Advancing Open Source AI in India:

Recommendations for Governments & Technology Developers

This project is a collaborative effort between Digital Futures Lab, NASSCOM, and the global initiative FAIR Forward – Artificial Intelligence for All. FAIR Forward is implemented by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) on behalf of the Federal Ministry for Economic Cooperation and Development (BMZ).

Jointly published with INDIAai Mission

Published by the
Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn, Germany
FAIR Forward – Artificial Intelligence for All
A-2/18 Safdarjung Enclave Delhi – 110029, India
T +91 49 49 5353 F +91 49 49 5391
fairforward@giz.de
https://www.bmz-digital.global/en/overview-of-initiatives/fair-forward/

In collaboration with
NASSCOM
Plot No. 7-10, Sector-126
Noida 201303, India
info@nasscom.in

As at
February 2026

Designed by
Devika Dwarkadas
Goa, India

Text
Aarushi Gupta, Digital Futures Lab
Anushka Jain, Digital Futures Lab
Urvashi Aneja, Digital Futures Lab
Shreeja Sen

On behalf of the
German Federal Ministry for Economic Cooperation and Development (BMZ)

Foreword

India stands at a critical juncture in shaping the trajectory of Artificial Intelligence (AI), by taking leadership in how this technology is developed, governed and deployed for public good. Today, AI systems influence access to and delivery of essential public and private services, from healthcare and education to agriculture and finance, making it imperative that they embody the values of transparency, affordability, inclusion, and accountability. In this context, open source AI has emerged as a key enabler of India's vision to democratise AI, lower entry barriers, and widen participation across startups, researchers, and public institutions.

Through the IndiaAI Mission, the Government of India is advancing a model that anchors AI in digital public infrastructure, interoperable building blocks, and shared resources, while upholding the principles of safety, responsibility, and trust. This approach seeks to ensure that core AI capabilities, compute, datasets, models, and platforms, are accessible as publicly provisioned infrastructure on which innovators across enterprises, startups and academia can build. India remains committed to an open, responsible, safe and secure AI ecosystem tailored to national development priorities and aligned with global best practices.

In alignment with this, the Open Source Policy Brief by the IndiaAI Mission, Nasscom, Digital Futures Lab, and the German Development Cooperation (GIZ) project FAIR Forward – AI for All (funded by the German Federal Ministry for Economic Cooperation and Development, BMZ) serves as a timely and practical resource on the role of openness in India's AI ecosystem. The Brief examines how different degrees of openness across data, code, and model weights shape innovation, adaptability, and public oversight, and how open source approaches can complement India's digital public infrastructure to support wider access and innovation. It also recognises the risks and barriers associated with open source AI, including quality assurance, security, and governance challenges, and sets out concrete, actionable recommendations for India and the wider global community to harness open source AI responsibly. I extend my sincere gratitude to the government officials, developers, civic tech leaders, and academic experts whose insights have guided this effort.

Abhishek Singh
Director General, National Informatics Centre
Additional Secretary, Ministry of Electronics
& Information Technology

Foreword

Artificial Intelligence today offers unparalleled potential to drive India's next phase of innovation and development. As we strive to harness AI for societal and economic advancement, open source AI offers a pathway to democratize access, make AI systems more transparent and ensure that AI innovations are better aligned with India's diverse linguistic, cultural, and social realities.

This report is a collaborative initiative of Nasscom, Digital Futures Lab, and GIZ (German Development Corporation) project FAIR Forward - AI for All (funded by the German Federal Ministry for Economic Cooperation and Development (BMZ). It arrives at a critical moment in India's AI journey, offering timely insights into India's journey towards an open source AI ecosystem. Importantly, it examines not only the promise of openness, but also the institutional capacity, governance models, and long-term sustainability required to make open systems truly effective.

The report explores the meaning of "openness" in the AI context and introduces a component–outcome matrix for the industry to illustrate how varying degrees of openness across an AI stack enable distinct outcomes such as transparency, affordability, customizability, and accountability. At the same time, it also recognizes the complexities of an open source approach, namely, the barriers, risks, and the trade-offs that different stakeholders must navigate. To this end, it presents concrete, actionable recommendations, outlining key policy levers for governments as well as practical steps for AI practitioners navigating complex questions around implementing and enabling the adoption of open source AI in the Global South.

For Nasscom, this initiative aligns with our broader mission to nurture a globally competitive, innovation-driven, transparent, and ethically grounded AI ecosystem. Open source AI approach represents not just a technical choice, but a strategic imperative, one that can help India and the Global South achieve greater technological autonomy and leadership in the age of AI.

We hope this report will serve as a practice-oriented guide for the policymakers and developers in India and will enable them to make better-informed decisions and catalyze the development of a collaborative, inclusive, and future-ready open AI ecosystem in India.

Rajesh Nambiar
President, Nasscom

Acknowledgements

We are grateful to the IndiaAI Mission for its engagement and support for this work, and in particular to Mr. Abhishek Singh and Ms. Kavita Bhatia for their guidance and engagement. We also extend our sincere thanks to the members of the working group for this project who shared their valuable time and insights with us and whose guidance and feedback have been instrumental in shaping the brief, Aman Taneja, Amrita Sengupta, Amritendu Mukherjee, Avik Sarkar, Gaurav Godhwani, Jigar Doshi, Meghna Bal, Prasanta Ghosh, Rama Devi Lanka, Shweta Gupta, Sivaramakrishnan Guruvayur, Swetha Kolluri, Venkatesh Hariharan, Vibhav Mithal, Ramakrishna Reddy Yekulla, Atul Gandre, and the Tattle team.

We would also like to thank members of this project's advisory board, Leonida Mutuku, Aaditeshwar Seth, and Maximilian Gahntz for their thorough review and feedback on our drafts.

Lastly, we extend our heartfelt thanks to team members from GIZ India (Philipp Olbrich & Aishwarya Salvi) and Nasscom AI (Ankit Bose, M. Chockalingam, Saikat Saha, Raj Shekhar, Simrandeep Singh, & Kritika Oberoi) who graciously provided us with detailed feedback and guidance throughout this project.

Table of **Contents**

Table of **Contents**

Executive Summary

India stands at a critical juncture in shaping how artificial intelligence (AI) is developed, governed, and applied. As AI systems begin to influence a wide range of public and private services, from healthcare to education and welfare delivery, questions of transparency, affordability, inclusion, and accountability are gaining prominence. Who builds and funds AI systems? On what terms are they shared or adapted? What forms of access and oversight are made possible — or foreclosed — by the way AI is designed and released? These aspects are also highlighted by the India AI Governance Guidelines released in November 2025, which emphasise the need to build more transparent and accessible AI systems.

Against the backdrop of these critical questions, this brief positions open source AI as a strategic policy option within India's evolving AI ecosystem. It unpacks the critical importance of openness for India's digital future, crystallising the key opportunities it offers to governments and developers alike. In doing so, it also explores the very meaning of "openness" in the context of AI, and how varying degrees of openness across data, code, and model weights shape possibilities for innovation, adaptability, and public oversight.

At the same time, the brief also recognises the barriers and risks that accompany an open source approach, particularly within India's institutional and regulatory realities, and the trade-offs different stakeholders may encounter. Building on this analysis, it presents **concrete, actionable recommendations, outlining key policy levers for governments as well as practical steps for AI practitioners navigating complex questions around implementing open source AI.** Both the analysis and the recommendations in this brief have been informed by and validated through a series of stakeholder interviews and workshops with government officials, developers, civic-tech organisations, and academic institutions.

Below is an overview of the key takeaways from this brief.

Understanding **Openness in AI**

The term “open source AI” is often used inconsistently, encompassing a wide range of release practices that differ not only in degree but in intent and substance. This is because openness in AI is not a binary attribute but a spectrum. Components of an AI model (data to model weights to documentation) can be made open to varying degrees and under different conditions.

Recognising this complexity, this brief proposes a component–outcome matrix to show how different levels of openness across an AI stack enable distinct outcomes such as transparency, affordability, customisability, and accountability. Rather than prescribing a singular definition of open source AI, the matrix serves as a practical heuristic to help Indian stakeholders navigate choices around openness in their respective contexts. It also allows for a more grounded policy conversation about what should be opened, by whom, under what safeguards, and towards what ends.

Opportunities of Open source AI for India

Broadening Participation, Innovation, and Strategic Autonomy

Open source AI holds significant potential to reshape who builds, adapts, and benefits from AI systems in India. By reducing the costs of access and experimentation, openness expands the range of actors who can meaningfully participate in AI development - small startups, academic institutions, state agencies, civic-tech groups, and independent researchers. This is particularly salient in the Indian context, where the needs of diverse linguistic communities, sector-specific use cases, and resource-constrained settings may not be catered to well by mainstream, proprietary AI models.

Beyond widening participation, open approaches can also serve as a lever of strategic autonomy and digital sovereignty, reducing dependence on foreign vendors. Much of today’s AI infrastructure is controlled by a small number of global firms, limiting India’s ability to adapt tools to its own regulatory, linguistic, and societal

needs. Open approaches offer a counterweight by enabling domestic institutions to inspect, modify, and govern AI systems independently. This flexibility is especially important in sectors like health, education, and agriculture, where local relevance and public trust are critical.

Openness as a Pathway to Responsible AI

Transparency is essential not just for technical scrutiny but for making AI systems more responsible, ethical, and aligned with public interest. As AI tools increasingly mediate decisions in health, education, financial services, and beyond, the need to understand how these systems operate, evaluate their impacts, and trace their decision logic becomes critical.

Open source AI helps create these conditions. When key components such as model code, datasets, evaluation protocols, and documentation are made accessible, they enable external actors, including researchers, developers, civil society, and regulators, to audit systems, replicate findings, and identify issues like bias, discrimination, or misalignment with stated objectives.

Crucially, **transparency supports responsibility by making it possible to attribute choices, test assumptions, and challenge outcomes.** Without it, efforts to ensure fairness, legality, or safety remain speculative. Open source AI is therefore not just an industrial policy lever, but a critical component of governing AI systems.

Risks, Trade-offs, & Barriers

While openness offers significant public benefits, it also introduces real risks. Open release can enable misuse, including the generation of disinformation, as well as biased or unsafe deployments. These risks are particularly pronounced in settings where institutions lack the legal, technical, or organisational capacity to monitor or mitigate such harms. Many of these risks are not unique to open source models. Closed and proprietary models also raise serious concerns around bias and malicious use. However, the risk profile of open source AI is shaped by its decentralised governance. This structure can be both a strength and a limitation. **Decentralisation may support transparency and wider oversight, but it can also leave gaps when misuse or security threats escalate more rapidly than existing safeguards can respond.**

That said, these risks do not negate the value of openness. Instead, they point to the importance of building institutional capacities and governance frameworks that can effectively mitigate them.

Key Recommendations: Policy and Developer Pathways

Realising the potential of open source AI in India will require tailored action from both policymakers and developers. Governments must play multiple roles: as promoters, regulators, users, and developers of AI systems, embedding openness where it creates the most value while safeguarding against risks.

Developers and data contributors, meanwhile, bear responsibility for building meaningful openness into their design choices, documentation, and release practices. The summary below distils key recommendations from this brief, offering a high-level guide for both sets of actors.

Policy Directions for Government

State as a Promoter of Open source AI

→ Support open-source AI projects through existing compute allocation schemes; prioritise access for projects committing to open source AI components.

→ Create long-term sustainability mechanisms (grants/ blended finance) for maintaining high-value open datasets, models, and tools.

→ Extend MSME-style designations to small open-source AI firms to improve procurement participation and access to finance.

→ Build awareness across government and ecosystem actors through information campaigns, short courses, and public repositories.

→ Integrate open approaches into national innovation programmes and hackathons.

Policy Directions for Government

State as a Regulator and Standard Setter

→ Establish minimum thresholds of openness in publicly funded AI — such as release of source code, weights, evaluation datasets, and adequate documentation.

→ Convene a community-led effort to clarify how licensing applies to AI (RAIL, Creative Commons, Apache) and develop India-specific templates if needed.

→ Complement licensing with safeguards such as oversight mechanisms, watermarking, and bug-bounty programmes, particularly for high-stakes deployments of general-purpose AI models.

→ Extend MeitY's existing Quality and Certification frameworks (STQC) to include AI systems by introducing evaluation benchmarks for aspects like robustness, reproducibility, cybersecurity, and documentation. These benchmarks would be especially valuable for AI systems built leveraging open source AI models, helping improve their credibility, production-readiness, and eligibility for government adoption.

Policy Directions for Government

State as a Procurer and User

→ Embed openness as a consideration in procurement, reward openness in procurement scoring where desirable.

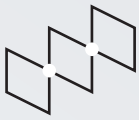
→ Relax prior deployment requirements so open source teams can demonstrate fitness via pilots or audits rather than past large deployments.

→ Build procurement capacity through the training of government officials.

→ Establish transparency baselines for public sector AI (registries, disclosure of evaluations, and risk assessments).

→ Model best practices when the state is itself a developer: release models/datasets under responsible licenses, ensure documentation and traceability, and ensure platforms like AI Kosh meet minimum quality and documentation standards.

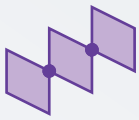
Practical Guidance for the AI Development Community



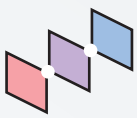
Prioritise meaningful openness
Focus on releasing components that enable genuine reuse and scrutiny, such as model weights, training code, training frameworks, evaluation datasets, and documentation, rather than symbolic or partial artefacts.

Strengthen documentation and traceability
Publish model cards, data cards, metadata, paradata, and lineage information to ensure transparency around development choices, dataset provenance, and performance characteristics.

Choose licensing fit for purpose
Select licences that suit your intended use, degree of openness, and risk profile. Use permissive licences where broad adoption is the goal, and purpose-limited licences (such as OpenRAIL) or custom licence terms where additional safeguards are needed to address misuse or context-specific risks.



Incorporate safeguards for responsible use
Use technical and legal safeguards such as system-level protections, use-based licensing clauses, and public statements of permitted use, to reduce the risk of misuse or harm, particularly for sensitive or high-risk deployments.



Plan for long-term stewardship
Build clear mechanisms for ongoing maintenance, community contribution, bug reporting, and external audit, to ensure open source artefacts remain usable, safe, and relevant over time.

Background: The Open Source AI Moment

Open source AI has gained significant traction over the past few years. Several model developers have embraced varying degrees of openness across different layers of the AI stack. This ranges from training data and source code to model weights and architectural design. A recent and notable example is OLMo, a large language model (LLM) released by the Allen Institute for AI. Its model weights, training data, training code and logs are publicly available, marking a significant moment in the evolution of open source AI.[1] Meta's AI models, such as OPT-175B and Llama, are also popular examples of open weight models.[2] However, the exact nature of their openness remains contested, given the limitations imposed by Meta on their use and redistribution.[3]

Despite contentions around its definitions, open source AI is widely recognised as a key enabler for making AI systems more transparent and accessible, particularly when compared to proprietary, black-box models.[4] It has been foundational to global AI trajectories, fuelling innovation and accelerating scientific discovery. Openly released models such as BLOOM, BERT, and Stable Diffusion have catalysed new directions in natural language processing and generative AI, often serving as baselines or building blocks for future models.[5] Similarly, publicly-released datasets like ImageNet and Common Crawl have played an instrumental role in

1 Team OLMo et al., '2 OLMo 2 Furious', arXiv:2501.00656, preprint, arXiv, 8 October 2025, https://doi.org/10.48550/arXiv.2501.00656.
2 'Democratizing Access to Large-Scale Language Models with OPT-175B', 3 May 2022, https://ai.meta.com/blog/democratizing-access-to-large-scale-language-models-with-opt-175b/.
3 'Meta's LLaMa License Is Not Open Source', Open Source Initiative, 20 July 2023, https://opensource.org/blog/metas-llama-2-license-is-not-open-source.
4 Open Source Initiative, 'The Open Source Definition', Open Source Initiative, 16 February 2024, https://opensource.org/osd.
5 'What's BLOOM and Why Is It Democratizing AI?', https://www.voiceflow.com/blog/bloom-ai.

training state-of-the-art models.[6] As of 2024, more than a billion contributions have been made to open source and public repositories on GitHub, which also include several generative AI projects.[7]

Open source AI is thus increasingly emerging as a compelling alternative to the prevailing Big Tech-dominated model of AI development, offering the promise of a more collaborative, democratised, and contextually sensitive AI paradigm.

In particular, it offers practical remedies for some of the key challenges faced by countries in the Majority World, including India, where accessing and using state-of-the-art AI systems, having mostly originated in the Global North, is prohibitively resource-intensive and ridden with issues of underrepresentation.[8] By lowering barriers to entry into the AI development ecosystem and offering pathways to customise large AI models, an open source approach can help such countries achieve greater technological autonomy while ensuring that AI innovations are better aligned with local linguistic, cultural, and social realities.

In India, AI initiatives underpinned by open source approaches are already taking shape. In July 2025, the IndiaAI Mission, the Indian government's flagship initiative on AI, confirmed that the large language model it is sponsoring will be released under an open source license.[9] Complementing this, the Bhashini Initiative and the wider Digital India Mission reflect a consistent policy orientation that values openness, interoperability, and shared digital infrastructure.[10] Yet these ambitions also surface important questions about sustainability, accountability, and the institutional capacity required to govern open systems effectively.

