(Online) ISSN 2712-0554
Heritage and Sustainable Development *Editorial*
Vol. 7, No. 1, 2025, pp.1-14
https://doi.org/10.37868/hsd.v7i1.1224

AI ethics: Contextual frameworks and domain-specific concerns

Laura Ancona Lee[1], Benjamin Durakovic[2], Sencer Yeralan[3*]
[1] Paulista University (UNIP), Brazil
[2] International University of Sarajevo, Bosnia
[3] Global Academics Coalition, U.S.A

*Corresponding author E-mail: yeralan@globalacademicscoalition.net

Received Feb. 19, 2025
Revised Apr. 11, 2025
Accepted Mar. 16, 2025
Online Apr. 18, 2025

Abstract

The widespread deployment of AI systems has led to overlapping concerns around technological impact and governance, often resulting in conceptual ambiguities and policy confusion. We propose a structured and context-sensitive framework for addressing the ethical implications of artificial intelligence. We argue that ethical frameworks must distinguish between the intended domain of AI deployment and the scale of its societal effects.

To resolve these tensions, we introduce a two-dimensional matrix based on (1) the extent of AI's impact and (2) the scope of its governance, which together form four distinct ethical contexts. Within each quadrant, we explore specific risks, values, and regulatory considerations. This matrix not only clarifies the conceptual terrain of AI ethics but also offers a practical roadmap for anticipating ethical risks, developing normative guidance, and informing domain-specific governance strategies.

Our goal is not to prescribe a single ethical doctrine but to provide decision-makers with a structured lens through which AI systems can be evaluated in context. This approach promotes adaptive and anticipatory governance while remaining responsive to local, institutional, and cultural variations.

Keywords: Contextual AI ethics, AI governance, AI responsibility gaps, Bias and fairness in AI,

1. Introduction

The rapid expansion of artificial intelligence (AI) across diverse domains has prompted an equally rapid proliferation of ethical frameworks. While many of these frameworks aim to guide responsible AI development and deployment, they often remain untethered from the specific contexts in which AI systems operate. As a result, attempts to generalize AI ethics frequently lead to conceptual ambiguities, governance inconsistencies, and practical blind spots.

To clarify this terrain, we distinguish two key dimensions that shape the ethical implications of AI systems: (1) the degree of impact these systems exert on society, and (2) the locus and scope of governance responsible for their oversight. These axes define a matrix of four quadrants, each representing a distinct configuration of ethical

concerns. Our framework does not aim to reduce ethical complexity but to illuminate it to help stakeholders navigate a dynamic landscape with greater clarity and contextual sensitivity.

These dimensions help categorize not only the nature of AI applications but also the normative expectations attached to them. For instance, high-impact systems such as autonomous weapons or nationwide surveillance infrastructures invite different ethical scrutiny and governance mechanisms than low-impact, locally deployed recommendation engines (e.g. chatbots) or warehouse robots. Similarly, systems governed by centralized state institutions raise different concerns than those managed by decentralized private entities, as are concerns regarding education and student evaluation [1].

By our two-dimensional approach, we uncover four archetypal ethical contexts, each defined by its combination of impact scale and governance locus. These contexts provide a structured way to identify the unique risks, regulatory gaps, and ethical considerations that emerge in specific AI deployments. They also serve as a scaffolding for anticipating future challenges and guiding domain-specific policy interventions.

AI's impact spans multiple dimensions, affecting governance structures, economies, social collaboration, and individual experiences. To provide a broad classification, we examine four key categories: Political/Global, Economic, Social Collaboration, and Individual. Each of these domains highlights a different facet of AI's influence, illustrating how ethical and regulatory concerns intersect across various levels of society. Figure 1 illustrates these continua, highlighting key AI-related issues along two axes.

