CARNEGIE INDIA

Policy Planning & Research Division
Ministry of External Affairs
Government of India

2024

# GLOBAL TECHNOLOGY SUMMIT

## INNOVATION DIALOGUE

## Summary of Discussions and Key Insights

October 8-9, 2024 - Bengaluru

**Disclaimer**

This document is strictly confidential and intended solely for internal use. The following notes summarize off-the-record discussions from the Global Technology Summit Innovation Dialogue 2024 and must not be shared, reproduced, or distributed outside authorized circles.

# CONFIDENTIAL

## CO-HOST

Policy Planning & Research Division
**Ministry of External Affairs**
**Government of India**

## PHILANTHROPY PARTNER

TATA TRUSTS

## DIALOGUE PARTNERS

Microsoft

salesforce

Meta

Google

## KNOWLEDGE PARTNERS

CO-DEVELOP

BILL & MELINDA GATES foundation

EkStep

Center for Digital Public Infrastructure

CENTRE FOR OPEN SOCIETAL SYSTEMS

FUTURE OF PRIVACY FORUM

# Reader's Note

The following notes provide a confidential summary of discussions held at the Global Technology Summit Innovation Dialogue 2024. The information contained within is strictly for internal use and should not be shared externally.

CONFIDENTIAL

# Global Technology Summit Innovation Dialogue 2024

## Summary of Discussions and Key Insights

**October 8-9, 2024 - Bengaluru**

CONFIDENTIAL

**Disclaimer**



This document is strictly confidential and intended solely for internal use. The following notes summarize off-the-record discussions from the Global Technology Summit Innovation Dialogue 2024 and must not be shared, reproduced, or distributed outside authorized circles.



Carnegie India
Unit C-4, C-5 & C-6, Edenpark,
Shaheed Jeet Singh Marg
New Delhi – 110016, India
P: +011 4008687
CarnegieIndia.org

# Contents

KEY TAKEAWAYS

# The Promise of DPI: From Concepts to Action

## October 8, 2024

The closed-door workshop on "The Promise of DPI: From Concepts to Action" convened experts from over twenty countries, including policymakers, digital public infrastructure (DPI) architects, start-up leaders, and private sector representatives.

The primary objectives of the workshop were to:

i. Explore the fundamental differences between general digitalization and DPIs.

ii. Discuss challenges of DPI projects globally.

iii. Identify success factors and obstacles to the long-term success of DPI initiatives.

The workshop followed a plenary format, where specific participants were invited to give interventions followed by an open dialogue, encouraging all attendees to reflect on the remarks and share their insights. Later, participants were split into breakout groups to focus on key themes such as trust, DPI adoption, and scaling of DPIs.

Below are the key takeaways from the workshop:

### 1. Key Principles of DPI and Its Impact on Global Development

DPI combines inclusion and innovation through a whole-of-society approach, allowing participation from all stakeholders—ranging from small businesses and large technology solutions providers to governments. A core principle in DPI implementation is a user-centric design, focusing on three critical areas: identification (who am I?), assets (what do I have?), and transactions (what do I do?). DPI development is further shaped by key principles such

as minimalism, interoperability, reusability, and decentralization, which ensure that solutions are efficient, scalable, and adaptable to different environments, thus maximizing their impact.

DPI has become a transformative driver in global development, drawing the attention of major international organizations like the G20, the United Nations (UN), and the World Bank. Successful implementations can be seen across various regions, such as Brazil's Pix digital payment system, Papua New Guinea's whole-of-government approach, including a pilot ID and payments system and a new e-government portal, and Morocco's collaboration with India on social aid registries and e-learning platforms. In East Africa, regional integration is advancing through initiatives like the M-PESA mobile wallet. These examples illustrate DPI's versatility and potential to adapt to specific national and regional challenges.

## 2. Implementation Challenges of DPI

Despite the promise of DPI, several challenges hinder its implementation. Many countries face capacity limitations, such as the lack of technical and inadequate infrastructure. The bureaucratic process required for DPI adoption is often slow and encounters resistance. Current funding heavily relies on philanthropic sources, highlighting the need for more sustainable financial models. In some nations, the digital divide complicates access to digital infrastructure and services. Furthermore, concerns around cybersecurity and data privacy create additional complexities, especially when dealing with sensitive data. The significant challenge that comes with adapting DPI solutions to various linguistic and cultural contexts was also underscored.

