Political Science Critique

Journal Homepage: https://so01.tci-thaijo.org/index.php/PSC

Volume 12 Issue 24

Artificial Intelligence Beyond Automation: Discourses of Risk, Opportunity, and Control*

Shamir Jakaria [a]

Article Info

Article history:
Received October 1, 2025.
Revised October 31, 2025.
Accepted December 12, 2025.

Keywords:
Artificial Intelligence; Automation; Governance

* This research article is part of Artificial Intelligence Beyond Automation: Discourses of Risk, Opportunity, and Control

Abstract

Artificial Intelligence (AI) has evolved from an automation tool into a transformative societal force, creating a polarized landscape of optimism and concern regarding its governance and impact. This study seeks to bring clarity by mapping the dominant narratives that shape our collective understanding of AI. Guided by technological securitization theory, this research employs a qualitative, critical discourse analysis of 44 influential academic, policy, and media documents published between 2019 and 2025. The analysis reveals three distinct yet interconnected discursive patterns: AI as an Opportunity for human development, economic growth, and institutional efficiency; AI as a Risk that amplifies systemic vulnerabilities like algorithmic bias, labor displacement, and disinformation; and AI as an object of Control that necessitates governance through ethical standards, regulations, and accountability mechanisms. The findings demonstrate that these competing narratives are not isolated but interact to shape political and regulatory priorities. The study concludes that the future of AI will be determined less by its technical capabilities and more by the political choices informed by these discourses. It consequently proposes a multi-level set of policy recommendations spanning national regulations, international cooperation, corporate accountability, and public foresight to proactively steer AI innovation towards upholding human dignity, democratic values, and global security.

[a] Lecturer, College of Politics & Governance, Suan Sunandha Rajabhat University, 10300 Thailand. Email: shamir.ja@ssru.ac.th

1. Introduction

In just a few years, Artificial Intelligence (AI) has shed its skin as a specialized tool to become a societal force, quietly weaving itself into the fabric of our institutions, economies, and daily lives. Its influence now extends far beyond simple automation, prompting a crucial question:
how do we make sense of a technology that promises to revolutionize healthcare and education, yet simultaneously threatens to amplify bias and disrupt the very foundations of the labor market? This duality lies at the heart of contemporary debates. On one hand, a wave of optimism frames AI as an unparalleled opportunity for human development and economic growth (Lengfelder & Alkire, 2025; Babina et al., 2024). On the other, a chorus of caution warns of its potential to deepen social inequalities and create novel dangers, from autonomous weapons to pervasive disinformation (Ukanwa, 2024; Frank et al., 2025).

Navigating this polarized landscape is a central challenge of our time. International bodies like the OECD (2023) and UNESCO (2021) call for "trustworthy AI," but public understanding often swings between hope and fear, complicating the path to sensible regulation (Rehak, 2025). As a researcher observing these debates, it becomes clear that the future of AI will be shaped not just by its technical capabilities, but by the stories we tell about it. Is it a tool for progress, an existential risk, or something that must be rigorously controlled?

This paper seeks to bring clarity to this conversation by mapping the dominant narratives that define our collective understanding of AI. Guided by technological securitization theory (Buzan et al., 1998), which helps explain how issues get framed as security threats, we analyze how AI is constructed in three distinct ways: as an opportunity, a risk, and an object of control. By tracing these threads through academic, policy, and media discourses, we move beyond a simple listing of pros and cons. While many studies have explored the ethical, economical, and political dimensions of AI, few have systematically examined how competing narratives interact to shape governance priorities. This gap limits our understanding of how disclosure itself influences regulatory and institutional outcomes.

Accordingly, this study seeks to fill this gap by addressing the following research questions:

1. How is artificial intelligence discursively constructed as an opportunity, risk, and an object of control in academic, policy and media texts?

2. In what ways do these discourses interact to shape contemporary approaches to AI governance?

3. How does the application of technological securitization theory help explain the political and regulatory implications of these narratives?

The paper is structured to first outline its research objectives and methodology, then present its findings according to the three discursive patterns, and finally discuss their interplay and propose concrete policy implications for a balanced and human-centric AI governance approach.

2. Research Objectives

This study is guided by three core research objectives, designed to systematically investigate the discursive construction of artificial intelligence and its profound implications for society and governance:

1. Examine how artificial intelligence is discursively framed as opportunity, risk, and control in academic, policy, and media narratives.
2. Apply technological securitization theory to analyze the societal and governance implications of these discourses.
3. Recommend strategies and policy measures that balance innovation with accountability to ensure AI supports human rights, democracy, and security.

3. Research Frameworks and Research Methodology

3.1 Research Frameworks and Conceptual Framework

Artificial Intelligence (AI) has progressed from its origins as a technical tool to a transformative force shaping healthcare, education, finance, and governance, where it influences institutional efficiency and social behavior (Whittlestone et al., 2019). Optimistic perspectives emphasize AI's potential to accelerate innovation, improve service delivery, and expand human capabilities, particularly in advancing diagnostics, decision-making, and economic growth (Akter et al., 2021). Yet, critical studies reveal systemic risks such as algorithmic bias, labor displacement, surveillance, and disinformation, which threaten fairness, equity, and democratic accountability (Azgin & Kiralp, 2024; Ferrara, 2024; Franklin et al., 2024). These competing perspectives frame AI as both an opportunity and a vulnerability, shifting the debate from technical efficiency to ethical, political, and security concerns, and calling for governance structures that balance innovation with accountability (Osasona et al., 2024).

