ICRIER PROSUS CENTRE FOR INTERNET AND DIGITAL ECONOMY

ETHICS AND GOVERNANCE OF AI
A SYNTHESIS REPORT

Reflections and Policy Insights from a Pre-Summit Webinar Series

FEBRUARY 2026

Ethics and Governance of AI
A Synthesis Report

Reflections and Policy Insights from a Pre-Summit Webinar Series

Author

Shiva Kanwar

Table of Contents

Abstract

This synthesis report examines critical dimensions of artificial intelligence (AI) governance, drawing on insights from a four-part webinar series convened by The ICRIER Prosus Centre for Internet and Digital Economy (IPCIDE) from October 2025 to January 2026. Each webinar featured leading international experts and Indian discussants, creating structured dialogue between global regulatory innovations and India's institutional contexts. The webinars explored the regulation of AI value chains; environmental and social sustainability throughout AI lifecycles; perspectives of AI researchers on AI, public engagement, and responsibility; and the evolution from digital to AI sovereignty in BRICS countries.

The discussions reveal cross-cutting themes in AI governance, including asymmetries in information and bargaining power, tensions between claims of technological universality and local realities, and concentrations of power in AI ecosystems. This report synthesises these discussions to identify opportunities for India's strategic positioning in global AI governance, emphasising digital public infrastructure principles as alternatives to concentrated private control, embedding sustainability across AI development, building institutional capacity for adaptive governance, and operationalising meaningful public participation in AI decision-making. With India assuming leadership in AI governance through the AI Impact Summit 2026 and the 2026 BRICS presidency, this report aims to contribute to deliberations on building robust AI governance frameworks aligned with India's developmental priorities and global positioning.

Acknowledgements: We are grateful to all keynote speakers, discussants, and attendees of the *IPCIDE Webinar Series on Ethics and Governance of AI* held monthly from October 2025 to January 2026. We would like to thank Payal Malik for her support and feedback, and the rest of our colleagues at ICRIER for their support during the webinars and the preparation of this brief. We would also like to acknowledge Raj Kumar for IT support during the webinars, Rajesh Chaudhary for report design, and Aswathy Gopinath for her editorial assistance. Finally, we would like to acknowledge the ICRIER Prosus Centre for Internet and Digital Economy for the research grant funding for this project.

Author email: skanwar@icrier.res.in

Keywords: AI governance; AI sustainability; AI and sovereignty; Policy insights

Ethics and Governance of AI
A Synthesis Report
Reflections and Policy Insights from a Pre-Summit Webinar Series

Shiva Kanwar

1. Introduction

The proliferation of AI in critical infrastructures, economic processes, and everyday life has raised questions about governance, ethical deployment, and societal implications and amplified the need for robust governance frameworks. However, global AI governance narratives are often dominated by perspectives from developed economies, thus reflecting regulatory priorities, institutional capacities, and socioeconomic contexts that diverge significantly from those of developing nations. This asymmetry has created a critical gap where governance frameworks, while sophisticated in their technical and legal architectures, frequently fail to address the distinct developmental imperatives, resource constraints, and democratic aspirations that characterise emerging economies like India.

India's positioning within the global AI landscape has evolved considerably. As one of the world's largest digital economies and a leading voice among developing nations, India confronts the dual challenge of harnessing AI's transformative potential while ensuring that its deployment aligns with the principles of equity, sustainability, and democratic accountability. The India AI Impact Summit, scheduled to be held in February 2026 in New Delhi, is a pivotal opportunity to propose India's strategic vision for AI governance and to contribute meaningfully to global conversations that have, until now, had limited perspectives from the Global South.

The webinar series on the ethics and governance of AI was conceptualised to foreground ethical and governance considerations as constitutive elements of the AI development trajectory and promote critical discussion on AI governance within India's research and policy communities. The intent was to create a structured space for sustained, analytical engagement with questions that are often acknowledged rhetorically but insufficiently examined in policy design and institutional practice. Each webinar featured a leading international expert whose research addresses critical dimensions of AI ethics and governance, alongside distinguished Indian discussants from academia, civil society, and policy research. This mix of participants avoided both uncritical importation of external models and insular domestic debates. International perspectives enabled comparative learning and exposure to regulatory and conceptual innovations, while domestic reflections ensured that discussions remained attentive to questions of feasibility, legitimacy, and local impact within India. The interaction between these perspectives was intended to surface convergence as well as tensions, thereby enriching the analytical depth of the conversations.

The series comprised four webinars, each addressing a distinct yet interconnected dimension of AI governance. The first webinar examined responsibility allocation within general-purpose AI value chains, analysing the European Union's (EU) AI Act and its implications for jurisdictions developing their own regulatory frameworks. The keynote speaker was Dr Martin Ebers, President of the Robotics & AI Law Society (RAILS), Germany and Professor of IT Law at the University of Tartu, Estonia, with Isha Suri, European AI Society Fund AI and Market Power Fellow, and Dr Abha Yadav, Assistant Professor of Law at National Law University Delhi, as discussants.

The second webinar investigated sustainability considerations across the AI lifecycle, exploring the environmental and resource implications of AI infrastructures, with Dr Friederike Rohde from the Berlin Ethics Lab for Responsible AI and Responsible Human-Machine-Interaction at Technische Universität Berlin as the keynote speaker and Prof. (Dr) Ayushi Srivastava, Lecturer at the Jindal School of Environment & Sustainability, O.P. Jindal Global University, as the discussant.

The third webinar presented findings from a comprehensive survey of AI researchers, examining how the visions and values of those developing AI systems can shape technological trajectories. Dr Jack Stilgoe, Professor of Science and Technology Policy at the University College London, was the keynote speaker, with Dr Anulekha Nandi, an independent researcher, and Sadhana Sanjay, Research and Policy Engagement Lead at IT for Change, as discussants.

The fourth webinar explored the concept of AI sovereignty, drawing lessons from BRICS countries' experiences in navigating digital and AI governance while maintaining strategic autonomy. Prof. Luca Belli, Professor at FGV Law School, Rio, and Director, CyberBRICS and Center for Technology and Society, FGV, was the keynote speaker, with Smriti Parsheera, a lawyer and public policy researcher, as the discussant. All webinars were moderated by Shiva Kanwar, Fellow at the ICRIER Prosus Centre and ICRIER.

The webinar series sought to demonstrate that AI governance cannot be adequately addressed through frameworks developed in isolation from the contexts in which AI systems will be deployed. The discussions that emerged, documented in the chapters that follow, reveal both the challenges inherent in governing AI and the context-specific considerations that must inform an effective governance framework. This report concludes with overarching reflections that synthesise cross-cutting themes across the discussions, identify persistent frictions requiring sustained deliberation, and considerations for India's strategic positioning in global AI governance.

2. Webinar Insights: Regulating the AI Value Chain in the EU and Beyond: Upstream, Downstream, or in Between

2.1 Introduction

The emergence of general-purpose artificial intelligence (GPAI)[1] has presented unprecedented challenges for legal frameworks worldwide. Unlike conventional digital products with defined purposes, linear supply chains, and predictable use cases, GPAI models operate within fluid, dynamic value chains characterised by multiple actors, co-creation relationships, and overlapping regulatory domains. The EU's Artificial Intelligence Act (AI Act), 2024,[2] represents the first comprehensive attempt to address these complexities, establishing a regulatory architecture that distributes obligations across various stakeholders in the AI value chain. This webinar examined how the EU's AI Act allocates responsibility among upstream model providers, downstream system developers, and intermediate actors. It also explored the implications of this approach for jurisdictions like India that are developing their own AI governance frameworks and must navigate similar regulatory challenges while confronting distinct institutional capacities, market structures, and developmental imperatives.

The discussion revealed fundamental tensions in allocating responsibility for AI safety and compliance. Dr Ebers's research demonstrates that the AI Act places substantial regulatory burdens on providers of GPAI models, particularly those classified as presenting "systemic risks", while imposing limited obligations on downstream providers who integrate these models into specific applications.[3] This asymmetric distribution of responsibility creates regulatory gaps, particularly for GPAI systems that do not qualify as "high-risk" under the Act's risk-based framework. The discussants contextualised these challenges within India's regulatory landscape, highlighting how India's piecemeal approach to AI regulation through amendments to existing legislation contrasts with the EU's comprehensive framework.

2.2 Understanding the GPAI "Value Chain"

The non-linear and fluid nature of the GPAI value chain represents a departure from traditional supply-chain models. Rather than following a sequential progression from raw materials to final distribution, it embodies a network of co-creation relationships where resources, capabilities, and risks move in multiple directions.[4] The use of the term "value chain", as opposed to "supply chain", underscores the complexity of GPAI development,

---

[1] GPAI refers to AI models designed to perform a broad range of tasks across multiple domains rather than being purpose-built for specific applications. These models, also known as foundation models in technical literature, serve as the basis for numerous downstream applications through finetuning or direct integration.

