Beyond the Giants: AI, Sovereignty, and National Security in the Emerging Global Order

Noah Fehr under the supervision of Prof. Matthias Leese
fehrno@ethz.ch

Master's in Science, Technology, and Policy: Case Study
Extension of *Data & Society*
ETH Zürich
May 16, 2025

Preface

Recent advancements in machine intelligence are rapidly reshaping national security considerations, shifting traditional power dynamics and intensifying geopolitical competition. This moment represents more than an inflection point for technological capabilities, raising broader questions about society's relationship with technology and its impact on geopolitics. Longstanding debates between technological determinism and social constructivism have resurfaced in response to the pace and scale of AI development.

This review was inspired in part by Professor Matthias Leese's *Data and Society* lecture series at ETH Zürich, which situates data-driven technologies within broader structures of political power and sociotechnical change. The implications of machine intelligence extend far beyond the innovations themselves: they raise existential questions about the future of governance, sovereignty, and global order.

While much of the discourse focuses on major technological powers, namely the United States, the European Union, and China, it is increasingly urgent for countries beyond this triad to develop and implement robust AI strategies. Dependence on foreign infrastructure and standards, limited technological autonomy, and mounting international pressure place these nations in a uniquely vulnerable position. A timely example of such considerations is Malaysia's recent consultation draft on *Sovereign AI and National AI Policy* (Shrier, Piotti, Pentland, & Faisal, 2024), which highlights the nation's attempt to articulate its concerns regarding digital sovereignty in the face of rapid technological change. Direct engagement with this draft and dialogue with Malaysian stakeholders informed this review's motivation to center perspectives from countries navigating similar constraints.

This literature review addresses the notable gap in such conversations by centering nations outside of the primary technological powers. It synthesizes academic scholarship on strategic action in the era of machine intelligence to inform countries' technology policy in relation to sovereignty and national security. As such, this review is targeted towards policymakers, international relations scholars, and technological strategists. In an effort to filter out the incessant noise around emerging machine intelligence, this review focuses exclusively on academic publications to answer the research question: How can nations outside of the traditional technological giants leverage strategic action to protect national security amidst emerging machine intelligence capabilities?

Methodology

The release of ChatGPT-3.5 in November 2022 amplified growing interest in machine intelligence, inspiring a surge of media coverage, public discourse, and academic publications. In order to both synthesize current scholarship and contextualize it within broader trends, this literature review focuses on academic publications from the last fifteen years, including sources from before and after November 2022. As noted above, this review is limited to academic journals in an attempt to limit the noise and better contextualize these trends in theoretical underpinnings. While this review itself does not introduce any novel theoretical frameworks, important constructs are discussed in detail throughout.

To narrow the scope of this review, keyword searches on Google Scholar were conducted using the terms “AI sovereignty” and “AI national security.” These search terms were selected after iterative testing of related keywords and proved especially effective at identifying relevant publications while maintaining enough breadth to include diverse disciplinary and geographic perspectives. Together, they target two critical concerns for countries outside of the dominant AI powers: the ability to govern and control technological infrastructure independently, and the strategic imperative to maintain national resilience amid AI-driven geopolitical shifts.

The search yielded over 60 peer-reviewed publications, which were then narrowed down to a final set of 24 core articles. These were drawn primarily from journals in economics, political science, international relations, and security studies. We prioritized academic publications, with a particular focus on contributions from universities and research institutions, rather than government reports or think tank outputs. Articles were retained based on the originality and relevance of their insights. Publications were excluded if they offered no substantive new perspective or if their contributions did not speak meaningfully to the concerns of countries outside the US-EU-China triad. The selection thus reflects both thematic relevance and analytical depth, grounded in academic rigor across multiple disciplines.

The final corpus consists of journal articles that address these themes directly or provide insight into adjacent issues. The review employs a qualitative, interpretive approach based on close reading and non-formalized inductive coding. Central themes were identified iteratively, ultimately consolidating into three clusters:

1. **Public Development Projects & Public Ownership of Infrastructure:** State-led or supported initiatives to build, leverage, or own different elements of the AI value chain
2. **Domestic Regulation & Governance**: National frameworks and regulatory instruments for the governance and standardization of machine intelligence
3. **International Cooperation**: Strategies and considerations for cross-border collaboration, coordinated capacity building, and global standards.

Together, these clusters represent the most commonly cited modes of strategic action discussed in the literature. They offer a useful framework for understanding how nations outside the dominant AI triad can assert agency in this evolving technological landscape. For the remainder of this review, each cluster will be reviewed in detail, exploring specific facets of sovereignty and national security with regard to its broader theme.

Public Development Projects & Public Ownership of Infrastructure

The past two decades have marked unparalleled progress in machine intelligence and data infrastructure; however, these advances have been centered in the hands of the technological elite (Bibi, 2024). For countries beyond the AI powerhouses, these advances demand action in order to protect and control their technological infrastructure. Through applications in the public sector, specific considerations regarding procurement, development of infrastructure, and broad innovation policy, such nations have the opportunity to reduce dependency on foreign platforms, assert stronger control over data flows, and shape technological development. Strategic public investments serve as a primary lever for states to bulwark against geopolitical vulnerabilities. For policymakers and scholars concerned with resilience and autonomy in a digitized global order, understanding these projects provides insight into how states can actively reconfigure their role in the machine intelligence value chain rather than remain passive recipients of external technologies.

Public Administration

As artificial intelligence technologies continue to proliferate, national governments face a pivotal opportunity and responsibility to apply machine intelligence within their own public administrations and projects. Public application of these advancements is not merely a matter of efficiency, but rather a question of the strategic capacity of the state as machine intelligence increasingly determines the action space of federal administrations both domestically and abroad. Engstrom, Ho, Sharkey, and Cuéllar (2020) found that upwards of 45% of federal and state agencies in the U.S. had begun experimenting with AI in 2020; such figures have likely increased given recent excitement around these technologies. The contracting system in the U.S., however, means that much of this integration has occurred vis-a-vis government contractors, reinforcing dependence on private firms and infrastructure. While the private sector of the U.S. leads the world in machine intelligence, its public sector lags behind its geopolitical rival, China, who has taken a more an integrated, assertive approach by embedding AI directly into state functions through centrally coordinated policies outlined in the *Next Generation Artificial Intelligence Development Plan*. For nations outside of these superpowers, a comparative understanding of these approaches is imperative to effectively define their strategic approach.

