Medical Research
Volume 7, Issue 3, 2025, Page 1-12
https://doi.org/10.6913/mrhk.070301

Medical Research

The Regulatory Trilemma of Medical AI: A New Framework for Navigating Innovation, Rights, and Sovereignty in Global Governance

Changkui LI
Hong Kong Integration Research Institute, lichangkui@gmail.com https://orcid.org/0000-0001-7446-0198

Abstract

The rapid rise of medical AI has outpaced global governance, with the US, EU, and China taking divergent regulatory paths. Existing theories such as regulatory competition and techno-nationalism cannot fully explain this fragmentation. This paper proposes the Regulatory Trilemma of Medical AI, which posits that no system can simultaneously maximize innovation velocity, rights protection, and national sovereignty. Using comparative case studies and framework analysis, we show that the US prioritizes innovation speed, the EU emphasizes rights and safety, and China asserts sovereignty through state-led data control. These choices generate systemic trade-offs, leading to a "Balkanization" of global medical AI markets, creating barriers to trade, complicating standard-setting by bodies like the WHO, and threatening health equity. The framework offers a more integrated lens to understand these dynamics and calls for a new global compact to establish baseline principles of safety, efficacy, and ethics in medical AI governance.

Keywords Medical AI governance; Regulatory trilemma; Innovation rights sovereignty trade-offs; US EU China comparison; Global health equity

1 Introduction: The Governance Gap in Global Health AI

The 21st century has witnessed the emergence of artificial intelligence (AI) as a paradigm-shifting force in medicine, with the potential to reshape diagnostics, accelerate therapeutic discovery, and fortify public health systems. From deep learning algorithms that interpret medical images with expert-level accuracy to generative models that design novel drug candidates, AI promises a future of healthcare that is more predictive, personalized, and efficient. Yet, this technological revolution has outpaced the development of a coherent global governance framework. The rapid proliferation of medical AI has created a critical governance gap, where divergent national and regional approaches to regulation are not merely technical in nature but reflect deep-seated geopolitical and ideological competition.

To understand this fractured landscape, scholars have often turned to established theories of international political economy and technology policy. The theory of regulatory competition, for instance, posits that jurisdictions compete to attract investment by offering different legal frameworks, potentially leading

to a "race to the bottom" on standards or, conversely, a "race to the top" as actors select the most efficient rules. This lens is particularly useful for analyzing the dynamic between the United States and the European Union. The US, with its agile, market-driven approach, appears to compete on speed and innovation, while the EU champions a "race to the top" on safety and ethical standards. A powerful manifestation of this is the "Brussels Effect," a term describing the EU's unilateral ability to export its stringent, rights-based regulations—such as the General Data Protection Regulation (GDPR) and the new AI Act—globally, as multinational corporations adopt these high standards across their operations to access the vast EU market.

In parallel, the concept of techno-nationalism has gained prominence to explain the strategic behavior of states that link technological capabilities and self-sufficiency directly to national security, economic prosperity, and social stability. This framework aptly describes China's approach, where a grand national strategy—the "AI+" Action Plan—aims to build a comprehensive, state-supported, and sovereign technological ecosystem, explicitly designed to reduce long-term reliance on foreign technology and assert control over data as a strategic national asset.

While these theoretical lenses are powerful, they often operate in separate analytical silos. Regulatory competition, with its focus on market-based incentives, struggles to fully account for the non-market, state-security logic of techno-nationalism. Conversely, techno-nationalism does not adequately capture the nuanced, standards-based competition occurring between market-oriented economies like the US and the EU. This theoretical fragmentation leaves a critical gap in our understanding of the integrated, global system of medical AI governance.

This paper seeks to fill this gap by proposing a new, integrated theoretical framework: The Regulatory Trilemma of Medical AI. Adapted from the "impossible trinity" concept in international economics, this framework posits that a single regulatory system cannot simultaneously and fully achieve three competing objectives: Innovation Velocity, Rights Protection, and National Sovereignty. Each of the world's three major techno-economic blocs has implicitly chosen to prioritize one vertex of this trilemma, necessarily accepting trade-offs on the other two. The United States has optimized its system for the rapid iteration and deployment of new technologies. The European Union has constructed a comprehensive legal architecture to serve as the global gold standard for protecting fundamental rights and ensuring safety. China has engineered a dual-pronged governance system to achieve its strategic goals of technological self-sufficiency and data sovereignty.

The central thesis of this paper is that the Regulatory Trilemma provides a more comprehensive and dynamic framework than existing theories for explaining the current fragmentation of global medical AI governance, predicting its future trajectory, and analyzing its systemic consequences for clinical practice, market structure, and global health equity. By placing the distinct logics of market-based competition and state-led sovereignty into a single, mutually constrained system, the trilemma reveals not just that these regulatory models are different, but why they are structurally divergent and why a simple convergence is unlikely. It demonstrates how the US and EU are engaged in one form of competition over global standards, while China's pursuit of a third, sovereign axis fundamentally alters the dynamics between the other two. This paper will first establish a rigorous methodology for this comparative analysis, then apply the trilemma framework to each case, substantiate the analysis with a novel semi-quantitative assessment, explore the tangible clinical and systemic consequences, and conclude by discussing the profound implications of this trilemma for the future of global health.