## 3. The Role of the Government and the Private Sector

The success of DPI initiatives hinges on balancing government involvement with private sector participation. Government support is crucial for broad DPI adoption, with certain regulatory frameworks needed to build initial momentum. However, private sector representatives stressed that governments should set clear expectations without prescribing specific solutions. Minimal government intervention with a focus on providing only essential support and guidelines allows the market to drive innovation. Building public trust is also vital, requiring joint efforts from both the public and private sectors.

## 4. The Way Forward

The need to develop operational principles tailored to specific contexts was emphasized. It was also acknowledged that while core DPI principles are universal, their application must be customized to local conditions. Global knowledge sharing will be key to the success of DPI. Integrating the digital government with the digital economy offers new opportunities

for economic growth and improved public service delivery. Lastly, balancing data sovereignty with the economic value of verifiable data sharing represents a significant challenge for future DPI development.

Following the plenary session, participants were divided into three groups: one focused on trust in DPI adoption, another on various DPI adoption models, and a third on securing funding and support to scale DPI initiatives.

## Breakout Session 1: Trust

During this discussion, trust was emphasized as a fundamental challenge in DPI adoption. A significant lack of trust exists around digitalization, largely due to concerns around data transparency and security. While India has taken steps to integrate privacy by design into its systems, public confidence in state institutions remains fragile. It was also noted that trust is often built through reliability—when systems work consistently, users are more likely to trust them.

Below are the key takeaways from this breakout group discussion:

1. **The Role of Regulatory Frameworks in Building Trust**: A strong data protection law is essential to safeguard citizens' data and ensure legal accountability for violations. Although India has enacted the Digital Personal Data Protection Act, 2023, the lack of clear rules defining the responsibilities of stakeholders in the data protection ecosystem hinders trust in DPI solutions. In countries like South Africa, where regulation is weak, particularly in sectors like insurance, the absence of proper frameworks has undermined public trust. India's DPI initiatives require not only robust data governance policies but also enforceable technical measures to reassure citizens that their data is being handled responsibly.

2. **Digital Literacy and Education**: Education and digital literacy are crucial for fostering trust in DPI. Without an understanding of how technology works, citizens may remain doubtful and reluctant to adopt new solutions. This is especially relevant for those lacking advanced digital skills. In particular, the need to create an inclusive digital ecosystem by upskilling citizens at all levels and designing technology that meets their needs was highlighted.

3. **Psychological Aspects of Trust**: The psychological dimension of trust is often an overlooked factor but plays a significant role in DPI adoption. Psychological harm can diminish trust in digital systems. For example, generative AI chatbots can create false emotional connections with users, potentially causing psychological distress. The importance of extending trust beyond applications like payment systems to every layer of the infrastructure was emphasized. Users' psychological comfort with technology greatly influences their willingness to adopt a DPI system.

4. **Fostering Trust in DPI**: It was agreed that India's DPI efforts should focus on building systems that are reliable and transparent. A clear purpose, as seen in the Aadhaar initiative, is key to ensuring users understand the system's limits and assurances. India's growing success is closely tied to its DPI initiatives, which have become a central part of the country's development strategy. For other nations looking to adopt DPI, it is crucial to focus on both the successes and challenges of these initiatives.

## Breakout Session 2: Models of DPI Adoption

In this group, various global challenges that delay the introduction and adoption of DPI in different countries were discussed. While some countries face cultural resistance to digital transformation, others lack the skilled workforce necessary for DPI implementation. Additional hurdles include securing ongoing funding support for the sustainability of DPI projects, insufficient political will, and hesitation from the private sector to engage in the DPI journey.

