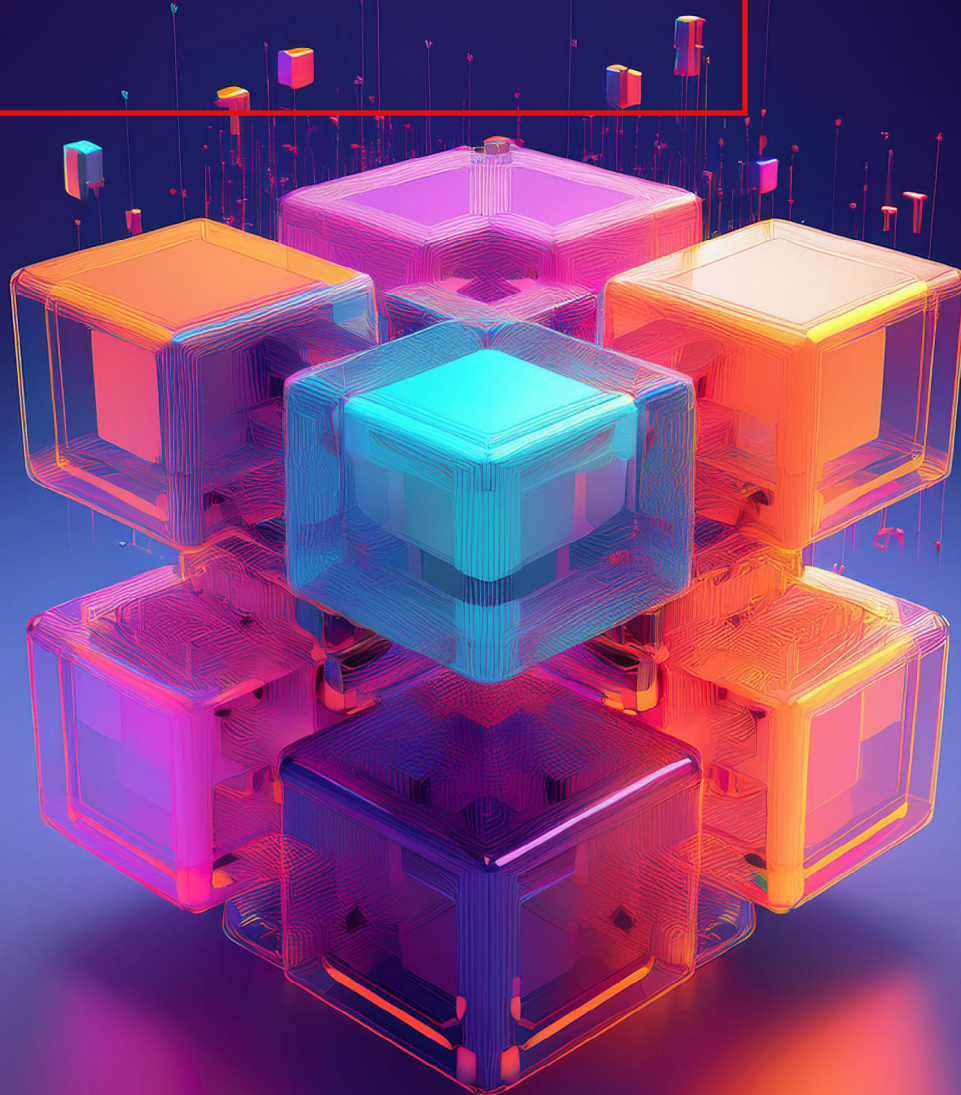


ECONOMIST
IMPACT

Open sourcing the AI revolution



Sponsored by



About the paper

This white paper, written by Economist Impact and sponsored by Meta, outlines the value proposition of the open-source approach to artificial intelligence (AI) across three key areas, while highlighting potential limitations and barriers. Economist Impact conducted extensive desk research and an interview programme to develop the paper.

Economist Impact wishes to thank the following experts for sharing their time and expertise:

- **Rishi Bommasani**, society lead, Stanford Center for Research on Foundation Models
- **Ray Perrault**, co-chair, Stanford University AI Index Steering Committee
- **Anand Rao**, distinguished services professor of applied data science and AI, Heinz College of Information Systems and Public Policy at Carnegie Mellon University
- **Rahul Roy-Chowdhury**, CEO, Grammarly
- **Priyanka Sharma**, executive director, Cloud Native Computing Foundation

This report was produced by a team of in-house researchers, writers, editors and designers, including:

- **Vaibhav Sahgal**, project director
- **Edwin Saliba**, project manager
- **Adam Green**, writer
- **Amanda Simms**, copy editor

Contents

4	Executive summary
6	Chapter 1: Access
	1.1 Tackling bias, toxicity and harm
	1.2 Linguistic diversity
	1.3 Limitations and barriers
	Malicious use
	Resource constraints
11	Chapter 2: Innovation and economic growth
	2.1 Pushing the science frontier
	2.2 Productivity gains
	2.3 Going local
	2.4 Limitations and barriers
	Quantifying value
	Cost of failure
16	Chapter 3: Transparency and trust
	3.1 Many eyes
	3.2 Transparency in business reporting
	3.3 Limitations and barriers
	Data quality
	Clarity and consensus
19	Conclusion
20	Endnotes

Executive summary

The early origins of free open-source software date back to the 1960s, but it was not until the late 1990s that the term was coined.¹ Since then, open-source has moved from the margins to the mainstream, powering the digital infrastructures of our lives, from code to cloud.² The Linux operating system, released in 1991, today drives the world's top 500 supercomputers, 90% of cloud infrastructure and 85% of smartphones.³

The generative artificial intelligence (AI) revolution marks a new chapter in a storied history; two-thirds of the large language models (LLMs) released in 2023 were open-source, and tools are emerging to help developers design and build an application ecosystem atop open-source foundation models.^{4,5} Governments have already experimented with open-source AI in everything from Covid-trackers and voting systems to chatbots for completing tax returns.⁶

"Ever since ChatGPT came out and gave AI such a broad appeal, the importance of having open-source underpinnings became part of the parlance faster than ever before," says Priyanka Sharma, executive director of the Cloud Native Computing Foundation. "The most exciting thing is the level of consciousness around open-source and how critical it is in the AI movement."

Advocates argue that open-source promotes access to AI, supports research, and boosts

innovation and economic competitiveness, as well as strengthening safety, transparency, privacy and trust. A more democratised ecosystem promotes diversity and inclusivity by broadening the distribution of a powerful technology. However, there are limitations, barriers and potential risks to contend with too. Open-source models still require time, expertise and resources to fine-tune and adapt. Users need to source high-quality data to ensure optimal performance, which could be a challenge for smaller entities as compared with more established closed source developers. Open-source tools can be used and modified by nefarious actors in ways that may be harder to govern and track than closed source. This Economist Impact report, sponsored by Meta, assesses the strengths and shortcomings of open-source AI and looks to its likely role in the future of the AI ecosystem.

Key findings

Open-source can promote the democratisation of AI. More open models allow users to interrogate data, weights and design to spot biases and adapt models to their needs, including cultural and linguistic suitability. Open-source developers are also creating and sharing tools to support those in the ecosystem to build responsible AI. However, open-source AI models could also be used by nefarious actors, such as producing toxic or harmful



content and pursuing cybercrime. Moreover, while barriers to entry are falling thanks to the emergence of innovations like small language models, there are financial, technical and resource constraints that may inhibit smaller companies and those in academia or the public sector from fully leveraging open-source models.