2 Methodology: A Comparative Framework Analysis

To provide a rigorous and transparent analysis of the complex global dynamics of medical AI governance, this study moves beyond a descriptive overview to employ a formal qualitative research design. The methodology is structured as a comparative case study, defined as the systematic comparison of two

or more cases to understand a larger phenomenon, detect causal patterns, and build or test theory. This approach is exceptionally well-suited for in-depth policy analysis where large-N statistical methods are not feasible, allowing for a rich, context-sensitive examination of the forces shaping regulatory divergence.

The cases selected for this study are the United States, the European Union, and China. This selection is purposive and justified on the grounds that these three entities represent the world's primary techno-economic blocs, are the leading sources of medical AI innovation and investment, and are the most active and influential actors in setting and exporting regulatory standards. They function as distinct "ideal types" of regulatory philosophy, and their interactions and divergences are the most consequential for shaping the global governance landscape.

The data collection strategy for this study involved a structured synthesis of evidence from three primary categories of sources. First, policy and regulatory documents form the core of the analysis. These include foundational strategic documents and specific legal statutes from each jurisdiction. For China, this includes the "AI+" Action Plan, the Personal Information Protection Law (PIPL), the Cybersecurity Law (CSL), the Data Security Law (DSL), and key guidelines from the National Medical Products Administration (NMPA) and the Cyberspace Administration of China (CAC). For the United States, sources include the Food and Drug Administration's (FDA) guidance on Software as a Medical Device (SaMD), the Predetermined Change Control Plan (PCCP) framework, and the Health Insurance Portability and Accountability Act (HIPAA). For the European Union, the analysis centers on the official texts of the EU AI Act and the General Data Protection Regulation (GDPR). Second, academic and clinical literature was gathered through a systematic search of peer-reviewed literature to find evidence on the clinical translation, efficacy, and implementation challenges of medical AI. Databases including PubMed and Web of Science were searched using a combination of keywords such as "medical AI," "AI regulation," "digital health," "FDA," "EU AI Act," "GDPR," "China NMPA," "physician de-skilling," "AI drug discovery," "algorithmic bias," and "regulatory sandbox." Inclusion criteria prioritized peer-reviewed articles, systematic reviews, and meta-analyses published between 2015 and the present that provided empirical evidence or rigorous analysis of the topic. Opinion pieces and non-peer-reviewed reports were excluded. Third, industry and governmental reports were consulted to gather data for the semi-quantitative analysis. These included official reports and public databases from regulatory agencies (e.g., the FDA's list of authorized AI/ML devices) and reputable industry analyses.

The analytical approach employed in this study is Framework Analysis, a systematic method for managing and interpreting qualitative data that is particularly well-suited for policy research involving large datasets. A defining feature of this method is the use of a pre-defined analytical framework to structure the analysis. For this study, the "Regulatory Trilemma" itself serves as the *a priori* analytical framework. The three vertices—Innovation Velocity, Rights Protection, and National Sovereignty—constitute the primary thematic categories. The analytical process involved systematically coding the collected data from all sources for each case (US, EU, China) according to these three themes. The coded data was then charted into a matrix, allowing for a structured and transparent comparison both within and across the cases. This systematic process ensures that the trilemma is applied not as a loose metaphor but as a rigorous analytical tool, providing a robust and defensible foundation for the paper's conclusions and transforming the analysis from a narrative description into a structured, evidence-based academic argument.

3 The Regulatory Trilemma in Practice: A Comparative Analysis of Global Powers

Applying the framework analysis reveals how the regulatory architectures of the United States, the European Union, and China each embody a distinct and deliberate prioritization of one vertex of the Regulatory Trilemma. This section deconstructs each model to demonstrate the philosophies, mechanisms, and trade-offs that define their positions within this constrained global system.

3.1 The United States: Prioritizing the Velocity of Innovation

The US regulatory model for medical AI, spearheaded by the Food and Drug Administration (FDA), is fundamentally market-driven, pragmatic, and optimized for innovation-centric agility. The overarching philosophy is to facilitate the rapid development and deployment of safe and effective technologies, recognizing that the iterative nature of AI requires a regulatory approach that differs from that for static hardware devices. This prioritization is most clearly exemplified by the FDA's risk-based framework for Software as a Medical Device (SaMD) and its pioneering Predetermined Change Control Plan (PCCP).

The SaMD framework applies a pragmatic, risk-based classification (Class I, II, III) that focuses on the device's intended use and its potential impact on patient care, rather than adopting a broad, categorical definition of risk. This allows for a more tailored and often faster pathway to market for lower-risk devices. The most significant innovation embodying the US focus on velocity is the PCCP. This unique regulatory mechanism allows a manufacturer to prospectively define the scope of anticipated modifications to an AI model, the methodology for validating those changes, and the protocol for their implementation within a single pre-market submission. Once this plan is authorized, the developer can deploy updates that fall within these pre-specified boundaries without needing to submit a new regulatory application for each change. This agile approach is explicitly designed to accommodate the dynamic, learning nature of modern AI systems, enabling rapid, continuous improvement and maintaining a technological edge in a fast-moving field. In doing so, the US system makes a clear trade-off: it accepts a degree of regulatory flexibility and reliance on manufacturer-led validation in exchange for accelerating the pace at which innovative tools can reach the clinic.

