International Journal of Communication 20(2026), 1190–1213 1932–8036/20260005

Between Duty and Discipline: Kantian Ideals and Techno-Governmental Realities in the European Union Artificial Intelligence Act and the United Nations Artificial Intelligence Framework

MIR HASIB
The University of Alabama, USA

LYOMBE EKO
Texas Tech University, USA

This study comparatively analyzes two major global AI governance frameworks—the European Union's AI Act and the United Nations' Governing AI for Humanity report—through the dual lenses of Kantian ethics and techno-governmentality. While previous research has focused on the ethical content of these policies, this study investigates the fundamental tension between their normative rhetoric and operational mechanisms. The findings reveal a critical paradox: While both documents articulate strong Kantian commitments to human dignity, autonomy, and rights, their implementation strategies rely heavily on techno-governmental logics—managing populations through surveillance, classification, and statistical risk assessment. The analysis highlights deep structural constraints, including the EU's reliance on internal market competence, which subordinates rights to communitarian market standardization, and the "responsibility gap" in autonomous systems, rendering strict Kantian duty ethics structurally unfulfillable. Ultimately, the study argues that without robust participatory mechanisms, ethical AI governance risks becoming a tool shaped by power dynamics and bureaucratic control rather than moral protection. This research concludes that restoring human autonomy requires moving beyond technical compliance to democratically legitimized, inclusive decision making.

Keywords: artificial intelligence, European Union AI regulation, United Nations AI report, EU Kantian AI ethics, UN risk-based AI regulation, techno-governmentality in AI, AI policy

Artificial intelligence (AI) has become a transformative force with significant implications for society, the economy, and systems of governance. The rapid progress in AI technologies has triggered vigorous debates concerning their ethical dimensions, societal consequences, and the necessity for robust governance. Policy makers across the globe face the challenge of creating regulatory frameworks that

Mir Hasib: mirhasib1822@gmail.com
Lyombe Eko: leo.eko@ttu.edu
Date submitted: 2026-01-08

maximize AI's benefits while minimizing risks. These deliberations have led to numerous policy initiatives at both national and international levels, each embodying distinct approaches to AI regulation and oversight. However, treating these policies merely as protective guidelines overlooks their deeper structural function. Instead of fostering moral autonomy, these systems embed normative assumptions within algorithmic infrastructures that discipline and guide populations, often invisibly, according to state or corporate imperatives.

From this perspective, AI policy documents function as statements of ethical intent and instruments of power, operationalizing ethical values through mechanisms of surveillance, automated decision making, translation/transcription, categorization, and control (Bender & Hanna, 2025). For example, while the European Union (EU) AI Act strongly affirms commitments to human rights and autonomy, its implementation strategies simultaneously construct algorithmic environments that shape human behavior in predetermined ways. This duality highlights the risk that Kantian ideals may be absorbed into, and ultimately subordinated by, the logic and contextual matrices of techno-governmentality. This neologism extends Foucault's (1994) concept of governmentality to the ICT domain. Techno-governmentality refers to diverse conceptualizations of the role, place, and governance of technology and its morally acceptable use in specific national and supranational, politico-cultural, and ideological contexts like the EU and the United Nations (UN). It is the national and transnational governance of ICTs and AI technologies within the power dynamics of geostrategic competition among global powers capable of deploying these technologies. The techno-governmentality of the Internet, social media platforms, and AI reflects power dynamics among and within countries and transnational organizations. Therefore, ethical AI governance must move beyond rhetorical commitments to address the infrastructural and epistemic power embedded within AI systems.

To investigate this tension between governmentality and ethics, this study performs a comparative ethical analysis of two major AI policy frameworks: the EU's AI Act (Artificial Intelligence Act, 2021) and the UN's Governing AI for Humanity Final Report (UN Advisory Body on Artificial Intelligence, 2024). By analyzing these policies through the conflicting lenses of Kantian ethics and techno-governmentality, this study aims to reveal how transnational governmental strategies use ethical language to shape the trajectory of global AI governance and power dynamics. This research is especially relevant amid rising public awareness and concern about the risks and benefits of AI. Following OpenAI's release of ChatGPT in 2022, a group of leading AI experts issued an open letter in 2023 urging AI labs to halt large-scale AI experiments for at least six months, specifically the training of systems more advanced than GPT-4. Their subsequent actions, including policy recommendations (Future of Life Institute, 2023b) and the AI Safety Index (Future of Life Institute, 2025), highlighted the pressing need to address ethical challenges in AI development and governance. By analyzing official policy documents, the research aims to enhance understanding of the relationship between ethical considerations and AI policy formulation.

Literature Review

History of AI, Governance, and Policy

While scholarly concern about automation and computer ethics dates back to the 1950s (Stahl, 2021) and early professional codes (Loui & Miller, 2008), comprehensive AI governance only gained

momentum with the deep learning breakthroughs of the 2010s. This era produced numerous principle-based frameworks from organizations like the Institute of Electrical and Electronics Engineers (IEEE) and major technology firms, establishing a global vocabulary of values including fairness, transparency, and safety (Chesterman et al., 2024; Corrêa et al., 2023; Habuka & Osa, 2024). By 2019, this normative consensus hardened into intergovernmental policy, marked by the Organisation for Economic Co-operation and Development's (OECD) global AI recommendation and UNESCO's human rights-centric framework (OECD.AI, 2021; UNESCO, 2021). Concurrently, nations formalized strategic interests: The EU pursued binding regulation via its AI Act, while the United States and China launched strategies aligning AI with economic and national security goals (Habuka & Osa, 2024; National Institute of Standards and Technology [NIST], 2019; Webster et al., 2017). Reconciling these diverse geopolitical approaches remains a persistent governance challenge (Corrêa et al., 2023).

The United Nations

The UN has historically shaped global governance through approaches that emphasize technical harmonization and consensus building rather than legally binding, prescriptive legislation. Structurally, it operates within a hierarchical political economy defined by recognition of the sovereignty of diverse member states within a system of unequal power dynamics: the five permanent members of the Security Council—China, France, the United Kingdom, Russia, and the United States—hold veto power over all resolutions and decisions of the Security Council while regular members do not. This system necessitates a governance model often characterized as "soft law" or techno-diplomacy because it is not legally binding on UN member nations. This orientation is evident in the work of the United Nations Commission on International Trade Law (UNCITRAL), which produced American-inspired, market-oriented model laws on e-commerce (1996) and electronic signatures (2001), at the beginning of the Internet age. These model laws were endorsed by the five permanent members of the Security Council and subsequently deterritorialized from this center of power to countries around the world through a system of policy transfers that reflected unequal international power dynamics (Eko, 2012). As earlier analyses have observed, this reflects a distinctive form of UN governmentality: Instead of advancing a single moral doctrine, potentially at odds with local cultural contexts, the UN prioritizes "functional equivalence" and "technological neutrality" to promote global interoperability (Eko, 2012, p. 45). This long-standing reliance on standardization and development-oriented frameworks, such as the UN's Sustainable Development Goals, predisposes the UN toward a Western techno-governmental approach. Its recent AI governance initiatives, including the *Governing AI for Humanity* report, extend this trajectory by framing AI-related risks as challenges to global coordination that need to be addressed through inclusive, multistakeholder dialogue rather than enforceable legal obligations. In this sense, the UN's strategy is historically grounded in managing global populations and reducing digital inequalities within the context of unequal power dynamics between the five permanent members of the UN Security Council and the remaining member states. This contrasts with the EU's multilateral, communitarian, rights-enforcement paradigm.