Simply making AI components publicly available does not automatically ensure responsible use or mitigate risks such as bias, security vulnerabilities, or ethical lapses. Realising the benefits of openness requires deliberate design, governance, and sustained institutional support. Enabling and sustaining

6 'Common Crawl - Open Repository of Web Crawl Data', https://commoncrawl.org/.
7 GitHub Staff, 'Octoverse: AI Leads Python to Top Language as the Number of Global Developers Surges', The GitHub Blog, 29 October 2024, https://github.blog/news-insights/octoverse/octoverse-2024/.
8 Gabriel Nicholas and Aliya Bhatia, Lost in Translation: Large Language Models in Non-English Content Analysis (Center for Democracy & Technology, 2023), https://cdt.org/insights/lost-in-translation-large-language-models-in-non-english-content-analysis/.
9 'Sarvam AI Will Open Source Its IndiaAI Mission AI Models - The Economic Times', accessed 6 November 2025, https://economictimes.indiatimes.com/tech/artificial-intelligence/sarvam-ai-to-open-source-ai-models-it-is-training-under-indiaai-mission/articleshow/122524232.cms?from=mdr.
10 Bhashini India, Field Guide for Inclusive Language AI in India (2025), https://bhashinimigrationns.sosnm1.shakticloud.ai:9024/bhashinistaticassets/bhashini-assets/website/Field%20Guide%20-%201st%20Edition%20%284%29%20%281%29.pdf.

openness in AI is also far from straightforward, involving complex conceptual and operational questions that continue to be debated globally.[11] At the definitional level, questions persist around what 'openness' means in the context of AI and how its thresholds can be formally defined.[12] On the operational front, issues related to the security, governance, long-term sustainability, and stewardship of open source AI models remain contested. As a result, operationalising openness in AI requires not only technical clarity but also a clear, actionable roadmap that can inform the efforts of policymakers and developers navigating this space.

Against this backdrop, this brief aims to provide a set of practice-oriented guidelines to support policymakers and developers in India in effectively harnessing the potential of open source AI, while delicately balancing the trade-offs it poses (see Appendix I for a detailed overview of our scope and methodology). By equipping these key stakeholders with actionable insights and a nuanced understanding of the open source AI arena, we hope to enable them to make better-informed decisions and catalyse the development of a collaborative, inclusive, and future-ready open AI ecosystem in India.

11 Adrien Basdevant et al., 'Towards a Framework for Openness in Foundation Models: Proceedings from the Columbia Convening on Openness in Artificial Intelligence', arXiv:2405.15802, preprint, arXiv, 17 May 2024, https://doi.org/10.48550/arXiv.2405.15802.

12 Edd Gent, 'The Tech Industry Can't Agree on What Open source AI Means. That's a Problem.', MIT Technology Review, 25 March 2024, https://www.technologyreview.com/2024/03/25/1090111/tech-industry-open-source-ai-definition-problem/.

What Makes An AI System Open?

Definitional challenges are intrinsic to both the concept of openness and to the very nature of AI models. In traditional software development, the assessment of openness is relatively straightforward: it primarily concerns the accessibility and licensing of the source code.[13] However, in the context of AI, this characterisation of openness becomes more complex.

Unlike other software, AI models are not composed of a single codebase alone. They are multi-component systems comprising datasets, model architectures, training algorithms, model weights, evaluation benchmarks, and downstream applications. Openness thus needs to be viewed as a composite quality, often resulting in conceptual ambiguity. Moreover, as AI models grow in complexity — evident in the advent of transformer architectures — defining openness as a static, binary quality is not always feasible or desirable.

The term open may be used to describe systems that offer transparency, reusability, and extensibility, i.e., they can be scrutinised, reused, and built on.[14] However, this may look different in practice. This includes defining what it means to claim that a specific component of a model is -open-. It also needs to be understood which components need to be open for the AI system to be characterised as open source. Concurrently, the question also arises if openness is a system-level

13 'The Open Source Definition', Open Source Initiative, n.d., accessed 6 November 2025, https://opensource.org/osd/.
14 David Gray Widder et al., 'Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI', SSRN Scholarly Paper no. 4543807 (Social Science Research Network, 17 August 2023), https://doi.org/10.2139/ssrn.4543807.

characteristic that should not be based on specific components but seen as the sum of the system's parts and what it enables.

An unfortunate result of the definitional ambiguity surrounding OS-AI is the phenomenon of open-washing, where AI companies claim their foundational models are open by sharing access to certain components, such as weights, in order to benefit from the open tag. However, they continue to restrict access to critical information and components required to build similar models, such as the underlying data, and escape the scientific scrutiny and legal exposure that comes with full openness.[15]

Such open-washing tactics obfuscate both the benefits and the limitations that different degrees of disclosures enable. Open weights, for example, help enable only a partial form of transparency, which often falls short of the disclosure standards needed to engender true accountability of system designers. This contestation is clearly evident in the case of Meta's Llama models: while Meta points to the release of weights as evidence of openness, critics argue that such disclosure protocols do not enable the level of transparency and accountability that have long been demanded of the company.[16]

This debate around Llama — coupled with the multi-component nature of AI — illustrates why the definition of openness in AI cannot be pegged to the disclosure of a single component or stay divorced from the outcomes such disclosures enable.

This is also the guiding lens that underpins recent efforts to define open source AI (see Appendix III). For example, the Open Source Initiative lays out four freedoms that an open source AI system must enable to qualify as such. Similarly, Mozilla Foundation's framework encourages developers, regulators, and civil society to think beyond whether something is open or closed, and instead ask: what is open, to whom, and for what purpose?

15 Andreas Liesenfeld and Mark Dingemanse, 'Rethinking Open Source Generative AI: Open Washing and the EU AI Act', in The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24: The 2024 ACM Conference on Fairness, Accountability, and Transparency, Rio de Janeiro Brazil: ACM, 2024), 1774–87, https://doi.org/10.1145/3630106.3659005; Edd Gent, 'The Tech Industry Can't Agree on What Open source AI Means. That's a Problem.', MIT Technology Review, 25 March 2024, https://www.technologyreview.com/2024/03/25/1090111/tech-industry-open-source-ai-definition-problem/.

16 Sarah Kessler, 'Openwashing', The New York Times, 17 May 2024, sec. Business, https://www.nytimes.com/2024/05/17/business/what-is-openwashing-ai.html; 'Meta's LLaMa License Is Not Open Source'.

A Practical Rubric for Understanding **What Open Source AI Enables**

Open source AI initiatives offer immense value to both policymakers and AI practitioners.[17] As a policy instrument, leveraged by the government as a regulator or procurer, open source can enhance transparency and, consequently, accountability in AI value chains. As a strategic lever, it can also offer various possibilities to reduce dependence on a smaller set of actors, thus supporting digital sovereignty.[18] Furthermore, the customisability afforded by open source AI can help ensure better representation of local contexts and values in AI systems.

As a tool for technological production or development, open source AI can enable greater collaboration across different entities in the ecosystem, lower entry barriers for smaller actors, complement safety and security measures, and accelerate innovation and cutting-edge research.

What openness enables depends on which component of the AI system is open. For instance, open data may enable greater scrutiny and contextual adaptation, while open weights may facilitate replication and benchmarking when included with other open components. These opportunities manifest across different layers of the AI ecosystem, offering distinct advantages to different stakeholders.[19]

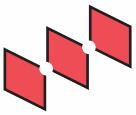
To reflect this, we propose a component-outcome matrix that disaggregates openness across key elements of a typical AI stack (model, data, evaluation results) and maps these to the kinds of affordances or outcomes that openness in each component makes possible.

17 Sayash Kapoor et al., 'On the Societal Impact of Open Foundation Models', arXiv:2403.07918, preprint, arXiv, 27 February 2024, https://doi.org/10.48550/arXiv.2403.07918.
18 Yacine Jernite and Lucie-Aimee Kaffee, 'Open Source AI: A Cornerstone of Digital Sovereignty', Hugging Face, 11 June 2025, https://huggingface.co/blog/frimelle/sovereignty-and-open-source.
19 Interview with a Working Group member.

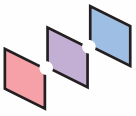
Structure of the Component-Outcome Matrix

The columns in this matrix represent the various outcomes that can be achieved through open source AI. These include transparency, reproducibility, customisability, and affordability. These outcomes were chosen because they reflect the core policy and governance goals for open source AI identified in the literature and our stakeholder consultations.

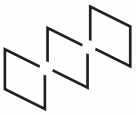
Transparency ensures that systems can be inspected and understood, supporting accountability and preventing open washing.



Reproducibility allows independent verification of results, building trust and credibility, especially in public or safety-critical contexts.



Innovation and customisability enable re-use, adaptation, and local innovation, linking openness to economic development and capacity building.



Affordability reduces barriers to entry, ensuring open source AI supports inclusive participation rather than benefiting only well-funded actors.[20]

It is worth noting that these qualities, in many instances, will also reinforce each other. For example, the customisability of existing pre-trained AI models will allow many developers to bypass the need to build a foundational model from scratch, thus making the development of AI applications more affordable.

The rows in the matrix represent the various components of an AI system, including training data, model source code, model weights, and evaluation results, among others. Many of these components may not apply to all types of AI models. To capture this variation, the last column in the matrix indicates whether each component is primarily relevant to generative AI, predictive AI, or broadly applicable to all types of AI models.

20 Affordability in this matrix is defined in relation to downstream developers, rather than consumer-facing affordability. It refers to lowering the entry barriers for smaller actors by providing access to reusable components such as pre-trained weights, thereby reducing the cost of building AI applications.

↓ Table 1
Mapping Opportunities of Open source AI: A Component x Outcome Matrix

Key Components and Information	Transparency	Reproducability	Innovation and Customisability	Affordability	Relevant Model Category
Model Source Code	✓	✓	✓	✓	**All**
Model Weights	✓	✗	✓	✓	**Generative**
Model Input or Variable Selection	✓	✓	✗	✗	**Predictive**
Model Training Process	✓	✓	✗	✗	**All**
Pre-Training Data or Training Data	✓	✓	✓	✓	**All**
Fine-tuning or Task-specific Data	✓	✓	✓	✗	**All**
Data Collection & Processing Protocols	✓	✓	✓	✗	**All**
↳Workers involved in data cleaning, translation, and annotation[21]	✓	✗	✗	✗	**All**
Model Evaluation Data or Test Data	✓	✗	✗	✗	**All**
↳Workers involved in reinforcement learning with human feedback[22]	✓	✗	✗	✗	**Generative**
Model Evaluation Results	✓	✗	✗	✗	**All**

21 This sub-category captures the human labour involved in preparing datasets, including tasks such as data cleaning, annotation, and translation which directly affects dataset accuracy, bias, and representativeness.

22 This category refers to the human evaluators who provide judgments on model outputs during reinforcement learning with human feedback (RLHF). Their feedback is used to train preference models and guide system alignment, directly influencing how the AI responds as well as its safety and reliability.

How To Interpret This Matrix

Each cell in the matrix, marked either by a check mark (✓) or a cross (✗), signifies whether a particular component contributes to enabling the corresponding outcome. The intent is not to quantify the extent of contribution, but to signal the presence or absence of an enabling role.

Unlike the more granular frameworks discussed in Appendix III, which aim to define and measure the degree of openness in AI systems, this matrix deliberately avoids a normative framing about what qualifies as an open source AI system.

Instead, it offers a pragmatic diagnostic tool to understand what open source AI makes possible, and how those possibilities differ based on which components are open. In doing so, we frame AI openness not as an end in itself, but as a means to advance downstream goals such as transparency, replicability, innovation, or local adaptability.

With respect to the extent of openness itself, emerging frameworks (see Appendix III) emphasise that openness is best understood as a spectrum rather than an absolute, binary characteristic. For instance, pre-training data could be fully disclosed, partially disclosed through documentation or summaries, or remain closed due to privacy and IP constraints. In this matrix, the check marks only represent the potential for a component to contribute to a given outcome, assuming a meaningful degree of openness. The actual extent of that contribution depends on where the component falls along the openness spectrum.

Notes:

i. Model weights facilitate transparency, affordability, and customisation, and can allow downstream developers to scrutinise or adapt models. However, they do not ensure reproducibility: without access to training data, preprocessing pipelines, and hyperparameters, downstream actors cannot reproduce the model itself, only replicate its behaviour at inference.

ii. While having model training processes openly available may not directly support innovation and customisability, knowing process details may support innovation by helping organisations understand process details better.

iii. Publicly disclosing model training processes may provide affordability gains by allowing model developers to bypass costs associated with trial and error during training processes.

Unpacking the Opportunities Offered by Open Source AI

Using our component-outcome matrix as an anchor, this section provides a detailed explanation of each of the opportunities that come with the opening of different components of an AI stack, distilling them from the vantage points of two key stakeholder groups, the government and the AI development community (model developers, data collectors, and downstream application developers).

Direct **Benefits Of Openness** in AI

Enhanced Transparency

One of the most significant opportunities that open source offers in the context of AI is that of transparency, a concept intrinsically linked to openness itself. Transparency in AI can have multiple meanings or interpretations[23] but broadly, it enables a range of stakeholders, including policymakers, regulators, researchers, downstream developers, and users, to understand how an AI system has been developed, trained, and deployed, and allows insight into the processes and choices that shape its outputs.[24]

Importantly, transparency is not an end in itself. It functions as an enabling condition for other properties such as interpretability, explainability, traceability, auditability, and safety & security (see Table 2). These properties, in turn, make it possible to attribute responsibility for errors or harms, thereby strengthening accountability in AI systems.[25]

There are several ways to ensure greater transparency in AI systems. Its extent directly varies with different levels of openness in an AI system, depending on which component is open and to what extent.[26] As indicated in the matrix above, appropriate public disclosures of any one or more components of an AI system can, in their own unique way, unlock one or more facets of transparency. We explain in more detail below.

Datasets

In the case of data (components 5 to 8), a variety of disclosures can be provided by the system developer or other entities involved in data collection and processing. The gold standard of such data disclosures is the public release of an AI model's training data itself, enabling individuals, organisations, and the wider community to understand and assess the quality of the inputs or raw material provided to the model. Such disclosures not only facilitate explainability but also strengthen traceability, auditability, and safety & security, by allowing independent scrutiny of the data sources and collection practices.[27]

23 Larsson and Heintz, 'Transparency in Artificial Intelligence', info:eu-repo/semantics/article, Alexander von Humboldt Institute for Internet and Society gGmbH, 5 May 2020, https://doi.org/10.14763/2020.2.1469.

24 OECD AI Principles, 'Transparency and Explainability', OECD.AI, accessed 30 April 2025, https://oecd.ai/en/dashboards/ai-principles/P7.

25 Carnegie Council for Ethics in International Affairs, 'AI Accountability', Carnegie Council for Ethics in International Affairs, accessed 8 September 2025, https://www.carnegiecouncil.org/explore-engage/key-terms/ai-accountability.

26 Solaiman, 'The Gradient of Generative AI Release'; Larsson, S. (2017). Conceptions in the Code. How Metaphors Explain Legal Challenges in Digital Times. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190650384.001.0001

27 Interview with a Working Group member.

However, in practice, full disclosures around training data are met with several constraints. Pre-training datasets are often treated as closely guarded trade secrets, making it less likely for developers to offer full access to such datasets.[28] Moreover, there may also be a variety of privacy or security-related challenges that emerge from the disclosure of such data (see Section 5). Even where privacy concerns do not apply, such as in the case of non-personal data, data disclosures may not be feasible due to the presence of copyright-protected material.

Given these concerns, providing summaries of such datasets has been proposed as an alternative approach for achieving transparency.[29] These summaries only provide information about the broad characteristics of a system's training data, while avoiding disclosures of the underlying data points.[30]

In addition to the pre-training data, publishing the data used to fine-tune (component 6) or instruct the model (more specialised datasets, prompts or annotated examples provided to the model) clarifies how the model's behaviour was shaped by its developers. This not only enhances traceability of design choices but also supports explainability and, in some cases, auditability and safety & security by enabling scrutiny of the alignment data itself. That said, it does not engender the same degree of transparency as disclosing the full training dataset does.

Model weights

Similarly, release of model weights (component 2) can provide information about the underlying values and biases that influence how the model perceives inputs and generates outputs. Open weights also help provide insights into the model's decision logic, which contributes to interpretability.[31] They also enable an enhanced form of auditability by allowing researchers and auditors to run and test the model actively, rather than a passive review of datacards and documentation.[32]

28 Xiangyu Qi et al., 'Fine-Tuning Aligned Language Models Compromises Safety, Even When Users Do Not Intend To!', arXiv:2310.03693, preprint, arXiv, 5 October 2023, https://doi.org/10.48550/arXiv.2310.03693.

29 'High-Level Summary of the AI Act', EU Artificial Intelligence Act, 30 May 2024, https://artificialintelligenceact.eu/high-level-summary/.

30 'Training Data Transparency in AI: Tools, Trends, and Policy Recommendations □□', accessed 22 May 2025, https://huggingface.co/blog/yjernite/data-transparency; '(PDF) Constructing an AI Value Chain and Ecosystem Model', ResearchGate, https://doi.org/10.13140/RG.2.2.11981.86246.

31 'Open Weights: Not Quite What You've Been Told', Open Source Initiative, accessed 24 April 2025, https://opensource.org/ai/open-weights.

32 'Definition/Definition.Md at Main · Open-Weights/Definition · GitHub', accessed 9 June 2025, https://github.com/Open-Weights/Definition/blob/maicodedependn/Definition.md.

Model Source Code

Disclosures of the model's source code (component 1) can enable transparency, but only to a limited extent. While the source code reveals information about the model's architecture or tokenisation strategies, it does not disclose the kind of instructions — restrictions or safety measures — the developers may have provided to the model. Such information is typically embedded in the training data.[33] Moreover, the level of detail the source code reveals and the ease with which it is uncovered mostly depends on how the code is structured. For example, codebases for LLMs can span millions of lines, with the ability to scrutinise such complex codes limited to highly trained AI experts.

Human Labour Inputs

Data workers and human evaluators (components 7.1 and 8.1) play a critical role in shaping the performance and safety of an AI system. In many generative AI systems, particularly those trained using Reinforcement Learning from Human Feedback (RLHF), model outputs are aligned with the preferences, judgments, and value systems of the humans involved in the evaluation and feedback loops. Similarly, data workers who collect, label, annotate, or curate training datasets influence what the model learns in the first place, determining what is included, what is excluded, and how information is categorised. Their decisions directly affect representation, fairness, and contextual relevance in the model's behaviour.[34]

Disclosing these human labour inputs, including the nature of the task performed and the instructions provided to the annotators and evaluators, enhances transparency, and, in turn, traceability and explainability, by making visible the human assumptions and values that are embedded in the AI system.[35] Such disclosures also reinforce the understanding that AI systems are socio-technical artefacts, not autonomous or value-neutral.