3.2 The European Union: Championing Fundamental Rights and Safety

In stark contrast to the US model, the European Union's approach is rights-centric, comprehensive, and guided by the precautionary principle. The primary objective of its regulatory architecture is not to maximize the speed of innovation but to ensure the robust protection of fundamental rights, safety, and democratic values, thereby building broad public trust in AI technologies. This philosophy is enshrined in two landmark pieces of legislation: the EU AI Act and the General Data Protection Regulation (GDPR).

The EU AI Act establishes a risk-based pyramid for all AI systems, but crucially, it automatically classifies most medical devices requiring a conformity assessment under the EU's Medical Devices Regulation (MDR) as "high-risk". This classification triggers a cascade of stringent and comprehensive *ex-ante* (pre-market) compliance obligations that are far more extensive than in other jurisdictions. Providers of high-risk medical AI must implement robust risk management systems, adhere to strict data governance practices to ensure training data is high-quality and free of bias, create extensive technical documentation detailing the system's design and performance, and build in capabilities for effective human oversight. This creates a high barrier to market entry, deliberately slowing the innovation cycle to prioritize safety and accountability.

This framework is further reinforced by the GDPR, which imposes some of the world's strictest rules on data processing. Health data is classified as a "special category of personal data," requiring an explicit legal basis and often specific, unambiguous consent for any processing activity. This dual compliance challenge—navigating both the technical requirements of the AI Act and the complex data protection mandates of the GDPR—firmly places the EU at the "Rights Protection" vertex of the trilemma. The EU model is designed to function as a global standard-setter through the "Brussels Effect," compelling global companies to adopt its high standards, even if it means a more deliberate and cautious pace of technological deployment.

3.3 China: Asserting Technological and Data Sovereignty

China's model for medical AI governance is distinct from both the US and the EU, prioritizing a third objective: the achievement of national strategic goals and the assertion of state sovereignty over technology and data. This approach is state-led, security-focused, and designed to cultivate a domestic, self-sufficient AI ecosystem while maintaining firm state control over what it deems to be critical national assets.

This sovereignty-focused approach is operationalized through a unique dual-pillar governance structure. The first pillar, the National Medical Products Administration (NMPA), functions similarly to its Western counterparts, regulating the safety and efficacy of AI software as a medical device through a risk-based classification system. The second, and more distinctive, pillar is the Cyberspace Administration of China (CAC). The CAC's mandate extends beyond patient safety to the broader governance of data security, algorithm content, and information control. Under a suite of laws including the Personal Information Protection Law (PIPL), Cybersecurity Law (CSL), and Data Security Law (DSL), the CAC imposes unique requirements that serve state interests. These include a mandatory algorithm filing and registration system, which provides the government with direct oversight of algorithms used in services with "public opinion attributes or social mobilization capabilities".

Furthermore, these laws establish a formidable "data moat" around the country. The PIPL imposes stringent requirements on data localization and establishes high barriers for cross-border data transfer, effectively treating data not as a corporate asset or personal property, but as a matter of national sovereignty. This dual strategy of promoting internal data circulation while restricting external data flow serves two strategic purposes: it provides domestic AI developers with a significant competitive advantage by granting them access to massive, centralized datasets, and it maintains state control over a critical resource. This entire regulatory architecture is the implementation arm of the broader "AI+" national strategy, a top-down, synergistic policy framework designed to build a complete, self-reinforcing, and globally competitive national AI industry, firmly placing China at the "National Sovereignty" vertex of the trilemma.

4 Operationalizing the Trilemma: A Semi-Quantitative Assessment

To move the analysis of the Regulatory Trilemma from a qualitative description to a more concrete, evidence-based validation, this section operationalizes each vertex of the framework using measurable proxy indicators. By systematically comparing the United States, the European Union, and China across these indicators, it is possible to create a semi-quantitative index that reveals the tangible trade-offs each regulatory system has made. This approach provides a "hard-core" demonstration of how the abstract priorities of innovation, rights, and sovereignty manifest in specific, observable policy outcomes and regulatory burdens. The following table synthesizes data on regulatory approvals, compliance requirements, and state control mechanisms to map each bloc' s position within the trilemma.

The data synthesized in this table provides compelling evidence for the Regulatory Trilemma framework. It demonstrates that each bloc' s policy choices result in a distinct profile of strengths and weaknesses across the three dimensions. The United States scores highly on all indicators of "Innovation Velocity," facilitated by its high volume of approvals, rapid review pathways, and the unique PCCP mechanism. However, this comes at the cost of lower scores on "Rights Protection" (less stringent pre-market burdens and data protection compared to the EU) and "National Sovereignty" (a permissive data transfer regime and minimal state control over algorithms).