The European Union

The EU has historically used its economic influence as the world's second largest single market to function as a "normative power" (Bradford, 2020, pp. 67–81) and a "regulatory superstate that rules the

https://doi.org/10.65476/zxewbx77

world" (Bradford, 2020, pp. 7–24), projecting its ethical standards through its communitarian market regulation—a dynamic commonly referred to as the Brussels Effect. In ICT, this expression denotes the globalization of European techno-governmentality through the promulgation of regulations that shape the international business ecosystem, leading to the Europeanization of data privacy, antitrust, intellectual property, social media platforms, content moderation, hate speech in cyberspace, and increasingly in AI (Bradford, 2020). The EU's AI regulatory strategy is not an isolated initiative, but a continuation of its broader project of digital constitutionalism, most prominently embodied in the General Data Protection Regulation (GDPR). The GDPR (European Parliament & Council of the European Union, 2016) established a significant precedent by recognizing data protection as a fundamental right rather than a mere commercial asset. This institutional background helps explain the EU's inclination toward a Kantian human rights ethical framework. The most far-reaching manifestation of the Brussels Effect is the Digital Services Act (DSA) of 2022. According to the EU (European Commission, 2026), the DSA:

The DSA empowers citizens by strengthening the protection of their fundamental rights online and giving them greater control and more choices when they navigate online platforms and search engines. The DSA also requires platforms to minimise the risks of exposing citizens, including children and young people, to illegal and harmful content (paras. 4–5).

A European civil society group, AlgorithmWatch sees the DSA as the EU's attempt "to make powerful tech platforms like YouTube, TikTok, Facebook, and X more transparent and accountable for the risks they pose to society" (Marsh & Podgorsek, 2026, p. 1). This is tantamount to bringing them within the ambit of European techno-governmentality. The companion legislation of the DSA is the Digital Markets Act (DMA), which aims to promote competition by preventing American and Chinese Big Tech companies from abusing their market share (Schwartz, 2023). These two pieces of legislation further globalize European governance of ICT in both real space and cyberspace.

Although the EU relies on the Single Market (Article 114 TFEU) as the primary legal basis for regulation, its political legitimacy is anchored in the post-World War II European Charter of Human Rights and Fundamental Freedoms of 1950. Accordingly, the EU AI Act (2021 proposal) follows the precautionary and rights-centered trajectory established by the 2019 Ethics Guidelines for Trustworthy AI. In contrast to the UN's consensus-oriented model, where hierarchies and thinly veiled unequal power dynamics are masked by the one nation, one vote system, the EU's communitarian, single-market regulatory tradition is binding and duty focused, seeking to reconcile the economic advantages of the Digital Single Market with a firm commitment to human dignity and the rule of law (Habuka & Osa, 2024). The divergence between the legally binding governance documents of the EU and the hortatory exhortations of UN resolutions, therefore, reflects not merely textual variations but also the distinct historical mandates and institutional specificities of the two entities.

Ethical Concerns on AI

The rapid evolution of AI has positioned transparency as a core ethical requirement (Sebastião et al., 2025). Following the release of ChatGPT in 2022, concerns rapidly escalated about disinformation (Hasib & Eko, 2025), academic misconduct (Hasib & Islam, 2025), economic inequality (Future of Life Institute,

2023a), biosecurity, and existential threats involving artificial general intelligence (AGI) and artificial superintelligence (ASI; IBM, 2026). However, scholars warn that prioritizing these futuristic risks must not obscure immediate, tangible harms, such as algorithmic bias and the exploitation of poorly paid data annotators in developing countries that process toxic content to train large language models (LLM; Heikkilä, 2023; Perrigo, 2023). These debates underscore a growing consensus that AI development is outpacing governance, creating an urgent imperative for robust safety measures and global regulatory cooperation.

Theoretical Framework

To examine this ethical debate more comprehensively, this study employs two theoretical approaches: techno-governmentality and Kantianism. These frameworks are deliberately chosen because they capture the central tension in contemporary AI governance: the divide between normative aspirations and practical implementation. Kantianism is adopted because it offers the core conceptual language of Western human rights discourse, autonomy, dignity, and duty, which underpins the stated objectives of regulations, such as the EU AI Act, DSA, and DMA. They address the question of how ICTs in general, and AI in particular, should relate to and treat individuals. In contrast, techno-governmentality is employed to explain the concrete mechanisms through which ICTs, AI systems, and regulatory regimes operate. These mechanisms include automation of decision making, information selection or recommendation, surveillance, statistical categorization or classification, translation and transcription, text and image generation, and population management (Bender & Hanna, 2025).

The importance of this comparison lies in its ability to illuminate the divergence between policy rhetoric and regulatory practice. A policy text may articulate a Kantian human rights orientation in its preamble, emphasizing trust and agency (empowerment), while embedding a techno-governmental logic in its operative provisions, such as requirements for monitoring, reporting perceived objectionable or illegal content, or behavioral modulation. An analysis grounded exclusively in Kantianism risks taking instances of 'ethics washing' at face value; one based solely on techno-governmentality may underappreciate genuine normative commitments to rights protection. By integrating both perspectives, this study demonstrates that ethical principles are not merely abstract ideals, but are implemented (or constrained) within existing power structures.

Techno-Governmentality

Michel Foucault (1991), a French philosopher, historian, and political activist, defined governmentality as the "conduct of conduct," referring to the ways power operates through institutions, procedures, and strategies to shape human behavior. This concept situates governance within historical, social, cultural, and philosophical contexts, illustrating how national and international authorities exercise power. It underscores the governance of populations through technological mechanisms and frames digital surveillance and AI regulation as tools of societal control (Eko, 2012). Techno-governmentality, derived from Foucault's notion, explores how technology functions as an instrument of governance and social control. Techno-governmentality focuses on how ICTs and AI systems, in practice, constitute an inseparable "capacity-communication-power" nexus (Foucault, 1994, p. 465). The title of Bradford's (2020) pathbreaking book, *The Brussels Effect: How the European Union Rules the World*, succinctly expresses the

https://doi.org/10.65476/zxewbx77

practical, worldwide, geostrategic dimensions and outcomes of the power dynamics and "power plays" of European techno-governmentality in trade and ICT.

Within AI policy, this framework sheds light on the power dynamics underlying governance, the influence of expertise in shaping regulations, and the ways AI technologies can be deployed to monitor and control populations. Evgeny Morozov's (2014) concept of "technological solutionism" expands on this critique, exposing the tendency to present complex social challenges as problems solvable through technological fixes. This perspective reveals that opinion leaders' involvement in AI governance often serves a symbolic role rather than a substantive one. Ethical guidelines and consultations frequently legitimize preexisting policy decisions rather than incorporate meaningful public input (Schultz et al., 2024). Authorities reflexively deterritorialize or transfer policies crafted to protect the individual's right to privacy and data protection from infringement by the media in physical space to ICTs and AI in cyberspace, often with little or no modification. Such practices reflect a top-down governance model where regulatory bodies and private corporations dominate AI policy making with limited democratic oversight. Techno-governmentality further emphasizes algorithmic governance as a power dynamic, a mechanism that shapes decision-making processes. Increasingly, AI systems automate and regulate decision making in domains as diverse as financial markets, law enforcement, the military, and social services, consolidating state power while reducing public accountability.