From international scientific collaboration to business productivity, organisations are reaping economic benefits from open-source AI. By fostering collaboration, open-access models are boosting scientific research, especially in medicine and biology. At the company level, firms are quickly integrating AI tools tailored to their needs, helped along by common frameworks and a pool of vendors offering integration support. At the national level, productivity boosts could help countries find new sources of economic competitiveness and growth, and support the development of innovation ecosystems to solve locally relevant problems, especially in developing countries. However, few estimates chart open-source AI's economic impact. Companies often use multiple models in any given project, and open-source is, by definition, hard to track in terms of usage. There could be economic downsides too. For instance, companies face operational risks such as choosing a sub-optimal open-source model or fine-tuning technique, which could damage their performance or reputation. There is also the issue of countervailing trends, like the rising cost of data acquisition, and potential fines and damage

caused by the nefarious use of AI or accidental compliance breaches.

Open-source models are subject to continuous peer review, improving reliability and security, and can support transparency in the data economy as a whole. Open-source can improve trust in AI by allowing users to interrogate and test systems and by enabling more secure data handling through localisation and on-premise utilisation. To deliver on its promise of more trustworthy and transparent models, however, all participants in the open-source ecosystem need to adopt best practices in data sourcing and management. Data governance challenges are not unique to open-source, but by increasing the number of entities able to develop AI systems, including smaller organisations, there may be a greater risk of flawed practices and errors.

Delivering open-source AI's full potential requires collaboration between industry and government; proactive support to ensure all users can fully benefit; and a lead institution to set standards, definitions and governance principles. Some actors may lack the resources and tools to fully exploit open-source AI. There could be a role for industry and government to support actors such as universities or smaller companies in areas like compute and appropriate tooling to move from experiment to scale. Added to this, a global entity could help forge common standards and governance, drawing inspiration from organisations such as the World Wide Web Consortium.

Chapter 1: Access

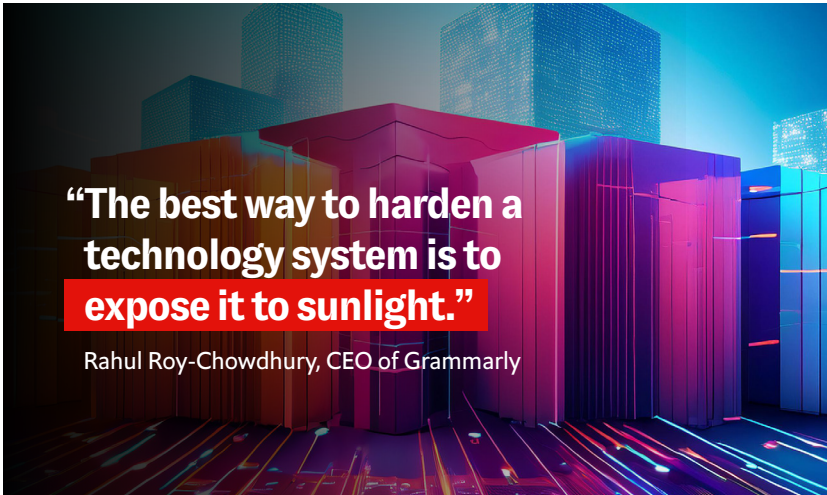
According to the Open Source Initiative (OSI), open-source AI offers four ‘freedoms’: to study, modify, use and share. It allows any user the freedom to use the AI system without needing permission, study its inner composition, modify the system and share it with others to suit their specific needs.⁷ This allows developers and users of all stripes to refine and iterate models and algorithms and apply them to diverse commercial, personal and social goals. But access is a double-edged sword, since the same tools can empower malicious actors.

Chapter one examines how these factors play out in specific domains, along with their limitations and barriers.

1.1 Tackling bias, toxicity and harm

Bias in AI training data, weights (the parameters that determine a model’s outputs) or model design can lead to unfair or exclusionary outcomes. This is a key concern, as model outputs can help inform decisions in sensitive areas like criminal justice and financial credit. Open-source models can be more robust to bias because more users, with more critical perspectives, can spot risks and vulnerabilities. “The best way to harden a technology system is to expose it to sunlight,” argues Rahul Roy-Chowdhury, the CEO of Grammarly, an enterprise communications company.

Organisations are also building open-source tools to help others tackle AI bias. Grammarly grappled with the challenge of LLMs producing inappropriate content suggestions due to a lack of context. For example, positive language is often a wise choice for letter writing, but not in the context of a condolence letter. Grammarly built a language sensitivity tool called Seismograph to identify and classify text that requires careful handling due to its content; it detects sensitive situations and limits the production of potentially offensive content.⁸ The company, a proponent of open-source, released Seismograph for public use. “We want to be positive contributors to the ecosystem through projects like Seismograph. Because these are public, they can get better all the time,” says Mr Roy-Chowdhury.



“The best way to harden a technology system is to expose it to sunlight.”

Rahul Roy-Chowdhury, CEO of Grammarly

EleutherAI, a grassroots collaborative project, released Pythia, a publicly available suite of tools to inspect LLMs for bias.⁹ To ensure their work was reproducible, the team used freely and publicly available code bases—the open-source GPT-NeoX and DeepSpeed libraries—for training. The platform includes 16 publicly accessible LLMs trained on public data, each with 154 checkpoints to allow researchers to study the model at various stages of the training process. The ability to modify pre-training data improved model performance in the area of gender bias.¹⁰ Other open-source bias toolkits include IBM AI Fairness 360, which lets users examine, report, and mitigate discrimination and bias in machine learning (ML) models throughout the AI application lifecycle.¹¹

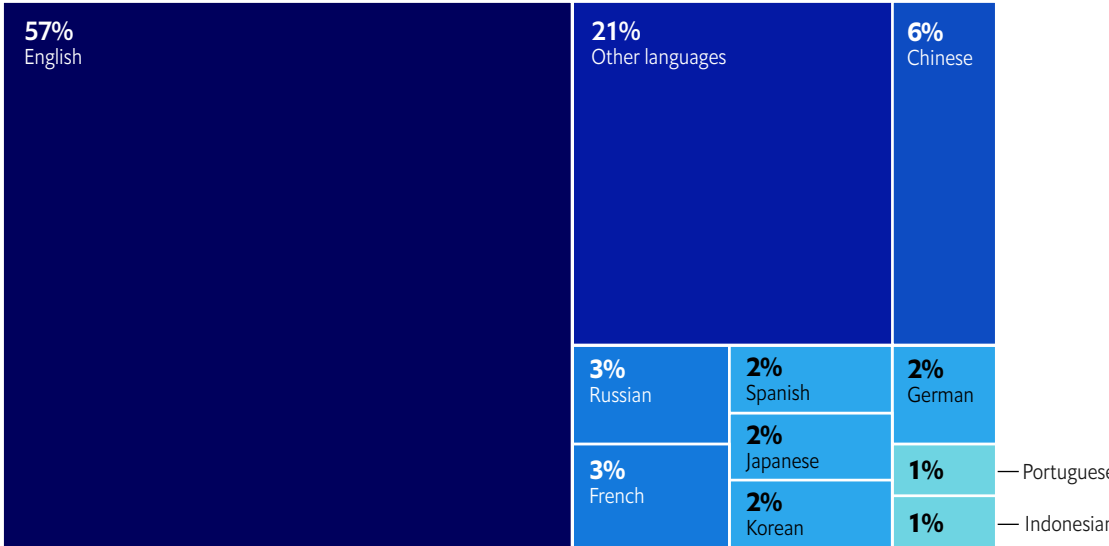
Open-source organisations can also tackle bias in the AI ecosystem by channelling their own resources to innovators, as evidenced by Mozilla. ai, an initiative from the organisation that created the Firefox browser.¹² The Mozilla Foundation launched Mozilla Ventures, which backs early stage

companies that align with the *Pledge for a Healthy Internet*, the Mozilla manifesto for an open and accessible internet.¹³ The portfolio includes Themis, which has developed software to help ML models recognise unreliable outputs, and Truepic, which is building tools to authenticate content and limit the spread of misinformation from AI-altered images.¹⁴

