

DIGITAL INCLUSION VS DIGITAL SOVEREIGNTY: RRI AS A PLATFORM FOR INTEGRATING ETHICS INTO GEOPOLITICS

OLEG N. GUROV*  

*PhD in Philosophy, MBA, Research fellow at AI Centre of Moscow State Institute of International Relations of the Ministry of Foreign Affairs of Russia,
Associate Professor at Moscow State University after M. V. Lomonosov, Russia*

Abstract

The rapid advancement of artificial intelligence (AI) has intensified the global tension between digital inclusion, which advocates for equitable access to technology, and digital sovereignty, emphasizing national control over data and infrastructure. This article examines how Responsible Research and Innovation (RRI) can reconcile these competing imperatives by embedding ethical principles, such as inclusivity, sustainability, precaution, and reflexivity, into technology governance. Through qualitative case studies in education (e.g., Kenya's eLimu and India's DIKSHA platforms) and healthcare (e.g., WHO's pandemic data-sharing protocols), the study demonstrates that RRI fosters participatory design, balances sovereignty with global collaboration, and mitigates systemic biases. Findings reveal that rigid sovereignty policies often exacerbate inequalities, while RRI-driven frameworks enable marginalized communities to co-create solutions, ensuring culturally relevant and ethically aligned technologies. The analysis highlights RRI's potential to transform geopolitical competition into equitable governance, advocating for its institutionalization through international mechanisms such as the UN's Global Digital Compact. By prioritizing social justice, RRI redefines sovereignty as a stewardship obligation, ensuring AI development uplifts, rather than undermines, vulnerable populations. The article concludes that integrating RRI principles into both global and local agendas is crucial for dismantling digital hierarchies and promoting inclusive innovation.

Keywords: *digital inclusion, digital sovereignty, responsible research and innovation, AI ethics, social justice, global governance.*

* Dr. Oleg Gurov is a Research fellow at AI Centre of Moscow State Institute of International Relations of the Ministry of Foreign Affairs of Russia. His research interests include digital immortality, autonomy loss in hyper-technologized societies, and the cultural anxieties reflected in neurotechnologies and AI integration.

Email: gurov-on@ranepa.ru ORCID iD: 0000-0002-8425-1338



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Received: 23.04.2025
Revised: 09.05.2025
Accepted: 20.05.2025

© The Author(s) 2025

Introduction

The swift advancement of artificial intelligence (AI) has triggered a global race for technological dominance, yet progress remains uneven. While high-income nations and corporations invest heavily in AI infrastructure, low- and middle-income countries (LMICs) face widening gaps in access, exacerbating technological inequality.¹ This disparity is compounded by the tension between two competing imperatives: digital inclusion, which advocates for universal access to technology as a fundamental right, and digital sovereignty, which prioritizes national control over data governance and infrastructure (e.g., GDPR in the EU, China's "Great Firewall").

This dichotomy is acutely evident in education and healthcare, two domains critical to social equity. For instance, restrictive data sovereignty policies, such as bans on foreign EdTech platforms in Global South nations, often limit access to global educational resources. As a result, marginalized rural populations are often the most affected.² Similarly, in healthcare, sovereignty-driven restrictions on medical data sharing, as seen during the COVID-19 pandemic, hinder collaborative research while failing to address systemic biases in AI-driven diagnostics, which frequently underrepresent minority groups.³ Ironically, policies designed to safeguard national interests sometimes deepen existing inequalities.

Social justice is key to reconciling these priorities. Incorporating principles like inclusivity and reflexivity into global strategies can help reduce the specific ethical challenges, such as algorithmic bias, data privacy breaches, and digital exclusion.⁴ The EU's ethical AI standards demonstrate that international cooperation, grounded in principles of fairness, can balance sovereignty and inclusion. This ensures technologies benefit the public rather than reinforcing inequality.

This study explores how Responsible Research and Innovation (RRI) can help reconcile the tensions between digital inclusion and sovereignty. Specifically, it aims to:

1. Evaluate RRI's potential as a transnational platform for harmonizing ethical technology development with divergent national priorities (e.g., data protection vs. equitable access).

¹ International Telecommunication Union (ITU), *Digital Inclusion of All* (ITU, 2023), <https://www.itu.int/en/mediacentre/backgrounders/Pages/digital-inclusion-of-all.aspx>.

² UNESCO, Global Education Monitoring Report 2023: Technology in Education: A Tool on Whose Terms? (Paris: UNESCO, 2023), 132.

³ World Health Organization, *Ethical Guidelines for Health Data Sharing in Pandemic Response* (WHO, 2022), 22, <https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc-79ca799dce4d.pdf>.

⁴ Jack Stilgoe, "Why Responsible Innovation?" in *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, ed. Richard Owen, John Bessant, and Maggy Heintz (Chichester, UK: Wiley, 2013), 63.

2. Demonstrate RRI's operational capacity to institutionalize ethics in AI systems, thereby preempting discrimination in education and healthcare digital systems exacerbated by algorithms trained on non-representative datasets.