China's release of the *Next Generation Artificial Intelligence Development* plan marked a broad and coordinated approach of the national government to bring machine intelligence into government operations. Historically, China has been a global leader of smart cities and digitization, with a "focus on surveillance technologies" (Roberts, Cowls, Morley, Taddeo, Wang, & Floridi, 2021, p. 20); this sweeping vision of coordinated AI development in China can be understood as an extension of this pattern. Notably, however, this approach encompasses more than pure technological strategy; the application of these technologies also serve the pre-defined social governance goals of the Chinese Communist Party. Roberts et al. (2021) highlight the enactment of a social credit system, enabled by applied machine intelligence, as an attempt to rectify the perceived moral decline in China's post-Mao period. Although such moral governance may clash with liberal democratic norms, it exemplifies the degree to which national governments' application of machine intelligence can serve both economic and political goals. Nations should cautiously look to China as an example of public influence over technological development, recognizing the serious risks of prioritizing control over individual rights.

In contrast to the centralized approach of the Chinese, Estonia presents a compelling democratic alternative. Despite its comparatively small technological sector, especially next to nations like China and the US, Estonia has transformed its public administration through e-citizenship and digital-first governance. The small Baltic state has online public services for typical administrative functions like tax filing, with over 70% of Estonian citizens registered with an e-citizenship and able to access this technology-enabled service (Kerikmäe & Pärn-Lee, 2021). Recent policy proposals and pilot projects extend these programs, with "regulation [enabling machines] to make administrative decisions without human interaction" already passed and test projects running in administrations like social benefits (Kerikmäe & Pärn-Lee, 2021, p. 5). The Estonian model is especially instructive for smaller states, demonstrating the value of applied machine intelligence in public administration to reinforce existing democratic norms. Furthermore, public procurement and application of these technologies, as discussed later in this review, accelerates technological development in such nations.

The American, Chinese, and Estonian approaches to artificial intelligence in public administration underscore the importance of how these technologies are applied. These decisions shape both technological development and social norms, in turn affecting the political trajectory of nation states. The efficiency gains, expanded capabilities, and technological spillovers of public sector application make a compelling case for integrating machine intelligence into existing government functions; the challenge for modern nation states is how to craft political action which serves both their technological goals and their broader vision for the country.

Build vs. Buy

As demands grow for more robust integration of artificial intelligence in federal governance and public agencies, states are faced with the classical tradeoff of procurement: to build or to buy. This decision between internal builds or procurement from external vendors introduces significant questions of institutional control, public accountability, and national capacity. Building solutions in-house not only enables governments to tailor solutions for complex tasks, it also promotes alignment with legal and policy norms. Engstrom et al. (2020) develops this view, noting that developing tools internally promotes systems that are "more likely to be designed and implemented in lawful, policy-compliant and accountable ways." As government teams build applications for their specific application, they expect greater principal agent alignment and therefore better adherence to public sector considerations, which are often more extensive than in the private sector.

More broadly, the choice to build internally extends benefits beyond the immediate output: it builds a technically literate civil service, strengthens internal institutional capacity, and positions governments to be active participants, rather than passive consumers, in machine intelligence. As states seek to modernize their digital infrastructure, the externality of stronger public-sector talent is a critical asset. Such development comes at a cost, however, and many governments now depend heavily upon external contractors for any technological innovation. For public institutions tasked with balancing innovation and public interest, navigating this reality is critical.

Cross-industry analyses of different procurement strategies emphasize the complementary behavior of procurement, rather than direct substitution (Hoffreumon, Forman, & van Zeebroeck, 2024). With the notable exception of traditional IT, all sectors "exhibit strong evidence of complementaries," indicating the value of developing and refining these technologies in tandem with procuring private solutions (Hoffreumon et al., 2024, p. 471). For governments and federal agencies alike, this model of refinement and application exemplifies the opportunities in procurement; instead of creating deep dependencies on external consultants and vendors, public institutions can leverage ready-made technologies and apply them to specific contexts, following with the principal agent alignment described by Engstrom et al. (2020). This model not only reduces long-term reliance on third-party vendors, but also creates a feedback loop of capacity building, where the process of internal adaptation itself enhances technical expertise. As Hoffreumon et al. (2024, p. 473) conclude, the widespread presence of such complementarities suggests that "investments in developing or adapting software may help [organizations] to develop skills that they use to customize and extend investments in ready-made software.”

For governments navigating the transition towards more robust AI adoption, private procurement and external talent may serve as a necessary bridge; however, long-term strategic autonomy will depend on cultivating in-house talent and retaining control over mission-critical algorithms. As such, procurement and hiring decisions across federal agencies must understand this as strategic action to protect sovereignty and build lasting capacity.

Infrastructure Development

Underlying all of the rapid innovation in machine intelligence lies an interconnected web of foundational infrastructure. We can conceptualize such networks in layers, from the physical infrastructure of the energy grid and data centers to the highly specialized applications for large language models. While much of this is sidelined from public discourse on machine intelligence, choices surrounding the development, ownership, and operation of infrastructure will shape strategic positioning. Each layer introduces new complexities and dependencies for countries to weigh when considering public action.

As technology shifted away from on-premise to cloud computing over the past two decades, dominant firms like Amazon Web Services and Microsoft Azure rose to power, serving as the backbone for subsequent innovation and development. With the promise of efficiency and reduced operational costs, governments have also shifted towards these technology giants, dramatically shifting the nature of relationships between states and private actors (Irion, 2013). The risks of such power have been clear from the beginning, as governments often were forced to weigh the potential risks to public IT systems often slowing their development, prompting Irion to argue in 2013 that national sovereignty is conditional upon adequate data sovereignty. The combination of recent concerns of data security and the promise of machine intelligence have heightened such concerns over data control, data ownership, and the nature of the relationship between public actors and private data services.