[2] The EU AI Act (Regulation 2024/1689). https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng#cit_1

[3] Ebers, M., & Penagos, E.V. (2025). Upstream, downstream, and in between: Navigating the GPAI value chain under EU law. http://dx.doi.org/10.2139/ssrn.5640910

[4] Ebers, M., & Penagos, E.V. (2025). Upstream, downstream, and in between: Navigating the GPAI value chain under EU law. http://dx.doi.org/10.2139/ssrn.5640910

highlighting that value and potential harm can arise from interactions among diverse stakeholders.[5]

At the core of this value chain are various tiers of stakeholders, each playing a crucial role in technological outcomes. Compute and cloud providers establish the foundational infrastructure necessary for model training and operation, data providers supply essential datasets for model development, and GPAI model providers create foundational models that serve as the basis for several downstream applications. Subsequent modifications by domain-specific developers refine these models for particular uses, such as legal applications or customer service automation. Finally, app developers and service providers integrate these models into user-facing systems, with distribution platforms, including social media companies designated as very large online platforms (VLOPs),[6] embedding GPAI capabilities within their services.

Within the GPAI value chain, individual actors can occupy multiple simultaneous roles. Modifications can occur at any stage, potentially undermining upstream safeguards. Additionally, information asymmetries are prevalent as upstream providers possess detailed knowledge of model architectures but have limited insight into downstream deployment contexts, while downstream actors understand application-specific risks but lack visibility into foundational model characteristics.[7] This asymmetry is exacerbated by market concentration, allowing powerful actors to dictate contractual terms that limit liability and stifle downstream innovation.

These challenges are magnified in the Indian context. Technology developers often depend on a small number of foreign foundation model providers, which constrains their technical autonomy and regulatory influence.[8] Many lack insight into data provenance, struggle to verify training data authenticity, and have limited capacity for auditing model behaviour. Unlike larger technology firms with robust trust and safety teams, smaller providers operate under significant resource constraints, making comprehensive safety protocols economically unfeasible. This disparity between regulatory expectations, which are often modelled on resource-rich enterprises, and the capabilities of developers in emerging economies creates a tension between ambitious governance frameworks and their practical implementation.

---

5 ICRIER. (2025, October 28). *Regulating the AI value chain in the EU and beyond: Upstream, downstream, or in between* [Video]. YouTube. https://www.youtube.com/live/KiOLhhJZcFc?si=DltcxwLLXkUv8sIz The deliberate use of the term "value chain" was to capture ethical and psychosocial dimensions often lost when technology is reduced to "supply chain" frameworks.

6 Under the EU Digital Services Act, 2022, VLOPs are platforms with more than 45 million monthly active users in the EU and are subject to enhanced regulatory obligations.

7 Ebers, M., & Penagos, E.V. (2025). Upstream, downstream, and in between: Navigating the GPAI value chain under EU law. http://dx.doi.org/10.2139/ssrn.5640910

8 ICRIER. (2025, October 28). *Regulating the AI value chain in the EU and beyond: Upstream, downstream, or in between* [Video]. YouTube. https://www.youtube.com/live/KiOLhhJZcFc?si=DltcxwLLXkUv8sIz

2.3 The EU's Regulatory Response: Upstream Focus and Downstream Gaps

The EU's AI Act adopts a risk-based approach that categorises AI applications into various tiers, from unacceptable to minimal risk. Initially, this framework focused on systems rather than individual models, presuming that risk assessments could be effectively conducted based on the intended purpose and deployment context of the technology.[9] However, this presumption was challenged by the general-purpose nature of many AI systems, which would have led to significant regulatory inadequacies. In response, in the final legislative stages of the Act, specific obligations for GPAI model providers were incorporated. These obligations include the preparation of technical documentation, the provision of information for downstream compliance, the implementation of copyright policies, and the publication of training data summaries. Such requirements aim to enhance transparency and facilitate information flow under the assumption that downstream providers require this information to assess and mitigate deployment risks.

Further obligations arise for GPAI models identified as posing "systemic risks". These obligations include standardised model evaluations, risk assessments, incident tracking, and cybersecurity measures. The General-Purpose AI Code of Practice ("Code of Practice")[10] seeks to operationalise these obligations through commitments related to transparency, copyright, and safety. Although the Code of Practice is ostensibly voluntary, it can be argued that it becomes mandatory for companies seeking regulatory certainty, particularly through its model documentation forms in the transparency chapter, which detail essential model properties and processes.

Despite these advancements, critical gaps remain in the regulatory framework. The AI Act places substantial obligations on model providers while imposing minimal requirements on GPAI systems, which often escape rigorous scrutiny. A notable example is Character AI, a service that was implicated in an incident involving a minor,[11] underscoring the regulatory blind spot concerning general-purpose systems. Thus, while the underlying model might qualify for systemic risk obligations, the system itself, being general-purpose rather than tied to a specific high-risk domain, escapes meaningful regulatory scrutiny beyond basic transparency requirements.

This regulatory asymmetry highlights a misalignment between the concentration of regulatory burdens and the actual emergence of harms. Upstream providers are tasked with conducting risk assessments despite lacking insight into downstream deployment contexts, hindering their ability to evaluate potential risks such as self-harm or discriminatory practices. Conversely, downstream providers, who possess the necessary contextual knowledge for

---

[9] ICRIER. (2025, October 28). *Regulating the AI value chain in the EU and beyond: Upstream, downstream, or in between* [Video]. YouTube. https://www.youtube.com/live/KiOLhhJZcFc?si=DltcxwLLXkUv8sIz

[10] European Commission. (2025, July 10). *The general-purpose AI code of practice*. https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai

[11] Montgomery, B. (2024, October 23). Mother says AI chatbot led her son to kill himself in lawsuit against its maker. *The Guardian*. https://www.theguardian.com/technology/2024/oct/23/character-ai-chatbot-sewell-setzer-death

effective risk assessment, face minimal obligations unless they exceed high-risk thresholds. This is identified as a fundamental misalignment between where obligations concentrate and where harms materialise.[12] While the AI Act acknowledges the interdependence of these parties through cooperation provisions, it falls short of specifying mechanisms to bridge this critical information gap.

As India contemplates its own approach to AI regulation, rather than pursuing comprehensive regulation across the entire AI landscape, which may be beyond current institutional capacities, India could benefit from concentrating its regulatory efforts on the highest-risk models and applications. Establishing clear thresholds for regulatory applicability, based on factors such as computational scale, deployment context, or inherent risk characteristics, would enable the allocation of resources to areas where the stakes are highest. Nevertheless, any regulatory approach must also address the critical need for cooperation and information flow among stakeholders, as downstream actors require upstream transparency to effectively assess risks.[13]

2.4 Cooperation Duties and Information Asymmetries

The integration of third-party AI systems into high-risk applications, as mandated by Article 25(4) of the AI Act, requires a structured framework for cooperation between providers and third parties.[14] This includes the establishment of written agreements that delineate essential information, capabilities, technical access, and assistance. The Code of Practice reinforces the significance of cooperation for effective systemic risk management. However, critical questions remain unaddressed, particularly regarding the extent of information disclosure required from upstream providers beyond mere technical documentation, the rights of downstream providers to audit models or contest upstream risk assessments, and the allocation of liability in light of complex interactions within the value chain.[15]

Market concentration exacerbates these uncertainties, as a limited number of foundation model providers dominate the ecosystem, exerting disproportionate bargaining power in the negotiation of cooperation terms. Consequently, their contractual terms can dictate the flow of information, permissible modifications, and liability distribution. Smaller downstream developers, often with no viable alternatives, may be compelled to accept terms that not only shield upstream providers from accountability but also impose compliance burdens that exceed their capabilities.[16]

---

[12] Ebers, M., & Penagos, E.V. (2025). Upstream, downstream, and in between: Navigating the GPAI value chain under EU law. http://dx.doi.org/10.2139/ssrn.5640910

[13] ICRIER. (2025, October 28). *Regulating the AI value chain in the EU and beyond: Upstream, downstream, or in between* [Video]. YouTube. https://www.youtube.com/live/KiOLhhJZcFc?si=DltcxwLLXkUv8sIz

[14] The EU AI Act (Regulation 2024/1689). https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng#cit_1

[15] Ebers, M., & Penagos, E.V. (2025). Upstream, downstream, and in between: Navigating the GPAI value chain under EU law. http://dx.doi.org/10.2139/ssrn.5640910

[16] ICRIER. (2025, October 28). *Regulating the AI value chain in the EU and beyond: Upstream, downstream, or in between* [Video]. YouTube. https://www.youtube.com/live/KiOLhhJZcFc?si=DltcxwLLXkUv8sIz

These challenges are further compounded in contexts where regulatory frameworks lag behind technological advancements. For instance, Indian developers negotiating with foreign foundation model providers face not only information asymmetries but also jurisdictional complexities. In instances of harm, questions arise regarding which regulatory authority holds jurisdiction and the legal framework applicable to dispute resolution. The lack of international cooperation mechanisms also exacerbates these challenges, facilitating forum shopping and creating opportunities for regulatory arbitrage that ultimately undermine accountability within the AI ecosystem.[17]

2.5 Implications for India's AI Governance Trajectory

India's trajectory in AI governance reflects a complex interplay of innovation facilitation, investment attraction, and harm mitigation, characterised by an incremental rather than a comprehensive regulatory approach. The postponement of the ambitious Digital India Act in favour of targeted legislative amendments illustrates this piecemeal strategy. Recent initiatives, such as the draft amendments to the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules mandating the labelling of synthetic content in 2025,[18] and the India AI Mission's emphasis on building computational infrastructure, developing indigenous foundational models, and creating national datasets, signify a shift towards facilitating development through investment and infrastructure rather than imposing stringent compliance requirements.

