* **2024**, *25*, 10. [CrossRef]
103. Barrance, E.; Kazim, E.; Hilliard, A.; Trengove, M.; Zannone, S.; Koshiyama, A. Overview and commentary of the CDEI's extended roadmap to an effective AI assurance ecosystem. *Front. Artif. Intell.* **2022**, *5*, 932358. [CrossRef]
104. Hickok, M. Lessons learned from AI ethics principles for future actions. *AI Ethics* **2021**, *1*, 41–47. [CrossRef]
105. Attard-Frost, B.; De los Ríos, A.; Walters, D.R. The ethics of AI business practices: A review of 47 AI ethics guidelines. *AI Ethics* **2023**, *3*, 389–406. [CrossRef]
106. Ray, P.P. Benchmarking, ethical alignment, and evaluation framework for conversational AI: Advancing responsible development of chatgpt. *BenchCouncil Trans. Benchmarks Stand. Eval.* **2023**, *3*, 100136. [CrossRef]
107. European Commission. *Artificial Intelligence for Europe; COM(2018) 237 Final*; European Commission: Brussels, Belgium, 2018.
108. Hamon, R.; Junklewitz, H.; Sanchez, I.; Malgieri, G.; De Hert, P. Bridging the gap between AI and explainability in the GDPR: Towards trustworthiness-by-design in automated decision-making. *IEEE Comput. Intell. Mag.* **2022**, *17*, 72–85. [CrossRef]
109. Winfield, A.F.; Jirotko, M. Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2018**, *376*, 20180085. [CrossRef] [PubMed]
110. Pagallo, U. The legal challenges of big data: Putting secondary rules first in the field of EU data protection. *Eur. Data Prot. Law Rev.* **2017**, *3*, 36–46. [CrossRef]
111. Vedder, A.; Spajić, D. Moral autonomy of patients and legal barriers to a possible duty of health related data sharing. *Ethics Inf. Technol.* **2023**, *25*, 23. [CrossRef]
112. Owczarczuk, M. Ethical and regulatory challenges amid artificial intelligence development: An outline of the issue. *Ekon. I Prawo. Econ. Law* **2023**, *22*, 295–310. [CrossRef]
113. Bolte, L.; Vandemeulebroucke, T.; van Wynsberghe, A. From an ethics of carefulness to an ethics of desirability: Going beyond current ethics approaches to sustainable AI. *Sustainability* **2022**, *14*, 4472. [CrossRef]
114. Lokshina, I.; Kniezova, J.; Lanting, C. On Building Users' Initial Trust in Autonomous Vehicles. *Procedia Comput. Sci.* **2022**, *198*, 7–14. [CrossRef]
115. Gibney, E. The scant science behind Cambridge Analytica's controversial marketing techniques. *Nature* **2018**, *555*, 559–560. [CrossRef]
116. Goggin, B. Inside Facebook's Suicide Algorithm: Here's How the Company Uses Artificial Intelligence to Predict Your Mental State from Your Posts. 2019. Available online: <https://www.businessinsider.com/facebook-is-using-ai-to-try-to-predict-if-youre-suicidal-2018-12> (accessed on 6 November 2025).

117. Hill, K. The secretive company that might end privacy as we know it. *The New York Times*. 18 January 2020. Available online: <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html> (accessed on 3 October 2025).
118. Goujard, C. Italian Privacy Regulator Bans ChatGPT. *Politico*. 2023. Available online: <https://www.politico.eu/article/italian-privacy-regulator-bans-chatgpt/> (accessed on 6 November 2025).
119. Claessens, S.; Frost, J.; Turner, G.; Zhu, F. Fintech credit markets around the world: Size, drivers and policy issues. *BIS Q. Rev.* **2018**, 29–49. Available online: https://www.bis.org/publ/qtrpdf/r_qt1809e.htm (accessed on 15 February 2026).
