Artificial Intelligence, Data Sovereignty, and Interoperability Challenges in Clinical Trials in Low-Resource Settings

Sophie McDonagh

Norman Patterson School of International Affairs, Carleton University

INAF5706: Global Health Policy

Professor Percival

November 30, 2025

Introduction

Artificial intelligence (AI) is increasingly woven into all facets of global health, frequently touted as a quick-fix solution to funding, knowledge, and technical deficiencies. Within the global health research and development sector, clinical health trials are key to testing medical, surgical or behavioral interventions. In low-resource settings in particular, where the burden of disease is high, clinical trials are essential for achieving equitable and effective health solutions, generating evidence for context-appropriate settings.

Despite the rapid expansion of AI across health sectors, a notable gap persists in understanding how AI models can be responsibly integrated into clinical research in low-resource environments. In particular, the intersection of AI with data sovereignty requirements poses significant challenges for protecting local populations while enabling meaningful scientific collaboration. This paper asks “In clinical trials in low-resource settings, how do data sovereignty requirements and cross-border restrictions shape the interoperability and communication protocols needed for AI analysis?”.

To explore this question, a review with qualitative synthesis was performed using peer-reviewed studies retrieved from major databases. This paper argues that Artificial Intelligence cannot be effectively or ethically integrated into clinical trials in low-resource settings because national sovereignty requirements directly conflict with the interoperability and communication standards required for AI systems to function.

Methods

A structured search strategy was developed to identify peer-reviewed and grey literature relevant to AI protocols, biobanking governance, data sovereignty, and global health applications of artificial intelligence. Searches were conducted across major academic databases including the MacOdrum Library, PubMed, Google Scholar, the National Library of Medicine, Frontiers, MDPI, and SpringerLink. To capture the policy and governance dimensions of AI, additional targeted searches were performed using institutional and organizational sources such as the World Health Organization, OECD,

IDRC, FDA regulatory statements, and the Cloud Security Alliance. Key search terms included combinations of: *AI protocol gaps in health care; AI governance challenges; FIPA; HL7-FHIR; global health AI; biobanking governance; data sovereignty; AI and public health; AI colonialism; cross-border data flows; digital health interoperability; and biobank regulation*. Boolean operators (AND/OR) were used to refine results.

Search results were screened in three stages: (1) review of the first 40 paper titles, (2) abstract screening of titles deemed relevant (10-20), and (3) full-text review (3-8). Inclusion criteria required that sources: (1) addressed AI systems, protocols, data governance, or health applications; (2) focused on global health, public health systems, biobanking, interoperability, or cross-border data issues; and (3) were published between 2015–2025 in English. Grey literature was included when it contributed essential regulatory, policy, or governance insights. Exclusion criteria eliminated studies focused solely on narrow clinical AI applications without broader relevance to governance, ethics, or system-level challenges. This search strategy ensured a comprehensive synthesis of both technical and governance-oriented literature to assess gaps in AI protocols and data governance in global health.

Artificial Intelligence

This paper will focus on machine learning AI platforms, using multi-agent systems trained using federated learning. Since the 1950s, artificial intelligence (AI) has referred to technologies that enable computers and machines to simulate human learning, comprehension, problem-solving, decision-making, creativity, and autonomy (Stryker & Kavlakoglu, 2025). In the 2020s, public and policy debates have shifted toward generative AI, which can produce text, images, and video. AI systems are built primarily through two approaches: machine learning, developed in the 1980s and reliant on large datasets to generate predictions and inferences, and deep learning, emerging in the 2010s as a subset of

machine learning that uses multilayered neural networks to mimic aspects of the human brain (Stryker & Kavlakoglu, 2025).

Within AI, there are Multi-Agent Systems (MAS), a core area of modern AI research, consisting of “multiple decision-making agents which interact in a shared environment to achieve common or conflicting goals” (The Allan Turning Institute, 2025). Simply put, instead of one large AI device trying to do everything, MAS can divide a problem into smaller tasks and assign them to different specialized agents (Malec, 2025). The two key components in MAS are agents which can be software bots, sensors, or robots, and the environment, which can be static or dynamic (Malec, 2025). When MAS is used, agents need to “talk”, they rely on a shared state channel and can share information in an entire chain of thought or through sharing the end results (Malec, 2025).

Machine learning models need a high volume of data to be trained and tested on, and this data is often shared across different places and devices which creates privacy, intellectual property rights and ownership concerns (Lareo, 2025). Federated learning may be a solution for machine-learning models where “each federated device shares its local model parameters instead of sharing whole datasets used to train it” (Lareo, 2025). Meaning, instead of devices sharing the data with one another, each device trains a small version of the model on its own data and sends model updates to others, not sharing the data itself (Lareo, 2025). The communication set up of the models or the topology, dictates how updates are shared. These updates can be centralized with every device sending updates to a central server, combining them to improve the model, or in a peer-to-peer or hierarchical way, where devices share updates with only some peers, with no central server (Lareo, 2025). The key takeaway here is that federated learning trains a shared AI model by sharing model updates opposed to sharing private training data, the data stays local. Federated learning is often presented as a promising approach to protecting privacy and reducing cross-border data transfers because it keeps data local and shares only

model updates. While it may alleviate some concerns about data exposure, its real-world effectiveness depends on technical, infrastructural, and governance conditions that are uneven across global health contexts.

