

DEMOCRATISING AI

TOWARDS OPEN,
DECENTRALISED
AI ECOSYSTEMS



Basu Chandola and Anirban Sarma

Editors

DEMOCRATISING

AI

**TOWARDS OPEN,
DECENTRALISED
AI ECOSYSTEMS**

© 2026 Observer Research Foundation. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means without permission in writing from ORF.

Attribution: Basu Chandola and Anirban Sarma, Eds., *Democratising AI: Towards Open, Decentralised AI Ecosystems*, Observer Research Foundation, February 2026.

ISBN: 978-93-49061-03-3

ISBN Digital: 978-93-49061-93-4

Editorial and Production Team: Vinia Mukherjee, *Editor and Producer*; Monika Ahlawat and Meryl Mammen, *Assistant Editors*; Rahil Miya Shaikh, *Design and Layout*.

Contents

INTRODUCTION	7
I. DECENTRALISING THE AI STACK COMPONENTS	
AI's Potential to Power Digital Public Infrastructure <i>Gaurav Sharma</i>	12
Reimagining India's AI Stack <i>Geetha Raju</i>	23
Democratising the AI Stack: Technical Exploration and Reflections <i>Sachin Kumar</i>	41
II. UNLOCKING SECTORAL APPLICATIONS	
AI in Agriculture: Connecting Urban Innovations with Rural Needs <i>Diwakar Kumar</i>	58
Democratising AI: The International Trade Perspective <i>Shailja Singh</i>	73
AI-Driven Healthcare Transformation: India's Approach to Open Innovation <i>Viola Savy Dsouza and Angela Brand</i>	85

III. RESPONSIBLE GOVERNANCE OF OPEN AI MODELS

**The Licensing Landscape for Responsible,
Open-Source AI** 98
Mohit Chawdhry

**The Global Challenges of Ensuring Ethical Regulations,
Governance, and Responsible AI** 112
Tanmay Agrawal

**Towards the Responsible Governance of
Open-Source AI** 121
Vibhav Mithal

IV. ENVISIONING A DECENTRALISED AND OPEN AI ECOSYSTEM

**Exploring Non-Technical Pathways to Building a
Decentralised and Open AI Ecosystem** 138
Aishani Rai and Gautam Misra

**Towards a Responsible AI Framework for
Open-Source AI Systems** 149
Kamesh Shekar and Meemansa Agarwal

Envisioning an Open AI Ecosystem in India 161
Sukriti

Introduction

*Basu Chandola and
Anirban Sarma*

Over the past decade, India has demonstrated what inclusive digital transformation can achieve. From driving digital financial inclusion and powering the world's largest vaccination programme to enabling secure e-commerce and strengthening direct benefit transfers, the country's digital public infrastructure (DPI) has shown how technology can serve citizens at scale and create public value. The India Stack, built on open standards, interoperability, and public-private collaboration, has become a global reference point for how digital ecosystems can unlock innovation while expanding access.

As India prepares to lead the Artificial Intelligence (AI) revolution, it is committed to building on this legacy. It envisages AI as “a big tool to solve many problems simultaneously”, one that can drive economic growth, strengthen public services, and address social challenges while containing the associated risks.¹ Boosting the accessibility of the technology and ensuring that no single player has a monopoly over it are two critical priorities of the strategy. Similar to the development of DPI, India aims to create a model in which the government invests in platforms, enabling everyone to use the technology to innovate, develop, and deliver products and services in a competitive and collaborative manner.²

This approach aims to uphold competition and ensure interoperability, promote the use of open-source technologies, install sufficient guardrails, and enhance transparency, inclusion, and collaboration.³ Offering AI-as-a-service to its citizens will help accelerate its adoption and foster innovation in this sector. India seeks to create a platform where computing power, high-quality datasets, and a common protocol and framework will be available for all.⁴

The IndiaAI Mission is currently building a comprehensive AI ecosystem that encourages innovation by democratising computing access, enhancing data quality, developing indigenous capabilities, attracting top talent, enabling industry collaboration, providing startup risk capital, ensuring socially impactful projects, and promoting ethical AI.⁵ It is guided by the fundamental principles of openness and accessibility across cloud platforms to foundational models to unlock public value and prevent the vendor lock-ins that are common in the AI space.

India firmly believes in this vision of an open and decentralised AI ecosystem. Indeed, democratising AI is a powerful pathway to improving millions of lives and transforming sectors. It is also a force for accelerating global progress towards the Sustainable Development Goals (SDGs).⁶

Through the India AI Impact Summit’s ‘Democratizing AI’ Working Group,⁷ India is helping lead the conversation on equitable access to foundational AI resources. The overall objective is to promote inclusivity, strengthen cooperation across the Global South, and ensure that the benefits of AI are shared widely and sustainably around the world.

The notion of democratising AI has multiple dimensions. It speaks not only to who can access and use the tools, but also to who has the resources and skills to build and adapt them, as well as who is included in the processes that will shape its growth trajectory. In essence, democratisation involves widening access to AI use, broadening people's participation in its development, and embedding inclusivity, accountability, and transparency in AI governance.⁸ Democratisation is not merely a technological aspiration but a societal imperative. It is central to building trust in digital systems, advancing equity, and ensuring that the transition to an AI-powered world is grounded in fairness and public value.⁹

This volume draws on the insights from a workshop organised by the Observer Research Foundation in Kolkata in February 2025, on the theme 'Democratising AI: Towards Open, Decentralised AI Ecosystems'. The chapters are based on the papers presented by leading Indian AI scholars during the workshop; they have been grouped according to four interlinked themes:

- Decentralising the individual components of the AI stack
- Unlocking downstream sectoral applications
- Responsible governance of open AI models
- Envisioning a responsible and open AI ecosystem

This edited volume, like the workshop that inspired it, reflects a shared belief that the promise of AI cannot be fully realised through isolated proprietary systems. Instead, real public value lies in building open, interoperable, and decentralised AI infrastructure. Wide-ranging efforts must be made across all stakeholder pools to retain AI as a public good that is governed transparently and aligned with public interest.

By anchoring these discussions around decentralisation, sectoral impact, responsible governance, and ecosystem-level thinking, this volume represents a comprehensive roadmap for how AI can evolve in a future-ready manner.

Basu Chandola is Associate Fellow, Centre for Digital Societies, ORF.

Anirban Sarma is Director, Centre for Digital Societies, ORF.

Endnotes

- 1 Nivash Jeevanandam, "Global IndiaAI Summit - Minister Ashwini Vaishnaw Highlights AI's Transformative Potential and Risks at Global IndiaAI Summit," *IndiaAI*, July 3, 2024, <https://indiaai.gov.in/article/minister-ashwini-vaishnaw-highlights-ai-s-transformative-potential-and-risks-at-global-indiaai-summit>.
- 2 Ministry of Electronics and IT, Government of India, <https://pib.gov.in/PressReleasePage.aspx?PRID=2030512#>.
- 3 Leslie D'Monte, "How India Plans to Make AI Accessible for All," *LiveMint*, July 5, 2024, <https://www.livemint.com/ai/how-india-plans-to-make-ai-accessible-for-all-11720155629511.html>.
- 4 Rashaad Ather, "India to Create UPI-like Public Platform for AI: IT Minister Ashwini," *StartUpPedia*, <https://startuppedia.in/trending/trending/india-to-create-upi-like-public-platform-for-ai-it-minister-ashwini-4793706>.
- 5 "IndiaAI," <https://indiaai.gov.in/>.
- 6 Narendra Modi, "Opening Address by Prime Minister Shri Narendra Modi at the AI Action Summit, Paris," (speech, Paris, February 11, 2025), https://www.mea.gov.in/Speeches-Statements.htm?dtl/39020/Opening_Address_by_Prime_Minister_Shri_Narendra_Modi_at_the_AI_Action_Summit_Paris_February_11_2025v.
- 7 "Democratizing AI Resources Working Group," <https://impact.indiaai.gov.in/working-groups/democratizing-ai-resources>.
- 8 Ta Lin, "Democratizing AI' and the Concern of Algorithmic Injustice," *Philosophy and Technology* 37 (2024) 103, <https://doi.org/10.1007/s13347-024-00792-2>.
- 9 R.R. Taylor, J.W. Murphy, W.T. Hoston et al., "Democratizing AI in Public Administration: Improving Equity Through Maximum Feasible Participation," *AI & Soc* 40 (2025): 3653–3662, <https://doi.org/10.1007/s00146-024-02120-w>.



I.

**Decentralising
the AI Stack
Components**

AI's Potential to Power Digital Public Infrastructure

Gaurav Sharma

New-age artificial intelligence (AI) avatars such as chatbots are becoming more popular in the Global South due to their utility, versatility, and human-like responses across domains and thematic areas. The January 2025 launch of China's DeepSeek¹ (an open-source, economical, and customisable model of chatbot) has caused even greater excitement about developments in the field. The most intractable challenges in Global South economies—among them, poverty, malnutrition, climate change, socio-economic inequalities—and the urgent need for digital inclusion with equitable development have

prompted these countries to seek and utilise the power of AI in public service delivery with the hopes of driving faster social impact.

Many countries have already adopted Digital Public Infrastructures (DPI) for faster, transparent, and efficient public service delivery. During India's presidency of the G20 in 2023, there was consensus on the fundamental role of DPI in improving people's lives and transforming governance. India's G20 task force report on DPI said, in part, "A DPI approach to craft the growth of AI would involve publication of open datasets via open API to train AI/ML models, creation of reusable AI toolkits and publication of open models."² Successful DPI deployment at population scale include: India (Aadhaar digital identity and Unified Payments Interface or UPI); Estonia (X-Road solution for data sharing, e-Estonia platform for government services); Brazil (Pix Fast Payment System); Thailand (National Digital ID, PromptPay digital financial transactions DPIs); and Togo (using DPI to enrol digital IDs with India, and collaborating with Estonia on data exchanges).

The integration of AI holds promise to further advance adoption, efficiency, and user-centric public service delivery—as the creation of DPI solutions for furthering financial inclusion and digital identity does not guarantee efficient usage. For example, the World Bank's Global Findex Database reported that more than one-third of bank account holders in India had an inactive account—AI tools can be tapped to plug these digital inadequacies.³

This rest of this article outlines the merits of AI and its potential to integrate with the success of India's DPI stack in addressing the prevalent digital and developmental challenges in countries of the Global South. It highlights how AI and DPIs together have the potential to power the future of digital economy growth. The article outlines specific recommendations towards facilitating a safe and responsible AI unification with DPIs.

DPIs and the New Age of AI

DPI, as defined by the G20 Digital Economy Ministers Meet (DEMM),⁴ "is a set of shared digital systems that should be secure and interoperable, and can be built on open standards and specifications to

deliver and provide equitable access to public and/or private services at societal scale and are governed by applicable legal frameworks and enabling rules to drive development, inclusion, innovation, trust and competition and respect human rights and fundamental freedoms.”

Perhaps the most comprehensible avatar yet of new-age AI technologies are the conversational AI agents—referred to as AI chatbots—such as ChatGPT. AI chatbots have become widely popular due to their ability to crunch large amounts of versatile datasets (voice, text, audio), and generate new knowledge and informational material in response to simple user prompts.⁵ Global South economies can harness the potential of AI to find solutions to their development challenges.

AI chatbots showcase the potential to learn, understand, and integrate complex socio-cultural datapoints, and provide personalised advisory and consumable knowledge content services, such as the Kumbh Sah’Al’yak^a and Ama Krushi^b chatbots that are serving diverse population groups. Chatbots have thus demonstrated strong use-cases for the uptake of AI technologies in Indian languages and local dialects. It is now natural to imagine the possibilities of embedding AI into the DPI ecosystem to open a new-age phenomenon of personalised public service delivery for citizens.

There are two questions that need to be addressed: Does AI hold potential to solve population-scale challenges such as water conservation, malnutrition, poverty, primary healthcare access, and climate change? If so, how can DPIs plug into AI technologies to achieve last-mile connectivity? These issues become more pertinent as there are at least 500 million people in India who do not have

-
- a Kumbh Sah’Al’yak is an AI-powered chatbot launched for Prayagraj Mahakumbh 2025 to enhance the pilgrimage experience. Mahakumbh is celebrated every 144 years and is the biggest Hindu festival with more than 450 million participating in bathing rituals. See: <https://pib.gov.in/PressReleasePage.aspx?PRID=2101679>
- b Ama Krushi is a free agricultural service provided by Odisha’s Department for Agriculture and Farmer’s Empowerment. See: <https://amakrushi.in/en/>

a digital identity nor access to financial services. The following two sections outline the challenges in scaling AI technologies in the social sector space in the Global South, and the advantages that the DPI tech stack presents towards a converging promise of DPI powering AI, and vice-versa.

AI Adoption and Societal Challenges in the Global South

Network Connectivity

The successes of AI weigh on credible, high-speed bandwidth network connectivity; the Global South, however, suffers from uneven networks in rural and semi-urban areas. Internet penetration in India stands at 52.4 percent.⁶ India's BharatNet,⁷ Digital Bharat Nidhi, and PM-WANI⁸ initiatives aim to provide high-speed internet connection to rural regions, especially the remote villages. However, distributed models of connectivity such as the state-led model, private sector model, and Central Public Sector Undertakings (CPSU) model, provide a bouquet of WiFi services and create a complex set of scattered internet service platforms without streamlined, consistent, and dependable internet services in rural and remote areas. AI-embedded broadband services can enable faster and 24x7 access by integrating internet services and reducing prompt response times. It is not about mobile coverage, but reliable 24x7 mobile connectivity over 4G/5G networks. This is a prerequisite for the real-time usage of AI applications such as farmer advisory services and medical diagnostics.

The need for broadband connectivity demands coverage not only at the gram panchayat centres, but also at farmlands where there is real need for a service advisory. Speed is the Achilles' heel in powering AI services at scale, and reliable public sector delivery depends on promptness and real-time information exchange—as delays in healthcare information can cost lives, and delays in farming information impact crop yields. Real-time information is a key to elevating socio-economic impacts and minimising financial and physical losses. The sparse integration of public datasets, coupled with distributed networks with varying network speeds are poised to complicate AI integration into public service DPIs. Unreliable internet services would risk reinforcing existing digital inequalities and creating AI dependencies on tech intermediaries.

High-quality socio-cultural datasets

The lack of high-quality datasets incorporating socio-cultural, linguistic, and traditional diversity of the Global South poses challenges for the deployment of AI. The absence of community-level social practices, norms and customs can only lead to low-quality or bad output-based (“hallucinating”) AI solutions. India’s diverse demography and rich linguistic heritage present a challenge for AI as the digital footprint is missing for these diverse datasets. Languages,⁹ food, cultures, and dressing styles change every few kilometres across the country, and AI models must be trained on these socio-cultural datasets to provide customisable and individualised services. Furthermore, the fragmented, detached, and non-interoperable public datasets in government departments stymie the creation of harmonised and systematic datasets required to make and train Indic AI models.

Additionally, the public datasets available are not of high quality and demand updating. They also need to be machine-readable: a fundamental requirement for use by AI models. The problem is particularly acute for datasets pertaining to vulnerable communities and marginalised groups, as they are limited in availability in the digital format, if at all they are existent. Thus, large, community-scale efforts and voluntary citizen contributions are needed to collect and create high-quality socio-cultural datasets in local languages and dialects with a higher level of data processing infrastructure. The need is also for the automation of inter-departmental data integration across socio-impact supply chains such as agriculture and education, as each individual generates data with different formats, structures, and diversity of representation for similar characteristics.

The ‘localisation’ challenge for AI

AI models are trained on internet datasets and lack a local context of cultures and social norms. Lack of local context creates hallucinations,¹⁰ biases, and inaccurate results. Localisation is critical for public service delivery in the Global South and must start from the design stage in the AI pipeline to incorporate all aspects of societal inclusivity.¹¹ Incorporating ‘localisation’, such as tailored human values specific to cultural, religious, and daily routine norms is critical. For example: to

answer a personalised query regarding agriculture in a particular locality, an AI chatbot requires highly trained models in the local dialect, with the ability to integrate the multilinguality of speech. The inclusion of localised cultural and societal contexts and norms with an inclusive user interface could be highly beneficial to individual users.

Further, if farmers represent a low-literacy population group, the chatbot will have to gauge the true interpretation of the nature of the query as it can often be morphed in the emotional and cultural tone of the question. Localisation presents a serious challenge in terms of identifying cultural contexts, codification, as well as training and customisation for the local population group in question. AI-integration into public service delivery without localised datasets and diversified knowledge from the Global South regions remains incomplete in itself.

Integrating AI with DPI and Data Pipeline: A Mutually Beneficial Proposition

The DPI architecture, data pipeline, and new-age AI technologies present a mutually beneficial confluence. The success of DPIs is rooted in interoperability and open standards, independent of geography and demography, and good data governance frameworks. The scale and societal impact of DPIs has been defined by the mobile revolution and the advancements in internet service infrastructure, such as low-cost bandwidth, hardware, and efficient data pricing. AI could take advantage of DPI frameworks and dataset pipelines, and integrate language technologies and computer vision to aid the personalisation of public service delivery, and invariably catalyse the next phase of DPI expansion. There are four points of confluence:

1. **Public investments in AI to power DPI capacities:** Investments made in natural language processing (NLP)^c to help understand and process languages and dialects, such as the BHASHINI^d programme

c NLP models are essential for processing diverse languages, especially in multilingual countries like India. Multilingual AI solutions help reduce the linguistic barrier to public services. See: S. Arora et al., "Challenges in Natural Language Processing for Multilingual Societies," *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2021.

d BHASHINI is an Indian government project developed by MeitY under its National Language Translation Mission (NLTM) to enable all Indians easy access to the internet and digital services in their own language and increase the content in Indian languages. See: <https://bhashini.gov.in/>

in India, and similar efforts by the Indian Institute of Sciences (IISc), Bangalore^e and Google's Vaani^f can support the development of Indic AI models and use-case-specific solutions. These projects can power customised AI Indic models and language datasets into social welfare digital programmes and public services. They also boost inclusivity and accessibility by providing customisable and individualised information in local dialects for the convenience of the those with lower literacy levels. India's investment in safe and trusted AI development will further enhance DPI capacities by building digital trust in the use of AI language technologies in public service delivery.

- 2. DPI datapoints to power AI pipeline, and vice-versa:** AI's strengths lie in the real-time analyses and processing of enormously complex and unstructured datasets; real-time predictive modelling and analytics and conveying the responses by undertaking human-like conversational tasks such as question and answer, reasoning, and search through a single interface, such as voice or text prompt. This can help scale DPIs and, in turn, DPI datapoints can be fed into a single AI data architecture layer. This would allow AI models to constantly learn and churn out meaningful, updated, and improved knowledge inferences that enable real-time information sharing and decision making.

For example, the use of AI in digital payments can refine the process of vetting the creditworthiness of an individual, based on alternative datapoints such as utility payments, e-commerce digital transactions, health data statistics or agricultural produce, thus making micro loans accessible for individuals in the unorganised

e Speech recognition in agriculture and finance for the poor in nine Indian languages by the Indian Institute of Sciences (IISc). See: <https://respin.iisc.ac.in/>

f Project Vaani, by IISc, Bangalore and ARTPARK, is expected to capture 150,000 hours of speech, part of which will be transcribed in local scripts, to ensure linguistic, educational, urban-rural, age, and gender diversity to propel language AI technologies and content for an inclusive Digital India. See: <https://vaani.iisc.ac.in/>

sector. The Data Empowerment and Protection Architecture (DEPA) allows third parties to use consented financial data for the stated purpose. The AI data layer on top can facilitate the advanced use of DEPA-consented datasets aligned to a user's profile from multiple social scenarios, such as healthcare data, education, and other DPI-enabled services. Thus, the congregation of multiple human-centred datasets can continuously feed into a single AI-interface data architecture layer, creating a human-reinforced learning loop with sharper outputs and reduced hallucinations.

- 3. AI to power efficiency and innovation in DPIs:** The reasoning and verification ability of AI models across domains could be utilised in amalgamating DPIs such as digital identity, digital payments and consent analysis into a single public service delivery architecture. For example, bringing together DPIs like Ayushman Bharat Digital Mission (ABDM),^g Agri Stack,^h and DIKSHAⁱ into an integrated public service suite broadens the scale and scope of DPI citizen engagement, and provides a wide range of advisory and information services across sectors through a single user interface. AI can accelerate and maximise the impact of existing DPIs by enabling alternate digital interfaces such as voice and real-time biometrics to access DPI services. The incorporation of public datasets onto a common layer of data architecture can also attract AI startups and SMEs to develop innovative DPI services on top. In addition, AI integration with other emerging technologies such as robotics, drones, and the internet of things (IoT) has the potential to propel economic growth by curating inter-sectoral public services over an integrated pipeline of datasets built upon optimised and efficient DPI architecture. AI integration with DPIs can advance context-specific public service availability, power innovation, and demonstrate social impact growth.

g Ayushman Bharat Digital Mission is aiming to develop the backbone necessary to support the integrated digital health infrastructure of the country. See: <https://abdm.gov.in/>

h Agri Stack is a digital agriculture platform developed by the Government of India to integrate all farmer-related data in one place. See: <https://agristack.net/>

i DIKSHA: Digital Infrastructure for knowledge sharing platform. The platform can be used to design and run programs for teachers, learners, and administrators. See: <https://diksha.gov.in/>

- 4. AI to take advantage of DPI governance framework:** Devising ethical and responsible AI standards that effectively address the spectrum of civil society concerns regarding public technologies demands robust governance models to ensure privacy, secure data-sharing, transparency, and citizen trust. In the essay, ‘Machines of Loving Grace: How AI Could Transform the World for the Better,’¹³ Dario Amodei (CEO of Anthropic) mentions that “both the AI companies and the developed world need to do their part as the moral imperative is too great and demands collective effort.”

To this effect, good governance demands monitoring, evaluation, and ensuring accountability of AI tools. AI technology can be integrated into the DPI governance ecosystem for solutions like digital payments or the Unified Payments Interface (UPI), which is under the supervisory control of the Reserve Bank of India (RBI), and adheres to national banking and financial frameworks. Thus, AI integration into the DPI governance ecosystem can be useful in cementing digital trust in the use and adoption of AI at scale. India’s IndiaAI mission approach^j has already outlined a commitment to drive equitable and responsible governance through its safe and trusted AI pillar.

Conclusion

The countries of the Global South are looking to adopt AI technology to scale up equitable digital socio-economic development. The successes of DPIs have demonstrated the power of technology in enabling public sector delivery in many developing economies. Imbibing AI can further power DPIs towards customised and contextualised public services. DPI frameworks have shown the promise of scale but have yet to completely overcome the challenges of equitable digital access, inclusion of socio-cultural norms, and targeted service delivery.

j The IndiaAI Mission aims to build a comprehensive ecosystem that fosters AI innovation by democratising computing access, enhancing data quality, developing indigenous AI capabilities, attracting top AI talent, enabling industry collaboration, providing startup risk capital, ensuring socially impactful AI projects, and promoting ethical AI. See: <https://indiaai.gov.in/>

AI holds the potential to transform this landscape. In turn, AI's need for a big data pipeline can be met by DPIs, and in addition, DPIs can provide avenues for ease of AI public adoption, effectively gaining their trust. Thus, the integration of AI with DPIs can provide avenues for responsible AI governance.

AI thinking in public service delivery can add value to new DPIs in the social sector. AI technologies hold the potential to streamline and customise DPI service by integrating public interest design principles, end-user needs, and local requirements. The government must remain the digital-trust bearer for all new-age AI technologies and encourage AI and DPI innovation through a trusted regulatory framework. AI integration with DPIs holds merit for an equitable digital transformation in the Global South and opens up a new world of opportunity for public service delivery. It remains to be seen how AI technologies integrate with DPI architecture and vision, and gain from each other in providing a credible, equitable, and responsible digital service delivery ecosystem built on digital trust.

Gaurav Sharma is a Research Associate with the TECHtonics project funded by the European Research Council (ERC) at Humboldt University, Berlin; a Visiting Researcher at the Centre for Digital Governance at Hertie School, Berlin; and a Policy Advisory Fellow at the Centre for Responsible AI (CeRAI.in).

Endnotes

- 1 Kelly Ng et al., “DeepSeek: The Chinese AI App that has the World Talking,” BBC, February 4, 2025, <https://www.bbc.com/news/articles/c5yv5976z9po>; Aditya Soni and Zaheer Kachwala, “DeepSeek’s Low-cost AI Spotlights Billions Spent by US Tech,” *Reuters*, January 29, 2025, <https://www.reuters.com/technology/artificial-intelligence/big-tech-faces-heat-chinas-deepseek-sows-doubts-billion-dollar-spending-2025-01-27/>
- 2 Ministry of Finance, Government of India, <https://dea.gov.in/sites/default/files/Report%20of%20Indias%20G20%20Task%20Force%20On%20Digital%20Public%20Infrastructure.pdf>
- 3 Asli Demircuc-Kunt et al., *The Global Findex Database, Financial Inclusion, Digital Payments and Resilience in the Age of COVID-19*, World Bank, 2022, <https://openknowledge.worldbank.org/bitstream/handle/10986/37578/9781464818974.pdf>
- 4 “The Global Findex Database, Financial Inclusion, Digital Payments and Resilience in the Age of COVID-19”
- 5 Sumit Kumar Dam, “A Complete Survey of LLM-based AI Chatbots,” Arxiv, November 18, 2024, <https://arxiv.org/abs/2406.16937>
- 6 “Internet Penetration Rate in India from 2014 to 2024,” Statista, <https://www.statista.com/statistics/792074/india-internet-penetration-rate/>
- 7 Digital Bharat Nidhi, Department of Telecommunications, Ministry of Communications, Government of India, “Bharatnet Project,” <https://usof.gov.in/en/bharatnet-project>
- 8 Research Unit, Press Information Bureau, Government of India, <https://static.pib.gov.in/WriteReadData/specificdocs/documents/2024/dec/doc20241226477201.pdf>
- 9 Hari Narayan, “India, a Land of Many Tongues,” *The Hindu*, August 7, 2017, <https://www.thehindu.com/thread/arts-culture-society/india-a-land-of-many-tongues/article19445187.ece>
- 10 IBM, “What are AI Hallucinations,” <https://www.ibm.com/think/topics/ai-hallucinations>
- 11 S. Agarwal, “Designing AI for Inclusivity: Addressing the Needs of Low-Literacy Populations,” *AI & Society Journal*, 2020.
- 12 Vikas Kathuria, “Data Empowerment and Protection Architecture: Concept and Assessment,” *ORF Issue Brief No. 487*, Observer Research Foundation, August 2021, <https://www.orfonline.org/research/data-empowerment-and-protection-architecture-concept-and-assessment>
- 13 Dario Amodei, “Machines of Loving Grace: How AI Could Transform the World for the Better,” October 2024, <https://darioamodei.com/machines-of-loving-grace>

Reimagining India's AI Stack

Geetha Raju

T

he rapid transformation of India's AI ecosystem is attended by unprecedented opportunities, including innovation, economic growth, and enhanced public service delivery. While AI holds immense potential for societal advancement, it imposes challenges, such as biased outcomes, data privacy, job displacement, and, more importantly, ensuring safe, ethical, and responsible use. Moreover, disparities in access to infrastructure, resources, and funding could exacerbate prevailing inequalities and hinder inclusive development.

The current concentration of data, algorithmic models, and compute facilities within big technology companies and research institutions risks marginalising smaller entities like startups and individual innovators. This also limits innovations at the grassroots. This article draws inspiration from India's Digital Public Infrastructure (DPI) model to argue for a decentralised AI architecture as a compelling pathway to navigate the challenges arising from technological monopolies and unequal resource distribution. This approach leverages DPI principles like interoperability, open-source collaboration, and decentralisation to foster an inclusive AI ecosystem in India.

Five components are required to develop and deploy an AI system:

- **Data:** This enables an AI model to learn, predict, or generate outputs
- **Models/AI algorithms:** Mathematical frameworks that interpret data and perform targeted tasks like prediction, classification, clustering, and generation
- **Compute Infrastructure:** Hardware infrastructure like GPUs and cloud servers that train, test, and deploy AI models
- **Tools and frameworks:** Libraries and platforms that enable AI developers to develop and deploy AI models and systems effortlessly
- **Silicon chips:** A hardware component that powers AI compute infrastructure

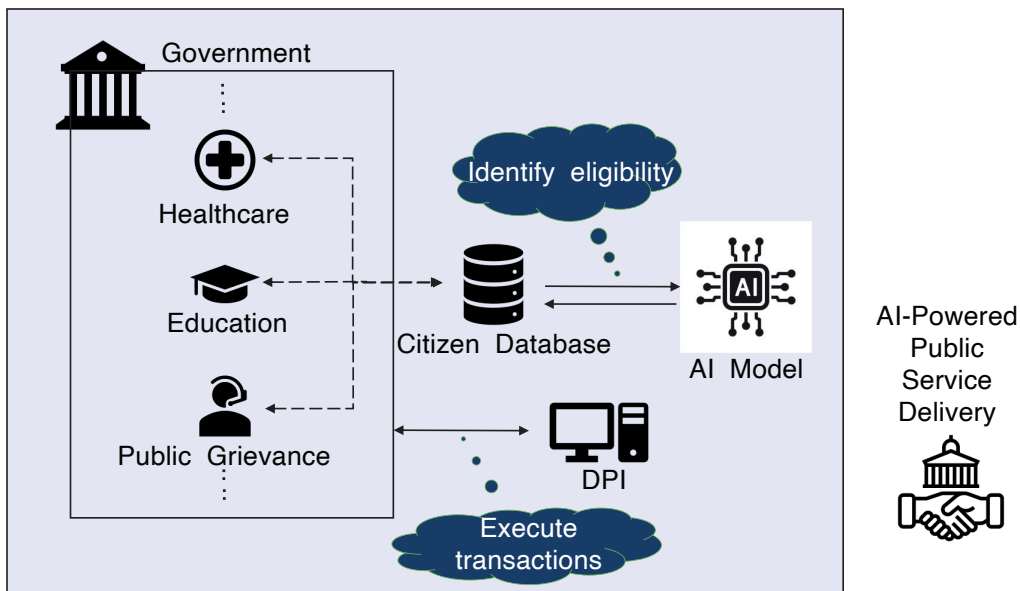
Most of these essential components are centralised in India's AI architecture: data lies with Big Tech companies and governments, state-of-the-art models are developed and owned by technology giants, and compute facilities are concentrated at elite research and development centres or well-funded organisations. This concentration of AI development creates technological monopolies, stifles innovation, and exacerbates inequalities in access to data, computing, and applications required for AI development, deployment, and monitoring.¹ Rather than relying on a centralised entity, it is thus essential that India adopts a decentralised approach that enables the development and deployment of AI models in a distributed fashion.

Integrating AI in India's DPI

India has already witnessed success in establishing a decentralised DPI^{2,3}—including the Unified Payments Interface (UPI),⁴ the Aadhaar card, the National Health Authority's Ayushman Bharat Digital Mission (ABDM), National Language Translation Mission (NLTM)'s Bhashini,⁵ and RBI-regulated⁶ account aggregators—that has allowed for the democratisation of various public services.

This DPI Stack can serve as a blueprint for reimagining India's AI stack. Extending the DPI principles of interoperability, open-source, collaboration, and decentralisation to AI can lead to its democratisation. Further, integrating AI in DPI for government and public services, particularly in India's priority sectors such as healthcare and education, where AI can be utilised to identify eligible citizens and AI-powered DPI can facilitate the seamless delivery of services or incentives to these individuals, presents numerous opportunities to address India's unique needs and challenges.

Figure 1. Improving Public Service Delivery by Integrating AI in India's DPI



Source: Author's own

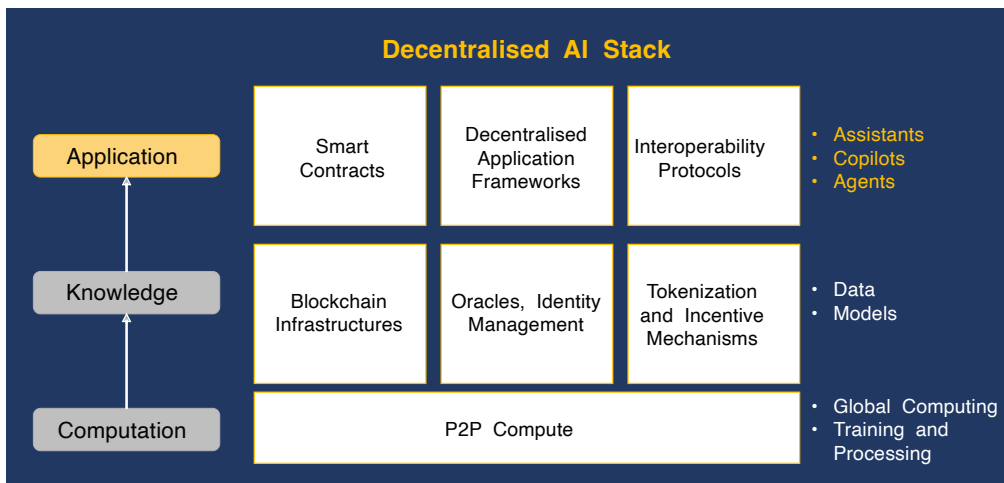
Decentralisation in AI ecosystems, too, has already been put into practice in some parts of the industry. The most essential aspects of decentralised AI are ensuring equally distributed access to computational resources, open-source, high-quality datasets,⁷ AI models, and applications for a wide range of stakeholders. The decentralised AI stack by SwarmZero,^a for instance, consists of three interdependent layers:

- 1. Knowledge layer:** This includes data and models stored in a decentralised manner. Decentralised data storage solutions offer enhanced privacy by implementing techniques such as encryption and secure multi-party computations over encrypted data. This ensures that data and models are kept private while collaboratively shared and accessed across the network.
- 2. Computation layer:** This includes compute and storage components. It provides the necessary capacities by enabling unused resource sharing through peer-to-peer (P2P) networks. This approach reduces reliance on centralised data centres and GPUs.
- 3. Application layer:** This includes application programming interfaces (APIs), Command Line Interface (CLIs), agents, assistants, and co-pilots that allow developers to build domain-specific AI solutions on top of the shared data, model, and compute capacities.

Additionally, decentralised AI also includes energy generation and distribution, silicon chips for hardware, and other external factors that are hidden in the AI lifecycle but impact AI diffusion and competition globally.

a SwarmZero is an early-stage decentralised AI platform designed to empower AI creators and AI providers in developing, deploying, and monetising AI agents. By leveraging blockchain technology and a user-friendly interface, SwarmZero aims to democratise AI development, fostering a transparent and collaborative ecosystem where AI agents can function autonomously and collaborate in “swarms” to tackle complex tasks.

Figure 2. Decentralised AI Stack



Source: SwarmZero⁹

The Current Landscape of Decentralised AI

Decentralised AI systems are built with numerous components to make them secure, efficient, reliable, and collaborative. These range from foundational technologies like blockchains, smart contracts, oracles, to advanced computational models facilitated by federated learning, coprocessors, peer-to-peer networks, zero-knowledge ML (zkML), trusted execution environment ML (teeML), optimistic ML (opML), model predictive controls (MPC), and homomorphic encryption methods. To broaden our understanding, the following section will delve into a few key technologies such as zkML and teeML—crucial for addressing the most pressing challenge of data privacy and verifiable computation within the decentralised AI ecosystem.

Zero-knowledge ML¹⁰ combines ML and zero-knowledge proofs (ZKPs) to arrive at a decision and prove its correctness, respectively. It allows the execution of ML models on remote hardware while generating cryptographic proofs that validate the correctness of computations without revealing sensitive model weights. ZKPs¹¹ enable one party to prove to another that a given statement is true without revealing any information beyond the validity of the statement itself. Thus, the

zkML framework ensures that the weights of an AI model remain confidential, thereby protecting sensitive information from unauthorised access/unintentional disclosures. Overall, incorporating zkML offers the advantage of model weight privacy, cryptographic proofs, and validation of trust in trustless environments, which are key to the decentralised AI lifecycle.

Yet, limitations like computational overheads, memory constraints, scalability issues (linearly growing proof sizes), energy costs, increased latency, and lack of fault tolerance make zkML implementations unsuitable for certain contexts, like developing large language models (LLMs).¹² Researchers have proposed alternative techniques, such as sampling consensus¹³ and optimistic ML (opML),¹⁴ to address the limitations of zkML. These techniques provide a more efficient means of developing and deploying AI solutions, especially in terms of cost and reliability. Another option is the split learning approach,¹⁵ which divides a model into segments based on the nodes on which it is processed thus preventing full data access at any single point.

Confidential computing (CC)¹⁶ is also an approach that provides a promising solution to bridge the privacy gap. CC leverages hardware-based TEEs^{17,18} to protect data confidentiality and code integrity on remote machines. It provides an isolated and secure environment for sensitive computations to take place without exposing model parameters or data to other participants in the distributed network.

LLMs can also employ other decentralised learning techniques like federated learning, where the model is trained across many local data sources without transferring them. The federated learning environment enhances data privacy and allows the model to learn from broader datasets, aggregate the knowledge to the global model, and distribute it back to the decentralised devices. Decentralising the data and models can also solve the problem of high compute and training data requirements of LLMs.