Guided by technological securitization theory (Buzan et al., 1998), this study conceptualizes AI across three interrelated dimensions. First, AI as Opportunity positions technology as a driver of innovation, efficiency, and human development. Second, AI as Risk highlights vulnerabilities such as bias, disinformation, and labor disruptions. Third, AI as Control situates governance mechanisms including laws, ethical standards, and oversight frameworks as mediating forces to reconcile innovation with accountability. These strands are mutually reinforcing: opportunities drive adoption, risks amplify regulation demands, and control emerges as the balancing force aligning technology with societal values. Thus, AI is not merely a neutral tool but a contested domain shaped by narratives, institutional design, and governance choices.

3.2 Research Methodology

This study adopts a qualitative, documentary-based design, employing critical discourse analysis in 44 key documents published between 2019 and 2025. These texts were specifically chosen for their influence and relevance, and they include:

1. Academic works on AI ethics, governance, and security;
2. Policy documents from organizations such as UNESCO, and the OECD;
3. Media and cultural narratives shaping public perception.

Purposive sampling criteria included influence (citations, relevance to AI debates), credibility (peer-reviewed or institutional reports), and representation (balanced across

optimistic, critical, and regulatory perspectives). Data were coded according to three dimensions: opportunity, risk, and control, following technological securitization theory. Coding was cross-checked to ensure consistency. The analytical framework below (Table 1) summarizes these patterns, their core narratives, and the key themes that emerged. This framework served as the guide for organizing and presenting the findings in the Results section. This multi-source, multi-level approach strengthens transparency and methodological rigor.

Table 1. Analytical Framework for Discursive Patterns in AI Narratives

Discursive Pattern	Core Narrative	Key Themes
AI as Opportunity	AI is a transformative force for positive progress, enhancing human capabilities, economic growth, and institutional efficiency.	Human Development and Capabilities Economic Growth and Innovation Healthcare and Pandemic Preparedness Environmental and Urban Governance Institutional and Public-Sector Efficiency
AI as Risk	AI generates systemic vulnerabilities that threaten social equity, economic stability, democratic integrity, and global security.	Algorithmic Bias and Inequality Labor Market Disruption Disinformation and Epistemic Instability Surveillance and Erosion of Democracy Weaponization and Autonomous Systems Media-Amplified Dystopian Narratives
AI as Control	AI must be actively steered through governance mechanisms, ethical standards, and regulations to ensure responsible and accountable development and use.	Accountability and Responsible AI Frameworks Risk-Based Regulation and Legal Safeguards (e.g., EU AI Act) International Coordination and Soft Law Organizational and Corporate Governance Multi-level and Multi-stakeholder Oversight

4. Results

Three major discursive patterns emerged from the analysis:

4.1. AI as Opportunity

Artificial intelligence (AI) is frequently framed as a transformative opportunity that extends far beyond automation, offering meaningful pathways for human development, economic expansion, and institutional efficiency. From a human development perspective, AI is increasingly understood as a tool to expand individual capabilities and improve well-being. Lengfelder et al. (2025) argue that AI, if directed by human development principles, can support people's ability to live fulfilling lives by enhancing access to education, healthcare, and decision-making resources. This perspective aligns with (Benvenuti et al., 2023) who emphasize that AI applications in behavioral and educational contexts foster learning and personal growth, suggesting that technology can become a catalyst for strengthening human capabilities when designed responsibly. In this sense, AI's promise lies not only in technical efficiency but also in its capacity to contribute to broader social development goals.

Beyond the human development lens, AI is also widely associated with its capacity to drive economic growth through innovation and productivity. Studies demonstrate that

firms adopting AI achieve higher levels of product innovation and experience accelerated growth trajectories compared to their counterparts (Babina et al., 2024). At the macro level, empirical evidence reveals that AI innovation positively correlates with economic growth, particularly when economies actively integrate AI into productive sectors (Gonzales, 2023). Broader analyses by Qin et al. (2023) confirm that AI influences economic development through multiple channels, including efficiency gains, improved capital allocation, and the creation of new markets. Filippucci et al. (2024) further highlight that AI adoption holds the potential to increase productivity and raise living standards, although they caution that distributional issues and policy frameworks will mediate these outcomes. Collectively, this body of work underscores AI's role as a driver of economic transformation, with implications for long-term competitiveness and inclusive growth.

In healthcare and pandemic preparedness, AI applications demonstrate critical societal value. Sharmin et al. (2025) report that AI-based systems can accurately predict outbreaks ($R^2 = .92$), improve resource allocation efficiency, and streamline vaccine delivery. McKee et al. (2024) highlight AI's utility in early outbreak detection, logistics optimization, surveillance, and policymaking support. Their insights are echoed by Alwakeel (2025) whose SmartHealth-Track framework integrates IoT, wearable data, pharmaceutical analytics, and AI modeling to enhance real-time disease monitoring and outbreak response.

Environmental and urban governance contexts further illustrate AI's capacity for monitoring and resilience. Sun et al. (2024) show how AI-driven smart city infrastructure, including intelligent transport, energy management, and environmental sensors, boosts operational efficiency and life quality. Additionally, AI-enabled environmental surveillance using satellite imagery and acoustic data, such as tracking plastic pollution, exemplifies how AI supports sustainable stewardship (Singh et al., 2025).