AI systems, both machine learning and deep learning models, depend on patterns learned from large, diverse data sets. In health contexts, these models need very large, high-quality data sets to distill patterns, where patterns may be rare (Wang, Pershing, & Lee, 2020). Further, data sets must be harmonized, meaning they are collected in similar ways, using the same variables, units, labeling conventions and with standardized definitions for clinical features, diagnoses and imaging (Wang, Pershing, & Lee, 2020). There is especially high risk of using AI in health care, as the AI models must be trained, tested and validated, across diverse populations, imaging devices, laboratory conditions and clinical environments (Wang, Pershing, & Lee, 2020). This ensures that AI models will work across different demographics and conditions. The more complex the data, which is the case in healthcare, the more example models are needed to detect patterns (Wang, Pershing, & Lee, 2020).

High and Low Income AI Divide

There is a growing divide in how AI is developed and deployed across high- and low-income countries. The economic and social benefits of AI are heavily concentrated in high-income countries that already possess advanced digital infrastructure, strong data ecosystems, and substantial resources for AI development (Schellekens and Skilling, 2024). Although AI has the potential to contribute up to USD $15.7 trillion to the global economy by 2030, the majority of these gains are expected to accrue to North America and China, while low-income regions capture only a small share of the benefits (Yu, Rosenfeld, and Gupta, 2023).

This readiness gap is reflected in global assessments. The Oxford Insights AI Readiness Index found that among 181 countries evaluated on their ability to integrate AI into public services, many of

the lowest scoring countries were in the Global South (Yu, Rosenfeld, and Gupta, 2023). The results highlight that governments require adequate operating environments to support effective AI development (Yu, Rosenfeld, and Gupta, 2023). These environments depend on a strong technology sector, reliable data infrastructure, and a clear strategic vision, along with attention to governance and ethics at the state level (Yu, Rosenfeld, and Gupta, 2023). Without these foundations, disparities in AI readiness will deepen existing global inequalities.

The unequal impacts of AI can be seen in the pervasiveness of digital colonialism. For example, in Africa, foreign technology companies control much of the continent's digital infrastructure and dominate its online platforms, giving them significant influence over political, economic, and cultural life (Salami, 2024). This external control allows these corporations to extract and store vast amounts of data from African users, reinforcing unequal data flows and creating dependency on Western technologies (Salami, 2024) As a result, AI systems and algorithms developed in the Global North are imposed on African contexts, often embedding harmful biases, enabling algorithmic oppression, and allowing powerful actors to exploit local populations (Salami, 2024). Together, these dynamics entrench existing inequalities and risk deepening Africa's marginalization in the global digital economy. These structural inequalities in AI development and digital ownership mirror the power imbalances already embedded within global biobanking networks shaping who ultimately benefits from clinical research.

Biobanking Data

Biobanks, dating back to the 1990s, is a critical research infrastructure in clinical trials, which rely on the retrieval, collection, storage, and preservation of human biological samples (Soares, Holzscheiter & Henrichsen, 2025). As biomedicine has expanded across borders, biobanking has become increasingly transnational, with human specimens and related data circulating globally despite the absence of international agreements governing these practices (Soares, Holzscheiter & Henrichsen,

2025). Biobanks now anchor national and international research ecosystems, supporting clinical studies across diseases and populations. Their rapid growth, from 145 biobanks in Europe in 2010 to 618 in 2021, demonstrates their rising importance (Soares, Holzscheiter & Henrichsen, 2025). This expansion, along with the movement of samples across borders, underscores how clinical trials depend on transnational flows of specimens and data. In the United States, access to specimens and data generally follows one of three models: open access, tiered access, or controlled access (Harrell & Rothstein, 2016). Open access allows unrestricted public availability, while controlled access limits use to approved researchers and protocols (Harrell & Rothstein, 2016). Tiered access occupies the middle ground, with restrictions based on donor consent, data sensitivity, or intended research use (Harrell & Rothstein, 2016).

Biobanking is undergoing a significant shift toward digitization. Biobanks are moving beyond physical sample repositories toward computational knowledge hubs, where large-scale datasets can be integrated with AI and machine learning tools (Frascarelli et al., 2023). As infrastructures grow more complex, combining both samples and expansive data resources, there is increasing emphasis on data utilization, digital transfer, and AI integration (Mayrhofer, 2025). One example of “digital biobanking” in practice and its power in AI research and development is in cancer research. Digital biobanking has the power to transform cancer research by pairing biological samples with AI-ready digital images and omics datasets (Frascarelli et al., 2023).