In response, countries like Indonesia are investing in colocation data centers and publicly funded digital infrastructure to "reterritorialize" data and regain control over compute resources (Irion, 2013; Nugraha, Kautsarina, & Sastrosubroto, 2015). Estonia exemplifies the provision of infrastructure as public goods, using its *X-Road* interoperability framework to link public and private sector systems: “X-road is the backbone of e-Estonia. Invisible yet crucial, it allows the nation’s various public and private sector e-service information systems to link up and function in harmony" (Kerikmäe & Pärn-Lee, 2021, p. 2). This project enables secure collaboration and coordination between the private and public sectors, laying the groundwork for future AI innovation from Tallinn and beyond. Prioritizing security and facilitating innovation exemplifies that infrastructure extends beyond raw compute and electrical power to systems that reinforce long-term resilience.

Reclaiming and developing digital infrastructure stands as a fundamental challenge for even the most advanced states. The convergence of distributed computing and superintelligent algorithms has challenged traditional state power. Kello (2017) argues that the rise of algorithmic capabilities creates a 'sovereignty' gap in which both state and non-state actors exploit these technologies (Timmers, 2019). This directly aligns with the conditionality of state power upon control of data, emerging technology and the underlying infrastructure expressed by Irion (2013). Timmers (2019) frames the action space for nation states in three main approaches: risk management, strategic partnerships, and the promotion of common goods. In the context of infrastructure development, we can map this to the policy recommendations of Irion (2013), including clear procurement guidelines, geographic restrictions on service providers, and domestic service providers. Without these measures, the integrity of the digital foundation for modern governance risks instability.

The need for control over computing capacity, data storage, and other critical infrastructure has moved to the center of policy debates across the globe. Existing literature builds upon dated paradigms ill fit to answer questions of training data, the accessibility of online services, and the use of open source models. Machine intelligence has challenged the fundamental definitions of such policy, blurring the specifics of what constitutes 'critical' infrastructure. The European Union, for example, has recognized the centrality of computing capacity and data storage to AI sovereignty, yet continues to struggle with implementation due to the complexity of multinational governance and evolving definitions of technological capacity (Mügge, 2024). These gaps leave smaller states especially vulnerable yet also offer a critical window for strategic investment and institutional innovation. For countries seeking to assert agency in a world increasingly defined by machine intelligence, building, governing, and securing infrastructure must be treated as a first-order strategic priority.

Innovation Policy

Global excitement about the possibilities of artificial intelligence has excited financial markets, individual companies, and governments alike. As nations grapple for the leading position in artificial intelligence development, they have enacted varied approaches to innovation policy in an attempt to spur domestic innovation. Innovation policy itself will serve a pivotal role in determining the future of this technology, the control over it, and the geopolitical consequences of such changes. Two pivotal tensions underpin national approaches to AI innovation: centralization and technological determinism.

Especially for larger nations with robust technology sectors, the question of national coordination is imperative. China's Artificial Intelligence Development Plan (AIDP), which aims to make China the "world centre of AI innovation by 2030", offers a seemingly nuanced approach to balance the need for coordination with the inefficiency of centralized control by offering direction and local autonomy for testing (Roberts et al., 2021, p. 5). Their plan incentivizes local projects, with "local governments becom[ing] hotbeds for testing and developing central government policy," in a broader effort to inform their unified strategy. Following traditional innovation policy, the government also has plans to select and elevate national champions, or promising private sector leaders in AI development (Roberts et al., 2021, p. 8).

While local sandboxes enable experimentation and meaningful innovation at the margins, they also risk policy asymmetries and subsequent disruption of innovation. Conversely, strong centralization can stifle bottom-up innovation and exacerbate bureaucratic inertia. Countries and multilateral institutions must therefore weigh the trade-offs between coherent national strategy and the inevitable inefficiencies and redundancies of decentralized experimentation. For smaller nations, the stakes are even higher: emulating centralized strategies without the same resources can lead to overreach, while full decentralization risks fragmentation and waste. Moreover, targeted innovation policy spurred by the growing development race shifts greater power to the state. Mügge (2024) argues that to maintain jurisdictional sovereignty, public authorities must champion and promote domestic AI companies, which may fail to serve the competing goals of citizen sovereignty. Mügge (2024) furthers that the EU's AI policy prioritizes jurisdictional independence, indicating a preference for state-centric control and global competitiveness rather than citizen empowerment.

This introduces broader concerns of technological determinism; nations must evaluate their degree of agency to shape the development and deployment of these technologies on a global scale. For nations outside of the primary AI superpowers, technological determinism may more accurately represent reality and their policy options. While Mügge (2024) advocates for norm setting in innovation policy, this perspective is largely eurocentric and may challenge other policies focused exclusively on global competitiveness. When norm setting stands at odds with technological progress, global markets will challenge these norms. Only the nations or coalitions with such broad regulatory influence will be able to set norms to a meaningful degree. Even with such constraints, states do maintain agency when making choices around investment in public research, supporting domestic infrastructure, and setting boundaries for applications.

By navigating the balance between central control and local innovation, and between norm-setting and technological adaptation, all states can carve out space for sovereign development in an increasingly asymmetric digital order.

Challenges and Unique Risks

The vast opportunities for public procurement, application, and development of machine intelligence is best understood in the context of the unique limitations and risks to public involvement. While private entities can afford to follow the Silicon Valley ethos of 'move fast and break things,' public actors' approach must be more deliberate and calculated. The risk reward tradeoff for public entities slants towards risk minimization over value maximization, slowing the adoption of new technologies. This tendency is a necessary feature of democratic governance with large legal liability; recognizing this is imperative to designing public involvement in these innovations (Desouza, Dawson, & Chenok, 2020). For democratic societies, questions of political feasibility and stakeholder support further complicate such application. Broader stakeholder groups and wider public involvement in decision making may create more deliberate outcomes but at the expense of speed (Desouza et al., 2020). Furthermore, this public accountability places additional requirements upon machine intelligence systems (explainability, transparency, rigorous oversight) that private sector entities do not face (Desouza et al., 2020; Engstrom et al., 2020; Hickok, 2024).