The urgency of this issue stems from the inadequacy of current governance models in addressing global inequities. For instance, while the EU's AI Act mandates risk-based assessments and transparency, its enforcement remains confined to member states, lacking mechanisms to address global inequities in AI access.⁵ Similarly, sovereignty-driven constructs, such as China's restrictions on cross-border data flows, often overlook the lack of access to digital literacy initiatives.⁶ RRI's focus on including stakeholders and ensuring fair outcomes provides a practical solution. Including the findings of an analysis of the needs of marginalized groups, such as rural communities and refugees, can help RRI reconcile digital sovereignty with inclusivity and avoid reinforcing bias within technology development processes.

Furthermore, the study will analyze RRI's role in the practical implementation of ethical norms. For example, the WHO's ethical data-sharing protocols during the pandemic exemplify how RRI's precautionary principle can reconcile sovereignty concerns (e.g., national security) with global health equity, enabling collaborative research while safeguarding privacy.⁷ These cases illustrate RRI's potential to mediate geopolitical conflicts while fostering equitable AI governance.

Theoretical Framework

Digital inclusion and digital sovereignty represent two competing yet interconnected paradigms that dominate current discussions surrounding technology governance.⁸ Digital inclusion focuses on ensuring equal access to technology, improving digital literacy, and breaking down barriers to participation in the digital economy. It involves expanding broadband infrastructure, ensuring affordable devices, and initiating programs designed to empower marginalized communities. For instance, the U.S. Federal Communications Commission's Lifeline Program aims to provide universal access by subsidizing broadband for underserved populations. Similarly, pilot initiatives in Michigan offer digital literacy training to refugees, while tele-

⁵ European Commission, *Proposal for a Regulation on Artificial Intelligence (EU AI Act)* (European Commission, 2023), art. 29, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A-52021PC0206>.

⁶ UNESCO, Global Education Monitoring Report 2023, 132.

⁷ WHO, Ethical Guidelines for Health Data Sharing, 16.

⁸ Samuele Fratini, Emmie Hine, Claudio Novelli, Huw Roberts, and Luciano Floridi, "Digital Sovereignty: A Descriptive Analysis and a Critical Evaluation of Existing Models," *Digital Society* 3, no. 3 (December 2024): 59, <https://doi.org/10.1007/s44206-024-00146-7>.

medicine programs in rural India have addressed healthcare access challenges, illustrating the importance of capacity-building efforts. Another vital dimension of digital inclusion involves addressing systemic disparities across sectors such as education and healthcare. This includes initiatives aimed at expanding access to digital learning through online platforms designed for underserved communities, as well as efforts like the World Health Organization's telehealth guidelines tailored to the specific needs of low- and middle-income countries (LMICs). These approaches reflect a broader commitment to ensuring that digital transformation benefits all, regardless of geographic or socio-economic status.

In contrast, the concept of digital sovereignty emphasizes the authority of individual nations or regions to govern their digital infrastructure, data flows, and regulatory frameworks. It underscores the importance of maintaining local control over technological ecosystems in the face of increasing global interdependence.⁹ This paradigm manifests through stringent data protection laws, such as the European Union's General Data Protection Regulation (GDPR), which prioritizes individual privacy and data security. However, while reinforcing the rights of citizens, it also complicates cross-border data flows that are essential for global collaboration, innovation, and the operation of transnational digital services.¹⁰ Geopolitical control constitutes yet another layer of digital sovereignty, exemplified by China's "Great Firewall," which restricts access to foreign technology platforms to safeguard national security and preserve ideological autonomy. While such measures aim to reinforce sovereignty, they often do so at the expense of access to global educational, scientific, and technological resources. Export controls further underscore the geopolitical stakes of digital governance. For instance, U.S. restrictions on the export of advanced AI chips to China have not only impacted domestic capabilities but have also inadvertently constrained access to high-performance computing resources in low- and middle-income countries (LMICs), thereby impeding efforts toward equitable technological development.

Tensions between digital inclusion and digital sovereignty frequently emerge, revealing deep-seated structural contradictions. The GDPR, for example, enforces rigorous data localization and privacy protections that, while crucial for individual rights, can obstruct cross-border initiatives such as international telemedicine partnerships, particularly in regions lacking interoperable infrastructure. Conversely, India's Aadhaar system illustrates how sovereignty-driven digital infrastructure, when underpinned by equity-focused design principles, can enhance inclusion, fa-

⁹ Fratini et al., "Digital Sovereignty," 61.

¹⁰ European Parliament, General Data Protection Regulation), Official Journal of the European Union L 119, (May 4, 2016): 45.

cilitate more efficient distribution of public resources, and expand access to essential services.

These enduring tensions underscore the urgent need for governance frameworks capable of harmonizing ethical commitments to inclusion with the practical imperatives of digital sovereignty—a balance that is increasingly vital in a globally interdependent digital landscape. However, as poststructuralist thinkers like Foucault and Derrida argue, deconstructing the binary opposition between inclusion and sovereignty reveals their potential complementarity. For instance, sovereignty can enable inclusion through localized data governance that protects marginalized groups while fostering global cooperation.¹¹ RRI emerges as a promising paradigm for addressing these tensions by embedding principles of inclusivity and reflexivity into governance structures to ensure that technologies serve both the public good and national interests without reinforcing or perpetuating exclusionary tendencies.