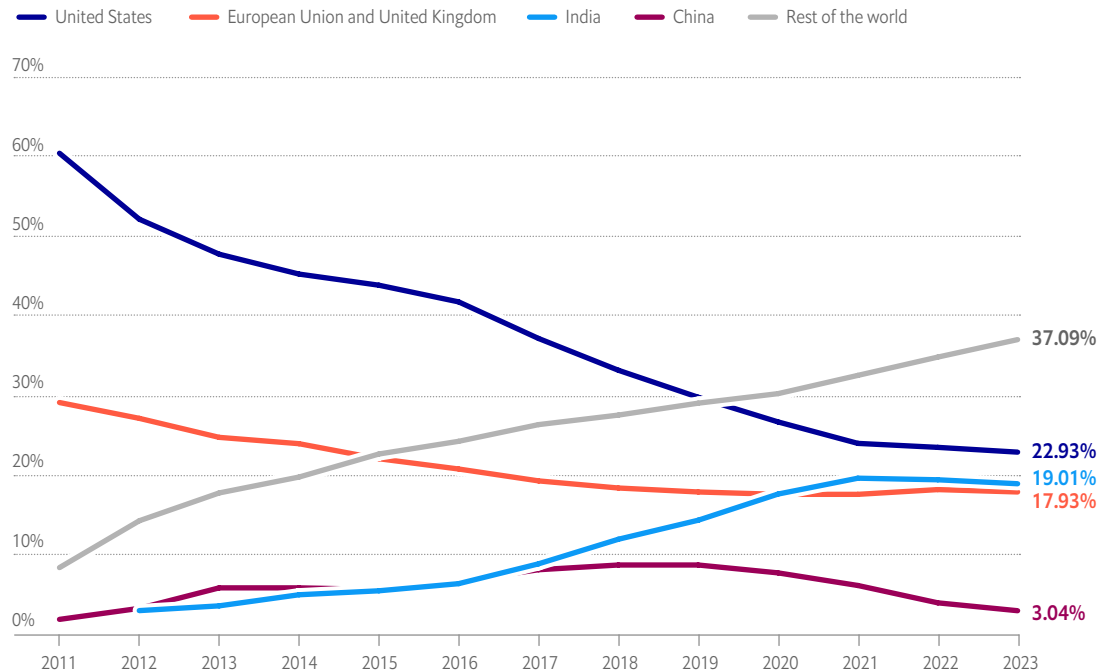
1.2 Linguistic diversity

The dominance of English language content online has led to a linguistic bias in AI models, which are often trained on, and therefore perform better in, English. LLMs perform better in “high-resource” languages due to a paucity of training data in other languages. ChatGPT-4 scored 85% in an English question-and-answer test compared with 62% in Telugu, an Indian dialect.¹⁵ While 57% of open-source AI training datasets are in English, there are initiatives afoot to improve linguistic representation (see Figure 1).¹⁶ A more accessible AI ecosystem could deepen support for greater linguistic representation in an Anglophone-dominated field.

Figure 1: Language divides in AI datasets
Breakdown of open-source AI training datasets on Hugging Face, by language, 2024



Source: OECD, p38¹⁷

Figure 2: GitHub AI projects (% of total) by geographic area, 2011-23Source: AI Index Report²³

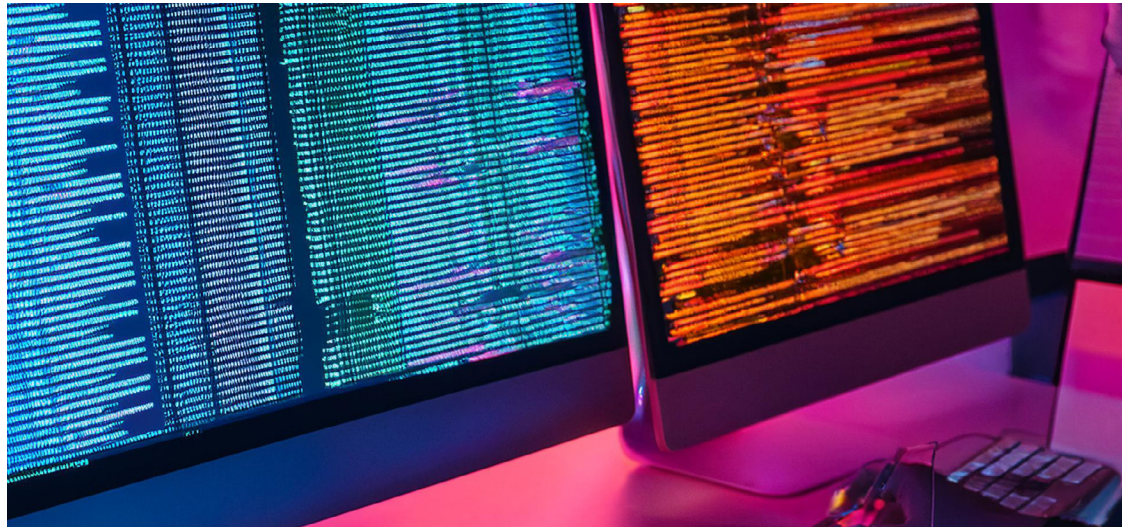
Various companies are already redressing the imbalance. Sarvam AI, an Indian start-up, modified the architecture of an open model from Hugging Face, a US software company, to train it in Hindi.¹⁸ Lelapa, a South African company, is seeking to improve access to AI services in South African languages with its Vulavula model.¹⁹ The Qatari Ministry of Communication and Information Technology is heading a multi-stakeholder collaboration to develop an Arabic LLM, called Fanar, based on open-source prototypes.²⁰ The model is trained on a combination of data from Islamic science, legal documents, and artistic, cultural, audiovisual and media content.²¹ Another influential project in the Middle East is JAIS 70B, an Arabic LLM to support heritage, culture and language and provide an AI foundation for Arabic-speaking nations.²² There are signs of greater global participation in open-source AI projects thanks to the availability of portals like GitHub, a developer platform (see Figure 2).

1.3 Limitations and barriers

Malicious use

Because open-source AI improves access for all, it can also be used by malicious actors seeking to perform criminal or harmful acts, from building bioweapons and launching sophisticated cyberattacks to producing misinformation or toxic content, such as non-consensual intimate imagery. Open-source models are, by definition, impossible to contain as they glide downstream from the guardrails their original developers put in place.²⁴ But the marginal risk posed by open-source AI depends on the use case.