The Clash of Techno-Governmentality and Ethics: The "Anthropic Affair" in the United States

Anthropic, an AI start-up prioritizing the "global good" (Anthropic, 2026), clashed with the Trump Administration's Department of War over ethical AI use. CEO Dario Amodei requested exceptions to prevent their Claude AI from being used for mass domestic surveillance and fully autonomous weapons, arguing that current AI lacks the reliability and critical judgment of human troops (Amodei, 2026a). In retaliation, the administration designated Anthropic a "supply chain risk," a label normally reserved for foreign adversaries, and canceled millions in government contracts (Frenkel, 2026). Amodei condemned this unprecedented intimidation (Amodei, 2026a). Anthropic sued, arguing First Amendment violations. A California federal court granted an injunction, ruling the government's actions as "classic illegal First Amendment retaliation" (Anthropic PBC v. U.S. Department of War, 2026). Another lawsuit is currently pending in the D.C. Circuit (Bordelon & Cheney, 2026).

Significance of this Study

This study is significant because it extends Foucault's concept of governmentality to the governance of technology and examines the tension between the power-based commands of techno-governmentality and ethical recommendations that AI be governed in the public interest (the greatest good). This study adds to the knowledge on the intersections of the contextual matrices of techno-governmentality and power. After extensive searches in Google Scholar, Scopus, and Web of Science using keywords like "techno-governmentality," "AI," "control," and "regulation" in titles, abstracts, and author keywords, no prior studies or conceptual models addressing techno-governmentality in AI policy were found. Addressing this gap, the study introduces a new theoretical framework for techno-governmentality in ICT and AI policy, incorporating

key dimensions such as biopower, surveillance, control, power relations, knowledge, privacy, fairness, security, and risk-based regulation (see Table 1).

Table 1. Frameworks of Techno-Governmentality in AI Policy.

Framework	Description
Biopower, Surveillance, Civilian Control, and Military Command and Control	Examination of how AI technologies expand institutional power and control through algorithmic governance, mass surveillance, law enforcement, data-driven normalization, and the enforcement of social standards, reshaping individual and collective behavior (Amodei, 2026a; Amoore, 2000; Bowker & Star, 2000; Cheney-Lippold, 2018; Danaher et al., 2017; Foucault, 1991; Lyon, 2018; Yeung, 2018).
Power Dynamics and Knowledge	Analyzing how AI systems influence and reinforce societal power structures and authority by generating, validating, and applying knowledge, examining whose interests are served and who is empowered or marginalized through AI practices (Gitelman, 2013; Noble, 2018; Pasquale, 2016; Perrigo, 2023).
Privacy, Fairness, and Security	Addressing risks and harms related to AI through robust frameworks safeguarding individual privacy, ensuring fairness and nondiscrimination and maintaining safety and security against misuse or failures (Amodei et al., 2016; Barocas & Selbst, 2016; Gebru et al., 2018; Mitchell et al., 2021; Solove, 2011; Wachter & Mittelstadt, 2018).
Risk-Based Regulation	Implementing regulatory frameworks tailored to AI systems based on their assessed risks to individuals and society, ensuring proportionate oversight and compliance standards (Artificial Intelligence Act, 2021; Cath, 2018; European Commission, 2026; Kaminski & Malgieri, 2021).

Kantianism

Kantianism originates from the philosophical work of Immanuel Kant, a German thinker and key figure in modern Western philosophy. It centers on the inherent dignity and autonomy of individuals that flows from their capacity to reason. At its core is Kant's (1981) categorical imperative: One should "act only according to that maxim whereby you can, at the same time, will that it should become a universal law" (pp. 30–31). Kant (1785) argued that moral actions must follow universal principles that uphold human dignity. This framework forms the foundation of modern human rights theory. In the realm of AI governance, ethical challenges remain, particularly around privacy, bias, automated decision making, and mass surveillance (Amodei, 2026a). The Stanford Encyclopedia of Philosophy identifies major ethical risks posed by AI, including opaque decision making, behavioral manipulation, and the possibility of AI autonomy surpassing human control (Johnson & Cureton, 2022).

Kantian ethics is grounded in the intrinsic dignity and autonomy of persons. In the field of AI, it is essential to differentiate between the three formulations of Kant's categorical imperative, as each provides distinct, and at times competing, guidance for governance. The first formulation, the *Formula of Universal Law*, requires that one act only according to maxims that could be universally applied. Within AI governance, this principle is frequently understood as demanding consistency in algorithmic reasoning

https://doi.org/10.65476/zxewbx77

and the universal applicability of decision-making rules (Powers, 2006). By contrast, the second formulation, the *Formula of Humanity*, instructs that humanity must never be treated merely as a means to an end, but always as an end in and of itself. This formulation points in a different normative direction, establishing a firm prohibition against using individuals instrumentally, for example, through data extraction, data labeling and annotation, or behavioral manipulation, issues that lie at the heart of tensions within the AI economy. The third formulation, the *Formula of Autonomy* or *Kingdom of Ends*, envisions a community of rational agents and highlights the broader social environment in which AI systems function. This study draws on these distinctions to assess whether policy documents emphasize the procedural universalism associated with algorithmic rules (Formula 1) or the substantive safeguarding of human dignity (Formula 2).

These unresolved issues underscore the complexity of embedding ethical theory into AI regulation. A major challenge in AI ethics is "ethics washing," where corporations and governments adopt ethical guidelines primarily for public relations rather than implementing meaningful regulatory measures (Schultz et al., 2024). This raises doubts about whether ethical principles genuinely shape AI governance or merely serve as rhetorical tools to legitimize power-driven policy decisions. After conducting extensive searches in Google Scholar, Scopus, and Web of Science using keywords such as "Kantianism," "Kantian Ethics," "AI," and "policy" in titles, abstracts, and author keywords, no prior studies or conceptual models addressing Kantian ethics in AI policy were identified. To address this gap, this study introduces a new theoretical framework for Kantian ethics in AI policy. This framework incorporates key elements such as human dignity, autonomy, freedom, rights protection, categorical imperative, duty, moral obligation, transparency, explainability, accountability, responsibility, and human oversight (see Table 2). The researchers also note that some frames overlap across both frameworks; for instance, transparency and accountability are central to Kantian ethics (as they uphold autonomy and moral responsibility) and are equally critical in techno-governmentality (as mechanisms for governing populations through visibility and auditability).

Table 2. Frameworks of Kantian Ethics in AI Policy.

Framework	Description
Human-Centered Ethics	Recognition, respect and prioritization of human dignity, autonomy, and fundamental rights within AI systems, ensuring technology enhances human well-being, freedom, and moral status, treating humans as ends in themselves (Amodei, 2026b; Coeckelbergh, 2020; Floridi & Cowls, 2019; Kant, 1785; McGregor et al., 2019; Ulgen, 2017; UNESCO, 2021).
Universal Moral Principals	Establishing universal ethical standards for AI governance based on duty and moral imperatives, ensuring that ethical obligations guide AI system design and regulation regardless of context or outcomes (Etzioni & Etzioni, 2017; Kant, 1981; Moor, 2006; Nyholm, 2018; Powers, 2006).
Transparency, Accountability, and Oversight	Mechanisms ensuring AI decision making is transparent, understandable, and accountable, combined with meaningful human oversight to maintain responsible and controlled AI operations (Amodei, 2026a; Coeckelbergh, 2020; Mittelstadt et al., 2016; Rudin, 2019; Shneiderman, 2020).