Evaluation Data & Results

Lastly, the release of evaluation datasets and results (components 8 and 9) allows scrutiny of the quality of the data and the robustness or validity of the process that was used to test the model.[36] For instance, open evaluation datasets allow independent replication of tests and benchmarking against comparable models, while published results make visible both the strengths and limitations of a system.

33 Interview with a Working Group member.
34 Amandalynne Paullada et al., 'Data and Its (Dis)Contents: A Survey of Dataset Development and Use in Machine Learning Research', Patterns 2, no. 11 (2021): 100336, https://doi.org/10.1016/j.patter.2021.100336.
35 Solaiman, 'The Gradient of Generative AI Release'.; Interview with a Working Group member.
36 Interview with a Working Group member.

↓ **Table 2**
Unpacking the Different Facets of Transparency and their Enablers

	Facets of Transparency	Enabling Components
Interpretability	Understanding why an AI system produces a given output based on its internal logic, mathematical relationships, or learned patterns.[37]	Model source code, model weights, training process, evaluation data, and results
Explainability	Providing comprehensible reasons or justifications for system outputs.[38]	Training or fine-tuning data, evaluation data and results
Traceability	Systematically recording and tracing how data is collected, processed, and applied, alongside the key design choices that shape an AI system across its lifecycle.[39]	Source code, training process, human labour inputs, data collection protocols, and evaluation results
Auditability	Enabling AI systems to be independently evaluated for compliance with ethical, legal, and technical standards across their lifecycle.[40]	Source code, training process, datasets, evaluation data, and results
Safety & Security	Ensuring AI components and processes are inspectable so that bugs, vulnerabilities, or unsafe behaviours can be detected and mitigated early.[41]	Model source code, model weights, training process, pre-training and fine-tuning data, data collection protocols, evaluation data, and evaluation results

37 Walter Haydock, 'The Complete Guide to AI Transparency, Explainability, and Interpretability', 16 March 2023, https://blog.stackaware.com/p/ai-transparency-explainability-interpretability-nist-rmf-iso-42001.
38 Haydock, 'The Complete Guide to AI Transparency, Explainability, and Interpretability'.
39 Please note that while access to training data documentation and evaluation results meaningfully improves understanding of model behaviour and reliability, such transparency does not imply full interpretability of every individual output. Foundation models remain probabilistic systems, and their internal representations are not fully explainable at the level of specific outputs. Nonetheless, structured transparency practices remain critical for enabling accountability, risk assessment, and responsible deployment.
40 Himanshu Verma et al., 'Can AI Be Auditable?', arXiv:2509.00575, preprint, arXiv, 30 August 2025, https://doi.org/10.48550/arXiv.2509.00575.; 'Preserving Agency: Why AI Safety Needs Community, Not Corporate Control', accessed 5 November 2025, https://huggingface.co/blog/giadap/preserving-agency.
41 'International AI Safety Report 2025 | International AI Safety Report', accessed 9 November 2025, https://internationalaisafetyreport.org/publication/international-ai-safety-report-2025.

Reproducibility

In the context of AI, reproducibility broadly refers to the ability of others to independently build, verify, re-run, or adapt a model using the same methods and components as the original developers. In a fast-moving and research-intensive field such as AI, reproducibility helps promote scientific rigour and accelerate scientific advancements.[42]

Reproducibility in AI systems hinges not on the disclosure of any single component, but on the availability of a constellation of interdependent elements, including model source code, training data, training processes, and model weights. While the release of individual components, such as model weights, can enhance transparency (as discussed above), they are insufficient on their own to enable full replication. Weights can be used to fine-tune or adapt a model for specific tasks, but without access to the original training data and procedures, replicating the original model becomes infeasible. This limitation also impedes efforts to audit the system, trace the provenance of specific behaviours, or correct biases introduced during the initial training process.[43] Similarly, with respect to the source code, many developers tend to only release their inference code. Such disclosures can help others deploy the model, but they cannot help them reproduce it.[44]

Even when developers do provide relatively comprehensive disclosures across the AI stack, reproducibility may still be out of reach for many due to the significant computational, financial, and technical resources required to recreate large-scale models from scratch. We expand on these barriers in Section 5.1.

Innovation and Customisability

Open source AI models offer critical flexibility, allowing the larger community of AI developers and enterprises to adapt these models to specific local contexts or domains. This flexibility enables them to build AI systems and applications that are more representative of regional languages, local cultures, and domain-specific considerations.[45] For example, an open source language model trained primarily on internet data from Westernised contexts may not generalise well to non-Western dialects or values, but reproducibility allows local teams to fine-tune and evaluate it for their own needs. Such customisability is made possible when model developers provide access to the model weights or to the model itself through an API.

42 Matt White et al., 'The Model Openness Framework: Promoting Completeness and Openness for Reproducibility, Transparency, and Usability in Artificial Intelligence', arXiv:2403.13784, preprint, arXiv, 18 October 2024, https://doi.org/10.48550/arXiv.2403.13784.
43 Open Source Initiative, 'Open Weights'.
44 Interview with a Working Group member.
45 Mark Craddock, 'Open Source AI as a Competitive Advantage', Medium, 28 January 2025, https://medium.com/@mcraddock/open-source-ai-as-a-competitive-advantage-45d59a159085.

This can be seen in the case of Sarvam AI's OpenHaathi model, which was developed by customising Meta's Llama, an open-weight model, and enhanced with Indic language capabilities.[46] Access to model weights can enable local developers to adapt the model for their specific purpose without retraining it from scratch — a critical advantage for actors in the Majority World, where computing resources are often limited.

Affordability and Cost-effectiveness

Open source AI can significantly lower the cost barriers to adopting and deploying advanced AI systems. When model weights are publicly released, particularly under permissive licenses, organisations are spared the substantial costs associated with model development, training, and licensing.[47] This makes cutting-edge AI tools accessible to a wider spectrum of actors, including governments, academic institutions, startups, and non-profits that may otherwise be excluded from proprietary AI ecosystems due to financial or infrastructural constraints.[48] A compelling example is InstructLab, an open source, model-agnostic platform that simplifies the fine-tuning of LLMs. InstructLab allows people without formal training in data science to contribute to model development. This opens the door to wider participation, particularly from communities typically excluded from AI development processes.[49]

In countries like India, where access to proprietary AI systems is often constrained by high licensing fees and infrastructural requirements, open source AI models serve as critical enablers of innovation. They allow educational institutions, early-stage startups, and public sector entities to experiment with, adapt, and deploy advanced AI systems at a fraction of the cost. For startups in particular, the ability to iterate using open source models without incurring prohibitive licensing costs can be vital to product development. For example, AI4Bharat's and the Indian Institute of Science's (IISc) release of high-quality text-to-speech models in Indian languages via the Bhashini platform allows local developers to create voice-enabled applications for underserved populations. Similarly, for government agencies, where budgetary constraints limit access to commercial models, open source AI models present a viable and scalable alternative.[50]

46 'Sarvamai/OpenHathi-7B-Hi-v0.1-Base · Hugging Face', accessed 29 April 2025, https://huggingface.co/sarvamai/OpenHathi-7B-Hi-v0.1-Base.

47 Francisco Eiras et al., 'Risks and Opportunities of Open-Source Generative AI', arXiv:2405.08597, preprint, arXiv, 29 May 2024, https://doi.org/10.48550/arXiv.2405.08597.

48 Eiras et al., 'Risks and Opportunities of Open-Source Generative AI'.

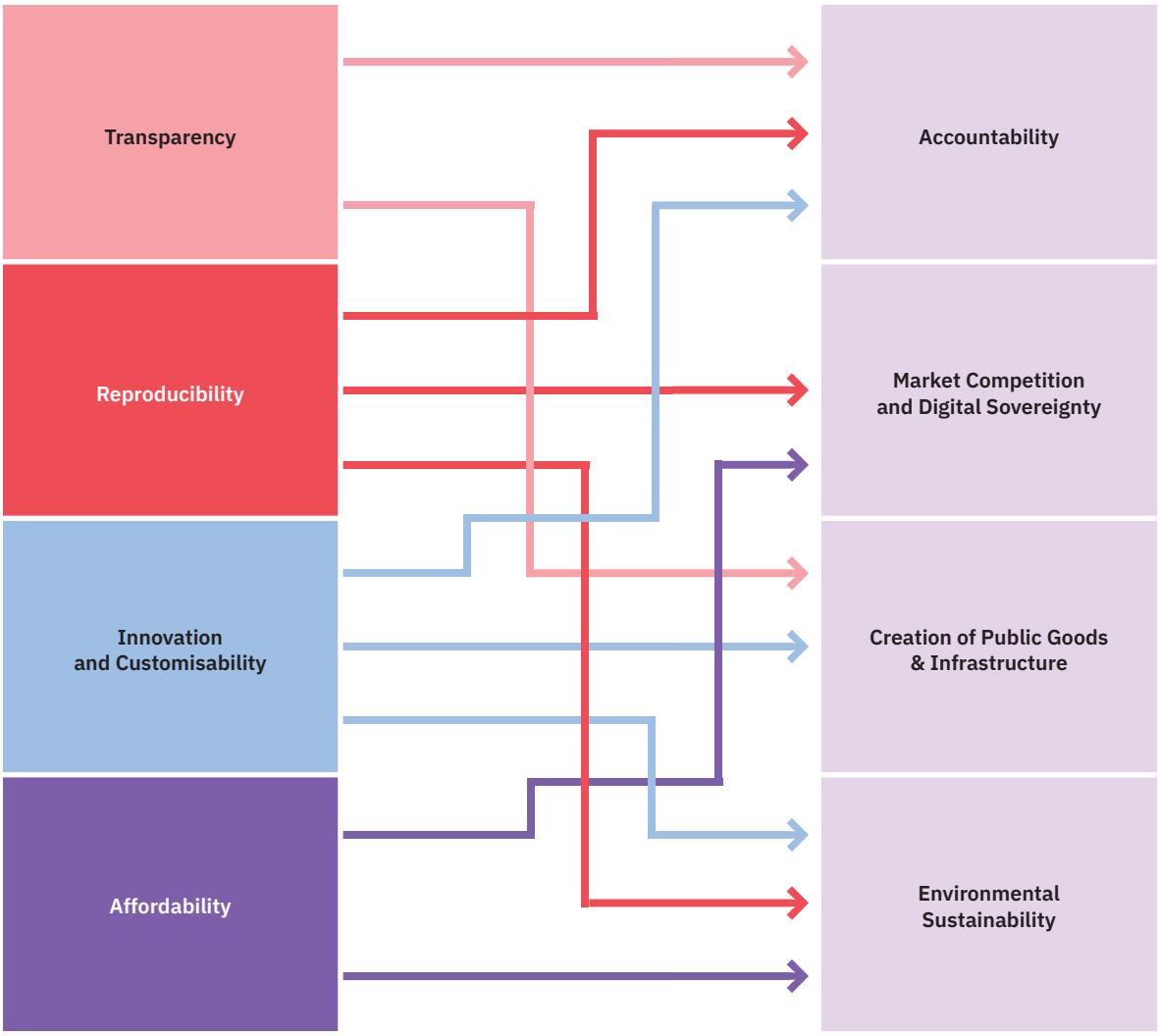
49 'Why Open Source Is Critical to the Future of AI', accessed 25 April 2025, https://www.redhat.com/en/blog/why-open-source-critical-future-ai.

50 Interview with a Working Group member.

Longer-term Impacts of Open source AI

The direct benefits outlined above serve as foundational building blocks for a broader set of longer-term impacts that open source AI could catalyse. These impacts are likely to unfold gradually, potentially driving systemic shifts within both global and national AI ecosystems (see Figure 1).

These longer-term impacts will play out differently across national and sectoral contexts. In India, open source AI can activate multiple strategic levers, ranging from increased accountability to enhancing digital sovereignty, catalysing structural shifts across the ecosystem. This section explores these possibilities in greater detail.



↑ **Figure 1**
Longer-Term Impacts of Open source AI

Accountability

Accountability serves as a vital safeguard in AI governance, helping protect citizens from algorithmic harms and ensuring that AI systems align with principles of equity, privacy, and human rights.[51] In the context of open source AI, accountability is typically enabled through transparency and scrutability, both of which allow stakeholders to examine how and why particular design choices or outputs have emerged, and to assign responsibility accordingly.[52]

Crucially, openness redistributes responsibility beyond the original developers. By making model components visible and modifiable, open source AI invites collective oversight and improvement. Developers, researchers, and users can identify flaws, suggest enhancements, and track unresolved issues, thus creating a distributed accountability mechanism.[53] As Eric Raymond observed, "given enough eyeballs, all bugs are shallow", referring to the ability of the open source community to ensure collaborative correction of system bugs.[54]

In practice, this collective oversight is operationalised through public issue trackers, open documentation, and active community forums, all of which facilitate real-time monitoring and iterative correction. This model of communal scrutiny has been foundational to open source software governance and is increasingly relevant for open models and datasets. However, it is not without limitations. Several high-profile incidents have revealed that open source systems remain vulnerable to security breaches, particularly when community oversight is inconsistent or under-resourced.[55] These limitations, and the need for complementary safeguards, are discussed further in Section 5.2.

Market Competition and Digital Sovereignty

Open source AI can play a transformative role in strengthening digital sovereignty, enhancing market competition, and expanding inclusive participation in AI innovation. By reducing dependence on proprietary systems controlled by a small group of foreign technology firms, open source approaches allow countries to exercise greater control over the development, deployment, and governance of AI systems. This is particularly significant for

51 Sustainability Directory, 'How Does Open Source Improve Accountability Of Systems? → Question', Sustainability Directory, n.d., accessed 29 April 2025, https://sustainability-directory.com/question/how-does-open-source-improve-accountability-of-systems/.
52 Eiras et al., 'Risks and Opportunities of Open-Source Generative AI'.
53 Shay David, 'Opening the Sources of Accountability', First Monday, ahead of print, 1 November 2004, https://doi.org/10.5210/fm.v9i11.1185.
54 Eric Raymond, 'The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary', Choice Reviews Online 39, no. 05 (2002): 39-2841-39–2841, https://doi.org/10.5860/CHOICE.39-2841.
55 Amanda Brock, ed., Open Source Law, Policy and Practice, 2nd edn (Oxford University PressOxford, 2022), https://doi.org/10.1093/oso/9780198862345.001.0001.

India and other Majority World countries seeking to build self-reliant, secure, and contextually relevant AI ecosystems.

Open access to high-quality models, tools, and infrastructure lowers the entry barriers for a wider set of actors (including startups, academic institutions, and public sector bodies) who may otherwise be excluded from proprietary ecosystems due to financial or infrastructural constraints. This enables a more pluralistic and decentralised innovation ecosystem, where entrepreneurship can flourish across a range of sectors and use cases, including those that are typically underserved or commercially unattractive to dominant players.[56] In turn, this supports a more vibrant domestic AI economy that can challenge monopolistic or oligopolistic market structures. India, with its established leadership in digital public infrastructure, is particularly well positioned to build shared open source AI assets that can underpin a resilient and competitive AI industry.[57]

Recent global developments also signal how open source releases can disrupt prevailing industry dynamics. The launch of DeepSeek's R-1 model in early 2025, for instance, demonstrated that open models can achieve competitive performance at significantly lower costs, prompting other developers to revisit their release strategies.[58]

Creation of Public Goods and Infrastructure

Open source AI contributes to the development of a shared digital commons — freely available software, datasets, and models — that can serve as the backbone of essential digital infrastructure. Governments, research institutions, and civil society actors can build and maintain public-interest applications using these common frameworks, without having to start from scratch or rely on proprietary platforms. This reduces duplication of effort and allows resources to be channelled towards adaptation, localisation, and responsible deployment.[59]

One example of such open digital infrastructure is PyTorch, a widely used open source machine learning library originally developed by Meta. Now governed under the Linux Foundation's AI initiative, PyTorch is actively maintained and improved by a broad community

56 Interview with a Working Group member.
57 'India Stack: Digital Public Infrastructure for All - Case - Faculty & Research - Harvard Business School', accessed 9 June 2025, https://www.hbs.edu/faculty/Pages/item.aspx?num=64379; 'Digital Public Infrastructure for the Developing World (SSIR)', accessed 9 June 2025, https://ssir.org/articles/entry/digital-public-infrastructure-developing-world.
58 Jinlin Wu, 'The Rise of DeepSeek: Technology Calls for the "Catfish Effect"', Journal of Thoracic Disease 17, no. 2 (2025): 1106–8, https://doi.org/10.21037/jtd-2025b-02.
59 'Generative AI and the Digital Commons', The Collective Intelligence Project, accessed 29 April 2025, https://www.cip.org/research/generative-ai-digital-commons.

of contributors. Its widespread adoption across academia, startups, and large-scale AI deployments illustrates how open source tools can form the foundational layer of national and global AI ecosystems, while remaining accessible, extensible, and adaptable for public and private sector needs alike.[60]

Environmental Sustainability

In the long term, the transparency and reproducibility of open source AI (OSAI) can also contribute meaningfully to environmental sustainability. Open access to models, datasets, and tools allows researchers and developers to fine-tune and repurpose existing systems instead of training new ones from scratch. This helps reduce redundant computational effort and the associated carbon footprint.[61] Furthermore, the collaborative ethos of open source communities can foster the adoption of efficiency-enhancing techniques such as model distillation (transferring knowledge from a larger model to a smaller, faster one) and quantisation (reducing the numerical precision of model parameters to improve performance). Both methods can significantly lower energy consumption during model training and inference.[62]

60 'PyTorch Foundation', PyTorch, n.d., accessed 11 June 2025, https://pytorch.org/foundation/.
61 Victor Sanh et al., 'DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter', arXiv:1910.01108, preprint, arXiv, 1 March 2020, https://doi.org/10.48550/arXiv.1910.01108; Sasha Luccioni and Regis Pierrard, 'Reduce, Reuse, Recycle: Why Open Source Is a Win for Sustainability', Hugging Face, 7 May 2025, https://huggingface.co/blog/sasha/reduce-reuse-recycle.
62 Luccioni and Pierrard, 'Reduce, Reuse, Recycle'.