However, this regulatory approach reveals significant gaps. While sector-specific interventions regarding AI, such as those in telecommunications, demonstrate progress, overarching issues related to algorithmic accountability, transparency in automated decision-making, and liability for AI-generated harms remain inadequately addressed. Moreover, the regulatory capacity of Indian institutions faces substantial challenges. Regulators must develop technical literacy and analytical capabilities to effectively engage with rapidly evolving technologies, which requires sustained investment rather than reactive measures. The Competition Commission of India's market study on AI[19] exemplifies the sophisticated analysis required; however, translating these insights into actionable enforcement necessitates enhanced coordination among diverse regulatory bodies, i.e., competition authorities, data protection authorities, and sectoral regulators, whose mandates intersect without clear mechanisms for collaborative governance.

The structure of the market imposes constraints on regulatory autonomy, as India's AI development heavily relies on foreign cloud computing infrastructure and foundation models.

---

17 ICRIER. (2025, October 28). *Regulating the AI value chain in the EU and beyond: Upstream, downstream, or in between* [Video]. YouTube. https://www.youtube.com/live/KiOLhhJZcFc?si=DltcxwLLXkUv8sIz

18 The Draft Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2025. https://www.meity.gov.in/static/uploads/2025/10/9de47fb06522b9e40a61e4731bc7de51.pdf; Ministry of Electronics and Information Technology. (2025). Explanatory note for the draft amendments to IT Rules, 2021 in relation to synthetically generated information. https://www.meity.gov.in/static/uploads/2025/10/8e40cdd134cd92dd783a37556428c370.pdf

19 Competition Commission of India. (2025). *Market study on artificial intelligence and competition.* https://www.cci.gov.in/economics-research/market-studies

This dependency raises vulnerabilities because access to critical resources is subject to external corporate policies and geopolitical dynamics, complicating domestic regulatory efforts. To navigate these multifaceted challenges, a shift away from binary framings such as regulation versus innovation and toward recognising necessary trade-offs and designing institutions capable of navigating them is essential. Several potentially productive principles emerged from the discussion:

- Focusing on multi-stakeholder approaches that distribute responsibility across value chains can align regulatory burdens with risk-mitigation capabilities.
- Co-regulatory models that integrate government oversight with industry self-regulation and civil society monitoring could address challenges of opacity.
- Establishing regulatory coordination mechanisms could facilitate information sharing and collaborative policymaking, thereby reducing inconsistencies and compliance burdens in the regulatory landscape.

The integration of labour considerations into AI governance frameworks is critical, yet remains systematically overlooked. The discussion highlighted the psychosocial harms inflicted on workers in the Global South due to the reliance on outsourced content moderation as well as the exploitation of open-source developer labour by commercial AI entities. These issues signify governance failures that are inadequately addressed by conventional consumer protection or competition frameworks.

While initiatives to create public infrastructure in AI present promising alternatives to concentrated private control, they necessitate vigilant accountability mechanisms. A mere transfer of power from private platforms to state-operated systems without accompanying robust safeguards risks perpetuating exclusionary dynamics under public auspices. A fundamental tension also emerged between socially desirable innovation and private profitable innovation, with the concentration of computational resources, training data, and model development among entities prioritising commercial gains over social benefits, creating structural misalignments. Addressing this tension will require not only the pursuit of regulatory constraints but also affirmative strategies that foster alternative development paths, such as publicly funded research, commons-based approaches, and cooperative ownership structures aimed at aligning incentive structures with broader welfare objectives.

2.6 Conclusion

The regulation of GPAI value chains presents multifaceted challenges that defy straightforward resolution. The EU's AI Act, despite being comprehensive, exposes several critical gaps, including the concentration of obligations upstream, complicating risk assessments due to unknown downstream applications, the imposition of minimal requirements on GPAI systems that may result in significant harm, insufficient specification of cooperation mechanisms necessary for overcoming information asymmetries, and the reliance on complementary frameworks with their own limitations.

For India, these challenges intersect with unique institutional realities. The country grapples with nascent regulatory capacities across domains, reliance on foreign infrastructure and models, and competing developmental priorities that emphasise economic growth alongside rights protection. Additionally, ongoing debates regarding regulatory philosophy, such as whether to pursue comprehensive frameworks or incremental adaptations, and centralised versus distributed governance, further complicate the landscape.

Moving forward, a hybrid approach appears necessary. This would involve targeted interventions for high-risk applications alongside horizontal principles applicable across multiple domains. Building capacity within regulatory institutions is crucial for enabling sophisticated analysis rather than mere reactive responses. Furthermore, multi-stakeholder governance models should be adopted to distribute responsibility while ensuring clear accountability. International cooperation mechanisms are also essential for addressing jurisdictional challenges inherent in transnational value chains.

Ultimately, effective governance must confront the underlying political economy questions surrounding AI development, deployment, and governance, particularly regarding power dynamics, equity, and the interests served by technological systems. While technical solutions, such as improved algorithms, enhanced safety protocols, and greater transparency, are vital, they are insufficient without addressing the deeper issues related to market structure, resource distribution, and institutional design. The challenge for India, as with any jurisdiction facing GPAI governance, extends beyond the formulation of appropriate rules, requiring the establishment of institutional ecologies capable of continuous learning, adaptation, and the balancing of competing imperatives as AI capabilities and associated risks evolve.

3. Webinar Insights: Navigating Sustainability Within the AI Lifecycle: Sustainability Criteria and Current Challenges of AI and Related Infrastructures

3.1 Introduction

The integration of AI systems into critical infrastructures raises significant sustainability concerns. While much of the discourse has focused on AI's potential to address and mitigate sustainability challenges, there is a parallel imperative to scrutinise the sustainability of AI systems themselves. This webinar explored the environmental, social, and economic impacts associated with the AI lifecycle, from resource extraction for hardware, to model training and deployment, to disposal. Dr Friederike Rohde introduced the Sustainability Criteria and Indicators for Artificial Intelligence Systems (SCAIS) framework, which is a product of interdisciplinary collaboration among the Institute for Ecological Economy Research, Technische Universität Berlin and AlgorithmWatch, among others. Prof. Ayushi Srivastava provided critical insights into sustainability challenges in the Indian context, highlighting supply-chain vulnerabilities and the precarious balance that developing nations must strike between technological competitiveness and environmental responsibility. The discussion explored how AI systems function as complex socio-techno-ecological systems whose sustainability impacts extend far beyond their immediate operational contexts.

3.2 The Materiality of AI Systems and Their Challenges

AI systems, often perceived as solely virtual, have significant material foundations that require substantial resources and energy. The infrastructure for large language models includes data centres, specialised chips, undersea cables, and electrical grids, and results in high energy consumption, exemplified by GPT-3's generation of approximately 552 tonnes of CO2 equivalent during training.[20] Models with similar sizes can have vastly different carbon footprints based on their energy sources. For instance, the BLOOM model, which uses nuclear energy, had lower emissions than GPT-3.[21] Global electricity consumption by data centres is on the rise, with water consumption projected to increase from 239 billion litres in 2024 to 664 billion litres in 2030 due to cooling needs.[22] AI hardware demands further exacerbate resource usage, with specialised GPUs consuming significantly more power than standard servers. Semiconductor manufacturing for AI also demands significant water resources, with a single AI query potentially using 500 millilitres, exacerbating water shortages in stressed areas and impacting local ecosystems and communities.[23] This creates conflicts over water usage, especially in stressed regions, and raises concerns about distributive justice as data centres are often located in economically disadvantaged areas.[24]

Further, the physical hardware layer underpinning AI systems and reliant on high-performance GPUs is dependent on the extraction of essential materials like cobalt, lithium, and other rare earth elements. This extraction process leads to environmental degradation, including deforestation and soil contamination, while also posing social challenges, exemplified in regions like the Democratic Republic of Congo, where cobalt mining is associated with hazardous working conditions.[25] The growing challenge of electronic waste also raises further concerns about safe disposal and environmental justice. Electronic waste

---

[20] ICRIER. (2025, November 21). *Sustainability criteria and current challenges of AI and related infrastructures* [Video]. YouTube. https://youtu.be/YZgMZL4MmLk