Institutional and governance contexts also illustrate AI's opportunities for improving organizational performance and public-sector service delivery. Research shows that governments are increasingly deploying AI to enhance administrative efficiency, streamline decision-making, and deliver citizen-centered services (Selten & Klievink, 2024). This adoption has begun to reshape managerial roles, as AI systems allow public managers to move beyond routine tasks and focus on strategic responsibilities (Giest & Klievink, 2022). Similarly, Mišić et al. (2025) propose that when governed effectively, AI can strengthen institutional values such as transparency, accountability, and fairness while simultaneously improving service delivery. These opportunities highlight AI's dual role as both a technical and institutional innovation capable of reshaping governance practices.

4.2. AI as Risk

While artificial intelligence (AI) has been heralded as a driver of progress, a growing body of scholarship frames it as a source of systemic risk, generating new vulnerabilities across social, economic, political, and security domains. One of the most widely discussed concerns relates to algorithmic bias, where AI systems reproduce or even amplify existing inequities embedded in data and design. Ukanwa (2024) emphasizes that algorithmic bias stems from underlying social and structural factors, warning that without strong theoretical foundations and cross-disciplinary solutions, such systems may reinforce existing inequalities. Similarly, Ebrahimi et al. (2024) point out that algorithmic opacity in organizational settings creates governance challenges that make bias harder to detect and correct. In high-stakes contexts such as healthcare, biased algorithms can lead to

discriminatory clinical outcomes, making bias recognition and mitigation essential (Hasanzadeh et al., 2025). Mackin et al. (2025) extend this concern to public sector safety nets, demonstrating how algorithmic decision-making can systematically disadvantage marginalized groups. These studies collectively illustrate how algorithmic systems, rather than offering neutrality, often encode and amplify systemic vulnerabilities.

Another major risk narrative concerns the labor market disruptions and unemployment risks associated with AI adoption. Frank et al. (2025) provide empirical evidence that occupations with higher exposure to AI technologies exhibit increased unemployment risks, underscoring how automation can create heterogeneous labor market vulnerabilities. At the household level, perceptions of AI-driven unemployment risks influence financial decisions, such as investment strategies, which suggests that anxieties about job loss extend beyond employment itself into broader economic behavior (Zhang, 2025). These findings reinforce longstanding concerns that AI may displace human labor at a scale that threatens socioeconomic stability, particularly in industries reliant on routine or codifiable tasks.

Concerns over disinformation, deepfakes, and epistemic instability further illustrate the risks of AI. Deepfake technologies are increasingly sophisticated, enabling the creation of hyper-realistic fabricated images, audio, and video content that can undermine trust in media and political institutions. Gambín et al. (2024) show how deepfakes can be weaponized for identity theft, scams, and political manipulation, while Kharvi (2024) argue that widespread exposure to AI-generated disinformation threatens societal trust and exacerbates polarization. Research consistently shows that human capacity to detect deepfakes without technological assistance is limited, making populations vulnerable to manipulation (Diel et al., 2024). Moreover, generative language models such as GPT-4 introduce new risks of persuasive disinformation, as Salvi et al. (2025) demonstrate in experiments showing AI-assisted conversational persuasion can alter opinions with high efficacy. These risks underscore the fragility of epistemic systems in the face of powerful AI content-generation tools.

AI-driven surveillance and erosion of democratic accountability present another dimension of risk. Curran (2023) frames these threats through the lens of surveillance capitalism, showing how data extraction and algorithmic profiling generate systemic risks for privacy and democratic governance. Black (2023) builds on this point by stressing that computational propaganda and surveillance technologies erode democratic equality, centralizing power and diminishing public accountability. The increasing use of AI in state and corporate surveillance infrastructures suggests that risks extend beyond privacy to the structural conditions of democracy itself.

Perhaps the most acute concerns emerge in the context of weaponization and autonomous systems. The potential deployment of Lethal Autonomous Weapons Systems (LAWS) raises profound ethical, legal, and security dilemmas. Guo (2025) argues that delegating lethal decision-making to machines fundamentally undermines moral accountability and increases risks of harm to civilians. Kohn et al. (2024) emphasize the difficulty of embedding reliable ethical decision-making in LAWS, pointing out the tension between operational autonomy and international humanitarian law. McFarland and Assaad (2023) show how “in-situ learning” by autonomous weapons complicates traditional legal reviews, raising questions about post-deployment predictability and accountability. Christie et al. (2023) highlights the growing international debate around governance

principles such as "meaningful human control," reflecting widespread recognition of the dangers posed by weaponized AI.

Media portrayals and public narratives often amplify dystopian scenarios, reflecting societal anxieties about losing control. Rehak (2025) argues that post-ChatGPT discourse swings between narratives of technological utopia and existential doom, shaping policy appetite and public imagination. Similarly, Zai et al. (2025) demonstrate that media framing frequently emphasizes risks such as job displacement, bias, and existential threats, thereby influencing public trust in AI. Williams (2025) notes that legacy U.S. media outlets in 2024 disproportionately centered on negative AI narratives, reinforcing fear-based discourses that contribute to public skepticism. These studies highlight how cultural and media narratives both reflect and magnify societal anxieties, shaping not only public perception but also regulatory responses to AI.