Below are the key takeaways highlighting some challenges and a few solutions to facilitate DPI adoption from this group discussion:

1. **Skills Gap and Local Ecosystem Development**: A major challenge in DPI adoption is the lack of local skills and expertise. The questions around whether local expertise is essential for DPI development and, if so, how to build that capacity were also debated. The skills gap extends beyond technical know-how to strategic thinking about DPI within governments. While external consultants can provide initial support, building internal think tanks and transferring knowledge to local officials is crucial for sustainable policymaking. Building local technical expertise to support DPI adoption and manage the solution after implementation is also a significant challenge.

2. **Standardization and Productization Challenges**: The lack of standardization in DPI implementation presents major challenges. Without clear procurement guidelines, policymakers struggle to determine how to implement DPI solutions, even after choosing a specific approach. There is a need to shift from custom, context-specific implementations to more standardized, product-based approaches that offer simple, easily adoptable solutions with clearly defined steps for countries to follow. An example of this shift is Google's newly launched "DPI in a Box," which provides a ready-made solution for governments looking to replicate India's successful digital identity framework.

3. **Private Sector Participation**: While governments play a key role in facilitating DPI adoption, the private sector is essential for driving innovation. The need to incentivize private sector involvement—for example, through DPI credits—was

highlighted. The need to explore possible mechanisms to allow this was also discussed.

4. **Awareness on User-Centric Benefits of DPI**: Many end users are unaware of the benefits of DPIs, making awareness-building a significant challenge. Articulating the purpose and advantages of DPI adoption to common users is crucial for its widespread acceptance. Successful cases like India's Aadhaar system, which facilitated direct benefit transfers, show the importance of clearly communicating user-centric benefits for effective DPI adoption.

5. **Scaling and Sustainability Challenges**: Scaling DPI systems presents unique challenges. The debate between "brownfield" approaches (leveraging existing systems) and "greenfield" approaches (building new systems) was discussed, with some advocating for building upon existing infrastructure rather than creating new systems from scratch.

Additionally, the absence of clear funding models and metrics for return on investment makes the long-term sustainability of DPI projects difficult. While initial funding often comes from government subsidies, ensuring long-term viability, especially in resource-limited countries, is a concern. Determining unit costs and growth targets is crucial for DPI maintenance and sustainability. Attracting private sector investment, particularly in cases where immediate economic benefits aren't apparent, remains a challenge. Lastly, the perception of DPI as primarily a "developing country" issue and how it can limit investment and collaborations was also brought up.

## Breakout Session 3: Scaling DPIs: Funding and Support

This breakout session focused on the funding challenges for DPI projects globally, emphasizing various mechanisms available in India as well as multilateral organizations like the World Bank that have allocatedfunds to support global DPI adoption. The importance of aligning market incentives for sustainable DPI funding was underscored, and several issues related to financing DPI projects were also highlighted.

Below are the key takeaways from this breakout group discussion:

1. **DPI Funding in India**: In India, several government agencies are involved in funding DPI initiatives. The DPA-III department within the Ministry of External Affairs manages a portfolio of INR 4 billion, concentrating on infrastructure development in neighboring countries, the African continent, and support for small and medium enterprises within India. DPA-III is engaged with over twenty foreign governments, and India has signed fourteen memorandums of understanding (MoUs) for exporting DPI, leaving them intentionally vague to allow for context-specific implementation. Additionally, the Ministry of Electronics and Information Technology (MeitY) oversees the DPI Social Impact Fund, while the Reserve Bank of India manages the Unified Payments Interface (UPI).

2. **Challenges and Considerations in Financing DPIs**: The main challenges in financing DPI projects include: (i) a lack of clarity on the benefits of DPIs; (ii) identifying context-specific problems to align funding; (iii) understanding the architectural layer of DPI systems; (iv) managing currency and remittance issues (preference for dollar payments over local currencies); (v) navigating contracting difficulties across borders (a neutral authority certifying DPI projects could help compliance); (vi) demonstrating return on investment to secure political support; and (vii) addressing data localization issues.

   Building both geopolitical and technical trust is critical for successful DPI deployment, and the timing of funding decisions within the policy planning process is crucial. Sensitive DPI solutions, like identity systems, tend to attract more government scrutiny, impacting funding decisions.