The global AI race has intensified the friction between digital inclusion and sovereignty, as nations prioritize control over infrastructure and data to secure competitive advantages. Sovereignty-driven policies, while ostensibly protective, often undermine transnational collaboration critical for addressing shared challenges. Strict data localization laws, like those in Russia and India during the COVID-19 pandemic, delayed the sharing of genomic data and slowed variant tracking.¹² Similarly, the EU's GDPR, despite its privacy safeguards, created barriers for LMICs seeking to leverage European health AI tools due to compliance complexities, exacerbating diagnostic inequities in regions with limited regulatory oversight.¹³

This conflict has a particularly negative impact on marginalized groups. Over-regulation in the name of sovereignty often manifests as blanket bans on foreign technologies, such as Ethiopia's restriction of global EdTech platforms, which left rural schools without alternatives for digital learning resources.¹⁴ On the other hand, under-regulated inclusion efforts, like India's Aadhaar system, show how flawed frameworks can exclude vulnerable populations. This creates a difficult situation: too much control limits access, while weak governance leaves vulnerable communities exposed to algorithmic bias, as evidenced by AI-enhanced medical devices like pulse oximeters that function less accurately on individuals with darker skin tones due to being trained on non-representative datasets.¹⁵

¹¹ Fratini et al., "Digital Sovereignty," 61.

¹² WHO, Ethical Guidelines for Health Data Sharing, 13.

¹³ ITU, *Digital Inclusion of All*, 28.

¹⁴ UNESCO, Global Education Monitoring Report 2023, 111.

¹⁵ Johannes Machinya, "Blog series on exploring the Intersections of Technology, Health, and Law: Data bias and the risk of algorithmic apartheid in South African healthcare," *SLSA Blog*, September

The social justice implications are stark. Sovereignty measures like China's tech export restrictions on AI chips deepen resource asymmetry, leaving LMICs dependent on outdated infrastructure. Meanwhile, exclusion from digital literacy programs, exemplified by the U.S. rural broadband gap, traps marginalized populations in cycles of disenfranchisement.¹⁶ These cases illustrate a paradox: policies framed as protective often prioritize state or corporate interests over equity, necessitating RRI to realign incentives.

RRI offers a practical platform to address the conflict between inclusion and sovereignty by integrating four key principles - inclusivity, sustainability, precaution, and reflexivity- into technology development. These principles enable stakeholders to navigate ethical dilemmas while balancing geopolitical priorities.

1. Inclusivity: By involving marginalized communities in design processes, RRI ensures technologies address localized needs without compromising sovereignty. This mirrors Gilligan's ethics of care, where moral decisions emerge from dialogue with specific communities, not universal axioms. For example, the EU's Horizon 2020-funded AI4People initiative engages civil society groups, policy-makers, and technologists in co-designing ethical AI tools, ensuring marginalized voices shape the governance ecosystem.¹⁷ Similarly, Kenya's telemedicine projects, developed with rural healthcare workers, demonstrate how participatory design bridges access gaps while respecting data sovereignty through localized consent protocols.¹⁸ Also, another illustrative example is the Africa PGI 2.0 project implemented by the Africa CDC, which involves the development of scenarios for organizing genomic surveillance and epidemiological control in African countries, considering regional specifics and updating data obtained by local communities of specialists.¹⁹ However, the principles of RRI require institutional democratization, which contradicts authoritarian tendencies in the management of social systems. For example, in China, the rhetoric of "inclusivity" is at times employed to justify expansive digital surveillance, illustrating how frameworks such as Responsible Research and Innovation (RRI) can be co-opted to legitimize control rather than uphold a genuine ethic of inclusion.

28, 2023, para. 9, <https://slsablog.co.uk/blog/blog-posts/blog-series-on-exploring-the-intersections-of-technology-health-and-law-data-bias-and-the-risk-of-algorithmic-apartheid-in-south-african-health-care/>.

¹⁶ Link Health, *Bridging the Digital Divide in Rural America* (Washington, DC: Link Health, 2024), <https://link-health.org/2024/07/16/bridging-the-digital-divide-in-rural-america/>.

¹⁷ European Commission, Proposal for a Regulation (EU AI Act), art 29.

¹⁸ WHO, Ethical Guidelines for Health Data Sharing, 39.

¹⁹ African Union Commission, *US\$100 Million Africa Pathogen Genomics Initiative to Boost Disease Surveillance and Emergency Response Capacity in Africa* (Addis Ababa: African Union Commission, 2020), <https://au.int/fr/node/39401>.

2. Sustainability: RRI emphasizes the long-term well-being of societies over immediate technological or economic gains. This is evident in the European Union's Ethics Guidelines for Trustworthy AI, which require environmental impact assessments for AI systems, thereby aligning innovation with broader goals of climate justice. A similar commitment is reflected in India's Digital Public Infrastructure, which integrates open-source technologies, such as the Aadhaar system, with energy-efficient data centers, aiming to ensure equitable access to digital services without exacerbating environmental degradation.