4.3. AI as Control

Beyond opportunity and risk, a growing scholarly stream casts artificial intelligence as a governance problem, a field that must be actively steered through norms, laws, standards, and organizational practice. Reviews and frameworks in the governance literature argue that "responsible AI" depends on structures such as oversight bodies, processes such as auditing and impact assessments, and relationships such as stakeholder participation that make systems answerable to the public interest (Papagiannidis et al., 2025; Batool et al., 2025). At the core of this discourse is accountability, which AI & Society scholars unpack as a multifaceted sociotechnical practice requiring answerability, the recognition of authority to question AI decisions, and real limits on power rather than a vague ideal (Novelli et al., 2023). Complementary work highlights how transparency and explainability support accountability yet face practical and legal frictions across sectors, motivating regulatory tools like documentation duties, model reporting, and rights to explanation (Cheong, 2024). Together, this research positions AI as an object of control that must be governed across multiple levels, including team, organizational, industry, national, and international, rather than left to technical optimization alone (Batool et al., 2025).

Legal-policy scholarship advances this "AI as control" framing by examining risk-based regulation and fundamental-rights safeguards. Analyses of the European Union Artificial Intelligence Act (EU AI Act) and allied approaches argue that categorizing use cases by risk, paired with duties for data governance, human oversight, and post-market monitoring, seeks to reconcile innovation with the prevention of harms to rights and safety (Kusche, 2024; Neuwirth, 2023). Work in humanities and social-science venues further stresses the need for international coordination, noting that cross-border deployment and supply chains outpace national law, and calling for emerging principles of international AI law and enforcement capacity Zaidan & Ibrahim (2024). In parallel, journals document how soft-law regimes such as the OECD AI Principles shape national policy and corporate practice by embedding values of human rights, transparency, and accountability into guidance and implementation reviews (OECD, 2023). Collectively, these articles treat regulation not as a brake but as a steering mechanism to align AI with security and human rights commitments.

A complementary stream examines organizational and corporate governance for AI. Business, information-systems, and public-administration journals propose models for

board-level oversight, risk committees, AI ethics policies, and audit trails, arguing that governance capacity inside firms and agencies is essential to make external regulations effective (Zhao, 2024; Camilleri, 2023). Public-sector studies show how AI can enhance financial accountability and participatory oversight when embedded in transparent processes, yet they also warn that without controls, automation can obscure responsibility (Shaban & Omoush, 2025). These works converge on the idea that control is distributed: regulators set guardrails, organizations operationalize them via lifecycle practices such as design controls, validation, monitoring, and incident reporting, and independent scrutiny by researchers, auditors, and civil society provides checks and balances.

Across these strands, the discursive positioning of AI as a governance issue reflects a normative project to keep innovation compatible with democratic values and human dignity. Governance scholarship argues that ethics statements are insufficient unless coupled to enforceable accountability and institutional transparency. It also stresses that rights-compatible AI requires ex ante design obligations and ex post remedies, and that international and multi-stakeholder coordination is necessary for problems that transcend jurisdictions (Cheong, 2024; Kusche, 2024; Papagiannidis, 2025). In short, "AI as control" reframes the debate from whether to innovate to how to govern innovation so that security and human rights remain non-negotiable.

5. Discussions

The analysis of academic, policy, and media narratives reveals three distinct yet interconnected discursive patterns for understanding artificial intelligence: as an opportunity, a risk, and an object of control. This section expands on their theoretical relationships and policy implications before proceeding to the conclusion.

The findings of this study underscore that discourses on artificial intelligence are neither singular nor static but emerge from overlapping narratives that position AI as opportunity, risk, and control. The opportunity discourse highlights AI's transformative potential for human development, economic growth, and institutional efficiency, resonating with historical accounts of technological optimism during earlier industrial and digital revolutions. What distinguishes AI, however, is the way it is constructed as a general-purpose technology that penetrates nearly all domains of life—from education and healthcare to urban governance and security (Lengfelder et al., 2025; Filippucci et al., 2024). This framing aligns with global development agendas that portray AI as a catalyst for inclusive growth, echoing the broader promise of technology as a vehicle for societal progress. Yet this promise is not self-executing, for the literature consistently demonstrates that benefits materialize only within enabling conditions of governance, equity, and responsible adoption (Mišić et al., 2025).

Alongside narratives of opportunity, the discourse of risk presents a sobering counterpoint. The persistence of algorithmic bias, disinformation, unemployment anxieties, and weaponization reveals that AI is not merely a neutral technical system but an amplifier of existing vulnerabilities. Echoing Buzan et al.'s (1998) securitization framework, these risks are framed as existential threats that justify extraordinary responses, such as regulatory restrictions, public moratoria, or international negotiations over lethal autonomous weapons (Guo, 2025; Kohn et al., 2024). The securitization lens clarifies how these risks transform political discourse: once AI is labeled as a threat to social stability, it becomes subject to exceptional governance responses, reshaping institutional authority and legitimizing stricter oversight.