The Challenges Involved in Open Source AI Systems: Key Risks & Barriers

Despite its promise in fostering innovation, accountability, and accessibility, the implementation of open source AI systems is fraught with a range of challenges. These challenges take two primary forms: barriers and risks. Barriers are obstacles that hinder developers, governments, and other stakeholders from effectively leveraging open source AI. These may be infrastructural, data-related, operational, or governance-related in nature. Risks, on the other hand, refer to potential harms or failures that could compromise safety, erode public trust, or diminish the long-term efficacy of open source initiatives. Addressing both is essential to ensuring that the promise of open source AI translates into meaningful and sustainable outcomes. We explore each of these categories and their implications below.

Barriers to Achieving Openness

Infrastructural Barriers

The development and deployment of large-scale AI models remain heavily dependent on advanced computational infrastructure, including high-performance computing. These resources are costly, often centralised in well-funded institutions or private corporations, and largely inaccessible to smaller developers, academic researchers, and public sector actors.[63] As a result, even when key components such as model weights or source code are openly released, significant structural barriers persist. The inability to access sufficient compute power constrains the capacity of smaller actors to meaningfully engage with, fine-tune, or repurpose open source models, thereby limiting the practical utility of openness.

In India, access to high-performance compute infrastructure remains heavily skewed toward a handful of large technology firms and premier academic institutions. While initiatives such as the IndiaAI Compute Portal seek to address this gap, several researchers, startups, and public sector bodies continue to face significant limitations in accessing affordable, on-demand compute.[64] These constraints are especially pronounced for those working in regional languages or under-resourced domains where commercial incentives for private investment are low.[65]

Data-related Barriers

Achieving meaningful openness in the AI ecosystem is also constrained by the availability, quality, and legal shareability of training data. While model weights and source code can often be released without friction, data introduces a qualitatively different set of challenges. First, high-value datasets are rarely open in practice. Training corpora for large AI systems are expensive to acquire and curate, and are frequently composed of material that is either protected by copyright or database rights or commercially sensitive. These constraints apply even when no personal or sensitive data is involved. Participants in our stakeholder consultations emphasised that copyright, not privacy, is often the binding constraint in deciding what can be released.

63 Jai Vipra, 'Computational Power and AI', AI Now Institute, 27 September 2023, https://ainowinstitute.org/publications/compute-and-ai.
64 Chase India, Navigating IndiaAI Mission (2025), https://www.chase-advisors.com/media/5d2hqyru/chase-report-navigating-indiaai-mission.pdf?utm_source=chatgpt.com.
65 Stakeholder Workshop, July 2025

Secondly, even where datasets are nominally open, they are not always reusable. Many are released without adequate documentation of provenance, annotation protocols, filtering and preprocessing steps, or known sources of bias and error. The absence of metadata and lineage documentation often results in open-washing at the data layer, i.e., the dataset becomes technically open but substantively unusable.[66] For example, recent research on open datasets in Africa has shown that while many datasets are intended to be open, practical access often involves requesting permission, navigating specific repository interfaces, or being part of a research collaboration.[67]

Lastly, maintaining datasets as shared infrastructure requires sustained institutional support. Without dedicated stewardship, periodic maintenance, and stable funding, datasets risk becoming one-off releases rather than durable public goods. This is especially difficult in low-resource environments, where the costs of hosting, storage, and curation are substantial.

Collectively, these factors make data openness structurally difficult to operationalise within current legal, financial, and institutional realities. In the absence of clear guidance on licensing, documentation standards, and sustainable stewardship models, openness at the data layer remains significantly fragile.

Operational Barriers

Open source tech projects typically depend on the sustained engagement of contributors to maintain code quality, ensure timely updates, and drive innovation. However, when contributions are primarily voluntary, participation becomes inherently fragile. Developers can disengage at any time, taking critical knowledge and momentum with them.

Unlike commercial AI models, which are backed by permanent engineering teams and predictable funding, open source AI projects may frequently lack guaranteed funding or staffing. This leads to high "developer churn," resulting in lost expertise, slower iteration cycles, and gaps in documentation and model accountability.[68] Without stable, earmarked funding for ongoing stewardship, maintenance, and

66 Open Source Initiative, Reimagining Data for Open Source AI: A Call to Action, 23 January 2025, https://opensource.org/blog/reimagining-data-for-open-source-ai-a-call-to-action/.

67 GIZ, Africa-Relevant Open Datasets: Catalysing Open AI Innovations (2025), https://www.giz.de/sites/default/files/media/els-document/2025-10/giz-africa-relevant-open-datasets-102025-1.pdf.

68 Martin P. Robillard, 'Turnover-Induced Knowledge Loss in Practice', Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (New York, NY, USA), ESEC/FSE 2021, Association for Computing Machinery, 18 August 2021, 1292–302, https://doi.org/10.1145/3468264.3473923.

evaluation, open source AI projects risk stagnation or abandonment, a dynamic commonly referred to as the 'tragedy of the commons.'[69]

Participants in the stakeholder workshop also stressed that sustaining openness in AI is even more complex than in traditional open source software. In software projects, developer communities can often maintain the core codebase. In open source AI, however, sustained openness requires parallel communities around datasets, annotation protocols, and evaluation benchmarks: each of which carries distinct labour, infrastructure, and governance demands. These communities are often fragmented, resource-intensive, and lack mature institutional support. As a result, the operational burden of keeping open source AI viable and high-quality is significantly higher than for conventional open source software.

Financial Barriers

Although open source AI systems may be available without licensing fees, adopting them is far from cost-free. Organisations face significant expenses related to data collection, compute, infrastructure, and specialised talent. Running large models locally or in the cloud can quickly accumulate high compute costs, particularly if systems are not optimised for efficiency. Additional financial burdens arise from the need to recruit and retain skilled professionals who can fine-tune, deploy, and maintain these models — expertise that is scarce and expensive. Customisation and integration for specific domains further add to the total cost of ownership. Even after deployment, sustainable operation requires ongoing investment in updates, security patches and infrastructure maintenance. Funding such expenses is often challenging. Participants in the stakeholder workshop highlighted that the absence of sustainable funding streams makes financial barriers particularly acute in India and the Global South.

Such cost-and-revenue dynamics in open source AI projects also sit uneasily with venture capital priorities, which emphasise rapid scaling and defensible market share rather than shared infrastructure and broad access.[70] Without stable revenue streams or institutional backing, projects risk accumulating "maintenance debt," where critical updates and support lag behind demand. That said, some observers argue that open source approaches could eventually ease cost pressures for AI startups by lowering

69 'Starting an Open Source Project', accessed 9 June 2025, https://www.linuxfoundation.org/resources/open-source-guides/starting-an-open-source-project; '(PDF) Open Source Software Maintenance Process Framework', ResearchGate, ahead of print, 14 December 2024, https://doi.org/10.1145/1083258.1083265.

70 Justin Hendrix, 'How Venture Capital Warps the World', Tech Policy Press, 4 May 2025, https://techpolicy.press/how-venture-capital-warps-the-world.

barriers to entry and improving margins.[71] The immediate reality, however, is that significant financial constraints continue to limit the effective uptake and sustainability of open source AI.

Governance-related Barriers

We use this category to describe the governance and institutional constraints that hinder the effective stewardship of open source AI systems. These barriers emerge when roles, responsibilities, and accountability mechanisms are unclear, or when regulatory structures are not equipped to support shared, community-driven infrastructure.

Assigning Responsibility

A key governance challenge in achieving openness in AI is the difficulty of assigning liability. The decentralised nature of open source AI development, involving multiple contributors across various jurisdictions, complicates traditional accountability structures. When a said harm occurs, it may often become difficult to discern where it has originated from - the initial model training process, future code or dataset modifications or fine-tuning, or downstream alignment or customisation processes. The diffused and layered nature of contributions makes pinpointing a specific actor for liability exceptionally difficult, and in some cases, practically impossible.[72]

Moreover, open source AI projects may often use permissive licenses (such as Apache, MIT[73]), which disclaim warranties and liabilities, shifting responsibility to end-user individuals and organisations. The MIT licence, for instance, specifies that the software is provided "as is," and that the authors may not be held liable for any claims, damages, or other legal consequences arising from its use.[74]

It is also important to emphasise that, historically, this absence of liability has been a core incentive driving open source collaboration. The ability for contributors to share code without assuming legal exposure encourages experimentation and participation. Imposing liability on developers would likely deter contributions and undermine the open source development ethos itself.[75]

71 DC, 'The Deepseek Effect: Why Open Source AI Models Are Good News for VC Returns', Medium, 28 January 2025, https://medium.com/@dcirl/the-deepseek-effect-why-open-source-ai-models-are-good-news-for-vc-returns-3aa732cafb47

72 D. Gragnaniello et al., 'Are GAN Generated Images Easy to Detect? A Critical Analysis of the State-Of-The-Art', 2021 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 5 July 2021, 1–6, https://doi.org/10.1109/ICME51207.2021.9428429.

73 Kruno Golubic, 'What Is MIT License?', Memgraph, 5 June 2023, https://memgraph.com/blog/what-is-mit-license.

74 'The MIT License', Open Source Initiative, accessed 28 April 2025, https://opensource.org/license/mit.

75 Lee Tiedrich et al., 'Open-Source and Open Access Licensing in an AI Large Language Model (LLMs) World', GPAI, 4 March 2024, https://gpai.ai/projects/blogs/open-source-and-open-access-licensing-in-an-ai-large-language-model-world.htm.

Limitations of Licensing Regimes

Licensing frameworks — legal agreements or sets of obligations that define how software, AI models, or related technologies can be used, modified, and distributed — play a crucial role in the context of open source AI. They define the rights and obligations of licensees, reduce uncertainty, lower transaction costs, and create a shared legal language that fosters collaboration.

For developers, licenses can also provide liability shields, offering reassurance that they will not be held responsible for downstream misuse. In recent years, use-restricted or "responsible AI" licenses — such as OpenRAIL-M — have emerged as a way to prohibit harmful applications, including the spread of disinformation targeting vulnerable groups.[76]

However, the effectiveness of such licensing frameworks is often constrained by a range of practical and systemic challenges. We discuss each of these below.

→ **Weak Enforcement**

Even when licenses explicitly restrict certain uses, enforcing these terms, especially across jurisdictions or with downstream users, is complex and often impractical.[77] This challenge is even more pronounced in the case of generative AI models. Once a user has access to a generative AI model, including its weights and inference code, they are able to deploy it on a large scale without alerting the licensor of the model. Unlike API-gated models such as OpenAI's GPT, where large-scale abuse can be detected through statistics available with the creators, such monitoring is not possible for fully open models. Once model weights are released, especially under weakly enforceable or permissive licenses, there are few effective mechanisms to prevent their misuse or appropriation for harmful purposes. Their centralised control allows model access to be revoked or monitored. On the other hand, fully open models are often deployed across decentralised servers and devices with little traceability. In such cases, licensing without accompanying technical, normative, or contractual safeguards offers only limited avenues to prevent or mitigate misuse.

76 Jonathan Cui and David A Araujo, 'Rethinking Use-Restricted Open-Source Licenses for Regulating Abuse of Generative Models', Big Data & Society 11, no. 1 (2024): 20539517241229699, https://doi.org/10.1177/20539517241229699.; 'Stable Diffusion Public Release', Stability AI, accessed 17 June 2025, https://stability.ai/news/stable-diffusion-public-release.

77 Kevin Klyman, 'Acceptable Use Policies for Foundation Models', arXiv:2409.09041, preprint, arXiv, 29 August 2024, https://doi.org/10.48550/arXiv.2409.09041; Interview with a Working Group member.

Additionally, as discussed earlier, it can also be very challenging to determine which model or actor was behind the said misuse. For API-gated models, where server logs can be searched, it is relatively easier to detect the source of the said malicious activity. The lack of such mechanisms for fully open models makes it difficult for licensors to gather evidence that can help them enforce legal liability.

Lastly, even in cases where evidence is available, it may not translate into meaningful or serious sanctions for the responsible actors. For example, the OpenRAIL license proposes ending access as a form of deterrence. While access can be restricted in the case of API-gated models, it is difficult to end access to models that have been made freely available and are accessible across several jurisdictions, servers, and devices.

Weak enforcement also arises from limited regulatory sensitivity to AI-enabled harms. In many deployment contexts, especially where regulatory frameworks are underdeveloped, instances of harmful or abusive content generation frequently go unpenalised.[78]

→ Fragmentation and Overlap

Licensing practices in AI are highly fragmented, with no standardised framework to determine what uses of a model are permitted, restricted, or prohibited. While some developers adopt traditional open source licences, others rely on bespoke acceptable-use policies or terms of service, which vary widely in interpretation and enforceability.[79] The result is a patchwork governance landscape that makes compliance and oversight difficult.[80]

This is further complicated by the modular nature of pre-trained AI models. A single system may involve source code, training data, model weights, and evaluation datasets, each governed by distinct and sometimes incompatible licences. When developers reuse or fine-tune these models, overlapping license terms can generate legal ambiguity and increase compliance costs. For smaller organisations without in-house legal capacity, navigating this complexity becomes particularly burdensome.[81]

78 Luke Munn, 'The Uselessness of AI Ethics', AI and Ethics 3, no. 3 (2023): 869–77, https://doi.org/10.1007/s43681-022-00209-w.
79 Klyman, 'Acceptable Use Policies for Foundation Models'.
80 Daniel McDuff et al., 'On the Standardization of Behavioral Use Clauses and Their Adoption for Responsible Licensing of AI', arXiv:2402.05979, preprint, arXiv, 7 February 2024, https://doi.org/10.48550/arXiv.2402.05979.
81 Moming Duan et al., 'ModelGo: A Practical Tool for Machine Learning License Analysis', Proceedings of the ACM Web Conference 2024, ACM, 13 May 2024, 1158–69, https://doi.org/10.1145/3589334.3645520.

→ Incompatibility with existing licenses

Certain open source licenses, such as GPL v3, require that downstream users retain the same freedoms granted under the original license, including the right to modify, distribute, and reuse the software without additional constraints. Consequently, developers cannot add new restrictions on use or distribution if they were not present in the original GPL terms. This incompatibility limits the applicability of use-restricted licences to any model, dataset, or codebase that has been released under a copyleft licence such as GPL v3.[82]

Taken together, these constraints point to a broader reality: licences in open source AI operate more effectively as normative signals than as mechanisms capable of enforcing responsible behaviour. This interpretation also aligns with guidance from the Digital Public Goods Alliance (DPGA). In its Community of Practice Recommendations for Assessing AI Systems as Digital Public Goods, the DPGA notes that "licenses as a tool are not fit for the purpose of inhibiting harm, given their lack of enforceability."

However, they can be a part of a broader strategy for responsible AI practices. In addition to relying on licences to police misuse, the DPGA recommends complementing them with other forms of risk mitigation. These can include technical measures such as model alignment and watermarking or contractual and institutional safeguards that embed accountability in deployment environments.[83]

82 Danish Contractor et al., 'Behavioral Use Licensing for Responsible AI', Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (New York, NY, USA), FAccT '22, Association for Computing Machinery, 20 June 2022, 778–88, https://doi.org/10.1145/3531146.3533143; 'The GNU General Public License v3.0 - GNU Project - Free Software Foundation', accessed 30 April 2025, https://www.gnu.org/licenses/gpl-3.0.html.

83 John Kirchenbauer et al., 'A Watermark for Large Language Models', Proceedings of the 40th International Conference on Machine Learning, PMLR, 3 July 2023, 17061–84, https://proceedings.mlr.press/v202/kirchenbauer23a.html; Daniel M. Ziegler et al., 'Fine-Tuning Language Models from Human Preferences', arXiv:1909.08593, preprint, arXiv, 8 January 2020, https://doi.org/10.48550/arXiv.1909.08593; Madhulika Srikumar et al., 'Risk Mitigation Strategies for the Open Foundation Model Value Chain', Partnership on AI, 11 July 2024, https://partnershiponai.org/resource/risk-mitigation-strategies-for-the-open-foundation-model-value-chain/.

Risks Involved in the **Development and Use** of Open source AI Systems

Open source AI can expand access, accelerate innovation, and enhance transparency. However, the same openness that enables these benefits can also introduce risks. These include the possibility of misuse, privacy violations, and the reinforcement of power asymmetries within the AI ecosystem.

It is worth noting that many of these risks are not exclusive to open source systems; closed and proprietary AI models exhibit similar concerns.[84] Yet, in open source environments, the risk profile can take a different form. Governance in such ecosystems is distributed rather than centralised. This can often prove to be a double bind. While decentralised mechanisms can enable accountability, as discussed earlier, they can also prove to be inadequate when misuse or security breaches proliferate faster than oversight structures can respond.

The subsections that follow examine these risks more closely and outline how they may affect developers, governments, and the broader ecosystem.

Misuse by Malicious Actors

Historically, open source initiatives have sought to democratise access to advanced technology, enabling knowledge sharing, development, and innovation. However, unrestricted access also lowers the barrier for malicious use. When model weights and code are openly released, non-experts, including actors with harmful intent, can modify, fine-tune, or redeploy systems without meaningful oversight.[85]

A key concern is the loss of control once a model is made public. Open releases are effectively irreversible: weights can be downloaded, mirrored, and redistributed across platforms and jurisdictions.[86] This was seen in the case of Microsoft's Wizard LM 2 model. It was released under an Apache license on Hugging Face but was soon deleted as it had not undergone robust toxicity testing. Although the model was not hosted on Hugging Face anymore, it had been downloaded and reshared, leaving Microsoft unable to retract access or revoke the license.[87]

84 Qi et al., 'Fine-Tuning Aligned Language Models Compromises Safety, Even When Users Do Not Intend To!'
85 Interview with a Working Group member.
86 Eiras et al., 'Risks and Opportunities of Open-Source Generative AI'.
87 Paul Gagnon et al., 'On the Modification and Revocation of Open Source Licences', arXiv:2407.13064, preprint, arXiv, 29 May 2024, https://doi.org/10.48550/arXiv.2407.13064.

Openness can also make misuse easier by enabling repurposing. Predictive models are generally constrained by their design as they generate outputs within a narrower domain. By contrast, generative models can be reconfigured to produce a wide range of content, including harmful, misleading, or deceptive outputs.[88] The subsections below outline the key forms that malicious use can take.