[21] ICRIER. (2025, November 21). *Sustainability criteria and current challenges of AI and related infrastructures* [Video]. YouTube. https://youtu.be/YZgMZL4MmLk

[22] ICRIER. (2025, November 21). *Sustainability criteria and current challenges of AI and related infrastructures* [Video]. YouTube. https://youtu.be/YZgMZL4MmLk

[23] ICRIER. (2025, November 21). *Sustainability criteria and current challenges of AI and related infrastructures* [Video]. YouTube. https://youtu.be/YZgMZL4MmLk

[24] As cited in Rohde, F., Wagner, J., Reinhard, P., Petschow, U., Meyer, A., Voß, M., & Mollen, A. (2024). Broadening the perspective for sustainable artificial intelligence: Sustainability criteria and indicators for artificial intelligence systems. *Current Opinion in Environmental Sustainability*, 66, 101411. https://doi.org/10.1016/j.cosust.2023.101411

[25] ICRIER. (2025, November 21). *Sustainability criteria and current challenges of AI and related infrastructures* [Video]. YouTube. https://youtu.be/YZgMZL4MmLk

from AI infrastructure is expected to reach 43.2 megatonnes by 2030, posing environmental justice challenges as e-waste typically ends up in regions with less regulatory oversight.[26]

Addressing these vulnerabilities requires robust frameworks and regulatory changes, some of which are exemplified under the OECD Due Diligence Guidance for Responsible Supply Chains of Minerals from Conflict-Affected and High-Risk Areas[27] and the EU's Critical Raw Materials Act (CRM Act) 2024.[28]

3.3 The Sustainability Criteria and Indicators for Artificial Intelligence Systems (SCAIS) Framework

The SCAIS Framework represents an effort to operationalise sustainable AI through concrete, assessable criteria spanning the full AI lifecycle. It emerged from recognising that existing approaches to AI and sustainability operated within silos, addressing either social concerns through ethics guidelines or environmental impacts through "green AI" initiatives, rarely integrating these dimensions systematically. The framework is composed of 19 criteria categorised into four primary dimensions: social, ecological, economic, and (organisational) governance.[29] These dimensions are operationalised through 67 measurable indicators, allowing for a robust evaluation of sustainability performance.

The social dimension emphasises justice, cohesion, and the importance of institutional arrangements. It seeks to address issues such as equity in AI access and the potential biases that can arise in algorithmic decision-making. The ecological dimension considers the environmental impact of AI systems, particularly in relation to planetary boundaries. This aspect highlights the necessity of minimising energy consumption and reducing the carbon footprint associated with AI technologies. The economic dimension underscores the importance of responsible decision-making that prioritises the common good over profit maximisation. It encourages organisations to consider the long-term implications of their AI initiatives on economic systems and to promote market diversity. Lastly, the organisational-governance dimension establishes cross-cutting criteria applicable throughout the AI lifecycle and addresses the need for transparency and accountability in AI development. It promotes governance structures that are adaptable and responsive to the rapidly evolving nature of technology. The framework also includes a self-assessment tool for organisations, which enables them to evaluate their sustainability performance and recognises the need for coordinated policy responses to tackle systemic issues effectively.

---

26 ICRIER. (2025, November 21). *Sustainability criteria and current challenges of AI and related infrastructures* [Video]. YouTube. https://youtu.be/YZgMZL4MmLk

27 OECD. (2016). *OECD due diligence guidance for responsible supply chains of minerals from conflict-affected and high-risk areas.* https://www.oecd.org/en/publications/2016/04/oecd-due-diligence-guidance-for-responsible-supply-chains-of-minerals-from-conflict-affected-and-high-risk-areas_g1g65996.html

28 European Union. Regulation (EU) 2024/1252, 2024. Critical Raw Materials Act (CRM Act) 2024. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401252

29 Rohde, F., Wagner, J., Reinhard, P., Petschow, U., Meyer, A., Voß, M., & Mollen, A. (2024). Broadening the perspective for sustainable artificial intelligence: Sustainability criteria and indicators for artificial intelligence systems. *Current Opinion in Environmental Sustainability*, 66, 101411. https://doi.org/10.1016/j.cosust.2023.101411

3.4 Implementation Pathways for Sustainability: From Principles to Practice

Translating sustainability criteria into practice will require multiple intervention strategies operating at different levels.

Technical and Organisational Measures: At the technical level, developers can implement various strategies to enhance the sustainability of AI systems, *inter alia*, emphasising data sparseness to ensure only essential data is collected and curated, prioritising efficiency in model architecture through techniques such as pruning,[30] quantisation,[31] and knowledge distillation,[32] and leveraging transfer learning[33] and pre-trained models instead of training large models from scratch to significantly reduce computational demands and energy consumption. Model cards[34] can also be used to document sustainability metrics, including energy usage and carbon footprint, during development. In fact, several cloud providers now offer tools that enable developers to monitor energy consumption and carbon emissions during model development.[35]

Organisational measures extend beyond individual projects to include company-wide practices, such as the establishment of AI ethics codes with monitoring mechanisms, implementation of stakeholder consultation processes, execution of environmental impact assessments before infrastructure decisions, and collaboration with certified data centres that adhere to energy and water efficiency standards. Integration of sustainability assessments into procurement decisions can also drive demand for more sustainable hardware and services.

Regulatory Approaches and Industry Standards: Policy interventions are essential for establishing baseline sustainability standards and addressing systemic challenges that extend beyond the control of individual organisations. Mandatory reporting requirements can enhance transparency regarding the environmental impacts of AI systems. For example, the EU's Energy Efficiency Directive mandates that data centres report specific metrics and promotes the utilisation of waste heat.[36] Additionally, location regulations for data centres can mitigate the risk of situating facilities in water-scarce regions or areas where vulnerable

30 Pruning involves removing unnecessary connections or neurons from neural networks to reduce computational requirements without significantly impacting performance.

31 Quantisation reduces the precision of numerical representations in models (e.g., from 32-bit to 8-bit numbers), decreasing memory requirements and computational costs.

32 Knowledge distillation is a machine-learning technique that aims to transfer the learnings of a large pre-trained "teacher model" to a smaller "student model".

33 Transfer learning refers to using knowledge gained from training a model on one task to improve performance on a related task, avoiding the need to train large models from scratch.

34 Model cards are documentation frameworks that provide standardised information about machine learning models, including intended use, training data characteristics, performance metrics, and potential biases.

35 ICRIER. (2025, November 21). *Sustainability criteria and current challenges of AI and related infrastructures* [Video]. YouTube. https://youtu.be/YZgMZL4MmLk

36 European Union. Directive (EU) 2023/1791 The Energy Efficiency Directive. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AJOL_2023_231_R_0001&qid=1695186598766.

communities could be disproportionately affected. Environmental impact assessments can also be mandated prior to the approval of significant AI infrastructure projects.

Due diligence frameworks, such as the OECD Due Diligence Guidance for Responsible Supply Chains of Minerals from Conflict-Affected and High-Risk Areas,[37] extend accountability to upstream impacts, including mineral extraction and component manufacturing. Standards and certification initiatives can also establish industry benchmarks and offer clear guidance for organisations seeking sustainability improvements. The EU's CRM Act[38] exemplifies regulatory approaches that can stimulate circular economy practices and decrease reliance on primary extraction. In India, several relevant policy frameworks, including the India AI Mission and the Semiconductor Manufacturing Mission, aim to develop domestic capabilities while addressing sustainability concerns. Nonetheless, substantial implementation challenges persist, particularly regarding the coordination of environmental objectives with imperatives of economic competitiveness.

Research and Collaboration for Capacity Building: Research is crucial for capacity building. However, while enhancing AI system capabilities remains a consistent focus, there has been relatively limited exploration of the environmental, economic, and social dimensions of AI sustainability. Expanding research in these areas can provide a robust evidence base for policy formulation, reveal previously overlooked impacts, and suggest innovative solutions. Further, increased international collaboration in research to share knowledge, methodologies, and best practices, rather than competing solely for AI supremacy, should be encouraged.

3.5 Policy Implications for India

India confronts distinctive challenges in navigating AI sustainability, balancing aspirations for technological advancement with environmental and social responsibilities. The tension between fostering innovation in emerging technologies and implementing sustainability safeguards is particularly acute for developing nations seeking to avoid being left behind in global AI development while simultaneously pursuing ambitious environmental goals. India has demonstrated substantial commitment to environmental sustainability through major initiatives around solar energy deployment, renewable energy expansion, and electric vehicle adoption. The challenge lies in extending this environmental consciousness to AI development without compromising competitiveness.

India's participation in international AI governance discussions presents opportunities for shaping global norms. Drawing parallels to the Paris Agreement's framework of common but differentiated responsibilities, India could leverage the AI Impact Summit 2026 to establish differentiated targets, enabling countries to contribute according to their capacities while

[37] OECD. (2016). *OECD due diligence guidance for responsible supply chains of minerals from conflict-affected and high-risk areas*. OECD Publishing. https://doi.org/10.1787/9789264252479-en

[38] European Union. Regulation (EU) 2024/1252, 2024. Critical Raw Materials Act (CRM Act) 2024. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401252

ensuring collective progress toward sustainability objectives.[39] This approach would acknowledge that developing nations face distinct challenges and should not be held to identical standards as industrialised countries with longer development histories.