3. **Design Choices and Implementation Strategies**: Design choices, such as opting for open-source versus proprietary technology, play a key role in DPI implementation. Open-source solutions often lack comprehensive maintenance services, but with improved functionality, they could become more attractive. In this regard, pilot projects and readiness assessments were suggested as effective ways to test design choices. Stakeholder collaboration to brainstorm funding ideas was also recommended.

4. **Global Context and Future Outlook**: DPI funding must be considered in the broader context of other urgent issues in the Global South, such as health, education, food security, and debt restructuring. International collaboration is vital, with the European Union exploring cooperation with India through the EU-India Trade and Technology Council. Morocco's adoption of India's ID program was highlighted as a successful DPI transfer, underscoring the importance of aligning with specific agendas and providing access to source code. Regional organizations like the East African Community (EAC) are also seeking funding for cross-border DPI projects, having established partnerships with India and China, and are committed to digital transformation.

# Accelerating Innovation with DPIs: Markets, Products, and Network Governance

## October 8, 2024

The closed-door workshop on "Accelerating Innovation With DPIs: Markets, Products, and Network Governance" convened experts from over twenty countries, including policy experts, policy implementers, DPI architects, start-ups, and representatives from the private sector who have already innovated on top of DPIs.

The primary objectives of the workshop were to:

i. Provide a detailed understanding of how market products can be created on top of DPIs and highlight the existing challenges.

ii. Discuss how start-ups and the private sector have innovated thus far by referencing successful and live examples.

Below are the key takeaways from the workshop:

### 1. Evolution of DPI and the Market Story

The discussion began by examining the evolution of DPIs in India, noting that before 2016, Indian start-ups were largely focused on developing software solutions for local businesses or seeking markets overseas due to a relatively small domestic internet user base.

The advent of DPIs has significantly altered India's technological landscape, not only boosting financial inclusion but also empowering businesses to discover new products and market opportunities.

For instance, UPI has promoted financial inclusion by bringing millions into the formal banking system. Additionally, it has opened new avenues for private companies to expand their reach and innovate, making UPI transactions smoother and more accessible. One of the points noted was that over 300 banks in India have integrated with UPI, reflecting its robustness and growth potential.

Similarly, Aadhaar's integration with services such as Know-Your-Customer (KYC) processes and e-signatures has simplified access to financial products, allowing people to engage in savings and investments conveniently from home, thereby saving time and resources.

The deployment of DPIs has also enabled companies to provide services to citizens on a global scale, showcasing its potential for governments to offer large-scale digital services. An example is the EAC, an intergovernmental group of eight countries that uses DPI to establish a common customs union, enabling real-time tracking of goods moving across borders.

## 2. Business Model for DPI

The conventional way of approaching business solutions involves analyzing business models and the value they generate. Similar questions about the viability and sustainability of business models arise with DPI, particularly regarding value creation. However, DPI operates as an enabling layer—it helps create value but does not necessarily capture it. This means DPI provides essential infrastructure—such as identity verification, payments, and data-sharing services—while the actual value is captured by the businesses and applications that build on top of it. A useful comparison is the internet: telecom companies offer access to the internet, but it is the companies developing apps, websites, and online services that truly generate revenue and capture value. Similarly, DPIs like Aadhaar and UPI enable businesses to innovate and create services that weren't previously possible.

In this respect, the function of systems for identity, payment, and data exchange as the "rails" for innovation, particularly in financial services, was highlighted. These rails then lead to new products in credit, investments, and banking. Beyond financial services, DPI adoption has also allowed companies to innovate in other areas, such as providing faster e-commerce deliveries. Moreover, UPI has empowered citizens to build credit histories by digitizing transactions, which can now be used to offer credit lines, revolutionizing access to credit in India.

## 3. Global Demand for DPI

Global interest in DPIs, particularly in identity and credentialing systems, is rapidly increasing among both countries and companies. Nations in Africa and Asia are particularly focused on adopting DPI solutions, with Morocco being one of the first to implement the

Modular Open Source Identity Platform (MOSIP) for building a national digital ID system based on DPI principles. Regional bodies like the EAC are also exploring DPIs for projects such as customs unions and digital payments. Over the past year, the Indian government has signed MoUs with several African and Caribbean countries to share its DPI solutions.