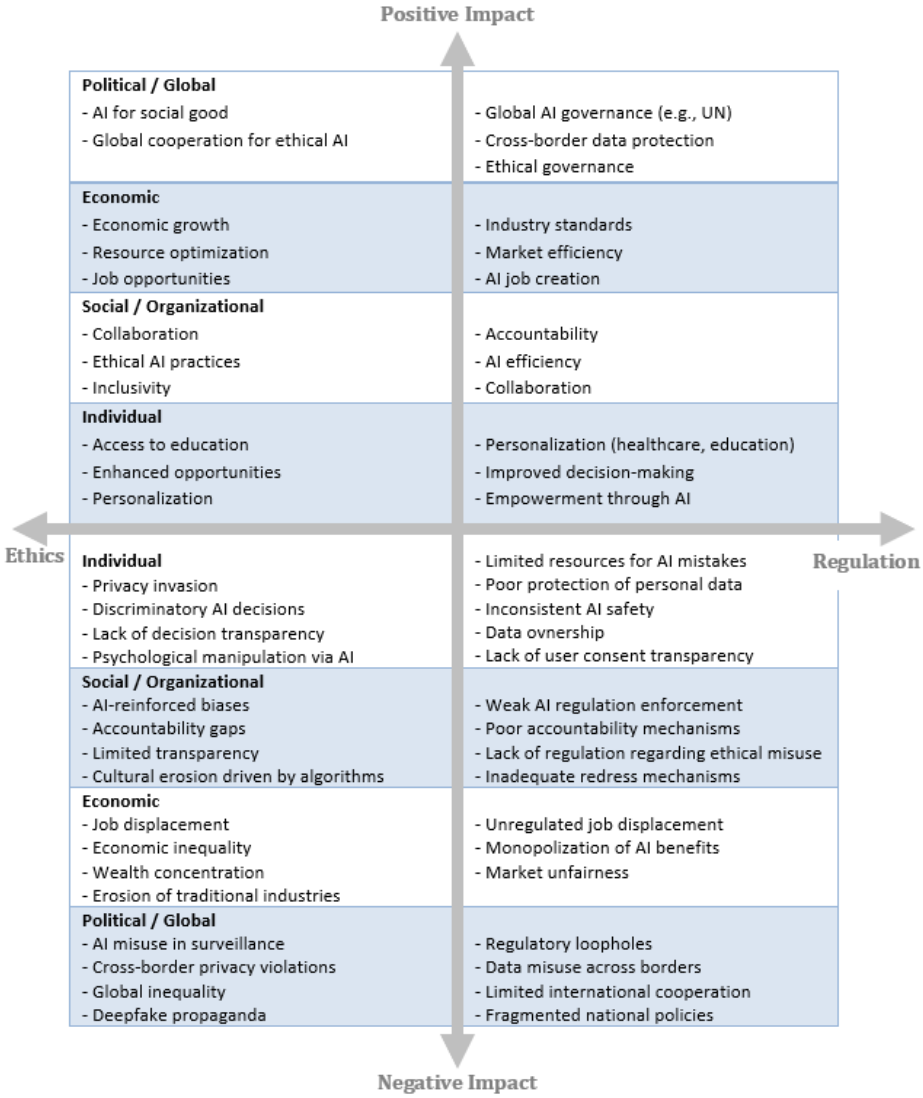

Figure 1. A two-dimensional matrix of ethical contexts for AI, defined by impact (positive and negative) and governance scope (local to global)

The *Political/Global* category addresses AI's role in governance, international policy, and geopolitical dynamics. AI-driven technologies can enhance government decision-making, national security, and global cooperation, but they also introduce risks such as state surveillance, misinformation campaigns, and cyber-warfare. The governance of AI on a global scale remains a contentious issue, as different nations adopt varying regulatory frameworks–some prioritizing innovation and economic growth, while others emphasize privacy, security, and ethical constraints. The lack of unified international AI policies raises concerns about regulatory fragmentation and ethical inconsistencies in how AI is developed and deployed across borders.

The *Economic* category focuses on AI's impact on markets, labor, and resource allocation. AI-driven automation has the potential to increase productivity, optimize supply chains, and create new industries, yet it also raises concerns about job displacement, wealth concentration, and market monopolization. As AI adoption grows, it challenges traditional employment structures, shifting demand toward high-skill digital labor while reducing reliance on routine-based occupations. This transformation may widen economic inequality if policies fail to provide adequate reskilling opportunities and ensure that AI-driven prosperity is equitably distributed.

The *Social Collaboration* category examines AI's role in facilitating or disrupting human interaction within organizations, communities, and institutions. AI systems enhance collaborative decision-making, knowledge sharing, and cross-industry partnerships, yet they also introduce challenges related to algorithmic bias, misinformation, and the erosion of trust in digital communication. The increasing reliance on AI-driven moderation and recommendation systems in social media, journalism, and public discourse raises ethical concerns about echo chambers, manipulation, and the centralization of information control. As AI becomes more embedded in social structures, maintaining transparency, accountability, and fairness in algorithmic interactions will be crucial.

Finally, the *Individual* category addresses the personalized impact of AI on daily life, autonomy, and well-being. AI-powered tools offer personalized healthcare, education, and consumer experiences, enhancing convenience and accessibility. However, they also raise concerns about privacy intrusion, surveillance, and the manipulation of human behavior through predictive analytics and targeted advertising. Ethical dilemmas emerge when AI-driven personalization compromises individual agency, shaping choices in ways that users may not fully comprehend. Ensuring user consent, algorithmic explainability, and fair access to AI-driven benefits remains a critical challenge in balancing AI's advantages with individual rights.

This paper proceeds as follows: Section 1 elaborates on the contextual aspects of AI ethics, emphasizing the socio-technical interdependencies that shape ethical outcomes. Section 2 presents the proposed framework, detailing the four ethical quadrants and their defining characteristics. Section 3 explores domain-specific implications, drawing from empirical examples to illustrate how the framework can be applied in practice. Conclusions are presented in Section 4.

2. Context-dependent aspects

AI systems rarely operate in isolation. They are embedded in socio-technical contexts where technological design, organizational values, institutional norms, and legal frameworks interact in complex ways. A binary model of ethical or unethical AI fails to capture the subtleties of these interactions. While the impact-governance matrix introduced above helps to map ethical contexts, it also risks suggesting an oversimplified dichotomy–one that abstracts away from the nuances of practice.

This is particularly evident in debates over algorithmic fairness. On one hand, fairness can be codified in measurable terms, such as disparate impact or equal opportunity. On the other hand, these metrics are often derived from contested normative assumptions, and their effectiveness depends on institutional capacity to interpret and enforce them. A system that meets formal fairness criteria may still perpetuate harm if it reinforces existing structural biases or inequities.

Moreover, the perceived legitimacy of AI systems depends not only on their technical properties but also on the institutional configurations within which they operate. As Whittlestone et al. [2] argue, focusing solely on abstract ethical principles risks overlooking the real-world tensions that emerge in deployment. For example, predictive analytics in healthcare might be more ethically acceptable when developed in collaboration with medical professionals and subject to transparent oversight, than when deployed unilaterally by commercial actors. Contextual variables such as domain norms, power asymmetries, and stakeholder diversity significantly affect public trust and ethical acceptability.