Beyond international frameworks, domestic solutions merit emphasis. Rather than continuously developing new giant models, focusing on finetuning existing models for specific applications would prove far less energy-intensive. Awareness and training programs addressing responsible AI sourcing and deployment constitute essential complements to technical and regulatory measures. Critically, India requires expanded research examining AI sustainability across environmental, economic, and social dimensions. While substantial research addresses AI development and applications, systematic investigation of sustainability implications specific to Indian contexts remains limited. Enhanced research capacity would enable evidence-based policymaking and contribute Indian perspectives to global discourse.

3.6 Conclusion

The sustainability challenges inherent in AI systems extend far beyond the technological, encompassing environmental impacts across the full lifecycle, social justice concerns from extraction through deployment, economic implications including market concentration and labour conditions, and governance questions regarding accountability and regulation. Solutions must operate at multiple levels, from technical interventions to regulatory frameworks, international coordination, and expanded research examining sustainability implications in diverse contexts.

Moving forward demands navigating inherent friction between technological advancement and sustainability imperatives, which are particularly acute for developing nations pursuing competitiveness while shouldering disproportionate environmental and social costs. Perhaps most fundamentally, the discourse must grapple with questions of necessity, distribution, and value, i.e., determining which AI applications genuinely serve collective welfare, ensuring equitable distribution of benefits and burdens, and critically examining whether anticipated gains justify environmental and human costs. As AI systems become increasingly embedded in critical infrastructures and everyday life, these questions become increasingly urgent, demanding sustained engagement from researchers, policymakers, industry actors, civil society, and affected communities.

---

[39] ICRIER. (2025, November 21). *Sustainability criteria and current challenges of AI and related infrastructures* [Video]. YouTube. https://youtu.be/YZgMZL4MmLk

4. Webinar Insights: Visions, Values, Voices: A Survey of AI Researchers

4.1 Introduction

As AI permeates critical infrastructures and everyday life, understanding the perspectives of those directly engaged in its development is crucial for responsible governance. This webinar discussed a comprehensive survey of AI researchers conducted by a team at University College London and offered insight into the values, concerns, and visions of the scientific community working at the forefront of AI innovation. The survey, representing the largest social science study of AI researchers to date, revealed a community far more diverse and uncertain than public discourse suggests.[40] The conversation examined how the technical community perceives the risks and benefits of AI, their views on public engagement, and the responsibilities they envision for themselves and others in the AI ecosystem. It extended beyond a presentation of survey findings to engage with fundamental questions about responsibility, inclusion, and the politics of AI development. The discussants brought critical perspectives on how these issues manifest in the Indian context, raising concerns about technical alignment in diverse social settings, distribution of responsibility across AI ecosystems, and the pedagogical foundations necessary for democratising AI governance. Their interventions highlighted the particular challenges facing countries like India, where AI development occurs within complex institutional landscapes and must serve diverse populations with varying needs and vulnerabilities.

4.2 Divergent Visions: Benefits, Risks, and Technological Inevitability of AI

The survey findings indicate that AI researchers are significantly more optimistic about AI's benefits compared to the general public, but both groups share similar concerns regarding its risks. Researchers largely believe that the benefits of AI either outweigh or balance the risks, whereas public opinion is more evenly divided. This optimism gap is particularly evident in areas such as education access, job facilitation, and household task automation, where researchers foresee transformative impacts that the public views with scepticism.[41] Conversely, there is notable consensus on negative impacts, with both groups prioritising concerns about disinformation, unauthorised data use, and cybercrime.

Regarding attitudes toward artificial general intelligence (AGI) and technological determinism, approximately half of the researchers surveyed believe AGI is inevitable, indicating a perception of technological progression beyond human control. This belief correlates with support for rapid AI development, a preference for expert responsibility over

---

[40] O'Donovan, C., Gurakan, S., Wu, X., Stilgoe, J., Bert, N., Dmitrichenko, E., Gorba, E., Liu, S., Zamborsky, T., & Zhao, T. (2025). *Visions, values, voices: A survey of artificial intelligence researchers*. UCL Centre for Responsible Innovation Working Paper 2025/01. https://doi.org/10.5281/zenodo.15080287

[41] O'Donovan, C., Gurakan, S., Wu, X., Stilgoe, J., Bert, N., Dmitrichenko, E., Gorba, E., Liu, S., Zamborsky, T., & Zhao, T. (2025). *Visions, values, voices: A survey of artificial intelligence researchers*. UCL Centre for Responsible Innovation Working Paper 2025/01. https://doi.org/10.5281/zenodo.15080287

public input, and reduced concern about unequal impacts across social groups.[42] However, such views may inhibit democratic deliberation about AI's trajectory as they frame its development as a natural process rather than a series of value-laden choices. Additionally, factors such as gender emerged as a variable, with female researchers more inclined to emphasise the importance of considering AI's effects on various social groups and express concerns about bias, revealing diverse viewpoints shaped by location, identity, and disciplinary orientation.

4.3 The "Deficit Model" and Public Engagement

A significant finding from an AI governance perspective is the widespread adherence among researchers to the "deficit model", which posits that public scepticism arises from a lack of technical knowledge rather than legitimate value differences or experiential insights. A large number of researchers believe that enhanced public understanding of AI would lead to increased trust, viewing the public as lacking awareness of both the benefits and risks associated with AI.[43]

This deficit thinking extends to researchers' perspectives on public involvement in AI development. While there is general support for public engagement in downstream activities such as regulation and impact assessment, there is significantly less endorsement for upstream participation in the development of AI systems or model training. This delineation reflects a division of labour where technical experts make core design decisions, relegating public input to implementation and governance, under the assumption that technical development can progress independently of social values. This dynamic presents a critical challenge for democratic AI governance by perpetuating an asymmetry in valuing different forms of knowledge and concern. Researchers often interpret public hesitance as an information deficit, overlooking opportunities to learn from public experiences and values. The discussion highlighted that concerns of the public frequently stem from awareness of the political economy, including the externalisation of social and economic costs to data workers in the Global South, concentration of computational infrastructure, and capture of research agendas by corporate interests, rather than ignorance of technical details.

4.4 Technical Alignment, Distributed Responsibility, and Industry Influence

The discussion also addressed questions about how AI systems can be aligned with social values and how responsibility should be distributed across complex AI ecosystems. These questions take on particular urgency in the Indian context, where AI deployment occurs across diverse social, linguistic, and cultural settings.

---

[42] ICRIER. (2025, December 19). *Visions, values, voices: A survey of artificial intelligence researchers* [Video]. https://youtu.be/o8eGMGK2S3g

[43] O'Donovan, C., Gurakan, S., Wu, X., Stilgoe, J., Bert, N., Dmitrichenko, E., Gorba, E., Liu, S., Zamborsky, T., & Zhao, T. (2025). *Visions, values, voices: A survey of artificial intelligence researchers*. UCL Centre for Responsible Innovation Working Paper 2025/01. https://doi.org/10.5281/zenodo.15080287

Technical alignment is a major focus within AI research communities.[44] However, current interpretations of alignment often regard values as programmable constraints, overlooking their contested, contextual, and evolving nature.[45] The discussion highlighted tensions between viewing alignment as a technical engineering challenge and as a dynamic social and political process. In organisational information systems, alignment requires continuous iteration in response to changing contexts and emerging information. AI systems, however, are designed for scalability and deployment across diverse contexts, creating friction with context-specific alignment needs. These challenges are particularly pronounced in India due to extreme linguistic diversity, varying legal traditions, differing community norms, and significant disparities in digital literacy and access. This raises critical questions regarding whose values AI systems should embody and how local contexts can influence systems intended for widespread deployment. While certain principles, such as non-discrimination, may be universally applicable, many governance issues require navigating competing priorities, such as content moderation that must reconcile legal obligations, platform governance norms, and individual liberties across diverse cultural and political contexts.

The question of **responsibility for AI systems** is also complex, given the distributed development structure and involvement of multiple actors, *inter alia*, foundation model developers, software engineers building applications on those models, deployers making implementation decisions, and end users. Each layer introduces both capabilities and constraints, making it difficult to trace where responsibility for outcomes should lie, and traditional frameworks for attributing responsibility struggle to accommodate these layered relationships. This distributed responsibility challenge becomes more acute when AI systems are opaque, not just in their internal workings but in how foundational models interact with layered software and interfaces built atop them. Determining responsibility will require unravelling complex threads of causation across technical layers and organisational boundaries. AI governance frameworks must grapple with this complexity while ensuring accountability mechanisms that are clear, enforceable, and fair.