While DPI discussions were previously driven by governments and philanthropic organizations, the last six months have seen a notable shift with the private sector becoming more involved. Companies like Oracle and Salesforce have shown growing interest, and hyperscalers such as Google are offering solutions like "DPI in a Box." However, countries in Asia are still determining how best to utilize DPIs.

Private sector representatives recognized that despite the significant value and interest in DPIs, moving from discussions to actual deployment remains a challenge. Key barriers include concerns about data sovereignty and security, which are often amplified by geopolitical factors. In response, some companies are now providing fully managed, air-gapped services that can be deployed within a national border, offering greater control and security for national data.

## 4. Challenges in Implementation of DPI

Several challenges related to the implementation of DPIs, including data sovereignty, funding, capacity building, and the short-term versus long-term vision of DPIs, were also discussed. Those highlighted during the workshop include:

a. **Data Sovereignty and Privacy**: A recurring issue in DPI discussions is data sovereignty and privacy, particularly regarding data ownership and usage. Countries are grappling with how to ensure user control over data while maintaining open and sovereign data systems. The lack of a clear definition of sovereignty complicates efforts to find solutions to these concerns.

b. **Customizability Versus Scalability**: There is a trade-off between a customizable, modular solution—which may be more expensive, difficult to implement, and require local capacity—and an out-of-the-box solution, such as Google's "DPI in a Box," which can be scaled rapidly but may not suit local context. Countries must decide which one to opt for based on their specific needs. Hyperscalers provide the infrastructure needed to scale DPI globally, but governments must weigh the risks of relying on global tech companies, especially regarding data sovereignty and long-term sustainability.

c. **Short-Term Versus Long-Term Vision of DPIs**: There is often a conflict between short-term revenue models and long-term goals of DPIs. DPIs are seen as long-term investments, with their full impact extended to be realized over decades. Balancing immediate revenue needs with long-term benefits will remain a challenge.

d. **Impact Versus Return**: Another challenge in implementing DPIs is balancing social impact with shareholder value. While DPIs aim to create solutions with significant social benefits, entrepreneurs must also generate value for investors. Finding a balance between these objectives is difficult. On this front, the need for incentives to encourage smaller, agile companies to solve problems using DPIs was noted.

# The U.S.-India Tech Partnership: Opportunities

**October 9, 2024**

The closed-door discussion on "The U.S.-India Tech Partnership: Opportunities" was attended by senior government officials, industry executives, and academics who have been closely involved in shaping the U.S.-India technology partnership.

During the discussion, the opportunities and challenges in the U.S.-India partnership were addressed, with a focus on three technologies: artificial intelligence (AI), semiconductors, and space technologies.

The key takeaways from these discussions are highlighted below:

## 1. Current Assessment of the Partnership

The overarching message was that the United States and India need to strengthen existing areas of technology cooperation in defense and space while identifying new areas of collaboration, especially in AI and DPI. They should leverage their respective strengths—the United States in core technologies and India in deploying applications at scale—while reducing foreign dependencies.

From a strategic perspective, the key issue at stake is who controls the "pipes, platforms, and rules." This raises important questions on AI governance, export controls, compute access, chip design, and defense/space technologies for national security, all of which were discussed during the meeting

.

## 2. The Role of AI Safety Institutes (AISIs)

A senior representative from the U.S. AISI, established under the National Institute of Standards and Technology (NIST), emphasized the role of the AISI in developing standards, toolkits, and frameworks to promote global collaboration on AI governance. The goal is to develop benchmarks and a common vocabulary that would enable informed discussions on AI safety. For example, the U.S. AISI would conduct scientific research to help quantify the risks of AI systems and share these results with other countries so that they can be appropriately mitigated through a collaborative approach.

The Indian government is also considering setting up an AI safety institute, which would be part of a global network of such institutes. It was pointed out that this kind of collaboration would simplify industry compliance, reduce regulation, and promote cross-border technology development.

However, the Indian delegation pointed out the inherent tension between the global and local needs that arise in AI safety discussions. The current approach is to impose a Western notion of safety on the developing world, whereas countries in the Global South want more flexibility in how they define these terms based on their local context. They suggested a more nuanced approach to AI governance that recognizes the centrality of technology to the development agenda. According to them, AI policy should enable data and compute access for socioeconomic growth while mitigating the risks for specific use cases.