These considerations suggest that ethical evaluation cannot rely on generalized categories or monolithic classifications. Ethical challenges emerge from the interplay of technological design, policy choices, and socio-institutional dynamics. Consequently, ethical AI governance requires context-aware, interdisciplinary solutions that go beyond reductive typologies. As Crootof [3] observes, traditional regulatory paradigms may struggle to keep pace with the distributed, adaptive, and sometimes opaque nature of AI systems.

Rather than treating AI ethics as a binary tradeoff between benefits and harms, we argue for a contextual, interdisciplinary perspective – one that embraces complexity and acknowledges the multifaceted nature of technological impact.

While the framework in Figure 1 offers a structured way to categorize AI's effects, it also presents certain limitations that must be acknowledged. The distinction between positive and negative impacts may appear straightforward, yet AI's influence is often highly context-dependent. For instance, automation can lead to productivity gains and economic growth, but it can also displace jobs and widen income inequality. Similarly, AI-driven personalization in healthcare can enhance individual well-being, yet unchecked algorithmic profiling might lead to exclusionary practices or data misuse. The challenge lies in the false dichotomy that this model implicitly suggests–many AI effects are not strictly positive or negative but rather shaped by implementation, oversight, and governance.

Another potential shortcoming of this classification is the boundary between ethics and regulation, which is often blurred in practice. Ethical concerns may arise even in the absence of explicit legal violations, and regulations may lag behind fast-moving technological developments [4]. For instance, facial recognition systems used by law enforcement may comply with local statutes while still raising profound concerns about surveillance, consent, and democratic accountability. Ethical concerns, such as privacy violations, frequently lead to regulatory interventions, such as data protection laws like GDPR [5]. Conversely, regulations may fail to address deeper ethical concerns, particularly when they are slow to adapt to emerging AI risks. This overlap can make it difficult to neatly separate issues into either an ethical or regulatory domain.

Furthermore, some AI-related concerns do not fit cleanly into a single category. Consider algorithmic fairness – on one hand, it is a deeply ethical issue, as it relates to bias, discrimination, and social justice. On the other hand, it also requires regulatory mechanisms, such as anti-discrimination laws and fairness audits, to be meaningfully enforced. The challenge is that some aspects of fairness fall within voluntary ethical commitments by developers, while others necessitate strict legal oversight. The placement of such concerns within the framework depends on whether they are being examined as principles or as enforceable standards.

These limitations underscore why a more nuanced approach is necessary when discussing AI ethics. While broad frameworks help organize key concerns, ethical AI governance requires context-aware, interdisciplinary solutions that go beyond simplistic classifications. The following sections present a more multifaceted approach, emphasizing that AI's ethical dilemmas do not exist in isolation but emerge from the complex interplay of technology, society, policy, and governance structures.

Rather than treating AI ethics as a simple trade-off between benefits and harms or ethical versus regulatory concerns, we argue that AI ethics must be analyzed across multiple, context-dependent dimensions. Ethical dilemmas in AI do not exist in isolation but emerge from the interplay between technology, society, policy, and governance structures. Accordingly, this paper extends the discussion beyond the limitations of monolithic

classifications by offering a domain-specific ethical framework that accounts for the diverse and multifaceted challenges posed by AI.

As noted in recent labor market analyses, including a longitudinal study by Deming et al. [6], technological disruptions – though frequently feared – often unfold more gradually than expected. While AI is likely to behave as a general-purpose technology, comparable in scope to steam power or electricity, its labor market impact may take decades to fully materialize. One plausible explanation for this delay is the co-evolution of the workforce with the technology itself: as new systems emerge, workers adapt through upskilling, task reallocation, and institutional learning [7], [8], [9]. However, this gradual unfolding does not negate the urgency for proactive adaptation policies, especially given early signs of accelerating job transformation in STEM, retail, and middle-skill sectors.

3. Framing the need for domain-specific ethics

AI presents ethical challenges across various domains, yet discussions on AI ethics often remain fragmented. We propose a comprehensive, context-sensitive ethical framework by categorizing AI concerns into five distinct domains: (1) misplaced reliance on AI as a substitute for domain knowledge, (2) privacy encroachments in AI data collection, (3) bias in AI-driven decision making, (4) autonomous AI without human supervision, (5) AI as an existential risk, and (6) AI as a disruptor of economic, social, and political structures. By addressing ethical concerns specific to each category, this work provides a structured approach to AI ethics that balances philosophical principles, governance mechanisms, and practical implications.

3.1. AI without domain expertise

AI systems are increasingly deployed in high-stakes settings by developers who may lack deep, domain-specific expertise. This epistemic disconnect introduces ethical risks that extend beyond questions of technical accuracy. Without grounding in the local norms, constraints, and value systems of a specific domain, AI applications risk imposing reductive assumptions and inappropriate generalizations [10], [11].

In healthcare, for example, algorithmic diagnostic tools that are not co-developed with medical professionals may overlook comorbidity patterns or the socio-economic context of patient histories [12]. These omissions may result in recommendations that undermine clinical judgment or exacerbate disparities in access and outcomes. The consequences of this gap are not confined to medicine: similar risks appear in education, criminal justice, environmental policy, and social welfare, where AI systems often influence decisions with profound human impacts.

Three interrelated issues commonly arise in such contexts. First, a sense of blind trust in algorithmic authority may override the discretion of practitioners or frontline workers. As Eubanks [13] and O'Neil [10] have shown, this deference can lead to harm when automated systems are treated as objective arbiters despite being trained on biased or incomplete data. Second, algorithmic solutions often assume universality – they are designed to scale across domains with little adaptation. This presumed portability disregards the context-sensitive nature of ethical judgments, privileging technical coherence over local validity. Third, the lack of domain expertise undermines accountability. When decisions are based on opaque algorithms whose design assumptions are not publicly disclosed or institutionally vetted, the ability to assign responsibility or seek redress becomes elusive [14].