In this context, the concept of decentralised business and decentralised financial (DeFi) systems,¹⁹ which can operate without intermediaries by using blockchains for transparency and security, has been widely explored and implemented. DeFi models are now expanding beyond

finance into other sectors. In healthcare, too, given its sensitive nature, researchers are working to develop decentralised AI models using a federated learning approach, such as for melanoma-nevus classification.^{20,21} This approach allows each participating hospital to train the AI model locally with their data, ensuring that sensitive patient information remains within the institution, and model weights are periodically aggregated using a weighted average based on the amount of data each hospital contributes. Further, federated learning is being explored for biomedical applications, such as emotion recognition from electroencephalography (EEG) data, to address privacy concerns.

Similar ideas on implementing decentralised identity solutions for healthcare systems are gaining traction too—for instance, Hyperledger Fabric²² to store a patient's health records with privacy protocols, or openPHR Protocols for Personal Health Records (PHR) that divide information into data blocks among participating devices. Experiments have shown that these strategies improve security and privacy in electronically stored and distributed health records.

Decentralised AI Platforms

SAKSHI,²³ a decentralised AI platform, separates data, control, and transaction paths using a 'proof of inference' layer for cryptographic resistance against misbehaviours. This architecture of SAKSHI can be visualised as a layered architecture that includes a service layer (AI inference), a control layer (networks, compute, and storage), a transaction layer (billing and metering), a proof layer (monitor and resolve billing and metering—proof inference, proof ownership, and proof of service delivery), a transaction layer (tokenless and incentives), and a marketplace (price discovery) to efficiently address the challenges in centralised AI services.

SingularityNET²⁴ is a notable example of a decentralised AI marketplace in which buyers and sellers exchange AI services via blockchain and AI agents transact with each other using smart contracts. Innovative features, such as the native token and decentralised governance model, offer a more equitable and collaborative AI ecosystem, and contribute to the robustness and ethical development of AI.

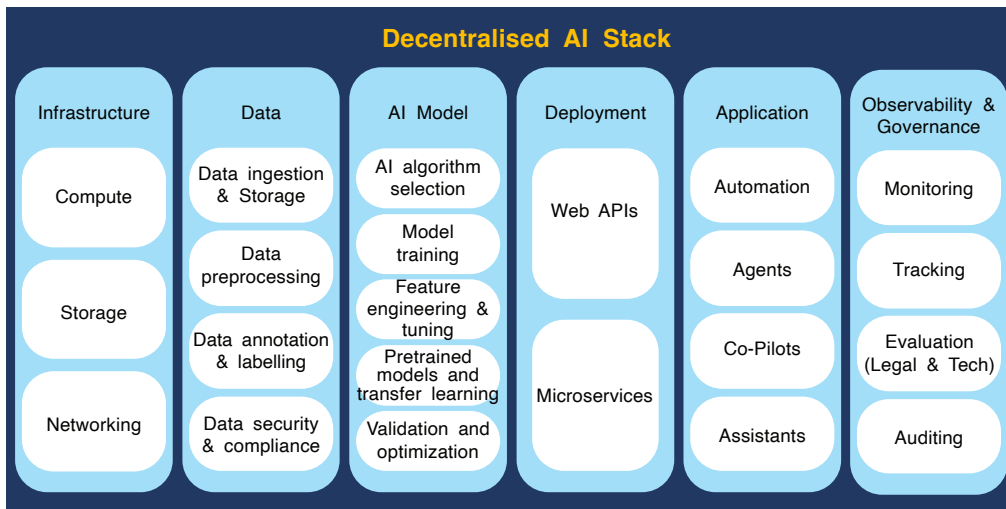
Decentralised AI marketplaces like Bittensor²⁵ promote autonomous AI systems and collaboration. Bittensor offers an ‘intelligence marketplace’ where AI models, (referred to as ‘miners’ or ‘neurons’) compete to provide the best solution to various AI tasks. The bittensor network is structured into specialised subnets focused on AI tasks or domains such as text prompting, image generation, etc. It also has a validator to evaluate the quality of miners and a proof of intelligence (PoI) as a consensus mechanism. This creates a global, open-source ecosystem for AI development.

Enablers of Decentralised AI

Recent breakthroughs in efficiency have made decentralised AI computing more feasible on smaller hardware. This shift is further supported by the development of custom AI chips and software abstraction frameworks, such as MLX and Triton, that reduce dependencies on centralised hardware providers and enable AI workloads to run efficiently on diverse platforms. Additionally, tools like Ocean Protocol²⁶ are pioneering decentralised data exchange protocols, enabling secure and transparent data sharing and monetisation. This approach empowers data owners to control and benefit from their data directly.

Challenges to Transitioning to Decentralised AI

The transition from a traditional AI architecture to a decentralised AI architecture does not only entail replacing a centralised approach with open-source and distributed architectures. It invites a paradigm shift in how AI systems are designed, developed, deployed, managed, and governed. Figure 3 illustrates the various AI components and their corresponding subcomponents and tasks that form the fundamental blocks of a decentralised approach.

Figure 3. Decentralisable Components of the AI Stack

Source: Author's own

The inherent complexity of decentralised approach holds promising potential for AI development and mitigation of risks associated with a centralised approach. However, implementing it also poses systemic challenges, ranging from technical, regulatory, and economic aspects to socio-political ones.

Table 1 summarises some of the challenges presented by the adoption and operationalisation of decentralised approaches.

	Challenge	Description
1.	Technical Challenges	
1.1	Algorithmic fragility in distributed environments	Due to the phenomenon of catastrophic forgetting, ^b algorithms can lose previously learnt information during incremental updates and may encounter unexpected failures when trained in a distributed setup. This issue is often exacerbated in decentralised setups due to factors like data heterogeneity across distributed sources and communication constraints, leading to model drifts in local training environments.
1.2	Scalability issues—increased latency during training and inference tasks	Training tasks of large models like GPT-3 across decentralised GPUs require synchronisation of gradients across thousands of nodes, increasing latency when integrated with real-time applications. Inference tasks might also require more computation while processing task-specific queries.
1.3	Data privacy	Model training and inference use cases involve personal or proprietary information, which, when served without privacy guarantees, can increase model theft or data disclosures.
1.4	Interoperability and Integration Challenges	Data schemas and metadata format mismatches, different model interfaces (proprietary AI vs. open-source AI), and hardware heterogeneity in decentralised networks increase development complexities.
2	Regulatory and Governance Challenges	
2.1	Jurisdictional boundaries	Compliance with different legal regimes when operating on a decentralised AI stack is critical. For instance, data anonymisation requirements differ between the EU’s AI Act and India’s DPDP Act.
2.2	Subjectivity in AI Governance	A lack of centralised oversight across decentralised units leads to increased susceptibility to subjectivity in AI outcomes (bias and fairness issues).

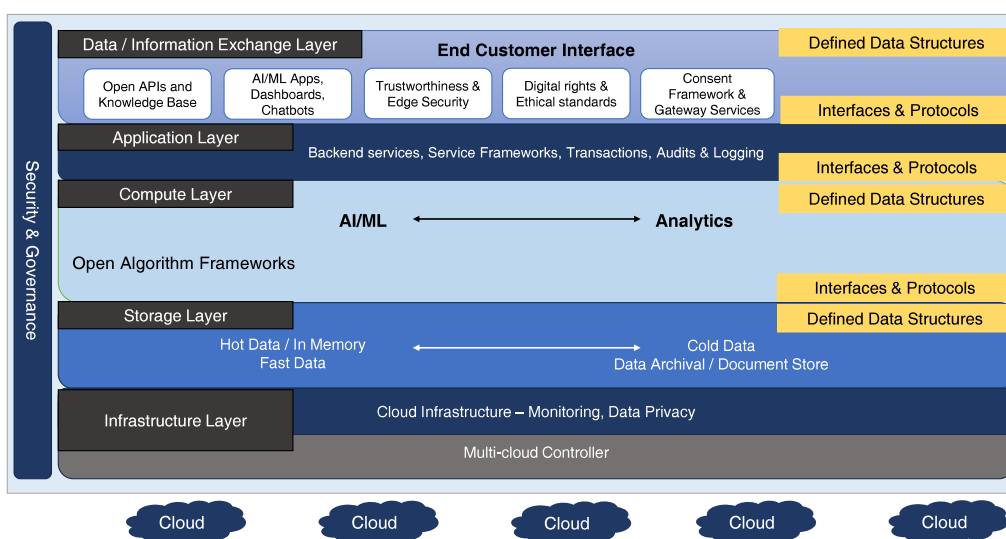
b Catastrophic forgetting: AI model trained to perform Task A will forget Task A when it is later trained on Task B.

3		
Economic Challenges		
3.1	Tokenisation pitfalls and resource concentration	Compute infrastructures are controlled by Big Tech, which will impact vendor lock-in situations.
3.2	Data contribution incentives	Inadequate or no compensation for data contributors can increase the gap in acquiring contextual data to train AI models for a decentralised stack.
3.3	Algorithmic complexities	The complexity of LLMs makes model training and querying very expensive. As the complexity of the algorithms increases, training or querying tasks incur more resources, which eventually increases the cost of establishing decentralised AI stacks.
4		
Environmental Constraints		
4.1	Energy consumption and carbon footprint	Emerging technologies like AI, IoT, and blockchain are energy-intensive. If devices used for decentralised data storage and compute capacity are not optimised for efficiency, it can lead to adverse environmental impacts.
4.2	E-Waste generation	Rapid advancements in AI and blockchain technologies can result in shorter device lifecycles, increasing e-waste generation.

IndiaAI: A Framework for a Decentralised AI Stack for India

Given the numerous challenges facing AI use worldwide, India must develop a safe, responsible, ethical, and trustworthy AI ecosystem that benefits its diverse population. The AI Standardisation Committee of the Department of Telecommunications, Government of India, has proposed a framework for an AI stack²⁹ that addresses the issues and challenges related to AI development and deployment in India's AI ecosystem. The stack, presented in Figure 4, aims to be implementable across all sectors while ensuring critical features of equitable AI like well-defined data structures, data privacy, data protection, data federation, data minimisation, open algorithm frameworks, protocols and interfaces, robust monitoring, auditing mechanisms, ethical values, digital rights, and trustworthiness.

Figure 4. DOT’s Indian AI Stack



Source: *Indian AI Stack*³⁰

DoT’s AI stack is organised into five primary horizontal layers—infrastructure, storage, compute, application, and data—with a dedicated vertical layer for security and governance. This layered approach aims to promote transparency, interoperability, and inclusivity while addressing concerns related to bias, data protection, and model explainability.

The IndiaAI mission³¹ seeks to operationalise India’s AI stack through public-private partnership-based investments and innovations in AI. One key aspect of the mission is the empanelment of 10 cloud service providers,³² which will pave the way for AI researchers and practitioners to access compute resources in a distributed environment at subsidised rates, allowing them to process data and train AI models of their interest. Similarly, the IndiaAI mission also encourages the development of indigenous AI models³³ that will be made available for the public to access, modify, or improve according to the unique needs of India’s AI ecosystem.

Policy Recommendations

As the Indian government undertakes initiatives and efforts to shape the national AI ecosystem, it must explore and experiment with various socio-techno-legal frameworks for a safe and sustainable decentralised

AI stack for India. The following section presents recommendations for governments, policymakers, and AI practitioners as they reimagine democratising AI through the lens of a decentralised AI stack.

1. The following aspects must be kept in mind when implementing a decentralised AI stack for the country:
 - a. Decentralised AI systems will require standardised interfaces and protocols to coordinate AI tasks across distributed nodes. Communication standards must also be established for data and model synchronisation during federated learning operations.
 - b. Organisations and researchers will need clear guidelines and training for adopting these protocols, as technical challenges and policy compliance issues could arise during the transition from existing centralised AI/ML ecosystems to decentralised models.
 - c. The government will need to develop and establish transparent policies for dynamic resource allocation. Further, the operational expectations for the distributed compute resources must be formally defined and disclosed to all relevant stakeholders:
 - i. **Availability** of data, compute, and storage resources must be presented to the requesting entities.
 - ii. **Accessibility standards** for diverse user groups must be defined.
 - iii. **Costing and technical** readiness must be assessed for operating and maintaining an AI system in a decentralised mode.
 - iv. **Metering mechanisms** must be built to track GPU utilisation, energy consumption, and service health across the decentralised nodes. These kinds of resource monitoring mechanisms can be built on the transaction layer in the AI stack.
 - d. Along with numerous opportunities, decentralisation presents numerous risks. Governments and practitioners will need to develop guidelines and tools to prevent misuse of openly available AI models and data, and conduct impact assessment and risk compliance and assessment in line with the DPDP Act

- and DPDP Rules 2025. In this context, it will also be essential to establish accountability for AI outcomes and failures in the decentralised setup and promulgate cybersecurity guidelines.
- e. Testing and auditing guidelines must be formulated to assess the models developed and deployed in a decentralised setup. An incident reporting and response mechanism to address the failures or defects encountered by the decentralised AI components would be highly beneficial in this regard.
 - f. Human-centric values and sector-specific practices must be integrated in the decentralised approach. For instance, while training a disease diagnostic model for a healthcare application, there must be provisions for the patients to:
 - i. Grant or revoke consent or permissions for using their data for AI model training. This would need a granular consent management system built on top of a consent management protocol or sector-specific policies or laws.
 - ii. Receive incentives or benefits when their anonymised information is added to the training data from an underrepresented community and has a significant impact on improving the model accuracy.
 - iii. Get notified and time-bound data usage beyond the consented period. Provisions for auto-data deletion upon consent expiry or withdrawal should be investigated.
 - iv. Share data across devices in a distributed setup with zero-knowledge proof protocols without compromising data privacy and security.
 - g. A participatory approach must be fostered in building a decentralised setup. The government must initiate a public consultation on public-sector cybersecurity governance, risk, and compliance to ensure applications developed for large-scale adoption are governed appropriately.
2. Reimagining the Decentralised AI Stack with an inclusive marketplace for AI components from the perspective of IndiaAI Mission:
- a. IndiaAI compute capacity, partially built along the lines of the decentralised model, can be optimally decentralised by considering the above aspects.

- b. IndiaAI Dataset Platform - AIKosh - offers domain-specific datasets curated for domain-specific applications and compliant with India's DPDP Act. These open datasets and models available can be leveraged for developing AI solutions using the decentralised AI stack.
- c. IndiaAI Application Development Initiative - aims to develop various sector-specific indigenous AI models and applications. There should be transparent policies to monitor, report, and reuse the models and applications in a decentralised setup.
- d. IndiaAI Safe and Trusted AI - aims to develop a safe and trustworthy AI ecosystem, and can focus on developing tools and frameworks required for managing the decentralised AI stack.

By addressing these aspects, India can develop a robust and decentralised AI stack that can achieve the vision of democratising AI in a safe, inclusive, ethical, and responsible manner.

Geetha Raju is Senior Policy Analyst - AI & Data, Centre for Responsible AI, Indian Institute of Technology Madras.

Endnotes

- 1 Tor Constantino, "How Big Tech, ChatGPT And DeepSeek Could Lose To Decentralized AI," *Forbes*, February 12, 2025, <https://www.forbes.com/sites/torconstantino/2025/02/12/how-big-tech-chatgpt-and-deepseek-could-lose-to-decentralized-ai/>
- 2 Anit Mukherjee and Ashwini Joshi, "Digital Public Infrastructure As a Catalyst for Private Sector Innovation: Lessons From Fintech Sector in India," Observer Research Foundation, January 20, 2025, <https://www.orfonline.org/research/digital-public-infrastructure-as-a-catalyst-for-private-sector-innovation>
- 3 Ministry of Information & Broadcasting, Government of India, <https://pib.gov.in/PressReleaseFramePage.aspx?PRID=2082144>.
- 4 Ashish Desai and Aroon P. Manoharan, "Digital Transformation and Public Administration: The Impacts of India's Digital Public Infrastructure," *International Journal of Public Administration* 47, no. 9 (2024): 575-578.
- 5 Siddharth Jindal, "After Aadhaar and UPI, This New Digital Public Infrastructure Will Change How Indians Communicate," *Analytics India Magazine*, July 23, 2024, <https://analyticsindiamag.com/it-services/after-aadhaar-and-upi-this-new-digital-public-infrastructure-will-change-how-indians-communicate/>
- 6 Arjun Goswami, Varun Mehta and Yashika Sachdeva, "RegTech and Digital Public Infrastructure: Navigating Compliance in India's Digital Landscape," Cyril Amarchand Mangaldas, November 18, 2024, <https://corporate.cyrilamarchandblogs.com/2024/11/regtech-and-digital-public-infrastructure-navigating-compliance-in-indias-digital-landscape/>
- 7 Ministry of Electronics & IT, Government of India, <https://pib.gov.in/PressReleasePage.aspx?PRID=2108810>
- 8 SwarmZero, "Understanding the Decentralized AI Stack," August 1, 2024, <https://swarmzero.ai/blog/understanding-the-decentralized-ai-stack>
- 9 SwarmZero, "Understanding the Decentralized AI Stack"
- 10 Hongyang Zhang et al., "Complete Security and Privacy for AI Inference in Decentralized Systems," arXiv, 2024.
- 11 Nitin Singh, Pankaj Dayama, and Vinayaka Pandit, "Zero Knowledge Proofs Towards Verifiable Decentralized Ai Pipelines," in *International Conference on Financial Cryptography and Data Security* (Cham: Springer International Publishing, 2022), pp. 248-275.
- 12 Bing-Jyue Chen, "Zkml: An Optimizing System for ml Inference in Zero-knowledge Proofs," in *Proceedings of the Nineteenth European Conference on Computer Systems* (Athens, Greece, 2024), pp. 560-574.
- 13 Xiaoxiao Liang, "Rscfed: Random Sampling Consensus Federated Semi-Supervised Learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10154-10163.
- 14 K.D. Conway et al., "Opml: Optimistic Machine Learning on Blockchain," arXiv, 2024.

- 15 Praneeth Vepakomma and Ramesh Raskar, "Split Learning: A Resource Efficient Model and Data Parallel Approach for Distributed Deep Learning," in *Federated Learning: A Comprehensive Overview of Methods and Applications* (Cham: Springer International Publishing, 2022), pp. 439-451.
- 16 Dayeol Lee, "Privacy-Preserving Decentralized AI with Confidential Computing," arXiv, 2024.
- 17 "Trusted Execution Environment (TEE)," Microsoft Learn, March 2025, <https://learn.microsoft.com/en-us/azure/confidential-computing/trusted-execution-environment>
- 18 Tim Geppert et al., "Trusted Execution Environments: Applications and Organizational Challenges," *Frontiers in Computer Science* 4 (2022): 930741.
- 19 Infosys BPM, "Decentralised Finance (DeFi) and Its Potential to Disrupt Traditional Finance," March 5, 2025, <https://www.infosysbpm.com/blogs/financial-services/decentralised-finance-and-its-potential-to-disrupt-traditional-finance.html>
- 20 Sarah Hagggenmüller, "Federated Learning for Decentralized Artificial Intelligence in Melanoma Diagnostics," *JAMA Dermatology* 160, no. 3 (2024): 303-311.
- 21 Kallista Bonawitz, "Federated learning and privacy," *Communications of the ACM* 65, no. 4 (2022): 90-97.
- 22 Charalampos Stamatellis et al., "A Privacy-preserving Healthcare Framework Using Hyperledger Fabric," *Sensors* 20, no. 22 (2020): 6587.
- 23 Suma Bhat et al., "Sakshi: Decentralized Ai Platforms," arXiv, 2023.
- 24 Gabriel Montes et al., "Distributed, Decentralized, and Democratized Artificial Intelligence," *Technological Forecasting and Social Change* 141 (2019): 354-358.
- 25 Yue Wang et al., "Blockchain-Enabled Decentralized Ai Ecosystems: A Conceptual Framework and Bittensor Case Study," *SSRN*, August 27, 2024, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4938275.
- 25 Ocean Protocol Team, "Ocean Protocol Joins Open DeFi Alliance, Connecting Members With New Sources of Liquidity in Data Tokens," 2021, <https://blog.oceanprotocol.com/ocean-protocol-joins-open-defi-alliance-connecting-members-with-new-sources-of-datatoken-liquidity-dd0669a9b2be>.
- 27 Botao Hu, "Is Decentralized Artificial Intelligence Governable? Towards Machine Sovereignty and Human Symbiosis," *SSRN*, January 31, 2025.
- 28 Zhen Dong et al., "Addressing Challenges in Large-scale Distributed AI Systems," UC Berkeley, 2022.
- 29 Department of Telecommunications, Government of India, *Indian Artificial Intelligence Stack* (New Delhi: Department of Telecommunication, 2020), <https://www.tec.gov.in/pdf/Whatsnew/ARTIFICIAL%20INTELLIGENCE%20-%20INDIAN%20STACK.pdf>
- 30 "Indian Artificial Intelligence Stack"
- 31 Cabinet, Government of India, <https://pib.gov.in/PressReleaselframePage.aspx?PRID=2012355>

- 32 IndiaAI, “Union Minister of Electronics & IT, Railways, and I&B Announces the Availability of 18,000+ Affordable AI Compute Units,” January 31, 2025, <https://indiaai.gov.in/article/union-minister-of-electronics-it-railways-and-i-b-announces-the-availability-of-18-000-affordable-ai-compute-units>
- 33 “Call for Proposals for Building India’s Foundational AI Models IndiaAI Mission,” IndiaAI, January 30, 2025, <https://indiaai.gov.in/article/call-for-proposals-for-building-india-s-foundational-ai-models>
- 34 Ministry of Law and Justice, Government of India, *The Digital Personal Data Protection Act* (New Delhi: Ministry of Government, 2023), <https://egazette.gov.in/WriteReadData/2023/248045.pdf>
- 35 Ministry of Electronics & IT, Government of India, <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2090048>
- 36 Ahmed M. Shamsan Saleh, “Blockchain for Secure and Decentralized Artificial Intelligence in Cybersecurity: A Comprehensive Review,” *Blockchain: Research and Applications* (2024): 100193.
- 37 “AlKosh – IndiaAI Dataset Platform,” Government of India, March 20, 2025, <https://aikosh.indiaai.gov.in/home>

Democratising the AI Stack: Technical Exploration and Reflections

Sachin Kumar

Artificial Intelligence (AI) is regarded as a transformative technology similar to electricity. AI solutions are projected to add US\$15.7 trillion to global GDP by 2030, potentially increasing it by 14 percent through increased productivity and new market opportunities across various sectors and industries.¹ In July 2023, a Workday survey covering 1,000 global organisations found that over 90 percent of organisations currently use AI to effectively manage personnel and finances, automate a host of internal functions and operations, and augment decision-making.² AI-powered

chatbots and virtual assistants provide 24/7 customer support, improving response times and customer satisfaction. Generative AI, in particular, has been found to increase highly skilled workers' productivity by nearly 40 percent compared with workers who do not use it.³ However, AI potentially displaces jobs with highly repetitive tasks in various sectors while creating new job opportunities—this necessitates skill upgrades and trainings worldwide.⁴

The landscape of AI technology design and development is also undergoing a profound transformation due to widespread challenges from external and internal requirements. While centralised cloud-based AI models have dominated recent advancements, a growing movement advocates decentralising AI components due to concerns around ethics, bias, transparency, accountability, privacy, security, and governance, seeking to distribute computational power, data ownership, and model development across a network of participants and stakeholders.⁵ For example, the concentration of vast amounts of sensitive data within a few organisations creates monopoly and privacy risks, limiting the AI potential for the larger good.

Indeed, there are a number of issues in current AI technological developments. First, according to the California-based Identity Theft Resource Center, there was a record number of reported data breaches globally in 2023, with sensitive personal data exposed in numerous incidents.⁶ This illustrates the vulnerability of centralised data storage for massive AI datasets. Second, the dominance of a few tech giants in AI development raises concerns about algorithmic bias and a lack of transparency in technology design and development. Studies highlight the pervasive bias problem in AI systems, often stemming from skewed training data.⁷ Third, the centralised AI design and development model can limit access to AI technologies for smaller organisations and individuals.⁸ The lack of adequate resources required for the deployment of AI potentially hinders innovation in less technologically advanced, underrepresented, and low-income stakeholders and countries, resulting in perpetuated inequalities in harnessing AI's benefits regionally.⁹

To boost the democratisation of the technology, the development process of AI systems must be modular and interdependent in a layered format called the AI stack to address its black box nature,^a the challenges of centralisation, and transparency issues.¹⁰ A stack is a collection of technologies, frameworks, infrastructure aspects, services, and modules that facilitate the development of AI products and services by layering components to support the end-to-end AI product lifecycle.¹¹ It comprises several layers: data collection, model development, deployment, governance, and utilisation.

Traditionally, the functions associated with these components are controlled by a handful of entities in North America, Europe, and East Asia, raising concerns about digital sovereignty and fair AI access, particularly for stakeholders having fewer resources and access to technology globally, and specifically in the Global South.¹² The development of decentralised AI systems and democratising access to several parts of the stack present a compelling alternative. By distributing data, computational power, and model development across a network of participants, decentralisation aims to enhance privacy, improve security, and democratise access to AI.¹³ If clear commitments to “democratising AI” and productive discussions of concrete policies and trade-offs are to materialise, it will be imperative to recognise the principal role of democratising the AI pipeline. In other words, the AI Stack and working on a modular approach in navigating trade-offs and risks across decisions around the AI use, design, and development, and sharing profits will be key.¹⁴

Overview of the Current Landscape and AI Stack

Over the last couple of years, discussions surrounding AI democratisation have been amplified worldwide due to the transformative nature of the technology and its far-reaching implications on individuals and societies.¹⁵ Leading AI companies such as Stability AI,¹⁶ Meta,¹⁷

a The ‘black box’ nature of AI refers to the fact that the internal decision-making and modelling processes of AI systems are opaque and difficult to understand, even for the developers who built them.

Microsoft,¹⁸ and Hugging Face¹⁹ are talking about their commitment to democratising AI. Yet, it is not always clear what they mean. The term 'AI democratisation' is being employed in various ways, causing stakeholders to speak past one another when discussing the goals, methodologies, risks, and benefits.²⁰ Several types of AI democratisation are commonly discussed, such as the democratisation of use, the democratisation of development, the democratisation of access and profits, and the democratisation of AI governance.²¹ AI democratisation is a multifarious and sometimes conflicting concept that should not be conflated with improving AI accessibility alone. According to Stanford University's AI Index 2023 report, over 75 percent of large-scale AI models are owned and controlled by just five technology companies.²²

Furthermore, 80 percent of AI computing power is concentrated in a handful of cloud providers, highlighting the risk of monopolistic control over AI development and access.²³ Like other tech stacks in software development, the AI stack organises the elements into layers that work together to enable efficient and scalable AI systems to be deployed in real-world situations. The layered approach breaks down the complex process of building AI solutions into manageable components, allowing teams to focus on individual aspects without losing sight of the bigger picture and making it easier to identify dependencies, allocate resources, and address challenges systematically over the AI product lifecycle. The detailed AI stack or AI pipeline layer information is as follows:

1. **AI Infrastructure Layer:** This foundational layer aggregates all Infrastructure as a Service (IaaS) resources, including specialised accelerators such as graphics processing units (GPUs), tensor processing units (TPUs), and field-programmable gate arrays (FPGAs). It incorporates the underlying hardware environment that delivers the computational power, storage capacity, and network communication essential for supporting AI-driven applications.²⁴
2. **AI Platform Layer:** Situated above the infrastructure, this layer integrates Platform as a Service (PaaS) functionalities, MLOps (Machine Learning Operations) practices, and the Intelligent Engagement Platform (IEP) to manage the full lifecycle of AI applications. It enables model development, evaluation, deployment, and monitoring, while ensuring continuous integration (CI), continuous

delivery (CD), and continuous testing (CT), thereby streamlining the sustained execution and maintenance of AI applications.²⁵

3. **AI Framework Layer:** This layer encompasses a wide range of AI frameworks designed to accelerate the creation and deployment of AI-enabled applications. It includes tensor-based computing with GPU acceleration, automatic differentiation systems, and pre-built libraries such as PyTorch, Scikit-learn, TensorFlow (developed by Google for deep learning), NumPy (for multi-dimensional arrays and mathematical operations), Keras (for neural network development), Pandas (for data handling and analysis), Matplotlib (for data visualisation), SciPy (for scientific computing), and OpenCV (for computer vision tasks such as facial recognition). Additionally, this layer offers pre-built models, such as neural networks (NNs), to facilitate faster development.²⁶
4. **AI Algorithm Layer:** At this level, a collection of open-source or customised algorithms is provided, covering domains such as supervised learning, unsupervised learning, and reinforcement learning. These algorithms form the foundation for addressing problem-solving, decision-making, and predictive tasks in AI systems.²⁷
5. **AI Data Layer:** This layer focuses on data management through Data as a Service (DaaS) and DataOps platforms. It unifies diverse data architectures to support the complete data lifecycle, offering preprocessing tools, feature engineering capabilities, and mechanisms for handling both internal and external data sources. It also accommodates heterogeneous data types, including stationary and non-stationary datasets.²⁸
6. **AI Service Layer:** This layer delivers a variety of ready-to-use, general-purpose application programming interfaces (APIs) for AI-enabled services. These APIs facilitate the transfer of information or raw data, allowing integration with existing IT systems and enterprise applications to provide solutions at higher operational levels.²⁹
7. **AI Solution Layer:** Positioned at the top, this layer delivers AI-enabled solutions tailored to specific business domains. It allows business analysts and domain experts to design and deploy sector-specific applications, thereby enhancing the adoption of AI capabilities across industries and organisational settings.³⁰

The AI stack model delineates the functional responsibilities of each layer and enables interoperability across vendors. By conforming to common standards and requirements, vendors can design products and services at a specific layer, utilise services from the underlying layer, and provide consistent service interfaces to the layer above. This approach reduces reliance on a single vendor for end-to-end solutions spanning hardware to applications, thereby mitigating the risks of vendor lock-in and opportunistic exploitation.

In recent years, major cloud service providers have intensified their involvement in the AI-as-a-service (AlaaS) domain by integrating cloud offerings with fundamental AI components, including large-scale datasets, advanced learning algorithms, and high-performance computing hardware.³¹ Although AlaaS allows organisations to access AI capabilities without substantial initial investment, several pertinent barriers continue to impede the development of effective AI systems.³² For instance, current AlaaS solutions are predominantly delivered as proprietary, bundled packages, which restrict interoperability among vendors, reinforce vendor lock-in, and exacerbate concerns over centralisation and limited democratisation of AI development.³³

Moreover, the tightly coupled nature of components across layers constrains extensibility, reducing developers' ability to flexibly adopt the most suitable AI components for practical deployment. Such bundled offerings are frequently regarded as closed ecosystems that limit engagement from the open-source community, elevate switching and lock-in costs, and generate long-term risks of incompatibility and migration challenges across vendors.³⁴

Assessment and Analysis

When analysing the AI stack to democratise the technologies, there is a need to integrate the principle of decentralisation and democratisation in each layer. The decentralisation of the AI Stack is a rapidly evolving field that aims to address critical limitations inherent in centralised AI and intertwined architectures. Traditional AI systems are predominantly structured around centralised models, where data storage, model training, and inference occur within the centralised infrastructure of a few powerful and limited entities. While this has facilitated progress

in AI research and deployment, it raises concerns about scalability, privacy, security, fairness, as well as the monopolisation of data, AI models, technology, and computational resources.

There are several aspects to achieving AI democratisation through decentralisation. Initially, the democratisation of AI use was considered the end product. The true democratisation of AI technology is the democratisation of data, AI design and development, the democratisation of AI profits, and the democratisation of AI governance.³⁵ Several organisations are expressing their desire to democratise AI, but it is not clear which aspect of the democratic stack they are referring to.

1. **Towards Democratisation and Empowerment:** The democratisation of AI usage makes AI technology more manageable for many people. For example, Stability AI³⁶ has been a champion of AI democratisation. The company proudly describes its main product, the image generation model Stable Diffusion, as “a text-to-image model that will empower billions of people to create stunning art within seconds.”³⁷ Some AI tools have limited benefits but pose manifold risks if widely accessible. For instance, AI for drug discovery can be misused to create toxins. Controlled access ensures benefits without enabling harm. Restricting access to high-risk AI does not necessarily hinder democratisation.
2. **Expanded Participation in AI Design and Development:** AI democratisation focuses on expanding participation in AI design and development, enabling a broader range of people to contribute. This approach aims to accelerate innovation within a safe AI ecosystem by implementing policies that prevent harm, misuse, and misalignment. Concerns about monopolisation by a few leading AI labs and their narrow developer demographics drive calls for democratisation. Expanding participation can help ensure that AI systems perform equitably across diverse ethnic, geographic, cultural, and professional backgrounds, ultimately fostering AI applications that address a wider range of global needs.
3. **The Promise and Peril of Openness:** The open-source movement seeks to advance AI development by creating openly available frameworks and datasets, as well as designing tools accessible to a wider community. Such openness allows for a larger and more diverse range of contributors to engage in the innovation

process. However, it simultaneously introduces risks, including malicious exploitation, model tampering, and the difficulty of enforcing effective safeguards, thereby necessitating external evaluation mechanisms. Broad participation, including academics, independent developers, and smaller research groups from varied disciplines and geographic regions, offers critical opportunities for external oversight and rigorous auditing. By enabling wider access to models for study and evaluation, AI laboratories can distribute auditing responsibilities across a larger and more diverse pool of developers than would be feasible within a single institution. This broader scrutiny increases the likelihood of identifying weaknesses or flaws, ultimately contributing to the development of safer and more reliable AI systems.

4. **Good Practices and New Risks:** To foster meaningful participation in AI design and development, several practices can be adopted. These include model sharing, providing computational resources, and supporting collaborative projects. Model sharing encompasses access to their code and weights, as well as the ability to query, modify, and analyse them. While such transparency is crucial for enabling external auditing and research, it also heightens the risk of misuse by malicious actors. For this reason, access to certain high-risk models may need to be restricted. For instance, Meta opted to regulate the distribution of its large language model LLaMA, granting access to academic researchers and a few others “to maintain integrity and prevent misuse.”³⁸
5. **Improving Access to Compute:** Large AI models need to have tremendous computing access (and other technical infrastructure.)³⁹ Accordingly, democratising development may also require improvements to computing access. However, restrictions on computing can also be leveraged to help minimise the misuse of powerful AI by limiting the ability of prospective malicious actors to build or modify large models. Therefore, decisions to provide multiple compute resources should involve adequate risk-benefit analysis, just like decisions to open-source AI models.⁴⁰
6. **Project Collaboration and Coordination:** Democratising AI development is not just about providing resources and assuming that people will participate. Effective input elicitation often benefits from dedicated project coordination and support. For example, the BigScience project was a collaborative effort coordinated by the AI

startup Hugging Face.⁴¹ BLOOM, another collaborative project, was created for a year by a global coalition of over 1000 volunteer AI developers, yielding a functional LLM in 46 languages.⁴²

7. **Educational and Upskilling Opportunities:** The democratisation of AI development can also be furthered by expanding the community of people capable of contributing to AI design and development processes. For instance, investment in computer science and machine learning (ML) educational resources is essential for establishing talent pipelines and narrowing the AI divide between the Global North and South.⁴³
8. **Assistive Tools:** Another option for expanding the community of prospective contributors is to lower barriers to participation in AI development activities by making it easier for people with minimal programming experience and little familiarity with ML to participate. This could be done by providing tools to enable those with less experience and expertise to create and implement machine-learning applications. For example, Microsoft, Google, H2O, and Amazon have developed “no-code” tools that allow people to build personalised models for their needs without prior coding or ML experience.⁴⁴
9. **Emerging Decentralised Technologies:** Emerging decentralised technologies, such as blockchain, federated learning (FL), peer-to-peer (P2P) networks, and decentralised autonomous organisations (DAOs), should be proactively explored as alternatives that distribute AI-related processes across multiple nodes, mitigating the risks associated with centralisation. Federated learning, for instance, allows model training on distributed devices without sharing raw data, addressing privacy concerns.⁴⁵ Blockchain technology offers potential solutions for secure data storage, transparent model auditing, and the creation of decentralised AI marketplaces.⁴⁶ According to research by MarketsandMarkets, the federated learning market is seeing growth, with an expected increase in its market size. However, the path to a decentralised AI ecosystem has its obstacles. Technical challenges include data distribution and management, distributed model training, and resource constraints on edge devices. In healthcare, Google’s Federated Learning for Mobile Health initiative has demonstrated a 40-percent reduction in data transfer needs while maintaining high predictive accuracy.⁴⁷ However, FL is not without challenges. Synchronising models

across distributed nodes leads to increased network overhead and computational costs. FL models can experience accuracy degradation due to heterogeneity in data distributions across nodes.⁴⁸

Limitations and Challenges of AI Stack Democratisation

Unlike centralised AI systems, where bias detection and mitigation techniques can be applied to a single data source, decentralised AI models aggregate data from multiple, often heterogeneous, sources, increasing the risk of algorithmic bias. A recent study from Stanford University's Human-Centred AI Institute found that federated learning models trained on diverse datasets exhibit a higher demographic bias than centralised models, emphasising the need for fairness-aware decentralised training mechanisms.⁴⁹ The economic and financial implications of decentralising AI are also noteworthy. Implementing decentralised AI solutions requires higher investment in distributed infrastructure, edge computing capabilities, and novel data and model-sharing incentive models. The World Economic Forum estimates that transitioning to decentralised AI architectures could increase operational costs by up to 50 percent in the short term before cost efficiencies emerge.⁵⁰

Ensuring trust in decentralised AI requires technical solutions and an understanding of human psychology and decision-making. Research suggests that trust in decentralised AI systems is 30-percent lower than in centralised AI models, mainly due to the lack of a single accountable entity.⁵¹ Bridging this trust gap necessitates transparent model explainability, decentralised audit mechanisms, and community-driven oversight frameworks. Another challenge lies in energy consumption and sustainability. Decentralised AI models, particularly those leveraging blockchain or distributed ledger technology, often require extensive computational power, leading to high energy consumption.⁵² This raises concerns about the environmental impact of decentralising AI, necessitating research into energy-efficient consensus mechanisms and sustainable computing practices.