120. Calo, R. Artificial intelligence policy: A primer and roadmap. *UCDL Rev.* **2017**, 51, 399.
121. Truby, J.; Brown, R.; Dahdal, A. Banking on AI: Mandating a proactive approach to AI regulation in the financial sector. *Law Financ. Mark. Rev.* **2020**, 14, 110–120. [CrossRef]
122. Arvan, M. Mental time-travel, semantic flexibility, and AI ethics. *AI Soc.* **2023**, 38, 2577–2596. [CrossRef]
123. Sanderson, C.; Douglas, D.; Lu, Q.; Schleiger, E.; Whittle, J.; Lacey, J.; Newnham, G.; Hajkowicz, S.; Robinson, C.; Hansen, D. AI ethics principles in practice: Perspectives of designers and developers. *IEEE Trans. Technol. Soc.* **2023**, 4, 171–187. [CrossRef]
124. Song, L.; Shokri, R.; Mittal, P. Privacy risks of securing machine learning models against adversarial examples. In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security; Association for Computing Machinery: New York, NY, USA, 2019.
125. Zhu, Y.; Lu, Y. Practice and challenges of the ethical governance of artificial intelligence in China: A new perspective. *Cult. Sci.* **2024**, 7, 14–23. [CrossRef]
126. Bleher, H.; Braun, M. Reflections on putting AI ethics into practice: How three AI ethics approaches conceptualize theory and practice. *Sci. Eng. Ethics* **2023**, 29, 21. [CrossRef]
127. Bryson, J.J.; Malikova, H. Is there an AI cold war? *Glob. Perspect.* **2021**, 2, 24803. [CrossRef]
128. European Commission. *Artificial Intelligence: A European Perspective*; Publications Office of the European Union: Luxembourg, 2018.
129. Bächle, T.C.; Bareis, J. “Autonomous weapons” as a geopolitical signifier in a national power play: Analysing AI imaginaries in Chinese and U.S. military policies. *Eur. J. Futures Res.* **2022**, 10, 20. [CrossRef]
130. Lingeivicius, J. Military artificial intelligence as power: Consideration for European Union actorness. *Ethics Inf. Technol.* **2023**, 25, 19. [CrossRef]
131. Yee, K.; Sebag, A.S.; Redfield, O.; Eck, M.; Sheng, E.; Belli, L. A keyword based approach to understanding the overpenalization of marginalized groups by English marginal abuse models on Twitter. In *Proceedings of the 3rd Workshop on Trustworthy Natural Language Processing (TrustNLP 2023)*; Association for Computational Linguistics (ACL): Kerrville, TX, USA, 2023; pp. 108–120.
132. IBM Security. *Cost of a Data Breach Report 2024*; IBM Corporation: Armonk, NY, USA, 2024. Available online: <https://www.ibm.com/downloads/documents/us-en/107a02e94948f4ec> (accessed on 16 February 2026).
133. Italian Parliament. Legge 23 settembre 2025, n. 132: Disposizioni e delega al Governo in materia di intelligenza artificiale (Law No. 132 of 23 September 2025 on Artificial Intelligence). *Gazzetta Ufficiale Della Repubblica Italiana*, No. 223. 23 September 2025. Available online: <https://www.julia-project.eu/database/legislation/267> (accessed on 11 February 2026).
134. OECD. *Revised Recommendation of the Council on Artificial Intelligence*; OECD: Paris, France, 2024. Available online: [https://one.oecd.org/document/C/MIN\(2024\)16/FINAL/en/pdf](https://one.oecd.org/document/C/MIN(2024)16/FINAL/en/pdf) (accessed on 11 February 2026).
135. Huang, X.; Kou, T.; Zhou, Q. Embedding AI ethics in the data lifecycle: A framework for enterprise AI governance. *Technol. Soc.* **2026**, 86, 103261. [CrossRef]
136. OECD. *Recommendation of the Council on Artificial Intelligence*; OECD: Paris, France, 2019. Available online: <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449> (accessed on 11 February 2026).
137. UNESCO. *Recommendation on the Ethics of Artificial Intelligence*; UNESCO: Paris, France, 2021. Available online: <https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence> (accessed on 11 February 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.