Discussions of the challenges faced by public entities in their deployment of artificial intelligence tend to focus on the comparison between public and private actors. To extend this distinction, we can further divide public actors into democratic and autocratic regimes. Without overcomplicating the semantics, this allows us to compare nations with a large body of stakeholders in decision-making versus nations with a small body of stakeholders. For democratic regimes with diverse stakeholders, governance requires a society-in-the-loop architecture with open discourse and many ideas (de Almeida, dos Santos, & Farias, 2021); autocracies, centralized governments or even other government agencies without this larger set of stakeholders will be able to move faster and deploy solutions more quickly. While this approach introduces greater risk, such agility may favor these entities to build and innovate more quickly. Democratic regimes must evaluate how to integrate such agility while maintaining the necessary guardrails of public governance. Leaning too far away from citizen oversight will inevitably doom any public projects, as "political support for a more effective and tech-savvy government will evaporate quickly" (Engstrom et al., 2020). Political norms and institutional structures will dictate the policy opportunities for states exploring machine intelligence capabilities.

Ultimately, the public sector's success in AI will depend not only on what technologies it adopts, but on how it adopts them and who it includes in the process. The path forward requires more than just technical capacity but also institutional buy-in and the creation of long-standing systems.By aligning innovation strategies with democratic values, states can create public AI systems that are not only efficient but also trustworthy, inclusive, and resilient.

Domestic Regulation and AI Governance

In parallel to the growing arms race of AI development, countries are scrambling to establish guardrails and domestic governance of machine intelligence in order to protect their governments and their citizens. Regulatory action can serve to mitigate external dependencies and align AI deployment with national values. Although regulation on such innovations is shaped by international standards and influence (Brussels effect, California effect), the mechanisms for enacting such controls vary widely across national and cultural borders. The details of such legislation will govern data rights, algorithmic transparency, and institutional oversight for coming decades, in turn determining the extent to which countries can maintain sovereign control and national security in this new era.

Regulatory Aims

As detailed by Daniell (2014), public policies at all levels are directly informed by cultural norms; AI regulation has perfectly exemplified this with the diverging aims of and diverse values embodied in global legislative responses to this technology. Two key axes of divergence, Eastern versus Western philosophical traditions and corporate versus individual interests, highlight how governance structures encode deeper questions about power, identity, and control.

Legal scholars' and politicians' critiques of emerging technology and its subsequent regulation often include normative statements about the importance of "ethical principles" (Taeihagh, 2021). Such arguments are best understood in the context within which they were issued; the governing ethical principles in the EU AI Act fail to hold the same weight for lawmakers in Beijing. Díaz-Rodríguez et al. (2023) outline the liberal, Western perspective on AI in their conceptualization of trustworthy AI. Their analysis focuses heavily on the individual, with strong emphasis on human agency, fairness, transparency, and accountability. The auditability, explainability (XAI), and non-discrimination requirements center the individual in discussions of rights in the era of artificial intelligence (Díaz-Rodríguez et al., 2023). This model reflects a broader political commitment to personal freedom and checks on state and corporate overreach: a measured, cautious approach that seeks to preserve individual sovereignty in a rapidly digitizing world (Díaz-Rodríguez et al., 2023).

By contrast, many Eastern perspectives, particularly in the case of China, frame AI development through a collectivist lens. China's National Artificial Intelligence Development Plan, for instance, "emphasises that, above all else, AI development should begin from enhancing the common well-being of humanity" (Roberts et al., 2021, p. 15). Understood through this framework, China's limited privacy protections and expansive data collection practices are not regulatory oversights; they are deliberate choices that prioritize societal outcomes, even at the expense of Western values surrounding individual autonomy. This approach asserts a different vision of sovereignty, one in which the state exercises centralized control to shape collective prosperity, resurfacing questions from Mügge (2024) in his discussion of jurisdictional versus individual independence.

Beyond cultural norms, nations are divided in their ideal balance between corporate and individual interests. While most countries consult experts when crafting regulation, the selection and influence of those experts, especially those coming directly from industry, varies greatly. In the U.S., for example, the close relationship between regulators and major tech firms has sparked concerns about regulatory capture: "The deep involvement of industry stakeholders in developing ethical principles and regulations for AI raises concerns that corporate interests dominate AI regulations" (Taeihagh, 2021, p. 4). While too strong of private influence risks regulatory capture and misalignment of regulation, industrial expertise is a necessary element of informed policy, especially given the growing asymmetry of information between innovators and regulators. All nations, but especially nations with less robust technology sectors, must critically formulate plans for the integration of technological expertise into regulatory decision making without risking misaligned policies.

Domestic regulatory environments will continue to dictate the flow of capital, innovation, and power as the relevance of machine intelligence grows. While policy will diffuse from primary actors like the EU, the U.S. and even China, sovereign states must evaluate the cultural and industrial context supporting such norms. Intentional adaptation to these global standards with respect to the domestic environment is critical for effective domestic governance.

Regulatory Instruments

The rapid rate of innovation and change as a result of the proliferation of machine intelligence has overwhelmed the conventional mechanisms of policy making. Without effective mechanisms for AI governance, states risk ceding regulatory influence to foreign powers or multinational corporations, compromising both national security and democratic legitimacy. This explosion of novel capabilities pushed regulators to explore novel mechanisms for regulatory enforcement in order to keep up with these rapid developments. Many of these new mechanisms remain in their infancy as pilot programs in testing, although some frameworks have become established with institutional backing. An evaluation and definition of these regulatory mechanisms is critical for nations formulating their own governance frameworks.