Framing Analysis of Regulatory Texts

The researchers employed a framing approach in this study. Although framing analysis emerged within media studies as a tool for examining how news discourse influences public understanding, it is especially well suited to the analysis of policy documents. Entman (1993) conceptualizes framing as the selection of "certain aspects of a perceived reality" (p. 52) and their elevation within a communicative text to advance a particular definition of the problem, causal explanation, moral judgment, and/or proposed remedy. Regulatory instruments operate precisely in this way. They define what constitutes the "problem" of AI—for instance, whether it is primarily a risk to fundamental rights or a source of economic innovation—articulate normative assessments of its consequences, and prescribe corresponding responses, such as outright bans, risk-based controls, or voluntary standards. Through this analytical lens, these documents are examined not simply as sets of legal provisions, but as persuasive texts that actively construct the reality they aim to govern, foregrounding certain ethical commitments while marginalizing others.

Research Questions

Drawing on the literature review and theoretical frameworks, this study proposes the following research questions to guide the analysis of the EU's AI Act and the UN's Governing AI for Humanity Final Report:

RQ1: What ethical components are embedded in each of the selected policy documents?

RQ1 seeks to identify the specific ethical values emphasized in these and how they are situated within broader governance frameworks. This will help reveal commonalities and differences in ethical priorities across jurisdictions and institutions.

RQ2(a): To what extent do these AI policies reflect Kantian principles in their framing of AI governance?

RQ2(b): To what extent do these AI policies reflect techno-governmentality principles in their framing of AI governance?

RQ2 aims to examine the underlying ethical and governance philosophies present in the policy documents. This analysis will clarify how different theoretical perspectives influence approaches to AI regulation and development, offering insights into the philosophical foundations of AI governance strategies.

Together, these research questions provide a structured framework for exploring ethical considerations and theoretical underpinnings in AI policy. They will guide the study in uncovering nuanced approaches to AI governance across diverse global contexts.

Methodology

This study adopts a qualitative content analysis approach, combining deductive coding with interpretative strategies (Bryman, 2016; Krippendorff, 2018; Schreier, 2012; Sebastião et al., 2025), to

https://doi.org/10.65476/zxewbx77

thematically examine two distinct AI governance models: the EU's AI Act (Artificial Intelligence Act, 2021) and the UN's Governing AI for Humanity Final Report (UN Advisory Body on Artificial Intelligence, 2024; see Table 3). While structurally asymmetrical, the EU AI Act is binding "hard law," capable of imposing sanctions, whereas the UN Report is nonbinding "soft law" or hortatory guidance. Both function as normative prototypes emerging from distinct politico-cultural contexts (Guest et al., 2012). Juxtaposing these frameworks demonstrates how comparable ethical notions are operationalized through divergent governance mechanisms, whether coercive (EU) or aspirational (UN).

It is necessary to recognize the structural asymmetry between the two documents: the EU AI Act constitutes binding "hard law" capable of imposing sanctions, whereas the UN Report represents nonbinding "soft law." It is nonenforceable hortatory guidance that nevertheless comes with the moral authority of the international community. Evaluating both documents as instruments of governmentality within a shared analytical framework is methodologically essential. Both texts function as normative prototypes emerging from distinct politico-cultural contexts and advancing competing problem definitions and ethical solutions to the same issue: AI governance. Both texts function as normative prototypes emerging from different politico-cultural contexts and advancing competing problem definitions and ethical solutions to the same problem: AI governance. By juxtaposing a binding regional regulation with a global advisory instrument, this policy analysis study demonstrates how comparable ethical notions, such as safety or fairness, are operationalized through distinct governance mechanisms, whether coercive (EU) or aspirational (UN).

Table 3. Description of the Analyzed Policy Documents.

Document Title	Issuer	Year	Number of Pages (Without Appendix/Annex)	Nature of the Document	Language
EU's AI Act (Artificial Intelligence Act)	European Parliament and Council	Drafted 2021, Published in 2024	123	Regulation	Multilingual
UN's Governing AI for Humanity Final Report	United Nations High Level Advisory Body on AI	2024	72	Report	Multilingual

An initial codebook was developed deductively based on the theoretical frameworks of techno-governmentality and Kantian ethics and validated through a pilot coding phase of approximately 10%–15% of the text to ensure conceptual consistency (Krippendorff, 2018; Schreier, 2012). For instance, broad categories like "Accountability" were disaggregated into subcodes like "Regulatory Oversight" and "Public Redress." The entire corpus was then systematically coded line by line using an iterative and reflexive approach with secondary analytic memos. Codes were organized into thematic categories aligned with the research questions, comparing how themes like human dignity or algorithmic control were emphasized across jurisdictions. To ensure rigor and trustworthiness, frequent direct quotations were used to preserve the original policy context alongside an audit trail of methodological decisions (Bowen, 2009; Lincoln & Guba, 1985).

Results

The first research question (RQ1) examined the ethical components embedded in the selected policy documents. The analysis reveals that both frameworks prioritize dignity, autonomy, transparency, and accountability, though they operationalize these values through differing mechanisms.

The EU AI Act (2021)

The EU AI Act (Artificial Intelligence Act, 2021) embodies a duty-driven ethical framework centered on fundamental rights. It enforces universal moral principles through rigorous safeguards against discrimination and manipulation. The Act mandates that AI must serve as a tool for human well-being, embedding privacy protections such as data minimization and encryption throughout the AI life cycle. It further operationalizes accountability through "fail-safe" mechanisms and human oversight requirements, ensuring that high-risk systems remain subject to human control. Functionally, the EU defines an AI system as a machine-based tool operating with varying autonomy and adaptiveness that infers how to generate outputs influencing physical or virtual environments (Artificial Intelligence Act, 2021, p. 46). Table 4 summarizes the ethical components in the EU AI Act.

Table 4. Summary of Ethical Components in the EU AI Act.

Ethical Component	Key Features
Human Dignity and Autonomy	Prioritizing human well-being, autonomy, and dignity; prohibiting manipulative and exploitative practices
Universal Moral Principles	Duty-based ethics applicable across contexts; AI design guided by EU's foundational values and human-centric norms
Transparency	Mandated disclosure of deep fakes and artificially generated content; interpretable AI outputs
Nondiscrimination and Fairness	Protections against reinforcing historical biases and discrimination
Privacy and Data Protection	Ensuring data minimization, anonymization, and encryption; embedding privacy protections throughout AI life cycle
Accountability and Oversight	Human oversight required; fail-safe mechanisms; resilient, secure, and robust AI designs
Risk Management	Risk-tiered regulation based on societal impact; continuous iterative process for risk assessment and mitigation
Inclusivity and Stakeholder Engagement	Participatory approach including civil society, academia, and stakeholders for legitimacy and ethical integrity

The UNs' Governing AI for Humanity Final Report (2024)

The UN Advisory Body on Artificial Intelligence (2024) articulates an ethical framework grounded in international human rights law in their "Governing AI for Humanity: Final Report." It underscores the imperative to place human welfare at the core of AI development, advocating for accountability mechanisms that address geopolitical and societal implications. The report emphasizes "societal integrity," aiming to

https://doi.org/10.65476/zxewbx77

mitigate risks related to misinformation and the erosion of social norms. Unlike the EU's binding regulations, the UN emphasizes inclusivity and "knowledge equity," aiming to counteract systemic marginalization and the digital divide. For its regulatory scope, the report adopts the OECD's definition of an AI system, which functionally mirrors the EU's definition by emphasizing a machine-based system's capacity to infer inputs and generate environment-influencing outputs with varying autonomy (UN Advisory Body on Artificial Intelligence, 2024, p. 24). Table 5 summarizes the ethical components in the UN's Governing AI for Humanity Final Report.