3. Precaution: A forward-looking approach to risk is essential for preventing unintended consequences. The World Health Organization's Data Sharing Agreement during the COVID-19 pandemic serves as a case in point: it enabled global cooperation in health research and crisis response, while incorporating safeguards such as data anonymization and controlled access to protect national and individual privacy.²⁰ Similarly, the Montreal Declaration for Responsible AI advocates for algorithmic audits to mitigate biases in facial recognition technologies. This directly addresses sovereignty concerns, such as GDPR compliance, while simultaneously protecting minority rights.

4. Reflexivity: Continuous evaluation ensures adaptive governance. The Montreal Declaration for Responsible AI provides an alternative governance model, emphasizing stakeholder co-creation and algorithmic fairness audits. While less binding than the EU AI Act, it offers a more flexible architecture for LMICs with varying regulatory capacities. In education, UNESCO's Dynamic Coalition on Digital Inclusion uses iterative stakeholder consultations to refine EdTech policies, balancing open access with national content regulations.²¹

These examples show how RRI can turn ethical principles into practical governance. By promoting transparency and equity, the RRI paradigm provides a guide for balancing inclusion and sovereignty.

Methodology: The methodology used in this article to explore the intersection of digital inclusion, digital sovereignty, and ethical concepts such as RRI is primarily qualitative, based on a combination of case study and comparative analysis. The study synthesizes data from a variety of sources, including international policy documents, academic literature, and real-world cases, to create a theoretical and empirical framework.

²⁰ WHO, Ethical Guidelines for Health Data Sharing, 20.

²¹ UNESCO, Global Education Monitoring 2023, 81.

Balancing Digital Sovereignty with Inclusion and Global Collaboration

The analysis reveals that stringent digital sovereignty policies, while intended to protect national interests, frequently exacerbate educational inequality by limiting access to global resources. For instance, Ethiopia's 2023 ban on foreign EdTech platforms (e.g., Google Classroom) left rural schools without alternatives for digital curricula, widening the gap between urban and rural literacy rates by 18%.²² India's early focus on data localization delayed access to international STEM resources in public schools, affecting marginalized communities the most.

Concerning RRI as a Mitigation Strategy, co-developing open-source platforms with local communities emerged as a critical solution. Kenya's eLimu platform, designed through participatory workshops with teachers, parents, and students, exemplifies RRI's inclusivity principle. By incorporating indigenous languages and culturally relevant content, eLimu achieved a 32% increase in primary school engagement in underserved regions while complying with national data governance laws.²³ Similarly, India's DIKSHA initiative, a sovereign digital infrastructure co-created with state educators, provided localized curricula in 31 languages, reducing reliance on foreign platforms without sacrificing access to quality content.²⁴

These projects underscore RRI's capacity to balance sovereignty and inclusion. For example, reflexivity mechanisms in DIKSHA's design allowed continuous feedback from rural teachers, enabling iterative improvements to address connectivity challenges. Conversely, Ethiopia's exclusionary bans, lacking such participatory models, resulted in prolonged educational disparities, highlighting the risks of sovereignty-first approaches devoid of RRI principles.

The tension between data sharing for global health research and sovereignty-driven restrictions has proven particularly acute in crisis contexts like the COVID-19 pandemic. Beyond COVID-19, sovereignty concerns have also impacted genomic research for diseases like Ebola. During the 2014 outbreak, delays in sharing pathogen data due to intellectual property disputes hindered the development of vaccines, highlighting the need for globally recognized benefit-sharing mechanisms. Similarly, the EU's GDPR, while safeguarding privacy, created compliance barriers for LMICs seeking access to European AI diagnostic tools, exacerbating disparities in pandemic response capabilities.²⁵

Ethical data-sharing protocols, grounded in RRI's precaution and inclusivity principles, have emerged as viable compromises. The WHO's Data Sharing

²² UNESCO, Global Education Monitoring Report 2023, 32.

²³ UNESCO, Global Education Monitoring Report 2023, 122.

²⁴ Ministry of Education, Government of India, *DIKSHA Platform: Annual Report* (New Delhi: Ministry of Education, 2023), 17.

²⁵ ITU, *Digital Inclusion of All*, 28.

Agreement enabled anonymized aggregation of clinical data into a global repository while allowing nations to retain sovereignty over raw datasets.²⁶ This approach allowed real-time tracking of variants while preserving national security, cutting diagnostic delays by 34% in participating LMICs.

In telemedicine, India's National Digital Health Mission (NDHM) adopted RRI-aligned federated learning models, where AI algorithms are trained on decentralized data without transferring sensitive patient information across borders. This approach resolved sovereignty concerns under GDPR while improving rural diagnostic accuracy by 27%.²⁷ Similarly, the Africa CDC's Pathogen Genomics Initiative uses RRI's reflexivity principle, enabling member states to audit data usage and revoke access, fostering trust in cross-border collaborations.