In terms of AI governance in decentralised systems, it presents a unique challenge. Pertinent questions include: Who owns the AI models? And who is responsible for biases, errors, or unethical decision-making?⁵³

The European Union's AI Act and the United States' (US) AI Bill of Rights are beginning to address these questions, but decentralised AI further complicates legal accountability. One hurdle in a decentralised AI stack is the lack of standardised protocols and interoperability among AI models, datasets, and computing environments. According to a report, over 60 percent of AI-driven organisations struggle with integrating decentralised AI solutions due to compatibility issues.⁵⁴

The Global South often faces data colonialism, where indigenous data is harvested without benefiting local communities. Decentralised data collection through data trusts and cooperative ownership models can provide a solution. AI training requires immense computational resources, typically housed in hyperscale data centres in the Global North. Countries in the Global South lack such infrastructure, making AI research and application development dependent on external resources. Developing countries in the Global South spend billions annually on AI services from foreign providers. By fostering local innovation through decentralised frameworks, nations can retain economic value, reduce dependency, and promote domestic AI entrepreneurship. Governments should mandate open-access AI models trained on regional datasets.

Recommendations

This study explores decentralised AI in the stack, emphasising the need for interoperability, transparency, inclusivity, security, and resource-efficient AI deployment.

1. It introduces a layered AI tech-stack model that offers flexibility in implementation by reducing vendor lock-in, supporting open-source AI tools, and facilitating seamless integration with enterprise IT systems. The model prevents dependency on proprietary AI solutions, enabling organisations to select AI tools based on specific needs. The structured layering helps define Service Level Agreements (SLAs), ensuring fault tolerance and smooth deployment. On-demand, pay-per-use AI resource allocation reduces financial and operational burdens. Open-source support and MLOps/DataOps frameworks enhance AI system updates and maintenance.
2. The proposed analysis offers managers a framework to assess AI capabilities, make informed decisions on in-house development

vs. outsourcing, and streamline AI adoption across organisational units. The model clarifies AI component layers, enabling targeted assessments for enterprise needs. Stakeholders can evaluate AI readiness and determine the required resources for implementation. Managers can leverage this framework to identify gaps, manage AI risks, and optimise affiliated investments.

3. The AI stack should ensure advancing secure and interoperable systems to reduce reliance on centralised systems; this study suggests decentralising AI components and providing secure, transparent, and equitable access. Key strategies include standardised protocols for interoperability, developing open-source standards (e.g., APIs, communication frameworks) to prevent vendor lock-in, and enhancing AI stack integration. Secure, decentralised platforms allow data owners to monetise and control access, ensuring ethical AI usage. Federated learning standardisation establishes protocols for privacy and AI training, which enhances decentralised AI adoption. Decentralised access control, audit mechanisms, and transparent AI governance strengthen trust and accountability.
4. Decentralised AI stacks and systems must address infrastructure limitations, particularly in resource-constrained environments like the Global South. Shifting AI processing to edge devices (IoT, mobile) reduces reliance on centralised servers. It develops local and lightweight AI models that function efficiently on low-power devices, supporting broader adoption. Collaborative AI development with Open-source tools and cooperatives encourages shared model training and data access, bridging AI disparities between the Global North and Global South.

Organisations can efficiently develop and deploy AI by leveraging AI tech-stack models, decentralised frameworks, and collaborative ecosystems while ensuring equitable access and security. Policymakers, researchers, and industry leaders must work together to standardise AI governance, enhance accessibility, and reduce the technology's dependency on centralised infrastructures.

Sachin Kumar is Assistant Professor of Computer Science at Cluster Innovation Centre, University of Delhi.

Endnotes

- 1 PwC, *Sizing the Prize. What's the Real Value of AI for Your Business and How Can You Capitalize?*, 2023, <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizingthe-prize-report.pdf>.
- 2 L Ms et al., "How AI Use in Organizations Contributes to Employee Competitive Advantage: The Moderating Role of Perceived Organization Support. Technological Forecasting and Social Change," *Elsevier* 123801, no. 209 (2024), <https://www.scribd.com/document/795241384/Ma-2024-How-AI-use-in-organizations-contributes-to-employee-competitive-advantage-The-moderating-role-of-perceived-organization-support#:~:text=use%20and%20employees'%20competitive%20advantage,competitive%20advantage>
- 3 Meredith Somers, "How Generative AI Can Boost Highly Skilled Workers' Productivity," *MIT*, October 19, 2023, <https://mitsloan.mit.edu/ideas-made-to-matter/how-generative-ai-can-boost-highly-skilled-workers-productivity>.
- 4 Araz Zirar et al., "Worker and Workplace Artificial Intelligence (AI) Coexistence: Emerging Themes and Research Agenda," *Elsevier* 102747, no. 124 (2023), <https://www.sciencedirect.com/science/article/pii/S0166497223000585>.
- 5 Abhishek Singh et al., "A Perspective on Decentralizing AI," *MIT*, 2024, https://decai-resources.s3.us-east-1.amazonaws.com/decentralized_ML_perspective-16.pdf.
- 6 Shanitamol Gracy, "A Global Analysis of Data Breaches From 2004 to 2024," *Arxiv* 2502, no.05205 (2025), <https://arxiv.org/abs/2502.05205>.
- 7 Emilio Ferrara, "The Butterfly Effect in Artificial Intelligence Systems: Implications for AI Bias and Fairness," *Elsevier* 100525, no. 15 (2024), <https://www.sciencedirect.com/science/article/pii/S266682702400001X>.
- 8 Emmanouil Papagiannidis et al., "Toward AI Governance: Identifying Best Practices and Potential Barriers and Outcomes. Information Systems," *Springer*, no. 25 (2023), <https://link.springer.com/article/10.1007/s10796-022-10251-y>.
- 9 Muhammad Khan et al., "Artificial Intelligence for Low-Income Countries. Humanities and Social Sciences Communications," *Humanities and Social Science Communication* 1422, no. 11 (2024), <https://www.nature.com/articles/s41599-024-03947-w>.
- 10 Rua Tsaih et al., "The AI Tech-Stack Model," *Communications of the ACM*, no. 66 (2023), <https://cacm.acm.org/research/the-ai-tech-stack-model/>.
- 11 Tsaih et al., "The AI Tech-Stack Model."
- 12 Huw Roberts et al., "Global AI Governance: Barriers and Pathways Forward," *International Affairs* 100, no. 3 (2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4588040.
- 13 Lingjuan Lyu et al., "How to Democratise and Protect AI: Fair and Differentially Private Decentralised Deep Learning," *Arxiv*, no. 19 (2020), <https://arxiv.org/abs/2007.09370>.
- 14 Erwan Moreau et al., "A Paradigm for Democratizing Artificial Intelligence Research. Innovations in Big Data Mining and Embedded Knowledge," *Hal* 02281202ff, 2019, <https://hal.science/hal-02281202v1/document>.

- 15 Elizabeth Seger et al., "Democratising AI: Multiple Meanings, Goals, and Methods," *Arxiv*, 2023, <https://arxiv.org/abs/2303.12642>.
- 16 Scale Events, "Emad Mostaque (Stability AI): Democratizing AI, Stable Diffusion & Generative Models," <https://exchange.scale.com/public/videos/emad-mostaque-stability-ai-stable-diffusion-open-source>.
- 17 Susan Zhang, Mona Diab and Luke Zettlemoyer, "Democratizing Access to Large-Scale Language Models with OPT-175B," *Meta*, May 3, 2022, <https://ai.facebook.com/blog/democratizing-access-to-large-scale-language-models-with-opt-175b>.
- 18 Microsoft News Center, "Democratizing AI for Every Person and Every Organization," *Microsoft*, September 26, 2016, <https://news.microsoft.com/features/democratizing-ai/>.
- 19 Hugging Face, "Sergei," <https://huggingface.co/about>
- 20 Elizabeth Seger et al., "Democratising AI: Multiple Meanings, Goals, and Methods," *Arxiv* 2303, (2023), <https://arxiv.org/abs/2303.12642>.
- 21 Seger et al., "Democratising AI: Multiple Meanings, Goals, and Methods."
- 22 Nestor Maslej et al., "Artificial Intelligence Index Report 2023," *ArXiv* 2310.03715, (2023), <https://arxiv.org/abs/2310.03715>.
- 23 Fernando van der Vlis et al., "Big AI: Cloud Infrastructure Dependence and the Industrialisation of Artificial Intelligence," *Sage Journal* 20539517241232630, no.11 2024), <https://journals.sagepub.com/doi/10.1177/20539517241232630>.
- 24 Tsaih et al., "The AI Tech-Stack Model."
- 25 Tsaih et al., "The AI Tech-Stack Model."
- 26 Tsaih et al., "The AI Tech-Stack Model."
- 27 Tsaih et al., "The AI Tech-Stack Model."
- 28 Tsaih et al., "The AI Tech-Stack Model."
- 29 Tsaih et al., "The AI Tech-Stack Model."
- 30 Tsaih et al., "The AI Tech-Stack Model."
- 31 Sebastian Lins et al., "Artificial Intelligence as a Service," *Springer*, no. 63 (2021), <https://link.springer.com/article/10.1007/s12599-021-00708-wf>
Classification and Research Directions.
- 32 Naeem Syed et al., "Artificial Intelligence as a Service (AlaaS) for Cloud, Fog and the Edge: State-of-the-Art Practices," *ACM Digital Libraries* 211, no. 57 (2025), <https://dl.acm.org/doi/10.1145/3712016>.
- 33 Desire Harauzek, "Cloud Computing: Challenges of Cloud Computing from Business Users Perspective-Vendor Lock-In," *Digitala Vetenskapliga Arkivet*, July 1, 2022, <http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1679817&dswid=2750>.
- 34 Justice Opara Martins et al., "Critical Analysis of Vendor Lock-In and Its Impact on Cloud Computing Migration: A Business Perspective," *Springer* 4, no.5 (2016), <https://link.springer.com/article/10.1186/s13677-016-0054-z>.
- 35 Tsaih et al., "The AI Tech-Stack Model."

- 36 Scale Events, "Emad Mostaque (Stability AI): Democratizing AI, Stable Diffusion & Generative Models," Scale Events, <https://exchange.scale.com/public/videos/emad-mostaque-stability-ai-stable-diffusion-open-source>.
- 37 "Stable Diffusion," *Diffusion News*, August 24, 2022, <https://diffusion-news.org/stable-diffusion#:~:text=Stable%20Diffusion%20is%20a%20text,can%20run%20on%20consumer%20GPUs>.
- 38 Susan Zhang, Mona Diab and Luke Zettlemoyer, "Democratizing Access to Large-Scale Language Models with OPT-175B," *Meta*, May 3, 2022, <https://ai.facebook.com/blog/democratizing-access-to-large-scale-language-models-with-opt-175b>.
- 39 Sukhpal Singh Gill et al., "AI for Next Generation Computing: Emerging Trends and Future Directions," *Elvesier* 100514, no. 19 (2022), <https://www.sciencedirect.com/science/article/abs/pii/S254266052200018X>.
- 40 Tsaih et al., "The AI Tech-Stack Model."
- 41 Christopher Akiki et al., "Bigscience: A Case Study in the Social Construction of a Multilingual Large Language Model," *Arxiv* 2212.04960, (2022), <https://arxiv.org/abs/2212.04960>.
- 42 Melissa Heikkila, "Inside the Radical New Project to Democratize AI," *MIT Technology Review*, July 12, 2022, <https://www.technologyreview.com/2022/07/12/1055817/inside-a-radical-new-project-to-democratize-ai/>.
- 43 Manbo He, "Investment in R&D of AI: Evidence from the Global South," *Tech Transformation and AI Readiness*, 2025, https://www.researchgate.net/publication/389003317_Investment_in_RD_of_AI_Evidence_from_the_Global_South.
- 44 Mario Truss et al., "Human-Centered AI Product Prototyping with No-Code AutoML: Conceptual Framework, Potentials and Limitations," *International Journal of Human-Computer Interaction*, no. 41 (2024), https://www.researchgate.net/publication/389397216_Human-Centered_AI_Product_Prototyping_with_No-Code_AutoML_Conceptual_Framework_Potentials_and_Limitations.
- 45 Jie Wen et al., "A Survey on Federated Learning: Challenges and Applications," *Springer*, no. 14 (2023), <https://link.springer.com/article/10.1007/s13042-022-01647-y>.
- 46 Khaled Salah et al., "Blockchain for AI: Review and Open Research Challenges," *IEEE*, no. 7 (2019), <https://ieeexplore.ieee.org/document/8598784>.
- 47 Nicola Rieke, "The Future of Digital Health with Federated Learning," *NPJ*, no. 3 (2020), <https://www.nature.com/articles/s41746-020-00323-1>.
- 48 Mang ye et al., "Heterogeneous Federated Learning: State-of-the-Art and Research Challenges," *ACM Computing Surveys*, no. 56 (2023), <https://dl.acm.org/doi/10.1145/3625558>.
- 49 Daniel Yue Zhang et al., "FairFL: A Fair Federated Learning Approach to Reducing Demographic Bias in Privacy-Sensitive Classification Models," *Institute of Electrical and Electronics Engineers* 9378043, 2020, <https://experts.illinois.edu/en/publications/fairfl-a-fair-federated-learning-approach-to-reducing-demographic/>.

- 50 Xiuquan Qiao et al., “6G Vision: An AI-Driven Decentralized Network and Service Architecture,” *Internet of Things, People, and Processes*, no. 24 (2020), https://dsg.tuwien.ac.at/~sd/papers/Zeitschriftenartikel_2020_SD_6G.pdf.
- 51 Bo Li et al., “Trustworthy AI: From Principles to Practices,” *ACM Computing Surveys* 55, no. 9 (2023), <https://dl.acm.org/doi/10.1145/3555803>.
- 52 Zhiwei Cao, “Towards a Systematic Survey for Carbon Neutral Data Centers,” *Cornell University*, no. 24 (2021), <https://arxiv.org/abs/2110.09284>.
- 53 Dana Alsagheer et al., “Decentralized Machine Learning Governance: Overview, Opportunities, and Challenges,” *IEEE Access*, 2023, https://www.researchgate.net/publication/373675344_Decentralized_Machine_Learning_Governance_Overview_Opportunities_and_Challenges.
- 54 Moetez Abdelhamid et al., “A Review on Blockchain Technology, Current Challenges, and AI-Driven Solutions,” *ACM Computing Surveys* 3, no. 57 (2024), <https://dl.acm.org/doi/10.1145/3700641>.



II.

Unlocking Sectoral Applications

AI in Agriculture: Connecting Urban Innovations with Rural Needs

Diwakar Kumar

Artificial Intelligence (AI), particularly in the realm of Machine Learning (ML),^a presents immense potential for India, not only in terms of technological advancement but also as a novel approach to tackling some of the most pressing challenges confronting the country today.¹ It is a pivotal force in ushering in a new age of innovation, steadily embedding itself within the fabric of everyday existence. India is transforming the discourse on AI

a Machine Learning (ML) is a subset of AI that focuses on creating computer algorithms that enhance themselves through data analysis and experience. It allows computers to learn from data and make inferences or predictions without being programmed. These algorithms improve with use, processing more data to make them more precise and efficient.

by making it accessible to all, and not just a privileged few.² The government is implementing a future-forward AI strategy to make AI development accessible to every aspiring innovator, student, and startup across the country, intending to position India as a global leader in the domain.

This article delves into the revolutionary possibilities of AI and ML in India, specifically examining how they could propel technical advancement, social and economic progress, and inclusive prosperity.³ From agriculture to extension^b and science communication, it delves into how India's strategic AI projects are tackling national concerns, democratising access to tools, and nurturing grassroots innovation while assuring equitable and ethical adoption. The study also examines India's roadmap to become a worldwide leader in AI.⁴

From Research to Real-World Applications

The history of AI begins with English mathematician Alan Turing's questioning of machine intelligence in 1950. His work on code-breaking machines like the Bombe revealed that machines could perform tasks traditionally requiring human intuition, prompting him to explore whether such machines could be considered as "thinking." Turing's influential paper of 1950, "Computing Machinery and Intelligence," shifted the debate from metaphysical discussions to practical evaluations, introducing the Turing Test.⁵ Additionally, the limitations of early computers and the burgeoning complexity of postwar scientific disciplines fostered the development of AI, aiming to create machines that could reason and learn, addressing the limitations of traditional technology.⁶ Ultimately, Turing's question emerged from the intersection of philosophical reflection, wartime advancements, and the recognition that existing machines needed to evolve to meet the demands of modern rational governance and scientific enquiry.⁷

b Agricultural extension is an education and support system for farmers to access, understand and apply scientific knowledge, new technologies, and practices for improved livelihood of smallholder farm families. It serves to bridge the gap between research and farming communities in terms of information dissemination, troubleshooting with a view to capacity building.

The subject matter has evolved through two global phases, the AI winter and the AI spring.⁸ The AI winter (1950-90) was marked by reduced interest in AI research, failures in machine translation, the abandonment of networks and projects, and the withdrawal of funding from programmes like DARPA (Defense Advanced Research Projects Agency) and the Strategic Computing Initiative.⁹ However, since the '90s, there have been strides in AI, particularly through self-driving cars, chatbots, and digital assistants. Despite warnings of “digital slavery”, technology pioneers have urged countries to prepare for AI disruption and develop adequate safeguards to minimise potential risks. However, a separate stream of related literature is being developed to focus on the transparent and responsible deployment of AI.¹⁰

In India, AI research began in the 1960s when Prof. H. N. Mahabala, upon his return from the Massachusetts Institute of Technology (MIT), introduced a course in AI at IIT Kanpur. The Indian government, in partnership with the UN Development Programme (UNDP), launched the Knowledge-Based Computing System (KBCS) programme in the 1980s as part of the Indian Fifth Generation Computer Systems (FGCS) research programme. Institutes like the Indian Institute of Science (IISc), IIT Madras, Indian Statistical Institute (ISI) Kolkata, and the Tata Institute of Fundamental Research (TIFR) were established as nodal agencies to develop critical aspects of AI in India. Between 1986 and 1995, these centres received INR15 million in funding, producing approximately 15 PhDs and employing 20 to 35 full-time researchers.¹¹ A number of AI-based applications have emerged, including IIT Madras' 'Eklavya', a knowledge-based programme designed to support community health workers in diagnosing symptoms of illness in toddlers, CDAC's 'Sarani', a flight scheduling expert system, and IISc's Computer Vision-based image processing facility. India's R&D capabilities in AI have grown steadily from 2010-16, with national institutes like IISc, IIT (Bombay, Delhi, Madras, Kanpur, Kharagpur), IIIT Hyderabad, and ISI Kolkata ranking among the top universities and research institutes for AI in India.¹² India ranks 10th globally in terms of the number of PhDs in AI, and 13th in terms of presentations in top AI research conferences. However, challenges remain in developing, adopting, and using AI in India.¹³

Between 2010 and 2016, universities in the United States (US), China, and Japan dominated AI research by offering new courses, establishing research facilities, and forming industry partnerships. Chinese universities, particularly Peking and Tsinghua, followed suit by utilising public funding and extensive partnerships with private companies, demonstrating their early leadership in AI research.¹⁴ Countries are increasingly recognising the economic and social benefits of AI development and application. China and the United Kingdom (UK) predict that 26 percent and 10 percent of their respective GDPs will comprise AI-related activities and businesses by 2030.¹⁵

In the past 18–24 months, there has been an increasing public visibility in the global landscape of governments, multilateral organisations, technology companies, and research institutions that have been adopting a range of policy positions on AI, from regulation and ethical principles to national AI strategies and commitments to public investment. Over 60 countries (in North America, Europe, Asia, Africa and Latin America) are mobilising their policy communities to steer them toward the development of national AI ecosystems by devising strategies and funding schemes as well as institutional processes.¹⁶ The Indian government has announced a pact between the IndiaAI Mission and the Parliament to utilise parliamentary data to establish an internal AI system. The Indian Parliament has gathered massive datasets in numerous languages over the years, providing a solid foundation for training models.¹⁷ The transformative potential of AI arises from its capacity to address the needs of a multitude of sectors; and India's agriculture sector—the focus of the present article—has already begun to harness data for a wide range of AI solutions to be built.^c

c Multilingual public datasets like the Bhashini language corpus (National Language Translation Mission 2022) and the Soil Health Card database, which has accumulated data on over 230 million farmers since 2015, are increasingly being recognised as foundational for developing AI and Machine Learning applications in governance and agriculture. Data from the PM-Kisan programme launched in 2019, and the AgriStack pilots of 2021–2024, are also proving beneficial.

Effective AI at scale, however, is conditional not just on access to data and sectoral readiness, but also on wider political, legal and infrastructure environments that frame the development of AI. The Indian government and businesses have, for instance, begun working on creating a domestic graphics processing unit (GPU) chip.¹⁸ Datasets are essential for building large language models (LLMs), and the Government of India is working towards making more of them publicly available. Finally, to run large-scale computation, large amounts of energy are required, and the funding for and development of smart energy solutions will increasingly need to be explored.

Shaping AI Development and Implementation

The IndiaAI Mission, launched in 2024 with a substantial allocation of INR210,300 crore, aims to provide essential infrastructure and computational resources for AI-driven innovation. The establishment of the country's inaugural large-scale GPU infrastructure is a milestone, as it allows AI technology development and application to move beyond barriers among industries and promote innovation across different sectors.¹⁹ A shared computing facility featuring 18,693 GPUs is currently under development, representing almost two-thirds of the computational capacity accessible to OpenAI's ChatGPT. This development positions India alongside leading nations equipped with advanced AI computational resources. The infrastructure is currently under development at a rate of US\$1 per hour, compared to the global rate of US\$2.5-US\$3 per hour. This initiative is expected to enhance AI-driven innovation^d and enable smaller entities to compete with more dominant forces in the technology sector.²⁰

d In particular, it will support startups and researchers in India to develop scalable solutions addressing critical issues like agricultural productivity, climate-resilient farming, efficient resource use, renewable energy, personalised healthcare focusing on mother and child care, disease prediction, and disaster response. These efforts align with the national strategy for AI, which includes priorities such as large-scale soil health management utilizing IoT and AI, as well as multilingual natural language understanding for agricultural extension services.

Data serves as the fundamental support for AI. In the absence of high-quality datasets, even the most advanced AI models are unable to achieve optimal performance. The Indian government is working towards developing the IndiaAI Dataset Platform, a massive undertaking aimed at making non-personal datasets accessible for research and development purposes.²¹ India has established Centres of Excellence (CoEs) in critical domains such as healthcare, agriculture, sustainable cities, and education to maintain its leadership in AI innovation. These centres will focus on industry-specific applications and provide advanced AI expertise to the youth.

Subsequently, the country is developing indigenous AI models to promote a self-sufficient and language-inclusive ecosystem. These include BharatGen, Sarvam-1 AI Model, Digital India BHASHINI, Chitralekha, and SML's Hanooman–Everest. BharatGen aims to transform governance, enhance public service delivery efficiency, and foster citizen engagement. The Sarvam-1 AI Model is considered sophisticated, with two billion parameters and 10 prominent Indian languages. Digital India BHASHINI facilitates internet and digital services in Indian languages, while Chitralekha is an open-source video trans-creation platform. These initiatives look to cultivate an AI ecosystem reflective of India's rich linguistic diversity and reduce reliance on Western AI paradigms.

AI in Agriculture

AI has become a crucial tool in agriculture, enabling farmers to analyse market demand, manage risk, breed seeds, monitor soil health, protect crops, and observe crop maturity. Companies like Descartes Labs, a New Mexico-based entity, use AI to analyse satellite imagery and weather data to provide valuable insights on optimal planting times and the best crops to grow. Intello Labs, an Indian startup, uses AI to analyse fruits and vegetables, detect defects, and prevent crop failures. It helps breed seeds by collecting data on plant growth, identifying the best-performing plant varieties, and crossbreeding them to create better hybrids. Agrocared, a Dutch company, uses AI-powered hardware and software to collect data from soil samples, providing farmers with accurate estimates of missing nutrients and overall soil status. AI can also monitor plant health to spot and predict diseases, identify and remove weeds, and recommend effective pest treatment. Taranis, a

precision agriculture startup, uses computer vision and ML to analyse high-resolution images of crops, providing plant insights to identify signs of stress or disease. Using AI to predict crop maturity can lead to higher accuracy rates than human observers, resulting in cost savings and higher profits for farmers.²²

AI refers to programs that can think, acquire new knowledge, and resolve complex issues. It is synthetic, digital, and artificial, distinguishing it from biological intelligence. AI in farming helps farmers deliver real value and guides society towards the 'fourth industrial revolution' through smart farming practices and decision-making tools. Precision agriculture (PA) is one such strategy that combines information technology with farm machinery and farm management. PA serves four purposes: autonomous field navigation, field-based change detection, data mapping and reporting, and farm management zone recommendation. AI-enabled PA has the potential to enhance agricultural profitability while decreasing environmental consequences.²³ However, it also has undesirable side effects, such as concerns about data ownership and sharing, storage and the security of information, and the distribution of benefits between farmers and agritech corporations. These problems worsen preexisting ethical concerns over data ownership and sow seeds of mistrust in future AI solutions. India aims to make AI accessible to all by emulating its Digital Public Goods model. This will increase the availability of AI for the general public while discouraging monopolies and guaranteeing interoperability. It will also boost the usage of open-source technology, as well as transparency, inclusiveness and cooperation, while paving the way for adequate safeguards in this domain. Further, the widespread availability of AI-as-a-service can stimulate innovation in the field.

The technology can perform cognitive tasks like thinking, perceiving, learning, problem-solving, and decision-making. There has been a remarkable evolution with advances in data collection, processing, and computational power. The utility of AI has expanded, enabling it to potentially solve India's socio-economic challenges, such as improving agricultural yields, thereby enhancing productivity and driving growth in underserved geographies.²⁴ The AI-as-a-service approach suggests identifying sectors with the greatest potential for AI adoption and inviting the government to help co-develop implementation roadmaps. For instance, India's agriculture sector requires multilayered technology

infusion and coordination among stakeholders. Private sector efforts may not be financially optimal or efficient, necessitating sustained government intervention. The market size for AI in agriculture is expected to grow exponentially from US\$2.35 billion in 2020 to US\$10.83 billion by 2025 at a Compound Annual Growth Rate (CAGR) of 35.6 percent during the forecast period.²⁵

AI Transforming Farming for India's Smallholders

AI is revolutionising agriculture in India, helping smallholder farmers overcome challenges such as volatile weather, bug outbreaks, and resource scarcity.²⁶ AI-powered tools, such as crop disease identification apps and weather prediction models, are assisting farmers in making data-driven decisions that increase yields and save costs.²⁷ To make the technology accessible in rural areas, startups and government initiatives utilise chatbots, satellite imaging, and IoT (Internet of Things)-enabled sensors to provide real-time data in local languages.²⁸ India has the potential to become a global leader in inclusive agritech innovation, improving rural livelihoods.

AI is being used in pest forecasting models and satellite-based NDVI (Normalised Difference Vegetation Index) analytics for soil health monitoring. State-led projects like Telangana's 'Rythu Bandhu' plan use AI and smart irrigation to reduce water wastage. However, challenges such as data governance policies, dispersed landholdings, and low digital literacy in rural areas persist.²⁹ Wadhvani AI's pest forecasting models and satellite-based NDVI analytics optimise input consumption and minimise yield losses. IBM's Watson Decision Platform and NITI Aayog's AI-powered 'Cropin' are among the scalable solutions prioritised under the National AI Strategy (2021) and Digital Agriculture Mission (2021-2025). State-led projects like the Rythu Bandhu plan to use AI and smart irrigation to reduce water wastage by 30 percent.³⁰ India can provide smallholders with FAIR (Findable, Accessible, Interoperable, and Reusable) data and democratise precision agriculture by integrating AI into the PM-KISAN infrastructure and AgriStack's federated databases. This tech-driven transition could lead to India becoming a global leader in inclusive agritech innovation, improving rural livelihoods.³¹

Key Challenges

India's agriculture and allied sectors remain an integral part of the country's workforce, contributing to 49 percent of the workforce and 16 percent of GDP. To maintain a growth rate of 8-10 percent, the sector must grow by at least 4 percent. The government has prioritised doubling farmers' income as a national agenda, focusing on supply chain perspectives and productivity augmentation.

Despite progress, the sector remains dependent on unpredictable variables and has a weak supply chain. AI has the potential to drive a food revolution and meet the global demand to produce 50 percent more food and cater to an additional 2 billion people by 2050. It can address challenges such as inadequate demand prediction, the lack of assured irrigation, and overuse/misuse of pesticides and fertilisers. Examples include real-time advisory for crop yield improvement, advanced pest detection, and crop price prediction for sowing practices.

India's dependence on resource-intensive agricultural practices, such as land degradation, soil fertility reduction, and increased reliance on inorganic fertilisers, has led to agricultural distress. The sector suffers from poor resource utilisation, with low production quantum and productivity. India's cereal yield is much lower than those of countries like China and the US, which have higher technology adoption and efficient resource usage.³² Water use in agriculture is also high and sub-optimal, making the sector a net exporter of water and posing questions about India's long-term agronomic sustainability, as it is responsible for 89 percent of extracted groundwater consumption.

AgriTech has emerged as an area of opportunity for AI-focused startups. For example, Intello Labs uses image-recognition software to monitor crops and predict farm yields, while Trithi Robotics deploys drone technology to allow farmers to monitor crops in real time and provide precise soil analysis. Sat-Sure relies on ML techniques to assess farm images and predict future yields. Image recognition and deep learning models enable distributed soil health monitoring without the need for laboratory testing infrastructure. AI solutions integrated with data signals from remote satellites and local image capture on farms enable farmers to take immediate action to restore soil health.

Vulnerability to varying weather patterns, such as an increase in temperature, changes in precipitation levels, and groundwater density due to its rain-dependent nature, are the other challenges prevalent in the sector. AI can predict advisories for sowing, pest control, and input control, ensuring increased income and stability for the agricultural community. Remote sensing allows for holistic crop monitoring and provides additional insights to extension workers/farmers. Berlin-based agricultural tech startup PEAT has developed a deep learning application called Plantix,^e which identifies potential soil defects and nutrient deficiencies.

AI can be used to predict advisories for sowing, pest control, and input control, ensuring increased income and stability for the agricultural community. Remote sensing can monitor agronomic factors like vegetation health and soil moisture, providing additional insights to extension workers and farmers. Image classification tools combined with remote and local sensed data can optimise the utilisation and efficiency of farm machinery, including weed removal, early disease identification, produce harvesting, and grading. AI tools provide round-the-clock monitoring of high-value products at all levels of plant growth.

Microsoft, in collaboration with ICRISAT, developed an AI Sowing App powered by the Microsoft Cortana Intelligence Suite, which includes ML and Power BI—its business intelligence platform. The app dispatches sowing advisories to participating farmers on the optimal date to sow,³³ eliminating the need to install sensors in fields or incur capital expenditure, as all farmers need is a feature phone capable of receiving text messages. The app also provides essential information, such as the optimal sowing date, soil test-based fertiliser application, farm yard manure application, seed treatment, and optimum sowing depth.

e The German startup PEAT (Progressive Environmental and Agricultural Technologies) developed the deep learning application Plantix.

AI has the potential to revolutionise agriculture in both rural and urban settings; however, there are persistent challenges, such as infrastructure gaps, accessibility, digital literacy, and the lack of a supportive regulatory environment. Accurate decision-making in AI systems requires high-quality, real-time data, which can be challenging due to inconsistent data collection methods and a lack of reliable sources. Traditional farming practices and the learning curve associated with AI adoption may also deter older generations from adopting the range of technologies.³⁴ Rural areas often lack the infrastructure for implementing AI, hindering real-time data collection and analysis. Further, the initial investment in AI technologies can be expensive for small-scale farmers, and low digital literacy hinders widespread adoption.

Additionally, the small landholdings in rural areas make it difficult to implement large-scale, uniform AI systems for optimising productivity and sustainability. Despite the potential benefits of AI, these challenges hinder the widespread adoption of AI-driven solutions in agriculture.³⁵ Most farmers worldwide, particularly smallholders, lack the necessary resources to implement these technologies. They typically have limited access to technical training and the financial resources needed to purchase the necessary equipment and software. They also lack the resources to effectively implement these technologies.

The rise of AI has prompted concerns about the use of complex systems without revealing the data used to train the model or the algorithm design.³⁶ This can lead to unfair or incorrect decisions if used in ‘un-designed contexts’.^f Without careful upfront design and safety precautions, some AI systems may be prone to errors or breakdowns when exposed to minor perturbations in data. Ongoing monitoring and fail-safe designs are vital, especially in safety-critical systems like the

f Un-designed contexts include situations, populations or decision-making environments that were not foreseen during a model’s original training, or system design, such as those caused by variations in socio-economic status, cultural practices, language use and legal systems and data distributions. When actual AI systems are used in such scenarios, these underlying assumptions might not be valid anymore leading to bias, misclassification or wrong decisions.

application of pesticides, insecticides, the sowing of seeds, harvest timings, and predicting weather. AI's nature as labour substitution can lead to inequality between labour and capital, and within the workforce, invariably raising public policy concerns. Ultimately, the impact on human employment depends on the way organisations deploy AI tools and training.³⁷

Conclusion

AI in precision agriculture allows farmers and farm managers to implement focused and accurate farming practices based on agroclimatic field measurements.³⁸ This technology has led to new possibilities for food production with minimal negative impact on ecology and climate. However, concerns include data ownership, privacy, security, and accountability.³⁹ To address these issues, farmers should be involved in AI development, ensuring that they are at the centre of the design process.

The accuracy and reliability of AI system recommendations are crucial; however, the AI model's design and usability are equally important.⁴⁰ Fostering inclusivity, trust, and long-term thinking can be achieved through innovation and governance outcomes. Reconnecting values, natural environments, and social contexts as starting points for dialogue can level the playing field between AI developers and farmers' knowledge production and implementation.⁴¹ Algorithmic bias is a potential issue due to digitalisation in agriculture, but overcoming farmers' scepticism can increase their trust in the technology.

Diwakar Kumar is DST Policy Fellow at the Centre for Policy Research, National Institute of Science Education and Research- Bhubaneswar.

Endnotes

- 1 *National Strategy for Artificial Intelligence: #AIforAll*, NITI Aayog, 2018, <https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf>.
- 2 Edyta Andrzejak, "AI - Powered Digital Transformation: Tools, Benefits and Challenges for Marketers - Case Study of LPP," *Procedia Computer Science* 219, 2023, <https://doi.org/10.1016/j.procs.2023.01.305>.
- 3 Diwakar Kumar, "Development of Agricultural Bioinformatics in India: Issues and Challenges," *Asian Biotechnology and Development Review* 20, no. 3 (2018), https://www.researchgate.net/publication/335586956_Development_of_Agricultural_Bioinformatics_in_India_Issues_and_Challenges.
- 4 Lefteris Benos et al., "Machine Learning in Agriculture: A Comprehensive Updated Review," *Sensors* 21, no. 11 (2021), <https://doi.org/10.3390/s21113758>.
- 5 Alan M Turing, "Computing Machinery and Intelligence," *Mind* 49: 433-460 (1950), <https://courses.cs.umbc.edu/471/papers/turing.pdf>.
- 6 Ryan Stock et al., "Arrays and Algorithms: Emerging Regimes of Dispossession at the Frontiers of Agrarian Technological Governance," *Earth System Governance* 100137, no. 12 (2022), <https://doi.org/10.1016/j.esg.2022.100137>.
- 7 Turing, "Computing Machinery and Intelligence."
- 8 Mark Ryan et al., "Identifying Key Ethical Debates for Autonomous Robots in Agri-Food: A Research Agenda," *AI And Ethics* 2, 2021, <https://doi.org/10.1007/s43681-021-00104-w>.
- 9 Purushottam Kaushik, "AI For India 2030: A Blueprint For Inclusive Growth and Global Leadership," *World Economic Forum*, January 22, 2025, <https://www.weforum.org/stories/2025/01/ai-for-india-2030-blueprint-inclusive-growth-global-leadership/>.
- 10 Ryan Stock et al., "Arrays and Algorithms: Emerging Regimes of Dispossession at the Frontiers of Agrarian Technological Governance," *Earth System Governance* 100137, no. 12 (2022), <https://doi.org/10.1016/j.esg.2022.100137>.
- 11 Emma White et al., "Report From the Conference: Identifying Obstacles to Applying Big Data in Agriculture," *Precision Agriculture* 22, no. 1 (2020), <https://doi.org/10.1007/s11119-020-09738-y>.
- 12 Ryan et al., "Identifying Key Ethical Debates for Autonomous Robots in Agri-Food: A Research Agenda."
- 13 Ryan et al., "Identifying Key Ethical Debates for Autonomous Robots in Agri-Food: A Research Agenda."
- 14 Zhao Litao, "Driving AI Excellence: How China's Top Universities Are Driving the Future of Technology," National University of Singapore EAI Background Brief No. 1828, https://research.nus.edu.sg/eai/wp-content/uploads/2025/05/EAIIBB-No.-1828-AI-Excellence_Chinas-Top-Universities-synopsis_exsum.pdf.