Further, recognition of the significant **influence of industry** on research priorities extends beyond the academic-industry divide. While the recognition is more pronounced among university researchers, it is shared by a substantial minority of industry-employed researchers. The concentration of AI research within large technology corporations shapes not only the focus of research but also how problems are framed and what solutions appear viable. When research agendas are driven by commercial interests, issues that threaten profitability or challenge existing business models are often disregarded. Research on alternatives to data-intensive methods, systems prioritising interpretability over performance maximisation, and equitable power distribution struggle to gain support when corporate

[44] Technical alignment here refers to the practice of ensuring that AI systems behave according to human intentions, values, and ethical standards.

[45] O'Donovan, C., Gurakan, S., Wu, X., Stilgoe, J., Bert, N., Dmitrichenko, E., Gorba, E., Liu, S., Zamborsky, T., & Zhao, T. (2025). *Visions, values, voices: A survey of artificial intelligence researchers*. UCL Centre for Responsible Innovation Working Paper 2025/01. https://doi.org/10.5281/zenodo.15080287

interests dominate.[46] However, as researchers possess both technical expertise and the authority to challenge oversimplified narratives and enrich public discourse, they can use this as an opportunity to "bring out the citizen in the scientist" and engage as democratic participants rather than merely as technical experts. They can highlight uncertainties overlooked by corporate communications, question trajectories constrained by venture capital, and advocate for democratic deliberation against claims of technological inevitability.[47]

4.5 Policy Implications for AI Governance in India

The survey findings and webinar discussion offer several insights relevant to India's evolving AI governance landscape.

The diversity of researcher perspectives challenges monolithic framings of AI development. The technical community holds varied views on development speed, risk priorities, responsibility allocation, and public engagement. Creating governance mechanisms that enable contributions from diverse technical voices rather than privileging a narrow set of perspectives aligned with commercial interests can enrich policy deliberation. The survey finding that most researchers worry about the industry's disproportionate role in setting research agendas suggests potential allies for governance approaches that balance commercial innovation with public interest considerations.

Challenges of technical alignment and the allocation of responsibilities in AI systems require governance structures that consider local contexts while ensuring accountability. AI systems are promoted as general-purpose technologies applicable across contexts, yet their impacts do not always apply uniformly, especially in local contexts, and are unevenly distributed. India's diversity means that AI systems cannot simply import alignment approaches developed for different social contexts. Governance mechanisms must enable ongoing negotiation of how AI systems should behave across different settings while establishing clear responsibility for harms. This may require moving beyond purely technical alignment approaches towards more adaptive, participatory approaches enabling communities to shape how AI systems function in their particular contexts, not through one-time consultations but through ongoing democratic engagement.

Pedagogical dimensions of AI governance deserve greater consideration. Building capacity for democratic AI governance in India requires educational reforms that integrate technical training with critical reflection on values, social impact, and democratic participation. This includes not just training for researchers and developers but also broader public education that enables informed engagement with AI governance questions. India's educational

[46] ICRIER. (2025, December 19). *Visions, values, voices: A survey of artificial intelligence researchers* [Video]. https://youtu.be/o8eGMGK2S3g

[47] ICRIER. (2025, December 19). *Visions, values, voices: A survey of artificial intelligence researchers* [Video]. https://youtu.be/o8eGMGK2S3g

institutions, civil society organisations, and policy bodies all have roles to play in developing this capacity.

AI governance cannot rely solely on science communication and literacy programs to build public trust. Instead, mechanisms must be developed that treat public experiences and concerns as legitimate inputs to governance decisions and not problems to be solved through better education. This requires institutional innovations that create genuine opportunities for upstream public participation in shaping AI development priorities and deployment contexts, drawing on India's own democratic traditions and participatory governance experiments.

4.6 Conclusion

As India advances ambitious AI development initiatives while preparing to host the AI Impact Summit in 2026, questions about whose voices shape AI trajectories, how responsibility is distributed across complex ecosystems, and what values AI systems should reflect have increasingly become crucial. These questions require a fundamental rethinking of how AI development is organised, governed, and integrated into democratic processes. Bringing researchers into democratic dialogue with affected communities, building coalitions around shared concerns about industry capture, creating institutional conditions enabling researchers to prioritise public interest, and developing governance frameworks that distribute power rather than concentrating it, offer pathways towards AI development serving broad public interest. The voices that should matter the most are those of communities whose lives AI systems will shape.

5. Webinar Insights: From Digital Sovereignty to AI Sovereignty: Lessons from BRICS Countries

5.1 Introduction

AI has intensified debates about national sovereignty in the digital age. While regulating AI systems to mitigate risks and harms has been a focal point in policy and regulatory discourse, it is equally important to examine whether nations can shape their own AI futures rather than simply adopting and consuming technologies created elsewhere. This discussion examined the transition from digital to AI sovereignty through the experiences of the BRICS countries, i.e., Brazil, Russia, India, China, and South Africa. Prof. Luca Belli presented a systemic framework for understanding AI sovereignty, drawing on seven years of comparative research mapping digital policies across the BRICS nations. Smriti Parsheera grounded these insights in the Indian context, examining questions of legitimacy, accountability, and citizen empowerment as India pursues technological development through digital public infrastructure (DPI) and prepares to assume leadership in global AI governance discussions. The conversation explored how achieving meaningful sovereignty requires not only regulatory frameworks but integrated approaches encompassing technological development, industrial policy, and citizen-centric governance that extend beyond immediate state interests to serve broader societal goals.

5.2 Conceptualising Digital Sovereignty

The discussion contextualised digital sovereignty as encompassing the capacity to understand, develop, and regulate digital technologies, thereby retaining self-determination, control, and power over such systems rather than being controlled through them by external actors.[48] This framework challenges conventional approaches that privilege the regulation of these technologies over their development. Sovereignty itself can manifest as protectionist control that restricts information flows and freedoms. Alternatively, it can appear to be "good digital sovereignty", i.e., approaches that empower citizens, foster competition, and ensure technologies operate under domestic law rather than on foreign corporate terms. Drawing on social contract theory, where individuals delegate authority to the state for the protection of rights and welfare, sovereignty frameworks should be evaluated not only by their effectiveness in asserting state power but by their capacity to serve and empower citizens.[49] India's 2021 Intermediary Guidelines can illustrate the tension that arises between these concepts. While the guidelines were conceptualised to address legitimate concerns about online harms and jurisdictional challenges with foreign-headquartered platforms, their enforcement has raised questions about free speech and the division between legislative and executive authority.[50]

5.3 The Governance of Data

In 2019, China designated data as a factor of production, on par with capital, labour, and land. This represented a shift with far-reaching policy implications as it reframed data governance from a purely rights-based framework centred on individual informational self-determination to frameworks that acknowledge the collective economic value and strategic importance of data.[51] While the capacity of individuals to control how their personal data is used remains essential, it is insufficient to address systemic dynamics of data-intensive economies by itself. For instance, the EU General Data Protection Regulation (GDPR) establishes robust rights but relies on individuals being aware of these rights and willing to petition authorities and on regulators understanding complex systems while possessing the political will to enforce against powerful interests. These mechanisms frequently falter, especially in developing

48 ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ
49 ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ
50 ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ
51 ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ

economies. Further, solely rights-based frameworks do not address how societies can leverage data as a collective resource for development while respecting individual rights.[52]

China blended regulation with industrial policy to develop a multi-trillion-dollar data industry, while India pursued a different strategy of leveraging software development capabilities to create DPI. India's Unified Payments Interface (UPI) exemplifies this approach as a software-based DPI that has served to challenge the Visa-Mastercard duopoly for payment systems. Brazil's PIX system refined this model by operating as a pure public service through the nation's central bank.[53] These examples demonstrate that sovereignty can be built through state initiatives.

Yet, DPIs raise sovereignty questions of their own. In India, these infrastructures have become central to domestic policy as well as international diplomacy and development cooperation, extending Indian influence globally. This, however, raises concerns about the creation of an "alt big tech" with similar patterns of power concentration. Combating this requires vigilance about accountability, power concentration, and lock-in effects as well as risks of excluding actors who do not conform to dominant frameworks.[54] The tension between building state-supported infrastructures for sovereignty and avoiding recreating monopolistic control structures remains an open challenge.

5.4 Cross-Border Data Flows and Data Localisation

India's journey through various iterations of its data-protection legislation reflects competing imperatives between facilitating the data economy and asserting sovereign control. The AI development argument is often invoked to justify data localisation, arguing that domestic data access is necessary for competitive AI systems. However, this link remains empirically weak as the mere physical location of data on domestic servers does not automatically translate into superior AI capabilities. Without a comprehensive surveillance law, localised data may become more vulnerable to law enforcement access.[55] This affirms the necessity of evaluating sovereignty measures not merely by assertion of state control but by implications for individual rights.