Table 5. Summary of Ethical Components in the UN's Governing AI for Humanity Final Report.

Ethical Component	Key Features
Human-Centered Ethics	Protection against manipulation, exploitation, discrimination, surveillance, and loss of agency
Human Rights and Civil Liberties	Rights to privacy, expression, assembly, due process, and equal treatment
Moral Imperatives	Ethical boundaries, particularly the prohibition of automating lethal decisions
Universal Moral Standards	Anchored in UN Charter, international law, and Sustainable Development Goals (SDGs)
Accountability and Transparency	Mechanisms ensuring explainability, oversight, and meaningful redress of harms
Societal Integrity and Trust	Addressing misinformation, disinformation, and algorithmic governance effects on social norms
Economic and Social Inclusion	Mitigating digital divide, employment displacement, and marginalization risks
Power Dynamics and Knowledge Equity	Ensuring equitable access to AI technology and expertise, counteracting systemic marginalization
Existential and Control Risks	Managing systemic risks, preventing loss of control, and ensuring oversight of advanced autonomous AI

The second research question, RQ2(a), was concerned with the extent to which AI policies reflect Kantian principles in their framing of AI governance. The analysis reveals that both documents rely heavily on Kantian rhetoric, specifically the *Formula of Humanity* (treating humans as ends, not means) and the duty of transparency. The EU AI Act explicitly frames AI not merely as a tool for efficiency, but as a technology that must be subordinated to human moral status. The regulation asserts that AI "should be a human-centric technology. It should serve as a tool for people, with the ultimate aim of increasing human well-being" (Artificial Intelligence Act, 2021, p. 2). This Kantian commitment to human dignity is operationalized through strict prohibitions on systems that instrumentalize human psychology. For instance, the Act bans manipulative practices because "they contradict Union values of respect for human dignity, freedom, equality, democracy and the rule of law" (Artificial Intelligence Act, 2021, p. 8). Furthermore, the Act operationalizes the Kantian duty of truth-telling through transparency mandates. To prevent deception—a violation of autonomy—the Act requires that deployers of deep fakes "shall disclose that the content has been artificially generated or manipulated" (Artificial Intelligence Act, 2021, p. 82). This ensures that citizens interact with AI as informed rational agents rather than as manipulated subjects.

Similarly, the UN Report anchors its framework in universal moral duties, asserting that "human rights must be at the centre of AI governance, ensuring rights-based accountability across jurisdictions" (UN Advisory Body on Artificial Intelligence, 2024, p. 39). It establishes categorical imperatives that transcend utility, most notably in its stance on lethal autonomous weapons. The report argues that "On legal and moral grounds, kill decisions should not be automated through AI," establishing a deontological boundary that machine efficiency cannot override (UN Advisory Body on Artificial Intelligence, 2024, p. 30). Table 6 summarizes these Kantian categories:

Table 6. Comparative Summary of Kantian Categories in Each Policy Document.

Kantian Category	The EU AI Regulation	The UN's Governing AI for Humanity Final Report
Human-Centered Ethics	The Act prioritizes human dignity, autonomy, and well-being, explicitly prohibiting manipulative AI practices that exploit or control individuals. It ensures technology serves humans as ends rather than means (Artificial Intelligence Act, 2021, pp. 1–2, 8).	The report emphasizes human dignity, autonomy, and fundamental rights, warning against manipulation, exploitation, and surveillance. It highlights the need to design AI that respects and protects human agency (UN Advisory Body on Artificial Intelligence, 2024, pp. 27, 31, 35).
Universal Moral Principles	The Act enshrines duty-based ethics across contexts, embedding fundamental rights, democracy, and rule of law throughout. Seven nonbinding ethical principles guide providers and deployers, aligning with deontological ethics (Artificial Intelligence Act, 2021, p. 8).	The report grounds its governance model in the UN Charter, international human rights law, and SDGs, establishing a universal ethical foundation that transcends specific applications or contexts (UN Advisory Body on Artificial Intelligence, 2024, p. 38).
Transparency, Accountability, and Oversight	The Act mandates disclosure for deepfakes, interpretability of AI outputs, and human oversight, ensuring humans remain in control of AI decisions. Technical documentation supports transparency and accountability throughout the AI life cycle (Artificial Intelligence Act, 2021, pp. 20, 59, 82).	The report prioritizes explainability, oversight, and redress, mandating transparent AI operations and accountability frameworks to ensure human control over AI systems and prevent harm (UN Advisory Body on Artificial Intelligence, 2024, pp. 7, 51, 78).

The next research question, RQ2(b), asked: To what extent do these AI policies reflect Techno-governmentality principles in their framing of AI governance? While the documents speak the language of Kantian ethics, the textual analysis reveals that they simultaneously operate through the logic of techno-governmentality—managing populations through surveillance, categorization, and risk assessment. The EU AI Act demonstrates how governance is exercised through the classification of life and behavior. The text acknowledges that AI systems used for "real-time remote biometric identification evoke a feeling of constant surveillance and indirectly dissuade the exercise of the freedom of assembly" (Artificial Intelligence Act,

https://doi.org/10.65476/zxewbx77

2021, p. 9). Despite this acknowledgement, the Act does not universally ban these technologies, but rather regulates them, effectively managing the conditions under which the state may exercise biopolitical control. Furthermore, the Act reveals how AI is used to automate the determination of citizens' eligibility for social services, noting that systems used "to evaluate eligibility for essential public assistance benefits may have a significant impact on persons' livelihood" (Artificial Intelligence Act, 2021, p. 16). Here, governance is not about moral duty, but about the administrative sorting of populations based on data-driven assessments of risk and worthiness.

The UN Report offers a more systemic critique of techno-governmentality, focusing on how AI concentrates power and knowledge. It explicitly states that "the accelerating development of AI concentrates power and wealth on a global scale, with geopolitical and geoeconomic implications" (UN Advisory Body on Artificial Intelligence, 2024, p. 37). The report further details how algorithmic governance shapes social norms, warning of "closed loop information ecosystems" that can manipulate societies, "potentially making them more accepting of intolerance and violence" (UN Advisory Body on Artificial Intelligence, 2024, p. 35). Unlike the EU's focus on market regulation, the UN highlights the epistemic violence of AI, noting that "AI capabilities in low- and lower-middle-income countries cannot be achieved without securing reliable electricity and Internet connectivity," creating a divide where the Global North governs the digital infrastructure of the Global South (UN Advisory Body on Artificial Intelligence, 2024, p. 61). Table 7 summarizes these governmentality themes:

Table 7. Comparative Summary of Techno-Governmentality Categories in Each Policy Documents.