Conversely, the European Union excels in "Rights Protection," with its very high pre-market compliance burdens, stringent GDPR framework, and demonstrable enforcement record. This leadership in rights and safety, however, creates significant friction for "Innovation Velocity," as evidenced by slower review times and highly restricted mechanisms for agile updates.

Table 1: The Medical AI Regulatory Trilemma Index: A Comparative Assessment of the US, EU, and China

Trilemma Dimension	Proxy Indicator	United States	European Union	China
Innovation Velocity	Number of AI/ML Devices Authorized	High: Over 700 devices authorized by the FDA as of late 2023, with a public list of 555 devices by mid-2025.	Moderate: Approval is decentralized through Notified Bodies under the Medical Devices Regulation (MDR); no centralized, comparable public list exists. The process is widely considered more lengthy than the FDA's primary pathways.	Growing: 92 AI tools approved as high-risk Class III medical devices by the NMPA as of June 2024, indicating a maturing but smaller market of approved high-risk devices.
	Avg. Regulatory Review Time	Fast (for most devices): The 510(k) pathway, used for most moderate-risk devices, has a statutory review target of 90 days. The high-risk Premarket Approval (PMA) pathway is longer, officially 180 days but often extending further.	Slow: The MDR conformity assessment process is notoriously lengthy, frequently exceeding one year, a situation exacerbated by a persistent shortage of accredited Notified Bodies.	Variable: NMPA review times for Class III devices are not consistently published but are known to be rigorous and time-consuming, comparable to or longer than the FDA's PMA process.
	Agile Update Mechanism	Explicitly Enabled: The Predetermined Change Control Plan (PCCP) is a unique framework that allows for pre-approved, iterative model updates without requiring new submissions for each change, designed specifically to accelerate AI innovation.	Highly Restricted: Any significant change to a high-risk AI system, including performance-enhancing updates, requires a new conformity assessment and re-certification by a Notified Body under the AI Act, creating a significant barrier to rapid iteration.	Restricted: Modifications to approved medical devices require formal approval from the NMPA. No agile framework comparable to the US PCCP exists, prioritizing stability over iterative speed.
Rights Protection	Pre-Market Compliance Burden	Moderate: The FDA focuses on demonstrating safety and effectiveness for a specific intended use. Documentation is substantial but generally less extensive and prescriptive than the EU's requirements for most devices.	Very High: The EU AI Act mandates that all high-risk systems undergo a third-party conformity assessment and produce extensive technical documentation covering risk management, data governance, human oversight design, and cybersecurity, representing the highest pre-market burden.	High: The NMPA requires extensive clinical and technical documentation for Class III devices. The separate CAC algorithm filing process adds another layer of compliance focused on data and content, creating a dual burden.
	Data Protection Stringency	Sector-Specific: The Health Insurance Portability and Accountability Act (HIPAA) governs Protected Health Information (PHI) but is less comprehensive in scope and individual rights than the GDPR.	Very High: The GDPR establishes a global benchmark, providing broad data subject rights and imposing strict requirements for consent and a clear legal basis for processing "special category" health data.	Very High: The Personal Information Protection Law (PIPL) mirrors the GDPR's stringency, requiring "separate consent" for processing sensitive health information and for any cross-border data transfers, establishing a strong rights-based framework.
	Enforcement (Financial Penalties)	Variable: Penalties for HIPAA violations exist and can be substantial, but enforcement actions are less frequent and publicly prominent than GDPR fines.	High & Demonstrable: The GDPR has a proven track record of enforcement, including significant fines levied within the healthcare sector for insufficient data security (e.g., €440,000 for OLVG hospital in the Netherlands) and unlawful data processing. Fines can reach up to 4% of a company's global annual turnover.	Emerging: Enforcement of the PIPL is active, but a public record of large, deterrent fines comparable to those under the GDPR is still developing as the legal framework matures.
National Sovereignty	Cross-Border Data Transfer	Permissive: The US model generally supports the free flow of data, with restrictions primarily targeted and based on specific national security concerns rather than as a default industrial policy.	Conditional: Data transfers outside the EU are highly restricted unless the recipient country has an official "adequacy decision" or other legal safeguards, such as Standard Contractual Clauses (SCCs), are in place to ensure equivalent data protection.	Highly Restrictive: The PIPL establishes the world's strictest default data control regime. It triggers a mandatory, state-run CAC Security Assessment for transfers of personal information exceeding specific thresholds (e.g., >1 million individuals, or >10,000 individuals' sensitive PI).
	State Control over Algorithms	Minimal: There is no general government registry for algorithms. Regulation is product-specific and focused on safety and efficacy, conducted by the FDA.	Indirect: The AI Act requires transparency through extensive technical documentation for high-risk systems, which must be available to authorities upon request, but it does not create a central state registry for all algorithms.	Explicit & Centralized: China operates a mandatory algorithm filing and registration system with the CAC for services deemed to have "public opinion attributes or social mobilization capabilities," providing the state with direct and proactive oversight of algorithmic systems.
	State-Driven Industrial Policy	Indirect: The government supports innovation through research funding via agencies like the National Institutes of Health (NIH), but the ecosystem is fundamentally market-driven and decentralized.	Supportive: The EU funds research and innovation through programs like Horizon Europe, but its role is to foster a competitive market within a strong regulatory framework, not to direct industrial outcomes.	Direct & Central: The "AI+" Action Plan represents a formal, top-down national strategy that uses state investment, procurement, and policy coordination to create a self-sufficient, state-supported industrial ecosystem and cultivate national champions.