Like other novel technologies, the broad application of machine intelligence requires standardization to prompt stability and security for all actors. Nation states, private companies, and international organizations have raced forward in their development, following the game theoretic first-mover advantage in standard setting. Such guidelines apply to all levels of the technology stack, from standards for encryption schemes to frameworks for connecting autonomous agents. Indonesia, for instance, has centered their efforts lower on the technology stack, focusing on cybersecurity and networking standards. In total, the Indonesian government has defined 25 specific cyber defense requirements including System and Communication Protection (SCP) and National Cryptographic Standards (NCS) in order to mitigate risk from foreign intelligence services, protect data sovereignty and in turn maintain national security (Nugraha et al., 2015). For other nations, especially those who depend on other countries for innovation in machine intelligence, scoping specific requirements and regulation in the form of baseline standards can help to protect their domestic interests from systemic risk.

The European Union has implemented the largest body of AI-specific standards in the EU AI Act, detailing a risk-based framework with stringent obligations for risk and conformity assessments before market entry (Díaz-Rodríguez et al., 2023). While early drafts of this legislation remained relatively high level and subsequent discussion questioned the efficacy of these efforts, the AI Act prompted calls for further standards including standardized auditing practices to evaluate algorithms (Desouza et al., 2020) and forcing the inclusion of mechanisms for human oversight (Díaz-Rodríguez et al., 2023). Such standardization of requirements for deployed AI systems enables the European Union to maintain control over these applications and mitigate underlying risk. Conversely, Roberts, Babuta, Morley, Thomas, Taddeo, and Floridi (2023) argue that the United Kingdom's lack of shared standards and certification schemes pose significant challenges to British leadership in AI. Even as Roberts et al. (2023) advocate for a market-based approach for assurance and regulatory oversight, they still emphasize this need for standards. The contrast between European leadership and British inaction emphasize the value of standard setting as a regulatory aim.

Extending the action space of regulators, scholars have echoed calls for an update to existing legal liability frameworks in response to widening acceptance of machine intelligence. While conventional legal systems focus on people or groups of people as the subject of legal liability, law is broadly unprepared to answer questions of liability allocation with autonomous AI systems (Timmers, 2019). Such a gap continues to grow as emerging applications, like fused management systems and the presence of AI on corporate boards, continue to guide machine intelligence into positions of greater power. Both to protect the interests of the public and to establish consistent guidelines for AI deployment, judicial systems must designate liability throughout the AI supply chain, from manufacturers and developers to end users (Taeihagh, 2021). Without clear ownership and liability, enforcement becomes murky, and sovereignty erodes as states lose their ability to respond to harm effectively.

In the EU AI Act, the authors specify that civil liability may arise from harm caused by high-risk AI systems in which “[the victim] may claim compensation from the provider of that system” (Schuett, 2024, p. 17). This design is intended to ensure accountability, but still leaves unanswered questions as to the constitution of a ‘provider.’ Estonia has extended this with a proposed framework in which the end user of an AI application holds the responsibility (Kerikmäe & Pärn-Lee, 2021). While clearer in its definition of liability, this approach raises questions as to the role of model developers in this framework. Nations must grapple with these dilemmas to create a clear definition of liability amidst the growing prevalence of deployed, autonomous AI systems. This definition of ownership will ease administrative burden and facilitate innovation by providing transparent guardrails and responsibility.

Regulatory Oversight & Implementation

Amidst broader conversations of regulatory responsibility and subsequent legal liability, the administrative burden of such governance looms over many efforts. Such a burden demands the provision of additional resources for the codification and enforcement of these norms. These new administrative structures generally follow one of two patterns: a centralized, dedicated agency to machine intelligence or a distributed effort across existing agencies. The United Kingdom has followed this sector-led approach, delegating AI governance to existing regulators within their specific domains. Roberts et al. (2023) argue that this decentralized approach facilitates the development of specific governance measures that reflect the unique risks associated with different AI applications. While this gives subject matter experts agency to govern their expertise, a decentralized approach also risks inconsistencies and regulatory gaps, a notable consideration with the emergence of general purpose AI systems (Roberts et al., 2023). In an effort to address such concerns, de Almeida et al. (2021) advocates for the formation of a dedicated regulatory agency distinct from existing governmental organizations. The regulatory agency would need strong relationships with the legislature, industries, and service providers to enable agile regulation and accurate consideration of regulatory trade-offs.

Even after the establishment of such organizations, the enforcement of AI governance policies is cumbersome. The distributed nature of machine intelligence challenges traditional forms of enforcement and demands novel policies to guide its development. Several parties have noted a growing interest in the creation and application of regulatory sandboxes as a means to test and validate AI systems in a controlled environment before market entry (Díaz-Rodríguez et al., 2023; Roberts et al., 2023). As most machine intelligence deployments are binary (the technology is deployed or not), regulatory sandboxes allow for experimentation and compliance checks prior to deployment (Díaz-Rodríguez et al., 2023). The United Kingdom has already begun experimenting with these sandboxes, seeing success in the initial stages (Roberts et al., 2023). This gives an environment for innovators to ensure compliance and for regulators to experiment with the actual impact of their policy designs. Regulation will need to explore innovative approaches like this in order to effectively shape the course of AI development; the constant novelty and emergence of fresh capabilities will continue to challenge both existing and future governance efforts.

Existing scholarship on technology policy conceptualizes this uncertainty and regulatory risk as the Collingridge dilemma. Many experts point towards a need for dynamic regulation that can adapt with the emergence of machine intelligence capabilities (Díaz-Rodríguez et al., 2023; Timmers, 2019). Justo-Hanani (2022) emphasizes the need for incrementalism for policymaking as a key element to facilitate this agility while maintaining consumer protection. Even with incrementalism and this hypothetical agility, the Collingridge dilemma persists. This tradeoff must remain at the center of conversations around national AI policy. To deepen the nuance of these regulatory decisions, Roberts et al. (2023) challenge conventional thinking on the tradeoff between regulation and innovation: he posits that less regulation does not always facilitate greater innovation as the weakening of regulatory bodies may harm innovation by eroding public trust in AI technologies. Smuha furthers that consumer trust in AI technologies "is not only a prerequisite for the adoption of the technology but can also be priced." (Smuha, 2021, p. 7).