Techno-Governmentality Category	The EU AI Act	The UN's Governing AI for Humanity Final Report
Biopower, Surveillance, and Control	The Act recognizes AI's potential for algorithmic governance, highlighting risks of surveillance, social control, and data-driven decision making in law enforcement, benefits administration, and employment. Provisions address these risks through regulatory constraints (Artificial Intelligence Act, 2021, pp. 9, 16–17).	The report highlights risks of AI-driven surveillance, algorithmic governance, and societal manipulation, particularly in law enforcement, migration, and border control (UN Advisory Body on Artificial Intelligence, 2024, pp. 30, 35).
Power Dynamics and Knowledge	The Act acknowledges AI's role in shaping social status, opportunity, and exclusion, particularly in high-risk areas like credit scoring and eligibility for public benefits. These knowledge systems reinforce structural power relations (Artificial Intelligence Act, 2021, p. 16).	The report exposes how AI can concentrate power, shape epistemological authority, and exacerbate global inequalities, notably through selective funding and control of data and infrastructure (UN Advisory Body on Artificial Intelligence, 2024, pp. 25, 28, 37, 51).
Privacy, Fairness, and Security	The Act embeds privacy protections (e.g., data minimization, encryption) and addresses risks of bias, discrimination, and cybersecurity vulnerabilities, ensuring AI systems do not exacerbate inequalities or compromise individual rights (Artificial Intelligence Act, 2021, pp. 16, 20, 22).	The report emphasizes protecting privacy, due process, and fairness, particularly concerning state AI use in enforcement and migration. It calls for safeguards against discrimination and biases (UN Advisory Body on Artificial Intelligence, 2024, pp. 27, 30).
Risk-Based Regulation	The Act embeds privacy protections (e.g., data minimization, encryption) and addresses risks of bias, discrimination, and cybersecurity vulnerabilities, ensuring AI systems do not exacerbate inequalities or compromise individual rights (Artificial Intelligence Act, 2021, pp. 16, 20, 22).	The report proposes a risk-based, proportionate approach to AI governance, tailoring regulation based on system risks, with global coordination of classification frameworks (UN Advisory Body on Artificial Intelligence, 2024, pp. 8, 76).

https://doi.org/10.65476/zxewbx77

Discussion

The analysis of the EU and UN policy documents reveals a fundamental tension between two contrasting ethical-political logics: the rhetoric of Kantian moral duty and the operational reality of techno-governmentality. While both frameworks consistently invoke human-centered ethics—prioritizing dignity, autonomy, and fundamental rights—the mechanisms used to enforce these values frequently subordinate moral autonomy to the logic of administrative control (Sebastião et al., 2025). Kantian ethics implies that governance should respect individuals as ends in themselves, enabling rational moral judgment. In contrast, the analysis confirms that these regulatory regimes increasingly function as instruments of techno-governmentality, where AI systems preemptively shape choices, enforce norms, and engineer behavioral conformity (Eko, 2012; Foucault, 1991).

The Trustworthiness Paradox and Market Logic

This tension is most visible in the EU AI Act's structural limitations. While employing fundamental rights language, the Act is actually grounded in the EU's internal market competence (Article 114 TFEU), framing rights protections as market-entry conditions rather than categorical moral obligations (Ebers, 2025). This creates a "trustworthiness paradox." While governance should focus on demonstrable trustworthiness (O'Neill, 2018), the Act's reliance on self-assessment for high-risk systems creates a compliance loophole (Wachter, 2024). Relying on harmonized technical standards fosters a "false sense of safety," enabling providers to bypass complex normative trade-offs (Laux et al., 2024, p. 3). Thus, the Act operates as a mechanism of techno-governmentality, prioritizing procedural market compliance over the substantive Kantian obligation to treat citizens as informed, rational agents.

Ontological Limits and the Responsibility Gap

The application of Kantian ethics is further constrained because AI agents lack the "dialectical" capacity for genuine moral agency, shifting the entire burden of responsibility to human oversight (Chakraborty & Bhuyan, 2023). However, this shift is complicated by the "responsibility gap" (Matthias, 2004). As AI systems evolve into adaptive learning automata operating beyond direct programmer control, strict Kantian duties become structurally unfulfillable because blame cannot be coherently assigned. Consequently, both the EU and UN frameworks retreat from Kantian moral accountability toward techno-governmentality. In this paradigm, governance manages the statistical probability of harm through risk-based regulation rather than assigning moral blame (Cath, 2018; Kaminski & Malgieri, 2021). Ethical failures are treated as system errors to be minimized, effectively managing the "accident" rather than upholding the maxim.

The Democratic Deficit and Participation

The concentration of epistemic authority in risk-based models exposes a significant "participation gap" (UN Advisory Body on Artificial Intelligence, 2024, p. 42). While the UN Report advocates for "multistakeholder cooperation" to address power imbalances (UN Advisory Body on Artificial Intelligence, 2024, p. 77), the EU's reliance on technical bodies insulates critical value judgments from democratic

scrutiny. From a Kantian perspective, defining "safety" and "fundamental rights" solely through technical experts treats the public as a managed population rather than as autonomous agents. This concentration of technical power exacerbates global inequality, reflecting the "epistemic violence" of the Global North in determining digital infrastructures for the Global South (Future of Life Institute, 2023b). Reconciling the efficiency of techno-governmentality with Kantian moral demands requires participatory governance. Authentic ethical governance demands democratically legitimized maxims, ensuring acceptable risk definitions reflect the collective will of the affected public rather than the calculated trade-offs of a bureaucratic elite.

Conclusion

This study offered a critical comparative analysis of the EU's AI Act and the UN's Governing AI for Humanity report, examining them not merely as regulatory instruments but as competing normative prototypes for global governance of AI. By juxtaposing Kantian ethics with techno-governmentality, the research revealed a fundamental tension at the heart of AI policy: the conflict between the stated intent to protect moral autonomy and the operational reality of managing populations through algorithmic control. The analysis demonstrated that while both frameworks invoke Kantian ideals of dignity and duty, these values are frequently subordinated to the logic of techno-governmentality. In the EU context, the reliance on internal market competence and self-certification creates a "trustworthiness paradox," where technical compliance often supplants genuine rights protection. Similarly, the UN's focus on harmonization and soft law risks reducing ethical governance to a coordination problem rather than a moral imperative. Furthermore, the study identified a critical "responsibility gap," where the autonomy of learning systems renders strict Kantian accountability structurally impossible, forcing a retreat into probability-based risk management.

These findings have profound implications for the future of AI governance. They suggest that current "human-centric" policies function simultaneously as instruments of power that discipline behavior and concentrate epistemic authority in the hands of policy makers and technical experts. The EU Digital Services Act, which set out to empower European citizens, increasingly appears to be another manifestation of the Brussels Effect—the globalization of European techno-governmentality and the incorporation of Big Tech companies within the ambit of European Single Market-regulated self-regulation (Marsh & Podgorsek, 2026; Schultz & Held, 2004). To counter this democratic deficit, this study concludes that policy makers must look beyond risk metrics and prioritize participatory decision making. This is what Habermas (1984) calls communicative action, discursive practices, and ethics that allow for the common elaboration of political, social, and cultural norms between citizens and nations in nonhierarchical democratic spaces. Strengthening ethical governance requires shifting the power to define "acceptable risk" from bureaucratic bodies back to the affected public. Only by anchoring technical standards in inclusive, democratic deliberation can governance frameworks bridge the gap between the administrative efficiency of techno-governmentality and the moral demands of Kantian autonomy.