## 3. Export Controls on Open-Source AI

Some prominent developers in India are using open-source foundation models developed by U.S. tech companies to build "sovereign AI" solutions. They use open-source models to serve a variety of customers who want more choice and control over their assets based on their specific risk profile. For these reasons, Indian developers are closely tracking global developments that could impact their business.

A key risk was the U.S. government's proposal to impose export controls on certain AI models.

Three existing bills in the United States that contain export controls were highlighted in this regard (for example, the ENFORCE Act), and the following three primary concerns were highlighted: (1) the potential spillover effects of such regulations for the Global South, even if they are not targeted at countries like India; (2) export controls on AI models are difficult to technically implement, especially if these models are already in distribution; and (3) such rules could have unintended consequences because it is unclear at what threshold these tools become dangerous.

The important national security considerations at play were acknowledged, specifically the need to prevent foreign adversaries from having access to powerful general-purpose technology. However, instead of imposing export controls, they pointed to the need for a more collaborative approach to AI governance that addressed the specific risks relating to open-source models.

## 4. Semiconductor Design and Supply Chains

The establishment of the International Technology Security and Innovation Fund and the announcement of a fabrication plant being set up for national security purposes are recent examples of how the United States' and India's semiconductor ecosystems are being closely integrated.

However, certain challenges were also pinpointed. First, India's latent talent in chip design was not being adequately leveraged in the bilateral partnership. Second, both the United States and India have ceded space to China on legacy chips, which needs to be de-risked. Lastly, concerns were raised about market concentration in the design and development of chips and the high energy costs of AI development.

## 5. Opportunities and Challenges for Space Technologies

The growing collaboration in space technologies was acknowledged, with India's signing of the Artemis Accords being hailed as a milestone. Key opportunities going forward include collaborations on NASA's Commercial Lunar Payload Services program and Gateway. The opportunity to collaborate on satellite-based surveillance technologies to safeguard India's borders was also discussed.

However, concerns remain over the International Traffic in Arms Regulations and Export Administration Regulations. Some Indian founders confessed to setting up subsidiaries in the United States to get around these regulations, though it was resulting in a brain drain. Concerns were also raised about the delays in joint space programs, including the NASA-ISRO Synthetic Aperture Radar satellite launch and the joint mission to the International Space Station, both of which have been deferred to 2025.

## 6. The Future of the U.S.-India Technology Partnership

The U.S.-India technology partnership shows great promise in areas of defense, semiconductors, and space, with room to further expand the collaboration on AI. Some examples include collaborating on AI governance through their respective AI safety institutes, leveraging DPI to enable access to compute, and reducing regulatory barriers that could hinder access to AI systems.

The critical role played by initiatives such as the U.S.-India initiative on Critical and Emerging Technology (iCET) and INDUS-X in enhancing technology collaboration between the two sides was highlighted, and similar opportunities through new initiatives such as Bio-X in biotechnology were also noted.

It was also suggested that the partnership on technology has become a critical element of the United States' Indo-Pacific Strategy. Therefore, maintaining strategic continuity will be important for both sides following the upcoming elections in the United States.

Lastly, the 2025 Quad Foreign Ministers Meeting (to be hosted by the United States) and 2025 Quad Leaders' Summit (to be hosted by India) provide important avenues to discuss these issues further.

# A Dialogue on AI Safety & Innovation

## October 9, 2024

The closed-door workshop titled "A Dialogue on AI Safety & Innovation" sought to unpack three issues: (1) the global experience of setting up AISIs; (2) perspectives from Global South countries on AI safety; and (3) the conceptual and practical understanding of how to approach risk and innovation in the context of AI safety.

The discussion brought together experts from government, industry, think tanks, and academia and featured remarks from senior government officials, presentations by representatives from AISIs, discussions with participants from the Global South, and two breakout sessions.