Generating Harmful Content

When core components of an AI system are freely accessible, they may be repurposed to generate harmful or unlawful content. For example, in 2022, a YouTuber trained an AI model on racist, misogynistic, and antisemitic content scraped from 4chan's "/pol/" board and uploaded the model to Hugging Face for public download. Although the platform later restricted access, the incident illustrates how harmful models can be easily created and distributed in open source ecosystems.[89]

Opening components such as model weights also makes it easier to remove safety constraints (for instance, by disabling content filters or prompt safety layers), and enables malicious actors to tailor models for disinformation, harassment, and other harmful uses.[90]

Privacy Violations

Access to models and their internal parameters can also enable privacy attacks. Open access may enable the retrieval of sensitive information contained in the underlying training data through model inversion attacks (MIAs).[91] MIAs, for instance, can be used to extract facial features from the underlying data of a facial recognition system, personal health information from the underlying data of a medical diagnostics system, and personal preferences and interests from targeted advertising systems.[92]

Attackers may also use membership inference attacks to determine whether a particular individual's data was included in the training dataset. In sensitive domains such as healthcare, this can reveal personal information simply by confirming whether a specific record was used during training.[93]

88 Interview with a Working Group member.

89 James Vincent, 'YouTuber Trains AI Bot on 4chan's Pile o' Bile with Entirely Predictable Results', The Verge, 8 June 2022, https://www.theverge.com/2022/6/8/23159465/youtuber-ai-bot-pol-gpt-4chan-yannic-kilcher-ethics.

90 National Telecommunications and Information Administration, Dual-Use Foundation Models with Widely Available Model Weights (National Telecommunications and Information Administration, 2024), https://www.ntia.gov/sites/default/files/publications/ntia-ai-open-model-report.pdf.

91 'Model Inversion Attacks | A New AI Security Risk', Michalsons, 8 March 2023, https://www.michalsons.com/blog/model-inversion-attacks-a-new-ai-security-risk/64427.

92 Zhanke Zhou et al., 'Model Inversion Attacks: A Survey of Approaches and Countermeasures', arXiv:2411.10023, preprint, arXiv, 15 November 2024, https://doi.org/10.48550/arXiv.2411.10023.

93 Hongsheng Hu et al., 'Membership Inference Attacks on Machine Learning: A Survey', arXiv:2103.07853, preprint, arXiv, 3 February 2022, https://doi.org/10.48550/arXiv.2103.07853.

Security Attacks

Unrestricted access to open source AI systems may also facilitate security breaches. Malicious actors can study model architectures, identify weak points, and exploit them through attacks such as data poisoning or the insertion of backdoors. In data poisoning, manipulated inputs are introduced during training so that the model internalises a hidden behaviour or trigger. Backdoor attacks take this further: a small set of poisoned examples allows the model to activate an additional, unintended function when a specific pattern appears in the input.[94]

While open source communities often identify vulnerabilities through collective scrutiny, this is not guaranteed. Openness does not substitute for structured, accountable maintenance. The absence of a dedicated security team, an active open source community or predictable funding can delay patches and create prolonged exposure to risk. The 2014 Heartbleed vulnerability in the widely used OpenSSL encryption library made this clear. The flaw went undetected for months, partly because maintenance capacity had eroded due to resource constraints.[95]

Risk of Capture by Dominant Actors

Despite the promise of democratised access, open source AI does not automatically redistribute power. Developing, deploying, and maintaining state-of-the-art AI models requires specialised talent, and sometimes, vast compute resources and significant financial investment. In any case, these capabilities are currently concentrated within a small number of large technology companies.[96] Even when components of AI systems are made open, these firms can continue to exert control over how the ecosystem evolves, who benefits from openness, and who is able to meaningfully participate.[97] We provide a detailed explanation below.

Infrastructural Dependency

Even when model code or weights are openly released, meaningful participation requires access to cloud infrastructure, GPUs, datasets, and engineering talent. Especially when developing large-scale models, these inputs are controlled by a small set of firms, creating a structural dependency: developers and smaller organisations must rely on corporate cloud platforms, proprietary APIs, or managed services to

94 Younis Al-Kharusi et al., 'Open-Source Artificial Intelligence Privacy and Security: A Review', Computers 13, no. 12 (2024): 12, https://doi.org/10.3390/computers13120311.
95 Brock, Open Source Law, Policy and Practice.
96 Widder et al., 'Open (For Business)'.
97 Open for Good Alliance, 'Open source AI Data Sharing: Yes! Data Colonialism: No!', Medium, 3 November 2023, https://medium.com/@openforgood/open-source-ai-data-sharing-yes-data-colonialism-no-3062a922de03.

train or deploy models. In such cases, openness exists at the software layer, but monopolistic control persists at the infrastructure layer.

Control through Standard-setting

Larger technology firms can strategically shape the direction of open source AI not by owning all components, but by defining the standards everyone else must follow. By open-sourcing key frameworks (such as PyTorch and TensorFlow), they set the technical architecture around which others build. This anchors the ecosystem to their platforms, enabling them to retain control over the most valuable layers such as compute, data, and hosting, while also benefiting from community contributions.[98]

Ability to Extract Value

Since large tech firms already control the infrastructure and distribution channels, they are better positioned to commercialise open source projects, offering paid hosting, proprietary add-ons, or vertically integrated services around open models. These companies also generate intellectual rents by capturing open data and transforming it into proprietary assets to reinforce their competitive advantage.[99]

Absorption of Emerging Competitors

For smaller model developers, the prospect of acquisition by larger corporations represents a risk. Open source innovation can be vulnerable to "killer acquisitions" or strategic partnerships between large firms and promising open source labs. These arrangements give larger firms a coordinating role in development direction, often without formal acquisition. This reduces competitive diversity and limits the emergence of independent alternatives.[100]

98 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," Open Markets Institute and Mozilla Foundation, https://blog.mozilla.org/wp-content/blogs.dir/278/files/2024/10/Stopping-Big-Tech-from-Becoming-Big-AI.pdf; '(PDF) Big Tech, Knowledge Predation and the Implications for Development', accessed 17 June 2025, https://www.researchgate.net/publication/346677643_Big_Tech_knowledge_predation_and_the_implications_for_development.

99 Ravindra Kirti Founder GGF, 'Why Big Tech Is in a Tizzy Over Open Source AI Models', GlobalGrowthForum, 28 August 2024, https://globalgrowthforum.com/why-big-tech-is-in-a-tizzy-over-open-source-ai-models/.

100 'Big Tech-Small AI Partnerships | Oxford Law Blogs', 19 July 2024, https://blogs.law.ox.ac.uk/oblb/blog-post/2024/07/big-tech-small-ai-partnerships.

Navigating the Trade-offs between Opportunities and Risks of Open Source AI

A variety of challenges and risks may accompany the opportunities offered by open source AI, presenting complex trade-offs for stakeholders to navigate. While some challenges may be solvable, others may manifest as deadlocked trade-offs wherein certain risks may be accepted in order to take advantage of certain attractive opportunities, or certain benefits may be foregone if the risks associated are too high. Since the nature, likelihood, and severity of such risks continue to remain contested, we conceptually map the interlinkages between open source AI's benefits and risks in this section. In practice, these trade-offs will be shaped by the unique motivations, incentives, and capacities of different actors such as governments, global tech companies, Indian start-ups, developers, and end users. Unpacking each of these trade-offs is therefore crucial for designing governance approaches that balance competing interests while maximising value for all stakeholders.

Ease of **Innovation** versus **Misuse**

Open source AI systems enable developers to innovate more easily and accessibly. However, open sourcing may also pave the way for misuse by malicious actors, as it lowers the entry barriers to replicating, customising, or integrating available models or datasets.[101]

Implications for Different Stakeholder Groups

Government

The conversation around innovation versus misuse is not limited to open source AI and is part of a larger ongoing debate on the role governments play within the AI ecosystem. In India, the government has come to assume a critical role in shaping the AI industry through access to computing infrastructure, research funding, and skilling programmes.[102] At the same time, regulatory frameworks to mitigate and address AI risks remain under development, with the government wary of stifling innovation through premature regulation.[103]

This debate may get further complicated in the context of open source AI. On one hand, open source can enable better self-regulation, not only at the enterprise level but at the community level. Openness and transparency can support distributed oversight, enabling better discovery and monitoring of system vulnerabilities and reducing the burden placed on top-level or centralised regulation. In this regard, enabling more openness in the AI ecosystem is in alignment with the government's current light-touch approach to AI regulation, which is reliant on non-binding guidelines or voluntary adoption of ethics codes.[104]

On the other hand, openness can also amplify the risk of misuse or security breaches within the AI ecosystem. However, the manifestation of these risks, particularly within the Indian context, has not yet materialised at scale and will depend on the governance models instituted within open source communities, their level of documentation and transparency, and the robustness of licensing protocols.

101 Dominik Hintersdorf et al., 'Balancing Transparency and Risk: The Security and Privacy Risks of Open-Source Machine Learning Models', arXiv:2308.09490, preprint, arXiv, 18 August 2023, https://doi.org/10.48550/arXiv.2308.09490.
102 'INDIAai | Pillars', IndiaAI, accessed 17 June 2025, https://indiaai.gov.in/.
103 Shaoshan Liu, 'India's AI Regulation Dilemma', 2023, https://thediplomat.com/2023/10/indias-ai-regulation-dilemma/.
104 'Use of AI: MeitY Readies Voluntary Ethics Code for Artificial Intelligence Firms', The Times of India, 19 November 2024, https://timesofindia.indiatimes.com/business/india-business/use-of-ai-meity-readies-voluntary-ethics-code-for-artificial-intelligence-firms/articleshow/115433686.cms.

Large Technology Corporations

Large technology companies play a prominent role within open source AI ecosystems, particularly in providing access to pre-trained models that are otherwise prohibitively expensive for smaller actors to develop independently. For example, Meta's open-weight LLaMa models have been used to build a range of downstream applications. The trade-off between such openness and the risk of misuse may also motivate developers to not release their models altogether.[105] In case there is rampant misuse of available models and their accompanying documentation, companies may respond by limiting future access or modifying their release strategies altogether. For instance, after Meta's LLaMA model and deployment methods were leaked on 4chan, company leadership indicated that in the face of increased safety risk, future model releases may be restricted to smaller groups or "known academic partners with very strong credentials".[106]

Short-term **Monetisation** versus Long-term **Community Building**

Many companies weigh the trade-off between securing profits through proprietary systems and licensing fees, and investing in the open source community. As discussed earlier, maintaining and sustaining open source systems can be costly. When companies withdraw their support, the viability of these systems often declines.[107] An increasing number of open source projects are being made proprietary as a strategy to ensure profitability and sustainability, a trend that has drawn criticism from the open source community and developers involved in these initiatives.[108]

105 Will Douglas Heaven, 'The Open source AI Boom Is Built on Big Tech's Handouts. How Long Will It Last?', MIT Technology Review, accessed 15 June 2025, https://www.technologyreview.com/2023/05/12/1072950/open-source-ai-google-openai-eleuther-meta/.

106 Will Douglas Heaven, 'The Open source AI Boom Is Built on Big Tech's Handouts. How Long Will It Last?'

107 Yuxia Zhang et al., 'Corporate Dominance in Open Source Ecosystems: A Case Study of OpenStack', Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, ACM, 7 November 2022, 1048–60, https://doi.org/10.1145/3540250.3549117.

108 The Trade-Offs of Open Source Going Private - Nocturnalknight's Lair, 26 December 2024, https://nocturnalknight.co/the-trade-offs-of-open-source-going-private/.

Implications for Different Stakeholder Groups

Large Technology Corporations

Large corporations are often the most prominent actors supporting open source initiatives.[109] Such corporations have come to establish themselves as providers of foundational platforms on which entrepreneurs and developers innovate. As discussed earlier, these initiatives allow them to create a broad ecosystem where successful innovations can be acquired or integrated, extending their influence and future growth opportunities. AI provides scope to identify and capitalise on new opportunities, making open source a strategic move to stay at the cutting edge of technology.[110] It positions these companies to profit from cloud hosting, infrastructure, and ancillary services where the open source tools run.[111]

At the same time, firms may shift toward proprietary releases to preserve competitive advantage and create predictable revenue streams from licensing and enterprise partnerships. OpenAI illustrates this shift clearly. Although founded as a non-profit with a commitment to openness, transparency, and public benefit, the organisation gradually moved toward a closed and commercially gated model. Beginning with GPT-3 and subsequent generations, detailed model information ceased to be released. Openness was replaced with selective disclosure and API-based access, signalling a transition from open research to proprietary control.[112]

Small Start-ups

Smaller organisations may be motivated to open source their tools to enable more visibility as well as attract more collaborators. However, such visibility can come at the cost of financial viability. Without robust monetisation models or IP protections, smaller start-ups face the risk of being outpaced or undervalued, especially if large actors — arguably more well-resourced and discoverable — adopt, scale, or integrate their tools into their existing product offerings.

109 Will Douglas Heaven, 'The Open source AI Boom Is Built on Big Tech's Handouts. How Long Will It Last?'
110 Patrick Shafto, 'Why Big Tech Companies Are Open-Sourcing Their AI Systems', The Conversation, 22 February 2016, http://theconversation.com/why-big-tech-companies-are-open-sourcing-their-ai-systems-54437.
111 Adrian Bridgwater, 'The Impact Of The Tech Giants On Open Source', Forbes, accessed 17 June 2025, https://www.forbes.com/sites/adrianbridgwater/2019/09/07/the-impact-of-the-tech-giants-on-open-source/.
112 OpenAI: Was the Shift to Closed Source Justified? – Hadron, 3 February 2021, https://sites.imsa.edu/hadron/2021/02/03/openai-was-the-shift-to-closed-source-justified/.

Reduced Vendor Dependence versus Ease of Integration

Open source AI reduces vendor lock-in by enabling organisations to adopt, customise, and maintain systems without being tied to a single provider's pricing, roadmap, or commercial priorities. By contrast, proprietary vendors often offer end-to-end services, including integration, support, and upgrades. For many actors, including government departments looking to procure AI systems, this convenience can make proprietary solutions attractive, even though it creates long-term dependence on a single provider.

Implications for Different Stakeholder Groups

Government

Adopting open source AI enables governments to avoid vendor lock-in. While it can be appealing to have a single vendor provide streamlined services, it may be difficult to switch vendors if issues arise. Dependence on a single vendor may also result in government actors being unable to migrate away from the technology roadmap of that vendor, essentially making them captive to their decisions.[113]

However, easily obtaining streamlined services and solutions offered by proprietary developers may be attractive to government stakeholders who are deploying AI technology at a large scale. Proprietary vendors may offer one-stop procurement and maintenance, which can be administratively attractive. In this case, the open, flexible, and voluntary structure of open source may discourage government stakeholders who may value ease of access and reliability over flexibility.

Further, large technology corporations may also provide philanthropic support and cloud/GPU credits that often come bundled with their AI solutions. Stakeholder consultations highlighted how several large technology firms also extend investment partnerships that accompany the AI solutions they offer: these might include infrastructure support, R&D investments, and skilling programmes. Together, these form an attractive package for governments. At present, open source alternatives are unable to offer such comprehensive bundles of products and services.

113 Michelle Laymon, 'The Hidden Costs of Vendor Lock-In: Why Open Source Value…', Suse, 5 May 2025, https://www.suse.com/c/the-hidden-costs-of-vendor-lock-in-why-open-source-values-matter/.

It is worth noting that while open source AI systems reduce vendor lock-in and offer adaptability to local needs, the government's choice of which system to procure is often influenced by perceptions as much as by technical considerations. Stakeholder consultations revealed that government actors can often hold misconceptions about open source AI. For instance, they may equate open with insecurity, assuming that data would be more vulnerable in open source systems. They may also conflate free with costless and therefore unreliable, assuming that no actor will be accountable for long-term maintenance or support. Such perceptions discourage adoption and reinforce reliance on proprietary vendors, even in cases where open source solutions may be viable. Addressing these barriers requires not only building technical capacity but also sustained communication and demonstration projects that clarify how open source can be secure, accountable, and cost-effective when properly implemented. It is against this backdrop that the brief proposes a decision matrix that can guide the government's procurement processes for AI systems (see Table 3).

Small start-ups

For early-stage start-ups, open source AI offers flexibility and avoids costly licensing arrangements. Open tools also allow rapid prototyping and reduce the risks associated with being tied to a vendor that may change pricing, restrict access, or discontinue products.

However, adopting open source solutions requires internal technical capacity. Organisations without prior experience may struggle due to limited documentation, unclear maintenance responsibilities, or the absence of streamlined onboarding processes. Similar to government departments, the trade-off lies between autonomy and the convenience of vendor-provided services.

Future Pathways: Recommendations for Open Source AI in India

Policy Directions for Open Source AI in India

The preceding sections of this brief highlight the myriad features, opportunities, and challenges associated with open source AI technologies. Particularly with respect to the government's role, policy debates in this domain are rich and multifaceted. On one hand, open source AI holds the potential to enhance transparency in AI systems deployed in the public sector, reduce dependence on foreign vendors, and catalyse innovation through shared infrastructures. At the same time, governments face legitimate concerns about the long-term sustainability of open source AI systems and the technical and institutional capacity required to implement and maintain such systems effectively.

These complex challenges underscore the need for policy responses that are adaptive and contextual, particularly within a rapidly evolving AI ecosystem such as India's. Rather than viewing openness as an all-or-nothing choice, this brief advocates for an affirmative, outcomes-driven approach to open source AI. The goal is not openness for its own sake, but openness as a means to enable public value: through shared infrastructure, reusability, local adaptability, and collective oversight.

To fully realise this potential, India requires a balanced policy approach, one that enables openness where it creates the most value, while also embedding the institutional support, quality safeguards, and incentives necessary for the long-term success of an open source AI approach.

Against this backdrop, the recommendations set out below propose differentiated pathways through which Indian policymakers can shape the future of open source AI in India. These recommendations have been co-developed and validated through consultations with stakeholders from government, industry, academia, and civil society.

The State as a Promoter of Open source AI

The Indian government, both at the centre and state levels, already plays a pivotal role within the AI ecosystem. The government is not only shaping but also actively creating the domestic AI market. Through the IndiaAI Mission and its seven pillars, it provides compute-related infrastructure, curates and makes available relevant training datasets, and defines the broader set of norms underpinning the AI ecosystem in India.