Finally, China achieves the highest scores on "National Sovereignty" through its highly restrictive cross-border data transfer regime, explicit state control over algorithms via the CAC registry, and a direct, state-driven industrial policy. This assertion of sovereignty is achieved through mechanisms that constrain the free flow of information central to Western rights-based models and limit the agile innovation prioritized by the US. The table thus makes the abstract concept of the trilemma concrete, illustrating the inescapable trade-offs that define the fragmented global landscape of medical AI governance.

5 Systemic Consequences and Clinical Realities

The high-level regulatory divergences captured by the trilemma framework are not merely theoretical; they generate profound and tangible consequences for the healthcare system, clinical practice, and the trajectory of biomedical innovation. This section examines three critical areas where these systemic effects are most apparent: the emergent risk of physician de-skilling as a threat to health system resilience, the accelerating paradigm of AI-driven therapeutic development, and the persistent chasm between the laboratory performance of diagnostic AI and its validated clinical utility.

5.1 Systemic Vulnerability: The Generalizable Risk of De-skilling

Beyond questions of algorithmic accuracy, the integration of AI into clinical workflows is surfacing a more subtle but potentially more pernicious challenge: the risk of physician de-skilling. This phenomenon was brought into sharp focus by a multi-center observational study published in *The Lancet Gastroenterology and Hepatology*, which provided the first compelling, real-world clinical evidence of its occurrence. The study found that after a period of routine use of an AI-powered polyp detection system, the unassisted Adenoma Detection Rate (ADR) of experienced endoscopists—a key quality metric—declined by a relative 20% from

their pre-AI baseline. This suggests that continuous reliance on AI assistance can erode the vigilance and core pattern-recognition skills that clinicians cultivate through years of repetitive practice.

This risk is not unique to medicine but is a generalizable consequence of human-automation interaction in high-stakes environments. A parallel and well-documented case exists in commercial aviation, a field with decades of experience managing human-machine collaboration. Studies on airline pilots have found that while extensive use of autoflight systems does not significantly degrade pilots' manual aircraft control skills, it can lead to the atrophy of crucial cognitive skills required for manual flight, such as tracking the aircraft's position, deciding on navigational steps, and recognizing instrument failures. This cognitive de-skilling creates a dangerous dependency on automation.

Framed in this broader context, physician de-skilling transcends an issue of individual professional development and becomes a latent threat to the resilience of the entire healthcare system. A resilient system is defined by its capacity to absorb shocks and disruptions while continuing to deliver essential services. In a future where clinicians become increasingly dependent on AI, the system becomes brittle—highly efficient under normal operating conditions but dangerously vulnerable to catastrophic failure during systemic shocks like widespread software failures, network outages, or malicious cyber-attacks that render AI tools unavailable. In such a scenario, the performance of a de-skilled clinical workforce could plausibly be *worse* than it was before AI was ever introduced, transforming de-skilling into a systemic patient safety and public health crisis.

5.2 The Acceleration of Therapeutics: A Demonstrable Trend

While diagnostics present a complex picture of benefits and risks, the application of AI in therapeutics is demonstrating more clear-cut and verifiable progress, fundamentally disrupting the economics and velocity of pharmaceutical innovation. The traditional drug development pipeline is notoriously long and expensive, often requiring over a decade and more than $2 billion per approved drug. AI is beginning to radically compress this timeline.

A powerful proof-of-concept is the candidate drug INS018_055, developed by Insilico Medicine for idiopathic pulmonary fibrosis (IPF). This compound is the first to have been discovered using a generative AI platform for target identification and subsequently designed *de novo* by a generative AI model to enter human clinical trials. The verifiable metrics are striking: the entire timeline from initial target identification to the commencement of Phase I clinical trials was less than 30 months, a dramatic acceleration compared to traditional methods. The drug has since successfully completed Phase I trials and is now advancing through Phase IIa clinical trials in both China and the US.

This is not an isolated success story but rather indicative of an emerging trend. Several other companies are successfully leveraging AI to build clinical-stage pipelines. Exscientia, a pioneer in the field, has advanced multiple AI-designed drug candidates into human trials, including DSP-1181 for obsessive-compulsive disorder, EXS21546 as an immuno-oncology treatment, and DSP-0038 for Alzheimer's disease psychosis. The company has also forged major collaborations with pharmaceutical giants like Sanofi to develop up to 15 novel small molecule candidates. Similarly, Recursion Pharmaceuticals has utilized its AI-driven platform to build an advanced clinical pipeline focused on oncology and rare diseases, with candidates such as REC-617 for advanced solid tumors and REC-4881 for familial adenomatous polyposis now in Phase 1/2 and Phase 2 trials, respectively. The concurrent success of multiple firms in translating computationally designed molecules into viable clinical candidates provides strong evidence that AI is catalyzing a genuine paradigm shift in therapeutic development, moving from a process of serendipitous discovery and slow screening to one of rational, accelerated design.