https://doi.org/10.65476/zxewbx77

References

Amodei, D. (2026a, February 26). Statement from Dario Amodei on our discussions with the Department of War. *Anthropic*. https://www.anthropic.com/news/statement-department-of-war

Amodei, D. (2026b, March 6). Where things stand with the Department of War. *Anthropic*. https://www.anthropic.com/news/where-stand-department-war

Amodei, D., Olah, C., Brain, G., Steinhardt, J., Christiano, P., Schulman, J., Dan, O., & Google Brain, M. (2016). *Concrete problems in AI Safety*. https://arxiv.org/pdf/1606.06565

Amoore, L. (2000). Machine learning political orders. *Review of International Studies, 49*(1), 20–36. https://doi.org/10.1017/S0260210522000031

Anthropic PBC v. U.S. Department of War, No. 26-cv-01996-RFL (N.D. Cal. 2026). https://cand.uscourts.gov/cases-e-filing/cases/326-cv-01996/anthropic-pbc-v-us-department-war-et-al

Anthropic. (2026). *Company.* https://www.anthropic.com/company

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.2477899

Bender, E., & Hanna, A. (2025). The AI con: How to fight big tech's hype and create the future we want. New York, NY: Harper-Collins.

Bordelon, B., & Cheney, K. (2026, March 9). Anthropic sues Trump admin over supply-chain risk label. *Politico*. https://www.politico.com/news/2026/03/09/anthropic-sues-trump-admin-over-supply-chain-risk-label-0081871

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40. https://doi.org/10.3316/QRJ0902027/FULL/XML

Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press eBooks. https://doi.org/10.7551/mitpress/6352.001.0001

Bradford, A. (2020, February). *The Brussels effect: How the European Union rules the world*. New York, NY: Oxford University Press. https://doi.org/10.1093/oso/9780190088583.001.0001

Bryman, A. (2016). *Social research methods* (5th ed.). New York, NY: Oxford University Press.

Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2133), 20180080. https://doi.org/10.1098/RSTA.2018.0080

Chakraborty, A., & Bhuyan, N. (2023). Can artificial intelligence be a Kantian moral agent? On moral autonomy of AI system. *AI and Ethics, 4*(2), 325–331. https://doi.org/10.1007/s43681-023-00269-6

Cheney-Lippold, J. (2018). *We are data*. Boston, MA: NYU Press eBooks. https://doi.org/10.2307/J.CTT1GK0941

Chesterman, S., Gao, Y., Hahn, J., & Sticher, V. (2024). The evolution of AI governance. *Computer, 57*(9), 80–92. https://doi.org/10.1109/MC.2024.3381215

Coeckelbergh, M. (2020). *AI ethics*. 229. Cambridge, MA: MIT Press. https://coeckelbergh.net/ai-ethics/

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., & de Oliveira, N. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns, 4*(10), 100857. https://doi.org/10.1016/J.PATTER.2023.100857

Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., Felzmann, H., Haklay, M., Khoo, S. M., Morison, J., Murphy, M. H., O'Brolchain, N., Schafer, B., & Shankar, K. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data and Society, 4*(2), 1–21. https://doi.org/10.1177/2053951717726554

Ebers, M. (2025). Truly risk-based regulation of artificial intelligence how to implement the EU's AI act. *European Journal of Risk Regulation, 16*(2), 684–703. https://doi.org/10.1017/err.2024.78

Eko, L. (2012). *New media, old regimes: Case studies in comparative communication law and policy*. Lanham, MD: Lexington Books. https://doi.org/10.5771/9780739167908

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51–58. https://doi.org/10.1111/J.1460-2466.1993.TB01304.X

Etzioni, A., & Etzioni, O. (2017). Incorporating ethics into artificial intelligence. *Journal of Ethics, 21*(4), 403–418. https://doi.org/10.1007/S10892-017-9252-2/METRICS.

European Commission. (2026, March 23). The digital services act. *Directorate-General for Communications Networks, Content and Technology.* https://digital-strategy.ec.europa.eu/en/policies/digital-services-act

https://doi.org/10.65476/zxewbx77

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1), 2–15. https://doi.org/10.1162/99608F92.8CD550D1

Foucault, M. (1991). The Foucault effect: Studies in governmentality with two lectures by and an interview with Michel Foucault. In G. Burchell, C. Gordon, & P. Miller (Eds.), *Foucault, M. (1991). The Foucault effect: Studies in governmentality. Harvester Wheatsheaf* (Number 2). Chicago, IL: The University of Chicago Press & Hemel Hempstead: Harvester Wheatsheaf. https://doi.org/10.1017/S0829320100002507

Foucault, M. (1994). *Dits et ecrits (Tome II)* [Sayings and writings, vol. II]. Paris, France: *Gallimard*.

Frenkel, S. (2026, March 17). U.S. says Anthropic is an 'unacceptable' national security risk. *The New York Times*. https://www.nytimes.com/2026/03/17/technology/anthropic-pentagon-national-security-risk.html

Future of Life Institute. (2023a, March 22). *Pause giant AI experiments: An open letter*. https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Future of Life Institute. (2023b, April 19). *Policymaking in the Pause*. https://futureoflife.org/document/policymaking-in-the-pause/

Future of Life Institute. (2025). *AI safety index winter 2025*. https://futureoflife.org/ai-safety-index-winter-2025/

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2018). Datasheets for datasets. *Communications of the ACM, 64*(12), 86–92. https://doi.org/10.1145/3458723

European Parliament & Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union, L 119*, 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Gitelman, L. (2013). *"Raw data" is an oxymoron*. Cambridge, MA: MIT Press. https://dsl.lsu.edu/nehtextualdata/wp-content/uploads/2017/11/RawData.pdf

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. Thousand Oaks, CA: SAGE Publications. https://doi.org/10.4135/9781483384436

Habermas, J. (1984). *Theory of communicative action, volume. I: Reason and the rationalization of society* (Thomas A. McCarthy, Trans.). Boston, MA: Beacon Press.

Habuka, H., & Osa, D. U. S. de la. (2024). Shaping global AI governance: Enhancements and next steps for the G7 Hiroshima AI process. *Cambridge Forum on AI Law and Governance, 1*(e15), 1–26. https://doi.org/10.1017/cfl.2024.5

Hasib, M., & Eko, L. (2025). Unraveling societal discourse on artificial intelligence through visual analysis of magazine covers. *Visual Communication Quarterly, 32*(4), 264–282. https://doi.org/10.1080/15551393.2025.2579258

Hasib, M., & Islam, M. S. (2025). How university students in Bangladesh engage with ChatGPT: A qualitative study. *PLoS One, 20*(9), 1–18. https://doi.org/10.1371/journal.pone.0333089

Heikkilä, M. (2023, September 23). What's changed since the "pause AI" letter six months ago? *MIT Technology Review*. https://www.technologyreview.com/2023/09/26/1080299/six-months-on-from-the-pause-letter/

IBM. (2026). What is artificial general intelligence (AGI)? https://www.ibm.com/think/topics/artificial-general-intelligence

Johnson, R., & Cureton, A. (2022). Kant's moral philosophy. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/kant-moral/

Kaminski, M. E., & Malgieri, G. (2021). Algorithmic impact assessments under the GDPR: Producing multi-layered explanations. *International Data Privacy Law, 11*(2), 125–144. https://doi.org/10.1093/IDPL/IPAA020