A useful way to analyze these emerging systems is through the size, site, access, and speed framework, which evaluates the volume of data, the physical or virtual location of repositories, access conditions, and the pace of AI integration (Mayrhofer, 2025). These dimensions help assess how effectively digital biobanks function. This transition supports improved risk prediction, disease stratification, biomarker discovery, multi-omics integration with phenotypic and clinical data, and more

efficient prioritization and quality control (Venturini, Faria & Cordeiro, 2025). It also enables scalability across samples and sites, improving pattern detection at population levels (Venturini, Faria & Cordeiro, 2025).

These developments, however, introduce new challenges. Data accessibility and variability remain major issues, as biobanks often rely on heterogeneous formats, incomplete datasets, inconsistent metadata, and differing sample processing protocols (Venturini, Faria & Cordeiro, 2025). Additional concerns include transparency and accountability, algorithmic bias, privacy and security risks, and gaps in governance and regulatory frameworks (Venturini, Faria & Cordeiro, 2025).

Biobanking in clinical trials raises significant policy challenges. National jurisdictions maintain divergent rules governing the use, storage, and circulation of human samples and data, creating persistent incompatibilities across borders (Soares, Holzscheiter & Henrichsen, 2025). The primary benchmarking instrument remains the *1997 Council of Europe Oviedo Convention on Human Rights and Biomedicine*, the only international treaty to establish binding rules relevant to biobanking (Soares, Holzscheiter & Henrichsen, 2025). Beyond this, protections rely on a patchwork of semi-related agreements, including *Article 12 of the Universal Declaration of Human Rights*, *Article 17 of the International Covenant on Civil and Political Rights, Article 8 of the European Convention on Human Rights, the 1980 OECD Privacy Guidelines* (renewed in 2013), *the Council of Europe Convention 108+, and the EU General Data Protection Regulation* (Soares, Holzscheiter & Henrichsen, 2025). These fragmented instruments do not fully address the scale of contemporary biobanking, which increasingly involves large volumes of cross-border sample and data flows that directly implicate questions of data sovereignty.

Biobanks form a direct bridge between global data inequalities and the technical demands of AI in clinical research. Because AI systems require large, harmonized datasets, the governance of biobanks defines whether biological samples and associated data can be standardized, linked, and shared across

trial sites. In low-and-middle-income countries, gaps in biobank infrastructure, inconsistent consent and export rules, and fragmented data standards determine not only who controls biological data but also whether AI protocols can function at all in multi-country trials.

Data Sovereignty

Data sovereignty is the principle that a jurisdiction has the authority to govern data generated within its borders (Imperva, 2025). In theory, states can and should regulate the collection, storage, processing, and distribution of data originating domestically, which directly affects AI use and transnational biobanking practices. Yet in the absence of clear international governance, countries lacking data localization requirements remain particularly vulnerable, as biological and genomic data can be stored, processed, or transferred abroad with limited oversight (Imperva, 2025). These vulnerabilities are already shaping international collaboration. China's 2021 Data Security Law restricts the export of designated categories of "important data," prompting several European research funders to pause collaborations with Chinese institutions due to uncertainty over data-transfer permissions (Silver, 2025). Similarly, in June 2025, the U.S. FDA announced new restrictions on exporting American biological samples to "hostile countries," largely low-income countries, framing genomic data security as a national security priority (U.S. FDA, 2025).

A further complication is that data residency does not necessarily guarantee data sovereignty. Hosting data within national borders does not ensure national control if the infrastructure is owned or operated by foreign firms (Richardson et al., 2025). Many (48%) of non-U.S. data centers, for example, are operated by U.S. companies, limiting the extent to which those centers fall under local legal authority and raising concerns about extraterritorial control (Richardson et al., 2025). This is particularly consequential for multi-country clinical trials that rely on biobanking infrastructure. One Cornell study, examining 775 of non-U.S. sites, found that many remain out of the control of foreign legal processes

due to the operator's nationality (Richardson et al., 2025). These asymmetries intensify when biological and genomic data from low- and middle-income countries are exported to high-income settings with advanced computational capacities, reinforcing global inequalities in data access, ownership, and benefit-sharing (Richardson et al., 2025).

Cloud Storage

` These sovereignty challenges are further compounded by the technical realities of cloud-based data storage, which weaken territorial control and complicates cross-border governance (Siry, 2019). Cloud architectures distribute data across multiple jurisdictions, making it difficult, even for service providers, to determine where specific data fragments physically reside (Siry, 2019). This undermines legal frameworks that rely on the physical location of data and fuels conflicting national claims over access and control (Siry, 2019). The absence of unified global rules on cloud based cross-border data flows creates legal fragmentation that exposes states to compliance risks and leaves data subjects vulnerable when information circulates through cloud environments (Ziyi, 2022). These cloud based challenges parallel the situation faced by AI-driven clinical trials that rely on biobank samples from low-resource settings, where sovereignty rules often restrict the movement of biological and genomic data even as AI systems require large, harmonized datasets. Siry (2019) also demonstrates that traditional mechanisms for cross-border data access have become incompatible with the speed and volume of cloud-era data flows, which has prompted states to adopt broad extraterritorial access laws like the U.S. CLOUD Act and similar EU proposals.