- 15 Arthur Mutambara, *Artificial Intelligence: A Driver of Inclusive Development and Shared Prosperity for The Global South* (Florida: CRC Press, 2025).
- 16 NASSCOM, "Implications of AI on the Indian Economy," <https://nasscom.in/knowledge-center/publications/implications-ai-indian-economy>.
- 17 Leanne Wiseman et al., "Farmers and Their Data: An Examination of Farmers' Reluctance To Share Their Data Through The Lens of The Laws Impacting Smart Farming," *NJAS - Wageningen Journal of Life Sciences* 90, 2019, <https://doi.org/10.1016/j.njas.2019.04.007>.
- 18 Daniel Varona et al., "Discrimination, Bias, Fairness, and Trustworthy AI," *Applied Sciences* 12, no. 5826 (2022), <https://doi.org/10.3390/app12125826>.
- 19 Sara Araújo et al., "Machine Learning Applications in Agriculture: Current Trends, Challenges and Future Perspectives," *Agronomy* 13, no. 12 (2023), <https://doi.org/10.3390/agronomy13122976>.
- 20 NASSCOM, "Implications of AI on the Indian Economy"
- 21 Shreshtha Verma, "How India is Democratizing AI: A Step Toward a Tech-Powered Future," *TICE News*, February 11, 2025, <https://www.tice.news/tice-trending/how-india-is-democratizing-ai-8710346>.
- 22 A Balkrishna et al., "Agricultural Mobile Apps Used in India: Current Status and Gap Analysis," *Agricultural Science Digest* 41, 2021, <https://arccjournals.com/journal/agricultural-science-digest/D-5140>.
- 23 Verma, "How India is Democratizing AI: A Step Toward a Tech-Powered Future"
- 24 Maaz Gardezi et al., "Artificial Intelligence In Farming: Challenges and Opportunities For Building Trust," *Agronomy Journal* 116, no. 3 (2023), <https://doi.org/10.1002/agj2.21353>.
- 25 Ryan et al., "An Interdisciplinary Approach to Artificial Intelligence in Agriculture," *NJAS: Impact in Agricultural and Life Sciences* 95, (2023), <https://www.tandfonline.com/doi/full/10.1080/27685241.2023.2168568>.
- 26 Mutambara, *Artificial Intelligence: A Driver of Inclusive Development and Shared Prosperity for The Global South*.
- 27 Sara Araújo et al., "Machine Learning Applications in Agriculture: Current Trends, Challenges and Future Perspectives," *Agronomy* 13, no. 12 (2023), <https://doi.org/10.3390/agronomy13122976>.
- 28 Balkrishna et al., "Agricultural Mobile Apps Used in India: Current Status and Gap Analysis," *Agricultural Science Digest* 41, (2021), <https://arccjournals.com/journal/agricultural-science-digest/D-5140>.
- 29 Diwakar Kumar et al., "Heritage Crop Diversity in Transition: Ragi's Role in Mitigating Food Security Challenges in India," *Journal of Heritage Management* 9, (2024), <https://journals.sagepub.com/doi/abs/10.1177/24559296241303042>.
- 30 Kumar et al., "Heritage Crop Diversity in Transition: Ragi's Role in Mitigating Food Security Challenges in India".
- 31 NASSCOM, "Implications of AI on the Indian Economy"
- 32 PRS India, State of Agriculture in India, 2017, <https://prsindia.org/policy/analytical-reports/state-agriculture-india#:~:text=Although%20India%20ranks%20third%20in,is%20the%20second%20highest%20producer>.

- 33 Johanna Thoma et al., *Risk Imposition by Artificial Agents: The Moral Proxy Problem* (Cambridge: Cambridge handbook of responsible artificial intelligence Interdisciplinary perspectives Cambridge University Press, 2022), pp 50.
- 34 Diwakar Kumar, "Climate Change, A Strong Threat to Food Security in India: With Special Reference to Gujarat," in *Climate Change Impacts on Natural Resources, Ecosystems and Agricultural Systems*, ed. Chaitanya B. Pande, Kanak N. Moharir, Sudhir Kumar Singh, Quoc Bao Pham and Ahmed Elbeltagi (Switzerland: Springer Cham, 2023), 153-173.
- 35 George Lawton, "Democratization of AI Creates Benefits and Challenges," *TechTarget*, August 26, 2024, <https://www.techtarget.com/searchenterpriseai/feature/Democratization-of-AI-creates-benefits-and-challenges>.
- 36 Emma White et al., "Report from the Conference Identifying Obstacles to Applying Big Data in Agriculture," *Precision Agriculture Springer* 22, (2020), <https://doi.org/10.1007/s11119-020-09738-y>.
- 37 Edyta Andrzejak, "AI - Powered Digital Transformation: Tools, Benefits and Challenges for Marketers - Case Study of LPP," *Procedia Computer Science* 219, (2023), <https://doi.org/10.1016/j.procs.2023.01.305>.
- 38 NASSCOM, "Implications of AI on the Indian Economy"
- 39 "AI in Agriculture: Pros, Cons and How to Stay Ahead," BPM, September 12, 2025, <https://www.bpm.com/insights/ai-in-agriculture/>.
- 40 Mutambara, *Artificial Intelligence: A Driver of Inclusive Development and Shared Prosperity for The Global South*.
- 41 Ryan et al., "Identifying Key Ethical Debates for Autonomous Robots in Agri-Food: A Research Agenda."

Democratising AI: The International Trade Perspective

Shailja Singh

Artificial Intelligence (AI) and international trade have an evolving, multi-faceted relationship, with AI rapidly becoming a key driver of global economic growth. Its projected contribution to the global economy is approximately US\$15 trillion by 2030.¹ Generative AI alone has the potential to add between US\$2.6 trillion and US\$4.4 trillion to the global economy annually.² Furthermore, the global AI market is also expected to see exponential growth, rising from US\$189 billion in 2023 to US\$4.8 trillion by 2033.³

AI is both reshaping the global trade landscape and being shaped by it in three key ways: It is transforming how trade is conducted, redefining what is being traded, and influencing how AI itself is governed through trade rules.

First, AI is transforming international trade operations and logistics across different verticals. In the realm of customs administrations, 25 percent of customs authorities^a currently use AI technologies, while an additional 25 percent have plans to adopt them soon.⁴ These tools reduce clearance times, enhance efficiency, and assist in combating fraud, making them particularly valuable for developing countries seeking to integrate more effectively into the global trading system.

Second, AI is reshaping the nature of goods and services being traded, as well as how they are being delivered. The rise in digitally delivered services and the increased demand for AI-related hardware illustrate this shift. For instance, the shift is evident in hardware, with the AI chips market experiencing rapid growth, projected to increase from US\$61.5 billion in 2023 to US\$621 billion by 2032.⁵

Third, countries are responding to these technological transformations by developing international rules that govern and aid in these developments. Digital trade is increasingly being addressed in international agreements, with 116 regional trade agreements (RTAs), or 33 percent of the total, now incorporating digital trade provisions that directly or indirectly impact AI adoption and governance. However, the scope and depth of these provisions vary across agreements.⁶

In parallel, countries are also racing to develop their domestic AI strategies. While these are typically broad and not limited to trade-specific issues, they reflect a growing appreciation of the need to regulate this general-purpose technology that has wide-ranging ramifications. Notably, as of 2023, 75 economies have unveiled

a Customs authorities in Brazil, China, Germany, Dubai, and Singapore are among those already employing AI technologies.

national AI strategies, a substantial increase from the three in 2017.⁷ Approximately 30 percent of developing economies have established national AI policies. However, among the least-developed countries, only Uganda has implemented such a policy to date.⁸

As AI becomes more embedded in trade systems, questions of equity and inclusion arise, particularly for developing countries like India. This article aims to examine AI from an international trade perspective, through the lens of democratisation, with a focus on the current landscape, key developments, and the associated challenges.

The Current Landscape

AI's Impact on the Conduct of International Trade

AI can optimise trade operations and logistics in several ways: it can help reduce costs and increase efficiency. For instance, AI can help with route optimisation by reducing shipping times and costs.⁹ It can also assist in making customs procedures more efficient and less prone to errors, as well as help both traders and governments in accurately classifying goods for customs purposes.¹⁰ AI can also aid in supply chain optimisation, making the entire chain of operations more efficient by predicting disruptions.¹¹ Through automated document processing, risk assessment, and compliance checks, AI can also be useful in trade finance.¹² This is significant given that an estimated 80 to 90 percent of world trade relies on instruments of trade finance like trade credit and insurance/guarantees.¹³ AI can also contribute to addressing market-access-related challenges through real-time translations powered by data analytics and AI-driven Natural Language Processing (NLP).^{14,15}

Dubai serves as a fitting example of AI's growing role in trade operations and facilitation, wherein a suite of AI applications is being used to enhance customs and trade processes.¹⁶ For example, travellers can now electronically file customs declarations using the iDeclare app, which also allows them to pre-declare products using images to ascertain their HS codes and the corresponding duties. The AI Munasiq app provides duty rates and limits, analyses item descriptions or photos, and assists customers in finding the relevant HS codes. Meanwhile, Dubai's Smart Refund System leverages Robotic Process

Automation to expedite refunds and claims. For authorised economic operators, customs clearance is expedited through AI-powered remote robotic inspections equipped with thermal and infrared cameras. Similar advancements can be seen in Abu Dhabi and Singapore, which host two of the busiest ports in the world. In 2022, Abu Dhabi developed an AI-based application for the real-time measurement of the time required to release goods at the border.¹⁷ These innovations illustrate how AI can have a transformative impact on trade operations.

India has also seen promising developments in the recent years. In 2024, Cusbuzz—the country's first AI-powered customs duties app, was launched.¹⁸ It offers users a real-time, mobile-friendly platform that provides accurate duty calculations, and provides valuable insights into industry trends and trade flows. The Central Board of Indirect Taxes and Customs has also developed an advanced AI-driven data analytics architecture designed to improve enforcement and streamline operations.¹⁹ This system is built around five interdependent layers, each working together to enhance targeting and detection capabilities across the customs process. The results have been game-changing. AI models have helped uncover smuggling attempts, including the detection of 3,000 kilograms of heroin smuggled from Afghanistan through the Mundra Port in Gujarat in 2021.¹⁰ In another case, 7.2 million sticks of foreign-brand cigarettes were seized at the Nhava Sheva Port, while poppy seeds concealed in a consignment were intercepted at the Chennai Port.²¹ These examples highlight how AI is becoming an asset in enhancing security and improving trade compliance in India.

AI's Impact on What Is Traded

AI is a key to boosting productivity and driving the creation of new, innovative services across sectors. As businesses adopt AI technologies, the demand for complementary infrastructure, such as improved Information and Communications Technology (ICT) networks and upgraded Information Technology (IT) equipment, including AI chips, is also expected to increase. According to the WTO, the intermediary services category is expected to see the most substantial increase, with cumulative growth projected to reach nearly 18 percentage points between 2023 and 2040.²² This trend reinforces the expanding role of digital trade and the need for up-to-date infrastructure and policy frameworks to support it.

International Trade Rules and AI Governance

Over the past decade, there has been a notable rise in governance issues pertaining to digital trade and electronic commerce. These are in addition to the traditional rules on goods, services, and intellectual property rights contained in existing agreements which impact AI creation and adoption. Traditional trade rules, while relevant, pose their own set of challenges when applied to AI, given their rigid segregation along the lines of goods and services. AI technologies and products do not often fit neatly into these binary classifications. The goods vs. services distinction in trade rules becomes especially important in the context of technical standardisation. While there are detailed rules regarding the technical standardisation of goods, the same is underdeveloped, or altogether absent, for services.

Additionally, digital trade rules are increasingly embedded in comprehensive RTAs as well as new-age and more specialised digital economy partnership agreements.²³ At the WTO, a subset of members has negotiated a plurilateral agreement on electronic commerce which is yet to be formally incorporated into the rules. At the multilateral level, the WTO decision to impose a temporary moratorium on customs duties on electronic transmission continues to be one of the most divisive areas of discussion.²⁴

At the heart of all international rule-making lies the tussle between a push for deregulation and open digital markets, and the demand for digital sovereignty and increased control over the data generated within a country for its own use and development. Several trade provisions can potentially impact the development and adoption of AI. While some agreements have low-ambition provisions specific to AI, such as requiring parties to work towards adopting ethical governance frameworks that support the trusted, safe, and responsible use of AI technologies,²⁵ others have more horizontal provisions. These relate to cross-border data flows, the location of computing facilities, non-discriminatory treatment of digital products, and disclosure of source code as well as ICT products that use cryptography. Personal data protection regulations are also becoming increasingly relevant, particularly as AI systems rely heavily on the collection and processing of large volumes of data.

The India–UK Comprehensive Economic and Trade Agreement and the India–UAE Comprehensive Economic Partnership Agreement are India’s only concluded free trade agreements with detailed chapters on digital trade. These do a delicate task of balancing digital sovereignty with openness and predictability in international trade rules in this area.²⁶

Key Challenges

As digital technologies permeate every aspect of life and play an increasingly critical role in trade, addressing issues related to the accessibility, availability, and affordability of digital infrastructure becomes more important. Studies indicate that a 1-percent increase in domestic digital connectivity correlates with a 1.5-percent rise in international trade, underscoring its importance.²⁷ The accessibility, availability, and affordability of digital infrastructure are foundational to AI development, use, and governance—the factors key to democratising the technology.²⁸ Therefore, the challenges associated with the democratisation of AI in international trade mirror those seen in other domains as they relate to basic infrastructural and connectivity issues.

Disparities persist, particularly in developing economies, where limited infrastructure and high costs impede digital inclusion. In 2024, 2.6 billion people—one-third of the global population—still did not have access to the internet.²⁹ Further, internet usage remains closely tied to a country’s level of development. In high-income nations, 93 percent of the population has access to the internet, nearing universal connectivity. In sharp contrast, only 27 percent of people in low-income countries are online.³⁰

In addition to basic infrastructural and connectivity issues, the cost of AI infrastructure is also steep, making AI adoption a tall task for many developing countries. As per industry estimates, the annual cost for cloud compute for training mid-sized models is estimated to be between US\$50,000 and US\$500,000, depending on usage patterns and model size.³¹ Further, establishing even a comparatively modest AI cluster can involve significant investment. A setup comprising approximately a dozen NVIDIA H100 GPUs, along with high-speed storage and the appropriate cooling infrastructure, typically ranges from US\$500,000 to

US\$1 million.³² Further, limited AI expertise and training programmes hinder workforce readiness in AI-driven trade sectors. Data quality and availability can also pose challenges in developing effective AI models for trade.³³

Big Tech firms dominate AI innovation, exercising disproportionate influence on its growth trajectory and usage, with patents and proprietary AI models limiting open access. This makes it difficult for firms in the Global South to use them, further widening the digital divide. The Indian government has made great efforts to disrupt this status quo by, inter alia, developing a comprehensive AI strategy addressing challenges related to compute and semiconductor infrastructure, AI talent, and workforce development, and by launching the IndiaAI Datasets Platform. As part of the IndiaAI Mission, the government has allocated INR10,300 crore (US\$1.24 billion) over five years to strengthen AI capabilities in the country.³⁴

Domestic as well as international trade rules also impact global AI diffusion. For instance, domestic rules in the form of export controls on AI-related chips and equipment can hamper supply and access to others.³⁵ Several nations have adopted diverse approaches to AI governance and are at different stages of the regulatory ladder, making it harder to establish global AI trade norms.

Lastly, the strict demarcation between goods and services in international trade rules appears to be ill-suited for the blurred boundaries that characterise AI technologies and products. An update to these rules is thus long overdue. Moreover, the digital trade-specific rules found in RTAs do not take into account the issues surrounding data sovereignty and the digital divide, highlighting the need for revision and reform to better align them to the needs of developing countries.

Recommendations and the Way Forward

Several steps can be taken both at the domestic and international level to harness the full potential of AI as a tool for improving international trade. When it comes to optimising the trade and customs process, the Indian government can consider integrating AI into front-end operations to reduce turnaround times, and facilitate trade. For instance, current

efforts are limited to e-filing through the Indian Customs Electronic Gateway (ICEGATE).³⁶ These could be updated along the lines of global best practices, as seen in the case of Dubai.

Further, there are a number of initiatives being undertaken under the aegis of the IndiaAI Mission³⁷ that could be utilised for international trade purposes. For example, AI-driven NLPs could be deployed for international trade operations purposes, streamlining the export and import of goods, as demonstrated successfully by Singapore. The government could make use of the already implemented Digital India Bhashini initiative that aims to provide AI-driven language technology solutions for customs processes.³⁸ The IndiaAI Mission has also allocated a large volume of funds to upgrade the infrastructure and capacity needed for properly harnessing AI-related developments. These could be utilised for enhancing the capacity of MSMEs to fully integrate into the AI ecosystem. Additionally, the embedding of AI tools in customs and trade operations processes will also reduce the costs for those MSMEs that participate in international trade, making AI an important tool in promoting and enhancing the capacities of Indian MSMEs and overcoming the domestic digital divide.

Efforts could also be made to encourage open-source developments and government-business partnerships to make trade operations and customs processes more efficient. Open-source AI tools can democratise access to advanced technologies and be game-changers by enabling both sides to work closely with each other for compliance and facilitation. To further boost innovation, the government could also increase the availability of datasets relevant for international trade and customs on the Open Government Data platform³⁹ and the IndiaAI datasets. This is an area where special efforts could lead to tangible outcomes for Indian businesses looking to build AI tools to streamline trade and customs operations.

Lastly, in the context of international trade rulemaking, India is extremely well-positioned to spearhead a coalition of like-minded developing countries in formulating a model framework of digital trade rules that address the unique interests and concerns of developing countries. This proactive approach would enable a shift in the current dynamic wherein developing countries often find themselves responding

to digital trade provisions proposed by others, towards one where they assert their own offensive interests and shape the negotiating agenda in RTAs. A jointly developed model framework could serve as a valuable reference point for developing countries, and better reflect the needs of the Global South.

In conclusion, India's AI ecosystem and governance approaches appear to be moving in the right direction. With a few targeted adjustments to better align with the specific demands of international trade, particularly in advancing the goal of democratising AI, India is well-positioned to play a leading role in shaping the future development and deployment of AI in international trade.

Shailja Singh is a Legal Consultant and Associate Professor at the Centre for Trade and Investment Law, Indian Institute of Foreign Trade, New Delhi.

Endnotes

- 1 PricewaterhouseCoopers, “Sizing the Prize: What’s the Real Value of AI for Your Business and How Can You Capitalise?,” PwC, 2017, <https://www.pwc.com/gx/en/issues/artificial-intelligence/publications/artificial-intelligence-study.html>.
- 2 McKinsey & Company, “The Economic Potential of Generative AI: The Next Productivity Frontier,” 2023, <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>.
- 3 United Nations Conference on Trade and Development (UNCTAD), “AI Market Projected to Hit \$4.8 Trillion by 2033, Emerging as Dominant Frontier Technology,” 2025, <https://unctad.org/news/ai-market-projected-hit-48-trillion-2033-emerging-dominant-frontier-technology>.
- 4 WTO, *Trading with Intelligence: How AI Shapes and Is Shaped by International Trade*, Geneva, World Trade Organisation, 2024, https://www.wto.org/english/res_e/booksp_e/trading_with_intelligence_e.pdf.
- 5 SNS Insider, *AI Chip Market Size, Share, and Segmentation by Chip Type (CPU, GPU, ASIC, FPGA, others), by Technology (Generative AI, Machine Learning, Natural Language Processing, Computer Vision), by Function (Training, Inference), by End-User and by Regions: Global Forecast 2024–2032*, 2024, <https://www.snsinsider.com/reports/ai-chip-market-4525>.
- 6 J. López-González, S. Sorescu, and P. Kaynak, *Of Bytes and Trade: Quantifying the Impact of Digitalisation on Trade*, Paris, Organisation for Economic Co-operation and Development, 2023, 13, https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/05/of-bytes-and-trade-quantifying-the-impact-of-digitalisation-on-trade_17cd5677/11889f2a-en.pdf.
- 7 Stanford Institute for Human-Centered Artificial Intelligence, *AI Index Report 2024*, Stanford CA, Stanford University, 2024, 391, https://hai-production.s3.amazonaws.com/files/hai_ai-index-report-2024-smaller2.pdf.
- 8 “Trading with Intelligence: How Artificial Intelligence Is Shaping International”
- 9 Ozcan Ozturk, “The Impact of AI on International Trade: Opportunities and Challenges,” *Economies* 12, no. 11 (2024), 298, <https://www.mdpi.com/2227-7099/12/11/298>.
- 10 Thibo Clicteur and Dries Bertrand, “Leveraging AI for Customs Classification Purposes,” *WCO News*, October 18, 2023, <https://mag.wcoomd.org/magazine/wco-news-102-issue-3-2023/leveraging-ai-for-customs-classification-purposes/>.
- 11 Sumit Dutta and Glenn Steinberg, “How Generative AI in Supply Chain Can Drive Value,” *EY*, October 18, 2023, https://www.ey.com/en_in/insights/supply-chain/how-generative-ai-in-supply-chain-can-drive-value.
- 12 Gurkirpal S. Ahluwalia, “The Powerful Duo of AI and Low Code in Transforming Trade Finance,” *Global Trade Review (GTR)*, January 15, 2024, <https://www.gtreview.com/magazine/the-commodities-issue-2024/the-powerful-duo-of-ai-and-low-code-in-transforming-trade-finance/>.

- 13 World Trade Organization, “Trade Finance,” https://www.wto.org/english/thewto_e/coher_e/tr_finance_e.htm.
- 14 Cole Stryker and Jim Holdsworth, “What is NLP (Natural Language Processing)?,” IBM, August 11, 2024, <https://www.ibm.com/think/topics/natural-language-processing>.
- 15 Elina Noor and Binya Kanitroj, *Speaking in Code: Contextualizing Large Language Models in Southeast Asia*, New Delhi, Carnegie Endowment for International Peace, 2025, <https://carnegieindia.org/research/2025/01/speaking-in-code-contextualizing-large-language-models-in-southeast-asia?lang=en¢er=china>.
- 16 Ahmed Mahboob Musabih, “Transforming Trade: How Dubai Customs Is Harnessing AI for Enhanced Trade Facilitation and Border Control,” *WCO News*, October 18, 2023, 48–51, https://mag.wcoomd.org/uploads/2023/10/WCO_NEWS_102-2.pdf.
- 17 Mike Squirrell and Dan Rochon, “Abu Dhabi Launches Application for Real-Time Measurement of the Time Required to Release Goods,” *WCO News*, February 15, 2023, 51, https://mag.wcoomd.org/uploads/2023/03/WCO_News100.pdf.
- 18 Cusbuzz, “Indian Customs Import Duty Calculator,” <https://www.cusbuzz.com/>.
- 19 M. Ramesh and Sruti Vijayakumar, “How Indian Customs Leverages Data Analytics for Efficient Risk Management,” *WCO News*, October 18, 2023, 58–61, https://mag.wcoomd.org/uploads/2023/10/WCO_NEWS_102-2.pdf.
- 20 Ramesh and Vijayakumar, “How Indian Customs Leverages Data Analytics for Efficient Risk Management”
- 21 Ramesh and Vijayakumar, “How Indian Customs Leverages Data Analytics for Efficient Risk Management”
- 22 WTO, *Trading with Intelligence: How Artificial Intelligence Is Shaping International Trade*, Geneva, World Trade Organization, 2024, 32, https://www.wto.org/english/res_e/booksp_e/trading_with_intelligence_e.pdf
- 23 IMF, OECD, UNCTAD, World Bank, and WTO, *Digital Trade for Development*, Geneva, WTO Publications, 2023, https://www.wto.org/english/res_e/booksp_e/dtd2023_e.pdf.
- 24 World Trade Organization, *Work Programme on Electronic Commerce – Ministerial Decision*, WT/MIN(24)/38, March 2, 2024, <https://docs.wto.org/dol2fe/Pages/SS/directdoc.aspx?filename=q:/WT/MIN24/38.pdf&Open=True>.
- 25 Ministry of trade and Industry Singapore, “Digital Economy Partnership Agreement (DEPA) between New Zealand, Chile, and Singapore, Article 8.2,” June 12, 2020, <https://www.mti.gov.sg/-/media/MTI/Microsites/DEAs/Digital-Economy-Partnership-Agreement/Text-of-the-DEPA.pdf>.
- 26 Government of India, Ministry of Commerce and Industry, *India-UAE Comprehensive Economic Partnership Agreement: Chapter 9 – Digital Trade* (New Delhi: Ministry of Commerce and Industry, 2022), <https://commerce.gov.in/wp-content/uploads/2022/03/Chapter-9.pdf>.
- 27 Organisation for Economic Co-operation and Development (OECD), “Opportunities and Benefits of Digital Trade,” <https://www.oecd.org/en/topics/sub-issues/opportunities-and-benefits-of-digital-trade.html>.

- 28 Elizabeth Seger et al., “Democratising AI: Multiple Meanings, Goals, and Methods,” in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, Association for Computing Machinery, August 2023, 715–722, <https://dl.acm.org/doi/10.1145/3600211.3604693>.
- 29 International Telecommunication Union (ITU), *Measuring Digital Development: Facts and Figures 2024*, Geneva, ITU Telecommunication Development Bureau, 2024, https://www.itu.int/hub/publication/D-IND-ICT_MDD-2024-4/.
- 30 “Measuring Digital Development: Facts and Figures 2024”.
- 31 Pure Storage, “The True Cost of AI: Beyond the Hype,” CDO Trends, April 20, 2025, <https://www.cdostrends.com/story/4522/true-cost-ai-beyond-hype>.
- 32 Pure Storage, “The True Cost of AI: Beyond the Hype,”
- 33 “Trading with Intelligence: How Artificial Intelligence Is Shaping International Trade”
- 34 Ministry of Electronics and Information Technology, Government of India, <https://pib.gov.in/PressReleasePage.aspx?PRID=2108810>
- 35 “Trading with Intelligence: How Artificial Intelligence Is Shaping International Trade”
- 36 Ministry of Finance, “ICEFATE 2.0 – Indian Customs National Trade Portal,” <https://www.icegate.gov.in/>.
- 37 Ministry of Electronics and Informational Technology, Government of India, <https://pib.gov.in/PressReleasePage.aspx?PRID=2113095#>. 2020.
- 38 Ministry of Electronics and Informational Technology, “National Language Translation Mission,” Digital India BHASHINI Division, <https://bhashini.gov.in/>.
- 39 Government of India, “Open Government Data (OGD) Platform India,” <https://www.data.gov.in/>.

AI-Driven Healthcare Transformation: India's Approach to Open Innovation

*Viola Savy Dsouza
and Angela Brand*

Artificial Intelligence (AI) holds promise as a catalyst for system-wide innovation in response to the growing demands for equitable and accessible healthcare. As healthcare systems globally strive to become more efficient, AI-driven technologies offer an opportunity to improve access, affordability, and accuracy of medical services. However, these benefits are not universal, as AI-driven solutions often remain concentrated within proprietary systems and are further limited by high costs, inadequate infrastructure, and regulatory barriers, restricting their widespread adoption.¹

India's approach to AI is anchored in the principles of digital public infrastructure (DPI) and looks to democratise access to AI-driven healthcare solutions by ensuring that innovation is inclusive and scalable. The National Digital Health Mission (NDHM) and initiatives such as the IndiaAI Mission are foundational steps toward integrating AI into India's health ecosystem.^{2,3} These initiatives highlight the importance of open AI models in improving healthcare accessibility, fostering innovation, and reducing urban–rural disparities. While progress is evident, challenges remain in areas such as ethics, data privacy, infrastructure, and interoperability. Open AI models offer a shared foundation for equitable innovation, ensuring that AI serves as a public good rather than being monopolised. These models empower startups, researchers, and institutions to create localised solutions while fostering education and capacity-building for all. By making AI accessible and adaptable, these models can bridge the gap between innovation and real-world healthcare implementation.

This chapter evaluates empirical case studies, analyses systemic barriers, and provides evidence-informed recommendations to promote fair and sustainable AI integration in healthcare settings. It reflects on how AI-powered digital health interventions can scale efficiently while balancing regulatory and ethical considerations. As AI adoption accelerates, India stands at a critical juncture where strategic policy interventions, robust public-private partnerships, and ethical AI governance frameworks will determine the success of AI-driven healthcare transformation. India's approach to AI in healthcare integrates open, equitably shared resources with market-driven innovation to foster both accessibility and competitiveness.

The Quintuple Helix Model promotes co-creation by integrating five key stakeholders: the government, industry, academia, civil society, and the natural environment or international partners, to identify and align with their respective roles.⁴ By recognising the diverse interests of all stakeholders and aligning the appropriate incentives, this model seeks to drive sustainable AI adoption while ensuring human-centred progress. A well-executed implementation could establish it as a global benchmark for using AI to advance both equitable healthcare outcomes and technological leadership.

The Current Landscape

AI in healthcare is increasingly employing deep learning and advanced machine learning (ML) algorithms to support clinical decision-making and encourage individualised treatment, enabled by the rise of low-cost or free, and open-source tools. AI has evolved from rule-based systems to powerful algorithms capable of diagnosing diseases, predicting risks, and planning treatments.⁵ This progress is driven by digitised health records, increased computational capacity, and large datasets. Telemedicine, hospital operational management, remote patient monitoring, and advanced diagnostics are among the few of the AI-enabled applications that have shown potential in reducing healthcare disparities between rural and urban areas.^{6,7} Public initiatives like the Ayushman Bharat Digital Mission (ABDM) have enabled streamlined health data access and interoperability, laying the groundwork for equitable AI adoption. Private enterprises, in turn, build on this infrastructure to scale innovative solutions. This public-private synergy ensures that AI-driven healthcare reaches underserved populations while promoting accessibility, innovation, and market vitality.

Key Drivers of AI Adoption in Healthcare: A Socio-Ecological Perspective

AI adoption in healthcare is not just a technological shift but part of a dynamic ecosystem shaped by policy, market forces, and social equity. The adoption of AI in healthcare can be understood through the Socio-Ecological Model (SEM), which examines multiple layers of influence: from individual users to broader policy frameworks.⁸ This perspective highlights how government initiatives, private-sector competition, and community needs intersect to drive AI's transformative impact across diverse healthcare settings, organisational, community, and policy-level factors.

1. Individual

The willingness of healthcare professionals to adopt AI is influenced by digital literacy, trust in AI, and perceived clinical utility.⁹ Adoption improves with targeted training, regulatory clarity, and the seamless integration of AI into clinical workflows. AI-powered tools like symptom

checkers and chronic disease management apps support patient self-care,¹⁰ but adoption depends on usability, transparency, and trust in AI-generated outputs. Patients' acceptance of AI services, such as telemedicine or diagnostics, is shaped by their familiarity with technology, concerns about data privacy, and cultural attitudes. Transparent risk communication, ethical governance, and culturally-sensitive design are vital for broader acceptance.^{11,12}

2. Interpersonal

AI enhances provider–patient communication and diagnostic accuracy, but fears of depersonalised care persist.¹³ Trust improves when AI tools complement and not replace clinical judgment, enabling physicians to focus on empathy and shared decision-making.¹⁴ Training clinicians in AI literacy through medical education and continuous professional development is key to ensuring the confident and effective use of AI in practice.¹⁵

3. Organisational

Hospitals and healthcare institutions are at the forefront of AI adoption, with investments in AI-powered infrastructure aimed at optimising clinical workflows, improving patient outcomes, and reducing operational inefficiencies. AI-driven hospital management systems enhance resource allocation, reducing patient wait times and optimising healthcare delivery. When aligned with environmental sustainability policies, they can further improve operational efficiency and reduce the facility's ecological footprint.¹⁶ In emergency settings, AI supports triage by prioritising cases based on severity, enhancing efficiency and accuracy.¹⁷

4. Community and Societal

Public perceptions, shaped by media portrayals, can influence trust in AI. Misconceptions must be addressed through targeted awareness campaigns that clarify AI's capabilities and limitations.^{18,19} At the community level, AI-powered public health initiatives support disease surveillance and outbreak preparedness, particularly in underserved areas. These systems depend on real-time analytics, reliable data collection, and coordination between public health bodies, civil society,

academia, and industry.²⁰ AI-enabled mobile health and telemedicine tools are instrumental in bridging rural healthcare gaps by providing access to diagnostics and consultations in low-resource settings.²¹

5. Policy and Regulatory

Government-led initiatives such as the ABDM have laid the foundation for digital infrastructure that will support AI integration and interoperability.²² Successful AI deployment depends on interoperability with Electronic Health Records (EHRs), necessitating standardised data governance and modernisation of IT infrastructure to support seamless integration.²³ Regulatory frameworks like the Digital Personal Data Protection Act (DPDPA) contribute to data privacy protection and support transparency in AI decision-making. However, they do not directly address informed consent or algorithmic bias.^{24,25} Policies must balance innovation with ethical safeguards. Additionally, public-private partnerships (PPPs) have been a catalyst for innovation, particularly in scaling up AI-based solutions.²⁶ For example, the IndiaAI Mission is a partnership between the public and private sectors to improve AI infrastructure and the generation of India-specific AI Models. Similarly, during COVID-19, the Health Equity Consortium (HEC) used cross-sector cooperation to combat health misinformation, enhance service accessibility, and illustrate how PPPs can advance data-driven and equitable healthcare.^{27,28}

Analysis and Challenges

The integration of AI into healthcare presents transformative opportunities, yet it also poses significant ethical, technical, and infrastructural challenges that must be addressed to ensure equitable and sustainable adoption.

1. Ethical and Social Implications

AI systems in healthcare are susceptible to algorithmic bias, particularly when trained on non-representative datasets. This may exacerbate health disparities for marginalised populations. For instance, healthcare expenses were used as an analogue for health status, i.e., the algorithm assigned Black patients with lower risk scores than equally

sick White patients. This routinely underestimated the health needs of Black patients, leading to unequal access to care.²⁹ Moreover, the opacity of AI decision-making raises concerns regarding accountability and patient safety. The Indian Council of Medical Research (ICMR) issued Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare.^{30,31} Consistent implementation of the guideline in real-world settings is limited by its descriptive rather than prescriptive nature. Recent editorials highlight that the lack of oversight permits biases, giving rise to concerns about stigmatisation, discrimination, and the reliability of digital health technologies. Public mistrust, fuelled by media sensationalism and fears of AI replacing human expertise, highlights the need for explainable and actionable AI (XAI)^a and stakeholder-inclusive development processes.^{32,33}

2. Data Privacy and Security

The reliance of AI systems on large volumes of sensitive patient data introduces vulnerabilities like breaches, unauthorised access, and misuse. Although the DPDPA establishes a legal foundation, implementation challenges remain, particularly in health data governance. Emerging privacy-preserving technologies, such as homomorphic encryption and federated learning, offer promising solutions but are yet to be widely adopted.^{34,35} Clear consent mechanisms and improved public awareness of data rights are essential for maintaining trust in AI-enabled health services. Furthermore, concerns about privacy and sovereignty in cross-border health data transfers also emphasise the necessity of robust legal protections and balanced data localisation. Interoperable standards, international collaboration, and rights-based regulation are some of the approaches that can facilitate the safe and fair exchange of global health data.³⁶

a The term “explainable artificial intelligence” (XAI) describes methods and strategies that improve trust, accountability, and usability in high-stakes industries, like healthcare, by making the results and decision-making processes of AI systems transparent, interpretable, and understandable.

3. Infrastructure and Digital Divide

Though the application of AI in fields like telemedicine, remote monitoring, and diagnostics has demonstrated promise in reducing health disparities between rural and urban areas, infrastructure gaps in many underserved areas continue to limit its wider deployment. Limited internet access, poor digital record systems, and insufficient computational resources hinder AI implementation at scale. The absence of structured capacity-building programmes also impedes AI adoption among healthcare professionals.³⁷ Government-led initiatives such as DPI offer a foundation, yet there is a pressing need to link these programmes explicitly with AI-readiness in low-resource settings.