5.3 The Evidence-Practice Chasm in Diagnostics

Despite the explosion of research and development in diagnostic AI, a persistent and critical gap remains between the performance of algorithms in controlled, retrospective laboratory settings and the demonstrated evidence of their value in prospective, real-world clinical practice. A landmark systematic review and meta-analysis published in *The Lancet Digital Health* found that while the pooled diagnostic accuracy of deep learning algorithms was statistically comparable to that of human clinicians, the vast majority of studies included were retrospective in nature, used highly curated datasets, and were at high risk of bias. The authors issued a strong cautionary note that these impressive results may not translate to the complexities of real-world clinical data and workflows.

This "lab-to-clinic" gap is a consistent finding across the literature. Subsequent systematic reviews of randomized controlled trials (RCTs)—the gold standard for clinical evidence—have repeatedly found that few high-quality RCTs for medical AI interventions have been conducted. A 2022 review in *JAMA Network Open* identified only 41 such RCTs and found that none fully adhered to the CONSORT-AI reporting guidelines, with common flaws including a failure to analyze performance errors or assess the impact of poor-quality input data. This scarcity of robust, prospective evidence raises significant concerns about the true clinical utility and generalizability of many diagnostic AI tools currently on the market. To provide a sober assessment of the current state of high-quality evidence, the table below provides an updated synthesis of key prospective or large-scale real-world studies that measure the impact of diagnostic AI.

As this updated table illustrates, while there is emerging high-quality evidence of benefit in specific, well-defined tasks—such as improving detection rates or reducing diagnostic turnaround times—the overall evidence base remains nascent. The persistent scarcity of large-scale, multi-center RCTs that measure patient-meaningful outcomes remains the central challenge for the field. This evidence-practice chasm highlights the critical importance of robust regulatory oversight and a commitment to evidence-based implementation, a challenge that each of the three regulatory models is attempting to address in its own way.

Table 2: Structured Synthesis of High-Quality Clinical Evidence for Diagnostic AI

Domain	Study Design	Sample Size	Primary Endpoint	Key Finding / Effect Size
Radiology - Lung Cancer Screening	Prospective, single-center, randomized controlled trial	10,476 participants	Detection rate of actionable lung nodules	The AI-assisted group had a significantly higher detection rate of 0.59% compared to 0.25% in the non-AI group (OR=2.4), without an increase in the false-positive rate.
Gastroenterology - Colonoscopy	Multi-center, prospective, randomized controlled trial	1,058 patients	Adenoma Detection Rate (ADR)	The use of a real-time AI system significantly increased the ADR from 29.1% in the control group to 36.6% in the AI group, corresponding to a relative increase of 25%.
Oncology - Breast Cancer Screening	Population-based, retrospective, paired-reader study	80,000 women	Cancer detection rate and radiologist workload	AI-supported screening was non-inferior to standard double reading by two radiologists and was associated with a 44.3% reduction in the screen-reading workload.
Neurology - Intracranial Hemorrhage	Real-world deployment study	N/A	Diagnostic turnaround time	A commercial solution (Aidoc) helped hospitals reduce the turnaround time for diagnosing critical conditions like intracranial hemorrhage by over 50%.
Gastroenterology - De-skilling Risk	Multi-center, observational study	1,443 non-AI-assisted procedures	Adenoma Detection Rate (ADR)	After routine introduction of AI, the ADR of experienced endoscopists in *unassisted* procedures decreased from a baseline of 28.4% to 22.4% (a 20% relative decrease).

6 Discussion: Theoretical Contributions and Policy Implications

The analysis of the divergent regulatory pathways and their systemic consequences provides a strong empirical foundation for the Regulatory Trilemma framework. This section elevates the discussion by explicitly articulating the paper's theoretical contribution, exploring the profound policy implications for global governance, acknowledging the study's limitations, and proposing a concrete agenda for future research.

6.1 Theoretical Contribution: Synthesizing Governance Theories

The primary theoretical contribution of this paper is the proposal and validation of the Regulatory Trilemma as a new, integrated framework for understanding the governance of transformative technologies like medical AI. Its novelty lies not in inventing new concepts from whole cloth, but in synthesizing and extending existing, and sometimes competing, theories of international political economy into a unified model that better captures the current global reality.

The framework explicitly integrates the economic logic of regulatory competition with the geopolitical logic of techno-nationalism. It recognizes that the US and the EU are engaged in a form of standards-based competition, where the EU's rights-centric approach and the resulting "Brussels Effect" represent an attempt to pull the global market toward a "race to the top" on safety and ethics. Simultaneously, it acknowledges that China's actions are better explained by techno-nationalism, where the primary goal is not to compete within the existing global market of rules but to build a parallel, sovereign ecosystem driven by state-security and industrial policy objectives.