The risk of open-source compared with closed depends on the threat. “We see clear evidence of an incremental risk of open models in the domains of child sexual exploitation and non-consensual intimate imagery,” notes Rishi Bommasani, society lead at the Stanford Center for Research



on Foundation Models. “Open weights, text and image models significantly increase the volume of this imagery on the internet.” But for bioweapons, there is no evidence yet. Moreover, Ray Perrault, co-chair of the Stanford University AI Index Steering Committee, says closed models can be used for nefarious purposes too, just to a lesser extent. “They are more restricted in what they allow to happen, but this is a problem that crosses the open-closed divide. Open-source models are easier to crack open, but it’s a matter of degree.”

“When the developer trains a foundation model, they implicitly or explicitly encode some model behaviour policy; what things the model can do, what types of requests might it refuse,” explains Mr Bommasani. “Because these weights are made available [in open-source models], this can be changed by third parties downstream, so there is more pluralism in determining what behaviours are acceptable.”

Open-source model developers could address malicious use through stronger conditions of use, but these take time to develop and lack legal enforcement. “It took a long time to develop standard licence agreements for software, which provided different sets of mutually acceptable use restrictions on the content,” argues Dr Perrault.

“We haven’t settled on a few dominant models for how AI can be licensed.”

He adds that while open-source models may include provisions on risks like divulging personally identifiable information and restricting applications like lethal weapons, they may be difficult to enforce. “Until there are more teeth in these agreements for what happens if conditions are violated, it will be hard to have any real control over [malicious use],” argues Dr Perrault. “There are always going to be bad actors that are not legal entities you can go after.”

Governments and regulators are developing legislation to punish the malicious use of open-source models but the key challenge is to craft rules that can realistically be enforced given the limits that open-source developers have on downstream use.

Creating a compliance framework that targets the individuals or organisations responsible is no easy feat. Once a model is shared, even if the model is benign, it can be altered and have its protocols removed by third parties. It is incumbent for all participants in the open-source ecosystem to act responsibly. They can be helped by the emergence of innovations that provide guardrails and checks

such as watermarking, stable signatures (which roots a watermark in an original AI image to prevent it from being removed), provenance tracking (tracing the history and quality of a dataset) and human authentication are being developed—some of which themselves are open-source initiatives.²⁵

Resource constraints

Open-source models are more financially accessible for actors with limited resources than the largest commercial enterprises, such as universities. “It has become so expensive to train a new model,” says Dr Perrault. “If you are an academic working on this, there is no way you can afford to train a model from scratch; this is beyond what most research budgets allow. If we are going to support research on this technology, [open] models are essential. If we forbade them, there would be a few companies that could afford them, and everything else would dry up, and that would be a disaster.”

However, while open-source AI models can broadly democratise access compared with a world of only closed-source models, the technical and resource requirements of adapting and fine-tuning AI models are not trivial. Fine-tuning refers to the process of customisation, to ensure policy compliance and data relevance for specific user needs.²⁶ This needs computation and resources to prepare and process relevant datasets, choose the right foundation model and fine-tuning technique, load and fine-tune the model, iterate and evaluate, and deploy.²⁷ This comes to a fraction of the cost of building a foundation model from scratch, but it is not necessarily a small cost, in terms of finance or human resources, especially for non-profit or public entities. Intensively performed fine-tuning activities could even result in higher AI energy consumption. However, more efficient fine-tuning techniques are being developed to remedy this issue.²⁸

“Even fine-tuning a reasonably small model can’t be done on a laptop and there is very little provisioning of real compute [in academia],” says Anand Rao, distinguished services professor of applied data science and AI at the Heinz College of Information Systems and Public Policy at Carnegie Mellon University. “Just having open-source models is insufficient for academics to make a dent.”

Mr Bommasani notes that while there are many open-source AI model providers, the market is concentrated, and if major actors change their approach to access or pricing, cost dynamics could shift significantly. The economic benefit of open-source is also contingent on the data environment. For instance, there are emerging restrictions on tools that trawl the public internet for content, which may require alternative data sourcing approaches. Data licensing companies and products may emerge, which could make data acquisition more costly, says Mr Bommasani.



Chapter 2: Innovation and economic growth

The economic benefits of generative AI could vary widely depending on the extent of use and uptake. Modelling by Oxford Economics, a consultancy, shows that it could raise the GDP of the US by 1.8% to 4%, depending on the speed of adoption.²⁹ Innovations that give users more choice and flexibility could help deliver the upper estimates of impact.

Open-source can spur the creation of more products and services, allowing organisations of all sizes to use AI in ways that the developers of closed-source systems had never conceived. “A lot of AI use cases tend to be localised, whether because of language or because the features are on edge devices such as phones and cars or they are connected to a local company or initiative; these last-mile use cases really encourage local innovation,” says Ms Sharma.

Chapter two explores how open-source AI can support innovation and commerce, as well as the practical limits and constraints.

2.1 Pushing the science frontier

A commitment to openness has already transformed scientific research through AlphaFold, an AI platform developed by Google DeepMind. It can rapidly predict a protein’s structure from its amino acid sequence, a process that used to take months or even years.³⁰

AlphaFold is freely available to the scientific community, who can access 200 million protein structure predictions, supporting research in fields from antibiotic resistance and cancer therapy to crop yields.^{31,32,33}

Insilico Medicine is collaborating with a lab at Harvard University to develop an AI model, Precious-3 GPT, focused on ageing and related diseases. It can be used to investigate results from DNA and ‘omics’-based tests (related to biological molecules), which it combines with reports and published research to predict drug sensitivity and cellular interactions responsible for the ageing process.³⁴ Another project, OpenCRISPR, combines LLMs with CRISPR, a gene-editing technique, to create customisable gene-editing proteins. The initiative aims to extend CRISPR’s reach by making the platform free for researchers.³⁵ Another area where open-source systems are especially impactful is diagnostics and imaging. Tesseract-Medical Imaging is an open-source platform providing image viewing and reporting schemes for the field.³⁶

A broader open-source science initiative is the BLOOM multilingual LLM by Hugging Face, which aims to make research accessible across the world by generating text in 46 natural languages and 13 programming languages.³⁷ A team of around 1,000 volunteers from over 70 countries were involved in sourcing and building data sets to ensure an

inclusive and collaborative process. Scientists in academia, non-profits and smaller companies' research labs can face difficulties gaining access to LLM resources, as they are often exclusive to large industrial labs.³⁸ BLOOM aims to change this by making AI research accessible to these smaller-scale researchers, particularly those in poorer, non-English-speaking countries who have historically been unable to benefit from AI.

Mr Bommasani hopes open-source will allow academic researchers to make a more meaningful research contribution to the field of AI itself. "Having access to the data in addition to weights is useful for transparency about issues such as bias, privacy or copyright. You can do some science with black-box access to an application programming interface (API), but naturally, for scientific inspection, deeper access is more valuable."

Open-access AI can allow researchers to understand the dynamics of AI in a way that is not possible in fields like psychology. "Academia's own capacity to train foundation models is limited," says Mr Bommasani. Having access to model weights is a richer resource for understanding what the model does than being able to query it as a black box. Interpretability is important for understanding AI in a way that we can't understand humans.