AI Multi-Agent Communication Protocols

When using AI in any form, systems depend on communication protocols to operate and interact with other systems, including the internet (Caballar and Stryker, 2025). These protocols include software engineering standards such as agent communication protocols, networking protocols, and data

exchange protocols. Agent protocols define how AI agents and other systems exchange information by specifying the syntax, structure, and sequence of messages, as well as conventions governing agent roles and when and how they respond to information (Caballar and Stryker, 2025). Without these standards, multi-agent systems cannot communicate. Effective agent protocols support interoperability, meaning that systems and software can exchange information in a consistent way, they reduce development complexity, and they support smoother system integration (Caballar and Stryker, 2025). Two of the most widely used agent communication languages are FIPA-ACL and KQML. These languages provide a structured way of sending messages between agents, which enables interoperability across different multi-agent systems (Kim Soon Gan et al., 2018).

A systematic review of multi-agent AI systems in health care identified several barriers that directly affect their use in clinical settings (Nweke et al., 2025). The review found persistent problems with data bias, inconsistent data formats, and difficulty integrating diverse data sources, all of which limited the performance of multi-agent systems in health care (Nweke et al., 2025). Ethical concerns were also common, including transparency, accountability, and trust issues. Because multi-agent systems involve multiple autonomous agents making decisions, it was often unclear which agent was responsible for a specific output, creating challenges for informed consent and ethical oversight (Nweke et al., 2025). The authors also noted significant difficulties integrating multi-agent systems into existing health care infrastructures such as electronic health records. Without standardized integration protocols, implementations caused disruptions or incompatibilities with legacy systems (Nweke et al., 2025). There were also concerns about real-world applicability because only a small number of systems in the review had been tested in clinical environments, while most remained at the simulation stage. This limited evidence about effectiveness, scalability, and safety (Nweke et al., 2025). Overall, the review shows that reliable, standardized communication protocols are essential for both AI and multi-agent systems. Without interoperability and consistent data foundations, the barriers identified in the review

become structural obstacles that interact with national data sovereignty rules and further disrupt the communication standards and cross-border data flows that AI-driven clinical trials depend on.

Interoperability and Ontologies

There are four levels of interoperability. Foundational interoperability represents the most basic tier, enabling one system to send or receive data from another, relying primarily on baseline IT connectivity rather than shared data structures (National Library of Medicine, n.d.). Structural interoperability builds upon this by aligning the format and syntax of exchanged data so that information preserves its structure, purpose, and framework during transmission. Semantic interoperability is more complex and involves maintaining consistent meaning across systems through unified terminologies, vocabularies, and ontologies; this level ensures that clinical, laboratory, epidemiological, or genomic information is interpreted consistently across sites. At the highest tier, organizational interoperability requires coherent governance, harmonized consent processes, legal alignment, and integrated operational workflows so that systems can exchange and use data in a coordinated and trustworthy manner (National Library of Medicine, n.d.). AI systems, particularly multi-country, biobank-supported, or multi-agent models, may function adequately when only foundational or structural interoperability is achieved, but they frequently fail at the semantic and organizational levels. Semantic failures arise when meaning is not preserved due to inconsistent metadata, incompatible clinical terminologies, or diverging formatting practices across health systems and biobanks, leading to misclassification, data loss, or reduced analytic validity. Organizational failures are even more pronounced in cross-border clinical research, where data sovereignty rules can block the transfer of biological or genomic data, consent requirements vary across jurisdictions, and legal frameworks conflict or lack clarity. These challenges become further compounded when cloud infrastructure is operated by foreign firms, when governance capacity is limited, or when institutional workflows remain siloed and

poorly integrated (National Library of Medicine, n.d.). These limitations reinforce why interoperability frameworks alone cannot guarantee seamless data exchange or reliable AI performance across regions. This is where ontologies become essential. Ontologies determine the meaning and relationships of concepts within a given domain and therefore provide the semantic structure by which clinical and biomedical knowledge can be contextualized and interpreted consistently (Ambalavanan et al., 2025). By delivering a shared, machine-readable vocabulary, ontologies serve as the semantic bridge between AI and healthcare, enabling harmonized data interpretation, automated reasoning, and precise clinical decision support (Ambalavanan et al., 2025). Standardized ontologies, aligned with international terminologies such as HL7 FHIR, SNOMED CT, LOINC, and ICD, ensure consistent representation of medical concepts and enable seamless data exchange across providers, research networks, and countries. In this sense, ontologies complement technical interoperability frameworks by embedding shared meaning into data and creating the semantic and organizational stability necessary for reliable AI training, cross-border analytics, and equitable integration of digital health innovations (Ambalavanan et al., 2025).