The Trilemma's explanatory power comes from modeling these different logics as interacting within a single, constrained system. It demonstrates that the choices are not binary (e.g., open vs. closed, high vs. low standards) but triangular. The existence of the "Sovereignty" axis pursued by China fundamentally alters the competitive dynamic between the "Innovation" and "Rights" axes championed by the US and EU, respectively. It creates a powerful gravitational pull toward a multi-polar regulatory world, a dynamic that single-theory explanations fail to fully capture. In this way, the Trilemma offers a more nuanced and predictive model for analyzing the complex interplay of market forces, normative values, and state power in the governance of 21st-century technology.

6.2 Policy Implications: The "Balkanization" of Global Health AI

The Regulatory Trilemma is not a static equilibrium but an engine of fragmentation, actively fostering a "Balkanization" of the global medical AI market. The divergent regulatory pathways are creating three distinct, and often mutually incompatible, techno-regulatory spheres of influence. This fragmentation has profound policy implications for international organizations tasked with overseeing global trade and health.

For the World Trade Organization (WTO), these divergent national standards function as significant non-tariff barriers to trade in digital services. A medical AI product developed and validated for the agile US market may be unable to meet the EU's extensive pre-market data governance and documentation requirements without a costly redesign. Similarly, a product trained on vast datasets within China's "data moat" faces nearly insurmountable legal and practical hurdles to being deployed in the EU or US due to stringent cross-border data transfer restrictions. This undermines the WTO's core mission of facilitating a predictable, rules-based global trading system.

For the World Health Organization (WHO), this Balkanization poses a direct threat to global public health. It severely complicates the WHO's ability to establish global standards, benchmarks, and best practices for the safe and effective use of AI in health. The lack of harmonized data formats and regulatory requirements impedes international medical research collaborations, undermines the effectiveness of global

disease surveillance, and creates barriers for sharing clinical insights across borders. This fragmentation could ultimately deepen global health inequities. Low- and Middle-Income Countries (LMICs), lacking the resources to develop their own indigenous AI ecosystems or regulatory frameworks, will be forced to align with one of the three dominant blocs. This choice, potentially driven by geopolitical pressures or trade dependencies, could lead to a new form of technological dependency, locking them into a specific sphere's standards, data protocols, and technologies, and creating a stratified global system of access to advanced health technology.

One potential mitigating strategy against complete fragmentation is the use of regulatory sandboxes. These are controlled, real-world environments where regulators and innovators can collaboratively test novel technologies and co-develop appropriate oversight mechanisms. Initiatives like the UK's "AI Airlock" and the FDA's precisionFDA platform provide a "soft law" approach that could foster pockets of international convergence on best practices for validation and oversight, even in the absence of formal treaty-based harmonization.

6.3 Limitations of the Study

This study, while comprehensive, is subject to several limitations that warrant acknowledgment and provide avenues for future research. First, the regulatory landscape for AI is extraordinarily dynamic. Policies and laws in all three jurisdictions are in a constant state of flux, and this analysis represents a snapshot at a particular moment in time. The framework's predictions will need to be continually tested against new regulatory developments. Second, this analysis relies primarily on publicly available policy documents, statutes, and academic literature. It cannot fully capture the nuances of internal enforcement practices, institutional cultures, or the informal political negotiations that also shape regulatory outcomes. Future research employing qualitative methods such as elite interviews with policymakers and industry leaders could provide a richer, more detailed understanding of these dynamics. Finally, while the Regulatory Trilemma provides a powerful explanatory framework, it necessarily simplifies a highly complex reality. The influence of other geopolitical factors, the role of other nations, and the agency of multinational corporations in shaping and navigating these regulatory environments are all areas that merit deeper investigation.

6.4 A Forward-Looking Research Agenda

To move the field toward more rigorous, evidence-based inquiry, this paper concludes by proposing a series of specific, testable, and falsifiable hypotheses that emerge directly from the challenges identified in this analysis. This serves as a concrete and forward-looking research agenda to guide future empirical work.

The De-skilling Mitigation Hypothesis: Healthcare organizations that implement a structured protocol of alternating AI-assisted and unassisted clinical work periods will observe no statistically significant decline in their clinicians' unassisted diagnostic performance (as measured by metrics like ADR) from baseline over a 24-month period.

The Regulatory Sandbox Efficacy Hypothesis: The establishment of national-level regulatory sandboxes for AI medical devices will lead to a reduction of more than 15% in the median time from submission to regulatory approval for novel high-risk diagnostic software within 36 months of the sandboxes' launch.

The Bias Audit Effectiveness Hypothesis: Diagnostic AI models that undergo prospective validation and calibration using a fairness-aware bias auditing framework will demonstrate at least a 50% reduction in the performance gap (e.g., in sensitivity or specificity) between majority and minority demographic subgroups in a real-world clinical environment, compared to models deployed without such an audit.

The Real-World Performance Decay Hypothesis: Due to factors including workflow friction, data drift, and concept drift, the real-world effectiveness of AI diagnostic tools—measured by their impact on patient-relevant outcomes such as time-to-treatment or length of hospital stay—will be at least 25% lower than the efficacy reported in their pivotal pre-market randomized controlled trials.