These cases demonstrate that RRI mechanisms do not merely balance competing interests but redefine sovereignty as a facilitator, not a barrier, to inclusion. By prioritizing equitable benefit-sharing (e.g., ensuring LMICs receive vaccines developed using their data), such frameworks align geopolitical priorities with social justice imperatives. Moreover, the introduction of such frameworks, focused on the values of diversity and non-exclusion, can help resolve the urgent problem noted by researchers of the imbalance between the consideration of the interests of the Global North and the Global South in the application of intelligent technologies and digitalization strategies.²⁸

The global AI race risks creating ethical disparities, as varying national standards can lead to technologies that harm marginalized groups. For example, concerns about “techno-racism” or “algorithmic apartheid” are substantiated by evidence showing AI-enhanced devices such as pulse oximeters work “less well in [individuals] with darker skin,” making it more difficult to detect dangerous drops in oxygen levels. This inaccuracy stems from algorithms often trained on datasets drawn predominantly from populations with European ancestry.²⁹ Similarly, facial recognition tools deployed in U.S. law enforcement, optimized for lighter skin tones, misidentified minorities with error rates up to 34%, perpetuating discriminatory policing practices.³⁰ Such cases illustrate how sovereignty-centric innovation, unchecked by universal ethical guardrails, institutionalizes structural discrimination.

²⁶ WHO, Ethical Guidelines for Health Data Sharing, 15.

²⁷ ITU, *Digital Inclusion of All*, 28.

²⁸ Cathy Roche, P. J. Wall, and Dave Lewis, “Ethics and Diversity in Artificial Intelligence Policies, Strategies and Initiatives,” *AI and Ethics* 3, no. 4 (November 2023): 1105, <https://doi.org/10.1007/s43681-022-00218-9>.

²⁹ Machinya, “Blog Series on Exploring the Intersections.”

³⁰ European Commission, Proposal for a Regulation (EU AI Act), art 29.

Regarding RRI as a Foundation for Global Standards, to counter this, its principles advocate for harmonized global norms akin to the Paris Agreement, ensuring AI development adheres to baseline ethical thresholds. The Montreal Declaration for Responsible AI (2018), influenced by RRI's inclusivity and reflexivity pillars, proposes binding requirements for algorithmic fairness audits and diverse dataset curation, measures that could prevent sovereignty from being weaponized to justify ethical compromises. Similarly, the WHO's Global Initiative on AI for Health leverages RRI's precautionary principle to standardize consent protocols for cross-border health data, ensuring marginalized communities benefit equitably from AI advancements.³¹

The EU's leadership in pushing for a Global AI Ethics Accord exemplifies this approach. By embedding RRI's four pillars into transnational mechanisms - such as mandatory bias mitigation in public-sector AI - the bloc aims to prevent a "race to the bottom" where nations sacrifice equity for competitive advantage.³² For instance, the accord proposes shared accountability mechanisms, requiring firms like NVIDIA to open-source fairness benchmarks for AI chips exported to LMICs, addressing resource asymmetries.

Such efforts redefine sovereignty not as a barrier but as a stewardship obligation. By aligning national AI strategies with RRI's global equity imperatives, policymakers can ensure the AI race uplifts, rather than undermines, social justice.

Fostering Global Trust and Social Justice via Responsible Innovation

The interplay between national interests and global inclusion poses a central challenge in contemporary tech governance. RRI's capacity to foster multilateral trust lies in its structured emphasis on shared ethical norms and participatory decision-making. For instance, the WHO's pandemic data-sharing protocols, grounded in RRI's precautionary principle, enabled nations to contribute anonymized health data to global repositories while retaining sovereignty over raw datasets.

This approach expedited the tracking of viral variants and promoted global trust by ensuring equitable benefit distribution, such as prioritizing vaccine access for low- and middle-income countries (LMICs) that shared epidemiological data. Similarly, the European Union's AI Act, though primarily region-specific, has catalyzed cross-border cooperation by promoting transparency standards aligned with RRI's principle of reflexivity. These standards have influenced regulatory developments beyond Europe, prompting countries, such as Canada and Brazil, to pursue compatible frameworks.³³

³¹ World Health Organization, *Global Strategy on Digital Health 2020–2025* (Geneva: WHO, 2023), 18.

³² European Commission, *Proposal for a Regulation (EU AI Act)*, art 29.

³³ European Commission, *Proposal for a Regulation (EU AI Act)*, art 29.

Nevertheless, the EU AI Act has not been without criticism. Its stringent compliance requirements risk placing a disproportionate burden on smaller nations and emerging startups, potentially deterring their participation in international AI collaborations and exacerbating existing innovation divides. Despite these challenges, such initiatives highlight RRI's potential to redefine digital sovereignty, transforming it from a barrier to cooperation into a foundation for collaborative governance rooted in reciprocity and shared ethical principles.

For technologically non-dominant nations, RRI also functions as a form of soft power, enabling them to exert normative influence in global digital policy arenas and contribute meaningfully to the shaping of inclusive, equitable innovation ecosystems.