2.2 Productivity gains

Commercially adopted open-source is already delivering economic gains at the micro and macro levels. Grammarly, for example, was able to ship newly released open-source LLMs into their experimental environment in as little as one day, according to Mr Roy-Chowdhury. It is becoming easier for companies to tinker with open-source models thanks to the development of integration software. ML-powered open-source software (MLOSS) can accelerate AI adoption, as they are shown to provide standardisation, experimentation and community creation.³⁹ An open-source developer community is more likely to make innovative code contributions,⁴⁰ while tools like GitHub Copilot are significantly boosting developers' productivity and efficiency (see Figure 3).

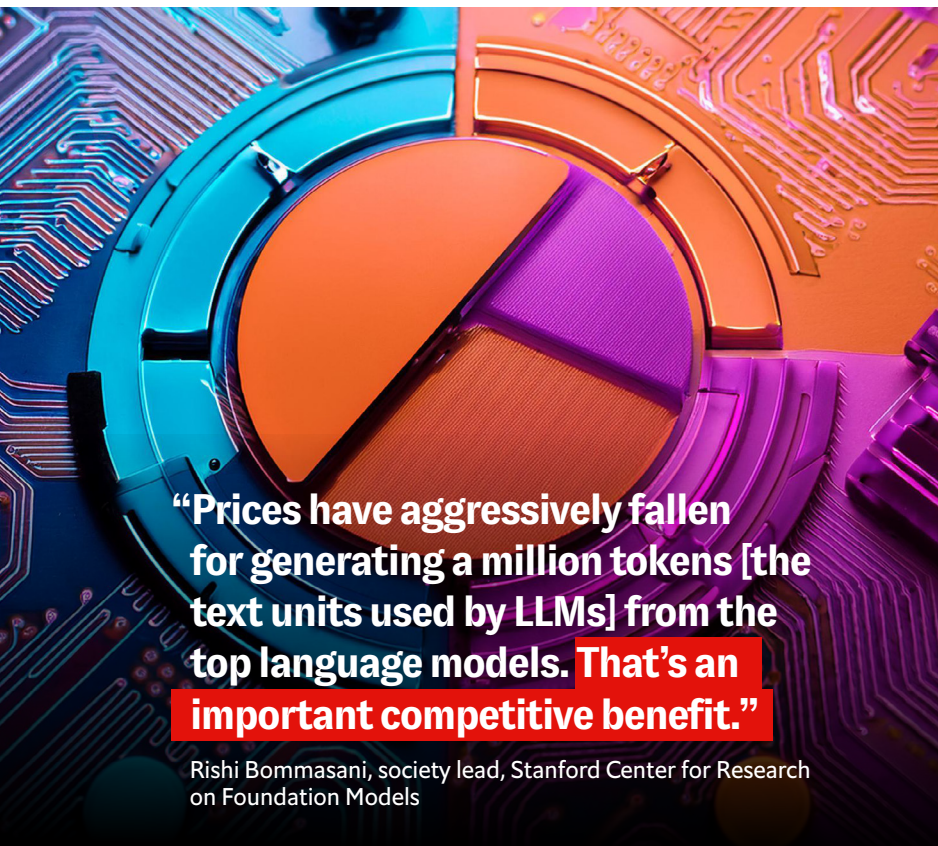
By standardising programming languages, frameworks and model designs, these tools lower the cost of training and skills among users and foster experimentation to find new use cases. In 2022 MLOSS models were estimated to have contributed US\$30bn to the global economy, representing a benefit-cost ratio of 100:1.⁴²

Start-ups and enterprises are building tools to help companies adopt and optimise open-source AI models. Unsloth, a language fine-tuning platform, is capable of making open-source models perform up to five times faster than competitors, with lower memory use. It makes LLMs specialised for specific tasks and industries

Figure 3: When using GitHub Copilot...



Source: GitHub⁴¹



“Prices have aggressively fallen for generating a million tokens [the text units used by LLMs] from the top language models. That’s an important competitive benefit.”

Rishi Bommasani, society lead, Stanford Center for Research on Foundation Models

in less time than other fine-tuning services, making innovative custom models easier to develop.⁴³ Landing AI is making computer vision accessible to a wider range of industries, from pharmaceutical research to automotives, via an open-source API.^{44,45}

Rasa AI is an open-source platform that has generated cost savings and commercial value by providing a platform for companies to build chatbots and access to an open-source community. One assessment of four companies found that the tool achieved a return on investment (ROI) of 181% over four years and improved developer productivity by 300-400%.⁴⁶

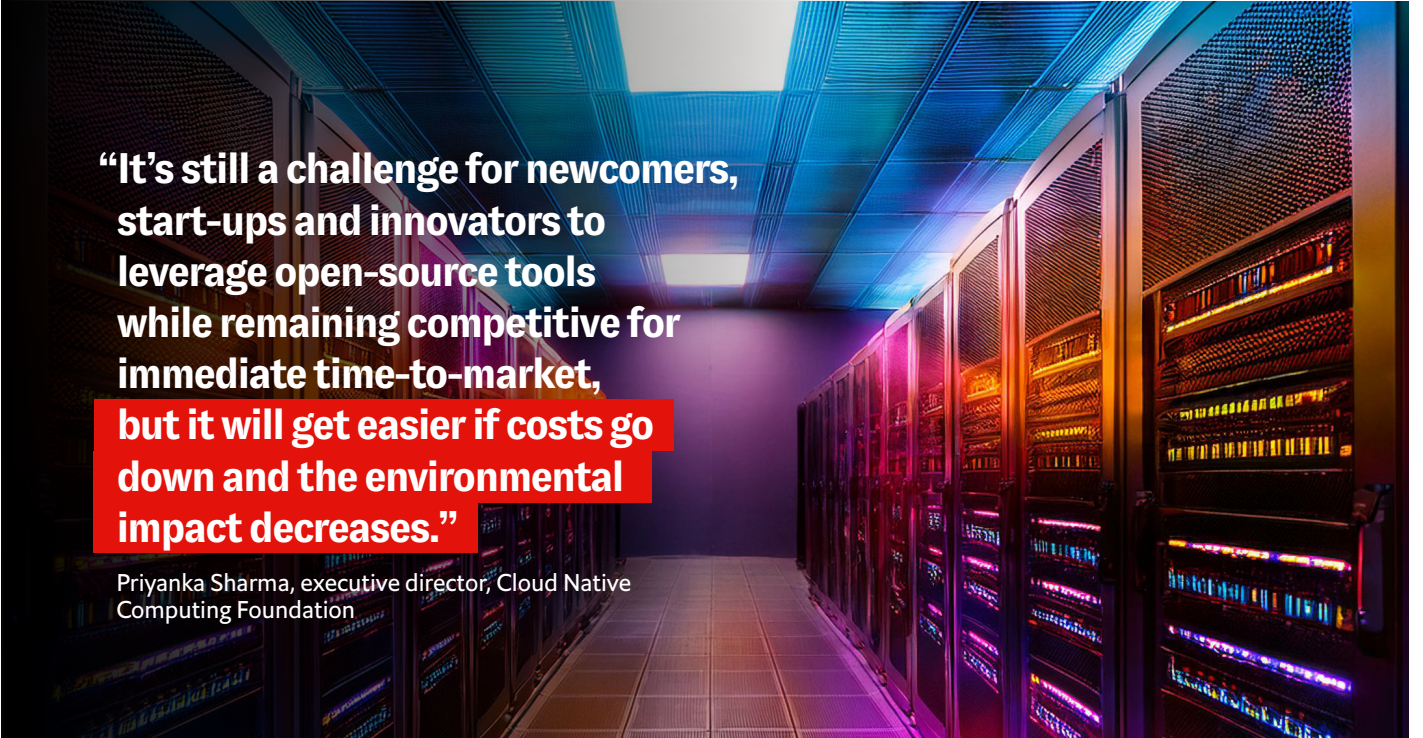
Open-source models may deliver a more efficient compute performance than proprietary ones, offering cost and environmental benefits. The carbon emissions generated during training were 20 times less for BLOOM compared with GPT-3, despite nearly identical model sizes.⁴⁷ Open-source providers are releasing pre-trained models with lower compute and thereby reduced emission

costs than closed source. The Hugging Face model was able to analyse a news corpus for US\$2.7 compared with GPT-4’s US\$3,061. Their CO₂ emissions stood at 0.12 kg CO₂ and 735-1,100 kg, respectively.⁴⁸

Looking forward, Mr Bommasani says open models could fuel a more competitive AI model landscape through more varied pricing solutions. “Prices have aggressively fallen for generating a million tokens [the text units used by LLMs] from the top language models. That’s an important competitive benefit.”