Kant, I. (1785). *Groundwork for the metaphysic of morals* (T. E. Hill & A. Zweig, Eds.). New York, NY: Oxford University Press. https://doi.org/10.12987/9780300128154

Kant, I. (1981). *Grounding for the metaphysics of morals* (J.W. Ellington, Trans.). Indianapolis, Indiana: Hackett Pub Co. https://hackettpublishing.com/grounding-for-the-metaphysics-of-morals

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed., T. Accomazzo, Ed.). Thousand Oaks, CA: SAGE Publications. https://doi.org/10.4135/9781071878781

Laux, J., Wachter, S., & Mittelstadt, B. (2024). Three pathways for standardisation and ethical disclosure by default under the European Union artificial intelligence act. *Computer Law & Security Review, 53*, 1–33. https://doi.org/10.1016/j.clsr.2024.105957

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury, CA: SAGE Publications. https://www.google.com/books/edition/Naturalistic_Inquiry/2oA9aWlNeooC?hl=en

https://doi.org/10.65476/zxewbx77

Loui, M., & Miller, K. W. (2008). *Ethics and professional responsibility in computing* (B. W. Wah, Ed.). New Jersey, NJ: Wiley Encyclopedia of Computer Science and Engineering. https://doi.org/10.1002/9780470050118.ecse909

Lyon, D. (2018). *The culture of surveillance: Watching as a way of life*. Cambridge, UK: Polity Press. https://www.politybooks.com/bookdetail?book_slug=the-culture-of-surveillance-watching-as-a-way-of-life--9780745671727

Marsh, O., & Podgorsek, E. (2026). A guide to the Digital Services Act, the EU's law to rein in Big Tech. *AlgorithmWatch*. https://algorithmwatch.org/en/dsa-explained/

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology, 6*(3), 175–183. https://doi.org/10.1007/s10676-004-3422-1

McGregor, L., Murray, D., & Ng, V. (2019). International human rights law as a framework for algorithmic accountability. *International & Comparative Law Quarterly, 68*(2), 309–343. https://doi.org/10.1017/S0020589319000046

Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application, 8*, 141–163. https://doi.org/10.1146/ANNUREV-STATISTICS-042720-125902/CITE/REFWORKS

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society, 3*(2), 1–21. https://doi.org/10.1177/2053951716679679

Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems, 21*(4), 18–21. https://doi.org/10.1109/MIS.2006.80

Morozov, E. (2014). *To save everything, click here: The folly of technological solutionism.* New York, NY: PublicAffairs. https://www.hachettebookgroup.com/titles/evgeny-morozov/to-save-everything-click-here/9781610393706/?lens=publicaffairs

National Institute of Standards and Technology. (2019). U.S. leadership in AI: A plan for federal engagement in developing AI technical standards and related tools. Gaithersburg, MD: U.S. Department of Commerce. https://www.nist.gov/artificial-intelligence/plan-federal-engagement-developing-ai-technical-standards-and-related-tools

Noble, S. U. (2018). Algorithms of oppression. New York, NY: New York University Press. https://doi.org/10.18574/NYU/9781479833641.001.0001

Nyholm, S. (2018). Attributing agency to automated systems: Reflections on human–robot collaborations and responsibility-loci. *Science and Engineering Ethics, 24*(4), 1201–1219. https://doi.org/10.1007/S11948-017-9943-X/METRICS

O'Neill, O. (2018). Linking trust to trustworthiness. *International Journal of Philosophical Studies, 26*(2), 293–300. https://doi.org/10.1080/09672559.2018.1454637

Organisation for Economic Co-operation and Development. (2021). *The OECD.AI policy navigator*. OECD.AI. https://oecd.ai/en/dashboards/overview

Pasquale, F. (2016). *Black box society: The secret algorithms that control money and information*. https://www.hup.harvard.edu/books/9780674970847

Perrigo, B. (2023, January 18). Exclusive: Open AI used Kenyan workers on less than $2 per hour to make ChatGPT less toxic. *Time*. https://time.com/6247678/openai-chatgpt-kenya-workers/

Powers, T. M. (2006). Prospects for a Kantian machine. *IEEE Intelligent Systems, 21*(4), 46–51. https://doi.org/10.1109/MIS.2006.77

Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts, COM (2021) 206 final. (2021). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/S42256-019-0048-X

Schreier, M. (2012). *Qualitative content analysis in practice*. London, UK: *Sage Publications*. https://doi.org/10.4135/9781529682571

Schultz, M. D., Conti, L. G., & Seele, P. (2024). Digital ethicswashing: A systematic review and a process-perception-outcome framework. *AI and Ethics, 5*(2), 805–818. https://doi.org/10.1007/S43681-024-00430-9

Schultz, W., & Held, T. (2004). *Regulated self-regulation as a form of modern government*. Eastleigh, UK: John Libbey Publishing.

Schwartz, W. (2023). *The EU's Digital Services Act confronts silicon valley*. Wilson Center. https://www.wilsoncenter.org/article/eus-digital-services-act-confronts-silicon-valley

Sebastião, S. P., & Dias, D. F. (2025). AI transparency: A conceptual, normative, and practical frame analysis. *Media and Communication, 13*(0), 1–19. https://doi.org/10.17645/MAC.9419

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction, 36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

https://doi.org/10.65476/zxewbx77

Solove, D. J. (2011). *Understanding privacy*. London, UK: Harvard University Press. https://fpf.org/wp-content/uploads/2021/05/Understanding-Privacy-CH1b.pdf

Stahl, B. C. (2021). From computer ethics and the ethics of AI towards an ethics of digital ecosystems. *AI and Ethics, 2*(1), 65–77. https://doi.org/10.1007/S43681-021-00080-1

Ulgen, O. (2017). Kantian ethics in the age of artificial intelligence and robotics. *Questions of International Law, 1*(43), 59–83. https://www.qil-qdi.org/kantian-ethics-age-artificial-intelligence-robotics/

UN Advisory Body on Artificial Intelligence. (2024). Governing AI for humanity: Final report. *United Nations.* https://digitallibrary.un.org/record/4062495?v=pdf

United Nations Commission on International Trade Law. (1996). *UNCITRAL model law on electronic commerce (1996) with additional article 5 bis as adopted in 1998.* United Nations Commission on International Trade Law. https://uncitral.un.org/en/texts/ecommerce/modellaw/electronic_commerce

United Nations Commission on International Trade Law. (2001). *UNCITRAL model maw on electronic signatures (2001)*. United Nations Commission on International Trade Law. https://uncitral.un.org/en/texts/ecommerce/modellaw/electronic_signatures

UNESCO. (2021). *Recommendation on the ethics of Artificial Intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000380455

Wachter, S. (2024). Limitations and loopholes in the EU AI Act and AI liability directives: What this means for the European Union, the United States, and beyond. *Yale Journal of Law & Technology, 26*(3). https://doi.org/10.2139/ssrn.4924553

Wachter, S., & Mittelstadt, B. (2018, October 9). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Oxford Business Law Blog*. https://blogs.law.ox.ac.uk/business-law-blog/blog/2018/10/right-reasonable-inferences-re-thinking-data-protection-law-age-big

Webster, G., Creemers, R., Kania, E., & Triolo, P. (2017, August 1). Full translation: China's 'new generation artificial intelligence development plan.' *DigiChina*. https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation and Governance, 12*(4), 505–523. https://doi.org/10.1111/REGO.12158