Within such a set-up, the government can extend its market-shaping role to incorporate explicit provisions for open source AI, ensuring that openness is embedded as a consideration across its investment, infrastructural, and innovation initiatives. Such initiatives can be modelled after the European Digital Infrastructure Consortium, which aims to support the open source community in Europe by facilitating access to funding and supporting development and scaling.[114] In this regard, we propose the following pathways:

Supporting Open source AI Projects through Existing Compute Allocation Public Initiatives: Existing government initiatives that provide compute resources to AI developers could incorporate dedicated provisions for open source AI projects. Allocation processes can be designed to incentivise openness by granting preferential access to compute or additional resources to projects that commit to open-sourcing their code, models, and/or datasets. Such an approach would also lower barriers for smaller open source teams, who often face acute constraints in accessing adequate compute.[115]

114 European Digital Infrastructure Consortium, 'Digital Commons European Digital Infrastructure Consortium (DC-EDIC)', European Digital Infrastructure Consortium, 2024, https://digital-strategy.ec.europa.eu/en/policies/edic.

115 It is worth noting that, at present, demand for government-provided compute remains lower than available supply. However, as one government stakeholder observed during our interviews, when the competition for these resources increases, the allocation processes are likely to privilege AI projects that incorporate open-source components.

Working Towards Longer-term Sustainability of Open Source AI Systems: A critical challenge for open source AI lies not in prototyping but in sustaining and scaling projects beyond the proof-of-concept stage. Unlike proprietary models backed by commercial revenue, open projects often depend on fragmented or short-term funding, leaving many of the system's components vulnerable to neglect.

To address this, the government, in partnership with private-sector and/or philanthropic actors, could establish grant schemes or blended finance mechanisms targeted at long-term maintainers of high-value, high-relevance open source AI tools, datasets, and code libraries. These initiatives can be modelled after existing programmes such as Germany's Sovereign Tech Fund, through which the government invests in open software components, particularly those that form critical digital infrastructure in the country.[116]

Providing Special Designations to Smaller Open source AI Firms: Government procurement processes and financing mechanisms can often disadvantage smaller, less recognised AI start-ups, particularly those working with open source models and tools. Current certification schemes for micro, small, and medium enterprises (MSMEs) in India create some exemptions and targeted support for smaller firms, but these have not yet been adapted to reflect the specific needs of digital and AI-based firms.

To address this gap, policymakers could extend MSME recognition frameworks to explicitly include firms engaged in open source AI development and maintenance. Such a designation would enable these firms to benefit from concessional finance, procurement preference, and capacity-building schemes, while lowering credibility barriers that currently prevent their participation in public tenders.

Building Awareness through Information Campaigns: In its role as a promoter of openness in AI, the government (in particular, the IndiaAI Mission) can play a critical role in building awareness about open source AI among different government departments, tech practitioners, and downstream adopters of AI models. Such efforts can include disseminating information about the benefits of open source AI, the range of tools and resources already available for adoption and reuse, and the pathways through which public and private actors can engage with them. The Mission can provide short courses or public repositories that familiarise officials, developers, and researchers with open datasets, models, and licensing frameworks. These initiatives can help mainstream open source literacy, particularly within the rapidly growing ecosystem of AI practitioners and adopters.

116 'Home', Sovereign Tech Agency, 29 October 2025, https://www.sovereign.tech/.

Integrating Open Source AI into National Innovation Initiatives: The government can strengthen its promotional role by scaling and institutionalising ongoing efforts that integrate open approaches within national innovation programmes. Recent examples such as the GenAI 2025 Hackathon, convened by the Indian Institute of Science and supported by the IndiaAI Mission, which encourages experimentation with open models and datasets, demonstrate the growing momentum for such integration.[117]

Building on this progress, national platforms such as Startup India, or the IndiaAI Innovation Challenge could include dedicated open source AI tracks, with specific incentives for participants to reuse, adapt, and contribute to open datasets, models, and tools.[118]

Such efforts can help normalise open source AI as a legitimate and high-value pathway for research and entrepreneurship, showcasing its role in driving innovation and creating public value within India's AI ecosystem.

The State as a Regulator and Standard Setter for Open Source AI

Preventing Open-washing in Publicly-funded Open Source AI Projects: Given the risk of open-washing, publicly-funded AI initiatives that seek to be open must be guided by clearly defined thresholds of openness.

This concern is particularly relevant in India, where the current AI governance framework — as reflected in the 2025 IndiaAI Governance Guidelines — places significant emphasis on voluntary measures and self-regulatory practices.[119] While these guidelines encourage industry actors to "publish transparency reports that evaluate the risk of harm to individuals and society in the Indian context", they are neither binding nor do they articulate clear expectations regarding the scope or level of detail of such reports.

With disclosure obligations largely left to organisational discretion, there is a high risk of uneven or selective compliance, a concern the guidelines themselves acknowledge. To mitigate this risk, we recommend defining baseline transparency requirements for projects funded by or deployed in the public sector. Stakeholder workshop discussions further suggested that such requirements can include

117 'Centre for Networked Intelligence', accessed 5 November 2025, https://cni.iisc.ac.in/hackathons/gen-AI-2025/.
118 'IndiaAI Innovation Challenge Launched to Foster Impactful AI Solutions Inviting Applications to Build AI Solutions for Critical Sectors', accessed 5 November 2025, https://www.pib.gov.in/www.pib.gov.in/Pressreleaseshare.aspx?PRID=2056991.
119 MeitY, India AI Governance Guidelines (2025), https://static.pib.gov.in/WriteReadData/specificdocs/documents/2025/nov/doc2025115685601.pdf.

disclosures related to source code, model weights, bias and fairness assessments, evaluation datasets, documentation of data collection protocols, and, where feasible, access to training and fine-tuning datasets.

Although such transparency requirements cannot be fully generalised across sectors, given differences in legal obligations, risk profiles, and deployment contexts, it is nonetheless in the public interest to articulate baseline disclosure expectations for publicly funded AI systems.

Defining such thresholds would help prevent superficial disclosures from being misrepresented as openness and ensure that public investments in AI deliver genuine transparency, accountability, and public value.

Consistent with the IndiaAI Governance Guidelines' recognition that voluntary compliance depends on clear and actionable guidance, India's upcoming AI Safety Institute could serve as a focal institution for developing guidance notes or model transparency frameworks that articulate baseline, non-negotiable disclosure expectations for publicly funded AI projects.

Convening a Community-led Effort to Develop Contextual Licensing Frameworks for Open Source AI: Licensing plays a foundational role in enabling open source collaboration, protecting contributors, and governing downstream use. However, most existing open source licenses, whether permissive or use-restricted, are not tailored to the layered, modular nature of AI systems. They also lack clarity on how usage rights and obligations apply across the AI pipeline, particularly in contexts like India, where state capacity for oversight is uneven and legal interpretations may differ.

That said, some experts in the working group emphasised that existing licensing regimes may already provide sufficient flexibility to govern AI development and deployment when appropriately combined or adapted. Frameworks such as Apache, Creative Commons, and RAIL include elements such as attribution requirements, use-based clauses, and redistribution controls that can, if layered thoughtfully, accommodate the composite architecture of AI systems. From their perspective, the challenge lies less in the absence of appropriate licenses and more in the need for clear guidance, harmonisation, and support on how these licenses can be applied consistently across different AI components.

To explore these questions in more detail, the government can convene a multi-stakeholder working group or initiate consultation

with leading industry experts and civic tech organisations. Such a process could assess the suitability of existing licenses for AI, identify gaps that warrant adaptation, and, where necessary, propose modular, India-specific templates aligned with the country's legal and institutional landscape.

This consultation could also serve as a venue to articulate what responsible use might mean in the Indian context, taking into account locally salient risks such as surveillance, misinformation, and marginalisation. India has precedent for similar efforts, wherein MeitY convened expert committees on open standards, with the aim to define key guidelines for adoption of such standards in e-governance.[120] A comparable mechanism could be initiated under the aegis of IndiaAI, MeitY's Standardisation Testing and Quality Certification Directorate (STQC). This would ensure legitimacy, continuity, and alignment with broader digital public infrastructure and AI governance efforts.

By investing in this kind of norm-setting procedure, without necessarily becoming the licensor or enforcer itself, the state can enable responsible open source AI ecosystems while respecting the autonomy and expertise of the developer community.

Governing Use and Deployment Beyond Licensing: Licensing is a foundational mechanism for governing open source AI. It defines how software, models, datasets, and related components can be used, modified, and redistributed. In the context of AI, licenses help clarify usage rights across different pipeline components, support value sharing by crediting contributors, and signal ethical boundaries through purpose-limited or use-restricted clauses.

However, as discussed earlier, there are also several limitations that licenses exhibit when applied to AI systems. Model pipelines often combine components with conflicting licenses. Permissive licenses like MIT or Apache 2.0 typically waive liability, and enforcement of use restrictions remains weak. Once model weights are released, they are difficult to retract or monitor. Use-restricted licenses like OpenRAIL aim to address this gap, but without robust oversight, they offer little more than normative signalling.

As such, while licensing is useful for clarifying ownership, enabling collaboration, and mitigating implementation harms (such as bugs or bias), it cannot by itself ensure the responsible use of open source AI

120 Ministry of Electronics and Information Technology, 29 March 2007, https://egovstandards.gov.in/sites/default/files/2021-07/Specialist%20Committee%20For%20the%20Policy%20Framework%20on%20Open%20Standards.pdf

systems, particularly when it comes to addressing use-based harms (such as the deployment of open models in surveillance systems, disinformation campaigns, or deepfake generation).[121]

Given these limitations, India's policy approach must avoid treating licensing alone as a sufficient safeguard, and additionally focus governance efforts at the point of deployment, particularly for high-risk use-cases or contexts. High-stakes deployments (such as healthcare or financial applications) of general-purpose AI models should therefore be subject to regulatory oversight and deployment-stage controls, and accompanied by additional safeguards to mitigate misuse and protect sensitive data. Sector-specific guidance can play an important role in clarifying permissible uses of open or widely available models in regulated domains, ensuring that openness at the model level does not translate into unchecked or inappropriate deployment in high-risk settings.

Specifically with respect to maintaining security in open source AI systems, the government can also launch "bug bounty" programs that provide monetary rewards to experts who volunteer to report bugs, security flaws, or other system vulnerabilities in open source AI systems.[122]

Extending MeitY's Quality Assurance Frameworks to AI: The Standardisation Testing and Quality Certification (STQC) Directorate under MeitY already provides testing and certification services that function as quality assurance signals for software and digital systems.[123] A similar assurance-oriented approach could be applied to AI systems at the point of deployment, with certification focused on deployment-relevant parameters such as robustness, cybersecurity, accessibility, and reproducibility, calibrated to the specific use case and sector.

Such certification can be understood as a voluntary assurance mechanism, rather than a regulatory approval or a prerequisite for market-entry. Used in this way, STQC-style certification could inform deployment and procurement decisions across both public and private sectors. In publicly funded deployments, embedding such assurance signals within procurement processes can help address concerns

121 David Gray Widder et al., 'Limits and Possibilities for "Ethical AI" in Open Source: A Study of Deepfakes', 2022 ACM Conference on Fairness Accountability and Transparency, ACM, 21 June 2022, 2035–46, https://doi.org/10.1145/3531146.3533779.

122 'Research Report: Bug Bounties and FOSS: Opportunities, Risks, and A…', Sovereign Tech Agency, accessed 5 November 2025, https://www.sovereign.tecopen-sourceh/publications/bug-bounties-and-foss.

123 Ministry of Electronics and Information Technology (MeitY), Standardisation Testing and Quality Certification (STQC) Directorate, Government of India, https://www.stqc.gov.in.

around system reliability and production readiness. This is particularly relevant for AI systems built using open source components, which are often perceived as less credible or less deployment-ready within public institutions. At the same time, this approach also helps avoid blanket or upstream certification requirements that could slow experimentation or adoption of AI models elsewhere in the ecosystem.

The State as a Procurer and User of Open source AI

As a large institutional adopter and consumer of AI systems in the domestic market, the government's procurement choices will decisively shape the ecosystem. Rather than blanket mandates, procurement frameworks can be structured to reward openness while still accommodating proprietary systems where no viable open alternative exists.

Embedding Openness as a Consideration in Procurement Frameworks: Open source AI can be attractive to governments because of its lower financial barriers, reduced vendor lock-in, adaptability to local needs, and inherent transparency. These advantages echo the Ministry of Electronics and Information Technology's Policy on Adoption of Open Source Software for Government of India, which identified open source software as an innovative, cost-optimising alternative to proprietary tools. A similar orientation could be applied to AI systems deployed in public governance.

That said, it is worth recognising that procurement choices must account for the fact that open source AI solutions are not always production-ready or suited to high-risk, time-sensitive contexts. Proprietary systems may provide faster deployment, vendor-backed maintenance, and established safeguards, which can be preferable in certain cases.

Technology decisions are best made with reference to the defined purpose and total cost of ownership across the project lifecycle, rather than prescribing one category of solution. In this approach, if open source is better suited, it should be adopted; if proprietary offers advantages, it may be preferred; and if both options are equivalent, openness can be prioritised. This kind of context-sensitive approach can be operationalised through a decision matrix that guides procurement officers on when to privilege open source systems versus proprietary ones. We present a non-exhaustive list of considerations that can go into such a decision matrix in Table 3.

↓ **Table 3**
Considerations for AI Procurement: Open Source versus Proprietary AI

Specific Considerations	Open source AI Systems	*Notes*	Proprietary AI systems	*Notes*
Sensitivity of data	**Data can be stored within government systems, giving authorities full control over storage and access.**	Retaining such autonomy requires strong in-house security and infrastructure.	**Data may need to be shared with third-party vendors, though some offer "sovereign cloud" or on-premise deployments.**	These still raise dependency and sovereignty concerns, since underlying infrastructure often remains foreign-controlled.
Resource constraints	**No licensing fees; potentially suitable where budgets are tight.**	Procuring open source AI models may still require upfront investment in skills and compatible infrastructure.	**Suitable when resources are ample: vendor support reduces immediate skill requirements.**	Recurring licensing costs apply, though short-term credits or philanthropic support can offset expenses; these may create future dependency risks
Need for customisation and adaptability	**Open source AI models can be highly adaptable to local languages, contexts, and needs.**	Adapting to local needs can be costly as it requires deep domain expertise and high-quality local datasets.	**Comes with predetermined technology roadmaps/ integration guides. These may suffice where customisation is not needed**	Vendor priorities may not align with local requirements.
Risk Mitigation	**Transparency enables better scrutiny, making it easier to identify risks and build safeguards. Open source AI models may work well where risks are known and safeguards exist.**		**Controlled deployment can help manage uncertainty where risks are less well understood (unknown-unknowns).**	Opacity in such systems can make systemic harms like bias or embedded assumptions harder to detect.
Time to deploy	**Deployment timelines vary: pre-trained open models and skilled teams can accelerate adoption, but in their absence, significant effort may be required for integration, fine-tuning, and testing, which can slow down deployment.**		**Best for urgent, turnkey deployment needs.**	
Maintenance requirements	**Dependent on community/ government support, continuity is not guaranteed.**	Openness allows the procurer to easily seek and accept new vendors if community maintenance efforts are not satisfactory/ available as needed.	**Vendor-backed continuity, though at risk if the vendor pivots or exits.**	Procurer becomes vulnerable to vendor lock-in.

Hence, a blanket mandate for open source AI procurement is neither feasible nor desirable. Instead, procurement frameworks should reward openness through additional points in Quality- and Cost-Based Selection (QCBS) models, while retaining the flexibility to accommodate proprietary solutions where appropriate. However, some experts in our stakeholder consultations highlighted that even when such preferential provisions for open source technologies are put in place, they often face implementation challenges. To strengthen compliance, procurement frameworks could therefore incorporate accountability mechanisms, such as requiring procuring agencies to justify departures from open source options where feasible. Such a mechanism can be embedded through amendments to the General Financial Rules, 2017, a compilation of rules and orders for managing the finances of government departments and organisations in India.

Relaxing Prior-Deployment Requirements: A common barrier for smaller open source developers is the requirement to demonstrate large-scale prior deployments, which they often lack. Procurement rules could be modified to allow pilot deployments or third-party audits as substitutes for prior experience. This reform would reduce credibility hurdles, allowing innovative open source solutions to enter government tenders without compromising risk management.

Building Procurement Capacity: Procurement officers often lack the technical expertise to evaluate open source AI offers, particularly in terms of licensing and sustainability. A dedicated module on open source AI within the iGOT Karmayogi platform, focused on demystifying openness and interpreting licensing terms, could build this capacity across the bureaucracy. While secondary to procurement reform itself, capacity-building would strengthen the state's ability to implement openness provisions effectively.

Enhancing Transparency and Auditability in Public Sector AI: Open source AI offers a comparative advantage when it comes to transparency and accountability, since its underlying code, models, and datasets can be more readily inspected and audited than proprietary systems. This advantage is particularly critical in the public sector, where concerns about opacity are acute and impacts on rights and citizen welfare can be most direct. At present, information on government use of AI in India is limited, with few institutionalised mechanisms for disclosing where and how systems are deployed, how they are evaluated, or what safeguards accompany them.

Establishing a transparency baseline would address this gap. Such a baseline could begin with simple registries of AI systems procured or deployed by the government, and gradually extend to include disclosure of evaluation results, risk assessments, and underlying datasets in the case of high-risk applications.

Modelling Best Practices through its Role as a Developer

Alongside its roles as a promoter, regulator, and procurer, the Indian state also acts directly as a developer of AI systems. For instance, through institutions such as the Bhashini programme, the government is actively building models, datasets, and platforms that serve as public infrastructure.

In this role, the state has a unique opportunity as well as responsibility to model best practices in openness. This includes releasing state-developed models and datasets under transparent and well-governed licenses, documenting training data and methods, and ensuring that such resources are accessible to researchers, startups, and smaller firms. For example, in the case of government-hosted datasets, many open data platforms rarely maintain historic records of downloads or usage beyond a few months, reducing both accountability and trust.