4. Interoperability and Integration Challenges

Many AI applications in healthcare face difficulties integrating with existing EHR systems, often due to legacy IT infrastructure and a lack of data standardisation. The adoption of interoperability frameworks such as Fast Healthcare Interoperability Resources (FHIR)^b remains limited.³⁸ Additionally, many healthcare institutions operate on legacy IT systems that are incompatible with modern AI applications.³⁹ Without concerted policy incentives and technical harmonisation, the seamless integration of AI into routine healthcare delivery will remain aspirational.⁴⁰

Recommendations

A multi-level, stakeholder-driven approach rooted in SEM is essential to foster ethical, inclusive AI in healthcare. This includes enforcing India's ethical AI guidelines, mandating XAI and actionable AI, and involving clinicians and patients in co-development. Robust implementation of the DPDPA, adoption of privacy-preserving technologies, and public education programmes on data rights are vital. Expanding digital infrastructure, promoting low-resource AI tools, and training healthcare workers will bridge the access, skill, and scale gaps. Lastly, standardising interoperability frameworks like FHIR, incentivising digital

b Fast Healthcare Interoperability Resources (FHIR) is built on HL7, i.e., a standard-developing tool to enhance the exchange of information between medical systems. FHIR seeks to advance semantic interoperability using modern, flexible data exchange technological advances.

upgrades, and strengthening PPPs will support scalable, AI-ready health systems across diverse care settings (Table 1).

Table 1: Stakeholder-Informed SEM Framework for AI in Healthcare in India

Challenges	Key Issues	Primary Stakeholders*	Strategic Recommendations
Ethical & Social Implications	Algorithmic bias, lack of transparency, public mistrust	Patients, clinicians, ethicists, regulators, technologists, IT professionals, developers	Enforce ethical AI guidelines; mandate XAI and actionable AI; involve patients and providers in participatory co-design.
Data Privacy & Security	Data misuse, weak consent mechanisms, limited Privacy-Enhancing Technology adoption ^c	Patients, health IT professionals, technologists, developers, legal experts, policymakers	Effectively implement the DPDPA; incentivise the adoption of PETs and federated learning; enhance public awareness of data rights through active engagement.
Infrastructure & Digital Divide	Limited access in rural areas, low AI literacy among professionals	Policymakers, rural communities, healthcare professionals, NGOs, local governments, telecom and tech providers, educational institutions	Expand DPI coverage, develop low-resource AI tools, integrate AI training into medical and public health curricula.
Interoperability & Integration	Legacy systems, poor EHR connectivity, a lack of data standards	Hospitals, IT vendors, digital health platforms, government agencies, standard setting, and international organisations	Institutionalise FHIR standards; offer incentives for digital upgrades; promote modular AI solutions via PPP.

*This framework applies the SEM to AI in healthcare by addressing challenges across multiple levels: (i) Individual and interpersonal- ethical concerns, patient trust, and clinician engagement; (ii) Community and Organisational- infrastructure gaps, AI literacy, and interoperability within healthcare systems; Policy: regulatory measures, data privacy frameworks, and national-level AI governance.

c Prevents outcomes that safeguard data privacy and promote data-driven innovation from being widely adopted.

As AI reshapes healthcare, ongoing policy innovation, capacity-building, and ethical oversight are essential to balance progress with public trust and access. With its talent pool, diverse healthcare needs, and digital transformation efforts, India is well-positioned to lead AI-driven healthcare for both local and global populations.

Government-led digitalisation is accelerating adoption, but evolving policy frameworks must ensure that AI delivers equitable, inclusive, and socially beneficial outcomes. Coordinated efforts are required to transform successful pilot projects into scalable, interoperable, and context-specific AI solutions across healthcare settings to fully utilise its potential.

Viola Savy Dsouza is an Assistant Professor at the NIMS Institute of Public Health and Governance, NIMS University, Jaipur and a PhD Candidate, Faculty of Health, Medicine, and Life Sciences (FHML), Maastricht University, Maastricht, the Netherlands.

Angela Brand is Professor, Faculty of Health, Medicine, and Life Sciences (FHML), Maastricht University, Maastricht, the Netherlands; and Professorial Fellow, United Nations University - Maastricht Economic and Social Research Institute on Innovation and Technology, the Netherlands.

Endnotes

- 1 Monika Nair et al., "A Comprehensive Overview of Barriers and Strategies for AI Implementation in Healthcare: Mixed-Method Design," *PLOS ONE* 19, no. 8 (2024): e0305949, <https://doi.org/10.1371/journal.pone.0305949>
- 2 Gerard Marshall Raj et al., "Inception of the Indian Digital Health Mission: Connecting the Dots," *Health Care Science* 2, no. 5 (2023): 345–351, <https://doi.org/10.1002/hcs2.67>
- 3 Ministry of Health and Family Welfare, Government of India, <https://pib.gov.in/PressReleaselframePage.aspx?PRID=2083189>
- 4 Hadi Naghavi Pour et al., "Pragmatic and Symbiotic Quintuple Helix Model Mitigating Emerging Technologies Disruption: A Vision, Strategy, and Policy," *TechRxiv* (2024), <https://doi.org/10.36227/techrxiv.170723152.28042295/v1>
- 5 Shuroug A. Alowais et al., "Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice," *BMC Medical Education* 23, September 22, 2023, <https://doi.org/10.1186/s12909-023-04698-z>
- 6 Kinalyne Perez et al., "Investigation into Application of AI and Telemedicine in Rural Communities: A Systematic Literature Review," *Healthcare* 13, no. 3 (2025): 324, <https://doi.org/10.3390/healthcare13030324>
- 7 Ajit Kerketta et al., "Leveraging AI Tools to Bridge the Healthcare Gap in Rural Areas in India," *medRxiv*, 2024, <https://doi.org/10.1101/2024.07.30.24311228>
- 8 EBSCO, "Social Ecological Model," *Research Starters: Environmental Sciences*, 2025, <https://www.ebsco.com/research-starters/environmental-sciences/social-ecological-model>
- 9 Mingyang Chen et al., "Acceptance of Clinical Artificial Intelligence among Physicians and Medical Students: A Systematic Review with Cross-Sectional Survey," *Frontiers in Medicine* 9, August 31, 2022, <https://doi.org/10.3389/fmed.2022.990604>
- 10 Yue You et al., "Self-Diagnosis through AI-Enabled Chatbot-Based Symptom Checkers: User Experiences and Design Considerations," *AMIA Annual Symposium Proceedings* (2021): 1354–1363.
- 11 Sally Moy et al., "Patient Perspectives on the Use of Artificial Intelligence in Health Care: A Scoping Review," *Journal of Patient-Centered Research and Reviews* 11, no. 1 (2024): 51–62, <https://doi.org/10.17294/2330-0698.2029>
- 12 Jingquan Li, "Security Implications of AI Chatbots in Health Care," *Journal of Medical Internet Research* 25, no. 1 (2023): e47551, <https://doi.org/10.2196/47551>
- 13 Junaid Bajwa et al., "Artificial Intelligence in Healthcare: Transforming the Practice of Medicine," *Future Healthcare Journal* 8, no. 2 (2021): 188–194, <https://doi.org/10.7861/fhj.2021-0095>
- 14 Aurelia Sauerbrei et al., "The Impact of Artificial Intelligence on the Person-Centred, Doctor-Patient Relationship: Some Problems and Solutions," *BMC Medical Informatics and Decision Making* 23, no. 1 (2023), <https://doi.org/10.1186/s12911-023-02162-y>

- 15 Mohammad Muzaffar Mir et al., "Application of Artificial Intelligence in Medical Education: Current Scenario and Future Perspectives," *Journal of Advances in Medical Education & Professionalism* 11, no. 3 (2023): 133–140, <https://doi.org/10.30476/jamp.2023.98655.1803>
- 16 Viola Savy Dsouza et al., "Building Perspectives for Resilient Health System: Lessons Learned from the Experience of Patients and Health Professionals during COVID-19," *Journal of Health Management* 26, no. 2 (2024): 203–213, <https://doi.org/10.1177/09720634241229267>
- 17 Samantha Tyler et al., "Use of Artificial Intelligence in Triage in Hospital Emergency Departments: A Scoping Review," *Cureus* 16, no. 5 (2024), <https://doi.org/10.7759/cureus.59906>
- 18 Karim Nader et al., "Public Understanding of Artificial Intelligence through Entertainment Media," *AI & Society* 39 (2024): 713–726, <https://doi.org/10.1007/s00146-022-01427-w>
- 19 Krzysztof Witkowski et al., "Public Perceptions of Artificial Intelligence in Healthcare: Ethical Concerns and Opportunities for Patient-Centered Care," *BMC Medical Ethics* 25 (2024): 74, <https://doi.org/10.1186/s12910-024-01066-4>
- 20 Ajit Kerketta et al., "Leveraging AI Tools to Bridge the Healthcare Gap in Rural Areas in India," *medRxiv*, 2024, <https://doi.org/10.1101/2024.07.30.24311228>
- 21 Kinalyne et al., "Investigation into Application of AI and Telemedicine in Rural Communities: A Systematic Literature Review," *Healthcare* 13, no. 3 (2025): 324, <https://doi.org/10.3390/healthcare13030324>
- 22 Ministry of Health & Family Welfare, Government of India, <https://www.mohfw.gov.in/?q=pressrelease-87>
- 23 Mohsen Khosravi et al., "Artificial Intelligence and Decision-Making in Healthcare: A Thematic Analysis of a Systematic Review of Reviews," *Health Services Research and Managerial Epidemiology* 11 (2024): 23333928241234863, <https://doi.org/10.1177/23333928241234863>
- 24 Sauradeep Bag, "Digital Personal Data Protection Act: Shaping India's AI-Driven Fintech Sector," Observer Research Foundation, 2024, <https://www.orfonline.org/expert-speak/digital-personal-data-protection-act-shaping-india-s-ai-driven-fintech-sector>
- 25 Rakesh, "Why We Need Data Protection Laws for AI in India," *De Facto Law Journal* 1, no. 1 (2025), <https://defactolawjournal.org/papers/why-we-need-data-protection-laws-for-ai-in-india/>
- 26 Samira Abdul et al., "Public-Private Partnerships in Health Sector Innovation: Lessons from around the World," *Magna Scientia Advanced Biology and Pharmacy* 12, no. 1 (2024): 45–59, <https://doi.org/10.30574/msabp.2024.12.1.0032>
- 27 Angel Arnaout et al., "Leveraging Technology in Public-Private Partnerships: A Model to Address Public Health Inequities," *Frontiers in Health Services* 3 (2023), <https://doi.org/10.3389/frhs.2023.1187306>
- 28 Shankar Krishnamurthy et al., "India's AI Ambitions: Can Public-Private Partnerships Lead the Way?," *S&P Global* (2024), <https://www.spglobal.com/en/research-insights/special-reports/india-forward/indias-ai-ambitions-can-public-private-partnerships-lead-the-way>
- 29 Zaid Obermeyer et al., "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science* 366, no. 6464 (2019): 447–453, <https://doi.org/10.1126/science.aax2342>

- 30 DHR-ICMR Artificial Intelligence Cell, "Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare," 2024, <https://www.icmr.gov.in/ethical-guidelines-for-application-of-artificial-intelligence-in-biomedical-research-and-healthcare>
- 31 Madhavi Bhargava et al., "Artificial Intelligence in Biomedical Research and Publications: It Is Not about Good or Evil but about Its Ethical Use," *Indian Journal of Community Medicine* 49, no. 6 (2024): 777–779, https://doi.org/10.4103/ijcm.ijcm_560_24
- 32 Soonyi Choi et al., "Unveiling Data Monopolies: How Information Control Shapes Finance, AI, and Healthcare," *Medium*, October 8, 2024, <https://medium.com/ai-ask/unveiling-data-monopolies-how-information-control-shapes-finance-ai-and-healthcare-2024-10-08-592640121f9b>
- 33 Viola Savy Dsouza et al., "Responsible Artificial Intelligence (AI) Governance Using a Relational Governance Framework," Observer Research Foundation, February 9, 2023, <https://www.orfonline.org/expert-speak/responsible-ai-governance-using-a-relational-governance-framework>
- 34 Hyunghoon Cho et al., "Privacy-Enhancing Technologies in Biomedical Data Science," *Annual Review of Biomedical Data Science* 7 (2024): 317–343, <https://doi.org/10.1146/annurev-biodatasci-120423-120107>
- 35 Ming Li et al., "From Challenges and Pitfalls to Recommendations and Opportunities: Implementing Federated Learning in Healthcare," *Medical Image Analysis* 101 (2025): 103497–97, <https://doi.org/10.1016/j.media.2025.103497>
- 36 Linhua Xia et al., "Paradigm Transformation of Global Health Data Regulation: Challenges in Governance and Human Rights Protection of Cross-Border Data Flows," *Risk Management and Healthcare Policy* 17 (2024): 3291–3304, <https://doi.org/10.2147/RMHP.S450082>
- 37 Ejike Innocent Nwankwo et al., "Integrating Telemedicine and AI to Improve Healthcare Access in Rural Settings," *International Journal of Life Science Research Archive* 7, no. 1 (2024): 59–77, <https://doi.org/10.53771/ijlsra.2024.7.1.0061>
- 38 Parinaz Tabari et al., "State-of-the-Art Fast Healthcare Interoperability Resources (FHIR)–Based Data Model and Structure Implementations: Systematic Scoping Review," *JMIR Medical Informatics* 12 (2024): e58445, <https://medinform.jmir.org/2024/1/e58445>
- 39 Shefali V. Bhagat et al., "Navigating the Future: The Transformative Impact of Artificial Intelligence on Hospital Management—A Comprehensive Review," *Cureus* 16, no. 2 (2024): e54518, <https://doi.org/10.7759/cureus.54518>
- 40 Jennifer Gaudet et al., "AI Can Bridge the Health Data Interoperability Gap," Booz Allen Hamilton, 2024, <https://www.boozallen.com/insights/data-optimization/ai-can-bridge-the-health-data-interoperability-gap.html>



III.

Responsible Governance of Open AI Models

The Licensing Landscape for Responsible, Open-Source AI

Mohit Chawdhry

Open-source Artificial Intelligence (AI) is increasingly seen as a catalyst for equitable technological development, especially in the Global South. By openly sharing models and tools, nations with limited resources can access and adapt cutting-edge AI for local needs in critical sectors like healthcare, agriculture, and education.¹ This collaborative model of innovation can help bridge the divide between the Global North and Global South—at present, most advanced AI models (like GPT-4) are developed by Western tech firms, and the benefits are unevenly concentrated. Notably, 59 percent of developers in emerging economies are of the view that open-source

software will drive their nation's economic growth in the next decade.² India itself has saved an estimated US\$2 billion per year by switching from proprietary software to open-source alternatives, and has become the second-largest source of contributors to open-source repositories on GitHub.³

As the development and deployment of open-source AI models accelerates, the question of how to govern their use has become increasingly urgent. Unlike traditional software, AI systems pose a unique set of downstream risks, like misinformation and surveillance, that challenge existing legal and ethical frameworks.⁴ At the centre of this debate is the role of licensing: the primary tool by which developers grant others the right to use, modify, and share their models.

This article examines current AI licensing practices from different parts of the globe, comparing legacy open-source software licences like MIT and Apache 2.0 with newer, AI-specific licences such as Meta's Llama Community License, Black Forest's FLUX, and the Responsible AI License (RAIL) framework. It analyses how these licences differ in their approach to openness, commercial use, harmful applications, and liability. It also considers the enforcement and accountability challenges that arise even when responsible-use clauses are included.

To address these gaps, the article explores a layered governance approach that builds on licensing but extends into regulation, provenance technologies,^a and community oversight. It highlights emerging best practices, including risk-tiered liability frameworks, blockchain-based provenance tracking, watermarking systems, and platform-driven moderation. Together, these interventions offer a more robust foundation for ensuring that open-source AI development remains both innovative and responsible.

a The term 'provenance technologies' refers to tools and systems designed to track the origin, ownership, and transformation history of digital assets—such as AI models or datasets—over time. These technologies can include watermarking, metadata tagging, and blockchain-based registries that log cryptographic hashes of model weights, training data, and licence terms.

Current Licensing Practices

In the context of AI models, ‘licensing’ refers to the legal terms under which developers release model weights, code, and associated assets. These licences define what others can, and cannot do with the model, including whether it can be used commercially, modified, redistributed, or integrated into other systems. As AI capabilities grow, licensing has become crucial to balancing openness, innovation, and responsible use.

Today’s AI licensing landscape spans a wide range—from permissive legacy software licences like MIT and Apache 2.0 to newly developed, AI-specific restrictive frameworks such as Meta’s Llama Community License and Black Forest’s FLUX. It also includes novel, responsible AI licences, such as the RAIL framework. These licences differ in how they address openness, commercial use, harmful applications, and liability. This section analyses these key dimensions.

Table 1. Various Open-Source Licences

Licence Name	MIT	Apache 2.0	LLaMa 3.1 (Meta)	FluxDev (Blackforest Labs)	Creative-ML Open RAIL M (Stable Diffusion)
Licence Type	Legacy OSS	Legacy OSS	AI-specific (Community Licencs)	AI-specific (non-commercial)	AI-specific, open-source
Number of Models Using the Licence (per HuggingFace)	55,724	133,920	1,207	21,506	27,408
Mandatory Attribution Notice in Derivatv Source Code	Ü	ü	Ü	Ü	Ü

Commercial Use Restrictions	Û	û	ü LLaMa cannot be used to train or improve other models	Û No commercial use allowed	ü Commercial use allowed (the model weights are open) but is subject to accepting licence terms, including ethical use
Misuse/Harmful Use Case Prohibitions	Û	û	ü	Û	Û
Limitation of Liability and Warranty Clause	Û	ü	ü	ü	Û

Source: Author's own, from open sources.

Openness and Commercial Use

Legacy licences, such as MIT and Apache 2.0, are known for their openness: they allow unrestricted reuse and commercial deployment. AI models released under these terms—such as EleutherAI’s GPT-J vis-à-vis Apache 2.0—can be integrated into proprietary products without payment or prior approval.⁵ This has helped democratise access to advanced AI systems. However, this openness creates a governance vacuum as these licenses lack mechanisms to restrict harmful applications like the creation of deepfakes or automated surveillance.⁶

In contrast, the more modern AI-specific licences introduce constraints on commercial utilisation. Meta’s Llama 2, for example, is released under a community license that restricts commercial use by organisations with more than 700 million monthly users. It also prohibits using Llama-generated outputs to train other models—a safeguard against producing synthetic data to develop competing systems.⁷ Similarly, the FLUX.1-dev licence from Black Forest Labs bans all commercial use unless a separate agreement is secured.⁸

Responsible AI licences (RAIL) seek to achieve a middle ground between legacy OSS licences and modern, restrictive AI licences by

combining a permissive base license and a separate ‘Responsible Use’ clause that outlines prohibited uses. This format was introduced by the RAIL Initiative in 2019 to encourage the responsible release of AI models. If the licence allows free redistribution and access while including such ethical restrictions, it is referred to as an OpenRAIL licence.⁹ They have been used in several high-profile AI model releases. The BigScience BLOOM multilingual language model was released under an OpenRAIL licence in 2022.¹⁰ Stable Diffusion, a widely used generative image model, uses the “CreativeML Open RAIL-M” licence, which combines broad accessibility with specific prohibitions, such as against extremist content and illicit uses.¹¹ Similarly, StarCoder, an open code-generation model, was released with a RAIL licence that bans uses, such as generating malware or exploiting minors.¹²

Guardrails Vs. Harmful Use

A critique of legacy open-source licences is that they do not prevent misuse. Originally designed for software tools like text editors or databases, these licences assume that the user will comply with existing laws and norms. However, AI models—especially generative ones—pose new risks, such as spreading disinformation or enabling biometric surveillance.¹³

Conversely, AI-specific licences seek to govern downstream use by prescribing acceptable and unacceptable use cases. For instance, Meta’s Llama 2 licence includes an Acceptable Use Policy that forbids harassment, misinformation, and unauthorised surveillance. Violations can result in loss of licence rights.¹⁴ The OpenRAIL family of licences prohibits outputs that defame individuals, incite hate, exploit minors, or promote misinformation. Licensees must also make reasonable efforts to filter such content.¹⁵ Black Forest’s FLUX licence bans use in legally prohibited domains and restricts export to sanctioned countries.¹⁶

However, enforcing these clauses remains a challenge. A determined actor can ignore the licence, especially if they remain anonymous. For instance, researchers have bypassed Llama 2’s safeguards to create a model dubbed ‘BadLlama,’ which generates toxic content, such as extremist propaganda and detailed cybercrime tutorials.¹⁷ While this violates the licence, accountability is difficult without identifiable actors.

Table 2 lists the instances where open-source AI models have been used for illicit purposes by downstream users.

Table 2. Malicious Use-Cases of Open-Source Models

Model/Dataset	Licence/Availability	Documented Illicit Use (downstream)	Year First Reported
'BadLlama 3' – community fork of Llama 3 8B ¹⁸	Redistributed under the original Llama community licence after the safety layers were removed	A quick fine tune stripped of guard rails, enabling unrestricted extremist propaganda and detailed cybercrime tutorials that can be deployed on any server	2024
Open source voice cloning toolkits (e.g., VAL E, Tortoise TTS) ¹⁹	Research checkpoints and code are freely downloadable from GitHub	Fraudsters clone relatives' or executives' voices for 'grandparent' and CEO impersonation scams, convincing victims to transfer funds	2023–24
Stable Diffusion and derivative image generators ²⁰	Weights publicly hosted; CreativeML Open RAIL M	Fine tuned models generate political deep fake ads and memes impersonating candidates, spreading election disinformation during the 2024 campaigns	2024
GPT Neo / GPT J language model weights	Apache 2.0, openly mirrored	Cyber criminal forums automate spear phishing emails and large scale spam templates that evade traditional filters	2023–24

Source: Author's own.

Liability and Accountability

Most open-source and proprietary software licences include disclaimers that shield developers from liability. MIT and Apache licences specify that software is provided “as is,” with no warranties. If an AI model produces harmful results, such as a flawed legal or medical recommendation, its creators bear no legal responsibility. AI-specific licences also follow this tradition. The Llama 2 license includes a liability waiver for Meta. FLUX.1-dev includes similar language shielding Black Forest Labs. This presents a governance challenge: model creators avoid liability, while users may also evade responsibility if they are difficult to trace.²²

The RAIL family of licences seeks to address accountability issues indirectly by banning certain uses. Thus, they create a contractual basis to hold users accountable on the violation of those terms. In theory, if a user employed an open model for disallowed purposes (e.g., generating extremist content that leads to harm), the model creator could take legal action for breach of contract. However, this requires identifying the violator and proving the breach. Moreover, monitoring compliance with licence restrictions is a resource-intensive and imperfect approach. For popular models downloaded by tens of thousands of people, it is nearly impossible to track all uses.²³

In sum, AI licensing today presents a difficult trade-off between innovation and accountability. Traditional open-source licences promote broad access and redistribution, enabling rapid innovation but failing to address the unique risks posed by advanced AI systems. In contrast, newer AI-specific licences attempt to curb harmful uses and protect commercial interests. The flipside remains that, in doing so, they may restrict who can build on foundational models—potentially centralising control in the hands of a few large players. Meanwhile, both types of licences protect original developers from liability. While this encourages experimentation and openness, it also creates a gap in accountability that could leave the public vulnerable and without recourse. Lastly, the RAIL framework marks a step in the right direction, allowing openness while seeking to prevent downstream misuse. However, monitoring licence violations and enforcing safeguards is onerous, especially where downstream users are not easily identified.

Emerging Best Practices

While open-source licences play a foundational role in distributing and governing AI models, they alone are not sufficient in preventing downstream misuse or mitigate harm. A more effective approach builds on licensing with complementary layers—regulatory frameworks, technical tracing tools, and active community oversight. Together, these mechanisms create a distributed system of accountability that is better suited to open-source AI’s scale, complexity, and evolving risks.

Risk-tiered Liability Frameworks

Policymakers increasingly agree that liability should reflect the context and scale of deployment—not simply the act of open publication. The European Union’s 2024 political agreement on the AI Act adopts this principle by distinguishing between open-weight general-purpose AI models and their downstream applications.²⁴ Developers who release model weights, documentation, and licence terms to the public must meet two core requirements: (1) publish a summary of the training data and (2) conduct a basic risk evaluation. If these steps are followed, such releases are exempt from the more stringent “high-risk system” rules. Liability and formal conformity assessments apply only when another party integrates the model into a high-risk product, such as a biometric ID system.²⁵ This approach helps safeguard open research while ensuring that commercial actors cannot externalise harm.

The United States (US) is moving in this direction. In March 2024, the Commerce Department’s National Telecommunications and Information Administration recommended a “qualified safe harbor” in its AI accountability report.²⁶ Developers who release models with licence terms banning unlawful use, disclose training data provenance, and publish red-teaming results would receive reduced litigation exposure in the event of a bad actor weaponising the model. In contrast, companies that integrate or commercialise such models in safety-critical settings would be required to carry risk insurance and conduct post-deployment monitoring, creating a clear incentive to manage downstream risks where they are most likely to occur.

International bodies have echoed this approach. A recent Global Partnership on AI discussion paper proposes a three-tiered liability framework—research, commercial, and critical—with each tier subject to increasing levels of documentation, auditing, and financial accountability.²⁷

Development and Implementation of Provenance and Tracking Technologies

Preventing misuse also depends on the ability to trace models and outputs after the weights leave the origin server. Image-based watermarking systems now embed imperceptible signatures at the point of generation. Stable Diffusion, for instance, includes such a watermark by default in its official release.²⁸ For text outputs, watermarking is being standardised by groups like Google DeepMind, which open-sourced its approach in October 2024. These techniques aim to mark AI-generated content without altering its meaning.²⁹ However, watermark robustness is limited; a 2024 paper shows that targeted paraphrasing or re-generation can reduce detection rates below 50 percent.³⁰ As a result, watermarks should be seen as probabilistic tools—useful for aggregate attribution but insufficient as a standalone safeguard.

To strengthen provenance, several research groups are developing blockchain-backed registries that track every derivative model checkpoint. Each entry logs a cryptographic hash of the parent and child weights, licensing terms, and the contributor's signature. Because these records are immutable and time-stamped, they allow investigators to identify when and by whom a misuse clause was removed, pinpointing accountability on the forker and not the original author. For instance, OpenAI and major media organisations participate in the Coalition for Content Provenance and Authenticity (C2PA), whose specification embeds signed cryptographic 'manifests' in each asset header. Although C2PA relies on standard public-key infrastructure and not a distributed ledger, the signatures render any post-release alteration of weights or generated outputs detectable, because tampering invalidates the chain of trust recorded in the manifest.³¹

Community-Driven Moderation and Tracking

The final layer of accountability is community-driven moderation, which provides the agility and responsiveness that statutory regulation often lacks. Platforms like Hugging Face, for example, allow users to flag repositories whose models or demos violate the attached license terms. In response, moderators can quarantine the project, add warning banners, or revoke access keys until the issue is resolved. The platform's policy explicitly cites "models facilitating extremist or non-consensual sexual content" as grounds for removal, closely aligned with the misuse clauses found in Responsible AI Licenses (RAIL).³² GitHub offers a parallel mechanism: its Acceptable Use Policy allows suspension of hosting for code or model weights that promote human trafficking or sexual exploitation.³³ Given that these platforms host much of the open-source AI ecosystem, community flagging, paired with platform enforcement, creates a fast and decentralised feedback loop for mitigating harm.

Public repositories have also emerged to track and document misuse. The AI Incident Database compiles verified cases in which AI systems caused or contributed to harm, including details such as model family, deployment context, and license type (when available).³⁴ The Partnership on AI's 2024 report outlines a pilot project: the Generative AI Misuse Observatory, which aggregates incident reports from platforms and security researchers into a searchable public feed.³⁵ These resources offer regulators and infrastructure providers early warning signals, allowing them to cross-reference new abuse cases with known model hashes, and suspend malicious deployments more efficiently.

Conclusion

Licensing remains the indispensable starting point for governing open-source AI because it establishes the legal boundaries for access, modification, and redistribution. By embedding responsible-use clauses, a licence can, in principle, deter malicious applications and set expectations of ethical conduct. Yet, licensing alone cannot shoulder the burden of responsible deployment for several reasons. First, misuse: once weights are downloaded, a bad actor can ignore contractual terms, fine-tune the model, and deploy it in ways the licence forbids.

Second, liability: most licences—legacy and AI-specific alike—contain broad warranty disclaimers, leaving the question of who compensates victims when an openly released system causes harm, unclear.

A more resilient governance architecture layers additional safeguards around the licence. At the legal level, risk-tiered regulation shifts the heaviest obligations to commercial or high-impact deployers rather than upstream researchers. On the technical front, provenance mechanisms, such as robust watermarking, cryptographically signed manifests, or tamper-evident ledgers, create auditable trails that make violations detectable. Lastly, platform and community oversight provide rapid, decentralised moderation through reporting portals, incident databases, and reputation systems. Together, they preserve the openness that fosters innovation while embedding the accountability necessary for safe and equitable AI deployment.

Mohit Chawdhry is a technology policy professional.

Endnotes

- 1 Kellee Wicker, “The Rise of AI in the Global South and the Need for Inclusion,” *Wilson Center*, September 3, 2024, <https://www.wilsoncenter.org/blog-post/rise-ai-global-south-and-need-inclusion#:~:text=facilitated%20biomedical%20research>.
- 2 “State of Open Source in the Global South Report,” *Eclipse Foundation*, 2025, <https://outreach.eclipse.foundation/open-source-global-south-developers#:~:text=Image%3A%20economic%20>.
- 3 Leslie D’Monte, “Open Source Software Can Save India \$2 Bn,” *Business Standard*, September 11, 2009, https://www.business-standard.com/article/technology/open-source-software-can-save-india-2-bn-109091200017_1.html.
- 4 David Evan Harris, “How to Regulate Unsecured ‘Open-Source’ AI: No Exemptions,” *Tech Policy Press*, December 4, 2023, <https://www.techpolicy.press/how-to-regulate-unsecured-opensource-ai-no-exemptions/>.
- 5 EleutherAI, “Gpt-J-6B,” Hugging Face, May 3, 2023, <https://huggingface.co/EleutherAI/gpt-j-6b>.
- 6 David Gray Widder et al., “Limits and Possibilities for ‘Ethical AI’ in Open Source: A Study of Deepfakes,” (*paper presented at the ACM Conference on Fairness, Accountability, and Transparency, Seoul, South Korea, June 20, 2022*), <https://doi.org/10.1145/3531146.3533779>.
- 7 Meta, “Llama 2 Community License Agreement,” *Meta AI*, 2023, <https://ai.meta.com/llama/license/>.
- 8 Black Forest Labs, “Flux1-Dev License,” *GitHub*, 2024, https://github.com/black-forest-labs/flux/blob/main/model_licenses/LICENSE-FLUX1-dev.
- 9 Organisation for Economic Cooperation and Development, “Responsible AI Licenses: A Practical Tool for Implementing the OECD Principles for Trustworthy AI,” *OECD.AI*, 2022, <https://oecd.ai/en/wonk/rails-licenses-trustworthy-ai>.
- 10 BigScience, “BigScience OpenRAIL-M,” *Hugging Face*, 2025, <https://bigscience.huggingface.co/blog/bigscience-openrail-m>.
- 11 CompVis, “Stable Diffusion License,” *Hugging Face*, 2025, <https://huggingface.co/spaces/CompVis/stable-diffusion-license>.
- 12 Bigcode, “StarCoder,” Hugging Face, August 17, 2023, <https://huggingface.co/bigcode/starcoder>.
- 13 Tom Simonite, “The AI Text Generator That’s Too Dangerous to Make Public,” *WIRED*, February 14, 2019, <https://www.wired.com/story/ai-text-generator-too-dangerous-to-make-public/>.
- 14 Meta, “Llama 2 Community License Agreement.”

- 15 Kate Downing, “AI Licensing Can’t Balance ‘Open’ with ‘Responsible,’” *Law Offices of Kate Downing*, July 14, 2023, <https://katedowninglaw.com/2023/07/13/ai-licensing-cant-balance-open-with-responsible/#:~:text=You%20agree%20not%20to%20Use,or%20Modifications%20of%20the%20Model.>
- 16 Black Forest Labs, “Flux1-Dev License.”
- 17 Pranav Gade et al., “BadLlama: Cheaply Removing Safety Fine-Tuning from Llama 2-Chat 13B,” *arXiv.org*, 2023, <https://arxiv.org/abs/2311.00117>.
- 18 Joe Whittaker, “AI and Extremism: Threats and Opportunities,” *VOX-Pol Network of Excellence*, April 24, 2024, <https://voxpath.eu/wp-content/uploads/2024/04/DCUPN0254-Vox-Pol-AI-Extremism-WEB-240424.pdf>.
- 19 Ashley Gold, “AI Voice Cloning Is Behind a Surge in Consumer Scams,” *Axios*, March 15, 2025, <https://www.axios.com/2025/03/15/ai-voice-cloning-consumer-scams>.
- 20 Marina Adami, “How AI-Generated Disinformation Might Impact This Year’s Elections—and How Journalists Should Report on It,” *Reuters Institute*, 2024, <https://reutersinstitute.politics.ox.ac.uk/news/how-ai-generated-disinformation-might-impact-years-elections-and-how-journalists-should-report>.
- 21 Fingerprint, “Large Language Models (LLMs), Fraud, and Malware: A Guide,” *Fingerprint Blog*, 2025, <https://fingerprint.com/blog/large-language-models-llm-fraud-malware-guide>.
- 22 Yaniv Benhamou, “Open Source AI – Definition and Selected Legal Challenges,” *Kluwer Copyright Blog*, April 15, 2024, <https://copyrightblog.kluweriplaw.com/2024/04/15/open-source-ai-definition-and-selected-legal-challenges/>.
- 23 Haiman Wong, “Mapping the Open-Source AI Debate: Cybersecurity Implications and Policy Priorities,” *R Street Institute*, 2024, <https://www.rstreet.org/research/mapping-the-open-source-ai-debate-cybersecurity-implications-and-policy-priorities/>.
- 24 AI Office, “General-Purpose AI Models in the AI Act – Questions & Answers,” *European Commission*, 2024, https://digital-strategy.ec.europa.eu/en/faqs/general-purpose-ai-models-ai-act-questions-answers?utm_source=chatgpt.com.
- 25 AI Office, “General-Purpose AI Models in the AI Act – Questions & Answers.”
- 26 Ellen P. Goodman, *Artificial Intelligence Accountability Policy Report* (Washington DC: National Telecommunications and Information Administration, 2024), <https://www.ntia.gov/sites/default/files/publications/ntia-ai-report-final.pdf>
- 27 Kelle Howson, “Scaling Responsible AI Solutions Challenges and Opportunities,” *Global Partnership on Artificial Intelligence*, 2023, <https://gpai.ai/projects/responsible-ai/RAI05%20-%20Scaling%20Responsible%20AI%20Solutions%20-%20Challenges%20and%20Opportunities.pdf>.
- 28 Jiacheng Liang, “WaterPark: Robust Image Watermarking Using Stable Diffusion,” *arXiv*, 2024, <https://arxiv.org/html/2411.13425>

- 29 Alison Snyder, “Google DeepMind Open Sources Its AI Text Watermarking Tool,” *Axios*, October 24, 2024, <https://www.axios.com/2024/10/24/google-deepmind-ai-text-watermarking-tool>.
- 30 Liang, “WaterPark: Robust Image Watermarking Using Stable Diffusion.”
- 31 Coalition for Content Provenance and Authenticity, *Technical Specification*, 2024, https://c2pa.org/specifications/specifications/1.2/specs/C2PA_Specification.html
- 32 Hugging Face, “Content Policy,” 2025, <https://huggingface.co/content-policy>.
- 33 GitHub, “Acceptable Use Policies,” 2025, <https://docs.github.com/en/site-policy/acceptable-use-policies/github-acceptable-use-policies>
- 34 “Artificial Intelligence Incident Database,” *Incidentdatabase.ai*, 2025, <https://incidentdatabase.ai>
- 35 Partnership on AI, *Annual Report 2024*, February 20, 2025, <https://partnershiponai.org/annual-report-2024/>.

The Global Challenges of Ensuring Ethical Regulations, Governance, and Responsible AI

Tanmay Agrawal

Like the rest of the world, India views Artificial Intelligence (AI) as an opportunity and a tool to address multiple challenges simultaneously. It seeks to leverage AI to respond to economic and social challenges while containing associated risks—an objective that can be pursued more effectively through the democratisation of AI,^a understood as ensuring equality, access, and robust governance.

a AI democratisation refers to the process through which AI technologies—such as data, computing infrastructure, and regulatory frameworks—are made accessible to a diverse set of individuals, organisations, communities, and countries, enabling equitable access, participation, oversight, and benefit-sharing, while advancing governance, fairness, and empowerment.

For AI to be democratised, it is imperative that systems are responsible and explainable and that ethical regulations are complied with and effectively governed so that outcomes remain accountable. This necessitates examining the legal aspects of AI, including questions around its legal personality, the applicable regulatory frameworks, and liability models, alongside international approaches to ethics, bias, privacy, and transparency.

Developers of AI should take necessary precautions to ensure data privacy, transparency, fairness, and accuracy in its use, thereby safeguarding the integrity of legal processes where AI is deployed and protecting the rights and interests of all stakeholders involved. This article identifies the lacunae in the existing legal framework and examines the regulatory challenges associated with applying the institution-based regulatory frameworks to generative AI models. It focuses on the intersection of data privacy and generative AI, especially in relation to data collection and use, the privacy implications of generated content, and mechanisms such as anonymisation and differential privacy.