Global Standard for Health Data Exchange

Clinical research data must be traceable, accessible, interoperable, reproducible, and high quality for research findings to be meaningful and generalizable across regions (Pétavy, Seigneuret and Hudson, 2019). When data is collected in inconsistent formats across different sites, it becomes difficult to compare or aggregate results, which is especially problematic in multi-country trials that operate across heterogeneous healthcare systems, regulatory regimes, and data infrastructures (Pétavy, Seigneuret and Hudson, 2019). Approximately 85 percent of research studies fail to translate into meaningful clinical discoveries (Pétavy, Seigneuret and Hudson, 2019). Adopting a common set of data standards is therefore essential for improving reproducibility, reducing evidence gaps, and enabling cross-regional data integration (Pétavy, Seigneuret and Hudson, 2019).

Health Level Seven (HL7) provides an internationally recognized suite of standards for exchanging and integrating electronic health information (Mahon, 2023). HL7 Version 2 structures the content of messages exchanged between health systems, while Version 3 introduces more complex modeling for clinical and administrative data. HL7 Fast Healthcare Interoperability Resources (FHIR) focuses on simplicity, modularity, and web-based integration, allowing diverse systems to exchange data more efficiently. The Clinical Document Architecture (CDA) standardizes the structure of clinical documents, and the Continuity of Care Document (CDC) summarizes patient information for transitions between care settings. Together, these standards supply a common language that supports accurate and consistent data exchange across laboratories, electronic health records, imaging systems, pharmacy systems, and pathology workflows (Mahon, 2023).

These global health standards support the large and harmonized datasets required for machine learning, but despite possible advantages, adoption remains uneven (Osamika et al., 2025). Many institutions continue to rely on legacy architectures or fragmented data systems, and persistent inconsistencies in semantics, structure, and governance limit the ability to pool or repurpose data for research or AI training. These challenges are even more pronounced in low-resource settings where infrastructure, regulatory alignment, and data-governance capacity are limited (Osamika et al., 2025).

Analyses of digital public health infrastructures show that many countries still struggle with fragmented data governance, inconsistent standards, and uneven data quality, which restrict the integration of advanced analytics and multi-agent AI tools (Slawomirski et al., 2023). Before the COVID-19 pandemic, governments attempted to establish foundational governance structures to support standardized data sharing, especially for immunisation and respiratory disease surveillance (Slawomirski et al., 2023). The pandemic revealed major limitations, including inadequate data availability for marginalized and Indigenous populations, limited linkability across decentralized systems, and

inconsistent reporting mechanisms that hindered real-time analysis (Slawomirski et al., 2023). These gaps showed that limited interoperability undermines both crisis response and the routine use of digital tools in public health practice.

In response, many governments are now pursuing long-term strategies to build secure and equitable digital infrastructures. These efforts include establishing independent data stewardship authorities, investing in digital skills for health workers and data specialists, and developing unified standards for high-quality and privacy-conscious data exchange. Examples from New Zealand and the United Kingdom, as well as Australia's national digital capacity-building initiatives and Canada's interoperability mandates, illustrate how governance, workforce capacity, and technical standards must evolve together to enable effective data use (Slawomirski et al., 2023). These developments highlight a central point for AI and biobank-supported clinical research. Without coherent governance and standardized datasets, health systems cannot maintain the data stability required for reliable AI training, validation, or deployment across clinical contexts.

A comparative study of FHIR and openEHR illustrates the real-world complexity of interoperability (Allwell-Brown, 2016). Both standards support robust Application Programming Interfaces, extensible modelling, and consistent data semantics. However, their usability depends heavily on implementation capacity. openEHR, an open standard for electronic health records, offers highly detailed and semantically rich modelling that allows precise representation of patient data, but this complexity creates steep implementation requirements that many systems, especially in low-resource settings, cannot meet (Allwell-Brown, 2016). FHIR is easier to deploy due to its lightweight design, but it offers less semantic depth, which limits the detail available for AI models that rely on harmonized and high-resolution data (Allwell-Brown, 2016). When tested in a controlled acute-care decision support scenario, both standards performed reliably only under stable and well-resourced

conditions. Mapping heterogeneous clinical data into either standard proved difficult due to the diversity of modern health data, which includes structured fields, free-text notes, imaging, and device outputs. These findings show that interoperability standards do not, on their own, resolve data fragmentation (Allwell-Brown, 2016). They require institutional capacity, consistent modelling practices, and coherent governance. In multinational or biobank-supported AI research, particularly in low-resource settings, these implementation burdens underscore that technical standards alone cannot overcome the structural and organizational gaps that affect cross-border AI reliability (Allwell-Brown, 2016).

Discussion

Artificial intelligence systems require stable communication protocols, harmonized data formats, and shared ontologies to function across institutions and national borders. Protocols such as FIPA ACL and KQML define how multi-agent systems exchange structured messages, while standards like HL7 and HL7 FHIR specify how clinical data should be formatted and transmitted so that AI systems can interpret it consistently. However, across global health settings and particularly in low-resource environments, these technical requirements clash with fragmented infrastructure, inconsistent data vocabularies, and limited interoperability capacity. Systematic reviews of multi-agent AI systems repeatedly show that inconsistent data formats, fragmented sources, and biased or incomplete datasets undermine AI performance and reliability in clinical contexts (Nweke et al., 2025). These are not isolated technical errors but structural characteristics of many health systems where governance capacity, digital infrastructure, and standardization remain uneven (Slawomirski et al., 2023; Mahmoud et al., 2025).