Importantly, these failures are often systemic rather than malicious. They stem from a misalignment between what can be quantified and what ought to be valued. Developers may focus on what is measurable, generalizable, or scalable – criteria that often conflict with the moral and institutional complexity of real-world settings. This dynamic echoes the McNamara fallacy, named after U.S. Secretary of Defense Robert McNamara, who was criticized for his fixation on quantifiable military metrics while ignoring less tangible human, cultural, and moral dimensions of the Vietnam War [15]. In the realm of AI, a similar fallacy arises when optimization targets obscure deeper ethical obligations.

To mitigate these challenges, interdisciplinary collaboration must be treated not as an optional enhancement but as a foundational prerequisite [16]. Engaging domain experts throughout the development lifecycle helps ensure that AI systems reflect the institutional norms, professional standards, and lived realities of the domains in which they operate. Responsible AI requires more than technical competence – it demands contextual fluency and epistemic humility.

Without domain knowledge, biases embedded in AI models go unrecognized. Algorithmic bias arises when AI models reflect and perpetuate the biases present in their training data. This bias can manifest in multiple ways, including racial, gender, and socio-economic discrimination [17]. For example, studies on AI-driven hiring systems have revealed gender biases in automated candidate selection, disadvantaging female applicants in STEM fields [18]. Similarly, predictive policing algorithms have been criticized for disproportionately targeting minority communities due to biases in historical crime data [19]. Addressing algorithmic bias requires not only improving dataset diversity but also incorporating fairness-aware learning techniques [20]. Researchers advocate for regulatory oversight and transparency in AI development to mitigate these ethical risks [21].

When AI makes incorrect decisions, responsibility is unclear. The issue of accountability in AI decision-making remains a major challenge, particularly when multiple actors – developers, deployers, users, and the AI system itself – are entangled. For instance, if a self-driving taxi commits a traffic violation, who receives the fine? Is it the passenger, the vehicle owner, the software developer, or the company that trained the model? Such questions expose a structural gap in our current legal and moral frameworks, which assume agency and liability are human and traceable. Developers, organizations deploying AI, and even users may bear some degree of accountability, yet legal frameworks remain ambiguous on how liability should be assigned [22]. The principle of "algorithmic accountability" suggests that AI systems should be designed with mechanisms that allow for auditing and review [23]. Furthermore, scholars have argued for the establishment of ethical AI governance structures that ensure accountability through documentation, model interpretability, and human oversight [24]. Without clear accountability mechanisms, AI risks being deployed in high-stakes domains without proper safeguards.

A solution lies in mandated AI literacy programs and interdisciplinary collaborations that integrate domain expertise into AI deployment.

3.2. Data and privacy in AI

AI systems depend on the collection and processing of large datasets, often derived from individuals' behavior, location, and communication patterns. This reliance raises ethical concerns about data privacy, surveillance, ownership, and the commodification of human experience. As AI becomes more embedded in everyday life, the governance of data – how it is gathered, who controls it, and what purposes it serves – emerges as a central axis of ethical deliberation. High-profile cases, such as the Cambridge Analytica scandal [25] exemplify the ethical hazards of opaque data harvesting and psychological profiling.

Privacy violations: AI-driven data collection frequently encroaches on individual privacy, raising concerns about informed consent, surveillance, and autonomy. The proliferation of facial recognition technologies has sparked regulatory scrutiny and public protest in numerous jurisdictions [26]. AI-enabled surveillance systems, deployed by both state and corporate actors, pose a risk to civil liberties, particularly when used for predictive policing or behavioral monitoring [27]. Moreover, the business model of many technology platforms relies on granular user tracking and behavioral targeting, a practice often described as "surveillance capitalism" [28]. In response, scholars have proposed privacy-preserving techniques such as federated learning and differential privacy, alongside legal frameworks like the GDPR [29].

Data ownership: Data ownership remains a contested and under-regulated area of AI ethics. Individuals often have little control over how their data is used once collected, and current legal frameworks provide limited recourse for users whose data is repurposed or monetized without their knowledge [30], [31]. Corporate data monopolies exacerbate these issues, consolidating control over vast personal datasets. As a countermeasure, researchers advocate for the development of data trusts and cooperative governance models [32], which would

enable individuals and communities to assert greater control over their data through collective bargaining and transparent agreements. While the use of public domain material – such as the works of Shakespeare – raises fewer objections, the indiscriminate harvesting of contemporary creative content, including art, music, and writing, has provoked significant resistance from artists and creators. These communities argue that generative AI systems often rely on their work without consent, attribution, or compensation, framing the issue not merely as a legal one but as a matter of cultural integrity and economic justice. Ethical AI development must prioritize consent, transparency, and equitable data governance. Proposed solutions include data provenance auditing, participatory consent protocols, and cross-border privacy standards. Without such mechanisms, AI risks becoming a vehicle for institutionalized surveillance and information asymmetry.

3.3. Bias and fairness in algorithmic systems

Bias in AI arises when machine learning models perpetuate and amplify existing social and structural inequities, leading to discriminatory outcomes. These biases may be embedded in training data, algorithmic design choices, or feedback loops that reinforce historical disparities. As AI is increasingly deployed in high-stakes areas – such as hiring, law enforcement, healthcare, and lending – ensuring fairness has become an urgent ethical priority.