To address these gaps, such data platforms should adhere to minimum documentation and quality standards, covering provenance, licensing terms, scale, filtering processes, and privacy safeguards. Embedding such standards into repositories like AI Kosh would enhance transparency, enable more effective monitoring, and increase the credibility, reproducibility, and adoption of open source AI systems built from them.

By embedding openness into its own development efforts, the government can set a benchmark for the wider ecosystem and reinforce India's digital sovereignty objectives. In India, creating such requirements would both improve accountability in public sector AI use and reinforce the attractiveness of open source approaches, which are often better positioned to meet transparency standards in practice.

Recommendations for the **AI Development Community**

Throughout this brief, we have identified the challenges, risks, and trade-offs that arise not only for organisations considering the adoption of open source AI models but also for developers and institutions releasing models or datasets into the open. For adopters, the decision to rely on open source solutions involves navigating uncertainties around quality, documentation, security, and long-term support. For those publishing AI components openly, the challenges are different: deciding what to open, how to safeguard against misuse, how to ensure responsible stewardship and sustainability, and how to create meaningful openness rather than symbolic disclosure.

At present, these decisions are shaped by a combination of sectoral regulations, horizontal legal frameworks, and organisation-specific practices. Larger developers with control over model training or data pipelines have begun to establish their own artefacts and processes, such as model cards, system documentation, and internal safety tooling.[124] These practices are largely voluntary and vary in scope and depth across organisations. Many downstream developers rely on upstream models and datasets and make use of the documentation and artefacts provided by upstream developers, while still needing to exercise judgement when adapting or deploying these systems in specific contexts.

It is against this backdrop of differentiated roles and distributed responsibilities across the AI value chain that the following recommendations are framed. They focus on the responsibilities and design practices of those who build, train, publish, or maintain open source AI models and datasets, as well as those who integrate and deploy them in downstream applications. Drawing on insights from technical experts and the stakeholder workshop, these recommendations are not intended to be exhaustive. Rather, they offer a practical starting point for addressing the design and stewardship choices that currently shape and constrain the responsible adoption and use of open source AI.

124 For example, Meta's releases of the LLaMA family of models have been accompanied by system cards, Responsible Use Guides, and technical papers. Together, these materials document training processes, evaluation methodologies, safety considerations, intended use, and relevant safeguards against misuse. For more details, see: https://ai.meta.com/blog/meta-llama-3-meta-ai-responsibility/ | https://www.llama.com/llama-protections/

Data[125] Contributors

Increasing transparency through robust metadata and paradata documentation:

Data Cards with Verifiable Checks: Use data cards to declare the purpose, methods, and conditions of collection.[126] These should go beyond descriptive notes to include minimal, programmatically verifiable checks such as consent logs, QC pass/fail markers.

Annotation and Process Transparency: Share the annotation guidelines, instructions given to data workers, and quality-control procedures.[127] Without this, downstream users cannot replicate or critically assess the production pipeline, even if the final dataset is technically open.

Synthetic Data Disclosure: Clearly flag where synthetic data is used and explain its intended role in addressing fairness, coverage, or balance issues. This guards against false assumptions of "natural" representativeness.[128]

Coverage and Representativeness Metrics: Embed coverage metrics as metadata, linking datasets to administrative units, postal codes, or comparable markers. This surfaces geographic or socio-linguistic gaps that otherwise remain hidden.[129]

Automated Toolkits for Metadata Creation: Use existing tools that can help automate common metadata or paradata documentation tasks such as pre-filling schema fields, flagging missing labels, or aligning with existing cataloguing standards.[130]

Cataloguing and Accessibility: Invest in organising and cataloguing practices that prevent data disarray. Metadata is not only about disclosure; it also enables searchability, discoverability, and interoperability across projects.[131]

125 Open-source datasets released either on a standalone basis or as part of a model release, as applicable.
126 Mahima Pushkarna et al., 'Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI', Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (New York, NY, USA), FAccT '22, Association for Computing Machinery, 20 June 2022, 1776–826, https://doi.org/10.1145/3531146.3533231.
127 Pushkarna et al., 'Data Cards'.
128 Peter Lee, 'Synthetic Data and the Future of AI', SSRN Scholarly Paper no. 4722162 (Social Science Research Network, 10 February 2024), https://papers.ssrn.com/abstract=4722162.
129 Rie Kamikubo et al., 'Data Representativeness in Accessibility Datasets: A Meta-Analysis', ASSETS. Annual ACM Conference on Assistive Technologies 2022 (October 2022): 8, https://doi.org/10.1145/3517428.3544826.
130 Anuar Ustayev, 'AI-Driven Metadata Enrichment in Open Data Portals: A Deep Dive', Datopian, 29 January 2025, h,ttps://www.datopian.com/blog/ai-driven-metadata-enrichment-in-open-data-portals-a-deep-dive?utm_source=chatgpt.com.
131 Riccardo Albertoni et al., 'The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake', arXiv:2303.08883, preprint, arXiv, 15 March 2023, https://doi.org/10.48550/arXiv.2303.08883.

Addressing misuse through measures such as:

Licensing: To help deter misuse of datasets being made open, they can be released under licenses such as OpenRAIL-D or RAIL-D that allow broad access but still help prohibit specified harmful uses.[132] However, as discussed above, it is worth noting that licenses cannot be solely relied upon as a safeguard against data misuse. In case of sensitive or high-stakes datasets where potential misuse carries a range of harms, data contributors should use data use agreements to provide controlled access (with strict conditions defined on a case-by-case basis).

Public commitment: Data contributors should issue public statements that clearly articulate the intended uses and explicit prohibitions on misuse of their datasets. Such commitments, when published alongside the dataset and metadata, increase the reputational and legal costs of violating the terms, even where formal enforcement is limited. By combining licensing restrictions with transparent public communication, contributors strengthen community norms, deter bad actors, and signal accountability to downstream users and stakeholders.

Model Developers

AI Bill of Materials (AIBOM):
Facilitate transparency and reproducibility by disclosing the decision-making processes that shaped development throughout the lifecycle of the AI model with existing disclosure methods such as AIBOM. An AIBOM should ideally include the following[133]:

Compute Transparency: Document the type of compute used (GPU/TPU specifications, cloud/on-premise), the amount of compute time (e.g., GPU hours), and associated energy/carbon footprint where possible.

→ **Transparency on compute makes replication feasible, clarifies accessibility barriers, and highlights inequalities in who can reproduce training runs.**

132 'Responsible AI Licenses (RAIL)', Responsible AI Licenses (RAIL), accessed 29 September 2025, https://www.licenses.ai.
133 Interview with a Working Group member.

Data Provenance & Lineage[134]

→ Document dataset sources, licenses, and collection protocols.
→ Log filtering, cleaning, or translation decisions, along with known biases and limitations.

Training Process Metadata

→ Specify training duration, number of runs, and experiments conducted (including ablation studies and ensemble methods).
→ Record and share training parameters: learning rates, optimisers, batch sizes, early stopping conditions, etc.
→ Indicate whether code for training pipelines is openly available, and provide links or repositories.

Model Artefacts and Accessibility

→ Disclose whether weights are released, and under what license.
→ Where weights cannot be released (e.g., legal or safety reasons), provide detailed cards documenting architecture, size, and reproducibility limits.

Evaluation Transparency

→ Clearly specify evaluation datasets (with links or metadata), tasks used, and benchmarks.
→ Document language, dialect, and cultural coverage, particularly critical in multilingual contexts such as India.

Experimentation and Negative Results

→ Encourage disclosure of failed runs, discarded experiments, and negative results where relevant. These are crucial for understanding model stability and preventing wasteful duplication of effort.

End-to-End Documentation

→ Adopt structured formats such as Model Cards, but extend them to include process metadata (compute, training runs, ablations, evaluation coverage).

134 Michelle Knight, 'What Is Data Lineage?', DATAVERSITY, 21 April 2025, https://www.dataversity.net/what-is-data-lineage/.

Addressing misuse through developer-side measures, such as:

Continuous monitoring: Establish feedback loops that monitor new adversarial attacks, vulnerabilities, or incidents. This may help developers keep pace with emerging threats, while civil society watchdogs and policymakers can rely on shared reporting hubs to prioritise interventions. Databases such as the ATLAS matrix[135] and AI Incident Database[136] may be utilised to monitor threats and incidents.

Incorporating Checks against Data Poisoning and Adversarial Attacks: To address data poisoning attacks, developers can incorporate data validation and sanitisation mechanisms,[137] anomaly detection,[138] data provenance,[139] and adversarial training for models.[140] For label flipping poisoning attacks, there are label sanitisation processes like the k-Nearest Neighbours (k-NN), which can mitigate such attacks on machine learning classifiers.[141]

Licenses: Model developers should adopt licensing frameworks that explicitly prohibit harmful or high-risk applications of open source AI models. RAIL provides a practical template by allowing free access for research and innovation while embedding enforceable restrictions on misuse (e.g., disinformation, surveillance, biometric profiling, or discriminatory systems).[142] By releasing models under RAIL or equivalent licenses, developers can balance openness with safeguards, reducing the likelihood of downstream harms without closing off legitimate uses.

Auditing: Developers releasing open source AI models or datasets should plan for independent audits to validate both technical robustness and responsible AI claims prior to public release. Given the likelihood

135 Mitre Atlas, 'Atlas Matrix', Mitre Atlas, accessed 12 September 2025, https://atlas.mitre.org/matrices/ATLAS.; ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems) is a living knowledge base of adversary tactics and techniques against Al-enabled systems based on real-world attack observations and realistic demonstrations from Al red teams and security groups.

136 'Artificial Intelligence Incident Database', Artificial Intelligence Incident Database, accessed 12 September 2025, https://incidentdatabase.ai/.

137 This process includes thorough verification and validation of training data before training, to ensure the model is trained on sanitised data. This process should be regularly conducted even after training stages.

138 Real time detection of differences and anomalies in data patterns can indicate if data poisoning attacks have been conducted.

139 Accurately identifying the origin and transformation of datasets used in model training can ensure that data poisoning can be identified; Nathalie Baracaldo et al., 'Mitigating Poisoning Attacks on Machine Learning Models: A Data Provenance Based Approach', Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security (New York, NY, USA), AISec '17, Association for Computing Machinery, 3 November 2017, 103–10, https://doi.org/10.1145/3128572.3140450.

140 Training models on adversarial examples can allow it to identify patterns and become resistant to data poisoning attacks; '(PDF) Data Poisoning -What Is It and How It Is Being Addressed by the Leading Gen AI Providers?', ResearchGate, https://doi.org/10.5281/zenodo.13318796.

141 Andrea Paudice et al., 'Label Sanitization Against Label Flipping Poisoning Attacks', in ECML PKDD 2018 Workshops, ed. Carlos Alzate et al. (Springer International Publishing, 2019), https://doi.org/10.1007/978-3-030-13453-2_1.

142 Responsible AI Licenses (RAIL), 'Responsible AI Licenses (RAIL)'.

of downstream reuse and adaptation, external audits can provide an impartial evaluation of data provenance, model performance across subgroups, and compliance with ethical or legal standards. This helps establish baseline accountability, builds trust among potential adopters, and offers credibility beyond internal review processes. This is particularly important when formal regulatory oversight is limited.[143]

Public communication: Model developers should make explicit public statements that define acceptable and prohibited uses of their models. Publishing these commitments alongside model weights, code repositories, and documentation increases the reputational and legal costs for actors who attempt to misuse the models. By combining transparent communication with licensing measures (such as RAIL), developers can signal accountability, strengthen community norms, and deter harmful downstream applications.

Tools to Empower Downstream Developers to Build Responsibly: To promote responsible AI development and deployment, it is recommended that, where possible, open source AI model developers provide integrated, open source safety tools such as content moderation filters, usage guidelines, alongside their foundational models. These resources should be designed to empower downstream developers to identify and mitigate potential risks, ensure compliance with ethical standards, and foster trust in AI systems across diverse applications. By embedding such tools and best practices into the open source ecosystem, the broader AI community can be better equipped to build and deploy AI technologies in a safe, transparent, and accountable manner.

Hosting Platforms

Facilitate accessibility and discoverability by encouraging adherence to data reporting/disclosure standards:

Contributor Standards for Metadata and Reporting: Require contributors to adopt baseline documentation standards, such as data cards covering provenance, purpose, scope, collection methodology, annotation protocols, and limitations.[144]

143 Jakob Mökander, 'Auditing of AI: Legal, Ethical and Technical Approaches', Digital Society 2, no. 3 (2023): 49, https://doi.org/10.1007/s44206-023-00074-y.

144 Timnit Gebru et al., 'Datasheets for Datasets', Commun. ACM 64, no. 12 (2021): 86–92, https://doi.org/10.1145/3458723.

→ **For text data, encourage contributors to follow domain-specific standards (e.g., consistent tokenisation, encoding formats) to ensure downstream NLP reproducibility.**[145]
→ **Provide schema templates and machine-readable formats (e.g., JSON-LD, DCAT) to enforce consistency.**[146]

AI-Supported Standardisation and Discovery

→ **Hosting platforms should leverage AI systems (including LLMs) to assist contributors in completing metadata requirements by pre-filling fields, flagging gaps, and aligning with recognised standards.**[147]
→ **AI can also enhance dataset discovery, allowing natural language search and query resolution across hosted datasets.**[148]

Transparency on Gaps and Limitations: Hosters should provide structured fields for contributors to disclose limitations (coverage gaps, demographic exclusions, biases introduced by refusal rates, etc.), ensuring that downstream users can critically interpret representativeness.

Multiple options for licensing and accessibility: Data hosting platforms should support multiple licensing models and access mechanisms to lower barriers for dataset sharing. By accommodating a range of licenses, including permissive, restrictive, and responsible AI licenses (such as RAIL), and offering flexible access options such as pay-per-download or tiered access, platforms incentivise organisations and contributors to make datasets available publicly. Platforms such as Mozilla Data Collective exemplify this approach, enabling broader participation while maintaining governance, discoverability, and compliance standards.[149]

145 Kaiser Sun et al., 'Tokenization Consistency Matters for Generative Models on Extractive NLP Tasks', in Findings of the Association for Computational Linguistics: EMNLP 2023, ed. Houda Bouamor et al. (Association for Computational Linguistics, 2023), https://doi.org/10.18653/v1/2023.findings-emnlp.887.
146 Mark A. Musen et al., 'Modeling Community Standards for Metadata as Templates Makes Data FAIR', arXiv:2208.02836, preprint, arXiv, 14 October 2022, https://doi.org/10.48550/arXiv.2208.02836.
147 Sowmya S. Sundaram et al., 'Use of a Structured Knowledge Base Enhances Metadata Curation by Large Language Models', arXiv:2404.05893, preprint, arXiv, 20 February 2025, https://doi.org/10.48550/arXiv.2404.05893.
148 Aivin Solatorio and Olivier Dupriez, 'Efficient Metadata Enhancement with AI for Better Data Discoverability', World Bank Blogs, 13 December 2024, https://blogs.worldbank.org/en/opendata/efficient-metadata-enhancement-with-ai-for-better-data-discovera.
149 'Mozilla Data Collective', Mozilla Data Collective, accessed 24 September 2025, https://datacollective.mozillafoundation.org/.

Open source consultancy to data providers: Data hosting platforms should provide guidance and support to contributors on legal, regulatory, and technical considerations related to dataset hosting. This includes advising on compliance with local data protection laws, addressing concerns about data localisation, and providing formal attestations or certifications when required.

Develop incentive-aligned marketplace models to sustain open data and model ecosystems: Platforms should explore and design marketplace mechanisms that enable fair value exchange within the open ecosystem while maintaining openness and accessibility. Such models can support the long-term sustainability of open source AI artefacts, encourage active data and model contribution, and reward high-quality, responsibly developed datasets and models. Platforms can consider approaches that allow monetary or non-monetary incentives (such as recognition credits) while ensuring that the core resources remain freely available.

Design for interoperability: Hosting providers and platform operators should implement open protocols and interface specifications that enable models, datasets, and tools to be easily transferred, reused, or integrated across different platforms and environments. Standardising formats, metadata schemas, and access mechanisms reduces dependence on proprietary systems, ensures greater technical flexibility, and supports the long-term sustainability and usability of open source AI components.

Establish joint data lifecycle management practices to prevent the accumulation of "data junk": Hosting platforms and data contributors should work together to ensure that open datasets remain accurate, relevant, and high-quality throughout their lifecycle. This requires implementing data hygiene practices, such as periodic reviews, versioning, the archival of outdated or redundant datasets, and maintaining clear metadata that records provenance, usage status, and update history. Hosting platforms should support these practices by providing tools and standards for dataset maintenance, while data contributors ensure compliance through transparent documentation and regular updates. By collaboratively managing dataset lifecycles, the open source AI ecosystem can avoid the buildup of obsolete or low-quality "data junk", improving trust, efficiency, and sustainability.

Addressing misuse through developer-side measures, such as:

Platform responsibility: Model hosting platforms (e.g., Hugging Face, GitHub, OpenAI) should handle safeguards like content filters, leakage prevention, and compliance checks. This avoids pushing the impossible task of legal compliance onto every developer. Platforms act as "safety intermediaries," a role regulators can formalise.

Downstream Application Developers

Acknowledging the Work of Others: Downstream developers should explicitly recognise and credit the contributions of other open source creators, maintainers, and communities whose work underpins their projects. This practice not only upholds ethical standards but also strengthens trust and encourages continued innovation within the open source AI ecosystem. By documenting dependencies, attributing datasets, and citing foundational models, developers reinforce a culture of transparency and mutual respect. Acknowledgement as a standard practice ensures that the collective effort behind open source AI is sustained.

Lineage Verification and Responsible Integration: Downstream developers should make use of lineage verification tools, checking that models and datasets are authentic, licensed, and compliant before integration. This reduces the risk of embedding untrustworthy components into applications, while also giving civil society and policymakers a practical mechanism to counter disinformation and copyright misuse.

Differential Watermarking: When synthetic content such as voices, images, or text circulates online, watermarking can support traceability and help reduce the risks of misinformation and misuse. Downstream application developers can embed watermarking or content provenance signals at the point of content generation or deployment whereas upstream model developers can support this by providing watermarking tools or default capabilities that downstream developers can adopt.