The control discourse offers a middle ground by conceptualizing AI as a governance object that must be actively steered. Here, the literature advances a normative project of aligning AI with democratic accountability, human rights, and international law (Novelli et al., 2023; Kusche, 2024; Papagiannidis et al., 2025). Unlike the polarized framing of opportunity versus risk, the control discourse emphasizes institutional design and regulatory experimentation as mechanisms for reconciling innovation with responsibility. This approach mirrors broader technological governance frameworks that stress the importance of oversight bodies, risk-based regulation, and stakeholder participation. It situates AI within a multi-level governance architecture where states, firms, and civil society actors co-produce accountability mechanisms, echoing debates in international relations about the governance of transboundary issues. Importantly, this discourse challenges the assumption that market logics alone can resolve the dilemmas posed by AI, instead positing that collective action and institutional foresight are indispensable. This reflects an emerging consensus in governance studies that complex technologies require multi-level coordination, with nation-states, firms, and civil society sharing responsibility for ethical oversight.

From a theoretical standpoint, these findings illustrate how technological securitization theory operates beyond traditional security contexts. AI is securitized not only in military terms but also in economic, ethical, and epistemic domains. By framing AI as both a source of risk and an object of control, policymakers create a justification for new regimes of surveillance, accountability, and legal constraint. This reaffirms that discourse is not a neutral medium of description but a mechanism of power that shapes what actions become politically possible.

6. Conclusion and Policy Implications

This study concludes that artificial intelligence is framed simultaneously as opportunity, risk, and control, with each discourse shaping policy, governance, and public perception in distinct yet interconnected ways. The opportunity narrative emphasizes AI's capacity to enhance human development, economic growth, and institutional efficiency, while the risk discourse highlights systemic vulnerabilities ranging from algorithmic bias and unemployment to disinformation and weaponization. The control discourse advances the view that AI must be actively governed through accountability mechanisms, regulatory frameworks, and international cooperation to reconcile innovation with democratic values and human rights. The integration of technological securitization theory has provided a deeper explanation of how these discourses operate as political tools that both reflect and construct the governance of emerging technologies.

In theoretical terms, the study demonstrates that AI governance debates mirror broader struggles over authority, legitimacy, and public trust in the digital age. In policy terms, it suggests that sustainable AI governance requires balancing innovation with institutional transparency, social justice, and collective foresight.

The following recommendations are structured across four levels of governance to provide a coherent roadmap for policymakers.

6.1. National-Level Policies: Building Foundational Trust and Capacity

To harness the Opportunity of AI while mitigating its Risks, national governments must act as foundational architects of trust and equity.

6.1.1. Mandate Algorithmic Auditing and Impact Assessments

Legislators should enact laws that require mandatory, transparent, and recurring algorithmic bias audits for high-stakes AI systems. For example, an AI system used for pre-trial risk assessment could be required to undergo a quarterly audit by an accredited third party, with results on key fairness metrics (e.g., Disparate Impact Ratio) submitted to a regulator. This directly addresses the discourse of AI as Risk by operationalizing the discovery of bias, as highlighted by Ukanwa (2024) and Mackin et al. (2025).

6.1.2. Launch National AI Literacy and Reskilling Initiatives

To counter labor market disruptions (AI as Risk) and ensure the equitable distribution of AI's benefits (AI as Opportunity), governments must fund massive digital and AI literacy programs. This includes integrating AI literacy into public education and creating robust reskilling subsidies for workers in AI-vulnerable occupations, a concern empirically validated by Frank et al. (2025).

6.1.3. Establish AI Safety Institutes and Sandboxes

Inspired by models for product safety, dedicated government bodies should be tasked with evaluating frontier AI models for emergent risks. Regulatory "sandboxes" can provide a controlled environment for businesses to test innovative AI applications under temporary regulatory relief, balancing the Control and Opportunity discourses. The UK's AI Safety Institute serves as a model, focusing on evaluating frontier models for risks like autonomy and deception. A complementary sandbox could be established for autonomous vehicle software. A company could test its AI in a designated urban zone under a temporary exemption from certain liability laws, provided it shares all safety and incident data in real-time with the regulator. This creates a controlled learning environment that balances the Control and Opportunity discourses.

6.2. International Governance: Managing Transboundary Risks

AI's risks, such as creating autonomous weapons or spreading disinformation, do not stop at national borders. Therefore, the "AI as Control" perspective argues that international cooperation is essential.

6.2.1. Develop a Global Framework for Lethal Autonomous Weapons (LAWS)

States must urgently converge on a legally binding international treaty that establishes "meaningful human control" as a fundamental principle for the use of force, directly responding to the ethical crises outlined by Guo (2025) and Kohn et al. (2024).

6.2.2. Promote Cross-Border Data and Model Governance Standards

Building on the EU AI Act's risk-based approach, international coalitions should harmonize standards for data provenance, model transparency, and liability. An international coalition, perhaps through the OECD or ISO, could develop a mandatory "AI Passport" for certain high-risk systems. This digital record would travel with the model, containing information on its training data provenance, performance across different demographics, known limitations, and results from standardized safety tests. A medical diagnostic AI developed in the U.S. and deployed in EU hospitals would require this passport, demonstrating compliance with shared transparency standards and addressing the jurisdictional challenges noted by Zaidan & Ibrahim (2024).

6.2.3. Create an AI Incident Monitoring and Sharing Network

Modeled on aviation safety networks, an international body should be established to confidentially share data on AI failures, near-misses, and malicious uses. This would

accelerate global learning and improve the collective ability to anticipate and prevent systemic Risks.

6.3. Organizational Accountability: Operationalizing Ethical AI

The Control discourse emphasizes that corporate responsibility is not optional. Policy must ensure that ethical principles are translated into organizational practice.