Kenya's eLimu platform, co-developed with rural communities, positions the country as a leader in inclusive EdTech, attracting partnerships with UNESCO and the World Bank. Likewise, India's DIKSHA initiative (a sovereign yet globally interoperable educational infrastructure) has become a model for LMICs seeking to balance data localization with access to quality content.³⁴ By championing RRI-aligned projects, these nations carve niches as ethical innovators, challenging the dominance of tech superpowers while advocating for equitable norms in such forums as the UN's Global Digital Compact.

However, challenges persist. Dominant players often resist RRI's distributive justice demands, as seen in the slow adoption of the Montreal Declaration's fairness audits by U.S.-based AI firms.³⁵ Resource constraints in LMICs further complicate implementation, e.g., Ethiopia's inability to replicate Kenya's eLimu success due to funding gaps.³⁶ Yet, the reflexive nature of RRI allows iterative adaptation, as demonstrated by South Africa's revised genomic data policies post-pandemic, which integrated stakeholder feedback to address sovereignty concerns without sacrificing inclusion.³⁷

RRI reshapes global competition in AI, encouraging collaborative approaches instead of one-sided technological dominance. By embedding ethics as a shared priority, RRI empowers non-dominant nations to reshape tech governance while compelling dominant actors to reconcile innovation with equity, which is a critical step toward dismantling systemic barriers to equality in the digital age.

The unchecked deployment of technologies risks cementing societal hierarchies unless explicitly designed to prioritize the public good over commercial or geopolitical interests. AI systems, trained on historically biased datasets, often replicate

³⁴ ITU, *Digital Inclusion of All*, 28.

³⁵ European Commission, *Proposal for a Regulation (EU AI Act)*, art 29.

³⁶ UNESCO, *Global Education Monitoring Report 2023*, 99.

³⁷ WHO, *Ethical Guidelines for Health Data Sharing*, 21.

and amplify discrimination, which is a phenomenon starkly evident in education. For instance, algorithmic admissions tools used in U.S. universities, which remarkably favor applicants from affluent school districts, have reinforced racial and socio-economic disparities in access to higher education.³⁸ Similarly, India's early AI-driven tutoring platforms, trained on urban student data, failed to accommodate rural dialects, widening performance gaps by 22%.³⁹ These outcomes underscore how technologies, when divorced from social justice imperatives, become tools of exclusion rather than empowerment.

RRI's principles counter this by mandating bias mitigation as a design prerequisite. This aligns with Rawls' difference principle, where justice requires technologies to maximally benefit the least advantaged, ensuring AI development prioritizes equity over profit. The EU's AI Act (2023), for example, prohibits high-risk educational AI systems unless developers demonstrate proactive steps to eliminate discriminatory outcomes, which acts as a direct application of RRI's precautionary principle. Kenya's eLimu platform operationalizes this by crowdsourcing content from rural educators to ensure cultural and linguistic relevance, reducing urban-rural learning disparities by 41%.⁴⁰ Such systems reject the notion of "neutral" technology, instead embedding equity into the innovation lifecycle.

RRI expands the concept of public good beyond simple access, emphasizing agency and representation. For instance, South Africa's updated telemedicine guidelines, co-developed with township healthcare workers, demonstrate how marginalized communities can influence AI tools to meet local needs, such as prioritizing maternal health alerts in areas with elevated maternal mortality rates.⁴¹ This approach resonates with Carol Gilligan's ethics of care, emphasizing context-driven solutions rooted in empathy rather than abstract universalism. Conversely, Ethiopia's exclusionary EdTech bans, which lacked participatory input, exacerbated gender disparities in rural education.⁴²

To dismantle hierarchies, technologies must democratize benefit-sharing. The WHO's equitable vaccine distribution infrastructure during COVID-19, which reserved doses for LMICs contributing genomic data, exemplifies RRI's reflexivity in action, rewarding collaboration rather than extraction.⁴³ Similarly, open-source initiatives like India's DIKSHA ensure that educational tools remain adaptable to marginalized groups, resisting privatization trends that commodify access.

³⁸ Link Health, Bridging the Digital Divide in Rural America.

³⁹ UNESCO, Global Education Monitoring Report 2023, 92.

⁴⁰ ITU, *Digital Inclusion of All*, 28.

⁴¹ WHO, Ethical Guidelines for Health Data Sharing, 20.

⁴² UNESCO, Global Education Monitoring Report 2023, 33.

⁴³ WHO, Ethical Guidelines for Health Data Sharing, 22.

In sum, social justice is not an ancillary concern but the core metric of ethical innovation. By institutionalizing RRI's pillars, policymakers can transform technologies from vectors of inequality into tools for empowerment.

Conclusion

At its core, RRI champions an innovation model that mandates ethical review and stakeholder engagement. This approach, by positioning ethics as fundamental to innovation, provides a practical means to reconcile the imperatives of digital inclusion and sovereignty. Through case studies in education and healthcare, this study demonstrates that RRI's principles transform geopolitical tensions into opportunities for equitable governance. For instance, Kenya's eLimu and India's DIKSHA platforms illustrate how co-developing technologies with marginalized communities can align sovereignty with access, while the WHO's ethical data-sharing protocols during COVID-19 prove that global collaboration need not compromise national control.