Biobanks illustrate these tensions most clearly. Although contemporary biobanking has evolved into a transnational, data-intensive research infrastructure, national data sovereignty laws increasingly restrict how biological samples and genomic data can move across borders. Sovereignty frameworks are

meant to protect privacy, prevent extractive practices, and ensure local control over valuable biological resources. Policies such as China's Data Security Law and the United States' restrictions on exporting American biological data demonstrate how governments are consolidating authority over genomic data and biological materials. These protections are justified by long histories of exploitation, unequal benefit-sharing, and digital colonialism. Yet they directly conflict with the data-hungry nature of AI systems, which require large, diverse, and harmonized datasets to produce accurate and generalizable models. As Mayrhofer (2025) notes, both AI and biobanking depend on scale and diversity, but sovereignty, consent requirements, and governance constraints increasingly limit the ability to combine or share this data across borders. In multi-country clinical trials, this creates practical obstacles for models that require pooled biospecimen data, harmonized metadata, or cross-site training datasets.

Interoperability gaps further amplify these constraints. Although HL7 and FHIR offer standardized data structures, implementation in real clinical systems remains highly uneven. Many institutions continue to rely on siloed databases or legacy systems that cannot communicate meaningfully with AI architectures that expect structured, machine-readable data. Allwell-Brown's comparison of FHIR and openEHR demonstrates that even modern standards work reliably only under well-resourced conditions, and mapping heterogeneous clinical data into these frameworks is technically complex (Allwell-Brown, 2016). In low-resource settings, the challenge is magnified by limited digital infrastructure, inconsistent metadata quality, and variable adherence to modeling standards (Osamika et al., 2025). Without semantic consistency, AI models trained in one context cannot be safely deployed in another, and federated approaches still require shared ontologies and governance agreements to align local data sources.

These technical barriers interact with deeper geopolitical and sovereignty concerns. Countries in the Global South face genuine risks of digital colonialism, unequal data extraction, and loss of

governance control over genomic data. Biobank networks and AI-driven research can reproduce historical patterns of extraction when data is centralized outside the originating country, governed by foreign institutions, or used for commercial gains without equitable benefit-sharing. As a result, governments impose stricter controls over data export, which protects local interests but restricts the cross-border data flows required for AI-supported clinical trials.

Taken together, the evidence reveals a structural bottleneck. AI in global health research and development depends on interoperability, harmonized data, and cross-site communication, yet the systems needed to support this are constrained by legal restrictions, inconsistent standards, legacy infrastructure, and sovereignty-based protections. These obstacles are most pronounced in low-resource settings, where interoperability frameworks are unevenly implemented, governance systems are fragmented, and digital infrastructure lags behind. The result is that AI models cannot be reliably trained, validated, or deployed across borders, particularly within multi-country clinical trials dependent on biobank specimens and genomic data.

Policy Recommendations

Effective and equitable integration of artificial intelligence into global health research requires policy responses that directly address the structural barriers identified in this paper. First, governments and research institutions should establish internationally aligned governance frameworks that clarify consent, data use, and benefit sharing for biobank supported AI research. This includes adopting dynamic or tiered consent models (such as in the U.S.) that allow participants to understand and authorize future AI uses of their biological samples and associated data. Second, multi-country trials should adopt interoperable data standards such as HL7 and HL7 FHIR and ensure that trial sites in low resource settings have the technical and regulatory capacity to integrate these standards into existing systems. Investments in digital infrastructure, standardized protocols, and workforce development are

essential for data harmonization and for supporting AI that is reliable and reproducible. Third, countries should create coordinated data sovereignty agreements that protect local control over sensitive health and genomic data while still enabling responsible cross border scientific collaboration. Policymakers should prioritize models such as federated learning and compute data arrangements which allow AI analysis without the movement of raw biospecimens or personal data outside national boundaries. Fourth, national governments and global funders should commit to building independent health data stewardship bodies that can oversee ethical data use, manage access requests, and ensure accountability and transparency. These bodies should reflect local values and protect marginalized and underrepresented groups whose data is often at risk of exploitation. Finally, global health agencies, including the WHO, should convene an international working group to align regulatory approaches for AI enabled clinical trials, since inconsistent national rules on data export, biospecimen handling, and cloud storage currently impede multi-country research. Policy leadership at the international level is necessary to prevent fragmented standards from reinforcing existing inequities in AI readiness across regions.