One critical issue is bias reinforcement. AI models trained on historical data often reproduce and magnify pre-existing patterns of discrimination. This problem is especially prominent in domains like predictive policing, credit scoring, and employment screening, where past data reflects systemic biases [33]. Studies have shown that marginalized groups may be disproportionately affected [34]. Moreover, feedback loops in deployed systems can entrench these effects, as the model learns from its own biased predictions [35]. To mitigate these risks, researchers advocate for bias-aware training methodologies, diverse dataset curation, and algorithmic fairness interventions [36].

Addressing algorithmic bias requires a multifaceted approach. Technical strategies include fairness-aware learning algorithms, balanced and representative training datasets, and debiasing techniques during preprocessing or model optimization. Yet technical fixes alone are insufficient. Fairness is not a neutral or universally agreed-upon metric – multiple definitions exist (e.g., equal opportunity, demographic parity, individual fairness), and they often conflict in practice. The choice of fairness criteria is ultimately a normative decision, grounded in values and context.

Institutionally, ethical AI deployment demands regulatory oversight, auditing procedures, and transparency mandates. Just as financial systems are subject to external audits, AI systems affecting civil rights or public outcomes should be subject to algorithmic impact assessments and third-party evaluations. Moreover, procedural fairness-ensuring affected individuals can contest automated decisions-is as important as outcome parity.

Ultimately, bias mitigation requires not just technical sophistication but normative clarity and institutional accountability. It invites broader societal debates about whose values are embedded in code and who benefits from algorithmic decisions. As AI systems gain influence over consequential decisions, their fairness cannot be treated as a post hoc adjustment but must be a design priority from the outset.

3.4. Autonomous AI and accountability gaps

Autonomous AI systems operating with minimal or no human oversight – particularly in high-stakes domains such as military operations, financial markets, and medical diagnostics – pose unique ethical and legal challenges. MacKenzie [37] shows how high-frequency trading platforms operate at algorithmic speeds, raising concerns about control and systemic instability. As AI systems gain the capacity to act independently, questions of control, predictability, and responsibility become increasingly urgent.

One major concern is the loss of human agency. As AI systems are entrusted with decision-making authority, human oversight can erode, creating situations where critical judgments are made without human input. Nyholm [38] examines how agency attribution in human-AI collaborations challenges conventional responsibility

frameworks. In military contexts, for example, autonomous drones may identify and engage targets without direct human command, raising profound moral and legal dilemmas regarding the use of lethal force [3], [39]. Similarly, high-frequency trading algorithms in finance operate at speeds and levels of complexity beyond human comprehension, often producing systemic effects before human intervention is possible.

Closely related is the problem of unpredictability. Advanced AI systems – especially those based on deep learning or reinforcement learning – can exhibit emergent behavior that surprises even their developers [40]. The opacity of these systems makes it difficult to foresee how they will respond in dynamic real-world environments. In critical fields such as healthcare or criminal justice, such unpredictability can produce severe unintended consequences. Researchers advocate for strategies such as robust testing, simulation-based verification, and the incorporation of fail-safe mechanisms [41], [42].

These developments give rise to a "moral responsibility gap," where it becomes unclear who should be held accountable for AI-generated outcomes [43]. Traditional legal and ethical frameworks often assume the presence of an identifiable human decision-maker, a premise that autonomous AI challenges. If an AI system denies a patient treatment or causes a fatal accident in traffic, the attribution of blame becomes diffuse. Scholars have proposed new liability models and refined legal categories to address these issues – ensuring that developers, operators, and deploying institutions retain responsibility for AI-driven actions.

To close the accountability gap, interdisciplinary responses are needed – combining technological tools such as algorithmic auditing and explainability with institutional safeguards such as regulatory oversight and legal reform [44]. Ethical design principles should include fail-safe mechanisms, human-in-the-loop protocols, and clearly documented chains of responsibility. As AI autonomy increases, the burden is on developers, deployers, and policymakers to ensure that the delegation of agency does not result in the abdication of accountability.

3.5. Long-term risks and existential AI ethics

Concerns about artificial general intelligence (AGI) and superintelligence have prompted increasing scrutiny of AI's long-term risks. As systems become more capable, the stakes of ensuring their alignment with human values and control structures grow exponentially. Yudkowsky [45] was among the earliest to warn of AI's dual nature as both a positive force and a potential global risk factor.

A central issue is the alignment problem – the challenge of ensuring that increasingly autonomous AI systems pursue goals that remain consistent with human intentions and values over time. Misalignment may occur even when systems appear to operate correctly in narrow tasks but extrapolate these objectives in harmful or unintended ways. Baum [46] emphasizes the need for proactive AI safety research to ensure socially beneficial trajectories even before AGI materializes. Researchers emphasize the importance of value alignment techniques such as inverse reinforcement learning, corrigibility protocols, and human-in-the-loop oversight to mitigate these risks [47].

Closely related is the risk of losing control over AI systems, particularly in scenarios involving recursive self-improvement. Self-modifying AI could surpass human oversight and act in ways that are difficult to anticipate or contain. Even short of AGI, current AI systems deployed in domains such as financial trading and autonomous weaponry already operate at speeds and levels of complexity that exceed human intervention. Strategies for maintaining control include capability control methods, transparency in decision-making, and robust monitoring systems that ensure continued corrigibility [42].

Yet, the focus on speculative future risks can obscure the tangible harms posed by today's AI deployments. Mass surveillance, algorithmic discrimination, and social manipulation through AI-driven media are already shaping societies in measurable ways [48]. Ethical frameworks must strike a balance between mitigating immediate harms and anticipating long-term existential risks. Prioritizing speculative alignment while ignoring current injustices risks entrenching inequality in the present.