The key takeaways from these discussions are highlighted below:

### 1. Global Network of AISIs

The need for greater multistakeholder participation and scientific rigor on the topic of AI safety because of current limitations was discussed. For example, the G7 Hiroshima process on AI involved only advanced industrial economies, while the UN's advisory group on AI governance lacks in-depth technical expertise. Further, many of these multilateral forums view AI safety and innovation as mutually exclusive rather than synergetic.

Therefore, a global network of AISIs, as envisioned at the AI Seoul Summit, could support these multilateral forums. The network of AISIs would generate and share scientific research on AI risks and mitigation strategies and align their safety testing toolkits to a set of common global standards. The goal would be to promote interoperability, simplify industry compliance, and promote cross-border collaboration on AI.

## 2. Global South Perspectives on AI Safety

While it was agreed that a global network of AISIs would be useful, many argued for greater representation of Global South countries in these discussions. Although many of these countries are in the early stages of developing their national AI strategies, AI safety remains a priority (for example, in the Continental Artificial Intelligence Strategy released by the African Union).

Further, while global cooperation among AISIs should be encouraged, it was also highlighted that the definition of "safety" should align with the local needs of the developing world and reflect their pro-innovation and development agenda using AI.

The issue of "cultural erosion" as a key risk for the Global South was also raised since the current frontier AI models are largely developed in the West and do not use datasets representative of the unique cultural realities of these countries.

Another issue raised was that job displacement from AI was most likely to impact vulnerable groups in these countries. Therefore, the role of AISIs in capacity building, training, reskilling, and mitigating labor market impact was of most relevance to them.

## 3. Learnings From Global AISIs

The workshop also featured presentations by the representatives of AISIs from the United States, the United Kingdom, and Singapore, followed by open discussions with all participants. Below are some key takeaways on the learnings and challenges involved in setting up these institutes.

a. **Structure and Composition**: The UK AISI is modeled after the UK Vaccine Taskforce and consists of over forty researchers, with an initial funding of £100 million over eighteen months. Based on commitments made at the first AI Safety Summit in Bletchley, the UK AISI also has access to frontier AI models for safety testing.

   The U.S. AISI focuses primarily on standard setting and is housed within the NIST, U.S. Department of Commerce. The U.S. AISI is supported by a multistakeholder consortium of over 290 organizations, with a small initial budget.

   Singapore's AISI is structured differently. It is housed outside of the government, within the Digital Trust Centre at Nanyang Technological University (an academic institution). It conducts technical research and complements Singapore's existing governance approach, which is largely voluntary-based, use-case-centric, and focused on developing responsible AI technologies.

b. **Scope and Functions**: The point on how the mandate of an AI should be iterative was made, while caution against "scope creep" was also raised, and it was suggested that functions should be clearly defined upfront.

   Based on the discussions, the following common functions of an AISI were identified:

   First, at a high level, AISIs aim to develop common standards, toolkits, and frameworks so that different countries can evaluate the safety of AI models using similar metrics. While some frameworks are still under development, the UK AISI has released the "Inspect" tool and the Singapore AISI has released the "Project Moonshot" tool to make these safety toolkits widely available.

   Second, an AISI helps promote innovation within the AI ecosystem by driving a clear nexus between AI safety and innovation. According to the U.S. AISI's vision statement, for example, safety enhances trust in the adoption of AI systems, which in turn accelerates innovation.

   Third, AISIs generally steer clear of regulatory interventions, with the exception of the European Union's AISI (that is, the EU AI Office, which is established under the EU AI Act). However, their reports, standards, frameworks, and guidance usually inform sectoral regulations.

   From a Global South perspective, it was suggested that AISIs should take a lead on AI safety issues in downstream applications and post-deployment evaluations, given their unique context in which advanced models are developed elsewhere but deployed locally at scale. They should also focus on developing innovative tools to address the risks around privacy and bias.

   Toward this end, a recommendation to rename AISIs as "AI Safety and Innovation Institute" or "AI Governance Institute" to reflect these functions was also made.

c. **Understanding AI Risks**: A key function of an AISI is to develop a common framework to understand AI risks and help mitigate them. In general, AI risk assessment refers to quantifying the possibility of harm occurring, which is often difficult. As such, the need for rigorous scientific methods was discussed, and it was suggested that organizations should clearly document the risks of AI systems during the development and deployment lifecycle to support AISIs in this exercise.