The BRICS countries have also collectively recognised the need for frameworks enabling interoperable data governance. The 2024 Kazan Declaration[56] called for "a fair and equitable global framework for data governance, including cross-border data flows', and to 'ensure(ing) the interoperability of data policy frameworks at all levels". This represents an attempt to

---

52 ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ

53 ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ

54 ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ

55 ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ

56 *Kazan declaration: Strengthening multilateralism for just global development and security*, para. 71. http://static.kremlin.ru/media/events/files/en/RosOySvLzGaJtmx2wYFv0lN4NSPZploG.pdf

articulate data governance alternatives to the United States and EU approaches of unregulated corporate control, as well as regulation without corresponding industrial policy, respectively. Whether this effort can translate into a concrete international instrument remains to be seen.

5.5 From Digital Sovereignty to AI Sovereignty

AI sovereignty extends digital sovereignty by focusing specifically on the capacity to understand, develop, and regulate AI systems, thus retaining self-determination, agency, and control over them.[57] A critical insight highlighted in the discussion was that risk-based regulation for AI, while necessary, is insufficient by itself. For instance, the EU's AI Act establishes important guardrails against high-risk applications, but without indigenous technological alternatives, it risks making EU actors mere consumers of systems developed elsewhere. Effective governance requires both constraining harmful uses and enabling the development of alternative systems that have different values and serve different needs.

The discussion presented the Key AI Sovereignty Enablers (KASE) framework[58] as a systemic approach to AI sovereignty. This approach comprises layered components forming an "AI sovereignty stack",[59] which includes:

- data governance, encompassing not merely personal data protection but frameworks for leveraging data as a collective resource;
- algorithmic governance and investing in research and development, while also addressing potential risks;
- computational capacity, increasingly concentrated in data centres controlled by a few firms;
- meaningful connectivity, allowing users to enjoy reliable, well-performing, universally accessible internet infrastructure for an affordable price;
- reliable electrical power as AI's energy intensity makes access to affordable, sustainable power strategic;
- digitally literate population, enabling them to be producers rather than mere consumers;
- strong cybersecurity, protecting against exploitation; and

---

[57] ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ

[58] Belli, L. (2023). Exploring the key AI sovereignty enablers (KASE) of Brazil, to build an AI sovereignty stack. In *The quest for AI sovereignty, transparency and accountability* (pp. 29–44). https://cyberbrics.info/wp-content/uploads/2023/08/AI-sovereignty-updated-CLEAN.pdf

[59] Belli, L. (2023). Exploring the key AI sovereignty enablers (KASE) of Brazil, to build an AI sovereignty stack. In *The quest for AI sovereignty, transparency and accountability* (pp. 29–44). https://cyberbrics.info/wp-content/uploads/2023/08/AI-sovereignty-updated-CLEAN.pdf

- comprehensive governance framework that encompasses ethical considerations, data-protection laws, and AI regulations.

Each element is critical, and weakness in one dimension can undermine strengths in others. Brazil's experience offers a cautionary example as, despite significant investments in computational capacity and algorithms, 78 per cent of the population lacks meaningful connectivity, i.e., affordable, reliable, sufficient bandwidth for full digital participation, which threatens to render these investments ineffective for the majority.[60]

Further, the prevalence of zero-rating practices in Brazil has resulted in the concentration of user traffic on foreign-controlled platforms. AI services have also recently adopted zero rating,[61] recognising that in markets permitting this practice, it becomes a prerequisite for competition, thus demonstrating how systemic vulnerabilities can be exploited to capture value and constrain sovereignty.

In contrast, India's 2016 prohibition of zero rating and affirmation of net neutrality represents a significant sovereignty achievement. By ensuring that all content and services compete on equal terms, this policy has fostered the growth of Indian technology companies and preserved space for innovation outside foreign platform control.[62] Yet, challenges remain around connectivity, specifically regarding affordability, reliability, and bandwidth. Drawing parallels to AI, the India AI Mission's focus on developing computational capacity, datasets, and foundational models addresses critical sovereignty stack components but must be complemented by continued attention to connectivity, digital literacy, and sociotechnical infrastructure enabling broad-based participation.

5.6 Need to Localise AI Governance Principles

AI systems are not universally applicable, and treating them as such can result in inefficiencies and harms. India's caste, religious, and ethnic diversity requires conceptions of fairness and bias that diverge from Western frameworks centred on race and gender. As an example, the discussion highlighted that research on facial recognition systems trained predominantly on Western datasets revealed significant performance disparities on Indian faces, particularly in gender classification.[63] Such examples also highlight the need for datasets, models, and evaluation frameworks that reflect local realities.

The lack of representative Indian datasets also poses technical and ethical challenges. Efforts to create such datasets raise questions about contextual integrity, such as whether data

---

[60] ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ; Meaningful connectivity refers to internet access that is not only available but also affordable, reliable, and of sufficient quality to enable full participation in the digital economy.

[61] Zero rating refers to practices wherein telecommunications providers exempt certain services from data charges

[62] ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ. India's 2016 net-neutrality regulations prohibited zero rating, ensuring equal competition for all content and services.

[63] ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ

collected for one purpose can be ethically repurposed for research or commercial AI development. For instance, the use of publicly available images in facial recognition systems raises concerns even when personal identifiers are removed and datasets are not publicly released.[64] Thus, responsible AI development requires not only compliance with the law but also ongoing ethical reflection and self-regulation by researchers, developers, and deployers.

5.7 Policy Implications and Insights for India

AI sovereignty requires a coordinated focus on technological and non-technological factors. Advances in computational capacity or algorithmic sophistication can be undermined by deficiencies in connectivity, energy infrastructure, or digital literacy. India's leadership in DPI and growing computational infrastructure position it favourably, but success depends on maintaining a sustained focus on non-technical factors such as meaningful connectivity and digital literacy initiatives.

Digital and AI sovereignty frameworks must be evaluated by the capacity to enhance citizen agency, ensure participatory governance, and deliver tangible welfare improvements and not merely by success in asserting state control or fostering corporate champions. Drawing on the social contract theory, sovereignty ultimately derives from citizens who delegate authority for protecting their rights and interests. At the same time, the tension between building powerful state-supported infrastructures and avoiding power concentration that enables exclusion and abuse of dominance remains a present concern.

India has the opportunity to advance South-South cooperation on AI governance. As India hosts the AI Impact Summit and assumes the BRICS presidency in 2026, it has the opportunity to build on existing work such as the Rio de Janeiro Declaration on Strengthening Global South Cooperation for a More Inclusive and Sustainable Governance[65] and the Kazan Declaration's call for a fair and equitable global framework for data governance.[66] Developing these commitments into substantive international instruments will require sustained political commitment and the creation of operational mechanisms, such as model contractual clauses for data transfers.[67] India's DPI experience positions it to contribute leadership on the global stage.

The transition from digital to AI sovereignty underscores that technology governance is inextricably linked to political economy, social justice, and democratic participation. Distribution of AI development benefits, preservation of space for diverse cultural

---

64 ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ

65 *Rio de Janeiro declaration: Strengthening Global South cooperation for a more inclusive and sustainable governance*. (2025, July, 6). https://www.mea.gov.in/bilateral-documents.htm?dtl/39770/Rio_de_Janeiro_Declaration_Strengthening_Global_South_Cooperation_for_a_More_Inclusive_and_Sustainable_Governance

66 *Kazan declaration: Strengthening multilateralism for just global development and security*, para. 71. (2024, October, 23). http://static.kremlin.ru/media/events/files/en/RosOySvLzGaJtmx2wYFv0lN4NSPZploG.pdf

67 ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ

expressions, and accountability of powerful systems to affected communities are not afterthoughts to be addressed once technical and economic challenges are solved. AI sovereignty that asserts state control while failing to empower citizens, address systemic inequalities, or enable meaningful participation in shaping technological trajectories would represent a hollow achievement.

5.8 Conclusion

The BRICS nations' journey from digital to AI sovereignty demonstrates that sovereignty cannot be achieved through regulation alone nor through technological development divorced from rights and participation. It requires integrated approaches that focus on the entire technology stack while grounding sovereignty claims in legitimacy derived from serving citizen interests rather than only asserting state or corporate power.

For India, the path forward involves building on existing strengths, such as DPI models, and digital sovereignty wins, such as the net neutrality protections, while addressing challenges in connectivity, equity, and digital literacy and attempting to develop governance frameworks that balance innovation with accountability. As AI systems become increasingly central to economic and social life, the stakes grow. The choices India makes about regulation and industrial policy, data governance and infrastructure, and participation and accountability will shape not only its global position but the lived experience of its citizens. The BRICS experience suggests that multiple paths exist, but the most important choices do not concern technical capabilities alone but also the values and interests that technological development serves.

6. Reflections on Ethics and Governance of AI: Insights from the Webinar Series

This webinar series traced an arc through critical dimensions of AI governance. Beginning with the regulatory architecture of AI value chains, where the EU AI Act's distribution of obligations revealed fundamental tensions between upstream model development and downstream deployment, the series progressed to AI sustainability, confronting the environmental and social costs embedded in the technology's physical infrastructure. The AI researchers then discussed their perspectives and explored deficit-model thinking and industry influence on research agendas, while the final discussion on digital and AI sovereignty discussed frameworks for technological self-determination. Together, these conversations frame AI governance as confronting challenges that are simultaneously technical, political, economic, environmental, and social and which demand integrated approaches rather than siloed interventions.