6.3.1. Enforce Board-Level AI Accountability

Regulators should mandate that corporate boards include members with AI governance expertise and formally oversee AI ethics and risk, as proposed by Zhao (2024). This makes accountability a top-level strategic issue, not just a technical compliance function.

6.3.2. Incentivize "Ethical by Design" Development

Public procurement is a powerful but underutilized lever. Governments can create preferential bidding criteria for companies that can demonstrably prove their adherence to ethical AI development lifecycle standards. For instance, a city government issuing a tender for a public-facing chatbot service could award additional points in the evaluation to companies that are certified under a recognized standard (e.g., based on the IEEE Ethically Aligned Design, which promotes human rights, accountability, and transparency in AI systems, or ISO 42001, which establishes requirements for managing AI systems responsibly and ensuring their ethical, safe, and reliable use). Furthermore, the government could offer a tax credit of 5% of R&D expenditure for companies that undergo independent, third-party audits of their AI development processes, focusing on fairness, transparency, and accountability. This directly ties corporate profit to public good, encouraging the responsible innovation highlighted in the Opportunity discourse.

6.3.3. Require Public AI Use Policies

All organizations deploying significant AI systems, especially in the public sector, should be required to publish clear policies on their use, including the system's purpose, data sources, known limitations, and human oversight procedures. This fosters the institutional transparency that Mišić et al. (2025) identify as crucial for good governance.

6.4. Public Investment and Foresight: Shaping a Desirable Future

Policy must be proactive, not merely reactive. It should actively shape the AI ecosystem towards positive goals.

6.4.1. Direct Public R&D towards Human-Centric AI

Governments should significantly increase funding for AI research focused on "co-intelligence"—tools that augment human capabilities—and applications in public health, climate science, and education. This steers the Opportunity discourse towards concrete, socially beneficial outcomes, as envisioned by Lengfelder et al. (2025).

6.4.2. Fund Independent Sociotechnical Research

A portion of national AI R&D budgets should be allocated to social scientists, legal scholars, and ethicists to study the long-term impacts of AI on society, democracy, and the economy. This builds the foundational knowledge needed for evidence-based policy.

6.4.3. Launch Public Awareness Campaigns on Digital Epistemology

To counter the Risk of AI-driven disinformation, public institutions should launch campaigns to educate citizens on how to critically evaluate digital information, the capabilities and limitations of generative AI, and the importance of verified sources. These campaigns must be proactive and skill-based, moving beyond warnings to active inoculation by teaching citizens not only to be skeptical but also to recognize the hallmarks

of AI-generated content and manipulation tactics. A national campaign, "Verify Before You Venerate," could be implemented in partnership with public libraries and schools to build this capacity. It would include interactive workshops that teach people to identify subtle deepfake artifacts such as unnatural blinking or mismatched audio using tools like the Witness' "Deepfake Lab," curriculum modules for schools that emphasize lateral reading by training students to open new browser tabs to verify the credibility of sources, and public service announcements that demonstrate how to use reverse image search and fact-checking websites to confirm the authenticity of viral AI-generated images or claims. This approach directly addresses the epistemic risks identified by Gambín et al. (2024) and Salvi et al. (2025).

In conclusion, navigating the competing discourses of AI requires policies that are as dynamic and multifaceted as the technology itself. The recommended actions, which span national capacity building, international cooperation, corporate accountability, and public foresight, provide an integrated strategy. The ultimate goal is to evolve AI governance from a reactive stance focused on Control to a proactive form of stewardship that confidently cultivates Opportunity, diligently mitigates Risk, and ensures that technological progress remains firmly anchored in the service of human dignity, democratic values, and global security.

Reference

Akter, S., Dwivedi, Y. K., Biswas, K., Michael, K., Bandara, R. J., & Sajib, S. (2021). Addressing Algorithmic Bias in AI-Driven Customer Management. *Journal of Global Information Management (JGIM)*, *29*(6), 1-27. https://doi.org/10.4018/JGIM.20211101.oa3

Alwakeel, M. M. (2025). AI-Assisted Real-Time Monitoring of Infectious Diseases in Urban Areas. *Mathematics*, *13*(12), 1911. https://doi.org/10.3390/math13121911

Azgin, B., & Kiralp, S. (2024). Surveillance, Disinformation, and Legislative Measures in the 21st Century: AI, Social Media, and the Future of Democracies. *Social Sciences*, *13*(10), 510. https://doi.org/10.3390/socsci13100510

Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, *151*. https://doi.org/10.1016/j.jfineco.2023.103745

Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: a systematic literature review. *AI And Ethics*. https://doi.org/10.1007/s43681-024-00653-w

Benvenuti, M., Cangelosi, A., Weinberger, A., Mazzoni, E., Benassi, M., Barbaresi, M., & Orsoni, M. (2023). Artificial intelligence and human behavioral development: A perspective on new skills and competences acquisition for the educational context. *Computers in Human Behavior*, *148*. https://doi.org/10.1016/j.chb.2023.107903

Black, A. (2023). AI and Democratic Equality: How Surveillance Capitalism and Computational Propaganda Threaten Democracy. In: B. Steffen (Ed.) *Bridging the Gap Between AI and Reality. AISoLA 2023. Lecture Notes in Computer Science, vol 14129* (pp. 333-347). Springer, Cham. https://doi.org/10.1007/978-3-031-73741-1_21

Buzan, B. G., Wæver, O., & de Wilde, J. H. (1998). *Security: A New Framework for Analysis*. Lynne Rienner.