Conclusion

AI, biobanking, and global health research are increasingly intertwined, yet this convergence exposes deep inequalities in data governance, digital capacity, and the ability of different countries to shape the direction of emerging technologies. This paper has shown that AI systems depend on harmonized, high quality, and interoperable datasets, but many low resource settings lack the infrastructure, regulatory coordination, and data governance systems required to support them. Biobanks have transitioned from physical repositories to complex computational infrastructures, and their value for AI supported research depends on the consistent application of standards, ethical oversight, and equitable approaches to data sharing. At the same time, national data sovereignty laws, concerns over cross border data flows, and geopolitical tensions increasingly shape how and where biological samples and genomic data can be used. These dynamics create obstacles for multi-country clinical trials and

undermine efforts to create AI tools that are valid, safe, and generalizable across populations. Without coherent governance frameworks and interoperable systems, AI risks amplifying global inequities rather than reducing them.

Strengthening governance, supporting equitable data infrastructures, and enabling responsible models for international data collaboration are essential if AI is to contribute to global health in meaningful and just ways. The policy recommendations outlined here offer a pathway toward aligning technical standards with ethical and geopolitical realities. By investing in data stewardship, harmonization, and sovereignty respecting collaboration, the international community can create conditions in which AI enhanced research benefits populations in low and high resource settings alike. In doing so, global health systems can move toward a future where innovation is matched with accountability, and where the use of AI supports, rather than undermines, principles of equity and trust.

Work Cited

Allwell-Brown, E. (2016). *A Comparative Analysis of HL7 FHIR and openEHR for Electronic Aggregation, Exchange and Reuse of Patient Data in Acute Care* (Master's thesis, Karolinska Institutet).https://www.researchgate.net/publication/308903413_A_Comparative_Analysis_of_HL7_FHIR_and_openEHR_for_Electronic_Aggregation_Exchange_and_Reuse_of_Patient_Data_in_Acute_Care/link/5ab4de4445851515f59924c3/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19

Ambalavanan, R., Snead, R. S., Marczika, J., Towett, G., Malioukis, A., & Mbogori-Kairichi, M. (2025). Ontologies as the Semantic Bridge Between Artificial Intelligence (AI) and Healthcare, Enabling Structured Knowledge Frameworks that Enhance Data. *Frontiers in Digital Health, 7*, 1668385. https://doi.org/10.3389/fdgth.2025.1668385

Caballar, R., Stryker, C. (2025). *What are AI agent protocols?* IBM. Retrieved from https://www.ibm.com/think/topics/ai-agent-protocols#:~:text=AI%20agent%20protocols%20offer%20a,for%20a%20smoother%20enterprise%20integration.

Fan, Q., & Qiang, C. Z. (2024). *Tipping the scales: AI's dual impact on developing nation*s. World Bank Blogs. https://blogs.worldbank.org/en/digital-development/tipping-the-scales--ai-s-dual-impact-on-developing-nations

Frascarelli, C., Bonizzi, G., Musico, C. R., Mane, E., Cassi, C., Guerini Rocco, E., Farina, A., Scarpa, A., Lawlor, R., Reggiani Bonetti, L., Caramaschi, S., Eccher, A., Marletta, S., & Fusco, N. (2023). Revolutionizing Cancer Research: The Impact of Artificial Intelligence in Digital Biobanking. *Journal of Personalized Medicine, 13*(9), 1390. https://doi.org/10.3390/jpm13091390

Goodluck, K. (2020). *Indigenous data sovereignty shakes up research*. High Country News. https://www.hcn.org/issues/52-11/indigenous-affairs-covid19-indigenous-data-sovereignty-shakes-up-research/

Harrell, H. L., & Rothstein, M. A. (2016). Biobanking Research and Privacy Laws in the United States. *The Journal of Law, Medicine & Ethics, 44*(1), 106–127. https://doi.org/10.1177/1073110516644203

Imperva. (2025). *What is data sovereignty? Challenges & best practices*. Retrieved from https://www.imperva.com/learn/data-security/data-sovereignty/

Kim Soon Gan, Kim On Chin, Anthony, P., & Hamdan, A. R. (2018). A FIPA-ACL ontology in enhancing interoperability in multi-agent communication. In *Computational Science and Technology* (pp. 151–160). Lecture Notes in Electrical Engineering. https://doi.org/10.1007/978-981-10-8276-4_15

Lareo, X. (2025). *Federated Learning.* European Data Protection Supervisor. Retrieved from https://www.edps.europa.eu/press-publications/publications/techsonar/federated-learning_en

Mahmoud, M. A., Abdelghaffar, H., Khalifa, E. A., & Ali, A. M. (2025). A multi-country review of the governance of hospital information systems interoperability. *Eastern Mediterranean Health Journal, 31*(9), 581–589. https://www.emro.who.int/emhj-volume-31-2025/volume-31-issue-9/a-multi-country-review-of-the-governance-of-hospital-information-systems-interoperability.html

Mahon, C. (2023). *HL7 explained: Health Level Seven standards, messages and integration in healthcare*. SciSure. https://www.scisure.com/blog/hl7-explained-digital-lab

Malec, M. (2025). *Understanding Agents and Multi Agent Systems for Better AI Solutions*. hatchWorksAI. Retrieved from https://hatchworks.com/blog/ai-agents/multi-agent-systems/#:~:text=A%20multi%2Dagent%20system%20in%20AI%20(or%20MAS%20for%20short,the%20dynamic%20interplay%20between%20agents.