Amidst the complexity of these tradeoffs, myopia in representative democracies, and the lack of clear mechanisms for agile policymaking, nations must chart an intentional course in their policy efforts. The definition, implementation, and enforcement of these policies is paramount for effective AI governance and maintaining agency amidst these technological shifts.

International Cooperation

The complexity and portability of recent developments in artificial intelligence render it elusive to traditional governance regimes (Judge, Nitzberg, & Russell, 2024). Stringent regulation and development limitations in one state do not exclude its development abroad or even the eventual deployment of new tools in the original state. Conversely, Desouza et al. (2020) argues that a unified approach to data sharing and AI governance can enhance global security and ethical standards. International cooperation is driven not by altruism or broader concerns for humanity, but rationality: countries' engagement in bilateral and multilateral efforts to address AI governance reflects an understanding of the systemic risk and downstream effects of this development (Smuha, 2021). As we accept that states are largely rational and have the capacity to engage internationally in order to bolster their national security and further domestic interests, nations must consider the efficacy and tradeoffs of global data sharing as well as international agreements.

Data Sharing & Data Sovereignty

Discussions of data sharing and date sovereignty date back decades, entering the spotlight with the emergence of global networking, and then with distributed computing. The Snowden revelations and legislation like the US PATRIOT Act highlighted the vulnerability of traditional networking architectures to powerful foreign actors (Polatin-Reuben & Wright, 2014). The concerns from this era were multiplied with the emergence of cloud computing, which at its core challenges the traditional notion of sovereignty by decoupling the owner of the data from the management and storage of this information (Irion, 2013). With insecure transmission of information and data can be stored across multiple jurisdictions, governments looked to prioritize data security.

With such context, the literature has characterized these efforts as 'data sovereignty.' Data sovereignty is a governments' effort to exclusively control its data and public assets, from storage to use and deployment

(Nugraha et al., 2015; Polatin-Reuben & Wright, 2014). Initiatives to promote data sovereignty can be broadly divided into weak and strong: weak data sovereignty allows for private-sector initiatives, emphasizing digital rights whereas strong data sovereignty focuses on the state with the top priority as safeguarding national security (Nugraha et al., 2015; Polatin-Reuben & Wright, 2014). As data lays the foundation for all recent developments in machine intelligence, a deep understanding of these efforts is necessary for global policymakers.The mechanistic differences here afford broader ideological differences which will be discussed in further detail below.

Under the USA PATRIOT Act and the broader American foreign surveillance complex, foreign data was subject to inspection by American authorities for decades (Irion, 2013, p. 20). Even beyond the legal avenues to inspect data, the weakness of the Border Gateway Protocol allows for hijacks by both state and non-state actors, meaning data transmission holds inherent risk (Butler, Farley, McDaniel, & Rexford, 2010). Even prior to the Snowden revelations in 2013, the Final Acts of the World Conference on International Telecommunications were proposed to establish standards and weaken the legal basis for countries in the Five Eyes to surveil the rest of the world. This resolution failed, with opposition from G20 states including the US and support from the nations which would come to form BRICS (Polatin-Reuben & Wright, 2014). The growing opposition to the Western surveillance regime was only strengthened with the outcry that followed Snowden's release of classified documents, empowering states to actively pursue robust information security measures in the name of national security (Nugraha et al., 2015; Polatin-Reuben & Wright, 2014): Brazil's "Marco Civil da Internet" attempted to force foreign cloud providers to store data domestically (later withdrawn); Russia and China both enforced strict data localization regimes; Indonesia forced all public service data to be stored domestically, although public service data has been left ill defined (Polatin-Reuben & Wright, 2014).

All of these efforts and subsequent discussions of reterritorialization can be understood as strong data sovereignty. As cloud providers often reveal little information about the underlying infrastructure supporting their offerings, data sovereignty laws provide legal basis to force this information into the public eye and, in turn, enforce some degree of control over the data (Hummel, Braun, Tretter, & Dabrock, 2021). Governments active interest in such control has spurred technological innovation with technologies like federated learning and differential privacy; these advances, along with broader data reterritorialization efforts, enable both public and private actors to use new advances in machine intelligence while maintaining compliance with the broader national security aims of the state (Díaz-Rodríguez et al., 2023). While data sovereignty and some of these technologies are more than a decade old, the relevance of these themes has only grown as machine intelligence increases societal appetite for data and exposes the systemic risks of sharing this data carelessly.

Countries must evaluate their domestic circumstances and their place in the broader global order when evaluating policy options of data sovereignty. Defining the objective as well as the primary beneficiary of data ownership is paramount. Hummel et al. (2021) point to the incompatibilities between data sovereignty of individuals and data sovereignty of the population, or a state. This tenuous balance between liberty and collective security is further discussed by Polatin-Reuben and Wright (2014): "A strong approach to data sovereignty would provide a pretext for expanding censorship activities within individual BRICS countries at the expense of privacy and freedom of expression." Certain policies of data ownership and control afford subsequent establishment or infringement of norms domestically, but also internationally. Complete data sovereignty of the state, by completely rejecting foreign technology companies, would subsequently weaken trade and international cooperation under the globalist regime of the early twenty first century (Polatin-Reuben & Wright, 2014). Polatin-Reuben and Wright (2014) describes this Internet fragmentation as 'Balkanization,' a set of isolated national intranets which suffer from problems of scale, as well as economic and social difficulties. They argue that stronger coalitions will need to be formed to address this, and that a potential organization of BRICS countries' efforts for data sovereignty would form a 'formidable bloc' in global Internet governance debates (Polatin-Reuben & Wright, 2014). Notably, NetMundial in 2014 marked an initial step in this direction, but as a non-binding, multi-stakeholder agreement it lacked the teeth to establish such a coalition.