Camilleri, M. A. (2023). Artificial intelligence governance: Ethical considerations and implications for social responsibility. *Expert Systems*, *41*(7). https://doi.org/10.1111/exsy.13406

Cheong, B. C. (2024). Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, *6*. https://doi.org/10.3389/fhumd.2024.1421273

Christie, E. H., Ertan, A., Adomaitis, L., & Klaus, M. (2023). Regulating lethal autonomous weapon systems: exploring the challenges of explainability and traceability. *AI And Ethics*, *4*(2), 229–245. https://doi.org/10.1007/s43681-023-00261-0

Curran, D. (2023). Surveillance capitalism and systemic digital risk: The imperative to collect and connect and the risks of interconnectedness. *Big Data & Society*, *10*(1). https://doi.org/10.1177/20539517231177621

Diel, A., Lalgi, T., Schröter, I. C., MacDorman, K., Teufel, M., & Bäuerle, A. (2024). Human performance in detecting deepfakes: A systematic review and meta-analysis of 56 papers. *Computers in Human Behavior Reports*, *16*. https://doi.org/10.1016/j.chbr.2024.100538

Ebrahimi, S., Abdelhalim, E., Hassanein, K., & Head, M. (2024). Reducing the incidence of biased algorithmic decisions through feature importance transparency: an empirical study. *European Journal of Information Systems*, 1–29. https://doi.org/10.1080/0960085x.2024.2395531

Filippucci, F., Gal, P., Jona-Lasinio, C., Leandro, A., & Nicoletti, G. (2024). The impact of Artificial Intelligence on productivity, distribution and growth: Key mechanisms, initial evidence and policy challenges. *OECD Artificial Intelligence Papers, 15*. https://doi.org/10.1787/8d900037-en

Ferrara, E. (2024). Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci*, *6*(1), 3. https://doi.org/10.3390/sci6010003

Frank, M. R., Ahn, Y., & Moro, E. (2025). AI exposure predicts unemployment risk: A new approach to technology-driven job loss. *PNAS Nexus*. https://doi.org/10.1093/pnasnexus/pgaf107

Franklin, G., Stephens, R., Piracha, M., Tiosano, S., Lehouillier, F., Koppel, R., & Elkin, P. L. (2024). The Sociodemographic Biases in Machine Learning Algorithms: A Biomedical Informatics Perspective. *Life*, *14*(6), 652. https://doi.org/10.3390/life14060652

Gambín, Á. F., Yazidi, A., Vasilakos, A., Haugerud, H., & Djenouri, Y. (2024). Deepfakes: current and future trends. *Artificial Intelligence Review*, *57*(3). https://doi.org/10.1007/s10462-023-10679-x

Giest, S. N., & Klievink, B. (2022). More than a digital system: how AI is changing the role of bureaucrats in different organizational contexts**.** *Public Management Review*, *26*(2), 379–398. https://doi.org/10.1080/14719037.2022.2095001

Gonzales, J. T. (2023). Implications of AI innovation on economic growth: a panel data study. *Journal of Economic Structures*, *12*(1). https://doi.org/10.1186/s40008-023-00307-w

Guo, J. (2025). The ethical legitimacy of autonomous Weapons systems: reconfiguring war accountability in the age of artificial Intelligence. *Ethics & Global Politics*, 1–13. https://doi.org/10.1080/16544951.2025.2540131

Hasanzadeh, F., Josephson, C. B., Waters, G., Adedinsewo, D., Azizi, Z., & White, J. A. (2025). Bias recognition and mitigation strategies in artificial intelligence healthcare applications. *Npj Digital Medicine*, *8*(1). https://doi.org/10.1038/s41746-025-01503-7

Kharvi, P. (2024). Understanding the Impact of AI-Generated Deepfakes on Public Opinion, Political Discourse, and Personal Security in Social Media. *IEEE Security & Privacy*, *22*, 115–122. https://doi.org/10.1109/MSEC.2024.3405963

Kohn, S., Cohen, M., Johnson, A., Terman, M., Weltman, G., & Lyons, J. (2024). Supporting ethical Decision-Making for lethal autonomous weapons. *Journal of Military Ethics*, *23*(1), 12–31. https://doi.org/10.1080/15027570.2024.2366094

Kusche, I. (2024). Possible harms of artificial intelligence and the EU AI act: fundamental rights and risk. *Journal of Risk Research*, 1–14. https://doi.org/10.1080/13669877.2024.2350720

Lengfelder, C., Tapia, H., & Biggeri, M. (2025). Navigating AI with a Human Development Compass – Shaping Tomorrow's Capabilities. *Journal of Human Development and Capabilities*, 1–10. https://doi.org/10.1080/19452829.2025.2520011

Mackin, S., Major, V. J., Chunara, R., & Newton-Dame, R. (2025). Identifying and mitigating algorithmic bias in the safety net. *Npj Digital Medicine*, *8*(1). https://doi.org/10.1038/s41746-025-01732-w

McFarland, T., & Assaad, Z. (2023). Legal reviews of in situ learning in autonomous weapons. *Ethics and Information Technology*, *25*(1). https://doi.org/10.1007/s10676-023-09688-9

Mišić, J., Van Est, R., & Kool, L. (2025). Good governance of public sector AI: a combined value framework for good order and a good society. *AI And Ethics*. https://doi.org/10.1007/s43681-025-00751-3