Mayrhofer, M. T. (2025). How the World of Biobanking is Changing with Artificial Intelligence. *Frontiers in Digital Health, 7*, Article 1626833. https://doi.org/10.3389/fdgth.2025.1626833

National Library of Medicine. (n.d.). *Health data standards and terminologies: A tutorial.* U.S. National Library of Medicine, National Institutes of Health. https://www.nlm.nih.gov/oet/ed/healthdatastandards/03-500.html

Nweke, I. P., Ogadah, C. O., Koshechkin, K., & Oluwasegun, P. M. (2025). Multi-agent AI systems in healthcare: A systematic review enhancing clinical decision-making. *Asian Journal of Medical Principles and Clinical Practice, 8*(1), 273–285. https://doi.org/10.9734/ajmpcp/2025/v8i1288

Osamika, D., Adelusi, B. S., Kelvin-Agwu, M. T. C., Mustapha, A. Y., Forkuo, A. Y., & Ikhalea, N. (2025). A critical review of health data interoperability standards: FHIR, HL7, and beyond. *World Scientific News, 203*, 194–233. https://worldscientificnews.com/a-critical-review-of-health-data-interoperability-standards-fhir-hl7-and-beyond/

Pétavy, F., Seigneuret, N., & Hudson, L. D. (2019). Global standardization of clinical research data. *Applied Clinical Trials, 28*(4). Retrieved from https://www.appliedclinicaltrialsonline.com/view/global-standardization-clinical-research-data

Richardson, A., Yi, H., Nie, M., Wisdom, S., Price, C., Weijers, R., Veld, S., & Baker, M. (2025). *How sovereign is sovereign compute? A review of 775 non-U.S. data centers* (arXiv:2508.00932). arXiv. https://doi.org/10.48550/arXiv.2508.00932

Salami, A. O. (2024). Artificial intelligence, digital colonialism, and the implications for Africa's future development. *Data & Policy, 6*. Cambridge University Press. https://www.cambridge.org/core/journals/data-and-policy/article/artificial-intelligence-digital-colonialism-and-the-implications-for-africas-future-development

Schellekens, P., & Skilling, D. (2024). *Three reasons why AI may widen global inequality*. Center for Global Development. https://www.cgdev.org/blog/three-reasons-why-ai-may-widen-global-inequality

Silver, A. (2025). *China's data protection rules prompt pause from major European research funders*. Reuters. Retrieved from https://www.reuters.com/sustainability/society-equity/chinas-data-protection-rules-prompt-pause-major-european-research-funders-2025-04-25/

Siry, L. (2019). Cloudy days ahead: Cross-border evidence collection and its impact on the rights of EU citizens. *New Journal of European Criminal Law*, *10*(3), 227–250. https://doi.org/10.1177/2032284419865608

Slawomirski, L., Lindner, L., de Bienassis, K., Haywood, P., Tiago Cravo Oliveira Hashiguchi, Steentjes, M., & Oderkirk, J. (2023). Progress on implementing and using electronic health record systems. *OECD Health Working Papers*. https://doi.org/10.1787/4f4ce846-en

Soares, M. W., Holzscheiter, A., & Henrichsen, T. (2025). Biobanking as a contentious issue in global health governance: Diversification and contestation of policy frames in international biobanking debates. *Social Science & Medicine, 369*, Article 117773. https://doi.org/10.1016/j.socscimed.2025.117773

Stryker, C., & Kavlakoglu, E. (2025). *What is artificial intelligence (AI)?* IBM. Retrieved from https://www.ibm.com/think/topics/artificial-intelligence

The Allan Turing Institute. (2025). *Multi-agent systems.* Retrieved from https://www.turing.ac.uk/research/interest-groups/multi-agent-systems

U.S. Food and Drug Administration. (2025). *FDA halts new clinical trials that export Americans' cells to foreign labs in hostile countries for genetic engineering*. Retrieved from https://www.fda.gov/news-events/press-announcements/fda-halts-new-clinical-trials-export-americans-cells-foreign-labs-hostile-countries-genetic

Venturini, P., Faria, P. L., & Cordeiro, J. V. (2025). AI and omics technologies in biobanking: Applications and challenges for public health. *Public Health (London), 243*, Article 105726. https://doi.org/10.1016/j.puhe.2025.105726

Wang, S. Y., Pershing, S., & Lee, A. Y. (2020). Big data requirements for artificial intelligence. *Journal of Big Data, 7*(1), 18. https://doi.org/10.1186/s40537-020-00320-2

Yu, D., Rosenfeld, H., & Gupta, A. (2023). *The 'AI divide' between the Global North and the Global South*. World Economic Forum. https://www.weforum.org/stories/2023/01/davos23-ai-divide-global-north-global-south/

Ziyi, X. (2022). International Law Protection of Cross-Border Transmission of Personal Information Based on Cloud Computing and Big Data. *Mobile Information Systems*, *2022*, Article 9672693. https://doi.org/10.1155/2022/9672693