The Imperative of Democratising AI

Before debating the rights and liabilities of AI in the context of its democratisation, it would be useful to reassess the legal personality of AI, as well as the regulatory frameworks and liability models governing its use. At present, the European Union AI Act, 2024 is the most comprehensive attempt in this direction, enacted to protect citizens' safety and fundamental rights while promoting innovation.¹ The Act classifies AI applications into high-risk, limited-risk, and minimal-risk categories and establishes stringent conditions, including prohibitions for certain high-risk systems. However, it does not offer detailed solutions for addressing ethical concerns such as bias and transparency.²

As AI democratisation “involves extending access to AI technologies beyond specialized technical experts to a broader spectrum of users and organizations,”^{3,4} there is a need to examine international regulations on ethics, bias, privacy, and transparency across jurisdictions. Legal professionals must therefore take necessary precautions to ensure data privacy, transparency, fairness, and accuracy in the use of AI, safeguarding the integrity of legal processes and the rights

and interests of all stakeholders. Other organisations, such as the Organisation for Economic Co-operation and Development (OECD), have developed principles to help ensure that AI is deployed fairly and with transparency.⁵ Similarly, the United Nations Educational, Scientific and Cultural Organization (UNESCO) has issued ethical guidelines for the use of AI.⁶ However, intellectual property issues surrounding AI—including authorship, ownership, and potential infringement—remain at a nascent stage, as foundational questions on legal personality and liability models have yet to be decided.

Another key policy intervention should favour recognising AI with a form of legal personality akin to corporations. If AI is treated as a legal entity, it could be held liable for its own actions under the Absolute Agent theory, which grants AI the authority to take its own decisions and bear its own liability. At the same time, questions arise around data ownership, since data primarily serves as the input on which AI systems are trained. Data scraping—the automated extraction of large volumes of data from online sources such as websites or databases, including text and images—raises particular concerns when such content is protected by copyright or other intellectual property (IP) regulations. Scraping data without permission may infringe the rights of content creators or owners, prompting novel questions on IP licensing and whether compulsory licensing frameworks for AI models are required. If such applications are deployed to advance financial literacy, they could have a transformative impact on financial inclusion.

Objectives and Challenges of Democratising AI

When discussions arise around the idea of “democratising technology”, they often focus on broadening public access—making technologies easier for a wider range of users to adopt and benefit from. In the early 2010s, the “democratisation of 3D printing” referred to printers becoming more affordable, user-friendly, and broadly available.⁷ A similar understanding today shapes debates on the democratisation of AI. Companies like Stability AI have actively promoted this narrative. It describes its flagship product, Stable Diffusion, as a text-to-image generation model designed to empower billions of people to create visually striking art within seconds.⁸ Likewise, Microsoft has articulated a bold vision to democratise AI, emphasising efforts to remove AI

from elite academic and corporate settings and make it universally accessible.⁹ Part of this strategy involves embedding intelligent features into every application, on any device, at any time.

Essentially, the objective of democratising AI is to make the technologies accessible and affordable to larger populations, ensuring portability, explainability, credibility, and fairness, with human oversight. Large-scale distributed AI systems are characterised by self-organising, hierarchical structures that can support the democratisation of AI through collaborative learning across systems, aiming to move beyond traditional federated learning by enhancing generalisation and specialisation. Today, understanding how an AI system reaches its conclusions is no longer merely a technical interest but central to responsible and ethical use.¹⁰

Transparency therefore forms the foundation of trust and fairness, helping ensure that AI technologies serve human interests rather than any malicious agendas or data hallucinations. When decision-makers can understand the reasoning behind an AI's recommendations, they are more likely to rely on them, while also uncovering hidden biases within AI systems. Such explainability enables stakeholders to trace, question, and improve AI-driven decisions, transforming the AI "black box" into a tool that empowers human judgement and social good with explainable AI with human oversight.¹¹

Despite the transformative potential of AI, the development of AI technologies remains largely controlled by a small number of corporations and advanced countries. This concentration of power creates challenges for the equitable distribution of AI's benefits, leaving smaller players—especially those in developing nations—at a disadvantage. As AI systems become embedded in critical areas like healthcare, finance, and criminal justice, they can cause harm if not developed or governed ethically and with transparency. Key challenges include algorithmic bias, discrimination, privacy violations, and limited transparency in decision-making processes. AI systems are trained on data that often reflect existing societal biases, whether originating from human inputs or emerging during data analysis—neither of which is AI's fault.

This can result in biased decision-making, particularly in sensitive areas like hiring, law enforcement, and lending. Democratising AI can reduce the risk of biased outputs by making development more inclusive and applications more responsive to diverse populations. However, if pursued without diligence and human oversight, it may also increase the bias, leading to discriminatory outcomes in areas such as hiring and criminal justice, and contributing to greater polarisation. At the same time, AI can be used ethically and responsibly if developers actively address societal biases embedded in data and algorithms.

Another key motivation for promoting the widespread distribution of AI tools is to collect diverse real-world data on how these technologies are used or potentially misused across a broader range of contexts than can be anticipated or tested in controlled environments. Such external feedback is vital, as it supports continuous model improvement and enables developers to identify and address emerging risks of misuse.¹² To address concerns around potential misuse, some developers have adopted a staged release approach to the democratisation of AI tools. In this method, increasingly powerful versions of a model are introduced incrementally, with deliberate pauses between releases to monitor usage patterns and conduct risk-benefit analyses. This strategy allows time for societies to adapt to new capabilities and to strengthen vulnerable systems, institutions, and processes against potential threats.¹³

However, where the risks of misuse are assessed as particularly high, responsible deployment may require placing access restrictions on certain AI functionalities. Although such restrictions inevitably curb the full democratisation of AI usage, they do not necessarily undermine the broader project of AI democratisation.¹⁴ As elaborated in other sections of this volume, democratising AI is not solely about universal access; it also entails embedding democratic principles in AI governance, shaping how technologies are developed, deployed, and regulated. In this sense, democratising AI governance may result in open access being only one possible outcome, depending on collective societal choices about how AI should be distributed and controlled.¹⁵

The Role of Non-State Actors in Democratising AI

One issue that arises is whether non-state actors—such as multinational technology companies, non-government organisations (NGOs), and advocacy groups—can address the democratic deficit in global AI governance. These actors are often recognised for their role in promoting accountability and transparency in AI governance by developing ethical frameworks and publishing transparency reports, particularly where AI-related decisions are taken without public oversight or inclusivity. However, these non-state actors lack formal authority and binding decision-making power under international law. In the absence of robust, regulated AI policy structures, their efforts alone cannot fully democratise global AI governance, which ultimately requires state-led mechanisms and intergovernmental cooperation. While democratic values such as accountability and transparency are advanced through these initiatives, they remain limited by the absence of enforceable authority.¹⁶

As powerful AI tools like ChatGPT and other Large Language Models (LLMs) become more common, the conversation is quickly shifting toward their social, political, and ethical effects. Many countries and regions are developing laws and policies to regulate the use of these technologies. At the same time, there is a growing expectation that civil society actors can represent public interest by ensuring that decision-making processes within international organisations and institutions are responsive to public opinion. These actors can function as watchdogs, holding those in positions of power accountable. Civil society groups have the potential to advance the democratisation of global governance by promoting greater inclusion, representation, and transparency.¹⁷

The international context is shaped by the fact that cutting-edge AI work is concentrated in international companies and global research laboratories, while AI's impacts increasingly extend across national borders, creating a growing need for international cooperation and regulation. This wider effort is often described as “global AI governance”, encompassing the various ways in which governments, international organisations, firms, and civil society actors seek to shape

the development and deployment of AI. Unlike domestic legal systems, which typically rely on a central enforcement authority, international AI governance lacks a single overarching system or leader.¹⁸

As AI technologies increasingly influence social, economic, and political contexts, it becomes imperative that those affected by these technologies have a legitimate and meaningful voice in determining their development and deployment. This article argues that non-state actors—particularly those with moral or epistemic standing—can play a constructive role in the democratic governance of AI. While not all non-state actors are equally equipped for this role, those with substantial market power, such as large technology companies, can and should be expected to strengthen their democratic legitimacy. In doing so, they may contribute to laying the foundations for a more inclusive, participatory, and open global AI governance regime.¹⁹

Recommendations

Democratising AI is essential to ensuring that its deployment remains ethical, transparent, and accountable. To this end, the legal personality of AI must first be recognised by states within existing legal frameworks, alongside clear and transparent regulatory and liability models to establish accountability for AI-driven actions and ensure responsible use grounded in ethics and transparency. At the same time, international regulations governing ethics, bias, and privacy must incorporate human oversight to ensure that AI continues to serve humanity, rather than replacing human judgement with algorithmic errors arising from flawed inputs.

It is widely agreed that a democratised AI ecosystem must reflect the plurality of human experience, including various kinds of cultural, social, and linguistic backgrounds to be accepted under human oversight. This requires training AI systems on diverse and representative datasets to reduce systemic bias and improve fairness in outcomes, thereby advancing ethical, transparent, and socially responsive objectives. Such an approach can support the development of more equitable and inclusive AI systems that address the needs of underrepresented communities while minimising risks of algorithmic bias. Further, to put checks in place, ethical AI governance frameworks need to be

established at both the national and global levels. These frameworks should have guidelines for ensuring fairness, transparency, accountability, and privacy in AI development and deployment. For example, AI systems should be subjected to thorough bias testing prior to being deployed in high-stakes sectors like healthcare and criminal justice.

Given the global impact of AI, collaboration is essential to establish shared norms, standards, and rules for its application. International organisations such as the United Nations and the OECD should work towards harmonising AI regulations worldwide, addressing concerns related to data privacy, transparency, and fairness. Discussions on AI regulation must include emerging economies to ensure that the benefits of AI are distributed beyond technologically advanced nations. The obstacles to democratising AI and implementing ethical governance are substantial but not insurmountable. By prioritising open access to AI, promoting diversity in AI development, investing in AI education, and establishing robust ethical governance frameworks, it is possible to envisage a future in which AI functions as a force for human advancement rather than a tool for manipulation or harm.

In conclusion, international cooperation is necessary to develop global standards and regulatory frameworks that ensure the fair distribution of AI's benefits. This way, AI can be utilised in solving some of the world's most critical challenges while minimising its harms and making sure that it serves the public interest.

Endnotes

- 1 EU Artificial Intelligence Act, “The EU Artificial Intelligence Act Up-To-Date Developments and Analyses of The EU AI Act,” <https://artificialintelligenceact.eu/>
- 2 European Commission, “Shaping Europe’s Digital Future AI Act,” <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
- 3 Carlo Costa et al., “The Democratization of Artificial Intelligence: Theoretical Framework,” *MDPI* 8236, no.14 (2024), <https://www.mdpi.com/2076-3417/14/18/8236>
- 4 Costa et al., “The Democratization of Artificial Intelligence: Theoretical Framework”
- 5 OECD, “AI Principles,” <https://www.oecd.org/en/topics/sub-issues/ai-principles.html>
- 6 UNESCO Artificial Intelligence and Emerging Technologies, “Ethics of Artificial Intelligence,” <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>
- 7 Jim Reitz, “3D Printing Today: Democratization of Technology and Disruptive Innovation Converge,” *Facefox*, March 19, 2018, <https://3dprint.com/207176/democratization-innovation/>
- 8 Stability.AI, “Stable Diffusion Launch Announcement,” August 10, 2022, <https://stability.ai/news/stable-diffusion-announcement>
- 9 Microsoft News Centre, “Democratizing AI: For Every Person and Organization,” *Microsoft*, September 26, 2016, <https://news.microsoft.com/features/democratizing-ai/>
- 10 “Democratized AI: Potential Benefits, Risks, and A Glimpse into the Future,” *Stefanini Group*, February 26, 2024, <https://stefanini.com/en/insights/news/democratized-ai-potential-benefits-risks-and-glimpse-to-future>
- 11 “Democratized AI: Potential Benefits, Risks, and A Glimpse into the Future”
- 12 Daniel Jeffries, “Let’s Speed Up AI,” *Future History*, February 4, 2023, https://danieljeffries.substack.com/p/lets-speed-up-ai?utm_medium=email
- 13 Jeffries, “Let’s Speed Up AI”
- 14 Irene Solaiman et al., “Release Strategies and the Social Impacts of Language Models,” *Cornell University* 14, no. 1908 (2019), <http://arxiv.org/abs/1908.09203>
- 15 Solaiman et al., “Release Strategies and the Social Impacts of Language Models”
- 16 Eva Erman et al., “The Democratization of Global AI Governance and the Role of Tech Companies,” *Nature Machine Intelligence*, no.6 (2024).
- 17 Patrizia Nanz and Jens Steffek, *Civil Society Participation in European and Global Governance* (Palgrave Macmillan, 2008)
- 18 Jeffries, “Let’s Speed Up AI”
- 19 Jeffries, “Let’s Speed Up AI”

Towards the Responsible Governance of Open-Source AI

Vibhav Mithal

“Open source” is widely considered to be an effective solution to challenges that may arise from a closed-source or closed-manner of functioning. The term owes its roots to software development.¹ Linus Trovalds, the creator of Linux and Git, has compared ‘open source’ to science, arguing that just as scientific progress builds on ideas developed openly and improved by others, open source operates in a similar way.^{a,2}

a Linus Trovalds has stated, “I often compare open source to science. Science took this whole notion of developing ideas in the open and improving on other peoples’ ideas. It made science what it is today and made the incredible advances that we have had possible. And I compare that to witchcraft and alchemy, where openness was something you didn’t do.”

Applied to Artificial Intelligence (AI)—particularly after the release of China’s DeepSeek^b—it is claimed that open-source AI “will help foster a collaborative environment and accelerate AI innovation.”³

This article avoids the use of a precise legal definition and conceptualises ‘open-source AI’ by focusing on the components of an AI system. Since an AI system contains many components, the present article highlights that open-source AI is not merely open-source software; software is only one component of an AI system. It also highlights that interpreting the term “open” in “open-source AI” requires exploring and answering what component of an AI system ought to be made open. The article explores this issue by highlighting the guidance, if any, provided by the European Union Artificial Intelligence Act (EU AI Act).⁴ It also highlights the unique governance challenges associated with open-source AI.

AI Governance

AI governance is defined as “a system of laws, policies, frameworks, practices and processes at international, national, and organisational levels. AI governance helps various stakeholders implement, manage, oversee and regulate the development, deployment and use of AI technology. It also helps manage associated risks to ensure AI aligns with stakeholders’ objectives, is developed and used responsibly and ethically, and complies with applicable legal and regulatory requirements.”⁵ AI governance extends beyond laws to include policies and frameworks that may be voluntary in nature, as well as practices and processes tailored to specific audiences.^{c,6} The present article adopts a national, India-specific perspective on AI governance, linked to the concept of Responsible AI as articulated in India.

b DeepSeek is a Chinese AI company that rose to international prominence in January 2025 following the release of its mobile chatbot application and the large language model DeepSeek-R1. DeepSeek was founded in December 2023 by Liang Wenfeng.

c Illustratively, AI Governance may involve assessment of issues that are broader than laws. For example, while compliance of data sources with privacy regulations, protection of the AI System by way of Intellectual Property laws and the ascertainment of liability when an AI output violates someone’s legal rights may be legal issues, AI Governance strategies involve addressing questions such as (a) Who decides which data is included in the training dataset, how is a model tested for bias and/or how is model performing. Addressing each of these aspects requires technical appreciation of an AI System.

Understanding Open-Source AI Through the Components of an AI System

Reaching agreement on a definition of artificial intelligence, particularly from a legal and regulatory perspective, presents clear challenges. The technical vocabulary used in the field often differs from its non-technical articulation. For example, while technical literature may refer to terms such as “machine learning” or “classical statistical methods”, a policy lens may subsume both under the broader category of “AI”.⁷ This divergence was evident in the evolution of the definition of AI in the 2021 draft of the EU AI Act, which was modified in early 2023. Articulating an acceptable definition of open-source AI has similar challenges, as there may be disagreement upon what exactly is open-source AI, and what exactly makes a model open-source in contrast to being closed-source.⁹

The present article frames an understanding of open-source AI, not by a functionality-based definition,^{d,10} but by contextualising the concept of AI by identifying the different components of an AI system. Moreover, with this focus of articulating open-source AI through the components of an AI system, it is also necessary that the term “AI model” and its interlink with an AI system is understood. The subtle distinction between an “AI model” and an “AI system” becomes important to better appreciate the term “open-source AI”.

Simply put, an AI model performs a specific task,¹¹ whereas an AI system is broader in scope, as it may integrate one or more AI models to utilise their outputs in support of a broader business decision.¹² The additional components that an AI system illustratively contains, includes processes such as data collection and processing, a user interface, and infrastructure.¹³ For example, an AI model in the healthcare space may analyse a patient’s chest scan to predict or arrive at an outcome as to whether a patient has COVID-19. The AI system may then utilise this prediction, along with other patient information, to arrive at a treatment plan.¹⁴

d A functionality based definition explains a concept/term by describing the specific capabilities or functions associated with it, rather than its characteristics.

At a practical level, the use of the term “open-source AI model” suggests that such a model would operate in conjunction with an AI system. On the other hand, the term “open-source AI” would refer to an AI system that incorporates one or more AI models. Since “open-source AI”, as used in common parlance, may in fact refer to AI systems, it becomes necessary to understand the concept of an AI system itself. As a corollary, to understand open-source AI, while it may be useful to have a definition of open-source AI, this term may also be appreciated by identifying the different components of an AI system. Additionally, as explored below, understanding these different components provides the necessary context for interpreting the term ‘open’ in open-source AI.

One way to identify the components of an AI system is through the lens of intellectual property law, as such a lens may enable innovators, businesses, and regulators to ascertain which aspects of an AI system require adequate legal protection to maximise the economic and legal value within each component. Viewed from an IP lens, coupled with an understanding of AI, the components of an AI system¹⁵ may be identified as hardware (e.g., semiconductor chips), training datasets, AI algorithms (i.e., the set of technical techniques to enable pattern recognition/creation of AI output), software in which these algorithms are embedded (such as running the application that embeds AI¹⁶), and outcome/output.

Based on the above components, it is evident that computer software is only one component of an AI System. Consequently, it would be misleading to refer to open-source AI as merely computer software. Open-source AI would necessarily involve hardware, training datasets, algorithms, and an output. Thus, open source AI is not open-source software.¹⁷

Refocusing the Open-Source AI Discourse on System Components

Parallel to conceptualising open-source AI through the components of an AI system, the meaning of the term ‘open’ in ‘open-source AI’ is equally critical. Simply put, the focus of open-source AI lies in determining which component of an AI System, i.e., hardware, training

datasets, AI algorithms or software (or their sub-components, as identified in relevant technical literature)—ought to be made open, and why, in order to realise the benefits and potential of openness.

As this discourse necessarily involves choosing between different components of an AI system and deciding which should be made open and why, it is possible that not every component of an AI system may be made open. Each component, including hardware such as semiconductor chips, serves a distinct function in building an AI system. A holistic understanding of the role of each component is therefore essential in determining whether it should be made open.

The Open Source Initiative (OSI), in 2024, articulated the first global definition of open-source AI as “an AI system made available under terms in a way that grant the freedoms to (a) use the system for any purpose and without having to ask for permission; (b) study how the system works and inspect its components; (c) modify the system for any purpose, including change its output; (d) share the system for others to use without modifications, for any purpose.”¹⁸ According to this definition, open-source AI systems can be used, studied, modified, and shared for any purpose, which includes changing an AI system’s output.

OSI provides additional context by stating that an open-source AI system is viewed “broadly” as a “fully functional structure” and its “discrete structural elements”.¹⁹ Therefore, recognising that an AI system comprises different components. Moreover, OSI clarifies that the same freedoms apply “whether applied to a system, a model, weights and parameters, or other structural elements”.²⁰ The fact that freedoms of open-source may apply equally to either the system, model, weights and parameters or other structural elements reinforces the view that different components of an AI system may independently be made “open”.

The focus thus remains on understanding open-source AI through the components of an AI system.

What Can Be Made ‘Open’ in Open-Source AI?

When the focus shifts to identifying which components of an AI system may be made open, it becomes evident that open-source AI may still contemplate restrictions on the use of different system components. Illustratively, the terms of use of Google’s Gemma open-source AI models²¹ make these models “freely available for individual developers, researchers, commercial users for access and redistribution, and users are also free to create and publish model variants”.²² Although access to model weights is provided, conditions relating to use, redistribution, and ownership of variants vary across models, and these terms are not necessarily grounded in an open-source license.²³ For instance, the terms of use²⁴ restrict applications of these models as per Gemma’s Prohibited Use Policy.²⁵ Simply put, even though Google Gemma models are open-source, Google does not permit any kind of use and restricts the way their models may be used.

Simultaneously, when open-source AI is examined from a component-based perspective, it is important to acknowledge that certain components of AI system—such as specific categories of AI algorithms—are already ‘open’. Examples include open-source libraries such as Scikit-Learn²⁶ for machine learning algorithms, Tensor Flow²⁷ (Google) and PyTorch²⁸ (Meta) for building and training deep neural networks. Similarly, open-source libraries exist for natural language processing,²⁹ computer vision,³⁰ and reinforcement learning.³¹

Instructive guidance on the question of ‘what ought to be made open’ in open source AI—particularly for jurisdictions outside the EU—may also be drawn from the EU AI Act. The Act defines “general purpose AI models”³² inclusively as models trained on large amounts of data that display “significant generality and are capable of competently performing a wide range of distinct tasks”.³³ Illustratively, ChatGPT may qualify as general-purpose AI models under the EU AI Act.³⁴ Under the Act, a “provider” includes a developer of a general-purpose AI model,³⁵ and the obligations applicable to such providers are provided in Article 52.

Guidance on ‘what ought to be made open’ is stated in Article 53(2), which provides exemptions from the obligations under Article 53(1) for

providers of AI models. Article 53(1) provides detailed obligations for providers of general-purpose AI models. In essence, where providers release their models under a free and open-source licence that (a) permits access, use, modification, and distribution of the model, and (b) makes publicly available the model's parameters, including weights, information on model architecture, and information on model usage, the obligations under Article 53(1) do not apply.³⁶ At the same time, AI systems released under free and open-source licences remain subject to regulation if they (a) qualify as high-risk systems; (b) fall within prohibited AI practice; or (c) are designed to interact with natural persons.³⁷ In answering what 'ought to be open', the EU AI Act thus points to two elements: free and open-source licences permitting access, use, modification, and distribution, and the publicly availability of parameters such as weights; model architecture, and information on model usage.

Scholars hold differing beliefs on what should be considered "open" in open-source AI. While 'open' is meant to suggest transparency, reusability, and extensibility, its practical application varies greatly.³⁸ Pertinently, many scholars converge on the traditionalist idea that open-source AI has its roots in the principles of open-source software (OSS), linking openness primarily to open-source code and permissive licensing.³⁹ In this context, an open-source AI model is one whose code and weights are freely accessible allowing for scrutiny, modification, and redistribution. Accordingly, models with openly available weights and licences permitting modification and reuse—such as Meta's Llama 2, Google's Gemma 2, Stability AI's Stable Diffusion, and EleutherAI's Pythia—are often cited as examples.⁴⁰ Additionally, model architecture disclosure, albeit without complete training data or process transparency, also counts as openness.⁴¹

Separately, many scholars believe that openness requires the release of more than model weights and should include other important components throughout the AI lifecycle, such as model architecture, training datasets, training methodologies, evaluation datasets and results, detailed documentation, source code, and licensing terms.⁴² This reflects a gradient of openness for AI models based on the levels of releases, the maximal extension being fully open.⁴³ As already stated above, given the article's component-based conceptualisation of open-

source AI, the author takes the view that a fully open form of open-source AI is not a pragmatic option.⁴⁴

The viable way to address the question of what 'ought to be open' is to view open-source AI as a gradient of openness. If viewed in this manner, scholars have also suggested that openness in AI may be addressed through a risk-mitigation paradigm that accounts for the multiple actors involved, including cloud providers, data providers, model providers, optimisers, application developers, distribution platforms, users, and evaluation providers.

To address this issue, a team of international researchers led by Matt White of the Linux Foundation, in their paper "The Model Openness Framework: Promoting Completeness And Openness For Reproducibility, Transparency and Usability in Artificial Intelligence," have proposed a pioneering three-level classification system, comprising an evidence-based framework of 17 components that may be made open.⁴⁵ This classification seeks to "classify the degree of completeness and openness of an AI model across all aspects of the model's development cycle."⁴⁶ The 17 components are then categorised into three classes, each building upon the previous one, "with Class III being the least complete and Class I being the most complete."⁴⁷

Simply put, components placed in Class I represent those that may be made open to reflect the lowest degree of openness. The seven components in Class I include: (1) model architecture; (2) model parameters (final checkpoint); (3) technical report or research paper; (4) evaluation results; (5) model card; (6) data card; (7) sample model outputs (optional). Where greater openness is sought, the components in Class II (along with components of Class I) may be made open. Class II includes: (1) training, validation, and testing code; (2) inference code; (3) evaluation code; (4) evaluation data; and (5) supporting libraries and tools. Finally, where maximum openness is sought, the components in Class III (along with components of Class I and Class II) may be made open. These include: (1) research papers; (2) datasets; (3) data preprocessing code; (4) model parameters; and (5) model metadata.⁴⁸

Effectively addressing what all should be made open in an open-source AI model is beyond the scope of this article. However, the above analysis indicates that (a) placing restrictions in existing open-source AI models does not disqualify an AI system from being open-source AI because in this scenario only the AI model within the AI system has been made open; (b) emphasising that when AI is viewed by way of components of the system, components such as certain algorithms are already open; (c) with respect to other components that are currently closed-source, the question is which component ought to be made open and why; (d) laws, such as EU AI Act, may provide instructive guidance and (e) demonstrating that scholars do confuse open-source AI with open-source software however ultimately agree that openness is a gradient.

Responsible Governance and the Challenges of Open-Source AI

The term “responsible” is interpreted in this article through the lens of Responsible AI, as framed by the Principles of Responsible AI laid down by the NITI Aayog.⁴⁹ These principles are fairly settled and include equality, inclusivity, and non-discrimination.⁵⁰ Illustratively, the principle of safety and reliability requires AI systems to be deployed with sufficient safeguards to minimise risks to relevant stakeholders, while the principle of privacy and security requires the protection of individuals’ data used in AI training.⁵¹ The principle of accountability states that all stakeholders involved in the design, development, and deployment of AI systems must be responsible for their actions, / conduct risk and impact assessments, and establish internal or external auditing mechanisms to oversee adherence to these principles.⁵² While the framing and objectives of these principles are broadly accepted and provide the context for “responsible”, the specific actions required to implement Responsible AI remain contextual.⁵³

In the sphere of open-source AI, where conceptualisation and definition are themselves contested, identifying concrete actions to implement the principles of Responsible AI requires an understanding of the associated governance challenges. Moreover, the author contends that the question of ‘what is to be made open’ in open-source AI is relevant primarily up to the point at which the open-source AI system, including

the AI model, is released. Once released, the governance challenges associated with open-source AI may, to some extent, converge with those faced by closed-source or proprietary AI systems. It is therefore imperative to examine the governance of open-source AI in this context, and the present article broadly highlights illustrative unique governance challenges that are distinctive to open-source AI.

Safety

Open-source AI poses serious safety concerns due to the absence of centralised monitoring and control, which makes it difficult for developers to monitor and contain misuse once such systems are released. Easy access presents unique challenges, as it makes it easier to share and customise these models, furthering the likelihood of unintended harms. Additionally, open access of these models may lead to malicious fine-tuning and/or circumvention of important safety filters and content moderation techniques, thus increasing the possibility of the generation and spread of offensive outputs.⁵⁴

Accountability

Another unique challenge rising from the open design of AI models relates to legal and regulatory complexity. Inflicting liability for downstream harm may be disproportionately borne by developers who release their models freely, even though they exercise little or no control over the diverse and potentially harmful uses to which others may put them. Further, establishing content provenance—an increasingly important concern in the context of AI-generated content—becomes considerably more difficult when models are freely shared and open to modification by numerous, often unidentifiable actors. This lack of traceability complicates efforts to identify the source of harmful content and to assign accountability.⁵⁵

Lack of Uniformity in the Licence Terms

Where components of an AI system are made open in open-source AI, those components are typically subject to licensing terms. The nature of such licences varies across different open-source models. For instance, state-of-the-art foundation models often impose licence restrictions based

on factors such as user base size, age, or geographical location, while other models incorporate different forms of use-based limitations. It is therefore important to understand that these models frequently operate under specific licensing conditions. Users may inadvertently expose themselves to legal risk if they fail to understand or comply with these restrictions.⁵⁸

Compute Challenges

Open-source AI also raises compute-related considerations that warrant independent assessment when evaluating its viability within the broader ecosystem. Such systems rely on large datasets, which in turn require massive computational power to process.⁵⁹ “Access to compute presents a significant barrier to reusability”⁶⁰ for open AI systems owing to the high cost involved in training and running inferences on large-scale AI models at scale.⁶¹ For instance, one study showed that “training the Llama 2 series of models required 3.3 million GPU hours on NVIDIA A100-80GB GPUs”⁶² and that, as per February 2024 rates of training, the cost would have been approximately US\$6 million.⁶³

Finally, akin to proprietary closed models, open-source AI may face issues such as bias (for instance, in training data), limitations in explainability, and risks of violating personal data protection and intellectual property rights.

Recommendations and Conclusion

For open-source AI, a critical starting point is clarity about what the term denotes. First, an AI Model is distinct from an AI System. An AI system includes an AI model and may also comprise additional components such as processes for data collection and processing, a user interface, and supporting infrastructure. Second, as the term “open-source-AI” is used in common parlance, it refers to AI Systems rather than models in isolation. Thus, to fully appreciate ‘open-source AI’, the components of an ‘AI System’ must be understood.

The governance of open-source AI should centre on identifying what is to be made “open,” and why, recognising that ‘openness’ in AI systems exists along a gradient. The governance of open-source AI

also presents a unique set of challenges, in conjunction with existing risks that apply to proprietary closed-source models. AI governance also encompasses a broad system of laws, policies, frameworks, and practices aimed at managing risks arising from the activities of various actors involved in the design, development, and deployment of AI systems. Therefore, while addressing the unique risks that arise from open-source AI, the overall governance of open-source AI would operate in a manner similar to AI governance initiatives focusing on overall risk identification and mitigation of AI systems.

Vibhav Mithal is an artificial intelligence and intellectual property lawyer and an Associate Partner at Anand & Anand; an ISO 42001 Lead Implementer, a ForHumanity Certified AI Auditor on the Digital Personal Data Protection Act, 2023, and co-author of the *NASSCOM Developer's Playbook for Responsible AI*.

Endnotes

- 1 Agnes Nduta, “A Brief History of Open Source,” *Free Code Camp*, April 3, 2023, <https://www.freecodecamp.org/news/brief-history-of-open-source/>
- 2 Sara Ana Cemazar, “Inspirational Quotes on Open Source from 10 Leading Experts,” Rocket Chat, October 6, 2021, <https://www.rocket.chat/blog/open-source-quotes>
- 3 Charlotte Edmond, “What is Open Source AI and How Could Deepseek Change the Industry,” World Economic Forum, February 5, 2025, <https://www.weforum.org/stories/2025/02/open-source-ai-innovation-deepseek/>
- 4 “2024/1689 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 And (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act),” *Official Journal of the European Union*, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689
- 5 IAPP, “Key Terms for AI Governance Essential terms and Explanations for AI governance,” <https://iapp.org/resources/article/key-terms-for-ai-governance/#:~:text=AI%20governance-,Definition,and%20use%20of%20AI%20technology>
- 6 Peter Wayner, “What is AI Governance?,” *Venture beat*, June 24, 2022, <https://venturebeat.com/ai/what-is-ai-governance-2/>
- 7 Matt O’Shaughnessy, “One of the Biggest Problems in Regulating AI Is Agreeing on a Definition,” Carnegie Endowment, October 6, 2022, <https://carnegieendowment.org/posts/2022/10/one-of-the-biggest-problems-in-regulating-ai-is-agreeing-on-a-definition?lang=en>
- 8 Luca Bertuzzi, “EU Lawmakers Set to Settle on OECD Definition For Artificial Intelligence,” *Euractiv*, September 30, 2024, <https://www.euractiv.com/section/tech/news/eu-lawmakers-set-to-settle-on-oecd-definition-for-artificial-intelligence/>
- 9 Edd Gentarchive, “The Tech Industry Can’t Agree on What Open-Source AI Means. That’s a Problem,” *MIT Technology Review*, March 25, 2024, <https://www.technologyreview.com/2024/03/25/1090111/tech-industry-open-source-ai-definition-problem/>
- 10 Merriam Webster, “Functionality-noun,” <https://www.merriam-webster.com/dictionary/functionality>
- 11 IBM, “What is an AI Model?,” <https://www.ibm.com/think/topics/ai-model>
- 12 Marine Boquien, “What is the Difference Between an AI System and an AI Model,” *Dastra*, July 17, 2024, <https://www.dastra.eu/en/article/difference-between-an-ai-system-and-an-ai-model/57721>
- 13 Boquien, “What is the Difference Between an AI System and an AI Model”
- 14 Anja Thieme and Cecily Morrison, “AI Models vs. AI Systems: Understanding Units of Performance Assessment,” Microsoft, September 19, 2022, <https://www.microsoft.com/en-us/research/blog/ai-models-vs-ai-systems-understanding-units-of-performance-assessment/>; Boquien, “What is the Difference Between an AI system and an AI Model”

- 15 Marta Duque Lizarralde, "A Guideline to Artificial Intelligence, Machine Learning and Intellectual Property," 4lp Council, September 2020, https://www.4ipcouncil.com/application/files/9016/0017/8691/A_Guideline_to_Artificial_Intelligence_Machine_Learning_and_Intellectual_Property.pdf)
- 16 Dmitry Baraishuk, "What is Artificial Intelligence? AI Vs Traditional Software," Belitsoft, January 21, 2024, <https://belitsoft.com/ai-development/artificial-intelligence-vs-conventional-software>
- 17 Rishi Bommasani et al., "Considerations for Governing Open Foundation Models," *Artificial Intelligence* 386, no. 6718, October 10, 2024, <https://www.science.org/doi/10.1126/science.adp1848>.
- 18 Open Source Initiative, "Why We Need Open Source Artificial Intelligence (AI)," <https://opensource.org/ai/open-source-ai-definition>
- 19 Open Source Initiative, "Why We Need Open Source Artificial Intelligence (AI)"; Rhiannon Williams, "We Finally Have a Definition for Open Source AI," *MIT Technology Review*, August 22, 2024, <https://www.technologyreview.com/2024/08/22/1097224/we-finally-have-a-definition-for-open-source-ai/>
- 20 Open Source Initiative, "Why We Need Open Source Artificial Intelligence (AI)"
- 21 Sean Michael and Kerner Dave Raffo, "What is Gemma? Google's Open Sourced AI Model Explained," Techtarget, March 14, 2025, <https://www.techtarget.com/searchenterpriseai/definition/Gemma>
- 22 Google Cloud, "Level Up Your AI Skills With Open-Source AI Tools," <https://cloud.google.com/use-cases/open-source-ai>
- 23 Google Cloud, "Level Up Your AI Skills with Open-Source AI Tools,"
- 24 Google AI for Developers, "Gemma Terms of Use," <https://ai.google.dev/gemma/terms>
- 25 Google AI for Developers, "Gemma Prohibited Use Policy"
- 26 Scikit-learn, "Machine Learning in Python," <https://scikit-learn.org/stable/>
- 27 Tensorflow, "An End-To-End Platform for Machine Learning," <https://www.tensorflow.org/>
- 28 PyTorch, <https://pytorch.org/>
- 29 Spacy, "Industrial-Strength Natural Language Processing," <https://spacy.io/>; Huggingface, "Transformers," <https://huggingface.co/docs/transformers/index>; Gensim, "Why Gensim," <https://radimrehurek.com/gensim/>
- 30 OpenCV, "Opencv is the World's Biggest Computer Vision Library," <https://opencv.org/>; PyTorch, "Torchvision," <https://pytorch.org/vision/stable/index.html>; TensorFlow, "Computer Vision With Tensorflow," <https://www.tensorflow.org/tutorials/images>
- 31 Stable Baselines3, "Stable-Baselines3 Docs - Reliable Reinforcement Learning Implementations," <https://stable-baselines3.readthedocs.io/en/master/>
- 32 Official Journal of the European Union, "2024/1689 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 And (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)"

- 33 Official Journal of the European Union, “2024/1689 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 And (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)”
- 34 European Parliament, “General-Purpose Artificial Intelligence,” [https://www.europarl.europa.eu/RegData/etudes/ATAG/2023/745708/EPRS_ATA\(2023\)745708_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2023/745708/EPRS_ATA(2023)745708_EN.pdf)
- 35 Official Journal of the European Union, “2024/1689 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 And (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)”
- 36 Official Journal of the European Union, “2024/1689 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 And (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)”
- 37 Official Journal of the European Union, “2024/1689 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 And (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)”
- 38 David Gray Widder et al., “Open (For Business): Big Tech, Concentrated Power and the Political Economy of Open AI,” *SSRN*, no. 2 (2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4543807
- 39 Widder et al., “Open (For Business): Big Tech, Concentrated Power and the Political Economy of Open AI”
- 40 Widder et al., “Open (For Business): Big Tech, Concentrated Power and the Political Economy of Open AI”
- 41 Widder et al., “Open (For Business): Big Tech, Concentrated Power and the Political Economy of Open AI”
- 42 Widder et al., “Open (For Business): Big Tech, Concentrated Power and the Political Economy of Open AI”
- 43 Widder et al., “Open (For Business): Big Tech, Concentrated Power and the Political Economy of Open AI”
- 44 Matt White et al., “The Model Openness Framework: Promoting Completeness and Openness For Reproducibility, Transparency, and Usability in Artificial Intelligence,” Cornell University, 2024, <https://arxiv.org/pdf/2403.13784>
- 45 White et al., “The Model Openness Framework: Promoting Completeness and Openness for Reproducibility, Transparency, and Usability in Artificial Intelligence”
- 46 White et al., “The Model Openness Framework: Promoting Completeness and Openness for Reproducibility, Transparency, and Usability in Artificial Intelligence”

- 47 White et al., “The Model Openness Framework: Promoting Completeness and Openness for Reproducibility, Transparency, and Usability in Artificial Intelligence”
- 48 White et al., “The Model Openness Framework: Promoting Completeness and Openness for Reproducibility, Transparency, and Usability in Artificial Intelligence”
- 49 Niti Aayog, *Responsible AI Approach Document For India Part 1 – Principles For Responsible AI*, February 2021, New Delhi, Niti Aayog, 2021, <https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf>
- 50 Niti Aayog, *Responsible AI Approach Document for India Part 1 – Principles for Responsible AI*
- 51 Niti Aayog, *Responsible AI Approach Document for India Part 1 – Principles for Responsible AI*
- 52 Niti Aayog, *Responsible AI Approach Document for India Part 1 – Principles for Responsible AI*
- 53 NASSCOM, *The Developer’s Playbook For Responsible AI in India*, November 2024, India, NASSCOM, 2024, <https://nasscom.in/ai/pdf/the-developer’s-playbook-for-responsible-ai-in-india.pdf>
- 54 Madhulika Srikumar, *Risk Mitigation Strategies For The Open Foundation Model Value Chain: Insights From PAI Workshop Co-Hosted With Github*, Partnership on AI, 2024, https://partnershiponai.org/wp-content/uploads/dlm_uploads/2024/07/open-foundation-model-risk-mitigation_rev3-1.pdf; Rishi Bommasani et al., “Considerations for Governing Open Foundation Models,” Stanford University HAI, 2023, <https://hai.stanford.edu/policy/issue-brief-considerations-governing-open-foundation-models>
- 55 Bommasani et al., “Considerations for Governing Open Foundation Models”
- 56 “Risk Mitigation Strategies for The Open Foundation Model Value Chain: Insights from PAI Workshop Co-Hosted with Github”
- 57 “Risk Mitigation Strategies for The Open Foundation Model Value Chain: Insights from PAI Workshop Co-Hosted with Github”
- 58 “Risk Mitigation Strategies for The Open Foundation Model Value Chain: Insights from PAI Workshop Co-Hosted with Github”
- 59 Widder et al., “Open (For Business): Big Tech, Concentrated Power and the Political Economy of Open AI”
- 60 Widder et al., “Open (For Business): Big Tech, Concentrated Power and the Political Economy of Open AI”
- 61 Widder et al., “Open (For Business): Big Tech, Concentrated Power and the Political Economy of Open AI”
- 62 Sayush Kapoor et al., “On the Societal Impact of Open Foundation Models,” *Cornell University*, 2024, <https://arxiv.org/abs/2403.07918>
- 63 Kapoor et al., “On the Societal Impact of Open Foundation Models”



IV.