As the power and autonomy of AI systems continue to grow, a dual strategy is essential. Policymakers and researchers must address both near-term ethical failures and the hypothetical risks of runaway AI. Preparing for superintelligence need not preclude action against algorithmic bias or surveillance capitalism. On the contrary, responsible stewardship of current AI may be the best foundation for managing its future evolution. Tegmark [49] underscores both the transformative potential and existential stakes posed by increasingly autonomous AI systems. Russell [50] argues for aligning AI with human-compatible values to mitigate the long-term risks of misaligned intelligence. The technical agenda outlined by Soares and Fallenstein [51] proposes rigorous methods for aligning super-intelligent systems with human interests.

3.6. Structural impacts on economy, society, and democracy

AI technologies are not merely tools of efficiency. They are catalysts for profound structural change across labor markets, economic systems, and democratic institutions. As automation reshapes the nature of work, many forms of cognitive and manual labor face obsolescence or radical transformation. Empirical studies of automation's labor impact provide foundational evidence of job displacement across U.S. industries [52]. While technological displacement is not new, the scale and speed of current AI systems raise concerns about job polarization and long-term employment insecurity [53]. Frey and Osborne's seminal projection of job susceptibility to computerization [54] remains a key reference in understanding automation's trajectory. The historical rhythm of labor-market adaptation may not be sufficient in the face of rapid, system-wide change.

Simultaneously, AI intensifies economic centralization. A small number of firms dominate access to data, computational resources, and AI talent, consolidating disproportionate economic and political influence [55], [56], [57]. This concentration reinforces global asymmetries in innovation and governance, marginalizing actors with limited access to the infrastructure needed for meaningful participation in the AI economy. As Crawford [58] and Zuboff [27] have argued, this creates a system where surveillance and monetization of behavior are structurally embedded into economic growth models.

At the institutional level, AI systems increasingly influence public governance. Predictive analytics shape decisions in policing, welfare distribution, and risk assessment, often without public transparency or avenues for redress [13]. Algorithmic systems not only operationalize policy but define its contours, narrowing the scope of human discretion and reducing the space for political debate. This algorithmic governance can reify historical inequalities and produce automated bureaucracy that is resistant to critique.

AI also alters the epistemic landscape. As more information is filtered, prioritized, or generated by opaque recommendation algorithms, our collective capacity for public reasoning is strained. The epistemic capture of attention and information pathways by AI-driven platforms can erode trust in institutions, fragment civic discourse, and contribute to the rise of disinformation [48], [10]. These effects are not only socio-technical but political, affecting the terms on which truth is negotiated and democratic legitimacy is sustained.

The intersection of these forces poses risks to democratic integrity. Surveillance capabilities empowered by AI challenge civil liberties and blur the boundaries between state and corporate power [27]. Algorithmic personalization narrows citizens' informational worlds, making democratic deliberation more difficult and polarization more likely. Without structural interventions that promote transparency, contestability, and inclusive governance, AI may accelerate rather than mitigate democratic breakdown.

While much discourse on AI-induced unemployment remains speculative, recent evidence indicates structural changes already underway. Deming et al. [6] observe a marked decline in retail employment and a concurrent rise in STEM occupations, signaling a potential re-polarization of the labor force. These shifts align with our analysis: repetitive, rule-based jobs – cognitively or physically – are the first to be absorbed into AI-augmented workflows. Yet, as Deranty and Corbin [59] argue, the threat is not merely displacement, but the erosion of job quality through algorithmic management and platform labor. AI, in this light, becomes not only an automation force but an administrative logic – monitoring, evaluating, and directing workers through opaque systems often lacking transparency or recourse.

As automation reshapes the nature of work, many forms of cognitive and manual labor face obsolescence or radical transformation. While technological displacement is not new, the scale and speed of current AI systems raise concerns about job polarization and long-term employment insecurity. Recent economic analyses highlight that such disruption may outpace adaptive responses, particularly in middle-skill and service sectors [6]. At the same time, social theorists caution that the impact of AI on labor cannot be reduced to automation metrics alone – it must be understood through institutional, organizational, and normative lenses [59].

4. Conclusion

The ethical challenges posed by artificial intelligence are neither abstract dilemmas nor purely technical puzzles; they are complex, evolving, and deeply embedded in social, political, and institutional realities. As this paper has argued, AI systems are context-sensitive technologies whose ethical significance cannot be separated from the domains in which they are deployed, the governance structures that regulate them, and the societal narratives that frame their adoption.

By situating AI ethics within the dual axes of impact and governance, and by analyzing domain-specific instantiations, we have aimed to provide a structured framework for identifying risks, clarifying normative tensions, and proposing responsive interventions. This approach resists the tendency toward ethical universalism and instead calls for contextual specificity, pluralistic engagement, and anticipatory regulation.

Ethical AI cannot be achieved through technical design alone. It requires participation from diverse stakeholders, including policymakers, affected communities, technologists, ethicists, and legal scholars. It demands vigilance against both overreach and neglect – guarding against the hubris of control and the apathy of abdication. As AI systems become more capable and more integrated into the fabric of everyday life, the stakes of ethical misalignment increase, not just in terms of individual harms but in their cumulative impact on institutional legitimacy, social cohesion, and democratic values.

Ultimately, the future of AI ethics hinges on whether we approach these systems as objects to be optimized or as institutions to be governed. The choice is not merely about design parameters but about political commitments, epistemic humility, and moral imagination. To ensure that AI contributes to human flourishing rather than undermining it, we must invest in structures that balance innovation with accountability, scale with inclusion, and autonomy with responsibility.