Much like the shift to cloud computing fuelled the emergence of new commercial juggernauts such as Salesforce, the demand for open-source AI has given rise to high-growth start-ups and unicorns, including Hugging Face and Mistral. It has also increased the commercial heft of players such as Databricks, which specialise in data infrastructure solutions. In 2023 Hugging Face was valued at US\$4.5bn, and is estimated to have seen its revenue jump by about fivefold from 2022 to 2023, while Databricks, valued at US\$43bn, saw year-on-year revenue increase by 50% for the year ending in January 2024.^{49,50}

“The business-to-business software-as-a-service sector is built on open-source data centres, powered by Linux,” says Mr Roy-Chowdhury. He believes that an open-source AI foundation could provide the engine for a similar commercial layering to emerge in AI. “The innovation will be at the application layer. We are going to have these incredible foundation models, which are the new platform on which people will build incredibly innovative applications... it is in all of our interests to make sure we have a playing field at the foundation model layer [through open-source].” He draws an analogy with the early years of cloud computing, when platforms came with vendor ‘lock-in’ and users could not easily move between clouds. Innovative applications only emerged later thanks to the emergence of a more open, multi-cloud environment.



“It’s still a challenge for newcomers, start-ups and innovators to leverage open-source tools while remaining competitive for immediate time-to-market, but it will get easier if costs go down and the environmental impact decreases.”

Priyanka Sharma, executive director, Cloud Native Computing Foundation

2.3 Going local

Open-source AI supports innovation at the grass-roots level in developing economies by allowing users to deploy and adapt models to solve their immediate challenges rather than respond to the priorities of technology developers in other markets and contexts. Jacaranda Health, a non-profit, developed PROMPTS as an AI and ML tool to support maternal health in Africa.⁵¹ The tool provides a sequence of messages at each stage of pregnancy and fast-tracks care for at-risk mothers. FoondaMate, or ‘study buddy’ in Zulu, is a learning support tool for middle and high school students, answering questions and providing revision resources and past practice tests.⁵²

2.4 Limitations and barriers

Quantifying value

A recent study estimated that open-source software and the supporting code-creation networks save firms around US\$8.8trn globally.⁵³ Without open-source software, companies would have to spend around 3.5 times more than they do already to create software and platforms that help run their business. In the EU, investments

of around €1bn in open-source software are estimated to have generated an economic impact of €65bn-95bn.⁵⁴

There are few hard numbers yet on the economic impact of open-source AI. Practically, companies often use multiple models, which complicates attempts to point to specific firm-level outcomes from open-source. “They might prototype with a closed-source model because the API is convenient and they don’t want dependence [on the provider], then, once committed, they will replace that pipeline by fine-tuning open models,” says Mr Bommasani.

Open-source is also hard to quantify “because it’s distributed and there is nobody who can naturally track it,” he adds. But the existence of firm numbers on the economic value of open-source software show it could be possible to develop the same in AI. “That is what I’d like to see; a more rigorous understanding of the magnitude of openness, especially when it comes to thinking about policy.” Mr Bommasani applauds attempts by the UK Competition and Markets Authority and the US National Telecommunications and Information Administration to focus on monitoring open models as a means of understanding how their benefits percolate

through the economy.^{55,56} “That is important for understanding the magnitude of the benefits of innovation. We can theorise how openness is good for innovation, but ultimately, we are going to want to quantify that,” he argues.

Companies need to develop ROI metrics to track the impact of open-source. There is excessive hype around AI, which even Silicon Valley-based venture capital firms like Sequoia believe has entered ‘bubble’ territory.⁵⁷ “Companies need to know that they’re actually gaining value from AI in general and open-source AI offerings in particular,” says Ms Sharma. “It’s still a challenge for newcomers, start-ups and innovators to leverage open-source tools while remaining competitive for immediate time-to-market, but it will get easier if costs go down and the environmental impact decreases.”

The process is far from easy. Companies have to make consequential decisions in terms of choosing between open models, the number of parameters and the “umpteen methods of fine-tuning,” says Professor Rao. “Quickly, it becomes a massive task of identifying; what are my needs and what am I trying to do? And therefore, what fine-tuning and models are needed. A wrong choice could put you on the wrong path for a couple of months, and that’s expensive from a corporate perspective.” Amplifying this issue, open-source systems often lack dedicated technical support.⁵⁸

Cost of failure

The history of open-source software suggests there are significant economic gains to be realised

from innovations that widen access and usage at the macroeconomic level. There might, however, be risks at the organisational level if companies take wrong turns in how they adapt and fine-tune models. They may, for example, fall foul of regulation through accidental or irresponsible use, for instance, risking fines, charges or legal action. Companies may use open-source to build models that later misfire. AI systems are domain-sensitive, which can lead to security or reliability issues when training them with data that are not representative of the original environment.⁵⁹ Although these risks are applicable more broadly to generative AI, they are accentuated in the context of open-source AI as, for instance, a larger number of organisations become subject to legal and regulatory risks.

If open-source models help nefarious actors target economic infrastructure or launch large-scale cyber-attacks, the knock-on effects could be significant; a cyber-attack against the US power grid could cost an estimated US\$240bn to more than US\$1trn.⁶⁰ Cyber criminals have already developed tools based on open-source models. FraudGPT and WormGPT, which can be used for phishing and malware attacks, are based on EleutherAI’s open-source LLM GPT-J.^{61,62} Malicious actors have used Stable Diffusion’s open-source image generation tool to create abusive and illegal imagery.⁶³ Other open-source packages like Python’s can be used to crack passwords and captchas.⁶⁴ All of these examples create commercial risk for companies caught up in the fallout.

Chapter 3: Transparency and trust

By making source code publicly available, open-source software is subjected to continuous peer review, thereby improving reliability and security.⁶⁵ One survey found that 89% of IT leaders believe open-source software is as secure or more secure for their enterprise than proprietary software.⁶⁶ Trust and transparency in AI models could also improve. The same hygiene effects seen in open-source software are evident in AI. Sixty-nine percent of respondents in a Linux Foundation survey believe that open-source generative AI leads to increased data control and transparency, resulting in more ethical and responsible AI.⁶⁷ For organisations in sectors with significant data

privacy and management requirements, such as healthcare and legal, open-source models offer stronger control and ownership.