Reflections and insights from this webinar series offer both cautionary lessons from other jurisdictions and suggestions to articulate India's distinctive approach towards AI governance by synthesising cross-cutting themes across conversations, outlining frictions that require deliberation, and translating these discussions into considerations for India's strategic positioning.

6.1 Cross-Cutting Themes: Identifying Structural Challenges in AI Governance

Asymmetries in information, capability, and bargaining power were a fairly consistent undercurrent across the series. For instance, upstream foundation model providers possess technical knowledge without downstream deployment context, while downstream developers seek to understand application risks without visibility into model characteristics. Market concentration amplifies these asymmetries, enabling powerful actors to dictate contractual terms limiting liability.[68] Similar dynamics appeared in researcher-public engagement, where deficit model thinking, i.e., the assumption that public concerns stem from inadequate understanding of AI rather than legitimate value differences, perpetuates hierarchies privileging technical expertise over lived experience. When experts working in AI assume that just enhancing public understanding of AI will bolster confidence in the technology, they risk overlooking valuable insights that could contribute to AI advancement.[69] The discussion on sovereignty illuminated how technological dependencies create asymmetries constraining policy autonomy as countries relying on foreign infrastructure face informational and jurisdictional challenges when regulating AI systems or attributing responsibility for harms.[70] Addressing these issues requires AI governance mechanisms to facilitate information exchange, redistribute bargaining power, and enable accountability across organisational and jurisdictional boundaries.

There exists recognition of AI's materiality challenges and framings that treat governance primarily as managing algorithmic behaviour or data flows. The discussion on AI sustainability foregrounded material foundations underlying these seemingly virtual technologies. Large language models require vast computational infrastructure, such as data centres that consume an immense amount of water and electricity, and require cooling systems that draw water from resource-stressed areas, specialised semiconductors manufactured through water-intensive processes, and rare earth minerals for manufacturing hardware often extracted under conditions that cause environmental degradation and social harm.[71] These material foundations have direct sovereignty implications, with computational capacity, electrical power, and connectivity constituting essential elements of the "AI sovereignty stack". Yet, these infrastructures are unevenly distributed and concentrated among a few actors.[72] AI systems are promoted as general-purpose technologies transcending physical constraints, but their development depends on specific material conditions that are neither universally available nor environmentally sustainable at current growth trajectories. The effective governance of AI must address the entire lifecycle, from

68 ICRIER. (2025, October 28). *Regulating the AI value chain in the EU and beyond: Upstream, downstream, or in between* [Video]. YouTube. https://www.youtube.com/live/KiOLhhJZcFc?si=DltcxwLLXkUv8sIz
69 ICRIER. (2025, December 19). *Visions, values, voices: A survey of artificial intelligence researchers* [Video]. https://youtu.be/o8eGMGK2S3g
70 ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ
71 ICRIER. (2025, November 21). *Navigating sustainability within the AI lifecycle* [Video]. https://youtu.be/YZgMZL4MmLk.
72 ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ

mineral extraction for hardware manufacturing to energy-intensive training, deployment infrastructure, and eventual disposal, thus confronting questions about who benefits from AI development and who bears its environmental and social costs.

The friction between claims of universality and local realities was also discussed. AI systems promoted as general-purpose technologies applicable across contexts produce impacts that do not conform to local realities. General-purpose models designed for broad deployment can create risk-assessment challenges because harms emerge in specific contexts that upstream providers cannot anticipate.[73] The environmental impacts of AI also vary based on whether electricity comes from coal or renewables, whether water is abundant or scarce, or whether regulatory capacity exists to enforce standards. Further, India's diversity requires conceptions of fairness and bias diverging from Western frameworks, the lack of which can result in claims of universality masking context-specific inadequacies with discriminatory effects.[74] AI governance must determine which principles operate transnationally versus which require adaptation to local contexts and balance global participation and collaboration against indigenous capability development serving local needs.

Underlying these themes is the concentration and distribution of power in AI ecosystems. The concentration of computational resources among a few technology companies contrasts with globally distributed environmental and social costs, disproportionately affecting Global South communities.[75,76] The concentration of research agendas within large corporations determines which problems are prioritised and what solutions appear viable.[77] India's DPI initiatives present an alternative to this challenge of concentrated private control while raising questions about whether state-supported infrastructures might reproduce concentration under public auspices. The concentration and distribution of power extends beyond market structure to AI governance architecture and the policy choices that affect whose interests AI serves and whose voices shape its trajectory.

6.2 Opportunities, Imperatives, and Strategic Questions for India

India's position as a developing economy and an emerging player in AI creates distinctive opportunities for shaping global AI governance.

Using DPI principles as an alternative to concentrated private control: India can position DPI governance frameworks as models for AI sovereignty that balance state capability building with citizen empowerment by, *inter alia*, documenting DPI governance mechanisms for

---

73 ICRIER. (2025, October 28). *Regulating the AI value chain in the EU and beyond: Upstream, downstream, or in between* [Video]. YouTube. https://www.youtube.com/live/KiOLhhJZcFc?si=DltcxwLLXkUv8sIz

74 ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ

75 ICRIER. (2025, November 21). *Navigating sustainability within the AI lifecycle* [Video]. https://youtu.be/YZgMZL4MmLk.

76 ICRIER. (2026, January 28). *From digital sovereignty to AI sovereignty: Lessons from the BRICS countries* [Video]. https://youtu.be/FlvN3YJ_yDQ

77 ICRIER. (2025, December 19). *Visions, values, voices: A survey of artificial intelligence researchers* [Video]. https://youtu.be/o8eGMGK2S3g

transparency and adaptability and sharing lessons about participatory governance in technology infrastructure. However, this requires confronting accountability challenges and addressing concerns about translating India's DPI experience to AI governance without recreating concentration dynamics under public control.

Embedding sustainability in AI systems across technical, social, ecological, economic, and governance dimensions: This can be operationalised by mandating sustainability impact assessments for the India AI Mission's infrastructure projects, requiring energy and water efficiency standards for data centres receiving public support and investing in research on energy-efficient model architectures and transfer learning approaches, thereby reducing computational demands. India should also explore if finetuning existing models for specific applications rather than developing new giant models could reduce the environmental footprint while serving development needs. This approach would challenge narratives equating AI leadership with developing the largest models and instead emphasise adapting for purpose and responsible resource use. However, prioritising sustainability from the start will require India to confront the friction between technological advancement and sustainability imperatives, which is particularly acute for developing nations pursuing global technological competitiveness.

Building institutional capacity for adaptive governance: India requires institutional innovations to enable governance that is responsive to evolving technologies. Suggested approaches include establishing multi-stakeholder coordination mechanisms bridging the coordination gap among regulators such as competition authorities, data protection bodies, and sectoral regulators; creating researcher-regulator exchanges to build technical knowledge within the government; developing capacity for sophisticated analysis and translating insights into enforcement; exploring the creation of regulatory sandboxes enabling controlled experimentation; and creating mechanisms for continuous stakeholder consultation to prevent capture by dominant interests. A major challenge will be building institutions capable of adaptation rather than defaulting to applying fixed rules.

Operationalising meaningful participation in AI governance: While public concerns about AI are often dismissed as information deficits rather than legitimate value differences, the technical complexity of the technology creates barriers for informed participation. Participatory governance offers a potential solution, but transposing this approach to fast-moving technological contexts requires institutional innovation. Major considerations include determining which decisions require broad participation, structuring participation to be empowering, and devising methods to bridge gaps between technical expertise and experiential knowledge without privileging one over the other. To address this, India could build coalitions between concerned researchers, civil society, and affected communities around shared governance interests.

6.3 Conclusion

As India assumes leadership in global AI governance through the AI Impact Summit and the BRICS presidency, the stakes extend beyond diplomatic positioning to fundamental questions about technological futures. Such questions, as exemplified by discussions in this webinar series, resist simple solutions and require sustained engagement across researchers, policymakers, industry, civil society, and affected communities, enabling mutual learning and collective navigation of complex trade-offs. This collaborative approach fosters a more inclusive governance framework and ensures that diverse perspectives inform the ethical and practical dimensions of AI development. Thus, India's opportunity lies in advocating for approaches that centre equity, sustainability, and democratic participation as key design principles shaping the development trajectories of AI.

ICRIER PROSUS
CENTRE FOR INTERNET AND DIGITAL ECONOMY

Our Offices:
4th Floor, Core 6A, India Habitat Centre, Lodhi Road, New Delhi-110003

The Isher Building, Plot No. 16-17, Pushp Vihar, Institutional Area, Sector 6, New Delhi-110017

O: +91 11 43112400 / 24645218 | **W:** https://icrier.org/ipcide/
Email: ipcide@icrier.res.in