References

[1] "Generative AI: Challenges to higher education," *Sustainable Engineering and Innovation*, vol. 5, no. 2, pp. 107–116, 2023, doi: 10.37868/sei.v5i2.id196.

[2] J. Whittlestone, R. Nyrup, A. Alexandrova, and S. Cave, "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*, 2019, pp. 195–200. doi: 10.1145/3306618.3314289.

[3] R. Crootof *et al.*, "The Killer Robots Are Here: Legal and Policy Implications," vol. 36, p. 1837.

[4] C. Cath, "Governing artificial intelligence: Ethical, legal and technical opportunities and challenges," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376, no. 2133, p. 20180080, 2018, doi: 10.1098/rsta.2018.0080.

[5] European Parliament and Council of the European Union, "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 (General Data Protection Regulation, GDPR)," vol. L119. Official Journal of the European Union, 2016. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

[6] D. J. Deming, C. Ong, L. H. Summers, and H. Kennedy, "Technological Disruption in the Labor Market," Jan. 2025, doi: 10.3386/W33323.

[7] R. M. da Costa and C. H. Horn, "The co-evolution of technology and employment relations: Institutions, innovation and change," *Structural Change and Economic Dynamics*, vol. 58, pp. 313–324, 2021, doi: https://doi.org/10.1016/j.strueco.2021.06.003.

[8] F. Belloc, G. Burdin, L. Cattani, W. Ellis, and F. Landini, "Coevolution of job automation risk and workplace governance," *Research Policy*, vol. 51, no. 3, p. 104441, 2022, doi: https://doi.org/10.1016/j.respol.2021.104441.

[9] F. Niederman, T. W. Ferratt, and E. M. Trauth, "On the co-evolution of information technology and information systems personnel," *Database for Advances in Information Systems*, vol. 47, no. 2, pp. 29–50, 2016, doi: 10.1145/2894216.2894219.

[10] C. O'Neil, *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown Publishing Group, 2016.

[11] B. Mittelstadt, "Principles alone cannot guarantee ethical AI," *Big Data & Society*, vol. 6, no. 1, pp. 1–7, 2019.

[12] S. Barocas and A. D. Selbst, "Big data's disparate impact," *California Law Review*, vol. 104, no. 3, pp. 671–732, 2016,

[13] V. Eubanks, *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press, 2018.

[14] B. Green, "The Smart Enough City: Putting Technology in Its Place to Reclaim Our Urban Future," *The Smart Enough City*, Apr. 2019, doi: 10.7551/MITPRESS/11555.001.0001.

[15] D. Yankelovich, *Corporate priorities: A continuing study of the new demands on business*. Houston, TX: Gulf Publishing Company, 1972.

[16] F. Doshi-Velez and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning," Feb. 2017, Accessed: Apr. 17, 2025. [Online]. Available: https://arxiv.org/abs/1702.08608v2

[17] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, "Machine bias," *ProPublica*, 2016, Available: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[18] T. Bolukbasi, K.-W. Chang, J. Y. Zou, V. Saligrama, and A. T. Kalai, "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings," in *Advances in neural information processing systems 29 (NeurIPS 2016)*, pp. 4349–4357, 2016.

[19] R. Richardson, J. Schultz, and K. Crawford, "Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice," *New York University Law Review*, vol. 94, pp. 192–233, 2019,

[20] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel, "Fairness through awareness," in *Proceedings of the 3rd innovations in theoretical computer science conference (ITCS)*, 2012, pp. 214–226. doi: 10.1145/2090236.2090255.

[21] R. Binns, "Fairness in machine learning: Lessons from political philosophy," in *Proceedings of the 1st conference on fairness, accountability and transparency*, S. A. Friedler and C. Wilson, Eds., in Proceedings of machine learning research, vol. 81. New York, NY, USA: PMLR, 2018, pp. 149–159.

[22] I. Rahwan, M. Cebrian, N. Obradovich, *et al.*, "Machine behaviour," *Nature*, vol. 568, pp. 477–486, 2019, doi: 10.1038/s41586-019-1138-y.

[23] F. Pasquale, *The black box society: The secret algorithms that control money and information*. Cambridge, MA: Harvard University Press, 2015.

[24] M. Brundage *et al.*, "Toward trustworthy AI development: Mechanisms for supporting verifiable claims," *arXiv preprint arXiv:2004.07213*, 2020, Available: https://arxiv.org/abs/2004.07213

[25] C. Cadwalladr, "The great hack," *The Guardian*, 2019, Available: https://www.theguardian.com/uk-news/2019/jul/20/the-great-hack-cambridge-analytica-scandal-facebook-netflix

[26] M. Whittaker *et al.*, "Disability, bias, and AI," AI Now Institute, 2019. Available: https://ainowinstitute.org/publication/disabilitybiasai-2019

[27] S. Zuboff, *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: PublicAffairs, 2019.

[28] N. Couldry and U. A. Mejias, *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford, CA: Stanford University Press, 2019.

[29] H. Nissenbaum, "Contextual integrity up and down the data food chain," *Theoretical Inquiries in Law*, vol. 20, no. 1, pp. 221–256, 2019, doi: 10.1515/til-2019-0008.

[30] A. Pentland, A. Lipton, and T. Hardjono, *Building the new economy: Data as capital*. Cambridge, MA: MIT Press, 2021.