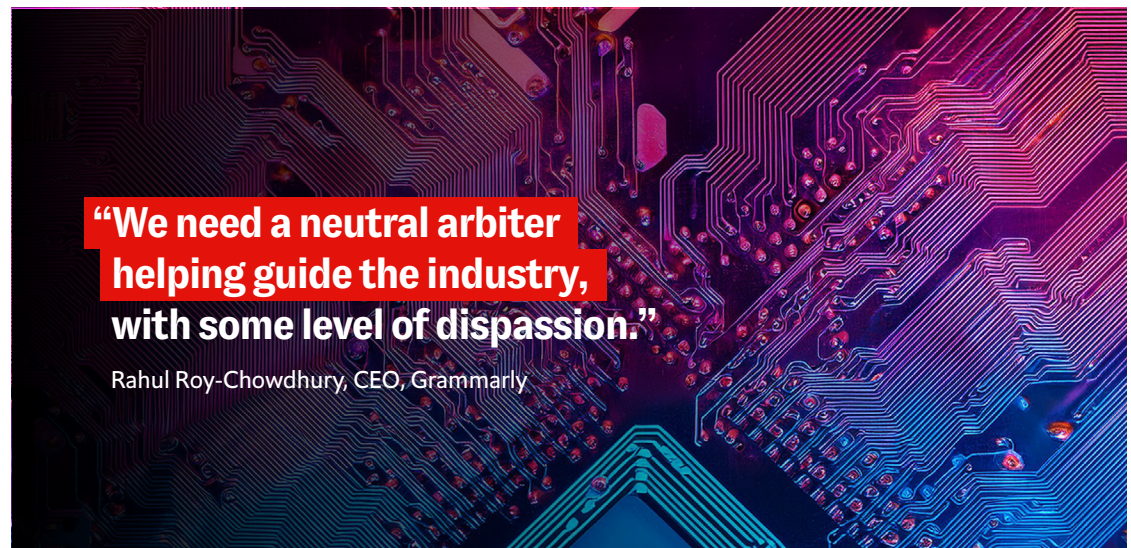
Chapter three details the salutary effects of open systems for reliability, scrutiny and security, as well as the benefits of open-source AI for transparency in wider fields like sustainability reporting. It then articulates how open-source systems could, without careful governance and management of issues like data quality, lead to less trustworthy outcomes.

3.1 Many eyes

Because models show pre-trained weights and the details of the architecture, metadata and algorithms, they let third parties inspect the inner workings, allowing more users to identify and respond to flaws.

“Having more eyes and more collaboration is the best way to solve problems, and that’s why open-source ends up being more secure,” says Ms Sharma. “In a small company with 50 employees, only 50 people can look, but if you’re open-source, then millions of people can look.” This also helps tackle the widening trust deficit in AI. Public trust in AI companies has decreased from 61% to 53% in the past five years.





Making a commitment to responsible AI can help companies overcome this obstacle.⁶⁸ In one study from Elsevier, a number of researchers and clinicians said that one of the top three disadvantages of AI is the lack of permission to use the data or information on which AI tools are trained; making their inner workings accessible for research and collaboration can solve this issue.⁶⁹

Exposing the inner workings of open-source systems also allows a cross-section of users and groups from fields such as ethics, sociology and psychology to collaborate with technology specialists to iterate on the quality and inclusivity of models.⁷⁰

3.2 Data ownership

Organisations must comply with both broad data management regulations like the EU’s General Data Protection Regulation (GDPR) and sector-specific frameworks, such as the Health Insurance Portability and Accountability Act (HIPAA) in the US. Further regulations may emerge to govern data management and commercialisation practices in specific verticals. By giving developers control of the models and data, and avoiding the need to share data with third-party model providers,

open-source AI strengthens ownership and management of data. Adjacent technology advances like on-device and edge AI offer further protection against the loss of data provenance, control and traceability.

3.3 Limitations and barriers

Data quality

The trust benefits of open-source AI depend heavily on users and developers identifying, curating and managing their data responsibly. Models alone cannot perform better than the data they are built on. Organisations can develop AI models based on their proprietary or unique data in order to gain a competitive edge, but they must invest in the appropriate checks and safeguards. Some organisations are turning to alternatives like synthetic data. While this has utility, research does call for caution due to the phenomena of ‘model collapse’ in which the quality of ML models degrades as models feed on their own outputs and become progressively dissociated from reality.⁷¹ While this is a challenge for the entire field, open-source might aggravate the problem because democratising AI allows more entities to build and run models and therefore risk flawed data practices.⁷²

Companies building or promoting open-source models that create synthetic data, can also put forward tools and guidelines for avoiding model collapse. For example, the use of LLMs to create labels and train models could be effective in detecting some forms of toxic content and improving content moderation.⁷³

Clarity and consensus

While open-source is associated with higher trust and transparency thanks to its decentralisation, there is a need for clear and shared definitions to ensure the terms are universally understood and used appropriately.⁷⁴ The EU AI Act defines free and open-source AI as systems where the source code or model is publicly available and can be used, modified and redistributed under licences. This went further than the common, narrower definitions, which tend to relate only to licensing.⁷⁵ Open Future, a European non-profit organisation, proposes a definition in which models qualify as open-source not simply by being publicly available, but by making various components public (including documentation, methods, weighting factors and architecture details).⁷⁶

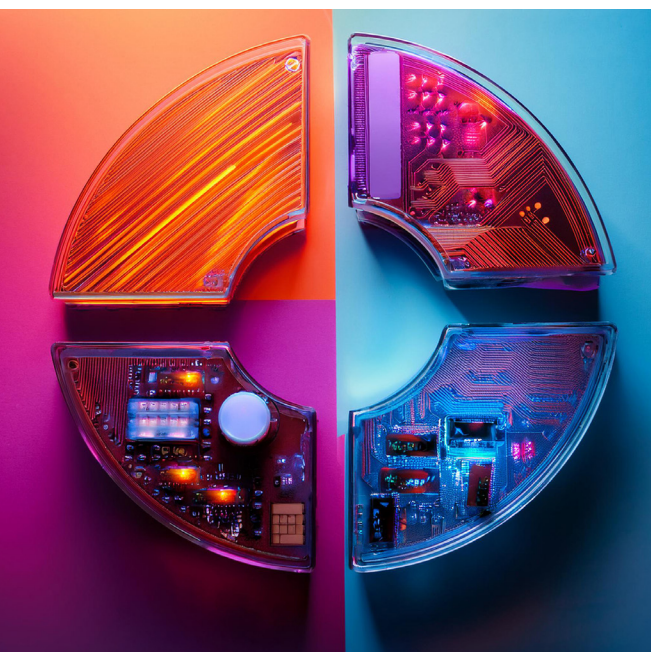
Some models have restrictions on usage or cannot be replicated, and critics point out that the computing power needed to build models makes the term 'open-source' somewhat meaningless without access to the computing power required to build and run AI models.⁷⁷ 'Open-weight' models

include information about pre-trained weights, which determine the model's behaviour, but the training data, algorithms and architecture are proprietary. Fully open-source models are fully transparent in both weights and the underlying data and architecture.⁷⁸ The Model Openness Framework refers to the most open category as Open Science, and adds Open Tooling and Open Models.⁷⁹

On August 22nd 2024, the Open Source Initiative, an industry body, issued a draft set of standards defining what counts as open-source artificial intelligence (AI). It stipulated that to qualify, developers of AI models must make available information about training data, source code and weights.⁸⁰ So far, a small number of models qualify on this definition and there remains debate within the industry about the evolution of the definition.