Novelli, C., Taddeo, M., & Floridi, L. (2023). Accountability in artificial intelligence: what it is and how it works. *AI & Society*, *39*(4), 1871–1882. https://doi.org/10.1007/s00146-023-01635-y

Neuwirth, R. (2023). Prohibited artificial intelligence practices in the proposed EU artificial intelligence act (AIA). *Computer Law & Security Review*, *48*. https://doi.org/10.1016/j.clsr.2023.105798

OECD. (2023). The state of implementation of the OECD AI Principles four years on. *OECD Artificial Intelligence Papers, 3*. https://doi.org/10.1787/835641c9-en

Osasona, F., Amoo, O., Atadoga, A., Abrahams, T., Farayola, O., & Ayinla, B. (2024). Reviewing the ethical implications of AI in decision making processes. *International Journal of Management & Entrepreneurship Research*, *6*(2), 322–335. https://doi.org/10.51594/ijmer.v6i2.773

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*, *34*(2). https://doi.org/10.1016/j.jsis.2024.101885

Rehak, R. (2025). AI Narrative Breakdown. A Critical Assessment of Power and Promise. *Association for Computing Machinery*, 1250–1260. https://doi.org/10.1145/3715275.3732083

Salvi, F., Ribeiro, M. H., Gallotti, R., & West, R. (2025). On the conversational persuasiveness of GPT-4. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-025-02194-6

Selten, F., & Klievink, B. (2024). Organizing public sector AI adoption: Navigating between separation and integration. *Government Information Quarterly*, *41*(1). https://doi.org/10.1016/j.giq.2023.101885

Shaban, O. S., & Omoush, A. (2025). AI-Driven Financial Transparency and Corporate Governance: Enhancing Accounting Practices with Evidence from Jordan. *MDPI*, *17*(9). https://doi.org/10.3390/su17093818

Sharmin, S., Biswas, B., Tiwari, A., Kamruzzaman, M., Saleh, M. A., Ferdousmou, J., & Hassan, M. (2025). Artificial Intelligence for Pandemic Preparedness and Response: Lessons Learned and Future Applications. *Journal of Management World, 2025*(2), 18-25. https://doi.org/10.53935/jomw.v2024i4.863

Singh, S., Singh, A., & Kumari, B. (2025). AI-DRIVEN ENVIRONMENTAL MONITORING SYSTEMS: A NEW FRONTIER IN CONSERVATION TECHNOLOGY. In D. Sahoo, S. Varalakshmi, S. Khairnar, & S. K. (Eds.) *Artificial Intelligence for Better Tomorrow Vol. 1* (pp. 16–29). https://www.researchgate.net/publication/394024562_AI-DRIVEN_ENVIRONMENTAL_MONITORING_SYSTEMS_A_NEW_FRONTIER_IN_CONSERVATION_TECHNOLOGY

Sun, J., Guan, X., Yuan, S., Guo, Y., Tan, Y., & Gao, Y. (2024). Public health perspectives on green efficiency through smart cities, artificial intelligence for healthcare and low carbon building materials. *Frontiers in Public Health*, *12*. https://doi.org/10.3389/fpubh.2024.1440049

Ukanwa, K. (2024). Algorithmic bias: Social science research integration through the 3-D Dependable AI Framework. *Current Opinion in Psychology*, *58*. https://doi.org/10.1016/j.copsyc.2024.101836

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Qin, Y., Xu, Z., Wang, X., & Skare, M. (2023). Artificial Intelligence and Economic Development: An Evolutionary investigation and Systematic review. *Journal of the Knowledge Economy*, *15*(1), 1736–1770. https://doi.org/10.1007/s13132-023-01183-2

Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. In *Nuffield Foundation*. Nuffield Foundation. Retrieved August 18, 2025, from https://www.nuffieldfoundation.org/wp-content/uploads/2019/12/Ethical-and-Societal-Implications-of-Data-and-AI-summary-WEB.pdf

Williams, L. (2025). *Artificial Intelligence in 2024: A Thematic Analysis of Media Coverage.* (Master's Thesis) Virginia Polytechnic Institute and State University. https://vtechworks.lib.vt.edu/server/api/core/bitstreams/ca6c755c-8f13-4e5f-b7ca-ce37a8518cf7/content

Zai, F., Rohrbach, T., & Fricker, R. H. (2025). Voices and media frames in the public debate on artificial intelligence: comparing results from manual and automated content analysis. *Frontiers in Communication*, *10*. https://doi.org/10.3389/fcomm.2025.1599854

Zaidan, E., & Ibrahim, I. A. (2024). AI governance in a complex and rapidly changing regulatory landscape: A Global perspective. *Humanities and Social Sciences Communications*, *11*(1). https://doi.org/10.1057/s41599-024-03560-x

Zeng, Y., Lu, E., & Huangfu, C. (2018). *Linking Artificial Intelligence Principles*. Arxiv. arXiv:1812.04814.

Zhang, Q. (2025). AI-driven unemployment risk and household financial decision: Evidence from China. *Journal of Asian Economics*, *99*. https://doi.org/10.1016/j.asieco.2025.101963

Zhao, J. (2024). Promoting more accountable AI in the boardroom through smart regulation. *Computer Law & Security Review*, *52*. https://doi.org/10.1016/j.clsr.2024.105939