The institutionalization of RRI principles under the auspices of international bodies like the United Nations is critical to preventing fragmented ethical standards.⁴⁴ The UN's Global Digital Compact, currently under negotiation, presents a pivotal opportunity to codify RRI's pillars—such as inclusivity and distributive justice—as universal norms. For example, integrating RRI into the Compact's provisions on AI governance could mandate transparency in algorithmic training data, ensuring technologies like diagnostic tools do not perpetuate racial biases.⁴⁵ Similarly, UNESCO's Recommendation on the Ethics of AI, which already echoes RRI's reflexivity principle, could evolve into a binding framework requiring member states to audit AI systems for equity impacts.⁴⁶

Nations must urgently embed RRI into their digitalization agendas, particularly in education and healthcare, where exclusion risks are highest. In education, this means adopting Kenya's model of participatory EdTech design, where rural communities co-create curricula to ensure cultural relevance and sovereignty compliance. In healthcare, India's National Digital Health Mission demonstrates how federated learning systems can balance data localization with global research needs, preventing LMICs from being relegated to "data colonies".⁴⁷ Policymakers should allocate resources for RRI-aligned capacity-building initiatives, including digital

⁴⁴ United Nations, *Our Common Agenda: Report of the Secretary-General (A/75/982)* (New York: United Nations, 2023), 37.

⁴⁵ WHO, Ethical Guidelines for Health Data Sharing, 16.

⁴⁶ UNESCO, Recommendation on the Ethics of Artificial Intelligence (Paris: UNESCO, 2021), sec. III.4.

⁴⁷ ITU, *Digital Inclusion of All*, 28.

literacy programs for refugees and bias-mitigation training for AI developers, to address and dismantle systemic barriers.

The AI race necessitates a paradigm shift: sovereignty should not justify exclusion, nor should inclusion compromise autonomy. Integrating RRI principles at both global and local levels can help ensure technologies promote equity rather than deepen divides. The RRI framework's commitment to the values of equality and non-exclusivity offers a novel perspective on established principles governing digital and intelligent technologies, such as controllability, responsibility, and reliability.⁴⁸

Considering the philosophical underpinnings of AI ethics, the regulatory shifts fostered by the RRI framework's broader adoption highlight the challenge of balancing autonomy and inclusion when developing ethical reasoning. The approach of dynamic value coordination, which considers moral reasoning as a parallel satisfaction of currently existing norms and restrictions, seems very promising in this sense: to put it simply, an intelligent system or a separate technology in this approach is considered as an agent that is in a situation of having to balance between different moral modules.⁴⁹ The stage of forming ethical modules and incorporating the steps necessary for their coordination and simultaneous implementation necessarily presupposes the simultaneous consideration of local value systems and large-scale principles of social justice. In addition, the approach of dynamic coordination of values provides a more complete disclosure of the question posed by L. Floridi about the possibilities of solving serious social problems by organizing cooperation between humans and (appropriately adapted) artificial agents.⁵⁰

Trends in the ethical regulation of artificial intelligence technologies can be characterized as a consequence of the growing complexity inherent in both the AI systems themselves and their functional environments.⁵¹ At the same time, complexity is revealed in many aspects: on the one hand, researchers note the complication of systems of norms and regulations, motivated by both socio-political and technical factors.⁵² On the other hand, a complex transformation of attitudes to

⁴⁸ Mariarosaria Taddeo et al., "Ethical Principles for Artificial Intelligence in National Defence," *Philosophy & Technology* 34, no. 4 (December 2021): 1718, <https://doi.org/10.1007/s13347-021-00482-3>.

⁴⁹ Linus Ta-Lun Huang, Gleb Papyshev, and James K. Wong, "Democratizing Value Alignment: From Authoritarian to Democratic AI Ethics," *AI and Ethics* 5, no. 1 (2025): 15, <https://doi.org/10.1007/s43681-024-00492-9>.

⁵⁰ Luciano Floridi, "AI as Agency without Intelligence: On Artificial Intelligence as a New Form of Artificial Agency and the Multiple Realisability of Agency Thesis," *Philosophy & Technology* 38 (2025): 35, <https://doi.org/10.1007/s13347-024-00725-z>.

⁵¹ Alexander V. Dumov and Viacheslav I. Kudashov, "Informatization and Digitalization: Complexity Approach to Assessing the Transformation of Education," *Professional Education in the Modern World* 9, no. 4 (2019): 3181.

⁵² Alexandra Kulikova, "Cyber Norms: Technical Extensions and Technological Challenges," *Journal of Cyber Policy* 6, no. 3 (September 2, 2021): 348, <https://doi.org/10.1080/23738871.2021.2020316>.

wards artificial agents is recorded, during which the methods of understanding an intelligent system as an instrumental supplement and as an agent cooperating with a person are complemented.⁵³ The epistemic and ethical guidelines suggested by the RRI framework can rightly be considered as reliable guidelines for the sustainable development of the practice of using intelligent technologies in the era of their rapid increase in complexity.

Conflict of Interests

The author declares no ethical issues or conflicts of interest in this research.