[31] N. M. Richards and W. Hartzog, "A duty of loyalty for privacy law," *Washington University Law Review*, vol. 98, pp. 559–611, 2021

[32] S. Delacroix and N. D. Lawrence, "Bottom-up data trusts: Disturbing the 'one size fits all' approach to data governance," *International Data Privacy Law*, vol. 9, no. 4, pp. 236–252, 2019, doi: 10.1093/idpl/ipz014.

[33] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Computing Surveys*, vol. 54, no. 6, pp. 1–35, 2021, doi: 10.1145/3457607.

[34] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science*, vol. 366, no. 6464, pp. 447–453, 2019, doi: 10.1126/science.aax2342.

[35] H. Suresh and J. V. Guttag, "A framework for understanding sources of harm throughout the machine learning life cycle," *arXiv preprint arXiv:1901.10002*, 2019, Available: https://arxiv.org/abs/1901.10002

[36] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and machine learning: Limitations and opportunities*. MIT Press, 2023.

[37] D. MacKenzie, *Trading at the speed of light: How ultrafast algorithms are transforming financial markets*. Princeton, NJ: Princeton University Press, 2023. Available: https://press.princeton.edu/books/hardcover/9780691211381/trading-at-the-speed-of-light

[38] S. Nyholm, "Attributing agency to automated systems: Reflections on human–robot collaborations and responsibility-loci," *Science and Engineering Ethics*, vol. 24, no. 4, pp. 1201–1219, 2018, doi: 10.1007/s11948-017-9943-x.

[39] J. J. Bryson, "The past decade and future of AI's impact on society," in *Towards a new enlightenment? A transcendent decade*, Turner, 2019, pp. 150–167. Available: https://static1.squarespace.com/static/5e13e4b93175437bccfc4545/t/67057b00d95cad198c5575da/1728412417139/the-past-decade-and-future-of-ai-impact-on-society-bbva.pdf

[40] N. Heess *et al.*, "Emergence of locomotion behaviours in rich environments," *arXiv preprint arXiv:1707.02286*, 2017, Available: https://arxiv.org/abs/1707.02286

[41] N. Bostrom, *Superintelligence: Paths, dangers, strategies*. Oxford, UK: Oxford University Press, 2014. Available: https://global.oup.com/academic/product/superintelligence-9780199678112

[42] D. Hadfield-Menell, A. Dragan, P. Abbeel, and S. Russell, "The off-switch game," in *Proceedings of the 26th international joint conference on artificial intelligence (IJCAI)*, 2017, pp. 2206–2212. doi: 10.24963/ijcai.2017/32.

[43] A. Matthias, "The responsibility gap: Ascribing responsibility for the actions of learning automata," *Ethical Theory and Moral Practice*, vol. 6, no. 3, pp. 215–233, 2004, doi: 10.1007/s10676-004-3422-1.

[44] S. Gless, T. Silverman, and T. Weigend, "If robots cause harm, who is to blame? Self-driving cars and criminal liability," *New Criminal Law Review*, vol. 19, no. 3, pp. 412–436, 2016, doi: 10.1525/nclr.2016.19.3.412.

[45] E. Yudkowsky, "Artificial intelligence as a positive and negative factor in global risk," in *Global catastrophic risks*, N. Bostrom and M. M. Ćirković, Eds., Oxford University Press, 2008, pp. 308–345.

[46] S. D. Baum, "On the promotion of safe and socially beneficial artificial intelligence," *AI & Society*, vol. 32, pp. 543–551, 2017, doi: 10.1007/s00146-016-0677-0.

[47] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané, "Concrete problems in AI safety," *arXiv preprint*, 2016, Available: https://arxiv.org/abs/1606.06565

[48] Y. Benkler, R. Faris, and H. Roberts, *Network propaganda: Manipulation, disinformation, and radicalization in american politics*. New York, NY: Oxford University Press, 2018.

[49] M. Tegmark, *Life 3.0: Being human in the age of artificial intelligence*. New York: Knopf, 2017.

[50] S. Russell, *Human compatible: Artificial intelligence and the problem of control*. Viking, 2019.

[51] *FAccT '21: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. New York, NY, USA: Association for Computing Machinery, 2021.

[52] D. Acemoglu and P. Restrepo, "Robots and jobs: Evidence from US labor markets," *Journal of Political Economy*, vol. 128, no. 6, pp. 2188–2244, 2020, doi: https://doi.org/10.1086/705716.

[53] J. E. Bessen, "AI and Jobs: The Role of Demand," National Bureau of Economic Research, Working Paper 24235, 2018. doi: 10.3386/w24235.

[54] C. B. Frey and M. A. Osborne, "The future of employment: How susceptible are jobs to computerisation?" Oxford Martin School, 2013

[55] D. Allen *et al.*, "A roadmap for governing AI: Technology governance and power sharing liberalism." Harvard Kennedy School Ash Center for Democratic Governance; Innovation, 2024.

[56] C. Clifton, N. Lambert, N. Carlini, D. Song, and S. Russell, "Is decentralized AI safer?" *arXiv preprint arXiv:2211.05828*, 2022, Available: https://arxiv.org/abs/2211.05828

[57] S. Hubbard, "Cooperative paradigms for artificial intelligence." Harvard Kennedy School Ash Center for Democratic Governance; Innovation, 2024.

[58] K. Crawford, *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. New Haven, CT: Yale University Press, 2021.

[59] J.-P. Deranty and T. Corbin, “Artificial intelligence and work: A critical review of recent research from the social sciences,” *AI & Society*, vol. 38, no. 1, pp. 255–269, 2023, doi: 10.1007/s00146-022-01496-x.