Envisioning a Decentralised and Open AI Ecosystem

Exploring Non-Technical Pathways to Building a Decentralised and Open AI Ecosystem

*Aishani Rai and
Gautam Misra*

Early 2025 set the stage for critical debates within the AI ecosystem. DeepSeek, a groundbreaking open-source AI model developed in China, has disrupted not only Silicon Valley but also the highest levels of regulatory discourse worldwide. The model, built at a fraction of the cost of its competitors, has demonstrated the immense potential of open AI by enabling others to innovate and build upon its foundational layer.¹

The impact of this development parallels India's approach to Digital Public Infrastructure (DPI), which serves as a cohesive framework linking socio-economic activities with digital

public goods. While the DPI story examines how use cases have been legitimised by incorporating the element of public good, there has also been some discussion around how AI-driven use cases can serve as a functional data layer within DPIs. Within the DPI discourse, terms such as ‘inclusivity’, ‘collaboration’, and ‘accountability’ are recognised as pillars for ensuring equitable and efficient digital ecosystems. These values align with India’s democratic principles and underscore the importance of embedding similar tenets within AI governance frameworks.²

Fostering a democratic AI ecosystem necessitates two interlinked priorities: decentralisation and the establishment of non-technical enablers that support an open AI paradigm. Decentralisation within AI ecosystems is not merely about distributing technological infrastructure; it involves diversifying stakeholder engagement, fostering cross-sector expertise, and ensuring equitable access to resources and knowledge. When these elements align, they create fertile ground for co-creation, knowledge-sharing, and collective problem-solving, ultimately leading to the development of more resilient and inclusive AI systems. Equally important is the need to build non-technical layers, such as open-source communities around AI solutions, and reshape financing practices. While technical layers constitute tech-oriented intricacies towards the development, deployment, and adaptation of AI systems and use cases, non-technical layers constitute elements that inform the interdependent elements of governance and sustainability of infrastructure that play a crucial role in shaping a democratic AI ecosystem through advocacy, capacity-building, and participatory decision-making processes.

This article looks to highlight the value that can be unlocked by the decentralisation of India’s AI ecosystem, while its core focus will be on building a case for pathways to non-technical enablers as essential drivers of a democratised ecosystem. It will also glean insights from India’s DPI model and global open-source AI developments to foster a more open, participatory, and responsible AI landscape.

Towards a Decentralised and Open AI Ecosystem for India

The concept of decentralisation carries similar significance in both India’s DPI ecosystem and the ongoing efforts to build an open AI

ecosystem. India's Unified Payments Interface (UPI), a real-time payment system that enables instant money transfer between banking institutions and mobile interfaces has taken individuals away from cash transactions and card payments. With interoperability at its core and a mission to reimagine tech as a tool for inclusive growth, UPI has become a global benchmark for innovations in digital infrastructure.^{3,4}

Drawing parallels from the UPI story, India is now gearing towards reducing dependence on global AI model developers by creating a sovereign, open-source GenAI Chatbot.⁵ Adapted from a Chinese AI model, this initiative ensures data security, sovereignty, and local control by hosting the model entirely on Indian servers. By shifting power away from a handful of global AI giants, India intends to empower local developers, startups, and research institutions to assume a more central role in shaping the AI landscape.⁶

From Monopolies to an Ecosystems Approach

Big Tech's aggressive competition in areas such as compute power, data, and widespread accessibility has left no stone unturned. During the past half-decade, the ecosystem has witnessed a sharp increase in startups creating unique AI use cases, only to be almost instantaneously acquired by big tech organisations as they are trained on the latter's infrastructure and networks. A prime target for these pseudo-acquisitions are AI startups that primarily focus on generative AI and natural language processing. Big tech entities are now racing to strike deals to license key models and tools, which allows them to capture the talent that built them. An example of this is how Google went through with one of the most renowned reverse acqui-hires where they secured a non-exclusive licence for Character.AI's technology.^{7,8} This upstream monopolisation stifles competition and innovation, steering AI development toward commercial interests rather than the public good.⁹ While smaller startups have historically disrupted industries, AI presents an even greater challenge—incumbents not only have a technological head start but also leverage their platforms to lock users, as in the aforementioned case.¹⁰

AI capabilities are heavily skewed in favour of big tech players, and this is by no means a level playing field. From exorbitant cloud

infrastructure, larger datasets, to specialised talent, there needs to be incentivisation for smaller players and other stakeholders across the AI value chain to push their weight in terms of negotiating for equal opportunities. The current scenario locks out critical stakeholders, and this imbalance is not just about market competition; it determines the future of India's AI landscape and for whom it is designed.

For smaller innovators, infrastructural limitations and restricted access to good-quality datasets make it nearly impossible to train competitive models without an overarching dependency on bigger players. Researchers and academia face a constant struggle to incentivise AI talent to stay dedicated to in-house initiatives, as AI labs by larger entities offer access to proprietary datasets and top-of-the-line infrastructure.¹¹ Regulators currently face multiple challenges, one of them being Indian tech leaders adopting a one-size-fits-all narrative that is built on models and datasets from the Global North. This is done without consideration of cultural and linguistic nuances and use cases wherein regulators need to account for diversity.^{12,13}

The downstream effects of such models and datasets for Indian applications are also left out of the equation. Users—the ultimate stakeholders—are increasingly nudged into AI ecosystems they cannot negotiate with, reinforcing surveillance capitalism—a method by which big tech companies harvest, analyse, and sell user data to better predict and influence user behaviour. This, in turn, limits AI's potential for public good. If the technology is to serve broader societal and developmental goals rather than just commercial interests, there is an urgent need to incentivise user participation across the AI value chain. This means creating open-source alternatives, lowering barriers to computing power, establishing fairer data access mechanisms, and embedding transparency and accountability into AI governance. Without these interventions, AI risks becoming a tool for reinforcing power hierarchies rather than a force for inclusive innovation.¹⁴

Unpacking a Socio-Technical Lens to Drive Democratised AI Ecosystems

India must adopt a socio-technical approach that empowers critical stakeholders across the AI value chain. This allows the ecosystem at

large to move beyond the narrow focus on technical performance and instead examine how AI systems interact with and reshape broader social structures.¹⁵

AI is not just a neutral tool; it is embedded within organisational hierarchies, regulatory frameworks, labour ecosystems, and power dynamics. The outcomes AI produces are shaped not only by its algorithms but also by who builds it, governs it, and benefits from it. Recognising AI as a socio-technical system highlights the need for an open and democratic AI ecosystem, where diverse stakeholders, including local innovators, civil society, regulators, researchers, and users, can actively shape its development and deployment rather than being passive adopters of proprietary models controlled by a few dominant names. This means moving beyond a reality where AI decisions are made solely within Big Tech and extending agency to users, communities, and institutions that interact with AI systems. If developers themselves are still grappling with the opacity and unpredictability of their models, communities should exercise their right to peremptorily contest and negotiate on how AI applications are developed and data is used.¹⁶

At its core, this approach would give Indian stakeholders greater control over AI development, governance, and commercial applications, rather than leaving these crucial domains in the hands of a few foreign entities and a volatile global market subject to disruptions. This is not just a technological shift but a strategic one; controlling AI models means controlling data, economic opportunities, national security, and the future of innovation—all of which have traditionally been concentrated in the Global North.

A decentralised approach to AI would democratise innovation by enabling broader participation in model development, reducing barriers to entry for AI startups, and fostering solutions tailored to India's socio-economic needs. Embedding layers of community and streamlined financing practices within the AI ecosystem catalyses the wider adoption of AI for public good, ensuring that the technology serves national interests and developmental goals rather than purely corporate profits.

The Non-Technical Layers to Support the Building of Open Technical Layers

Decentralisation, within the context of an AI ecosystem, is facilitated by strengthening the capacities of local stakeholders, augmenting their power to carry out their intended functions within the ecosystem. For Indian businesses, such power translates into enhanced abilities to build, access, and manage local datasets and computing power. For regulators, such power would boost their ability to shape the discourse of innovation and governance within the national ecosystem, ensuring that the overarching goals of leveraging AI for social and public good are met.

While there exists a preliminary understanding of the technical layers of how this ecosystem could be built, there is limited enquiry into the non-technical layers that bolster the development of a decentralised and open AI ecosystem. These non-technical layers outline practices, methods, and mechanisms that support the process of building decentralised and open technical layers. This article identifies two such layers that could strengthen the capacities of local stakeholders and attempts to unpack their transformative potential. With a strong imperative on building capacity, this analysis explores the potential of nurturing open-source communities and reshaping financing practices to enhance the resilience and sustainability of AI endeavours.

Nurturing Open-Source Communities within Open AI Ecosystems

The Indian AI ecosystem is rapidly advancing towards leveraging open-source to strengthen its AI capabilities. As attempts are made to decentralise the power dynamic, open-source has been championed as the way forward, enabling the optimal utilisation of limited resources—aligning with the overarching goals of promoting social and public good. The vision of open-source AI ecosystems is most compatible with the goals of the IndiaAI mission, which seeks to ensure that the discourse of AI innovation is shaped by diverse and local ecosystem actors.¹⁸

While developing open-source AI is pivotal in enhancing capabilities, building a community around it should be equally prioritised. Community building and developing the technical layer are not sequential, but

rather parallel to each other. An active and vibrant community ensures the sustainability of open-source AI solutions by making sure that developers and other members continue to contribute to public good solutions, even in the absence of tangible and material incentives.

The Indian DPG landscape provides examples where the building and participation of communities around DPGs was the first step to growing and sustaining the DPG reliably in the long run.¹⁹ Understanding the trajectories of global open-source projects sheds light on how building communities was never an afterthought, but always a key consideration from the outset.²⁰

Community building around open-source AI, characterised by values of transparency, openness, and collaboration, would be the foremost non-technical layer that aligns wholly with efforts of the State to decentralise and democratise AI and its ecosystem. A community presents a meaningful opportunity to bolster capacity and accelerate value creation, given that issues like skill gaps and talent retention, compounded by the effects of limited funding need to be addressed. While the Indian developer community exhibits an increased presence of AI skills, trends point to lower competencies.²¹ Active and vibrant communities offer constant and interactive support, creating spaces for deep engagement and robust learning.

Building communities involves a series of activities, all of which begin by unpacking the need and purpose for a community in the context of developing the solution. First, it is imperative that critical drivers of the community are identified. A stakeholder mapping exercise helps streamline efforts and resources in building communities by targeting interventions towards individuals who are significant contributors. This is followed by actively engaging with the community through the deployment of innovative measures to attract a diverse pool of contributors and sustain the growth of the community. Lastly, the needs and expectations of the community must be understood and met as reasonably as possible for the interest of its sustenance and evolution.

Reshaping Financing Practices within Open AI Ecosystems

While efforts by the government that seek to increase financing and public expenditure in this space are understood as imperative to accelerate democratisation, they need to be supplemented by streamlined funding practices that identify the universe of diverse actors and outline their newfound roles and responsibilities to facilitate decentralisation. Funding should be reinterpreted to involve capacity, expertise, and volunteering, and must include instruments beyond monetary funding.²²

While the need for public funding cannot be overstated, there must be restraints on where such funding can be utilised. Flows and sources of funding are often critical determinants of power dynamics in any landscape, and the AI ecosystem is no exception. It is only justified to include public actors, private entities, philanthropies, bilateral and multilateral agencies, CSOs, and non-profits, academia and users and leverage their potential to fund the open AI ecosystem in diverse ways. One can draw inspiration from understanding funding patterns in the Indian DPI landscape. For instance, in the case of ABDM,^a it was observed that initial public funding was crucial in kickstarting the ecosystem. However, to ensure the sustainability of the infrastructure, it was imperative to establish clear funding relationships between the government and private actors, helping identify the limitations on government funding and allowing private investment to flourish.²³

There is an immediate call to understand this landscape: breaking down funder types, identifying non-traditional funders, their funding potentialities and capabilities that serve the overarching needs of the ecosystem, as well as unpacking incentives and needs that will help sustain funding in the long run.

Funding types are not limited to financial funding, which typically includes grants from philanthropies or budget allocations by the State, but also make space for non-financial funding that must be increasingly

a The Ayushman Bharat Digital Mission is India's digital infrastructure responsible for health data collection, use, and transfer amongst healthcare stakeholders.

explored to support endeavours to build open AI ecosystems. For instance, academic capacity that cannot be replaced by financial resources to catalyse the creation of meaningful knowledge assets, which are unquantifiable. Spearheading efforts like this is the Centre for Responsible AI (CeRAI) at IIT Madras, a virtual interdisciplinary research centre that focuses on both fundamental and applied research towards the responsible development of AI. They have recently tied up with Swiss-based Roche Diagnostics Ltd. through an MoU that will look to expand research on digital health solutions. At the foundation of this partnership lies an opportunity to drive research, expand frameworks, and keep ethics in check. Both parties have stated the need to collaborate with academia to ensure equitable, responsible, and beneficial AI applications and use cases. Much like the above example, other experienced technical entities can volunteer their support and resources to aid the development of the ecosystem in India.²⁴

Conclusion

A socio-technical approach to building the ecosystem unpacks the intersection between the various technical layers and its non-technical components that helps to view the purpose of the open AI ecosystem beyond its technological utility. Understanding the underpinnings of governance and funding helps in assessing the impact of developing an open AI ecosystem from economic and social standpoints.

Exploring these non-technical layers enables a greater focus on inclusion, empowerment, and innovation that is rooted in lived experiences and localised realities. By embracing virtues such as transparency and collaboration embedded in these non-technical layers, India can shift from being a passive consumer of AI to a creator shaping solutions for critical sectors such as healthcare, education, finance, defence, and governance. These efforts to build non-technical layers, alongside democratising technical components of the open AI ecosystem, facilitate decentralisation and accelerate agency among all ecosystem players.

Aishani Rai is a Manager at Aapti Institute, with a background in law.

Gautam Misra is an associate researcher at Aapti Institute.

Endnotes

- 1 Yotta Data Services, “Yotta Launches myShakti: India’s First Sovereign B2C Gen AI Chatbot Utilizes DeepSeek Open-Source AI Model,” February 4, 2025, <https://yotta.com/media/yotta-launches-myshakti-indias-first-sovereign-b2c-gen-ai-chatbot-utilizes-deepseek-open-source-ai-model>.
- 2 Avani Airan et al., “The Governance of Digital Public Infrastructure,” Aapti Institute, June 2024, <https://aapti.in/wp-content/uploads/2024/06/AaptixONI-DPIGovernancePlaybook-compressed.pdf>.
- 3 Press Information Bureau, Government of India, <https://www.pib.gov.in/PressNoteDetails.aspx?NotelId=154912&ModuleId=3>
- 4 Beckn Foundation, “Home - Beckn Protocol,” <https://becknprotocol.io>.
- 5 YottaData Services, “Yotta Launches myShakti: India’s First Sovereign B2C Gen AI Chatbot Utilizes DeepSeek Open-Source AI Model,” <https://yotta.com/media/yotta-launches-myshakti-indias-first-sovereign-b2c-gen-ai-chatbot-utilizes-deepseek-open-source-ai-model/>
- 6 YottaData Services, “Yotta Launches myShakti: India’s First Sovereign B2C Gen AI Chatbot Utilizes DeepSeek Open-Source AI Model”
- 7 Kylie Robinson, “Character.AI Gave Up on AGI. Now It’s Selling Stories,” *Wired*, August 2025, <https://www.wired.com/story/character-ai-ceo-chatbots-entertainment/>
- 8 Max von Thun and Daniel A. Hanley, “Stopping Big Tech from Becoming Big AI: A Roadmap for Using Competition Policy to Keep Artificial Intelligence Open for All,” Mozilla Foundation, October 2024, <https://blog.mozilla.org/wp-content/blogs.dir/278/files/2024/10/Stopping-Big-Tech-h-from-Becoming-Big-AI.pdf>.
- 9 Anshika Mathews, “Old Employees, New Dollars – Google’s \$2.7 Billion Investment in Character. AI’s Reverse Acquire for AI Innovation!,” *AIM*, October 2024, <https://aimmediahouse.com/market-industry/old-employees-new-dollars-googles-2-7-billion-investment-in-character-ais-reverse-acquire-for-ai-innovation>
- 10 Pete Flint, “Startups vs Incumbents in the AI Era,” *NFX*, January 2024, <https://www.nfx.com/post/startups-vs-incumbents-ai>.
- 11 “The war for AI talent is heating up,” *The Economist*, June 8, 2024, <https://www.economist.com/business/2024/06/08/the-war-for-ai-talent-is-heating-up>
- 12 Max von Thun, “Monopoly Power Is the Elephant in the Room in the AI Debate,” *Tech Policy Press*, October 23, 2023, <https://www.techpolicy.press/monopoly-power-is-the-elephant-in-the-room-in-the-ai-debate>.
- 13 Thun, “Monopoly Power Is the Elephant in the Room in the AI Debate”
- 14 Joshua New, “Why We Must Protect an Open Innovation Ecosystem for AI,” *IBM Policy Lab*, December 5, 2023, <https://www.ibm.com/policy/why-we-must-protect-an-open-innovation-ecosystem-for-ai>.

- 15 Anulekha Nandi and Siddharth Yadav, "Digital Dreams, Real Challenges: Key Factors Driving India's AI Ecosystem," *ORF Occasional Paper No. 436*, Observer Research Foundation, May 2024, <https://www.orfonline.org/research/digital-dreams-real-challenges-key-factors-driving-indias-ai-ecosystem>.
- 16 Gautam Misra and Supratik Mitra, "Navigating AI Safety: A Socio-Technical and Risk-Based Approach to Policy Design," *Tech Policy Press*, December 19, 2024, <https://www.techpolicy.press/navigating-ai-safety-a-sociotechnical-and-riskbased-approach-to-policy-design>.
- 17 Misra and Mitra, "Navigating AI Safety: A Socio-Technical and Risk-Based Approach to Policy Design"
- 18 Pritam Bordoloi, "India Will Be an Open-Source AI Champion. Here's Why," *Analytics India Magazine*, March 13, 2024, <https://analyticsindiamag.com/ai-trends/why-will-india-champion-open-source-a>
- 19 Aishani Rai and Venkatesh Hariharan, "Building Communities Around Digital Public Goods: OSS4DPGs," Aapti Institute, June 15, 2024, <https://aapti.in/blog/building-communities-around-digital-public-goods-oss4dpg>
- 20 Rai and Hariharan, "Building Communities Around Digital Public Goods: OSS4DPGs."
- 21 Rai and Hariharan, "Building Communities Around Digital Public Goods: OSS4DPGs."
- 22 Nandi and Yadav, "Digital Dreams, Real Challenges: Key Factors Driving India's AI Ecosystem."
- 23 Ava Haidar and Dr. Sarayu Natarajan, "Financing Digital Public Infrastructures: A Playbook for Funders," Aapti Institute, June 2024, https://aapti.in/wp-content/uploads/2024/06/Financing-DPI-_compressed.pdf.
- 24 Press Release, "IIT Madras' Centre for Responsible AI teaming up with Switzerland-based Roche Diagnostics for R&D," IIT Madras, February 2025, <https://www.iitm.ac.in/happenings/press-releases-and-coverages/iit-madras-centre-responsible-ai-teaming-switzerland-based>

Towards a Responsible AI Framework for Open-Source AI Systems

*Kamesh Shekar and
Meemansa Agarwal*

As Artificial Intelligence (AI) becomes deeply embedded in governance, industry, and society, the question of how AI can be developed, deployed, and regulated responsibly and inclusively is increasingly critical. In this context, discourse on open AI holds rising importance.

There is growing interest in adopting open-source AI across industries. In a recent global survey of over 700 senior developers and leaders across 41 countries, 72 percent of respondents from tech organisations reported using an open-source AI model, with 63 percent across other sectors.¹ In 2023, 65.8

percent of foundation models were released as open access, while 18.8 percent remained entirely restricted, and 15.4 percent had limited availability.² This trend signals a strategic shift toward fostering greater innovation, competition, and decentralisation in AI.

Open AI systems promise to lower barriers to entry, enabling startups, researchers, and smaller enterprises to develop AI-driven solutions without requiring vast computational resources. They can thus help drive economic growth. This economic potential is underscored by a European Commission study, which found that an investment of around 1 billion euros in 2018 in open-source software generated an economic impact between 65 and 95 billion euros.³

Transparency is another key advantage of open foundation models. The 2023 Foundation Model Transparency Index⁴ shows that major open model developers are consistently and significantly more transparent than their closed counterparts, scoring an average of 20 percentage points higher. Open AI systems allow for third-party audits of datasets, training methodologies, and biases, which helps prevent the risks associated with opaque digital systems that have caused harm in the past, and is critical in sectors like finance, healthcare, and governance.⁵

For India, embracing open-source AI is not just an opportunity but a necessity, as the government champions digital public infrastructure (DPI) to foster innovation while ensuring equitable access to technology. Open-source AI systems, built on the principles of transparency, collaboration, and accessibility, align closely with these national priorities and have the potential to redefine India's AI landscape. Open models can help bridge critical gaps by offering shared, cost-effective resources that drive AI innovation, serving as viable alternatives to expensive proprietary systems. This is especially crucial for AI-driven agriculture, healthcare, and financial inclusion solutions, where local adaptation is key. Furthermore, an open AI ecosystem aligns with India's vision of technological self-reliance, reducing dependence on foreign technology while fostering homegrown innovation.

To fully unlock the potential of open AI, India must first establish a clear definition of what constitutes open-source AI. Alongside this, a robust governance framework is essential, one that upholds transparency and

accessibility while incorporating safeguards against misuse. Striking this balance will ensure that AI development remains ethical, inclusive, and accountable.

Defining ‘Open-Source AI’: A Spectrum of Openness

‘Open source’ is both a philosophy and a practice, built on the principles of transparency, collaboration, and shared innovation.⁶ It enables software engineers to share their work freely, contribute to each other’s projects, and collectively solve problems. As a result, open source has become a foundational pillar of the digital economy, as an estimated 70–90 percent of all software contains open-source code.⁷

In the context of AI, ‘open source’ refers to AI models that incorporate open-source components, such as software code, documentation, or training data, under licences that grant users the freedom to use, study, modify, and share them.⁸ However, there is no universally accepted definition of open-source AI, and its interpretation can vary depending on context and implementation.

Openness in AI exists along a spectrum. Just as open-source software licences range from highly permissive (e.g., MIT, Apache)⁹ to more restrictive (e.g., GNU GPL,¹⁰ BSD¹¹), AI models also vary in the degree of openness. The extent to which an AI model is considered open depends on how much of its underlying system is publicly accessible—whether that includes documentation, model architecture, weighting factors, or methods of use.¹²

Table 1. The Spectrum of Openness in AI Models

Degree of Openness	Attributes	Representative Model Examples
Full Openness	Source code and model weights are publicly available; clear documentation on training data, methodologies, and observed downstream impacts	<i>Mistral</i>
High Openness	Source code and model weights are accessible, but with limited or no transparency regarding training datasets or sources; partial visibility on downstream impacts	<i>Llama</i>
Partial openness	Source code is shared, but model weights are not; little to no clarity on training data or inputs	<i>Grok</i>
Limited Access	Code and weights are closed, but models are made available through public APIs or interfaces; some insights into training data and downstream implications are provided	<i>ChatGPT, Gemini, Claude</i>
Fully closed	Source code, model weights, and training data are proprietary; no external access or transparency into system inputs or outcomes	Proprietary enterprise models (e.g., internal systems used in logistics or operations)

Source: Adapted from 'New America, Openness in Artificial Intelligence Models'¹³

The diffusion strategies adopted by AI models influence their degree of openness. Some models provide limited interaction through hosted interfaces or cloud-based APIs. More open approaches include the release of downloadable model weights, enabling greater experimentation and adaptation. The most transparent models go further, offering not only weights but also source codes and training data, sometimes under usage restrictions and sometimes without, reflecting a commitment to broader accessibility and collaboration.¹⁴

Thus, the complexity of how AI systems are created, made available, and used is not adequately represented by the conventional binary classification of AI models as either ‘open’ or ‘closed’. In practice, most AI models fall somewhere between the two extremes, ranging from fully open systems, where all elements are publicly available, to selectively open or partially closed systems, where access is restricted by design or licence. Acknowledging this continuum is critical to developing meaningful and effective standards for open-source AI.¹⁵

Recognising that access to various components of AI systems, such as datasets, model weights, source code, and documentation, exists along a continuum allows for more context-sensitive governance and policy responses. This spectrum-based understanding helps stakeholders evaluate different degrees of accessibility, control, and risk. Such an approach enables more informed decisions regarding the use, collaboration, and oversight of AI systems, while also guiding the development of appropriate guardrails based on the degree of openness.

Unique Risks in the Open-Source AI Ecosystem

While open foundation models offer significant benefits, they also introduce distinct risks that differ from those associated with proprietary AI systems. The open nature of these models removes many of the barriers to access, allowing a broader range of actors, including those with malicious intent, to reproduce, modify, and deploy them with minimal oversight. This creates challenges in ensuring responsible AI development and use.¹⁷

Loss of Control

A fundamental risk of open-source AI is the developers’ loss of control once models are released. In proprietary systems, access is typically managed through APIs or other controlled interfaces, enabling developers to monitor usage, enforce safeguards, and respond to misuse. However, when models are made fully open, developers lose the ability to track or regulate their use. This means they cannot:

- block or restrict access to users engaging in harmful activities
- identify and respond to model misuse, including fine-tuning for illicit purposes
- adapt the model in real-time to prevent emerging threats, such as ‘jailbreak’ techniques that override built-in safeguards.

Once an open model is released, its proliferation becomes irreversible. Even if a newer, more secure version is introduced, earlier versions remain accessible on public repositories, alternative hosting platforms, or even the dark web.¹⁹ This persistence increases the difficulty of mitigating risks after the fact.

Manipulation and Misuse

The openness of generative AI models enables widespread innovation, but it also facilitates harmful applications. The risks include the following:

- **Removal of Safeguards:** AI model’s safety mechanism can be compromised in two primary ways. First, ethical safeguards embedded during development, such as restrictions on generating harmful or illegal content, can be deliberately removed or weakened by those with access to the model weights, enabling the generation of misinformation, extremist propaganda, or deepfakes, including pornography.²⁰

Second, even when safeguards remain in place, users can bypass them using techniques like adversarial suffixes or prompt injections, special character sequences, or cleverly crafted inputs that trick the model into generating prohibited outputs. Both approaches undermine safety-by-design and highlight the risks of releasing highly capable models without robust, enforceable guardrails.²¹

- **Fine-tuning for Malicious Purposes:** Open models can be fine-tuned with relatively little effort to generate harmful or illegal content, such as phishing emails, disinformation campaigns, or automated cyberattacks. Unlike proprietary systems, where developers can intervene, open models allow anyone with technical expertise to customise outputs freely.

Lack of Effective Enforcement Mechanisms

One of the biggest challenges in governing open-source AI is enforcement. Open-source AI operates without a centralised enforcement mechanism, making governance challenging. While developers may release models under ethical licenses, there is no practical mechanism to ensure that users comply. Unlike proprietary AI firms that can withdraw access or enforce terms of service, open-source developers lack the ability to impose restrictions on downstream users. This means:

- Malicious actors can still exploit unregulated versions of models (including pirated copies) or applications designed for harmful purposes, even after ethical concerns are raised.
- AI models with known vulnerabilities can remain in circulation indefinitely, exacerbating security risks.
- Developers have limited recourse to prevent misuse once a model is widely distributed.

Curating a Responsible AI Framework for Open AI Systems

At present, discussions on ‘ethical AI’ have centred around proprietary systems created by private companies.²² As a result, many of the design, policy, and tooling interventions aimed at promoting responsible AI are derived from studies of corporate AI ecosystems. These interventions may not be well-suited for open-source contexts, which operate under fundamentally different organisational structures, cultural norms, and incentive systems.

Moreover, emerging regulatory approaches risk disproportionately impacting open foundation models and their developers without effectively mitigating harm.²³ For instance, laws holding developers accountable for the content produced by their models or derivatives could unfairly penalise open-source contributors, as users can freely modify and fine-tune these models to generate illicit content.

Instead of reactive and overly restrictive regulations, a well-calibrated, responsible AI framework is needed, one that acknowledges the distinct challenges of open-source AI, balances innovation with risk mitigation, and ensures that policy interventions are grounded in evidence rather than speculation.

The Way Forward

The governance of open-source AI must strike a careful balance between upholding its core principles of collaboration, transparency, and shared innovation, and ensuring responsible development and use. The first step in developing a responsible AI framework for open-source systems is to establish clear metrics for assessing the openness of a system. Without a structured understanding of what constitutes openness, it becomes difficult to design governance mechanisms that are proportionate to the risks and opportunities posed by each model. Risk assessment must form the cornerstone of any open AI governance strategy. To address these issues, developers must adopt rigorous testing protocols to address the distinct vulnerabilities that open-source models introduce. These should include deterministic safety evaluations, probabilistic risk assessments, staged deployment models, and advanced red-teaming practices.^{24,25} Red-teamers, in particular, should be granted fine-tuning access before public release to help anticipate adversarial uses and surface potential vulnerabilities early.²⁶

Additionally, the safeguards established must be difficult to circumvent, ensuring that models cannot be easily modified for malicious purposes. A proactive approach, such as modifying training data to exclude harmful capabilities rather than relying solely on post-training safety filters, can further mitigate risks. Institutions like the proposed AI Safety Institute in India²⁷ can play a pivotal role in establishing standards while also enhancing capacity for independent third-party risk assessments.

Governance frameworks should also articulate clear model-sharing norms, established through standard-setting bodies and multistakeholder industry coalitions, that promote openness while accounting for risk. These standards must be sufficiently granular, allowing differentiated access based on the sensitivity of model components.

In addition, establishing an AI incident repository, as proposed by India's AI advisory group, will allow for timely and structured reporting of safety issues, help identify patterns, and improve collective response strategies.²⁸

To implement these governance recommendations effectively, it is necessary to strengthen institutional capacity within regulatory bodies

and open-source AI communities.²⁹ This includes building specialised regulatory expertise on AI risk assessment and ensuring that policymakers are equipped with the necessary technical knowledge to engage meaningfully in AI governance.

Beyond regulatory considerations, developing an ecosystem-wide understanding of the key players within the open-source AI landscape is crucial to fostering more effective and inclusive governance. Lessons from community-driven open innovation systems, such as decentralised software development models and collaborative security frameworks, can offer valuable insights into structuring AI governance. A distributed approach to governance, where oversight responsibilities are shared among maintainers, contributors, and rotating governance reviewers, can reduce reliance on centralised AI ethics teams and prevent bureaucratic bottlenecks. Additionally, self-regulation mechanisms, such as community-driven AI risk flagging, can complement formal oversight structures. Engaging the open-source community in regulatory discussions will help ensure that governance policies remain practical, enforceable, and aligned with real-world development practices.

Conclusion

Open-source AI offers immense value by fostering innovation, collaboration, and accessibility. However, as these technologies become more advanced, it is essential to comprehensively map their risks and develop governance frameworks that balance open-source principles and effective risk mitigation.

Importantly, it must be acknowledged that no model can offer perfect transparency or complete openness across all dimensions. Technical limitations, legal constraints, ethical considerations, and strategic interests will inevitably shape and limit the degree of access. Despite these variations, the priority must be to ensure that core AI capabilities are made available and used responsibly.

If the broader objective is to democratise access to AI and maximise its societal impact, then recognising and integrating the nuanced spectrum of openness into governance structures is not merely helpful; it is imperative. By adopting a structured and balanced approach to these complexities, it can be ensured that open-source AI continues to drive progress while remaining aligned with public interest and ethical responsibility.

Kamesh Shekar is Associate Director – Strategy & Research and Lead, Privacy and Data Governance Vertical & AI Vertical at The Dialogue.

Meemansa Agarwal is a Senior Research Associate at The Dialogue. Her work focuses on the regulation of Artificial Intelligence, with particular interest in legal and policy issues.

Endnotes

- 1 McKinsey Digital, “Open Source in the Age of AI,” February 2025, <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/open-source-in-the-age-of-ai>.
- 2 Nestor Maslej et al., *The AI Index 2024 Annual Report*, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, May 2024, <https://arxiv.org/abs/2405.19522>.
- 3 European Commission, *Commission Publishes Study: Impact of Open Source on the European Economy*, 2021, <https://digital-strategy.ec.europa.eu/en/news/commission-publishes-study-impact-open-source-european-economy>.
- 4 Stanford CRFM, “Foundation Model Transparency Index,” 2023, <https://arxiv.org/abs/2310.12941>.
- 5 Yaniv Benhamou, “Open Source AI – Definition and Selected Legal Challenges,” Kluwer Copyright Blog, April 15, 2024, <https://copyrightblog.kluweriplaw.com/2024/04/15/open-source-ai-definition-and-selected-legal-challenges/>.
- 6 Nick Botton and Mathias Vermeulen, *Generative AI’s Open Source Challenge: Policy Options to Balance Risks and Benefits of Openness in AI Regulations*, Digital Infrastructure Rights Fund, October 2024, [https://www.awo.agency/files/AWO-\(2024\)-Generative-AI-open-source-challenge-policy-options-to-balance-the-risks-and-benefits-of-openness-in-AI-regulation-041124.pdf](https://www.awo.agency/files/AWO-(2024)-Generative-AI-open-source-challenge-policy-options-to-balance-the-risks-and-benefits-of-openness-in-AI-regulation-041124.pdf).
- 7 Jason Perlow, “A Summary of Census II: Open Source Software Application Libraries the World Depends On,” Linux Foundation Blog, March 2022, <https://www.linuxfoundation.org/blog/blog/a-summary-of-census-ii-open-source-software-application-libraries-the->
- 8 Benhamou, “Open Source AI – Definition and Selected Legal Challenges”
- 9 Open Source Initiative, “The MIT License,” October 31, 2006, <https://opensource.org/license/mit>.
- 10 Open Source Initiative, “GNU General Public License Version 3,” October 23, 2007, <https://opensource.org/license/gpl-3-0>.
- 11 Open Source Initiative, “The 3-Clause BSD License,” <https://opensource.org/license/bsd-3-clause>.
- 12 Benhamou, “Open Source AI – Definition and Selected Legal Challenges”
- 13 Prem M. Trivedi and Nat Meysenburg, “Openness in Artificial Intelligence Models: A Key to Ensuring AI Serves Democratic Values and the Public Interest,” *New America*, 2024, <https://www.newamerica.org/oti/reports/openness-in-artificial-intelligence-models/the-spectrum-of-openness/>.
- 14 Rishi Bommasani et al., *Considerations for Governing Open Foundation Models*, Stanford HAI, 2023, <https://hai-production.s3.amazonaws.com/files/2023-12/Governing-Open-Foundation-Models.pdf>.