International cooperation continues to be incredibly challenging; however, a comprehensive foreign relations strategy must accompany any domestic efforts for data sovereignty. These issues cannot be understood only at a domestic level, especially as states evaluate systemic national security risks as a result of such dependencies. Excluding the U.S. and China, nations must pursue a collaborative approach if they hope to gain any leverate in global conversations of machine intelligence. As such, data sovereignty marks a meaningful step in the protection of national security interests; however, such efforts must be bounded by the realities of the global landscape.

Regulatory Collaboration, Competition

The global nature of machine intelligence and the subsequent demand for international coordination introduces critical considerations in states' foreign policy strategies. Larger states, like the U.S. and China, can afford to take unilateral action with a winner-take-all approach; however, for smaller nations and countries without the robust tech ecosystems of the U.S. and China, collaboration is necessary. Furthermore, within the global context, strategic games emerge with regulatory floors and ceilings as well as standard setting. A strong understanding of these dynamics can help state officials to shape their AI initiatives in accordance with their positioning in this complex game.

Especially with growing concerns of the military applications of machine intelligence, major powers like the U.S. and China have shifted towards unilateral approaches, aiming to supercharge their domestic development and hinder international competition (Timmers, 2019). President Xi Jinping's declaration that 'only the innovators win' encompasses the broader Chinese approach to leveraging the immense power of the state to accelerate innovation in machine intelligence, specifically for military applications (Roberts et al., 2021). Statements from senior officials from both China and the U.S. suggest a belief in cooperation and arms control (Roberts et al., 2021); however, a cynical perspective sees proliferation of machine intelligence systems in the military as a classic Prisoner's dilemma, not a coordination game: both superpowers have the incentive to signal cooperation while maintaining incentives to deviate from this signaling (Han, Santos, Pereira, & Lenaerts, 2021). Ultimately, the validity of this analogy will depend on the nature of the technology itself; open-source paradigms and increasingly efficient models challenge the perceived exclusivity of next-generation artificial intelligence, possibly disrupting this simplistic representation of this model.

Regardless of the accuracy of the Prisoner's Dilemma comparison, the vast majority of states are not engaged in such a dynamic. The massive scale of data and global nature of AI risk necessitate a coordinated approach to enhance global security and ethical standards (de Almeida et al., 2021; Desouza et al., 2020; Taeihagh, 2021). Even purely rational stakeholders only concerned with their self-interest have begun to engage in both bilateral and multilateral efforts in recognition of the global nature of AI's challenges (Smuha, 2021). Given the transnational nature of the supporting infrastructure for machine intelligence (cloud computing, data sharing, etc.) and the portability of the final technology itself, jurisdictional authority becomes extremely complicated (Irion, 2013). The existing legislation and international governance, like the Universal Declaration of Human Rights, may help us to outline broader principles but broadly lack the necessary specificity and enforcement mechanisms to address these novel risks (de Almeida et al., 2021). As such, the need for international cooperation in dictating AI policy is clear, but the details of implementation remain murky.

International coordination has always suffered from the same fundamental shortcoming: enforceability. Timmers (2019) draws parallels between cyberspace and the Montreal Protocol, one of the most successful examples of international cooperation: cyberspace should be treated as a global common good with supporting standards of ethics and privacy standards. Existing literature points towards the issuance of globally recognized security certificates to give legitimacy to products, either on a national or international level, and even the establishment of a global committee to coordinate regulatory efforts (de Almeida et al., 2021). All of these proposals add costs for developers and member states in order to support a broader global objective, leaving all individual states with clear incentives to deviate.

Despite such tendencies, there are mechanisms for the proliferation of these standards and regulations. The European Union is a strong example of a stronger international organization with the capacity to impose norms on its member states, but also broadly via the Brussels effect. In the case of cloud computing and artificial intelligence, the EU worked to establish an internal market with base standards thus preventing market fragmentation amongst member states and requiring all external actors to comply in order to engage with European customers (Irion, 2013; Justo-Hanani, 2022). Roberts et al. (2023) acknowledge the influence of such standards on British companies and British lawmakers as they respond to this emerging technology. The EU can be understood as a special case because of its strong competence in establishing standards and internal markets. Even Westminster lacks the jurisdiction to define explicit AI policy for all of its devolved nations, perhaps enabling dissonance in regulatory divergence within the UK (Roberts et al., 2023). The secondary effects and standard setting capabilities of a given state or organization vary widely based on market size, global relevance, and institutional jurisdiction; the EU and the UK exemplify how certain actors maintain standard setting power whereas other smaller actors are subject to align with the standards of larger players.

The question of regulatory and strategic alignment introduces two game theoretic mechanisms as states work to navigate the international machine intelligence development landscape. First, the competition for regulatory leadership has come to a head in international standard setting bodies as the U.S. and China grapple to set global standards for the creation and deployment of machine intelligence technologies (Smuha, 2021). In this coordination game, there is a great first mover advantage as the first standards to be established will likely follow for the rest of the innovation. This dynamic is critical for superpowers, but does not hold as much pertinence for smaller states. The second game, however, involves every state as a player. States' efforts to create AI policies create dynamics of regulatory competition, as nations work to balance data sovereignty and citizen protection with an attractive environment for innovation and deployment of machine intelligence (Smuha, 2021). Individual efforts to regulate AI or facilitate its development can therefore only be understood in the broader international landscape, not in isolation. For leaders of sovereign nations and development initiatives, these dynamics of political coordination and competition are critical to take into account when evaluating policy alternatives. Especially for nations with nascent technology sectors or relatively small market power, careful navigation of this environment is critical to protect their states' sovereignty and by extension national security interests.

Conclusion

Nations outside of the United States, European Union, and China face a dual imperative: to navigate the profound opportunities AI offers while actively mitigating its geopolitical risks. This review has examined a curated body of academic literature that highlights how countries on the periphery of AI superpower status can nevertheless exercise agency and shape their own technological futures.