Watermarking is not a fail-safe mechanism. Generated content can be modified, transformed, or re-encoded in ways that weaken or remove embedded signals.[150] Watermarking should therefore be understood as a complementary safeguard, alongside transparency, documentation, and licensing choices, rather than as a standalone control.

Embed "Shift-left" Responsible AI and Compliance Mechanisms Early in the Development Process: Developers should adopt a "shift-left" approach by integrating safety and compliance checks at the earliest stages of development, rather than as post-release evaluations. This involves embedding automated guardrails, data provenance validation, bias detection, and licensing verification directly into development environments, software development kits, and model training workflows. By moving accountability and quality assurance "left" in the lifecycle, developers can detect and mitigate risks proactively, streamline governance, and ensure that open source AI systems are responsibly designed from inception.

150 Partnership on AI, Risk Mitigation Strategies for the Open Foundation Model Value Chain, July 2024, https://partnershiponai.org/wp-content/uploads/dlm_uploads/2024/07/open-foundation-model-risk-mitigation_rev3-1.pdf.

Responsible AI Assessments: Downstream application developers should integrate structured assessments that evaluate ethical, social, and technical risks. These assessments should go beyond accuracy benchmarks to include fairness, explainability, safety, and societal impact, using recognised frameworks or external "stress-test" tools such as the Responsible AI Assessment Framework developed by the initiative of German Development Cooperation FAIR Forward - AI for All in collaboration with a community of inclusive AI experts.[151] Further, developers should embed responsible AI assessment frameworks that are tailored to the specific risks, standards, and regulatory contexts of individual sectors such as healthcare, finance, education, and agriculture. These assessments should examine domain-relevant risks, for instance, bias in medical diagnosis datasets, transparency in credit scoring models, or safety in agricultural automation tools. Incorporating sector-specific evaluation criteria will strengthen accountability, improve alignment with existing legal and professional standards, and support safer, context-aware deployment of open source AI systems.

→ **For developers seeking further guidance on how to operationalise responsible AI practices, several resources in India offer detailed and practical direction. These include NASSCOM's Developer's Playbook for Responsible AI in India, Digital Futures Lab's Practice Playbook on Responsible AI, and sector-specific frameworks such as the Indian Council of Medical Research's Ethical Guidelines for AI in Healthcare and Biomedical Research. The recommendations in this brief are indicative rather than exhaustive, and should be read in conjunction with such guidance to support a more comprehensive and context-sensitive approach to responsible AI development and deployment.**

Responsible AI Assessments: Downstream developers should document and disclose the fine-tuning or adaptation they perform on open source models, especially when targeting sensitive sectors like healthcare, finance, or agriculture. This ensures traceability of design choices and helps policymakers evaluate sector-specific risks.

151 Digital Global, 'Ethical Crash Test for AI? How to Navigate the Road to Responsible Innovation', BMZ Digital.Global, 14 August 2024, https://www.bmz-digital.global/en/news/ethical-crash-test-for-ai-how-to-navigate-the-road-to-responsible-innovation/.

Appendix I: **Scope & Methodology**

Objectives

The primary objective of this policy brief is threefold:

First, it aims to **demystify the concept of open source AI**[152] for both policymakers and AI practitioners in India, and **unpack its value proposition** for the Indian AI ecosystem. In this regard, it looks to answer the following questions:

→ What opportunities can be unlocked through open source AI in India?
→ What do these opportunities look like for different stakeholders - the government, the AI development community, and the Indian population at large?
→ To what extent, and at which layers, must an AI system be open for its advantages to materialise meaningfully in the Indian context?

Second, it seeks to explain the **challenges and tensions** involved in the adoption of an open source approach for AI development and use. To this end, it will cover the following questions:

For AI practitioners as well as policymakers:
→ What are the challenges or trade-offs involved in developing and releasing various components of an AI system 'openly'?

For policymakers, in particular:
→ What are the key challenges that will be involved in the governance of open source AI systems?

Lastly, it seeks to identify concrete steps that can be taken by policymakers and AI practitioners alike to effectively leverage the opportunities of open source AI in a safe and responsible manner. To achieve this, the brief will provide practical recommendations, structuring its guidance along the following set of inquiries:

→ What are the potential policy instruments that can be used by government bodies in India to both govern and steer open source AI initiatives?

152 For the purposes of this brief, the term "artificial intelligence" refers to predictive and generative systems. We do not address agentic or autonomous AI, which represent a distinct and emerging class of technologies. Within this scope, the analysis places greater emphasis on generative and general-purpose AI models, where questions of openness and governance are currently most salient, while predictive AI is discussed where relevant, including in the component–outcome matrix in Table 1.

→ What kind of policy interventions are critical to addressing the challenges faced by AI practitioners in developing, hosting, or using AI components under an open source regime?

→ What are some of the key lessons that can be drawn from international best practices surrounding open source AI and/or open source software?

→ What steps can be taken by AI practitioners (model developers, data collectors, or downstream application developers) to adopt an open source AI approach and navigate its tensions?

Methodology

We adopt a multi-pronged approach as part of our research and analysis, combining key informant interviews, stakeholder dialogues, and a rigorous literature review to develop grounded and contextually relevant policy recommendations.

Key elements of our approach are as follows:

Multistakeholder Dialogues: Formation of a 16-member working group comprising representatives from key stakeholder groups, including government bodies, multinational technology corporations, Indian start-ups, and civic tech organisations. The working group also consisted of key experts from academia, legal firms, and technology policy organisations.

→ One-on-one interviews were conducted with each working group member to gain an in-depth understanding of their perspectives on open source AI.

→ One in-person and two virtual workshops were conducted with all the working group members to present our emerging findings, as well as to enable consensus-building across stakeholders.

Secondary Research: A comprehensive review of existing scholarship and policy discourse on open source AI, with a focus on definitional debates, emerging opportunities, risks, challenges, and potential mitigation strategies.

Advisory Board: Establishment of a multidisciplinary advisory board to provide guidance and validate research directions and findings.

Appendix II:
List of Stakeholder Interviews

1 **Mr Avik Sarkar**
Indian School of Business, Mohali [Academia] *on March 11, 2025*

2 **Mr Prasanta Ghosh**
Indian Institute of Science (IISc), Bangalore [Civic-Tech] *on March 12, 2025*

3 **Ms Rama Devi Lanka**
Government of Telangana [Government] *on March 13, 2025*

4 **Dr Sivaramakrishnan**
Aaquarians.ai [Start-ups] *on March 17, 2025*

5 **Mr Venkatesh Hariharan**
[Civic-Tech] *on March 18, 2025 & June 9, 2025*

6 **Mr Vibhav Mithal**
Anand & Anand [Legal Expert] *on March 19, 2025*

7 **Ms Swetha Kolluri**
Gates Foundation [Civic-Tech] *on March 20, 2025*

8 **Mr Jigar Doshi**
ARTPARK [Civic-Tech] *on March 25, 2025*

9 **Mr Amritendu Mukherjee**
Neuropixel.ai [Start-ups] *on March 25, 2025*

10 **Ms Meghna Bal**
Esya Centre [Legal Expert] *on March 26, 2025*

11 **The Tattle team**
[Civic-Tech] *on March 27, 2025*

12 **Mr Gaurav Godhwani**
Civic Data Lab [Civic-Tech] *on March 28, 2025*

13 **Mr Aman Taneja**
Ikigai Law [Legal Expert] *on April 1, 2025*

14 **Ms Shweta Gupta**
Microsoft [Global Technology Companies] *on April 3, 2025*

15 **Ms Amrita Sengupta**
[Civil Society] *on April 7, 2025*

16 **Mr Atul Gandre**
Technology Head, Tata Consultancy Services, [Tech Expert] *on May 15, 2025*

17 **Mr Abhishek Singh**
Additional Secretary, Ministry of Electronics and Information Technology (MeitY) & Director General, National Informatics Centre India [Government] *on July 22, 2025*

Appendix III: **Existing Frameworks for Defining Open Source AI**

We use this section to provide a succinct overview of the various definitions of open source AI. We group them into two broad categories, binary and gradient-based approaches, a typology inspired by Basdevant et al.[153]

Type I - Binary approach:

In such frameworks, openness is viewed as an absolute: a system is either open or it is not. A few prominent examples include:

Open Source Initiative's (OSI) The Open Source AI Definition – 1.0[154]
Under this definition, an open source AI system is one which is made available under terms and in a way that grants the freedoms[155] to:

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

Further, they specify that to enable modification, certain elements such as data information, source code, and parameters must be made accessible. The definition of open source AI by OSI includes a description of data information.[156] Data information should include "(1) the complete description of all data used for training, including (if used) of unshareable data, disclosing the provenance of the data, its scope and characteristics, how the data was obtained and selected, the labeling procedures, and data processing and filtering methodologies; (2) a listing of all publicly available training data and where to obtain it; and (3) a listing of all training data obtainable from third parties and where to obtain it, including for fee."[157]

153 Basdevant et al., 'Towards a Framework for Openness in Foundation Models'.
154 Open Source Initiative, 'The Open Source AI Definition – 1.0', Open Source Initiative, 2024, https://opensource.org/ai/open-source-ai-definition.
155 These freedoms are derived from the Free Software definition. See Free Software Foundation, 'What Is Free Software?', GNU Project, n.d., accessed 30 April 2025, https://www.gnu.org/philosophy/free-sw.en.html.
156 'The Open Source AI Definition – 1.0', Open Source Initiative, n.d., accessed 16 June 2025, https://opensource.org/ai/open-source-ai-definition.
157 'The Open Source AI Definition – 1.0'.

However, this definition does not mandate the release of underlying training data and instead only requires data information to be made available. This has attracted criticism, with opponents arguing that the definition "fails to require reproducibility by the public of the scientific process of building these systems".[158] However, OSI has justified this exclusion by stating that they wanted to ensure the definition would also apply to sectors or fields where data cannot be legally shared, like sensitive personal data, like health data, and mandating such disclosures would severely restrict the scope of the definition.

Type II - Gradient-based Approaches

These approaches treat openness as a continuum, recognising varying levels of transparency and access across system components. A few key frameworks that exemplify this approach are:

Irene Solaiman's Gradient of Generative AI Release: Methods and Considerations[159]

Solaiman proposes a gradient framework for assessing levels of access to generative AI systems. The framework proposes that an AI system can be broken into three key parts when it comes to viewing openness: (i) the model itself, (ii) components that enable further risk analysis, and (iii) components that enable model replication. Solaiman proposes a framework to assess six levels of access to generative AI systems:

→ **Fully closed:** When all aspects and components of a system are inaccessible outside the developer organisation, or even closed outside a specific subsection of an organisation, the system is fully closed. Examples include Google's Imagen and DeepMind's Gopher.

→ **Gradual or staged access:** This method refers to releasing a system in stages or gradually over a predetermined amount of time. Examples include OpenAI's GPT-2.

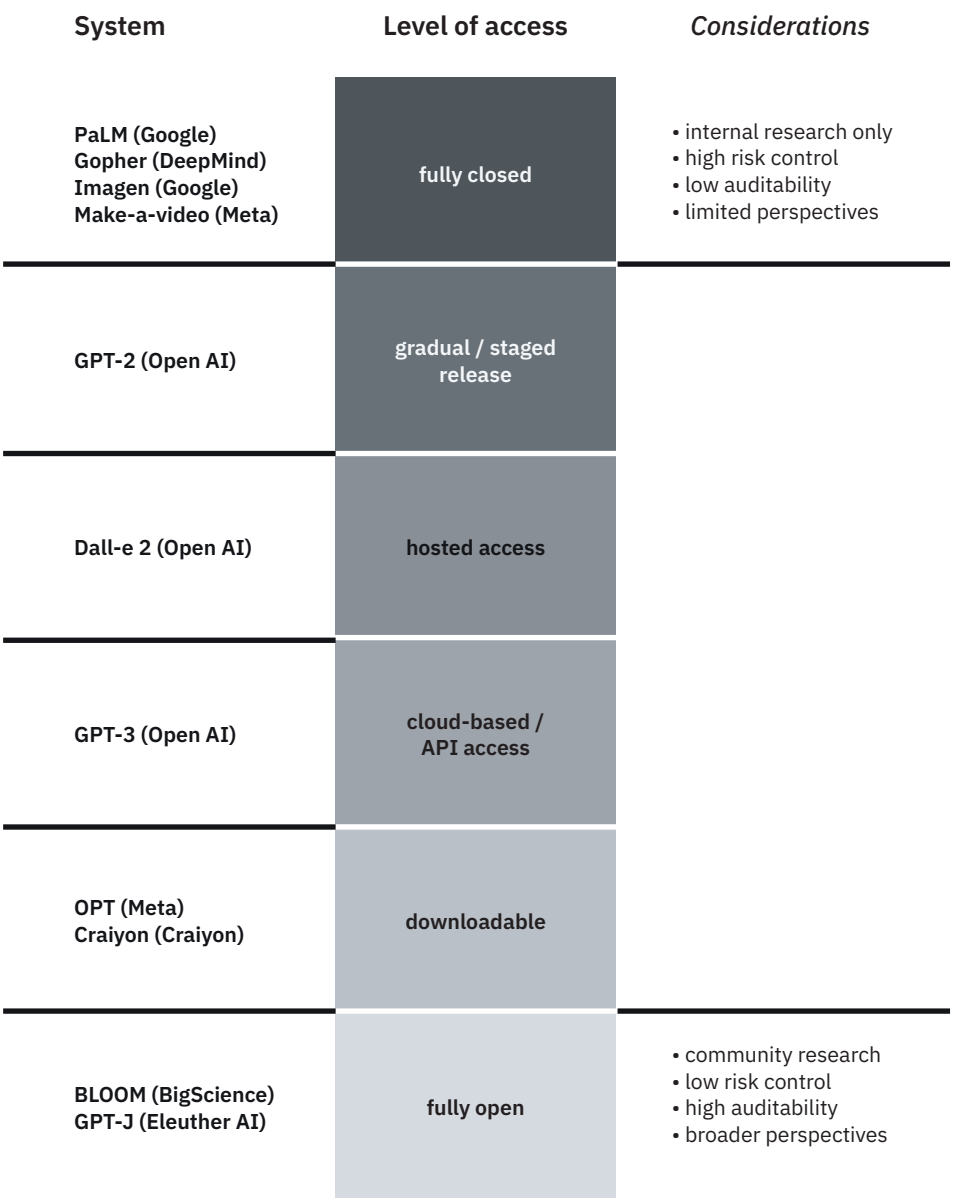
→ **Hosted access:** System deployers may provide access to the model itself by hosting the model on their servers and allowing surface-level interfacing. This method is specific only to model access, not access to other system components. Examples include Midjourney.

158 Bradley M. Kuhn, 'Open Source AI Definition Erodes the Meaning of "Open Source"', Software Freedom Conservancy, 31 October 2024, https://sfconservancy.org/blog/2024/oct/31/open-source-ai-definition-osaid-erodes-foss/.
159 Solaiman, 'The Gradient of Generative AI Release'.

→ **Cloud-based or API access:** Cloud-based access or access provided via application programming interface (API) provides more insight and researchability into a model than Hosting, but still allows for restrictive functionality. Examples include OpenAI's GPT-3 release via API.

→ **Downloadable access:** The system may be made available to users to download. However, its main distinction from a fully open system is that it may still withhold certain system components, such as training datasets.

→ **Fully open:** When all aspects of the system are accessible and downloadable, including all components, the system is fully open.



↑ **Figure 2**
Irene Solaiman's Gradient Approach to Open Source AI

Linux Foundation's Model Openness Framework (MOF)
The framework proposes a three-tier classification system to classify the degree of completeness and openness of AI models across all aspects of a model's development lifecycle. It uses the term "completeness" to measure the availability of components that are released with models (with the goal of full completeness) and the term "openness" to describe the usage of permissive licenses for components.[160] The framework identifies 17 components to fulfil completeness of model artefacts, which are categorised across three classes, with Class III being the least complete and Class I being the most complete.

↓ **Table 4**
Linux Foundation's Model Openness Framework

MoF Class	Components Included
Class III **Open Model**	**1. Model Architecture** **2. Model Parameters (Final Checkpoints)** **3. Technical Report or Research Paper** **4. Evaluation Results** **5. Model Card** **6. Data Card** **7. Sample Model Outputs (Optional)**
Class II **Open Tooling**	**1. All Class III Components** **2. Training, Validation and Testing Code** **3. Inference Code** **4. Evaluation Code** **5. Evaluation Data** **6. Supporting Libraries & Tools**
Class I **Open Science**	**1. All Class II Components** **2.Research Paper** **3. Datasets** **4. Data Preprocessing Code** **5. Model Parameters (Intermediate Checkpoints)** **6. Model Metadata (Optional)**

160 White et al., 'The Model Openness Framework'.

Mozilla Foundation's Framework for Openness in Foundation Models[161]

Mozilla's framework takes a layered and modular view of openness, grounded in four key principles:

- → **Openness should be assessed at both the model stack (e.g., weights, code) and system stack (e.g., APIs, deployment environments) levels.**
- → **Openness exists in multiple forms, and frameworks should allow for granular, flexible terminology.**
- → **Greater alignment is needed on the benefits, risks, and modalities of opening various AI components.**
- → **Safety considerations must be embedded into any effort to open AI systems.**

To operationalise these principles, Mozilla defines different "dimensions of openness" or specific components (such as datasets, model weights, and training code) and associated attributes (e.g., licensing, documentation quality, and accessibility). Each component can be made more or less open, depending on the chosen level of granularity. The framework encourages developers, regulators, and civil society to think beyond whether something is "open" or "closed," and instead ask: what is open, to whom, and for what purpose?

All the approaches presented above are a testament to the multidimensional nature of openness in AI — there is no one-size-fits-all definition. Each underscores different facets of openness: from access to source code and model weights to the availability of risk evaluation artefacts and replication documentation. Additionally, each of these definitions offers a unique perspective on openness. These frameworks are not competing definitions but complementary lenses that help exemplify the range of what "open source AI" can entail.

161 Basdevant et al., 'Towards a Framework for Openness in Foundation Models'.