- 15 Elizabeth Seger et al., “Open-Sourcing Highly Capable Foundation Models: An Evaluation of Risks, Benefits, and Alternative Methods for Pursuing Open-Source Objectives,” Social Science Research Network, October 9, 2023, <https://doi.org/10.2139/ssrn.4596436>.
- 16 David Gray Widder et al., “Open (for Business): Big Tech, Concentrated Power, and the Political Economy of Open AI,” Social Science Research Network, August 17, 2023, <https://doi.org/10.2139/ssrn.4543807>.
- 17 Seger et al., “Open-Sourcing Highly Capable Foundation Models: An Evaluation of Risks, Benefits, and Alternative Methods for Pursuing Open-Source Objectives”
- 18 Seger et al., “Open-Sourcing Highly Capable Foundation Models: An Evaluation of Risks, Benefits, and Alternative Methods for Pursuing Open-Source Objectives”
- 19 Toby Shevlane, “Structured Access: An Emerging Paradigm for Safe AI Deployment,” arXiv, April 11, 2022, <https://doi.org/10.48550/arXiv.2201.05159>.
- 20 J. Rando et al., “Red-Teaming the Stable Diffusion Safety Filter,” arXiv, November 2022.
- 21 A. Zou et. al, *Universal and Transferable Adversarial Attacks on Aligned Language Models*, arXiv, July, 2023, <https://arxiv.org/abs/2307.15043>
- 22 David Gray Widder et al., “Limits and Possibilities for ‘Ethical AI’ in Open Source: A Study of Deepfakes,” in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’22, June 21–24, 2022, Seoul, Republic of Korea)* (New York: Association for Computing Machinery, 2022), <https://dl.acm.org/doi/pdf/10.1145/3531146.3533779>.
- 23 Rishi Bommasani et.al, “Considerations for Governance of Open Foundation Model,” *Science* 386, no. 6718, October 10, 2024, <https://www.science.org/doi/10.1126/science.adp1848>.
- 24 Nazneen Rajani, Nathan Lambert, and Lewis Tunstall, “Red-Teaming Large Language Models,” Hugging Face Blog, February 24, 2023, <https://huggingface.co/blog/red-teaming>.
- 25 Seger et.al, “Open-Sourcing Highly Capable Foundation Models: An Evaluation of Risks, Benefits, and Alternative Methods for Pursuing Open-source Objectives”
- 26 M. Anderljung et al., “Frontier AI Regulation: Managing Emerging Risks to Public Safety,” arXiv, September 2023, <https://arxiv.org/abs/2307.03718>.
- 27 “India Takes the Lead: Establishing the IndiaAI Safety Institute for Responsible AI Innovation,” IndiaAI Blog, 2025, <https://indiaai.gov.in/article/india-takes-the-lead-establishing-the-indiaai-safety-institute-for-responsible-ai-innovation>.
- 28 Ministry of Electronics and Information Technology, *Report on AI Governance Guidelines Development* (New Delhi: Ministry of Electronics and Information, January 2025, <https://indiaai.s3.ap-south-1.amazonaws.com/docs/subcommittee-report-dec26.pdf>.
- 29 Seger et al., “Open-Sourcing Highly Capable Foundation Models: An Evaluation of Risks, Benefits, and Alternative Methods for Pursuing Open-source Objectives”

Envisioning an Open AI Ecosystem in India

Sukriti

W

ith increasing emphasis on AI globally, open-source or open AI has gained traction in both national and global discourse for its potential to distribute power, accelerate innovation, and ensure transparency. In the Indian context, there appears to be a move toward the values of open AI, through a narrative of democratising AI for all and for public good, with increasing recognition of open datasets and models for driving indigenous AI innovation.

While calls for the democratisation of AI have become more common, the phrase itself may signify a variety of goals, vis-à-vis its use, development, profits, or governance.¹ As a result, reference to democratisation without clarity on the goals may lead to ambiguous commitments and missed opportunities for concrete policy discussions.² For instance, scholars point to the need to recognise the principal role of the democratisation of AI governance in navigating the trade-offs and risks involved in decisions around use, development, and profits.³ Additionally, the different forms of democratisation^a may at times conflict with each other. For example, to ensure responsible AI deployment, if the risk and consequences of misuse are likely to be serious, the democratisation of governance may require access restrictions. This may, in turn, hinder the democratisation of development, which necessitates a certain degree of AI model accessibility.⁴

The desirability of AI democratisation is often assumed. Yet, democratisation of use, development, and profits could hold both beneficial and harmful consequences. This means that the desirability of AI democratisation should instead be “derived from alignment with the interests and values of those who will be impacted.”⁵

Seeger et al. note that the use of ‘democratisation’ for AI is “normatively loaded”, but often used to refer to “facilitating widespread AI use and development.”⁶ They argue, “invoking the term ‘democratisation’ [...] holds the hidden assumption that the decision to distribute or make accessible is what a democratic governance process would select. In other words, AI democratisation ultimately refers to the democratisation of AI governance. If by ‘AI democratisation’ all a speaker means is ‘make available to everyone’, then we would suggest less normatively loaded language (something like ‘broad accessibility’) be used.”⁷

a Here, this article relies on Seeger et al.’s (2023) identification of four kinds of ‘AI democratisation’: (1) “the democratisation of AI use”, (2) “the democratisation of AI development”, (3) “the democratisation of AI profits”, and (4) “the democratisation of AI governance”.

It is worth noting that while ‘responsible AI’, ‘open source AI’, or ‘democratic AI’ are often used alongside each other, and were “invented to serve the public interest,” they have been argued to “under-specify critical aspects of the development process which determine whether a technology serves the interests of the many or the few.”⁸ As a result, they are vulnerable to being co-opted by profit-driven corporations and giving way to concentrated market power.⁹

More broadly, efforts for democratising AI use involve an array of other measures such as reduction in costs of acquiring and running AI models, development of user interfaces that are accessible across different sections of the population, or improvement in internet infrastructure and accessibility.¹⁰ Democratisation efforts also need to account for the structural barriers, such as the fact that a small number of tech companies concentrated mainly in the global minority regions hold a lot of control over the resources necessary for AI development.

Such disparities in AI development, use, and profit distribution have led to open-source AI or open AI as an attractive way forward for democratisation efforts in AI. Openness enables expansion of access to the broader community of developers and academia, and allows them to scrutinise, reuse, and build upon these systems.¹¹ It further enables auditing and oversight to prevent privacy and security risks, and mitigate and reduce bias.¹² Open AI is viewed as an important means to reduce the concentration of power in a few technology corporations and regions (largely located in the Global North) in AI development and innovation.¹³

In India, there are several parallel and fragmented initiatives furthering AI development, education, research, and skilling. There appears to be a move toward the values of open AI through a narrative of democratising AI for all and for public good. However, it does not situate this within the broader critical discourse of what it means to achieve an open AI ecosystem or by referencing ‘democratisation of AI’ in this particular context. This would require an understanding of the goals and rationale of open AI contextualised to India, which go beyond ‘innovation’. This should further include working with a clear understanding of ‘democratisation’ that overcomes the entrenched structural and institutional power imbalances primarily owing to the

unequal distribution of resources in the development of technologies. This article will utilise existing international discourses on democratisation of and openness in AI to identify the gaps and challenges that need consideration for creating an open AI ecosystem in India.

What Is ‘Open AI’?

Openness in the context of AI is difficult to define, given the complexity of an AI system. Unlike traditional software, which is programmed with rules to perform a task, AI is programmed to *learn* to perform a task.¹⁴ AI has three broad components that enable this learning: training datasets, source codes to formalise the training task, and models to store the trained weights.¹⁵ Given its different interconnected components, ‘openness’ in AI can be present along a gradient,¹⁶ ranging from systems that offer minimal transparency and reusability—where only one of the components is open for scrutiny or reuse—to systems that are maximally transparent, reusable, and extendable.

Broadly, ‘open’ or ‘open source’ in the context of AI is used to refer to a range of capabilities that offer:

- transparency (the ability to access and vet source code, documentation, and data)
- reusability (the ability and licensing needed to allow third parties to reuse source code and/or data)
- extensibility (the ability to build on top of extant off-the-shelf models, ‘tuning’ them for one or another specific purpose).¹⁷

There are several contentious issues in terms of what components of the AI system may be made open and to what extent. For instance, “whether potentially sensitive training data should be made open, or whether offering powerful models for reuse is safe, or whether restricting the use of ‘open’ models to ‘safe’ or ‘ethical’ domains is acceptable under the banner of ‘open source,’ or not.”¹⁸ The challenges associated with determining openness for AI systems defy the traditional understanding of open source,¹⁹ which was conceived at the time to apply to software. Thus, traditional notions and assumptions associated with open-source do not translate to AI. After several rounds of revision

and feedback from developers and researchers, the Open Source Initiative has released its definition of open AI:

“An Open Source AI is an AI system made available under terms and in a way that grant the freedoms to:

- Use the system for any purpose and without having to ask for permission.
- Study how the system works and inspect its components.
- Modify the system for any purpose, including to change its output.
- Share the system for others to use with or without modifications, for any purpose.”²⁰

In regulatory and governance discourses, including those in India,²¹ open AI is being viewed as a way to disperse the benefits of AI for social good (such as with Digital Public Goods for Sustainable Development Goals)²² through a democratisation of the resources and access required to build and utilise AI systems. As will be discussed in the subsequent section, while often argued as a means to democratise AI development, open AI in itself does not enable democratisation, as access to the necessary resources required to build AI itself is associated with several structural barriers.

The Current AI Landscape in India

The Government of India launched the IndiaAI Mission in March 2024 under the public-private partnership (PPP) model for furthering India’s AI innovation ecosystem.²³ The mission aims to advance India’s AI innovation and “democratise the benefits of AI” across different sectors towards public good.²⁴ The primary objectives are: building a scalable AI ecosystem with increased compute capacity; developing, scaling, and promoting large-scale AI solutions for socio-economic transformation, and sectoral applications, such as governance, healthcare, and agriculture; creating a unified platform for access to high-quality non-personal datasets; expanding skilling and education;²⁵ startup financing; and ensuring responsible AI development and use by promoting fair, transparent, and ethical AI practices.²⁶

The Ministry of Electronics and Information Technology (MeitY) has also developed other collaborative initiatives such as Bhashini. Launched in July 2022 as part of the National Language Translation Mission, Bhashini aims to provide citizens with access to digital services and increase in content in different Indian languages. Bhashini is being developed as 'large open source datasets and models'.²⁷ It aims to create an ecosystem for developing products and services in Indian languages by utilising Bhashini's open repository of datasets and models. Additionally, Meta signed an MoU with the IndiaAI Mission to make Meta's open-source AI model available for use by the Indian ecosystem.²⁸

Other multi-stakeholder initiatives such as AI4Bharat at IIT Madras (co-sponsored by the Rohini Nilekani Foundation, Google, Microsoft, Yotta, and EkStep) have been working for "advancing AI technology for Indian languages through open-source contributions."²⁹ Similarly, Microsoft helped develop a generative AI chatbot called 'Jugalbandi' for access to information on government schemes, facilities, and procedures.³⁰ It utilised the LLMs developed by the AI4Bharat Initiative, combined with Microsoft Azure's Open AI Service, along with Bhashini. EkStep, a non-profit foundation co-founded by Nandan Nilekani (chairman and co-founder of Infosys), is also backing People+AI, which has been working to support the AI policy ecosystem by developing sectoral applications such as public service delivery, health, education and environment, enhancing communication accessibility for the hearing-impaired, and other open-source solutions.³¹

The national landscape is being heavily driven by the PPP model, with private players having the bigger say in how the AI landscape is shaped.³² For instance, organisations such as EkStep foundation are associated with technological projects that bear societal implications, and are making swift developments in welfare delivery mechanisms.³³ Given this landscape, the vision and goals of these private-driven endeavours become drivers of overall AI missions, and announcements by powerful organisations have a disproportionate influence on AI technological and policy discourse in India.

Further, there seem to be multiple government endeavours and initiatives for AI integration into government facilities, driven by private or non-governmental firms.³⁴ However, there is little information in the public domain about the execution of such projects. Given that the government has made critical investments in projects deploying AI in various crucial sectors for development,³⁵ it is important to have conversations around establishing accountability mechanisms and safeguards in AI systems. While the government has called for proposals towards building responsible AI, there is not enough foundational discourse to arrive at a common ground on what constitutes responsible, ethical and democratic development of open AI technologies.

What does it mean for an AI system to be open, and why does it matter?

Analysis and Challenges

With AI becoming crucial for countries to assert their geopolitical competitiveness, much of the push for AI has found itself cloaked under the banners of ‘democratisation’ and ‘social good’.³⁶ Open AI has therefore found a convenient spot in this narrative.³⁷ The current thrust on AI development in India and globally might chase the virtues of open AI, but in reality, it may be at odds with those virtues. The development of open AI continues to operate very strongly within the same structures of reliance created by Big Tech and the powerful economies that open AI appears to resist.³⁸ Resultantly, the development of an open AI ecosystem faces barriers, as discussed in the following paragraphs. Any discourse for enabling open AI should move forward with an understanding of these challenges.

a. Power Imbalances

The development of AI models requires large amounts of computational power for data processing. However, the computational resources needed for such a purpose are scarce, expensive, and provided by very few corporations.³⁹ This also poses a barrier to the reusability of existing open AI systems as training and running inferences on large-scale AI models at scale, and utilising them in a product or API for wider public use, is also extremely costly. Much of the production of

the resources required for AI development, such as cloud computing, chips, and microprocessing units, takes place in the minority world countries.⁴⁰ With most tech corporations originating in the Global North, such a power imbalance additionally disadvantages countries belonging to the Global South.

b. Obtaining Data and Its Challenges

While there has been increased focus on gathering inclusive, representative data and ensuring the appropriate documentation of data gathering and annotation processes,⁴¹ questions regarding privacy, power imbalances, and ethics remain unresolved.

Data collection and processing for AI development is labour-intensive.⁴² Given that most AI systems are built by companies located in the minority world, these models are built by exploiting the data and labour resources from the majority world. Therefore, for countries belonging to the majority world, notions of data sovereignty also become significant. For instance, the Te Hiku Media project recognises that “indigenous peoples may not have access to the resources that enable them to benefit from open source technology,”⁴³ and developed a Kaitiakitanga License as a way to enable guardianship of the data by the people from whom it originates.

Datasets may “launder others’ intellectual property, or commercially use data that was specifically licensed for non-commercial use, or was licensed under particular sovereignty mandates.”⁴⁴ Promoting the transparency of datasets remains key to the development of an open AI ecosystem to prevent value extraction and exploitation of proprietary data scraped from the internet.⁴⁵ Enabling open AI development would entail meaningful documentation and access to the training data for review. Many problems associated with fairness, accountability, transparency, and ethics in ML systems result from the decisions around the data collection and annotation process.⁴⁶ India should ensure that the development of open AI is based on principles of meaningful access while encouraging multidisciplinary research, studying and reviewing open AI development both in India and globally, and understanding challenges unique to the Indian social, cultural, linguistic, and economic context.

While maintaining and building datasets is crucial to an open AI ecosystem, it should not be removed from privacy and data protection considerations.⁴⁷ There are tensions between the need for vast amounts of data and the right to privacy that enables individuals to maintain control over their data.⁴⁸ India's longstanding deliberations on a legal framework for personal data protection culminated in the Digital Personal Data Protection Act, 2023 (DPDP Act) and the Rules, 2025.^b The Act and Rules provide for a notice and consent framework for the collection and processing of data. Ensuring transparency and collaboration in open AI development would require deliberation on responsible safeguards and privacy measures for the sharing of data while ensuring data protection towards ethical AI development. Data fiduciaries^c using data for training AI will need to develop policies in compliance with the DPDP Act's provisions such as outlining permitted uses and potential future uses of data, mechanisms for obtaining and withdrawing consent, specification of restricted uses and detailing the data access, collection, retention and sharing mechanisms (including cross-jurisdictional), along with action in cases of violation of provisions of the law.

c. Investigating the Benefits of Open AI

The current open AI discourse assumes that the underlying notions associated with the idea of 'open'⁴⁹ will inevitably result in benefits such as innovation, safety, transparency, and democratised access, and reduce dependence on dominant infrastructure and industry concentration. However, while these benefits are worth pursuing, they cannot be properly achieved without a clear understanding of what 'open' AI means and necessitates for its development, resources, as well as labour and maintenance, to truly democratise AI development and access, along with its limitations. For instance, even as open-

c Digital Personal Data Protection Act, 2023; Draft Digital Personal Data Protection Rules, 2025, <https://innovateindia.mygov.in/dpdp-rules-2025/>

d Section 2(i) of the DPDP Act defines "data fiduciary" to mean "any person who alone or in conjunction with other persons determines the purpose and means of processing of personal data."

source systems are assumed to enable safer AI through review by developers and researchers, open AI in itself does not guarantee this simply by being open.⁵⁰ To ensure this goal, any discourse on open source AI should also take into account the need for mandating auditing mechanisms that ensure maintenance and inspection, and consider where the responsibilities would lie if vulnerabilities and gaps were left unaddressed.

One should consider whether the investments being made for strengthening computing and data infrastructure are in alignment with the broader aim of democratising AI in a way that reduces corporate influence over and control of the AI ecosystem.⁵¹

Any deliberation on the implementation and deployment of AI raises normative questions such as what are the acceptable risk thresholds, what does justice and fairness mean in the context of AI applications, and what are the values that should underpin AI.⁵² These questions remain relevant for any discourse considering building an open AI ecosystem—despite its many benefits, open AI may be susceptible to malicious use, contributions, and modifications by certain actors. A democratic discourse among a diverse range of stakeholders can identify not only the areas of consensus but also the finer context-specific details of interpretation and implementation questions at local levels.⁵³

Recommendations and Conclusion

To achieve a democratised, open AI ecosystem, it is important that “AI research and resources are created and maintained by institutions meaningfully separate from large tech companies.”⁵⁴ Additionally, to counteract the discriminatory outputs that arise from large-scale AI models privately developed in and exported from the minority world,⁵⁵ investment into research could also take the form of funding scholarships and research fellowships for sociotechnical studies by people belonging to underrepresented groups and marginalised communities.⁵⁶ The following discussion focuses on two more immediate and actionable pathways for a move towards an open AI ecosystem, keeping in mind the current AI landscape in India.

Data Sharing, Collection, and Handling: Building Appropriate Standards and Frameworks

Given that implementing ethical forms of data collection and handling requires time, expertise, and resources, it serves to disadvantage institutions without adequate resources. One of the recommended measures to reduce this inequality is to create consortial arrangements for data sharing.⁵⁷ Another measure to address the lack of transparency in data collection is to have structured processes and institute record-keeping standards.⁵⁸ However, since data sharing can also lead to risks such as breach of privacy, data misrepresentation, and data collection without consent, data sharing arrangements should be made only after the formulation of a governance framework for ethical data sharing practices. Further, to prevent misuse of open data, there should be appropriate measures in place for access control, and ethics training for researchers and developers for the responsible use of data, especially sensitive data. The lack of well-annotated, feature-rich, local datasets, as is the case in India, further highlights the need for establishing appropriate standards, open-source licensing frameworks, formats, governance, and institutional frameworks for the collection, access, and management of data.⁵⁹

Accounting for Socio-cultural Values and Impacts: Enabling Public Participation in AI Development Processes

Expansion of access cannot be achieved by “going open” alone. Considering that many efforts in India towards openness in AI are geared towards public services, governance, and other sectors of public interest and welfare, it might be useful to consider public participation in AI development. Many of the barriers to access, innovation, and development result from geographical, cultural, linguistic, and economic divides.⁶⁰ Public participation may help break these barriers to some extent and enable the identification of and reduction in harmful, inaccurate, biased, or discriminatory outcomes of AI systems.⁶¹ Gilman defines public participation as “measures that offer opportunities for people most likely to be affected by a given system to have influence into the system’s design and deployment, including decision-

making power.”⁶² Public participation can help counter harmful impacts by enabling the inclusion of social-cultural values in technological development beyond technical parameters. Public participation can be built into various stages of AI development, i.e., problem formulation, data collection, model selection and training, deployment, and monitoring.⁶³ Public participation should ensure diversity among participants and the identification of communities and groups that will be most affected by the deployment of AI or have accessibility or other barriers to benefit from the deployment.⁶⁴

Public participation could be legally mandated to ensure “continuity within organisations over time, help build institutional norms around participation, and increase trust with affected communities.”⁶⁵ Such legal mandates should, among other things, include within themselves a definition of meaningful participation and its goals, specific stages of AI development for public participation, require culturally appropriate community outreach, and access for marginalised communities.⁶⁶

Additionally, synergies and networks should be built between universities, research centres, government, and industry as a way to create new ecosystems around open AI for driving innovation.⁶⁷ Such an ecosystem will also benefit startups and small and medium-sized companies in accessing dedicated infrastructure and capacities for AI development.⁶⁸

With resource and power imbalances influencing AI development, private resources and capabilities might be necessary in its nascent phases. However, this should not result in further entrenchment of the power imbalance as opposed to reducing inequalities. Therefore, India should consider what it means for the development of open AI systems outside of powerful private enterprise support and how it can transition to indigenous capabilities while reducing the present structures of dependency. This cannot be done without the development of governance and institutional frameworks for a truly democratic and open AI ecosystem.

A focus on reducing dependence on large corporate players will also enable an ecosystem that reduces the barriers of entry and promotes the sustainability of small-scale AI providers and open-source models. Smaller-scale models can also lead to innovation in trustworthy and safe design principles in their development.⁶⁹ More broadly, however, a focused approach with a clear rationale and goal for developing open AI systems requires a clear means of evaluating investment, resource use, and their downstream impact.

Sukriti is a Project Officer at the Centre for Communication Governance, National Law University Delhi.

Endnotes

- 1 Elizabeth Seger et al., “Democratising AI: Multiple Meanings, Goals and Methods,” arxiv, <https://arxiv.org/abs/2303.12642>.
- 2 Seger et al., “Democratising AI: Multiple Meanings, Goals and Methods”
- 3 Seger et al., “Democratising AI: Multiple Meanings, Goals and Methods”
- 4 Seger et al., “Democratising AI: Multiple Meanings, Goals and Methods”
- 5 Seger et al., “Democratising AI: Multiple Meanings, Goals and Methods”
- 6 Seger et al., “Democratising AI: Multiple Meanings, Goals and Methods”
- 7 Seger et al., “Democratising AI: Multiple Meanings, Goals and Methods”
- 8 Public AI Network, “Public AI: Infrastructure for the Common Good,” August 10, 2024, <https://publicai.network/whitepaper>
- 9 Public AI Network, “Public AI: Infrastructure for the Common Good”
- 10 Seger et al., “Democratising AI: Multiple Meanings, Goals and Methods”
- 11 David Gray Widder et al., “Why ‘Open’ AI Systems Are Actually Closed, and Why this Matters,” *Nature* 635 (2024): 827.
- 12 Alex Engler, “How Open-source Software Shapes AI Policy,” Brookings, August 10, 2021, <https://www.brookings.edu/articles/how-open-source-software-shapes-ai-policy/>
- 13 Mike Sexton, “What is Open AI?,” Third Way, <https://www.thirdway.org/memo/what-is-open-ai>.
- 14 Yash Raj Shrestha, Georg von Krogh and Stefan Feuerriegel, “Building Open Source AI,” *Nature Computational Science* 3 (2023): 908.
- 15 Shrestha et al., “Building Open Source AI”
- 16 Irene Solaiman, “The Gradient of Generative AI Release: Methods and Considerations,” arxiv, <https://arxiv.org/abs/2302.04844>.
- 17 Gray Widder et al., “Why ‘open’ AI Systems are Actually Closed, and Why This Matters,”; David Gray Widder, Sarah West, and Meredith Whittaker, “Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI,” *Nature*, August 18, 2023, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4543807.
- 18 Gray Widder, West, and Whittaker, “Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI”
- 19 Shrestha et al., “Building Open Source AI”
- 20 “The Open Source AI Definition – 1.0,” Open Source Initiative, <https://opensource.org/ai/open-source-ai-definition>.

- 21 Jyoti Panday and Mila T Samdub, “Promises and Pitfalls of India’s AI Industrial Policy,” in *AI Nationalism(s): Global Industrial Policy Approaches to AI*, eds. Amba Kak and Sarah Myers West (AI Now, 2024), pp. 85; Ministry of External Affairs, “Opening Address by Prime Minister Shri. Narendra Modi at the AI Action Summit, Paris (February 11, 2025),” February 11, 2025, https://www.mea.gov.in/Speeches-Statements.htm?dtl/39020/Opening_Address_by_Prime_Minister_Shri_Narendra_Modi_at_the_AI_Action_Summit_Paris_February_11_2025#:~:text=We%20are%20at%20the%20dawn,of%20responsibility%20must%20guide%20us.
- 22 “Digital Public Goods,” Digital Public Goods Alliance, <https://www.digitalpublicgoods.net/digital-public-goods>
- 23 Ministry of Electronics and Information Technology, Government of India, <https://pib.gov.in/PressReleasePage.aspx?PRID=2086605>; Ministry of Electronics and Information Technology, Government of India, <https://pib.gov.in/PressReleaselframePage.aspx?PRID=2012375#:~:text=%E2%80%9CIndiaAI%20Financial%20Outlay%20to%20Benefit,Startups%E2%80%9D%3A%20Union%20Minister%20Rajeev%20Chandrasekhar&text=The%20Cabinet%20has%20approved%20an,towards%20bolstering%20India’s%20AI%20ecosystem.>
- 24 Ministry of Electronics and Information Technology, Government of India, <https://pib.gov.in/PressReleasePage.aspx?PRID=2086605>.
- 25 Ministry of Electronics and Information Technology, Government of India, <https://pib.gov.in/PressReleasePage.aspx?PRID=2068251>; “Microsoft Partners With Govt’s India AI Mission To Skill 5 Lakh People By 2026,” *Economic Times*, January 9, 2025, <https://economictimes.indiatimes.com/tech/technology/microsoft-indiaai-join-hands-to-train-5-lakh-individuals-on-ai-by-2026-set-up-productivity-labs/articleshow/117070029.cms?from=mdr>; Ministry of Skill Development and Entrepreneurship, Government of India, <https://pib.gov.in/PressReleaselframePage.aspx?PRID=2064628>.
- 26 IndiaAI Mission, <https://indiaai.gov.in/>.
- 27 Bhashini, <https://bhashini.gov.in/about-bhashini>.
- 28 Nidhi Singal, “IBM, MeitY sign MoUs to Advance Innovation in AI, Semiconductor and Quantum Innovation in India,” *Business Today*, October 18, 2023, <https://www.businesstoday.in/latest/in-focus/story/ibm-meity-sign-mous-to-advance-innovation-in-ai-semiconductor-and-quantum-innovation-in-india-402508-2023-10-18>.
- 29 AI4Bharat, <https://ai4bharat.iitm.ac.in/>.
- 30 Chen May Yee, “With Help from Next-generation AI, Indian Villagers Gain Easier Access to Government Services,” *Microsoft*, May 23, 2023, <https://news.microsoft.com/source/asia/features/with-help-from-next-generation-ai-indian-villagers-gain-easier-access-to-government-services/>.
- 31 People+ai: Projects, <https://coda.io/@peopleplusai/people-ai-projects>.

- 32 “Indian startup Sarvam AI collaborates with Microsoft to bring its Indic voice large language model (LLM) to Azure,” *Microsoft*, February 8, 2024, <https://news.microsoft.com/en-in/indian-startup-sarvam-ai-collaborates-with-microsoft-to-bring-its-indic-voice-large-language-model-llm-to-azure/>; “India AI stack: Why Sam Altman’s Visit to India Matters,” *ETOnline*, February 5, 2025, <https://economictimes.indiatimes.com/industry/cons-products/electronics/india-ai-stack-why-sam-altmans-visit-to-india-matters/articleshow/117955027.cms?from=mdr>; “Union Minister Ashwini Vaishnaw Meets OpenAI’s Sam Altman, Discusses Collaboration,” *The Hindu*, February 5, 2025, <https://www.thehindu.com/sci-tech/technology/union-minister-ashwini-vaishnaw-meets-openais-sam-altman-discusses-collaboration/article69183215.ece>; “OpenAI Partners with Defence Firm Anduril to Use AI in National Security Missions,” *The Hindu*, December 5, 2024, <https://www.thehindu.com/sci-tech/technology/openai-partners-with-defence-firm-anduril-to-use-ai-in-national-security-missions/article68949356.ece>; “Google Cloud, EkStep Foundation Partner to Accelerate Adoptions of Digital Public Infrastructure,” *Economic Times*, May 30, 2024, <https://economictimes.indiatimes.com/tech/technology/google-cloud-ekstep-foundation-partner-to-accelerate-adoptions-of-digital-public-infrastructure/articleshow/110564514.cms?from=mdr>; Venkatesh Kannaiah, “India Should be the Use Case Capital for AI in the World: EkStep Foundation Co-founder Shankar Maruwada,” *Indian Express*, May 26, 2024, <https://indianexpress.com/article/technology/tech-news-technology/ekstep-foundation-shankar-maruwada-interview-ai-9351382/>; Peerzada Abrar, “Nandan Nilekani-backed people+ai Partners with 24 Tech Organisations,” *Business Standard*, May 7, 2024, https://www.business-standard.com/companies/start-ups/nandan-nilekani-backed-people-ai-partners-with-24-tech-organisations-124050701495_1.html.
- 33 Sarasvati NT, “10 Points Highlighting Nandan Nilekani’s AI Blueprint: DPI, Open Source Models, Use-Case Approach and More,” *MediaNama*, December 13, 2023, <https://www.medianama.com/2023/12/223-10-points-nandan-nilekani-ai-blueprint/>.
- 34 Sarasvati NT, “Medianama: Jugalbandi, A chatbot for Rural India by Microsoft and EkStep: What to Know and Think About?,” *MediaNama*, June 3, 2023, <https://www.medianama.com/2023/06/223-jugalbandi-chatbot-rural-india-what-to-know/>.
- 35 Vignesh Radhakrishnan et al., “Union Budget 2025: Artificial Intelligence Related Schemes Receive Significant Increase in Allocations,” *The Hindu*, February 1, 2025, <https://www.thehindu.com/data/union-budget-2025-artificial-intelligence-related-schemes-receive-significant-increase-in-allocations/article69163704.ece>.
- 36 AINOW, “AI Nationalisms: Global Industrial Policy Approaches to AI,” <https://ainowinstitute.org/publications/research/ai-nationalisms-global-industrial-policy-approaches-to-ai>.
- 37 Gray et al., “Why ‘Open’ AI Systems Are Actually Closed, and Why This Matters”
- 38 AINOW, “AI Nationalisms: Global Industrial Policy Approaches to AI,”
- 39 Gray et al., “Why ‘Open’ AI Systems Are Actually Closed, and Why This Matters”
- 40 Anulekha Nandi and Siddharth Yadav, “Digital Dreams, Real Challenges: Key Factors Driving India’s AI Ecosystem,” *ORF Occasional Paper No. 436*, Observer Research Foundation, May 2024, <https://www.orfonline.org/research/digital-dreams-real-challenges-key-factors-driving-indias-ai-ecosystem>.

- 41 Eun Seo Jo and Timnit Gebru, “Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning” (paper presented at the Conference on Fairness, Accountability, and Transparency (FAT* ’20), Barcelona, Spain, January 27–30, 2020).
- 42 Gray et al., “Why ‘Open’ AI Systems Are Actually Closed, and Why this Matters”
- 43 Te Hiku Media, “He Reo Tuku Iho, He Reo Ora: Living Language Transmitted Intergenerationally,” *MAI Journal: A New Zealand Journal of Indigenous Scholarship* 11, no. 1, (2022): 41, 45.
- 44 Gray et al., “Why ‘Open’ AI Systems Are Actually Closed, and Why this Matters”
- 45 Zuzanna Warso and Paul Keller, “Open Source, AI and the Paradox of Open,” *Open Future*, September 15, 2023, <https://openfuture.eu/blog/open-source-ai-and-the-paradox-of-open/>.
- 46 Seo Jo and Gebru, “Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning”
- 47 Bilal Mohamed, “Five Ways in Which the DPDPA Could Shape the Development of AI in India,” *Future of Privacy Forum*, September 6, 2024, <https://fpf.org/blog/five-ways-in-which-the-dpdpa-could-shape-the-development-of-ai-in-india/>.
- 48 Nick Vidal, “Reimagining Data for Open Source AI: A Call to Action,” *Open Source Initiative*, January 23, 2025, <https://opensource.org/blog/reimagining-data-for-open-source-ai-a-call-to-action>.
- 49 Fatih Bildirici, “Open Source AI: An Approach to Responsible AI Development,” *Reflektif Journal of Social Sciences* 5, no. 1 (2024): 73.
- 50 Gray et al., “Why ‘open’ AI Systems Are Actually Closed, and Why This Matters”
- 51 AI Now Institute and Data & Society Research Institute, *Democratise AI? How the Proposed National AI Research Resource Falls Short*, AI Now, October 5, 2021, <https://ainowinstitute.org/publication/democratize-ai-how-the-proposed-national-ai-research-resource-falls-short#:~:text=The%20NAIRR%20and%20related%20proposals,until%20these%20challenges%20are%20resolved;Amba%20Kak,Sarah%20Myers%20West,and%20Meredith%20Whittaker,Make%20No%20Mistake%20AI%20is%20Owned%20by%20Big%20Tech,MIT%20Technology%20Review,December%205,%202023,https://www.technologyreview.com/2023/12/05/1084393/make-no-mistake-ai-is-owned-by-big-tech/>.
- 52 Seger et al., “Democratising AI: Multiple Meanings, Goals and Methods”
- 53 Seger et al., “Democratising AI: Multiple Meanings, Goals and Methods”
- 54 Seger et al., “Democratising AI: Multiple Meanings, Goals and Methods”
- 55 Grace Browne, “AI Is Steeped in Big Tech’s ‘Digital Colonialism,’” *Wired*, May 25, 2023, <https://www.wired.com/story/abeba-birhane-ai-datasets/>.
- 56 AI Now and Data & Society, “Request for Information (RFI) on an Implementation Plan for a National Artificial Intelligence Research Resource,” October 1, 2021, <https://ainowinstitute.org/wp-content/uploads/2023/06/AINow-DS-NAIRR-comment.pdf>.
- 57 Seo Jo and Gebru, “Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning”
- 58 Seo Jo and Gebru, “Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning”

- 59 Nandi and Yadav, "Digital Dreams, Real Challenges: Key Factors Driving India's AI Ecosystem"
- 60 Seger et al., "Democratising AI: Multiple Meanings, Goals and Methods"
- 61 Michele Gilman, *Democratizing AI: Principles for Meaningful Public Participation*, Data and Society, September 27, 2023, <https://datasociety.net/library/democratizing-ai-principles-for-meaningful-public-participation/>.
- 62 "Democratizing AI: Principles for Meaningful Public Participation"
- 63 "Democratizing AI: Principles for Meaningful Public Participation"
- 64 "Democratizing AI: Principles for Meaningful Public Participation"
- 65 "Democratizing AI: Principles for Meaningful Public Participation"
- 66 "Democratizing AI: Principles for Meaningful Public Participation"
- 67 Shrestha et al., "Building Open Source AI"
- 68 Shrestha et al., "Building Open Source AI"
- 69 Derek Slater and Betsy Masiello, "Will Open Source AI Shift Power from 'Big Tech'? It Depends," Tech Policy Press, June 16, 2023, <https://www.techpolicy.press/will-open-source-ai-shift-power-from-big-tech-it-depends/>.
- 70 AI Now and Data & Society, "Request for Information (RFI) on an Implementation Plan for a National Artificial Intelligence Research Resource"

Acknowledgements

The early versions of the articles contained in this publication were presented at a workshop organised by the Observer Research Foundation (ORF) in Kolkata in February 2025.

The editors thank Debajyoti Chakravarty, Research Assistant, Centre for Digital Societies, ORF for supporting the management of the workshop, chairing a technical session, and offering editorial support. The editors also thank Anulekha Nandi, former Fellow at ORF, for chairing a session at the workshop; and Tanusha Tyagi, Research Assistant, Centre for Digital Societies, ORF for providing editorial support.



20, Rouse Avenue Institutional Area
New Delhi - 110 002, India
+91-11-35332000 Fax: +91-11-35332005
contactus@orfonline.org
www.orfonline.org