These hypotheses provide clear pathways for future research aimed at transforming the critical questions raised in this paper into a robust evidence base that can guide the responsible and effective integration of artificial intelligence into the future of global health.

7 Conclusion: Toward a New Global Compact for Health AI

This analysis has introduced and validated the Regulatory Trilemma as an essential new framework for understanding the divergent and competitive global governance of medical AI. It demonstrates that the distinct regulatory paths being forged by the United States, the European Union, and China are not arbitrary but are the logical outcomes of prioritizing different, and mutually constraining, objectives: innovation velocity, rights protection, and national sovereignty. A powerful policy catalyst, such as China's "AI+" Action Plan, can provide immense momentum for technological advancement, but its ultimate success—and that of its global counterparts—is contingent on navigating the complex trade-offs inherent in this trilemma.

The promise of diagnostic AI is tempered by the pressing need for real-world clinical validation and the insidious risk of physician de-skilling. The revolutionary potential of AI in drug discovery must be translated from computational success into proven clinical efficacy. The vision of personalized medicine hinges on our ability to overcome the obstacles of data bias, algorithmic opacity, and workflow integration. And the power of AI in public health must be wielded with a profound respect for privacy and individual rights.

The fragmentation driven by the trilemma makes addressing these challenges on a global scale significantly more difficult. It risks a future where the benefits of medical AI are unevenly distributed, where international collaboration is stifled by incompatible standards, and where global health equity is undermined. Ultimately, the future of medicine lies not in a binary opposition between human and machine, nor in the substitution of one for the other. It resides in the creation of a new paradigm of human-AI collaboration—a synergistic partnership where the computational power of AI augments the empathy, clinical intuition, and holistic judgment of the human physician.

Achieving this collaborative future in an increasingly Balkanized world is the central challenge ahead. It requires a new form of global dialogue, perhaps facilitated by the WHO, that moves beyond advocating for a single, monolithic regulatory model. Instead, it must start by acknowledging the reality of the trilemma and the legitimate, albeit conflicting, priorities it represents. The goal should be to forge a new global compact for health AI—one that seeks to establish a minimum baseline of shared principles for safety, efficacy, transparency, and ethics, creating a common ground upon which different regulatory systems can build. It is only through such a concerted effort to manage divergence and foster interoperability that the global community can hope to harness the full potential of artificial intelligence to build a more precise, efficient, and equitable standard of healthcare for all.

To Cite This Article Changkui LI. (2025). The Regulatory Trilemma of Medical AI: A New Framework for Navigating Innovation, Rights, and Sovereignty in Global Governance. *Medical Research*, 7(3), 1-12. https://doi.org/10.6913/mrhk.070301

References

[1] Bradford, A. (2020). *The Brussels effect: How the European Union rules the world*. Oxford University Press.

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

[3] Food and Drug Administration. (2021). *Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan*. U.S. Department of Health and Human Services. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

[4] Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. In *Artificial intelligence in healthcare* (pp. 295-336). Academic Press. https://doi.org/10.1016/B978-0-12-818438-7.00012-5

[5] Gless, S., Silverman, E., & Weigend, T. (2023). A transatlantic perspective on the EU's proposed AI Act. *European Journal of Criminology*, *20*(4), 1318 1338. https://doi.org/10.1177/14773708211035540

[6] Hao, Y., & Cui, W. (2024). Regulating generative AI in China: A focus on algorithm registration and security assessment. *Computer Law & Security Review*, *52*, 105921. https://doi.org/10.1016/j.clsr.2024.105921

[7] Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, *17*(1), 195. https://doi.org/10.1186/s12916-019-1426-2

[8] Liu, Y., Chen, P. C., Krause, J., & Peng, L. (2019). How to read articles that use machine learning: Users' guides to the medical literature. *JAMA*, *322*(18), 1806 1816. https://doi.org/10.1001/jama.2019.16489

[9] Obstfeld, M., Shambaugh, J. C., & Taylor, A. M. (2005). The trilemma in history: Tradeoffs among exchange rates, monetary policies, and capital mobility. *The Review of Economics and Statistics*, *87*(3), 423-438. https://doi.org/10.1162/0034653054638300

[10] 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. *AI & Society*, *36*(1), 59-77. https://doi.org/10.1007/s00146-020-00992-2

[11] Su, J. (2023). Techno-nationalism, the US-China tech cold war, and the future of globalization. In *The Palgrave handbook of globalization with Chinese characteristics* (pp. 427-448). Palgrave Macmillan, Singapore. https://doi.org/10.1007/978-981-19-8622-6_22

[12] Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7

[13] Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLOS Medicine*, *15*(11), e1002689. https://doi.org/10.1371/journal.pmed.1002689

[14] World Health Organization. (2021). *Ethics and governance of artificial intelligence for health: WHO guidance*. World Health Organization. https://www.who.int/publications/i/item/9789240029200

[15] Wu, E., Wu, K., Daneshjou, R., Ouyang, D., Ho, D. E., & Zou, J. (2021). How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nature Medicine*, *27*(4), 582-584. https://doi.org/10.1038/s41591-021-01312-x