Ethical Standards

The author affirms this research did not involve human subjects.

Bibliography

African Union Commission, *US\$100 Million Africa Pathogen Genomics Initiative to Boost Disease Surveillance and Emergency Response Capacity in Africa* (Addis Ababa: African Union Commission, 2020), <https://au.int/fr/node/39401>.

Dumov, Alexander V., and Viacheslav I. Kudashov. "Informatization and Digitalization: Complexity Approach to Assessing the Transformation of Education" [Информатизация и цифровизация: комплексный подход к оценке трансформации образования]. *Professional Education in the Modern World* 9, no. 4 (2019): 3176–86.

European Commission. *Proposal for a Regulation on Artificial Intelligence (EU AI Act)*. COM (2021) 206 final. Brussels: European Commission, 2021. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>.

European Parliament. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)*. Official Journal of the European Union, L 119 (May 4, 2016): 1–88. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.

Evans, Katie D., Scott A. Robbins, and Joanna J. Bryson. "Do We Collaborate with What We Design?" *Topics in Cognitive Science* 17, no. 2 (April 2025): 392–411. <https://doi.org/10.1111/tops.12682>.

Floridi, Luciano. "AI as Agency without Intelligence: On Artificial Intelligence as a New Form of Artificial Agency and the Multiple Realisability of Agency Thesis." *Philosophy & Technology* 38 (February 2025). <https://doi.org/10.1007/s13347-025-00858-9>.

Fratini, Samuele, Emmie Hine, Claudio Novelli, Huw Roberts, and Luciano Floridi. "Digital Sovereignty: A Descriptive Analysis and a Critical Evaluation of Existing Models." *Digital Society* 3, no. 3 (December 2024): 59. <https://doi.org/10.1007/s44206-024-00146-7>.

⁵³ Katie D. Evans, Scott A. Robbins, and Joanna J. Bryson, "Do We Collaborate With What We Design?," *Topics in Cognitive Science* 17, no. 2 (April 2025): 12, <https://doi.org/10.1111/tops.12682>.

Huang, Linus Ta-Lun, Gleb Papyshev, and James K. Wong. "Democratizing Value Alignment: From Authoritarian to Democratic AI Ethics." *AI and Ethics* 5, no. 1 (February 2025): 11–18. <https://doi.org/10.1007/s43681-024-00624-1>.

International Telecommunication Union (ITU). *Digital Inclusion of All*. Geneva: International Telecommunication Union, 2023. <https://www.itu.int/en/mediacentre/-/backgrounders/Pages/digital-inclusion-of-all.aspx>.

Kulikova, Alexandra. "Cyber Norms: Technical Extensions and Technological Challenges." *Journal of Cyber Policy* 6, no. 3 (September 2, 2021): 340–59. <https://doi.org/10.1080/23738871.2021.2020316>.

Link Health. *Bridging the Digital Divide in Rural America*. Washington, DC: Link Health, 2024. <https://link-health.org/2024/07/16/bridging-the-digital-divide-in-rural-america/>.

Machinya, Johannes. "Blog Series on Exploring the Intersections of Technology, Health, and Law: Data Bias and the Risk of Algorithmic Apartheid in South African Healthcare."

Ministry of Education, India. *DIKSHA Platform: Annual Report*. New Delhi: Government of India, 2023.

Roche, Cathy, P. J. Wall, and Dave Lewis. "Ethics and Diversity in Artificial Intelligence Policies, Strategies and Initiatives." *AI and Ethics* 3, no. 4 (November 2023): 1095–1115. <https://doi.org/10.1007/s43681-022-00218-9>.

SLSA Blog, September 28, 2023. <https://slsablog.co.uk/blog/blog-posts/blog-series-on-exploring-the-intersections-of-technology-health-and-law-data-bias-and-the-risk-of-algorithmic-apartheid-in-south-african-healthcare/>.

Stilgoe, Jack. "Why Responsible Innovation?" In *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, edited by Richard Owen, John Bessant, and Maggy Heintz, 51–74. Chichester, UK: Wiley, 2013.

Taddeo, Mariarosaria, Daniel McNeish, Alexander Blanchard, and Edward Edgar. "Ethical Principles for Artificial Intelligence in National Defence." *Philosophy & Technology* 34, no. 4 (December 2021): 1707–29. <https://doi.org/10.1007/s13347-021-00482-3>.

UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO, 2021. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>.

UNESCO. *Global Education Monitoring Report 2023: Technology in Education: A Tool on Whose Terms?* Paris: UNESCO, 2023. <https://unesdoc.unesco.org/ark:/48223-pf0000385723>.

United Nations (UN). *Our Common Agenda: Report of the Secretary-General*. A/75/982. New York: United Nations, 2021. <https://www.un.org/en/common-agenda>.

World Health Organization (WHO). *Ethical Guidelines for Health Data Sharing in Pandemic Response*. Geneva: WHO, 2022. <https://www.who.int/publications/item/9789240055747>.

World Health Organization. *Global Strategy on Digital Health 2020–2025*. Geneva: World Health Organization, 2021. <https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf>.