The three thematic clusters, public development and ownership, domestic regulation, and international cooperation, offer a roadmap for strategic action. Together, they reveal that sovereignty in the AI age is not solely a function of raw computational power or data volume, but also of vision, governance, and alliances. From state-backed data infrastructures to adaptive regulatory regimes and new forms of cross-border collaboration, nations are developing creative approaches to resist dependency and assert influence.

Importantly, this review underscores the need to move beyond reactive or copycat policy approaches. Nations that approach AI strategically, by tailoring development to local needs, cultivating regional alliances, and building regulatory capacity, can carve out meaningful space for sovereignty, even within an asymmetric global landscape. Ultimately, the future of machine intelligence will not be shaped solely by the technological giants. It will also be defined by the cumulative actions of countries that refuse to be passive subjects of technological change.

Works Cited

Bibi, P. (2024, October). Antitrust reform in AI-driven markets: Tackling the challenges of data monopolies. https://www.researchgate.net/profile/Palwasha-Bibi-2/publication/384967884_Antitrust_Reform_in_AI-Driven_Markets_Tackling_the_Challenges_of_Data_Monopolies/links/670fc42524a01038d0f04449/Antitrust-Reform-in-AI-Driven-Markets-Tackling-the-Challenges-of-Data-Monopolies.pdf

Butler, K., Farley, T. R., McDaniel, P., & Rexford, J. (2010). A survey of BGP security issues and solutions. *Proceedings of the IEEE, 98*(1), 100–122. https://doi.org/10.1109/JPROC.2009.2034031

Daniell, K. (2014). *The role of national culture in shaping public policy: A review of the literature* (Working Paper No. 1–42). Crawford School of Public Policy.

de Almeida, P. G. R., dos Santos, C. D., & Farias, J. S. (2021). Artificial intelligence regulation: A framework for governance. *Ethics and Information Technology,* 23, 505–525. https://doi.org/10.1007/s10676-021-09593-z

Desouza, K. C., Dawson, G. S., & Chenok, D. (2020). Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector. *Business Horizons, 63*(2), 205–213. https://doi.org/10.1016/j.bushor.2019.11.004

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy artificial intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion, 99*, 101896. https://doi.org/10.1016/j.inffus.2023.101896

Engstrom, D. F., Ho, D. E., Sharkey, C. M., & Cuéllar, M.-F. (2020, February 1). *Government by algorithm: Artificial intelligence in federal administrative agencies* (NYU School of Law, Public Law Research Paper No. 20-54). SSRN. https://doi.org/10.2139/ssrn.3551505

Han, T. A., Santos, F. C., Pereira, L. M., & Lenaerts, T. (2021, July 18–22). A regulation dilemma in artificial intelligence development. In *Proceedings of the ALIFE 2021: The 2021 Conference on Artificial Life* (p. 65). ASME. https://doi.org/10.1162/isal_a_00385

Hickok, M. (2024). Public procurement of artificial intelligence systems: New risks and future proofing. *AI & Society, 39*, 1213–1227. https://doi.org/10.1007/s00146-022-01572-2

Hoffreumon, C., Forman, C., & van Zeebroeck, N. (2024). Make or buy your artificial intelligence? Complementarities in technology sourcing. *Journal of Economics & Management Strategy*, *33*(1). https://doi.org/10.1111/jems.12586

Irion, K. (2013). Government cloud computing and national data sovereignty. *Policy & Internet, 5*(3), 303–319. https://doi.org/10.1002/poi3.10

Judge, B., Nitzberg, M., & Russell, S. (2024). When code isn't law: Rethinking regulation for artificial intelligence. *Policy and Society, 00*(00), 1–13. https://doi.org/10.1093/polsoc/puae020

Justo-Hanani, R. (2022). The politics of artificial intelligence regulation and governance reform in the European Union. *Policy Sciences, 55*, 137–159. https://doi.org/10.1007/s11077-022-09452-8

Kerikmäe, T., & Pärn-Lee, E. (2021). Legal dilemmas of Estonian artificial intelligence strategy: In between e-society and global race. *AI & Society, 36*, 561–572. https://doi.org/10.1007/s00146-020-01009-8

Mügge, D. (2024). EU AI sovereignty: For whom, to what end, and to whose benefit? *Journal of European Public Policy, 31*(8), 2200–2225. https://doi.org/10.1080/13501763.2024.2318475

Nugraha, Y., Kautsarina, & Sastrosubroto, A. S. (2015). Towards data sovereignty in cyberspace. In *2015 3rd International Conference on Information and Communication Technology (ICoICT)* (pp. 465–471). IEEE. https://doi.org/10.1109/ICoICT.2015.7231469

Polatin-Reuben, D., & Wright, J. (2014, July 7). An internet with BRICS characteristics: Data sovereignty and the Balkanisation of the internet. University of Oxford.

Roberts, H., Babuta, A., Morley, J., Thomas, C., Taddeo, M., & Floridi, L. (2023, May 1). Artificial intelligence regulation in the United Kingdom: A path to good governance and global leadership? *Internet Policy Review*. https://doi.org/10.2139/ssrn.4223964

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. In L. Floridi (Ed.), *Ethics, governance, and policies in artificial intelligence* (Vol. 144, pp. 69–94). Springer. https://doi.org/10.1007/978-3-030-81907-1_5

Shrier, D. L., Piotti, A., Pentland, A., & Faisal, A. A. (2024, November). *Considerations regarding sovereign AI and national AI policy* (V1.1 Draft). Imperial College London; Trusted AI Alliance.

Smuha, N. A. (2021). From a 'race to AI' to a 'race to AI regulation': Regulatory competition for artificial intelligence. *Law, Innovation and Technology, 13*(1), 57–84. https://doi.org/10.1080/17579961.2021.1898300

Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and Society, 40*(2), 137–157. https://doi.org/10.1080/14494035.2021.1928377

Timmers, P. (2019). Ethics of AI and cybersecurity when sovereignty is at stake. *Minds & Machines, 29*, 635–645. https://doi.org/10.1007/s11023-019-09508-4