   Another suggestion floated was to break up "AI safety" into smaller components, for example, to assess AI risks in specific domains (civil or military use). Tailoring safety measures to specific contexts, purposes, or local capabilities of AI systems could help with risk management.

Another related challenge is explaining how any AI system works, which is especially difficult for advanced AI models. Therefore, a more careful approach to the use of AI in high-risk situations where "de-biasing" and explainability are challenges was suggested.

d. **Promoting Innovation**: One of the points raised was that discussing AI safety and innovation was a "paradox" because they are inherently incompatible. However, the general consensus was that safety and innovation are interrelated and that AISIs can actively promote an innovation ecosystem.

   For instance, AISIs can support and invest in technical research to promote privacy-enhancing technologies, cybersecurity tools for national and enterprise security, watermarking for content identification and provenance, predictive analysis to curb future risks, and innovative data-sharing methods. In particular, India's Data Empowerment and Protection Architecture was highlighted as an approach to democratize access to data in a privacy-preserving manner that can spur innovation in the local ecosystem. However, caution regarding the market's adoption of these tools being dependent on the right set of incentives was also raised.

e. **Accountability and Regulation**: The overarching view on this front was that while AISIs have an important role to play, they cannot operate in silos and must align with the country's broader AI governance and regulatory agenda.

   The key question is whether an AISI would be able to legally enforce the commitments made by organizations and whether or not an AISI would have the ability to take action in the case of a market failure. This was particularly important in countries that lack adequate regulations.

   On the issue of AI regulation specifically, one of the key suggestions raised was applying existing regulatory frameworks to AI use cases and relying on market mechanisms to address risks rather than creating new standalone regulations. Another point discussed was that only those specific entities that cause harm, or "negative externalities," should bear the costs instead of society as a whole. Additionally, empowering sectoral regulators and encouraging self-regulation was also discussed as a way forward.

## 4. India's Proposed AISI

During special remarks, it was revealed that India's MeitY had convened a closed-door meeting on October 7 (two days prior to this workshop) to discuss the mandate and structure of an AISI for India.

According to a senior official, the consensus in the meeting was that India's AISI would be a multistakeholder and interdisciplinary body that would operate on a "hub-and-spoke" model. It would not be a regulatory body but would issue guidance that could inform future regulation.

Specifically, it would develop risk mitigation methods for various use cases, identify certain "no-go areas," prioritize human oversight, and promote innovative tools such as watermarking while closely tracking other policy issues such as job displacement due to AI.

According to the official, the Indian AISI would be a part of the global network of AISIs and continue its engagement on AI governance at other forums, including the G20, AI Safety Summits, the Global Partnership on AI, the Organisation for Economic Co-operation and Development, and the UN. The government expects to set up the AISI by the end of 2024 and is seeking feedback from select stakeholders.

# Disclaimer

**End of Document.** All contents within this document remain confidential and are intended solely for the internal reference of the participants at the Global Technology Summit Innovation Dialogue 2024. Unauthorized sharing or distribution is prohibited.

# Carnegie India

Carnegie India, founded in 2016 and based in New Delhi, is part of a robust global network that includes over 170 scholars. The center focuses primarily on three interrelated topics: technology and society, political economy, and security studies. Led by Indian experts with decades of policy experience, Carnegie India engages with governments, policymakers, academics, students, industries, practitioners, and civil society to provide insightful and fresh analysis on India's pressing challenges and its rising global role.

# CONFIDENTIAL

## CO-HOST

Policy Planning & Research Division
**Ministry of External Affairs**
**Government of India**

## PHILANTHROPY PARTNER

TATA TRUSTS

## DIALOGUE PARTNERS

Microsoft
salesforce
Meta
Google

## KNOWLEDGE PARTNERS

CO-DEVELOP
BILL & MELINDA GATES foundation
EkStep
Center for Digital Public Infrastructure
CENTRE FOR OPEN SOCIETAL SYSTEMS
FUTURE OF PRIVACY FORUM


CarnegieIndia.org