The open-source AI community can look to history to create an appropriate lead actor to set standards and direction in areas such as definitions and common frameworks. "We need a neutral arbiter helping guide the industry, with some level of dispassion," argues Mr Roy-Chowdhury, citing the Linux Foundation and Chromium, the project behind Google Chrome, guided by the World Wide Web Consortium. "These are industry experts who care only about making sure that as we evolve the tech, we are doing it in a way that is fair, equitable and transparent."

Conclusion



Open-source AI models have grown in number and potency in recent years, fanning optimism that they could augur a more democratised, innovative and competitive AI ecosystem. In a sector where progress has often been defined by the accomplishments of its most powerful participants, the promise of a more democratic model that enables smaller players to bring innovative ideas to market is understandably alluring.

Democratisation is not the only enticing aspect of open-source systems; in theory, at least, the very notion of an open and shareable coding structure allows researchers and users from a diverse range of backgrounds to take the model in new directions, boosting transparency and creating value. But there are practical limits and surmountable obstacles still to be overcome. Fine-tuning AI models requires expertise and resources. Some risk vectors may increase in a more open-source environment.

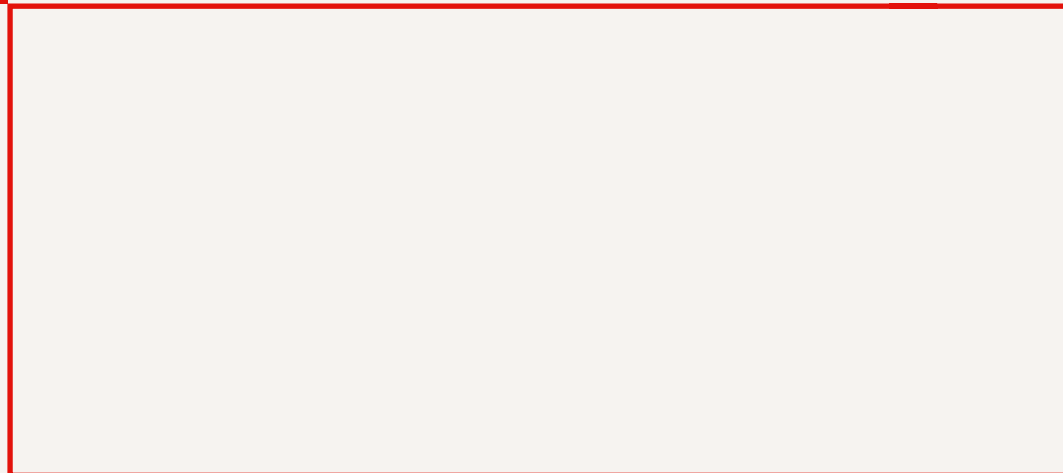
Regulators, developers, researchers and users must work together to develop a framework that minimises risk and bias, without sacrificing innovation. “Governments and regulators will need to make a lot of changes to address society’s needs as this technology grows at an unprecedented rate,” says Ms Sharma.

Endnotes

1. Computer, "A Brief History of Free, Open Source Software and Its Communities", February 2021, <https://www.computer.org/csdl/magazine/co/2021/02/09353517/1r8kwgBjU9W>
2. Open Source Initiative, "History of the OSI", September 19th 2006, <https://opensource.org/history>
3. European Parliament, "Challenges and limits of an open source approach to Artificial Intelligence", May 2021, [https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662908/IPOL_STU\(2021\)662908_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662908/IPOL_STU(2021)662908_EN.pdf)
4. HAI, Stanford University, "Artificial Intelligence Index Report 2024", aiindex.stanford.edu/wp-content/uploads/2024/05/HAI_AI-Index-Report-2024.pdf
5. TechCrunch, "How open-source software took over the world", <https://techcrunch.com/2019/01/12/how-open-source-software-took-over-the-world/>
6. European Parliament, "Challenges and limits of an open source approach to Artificial Intelligence", May 2021, [https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662908/IPOL_STU\(2021\)662908_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662908/IPOL_STU(2021)662908_EN.pdf)
7. Open Source Initiative, "The Open Source AI Definition – draft v. 0.0.3", <https://opensource.org/deepdive/drafts/the--ai-definition-draft-v-0-0-3>
8. Grammarly, "How Grammarly Is Writing New Standards for Responsible AI", August 16th 2024, <https://www.grammarly.com/business/learn/standards-for-responsible-ai/>
9. Hugging Face, "Pythia: A Suite for Analyzing Large Language Models Across Training and Scaling", April 3rd 2023, <https://huggingface.co/papers/2304.01373>
10. Stella Biderman, Hailey Schoelkopf, Quentin Anthony, et al, "Pythia: A Suite for Analyzing Large Language Models Across Training and Scaling", <https://arxiv.org/pdf/2304.01373v1>
11. See: <https://aif360.res.ibm.com/>
12. CNET, "Mozilla Launches Startup to Home In on 'Trustworthy AI'", March 22nd 2023, <https://www.cnet.com/tech/services-and-software/mozilla-launches-ai-focused-startup-called-mozilla-ai/>
13. See: <https://mozilla.vc/portfolio/?tag=13>
14. Mozilla, "Accelerating Progress Toward Trustworthy AI", February 22nd 2024, Accelerating Progress Toward Trustworthy AI - Mozilla Foundation
15. The Economist, "Why AI needs to learn new languages", January 24th 2024, <https://www.economist.com/science-and-technology/2024/01/24/why-ai-needs-to-learn-new-languages>
16. OECD, "OECD Digital Economy Outlook 2024 (Volume 1)", 14th May 2024, <https://www.oecd-ilibrary.org/docserver/a1689dc5-en.pdf?expires=1720712103&id=id&accname=guest&checksum=C5C45DFEE72B688A959C9653201067A4>
17. Ibid
18. See: <https://www.sarvam.ai/>
19. See: <https://vulavula.lalapla.ai/>
20. Ministry of Communications and Information Technology, "MCIT Announces the Participation of 4 Data Providers in Arabic GenAI Project (Fanan)", May 23rd 2024, <https://www.mcit.gov.qa/en/media-center/news/mcit-announces-participation-4-data-providers-arabic-genai-project-fanan>
21. Ibid
22. MSN, "Democratising AI: Here's how G42's new LLM JAIS 70B will help", <https://www.msn.com/en-ae/money/news/democratising-ai-here-s-how-g42-s-new-llm-jais-70b-will-help/ar-AA1og9Hd?ocid=BingNewsSerp>
23. HAI, Stanford University, "Artificial Intelligence Index Report 2024", p70, https://aiindex.stanford.edu/wp-content/uploads/2024/05/HAI_AI-Index-Report-2024.pdf
24. Partnership on AI, "Risk Mitigation Strategies for the Open Foundation Model Value Chain", July 11th 2024, <https://partnershiponai.org/resource/risk-mitigation-strategies-for-the-open-foundation-model-value-chain/>
25. ITI, "Authenticating AI Generated Content", January 2024, https://www.itic.org/policy/ITI_AIContentAuthorizationPolicy_122123.pdf
26. Turing, "Fine-Tuning LLMs : Overview, Methods, and Best Practices", <https://www.turing.com/resources/finetuning-large-language-models>
27. Towards Data Science, "Carbon Footprint of LLM Fine Tuning — A Case Study", February 22nd 2024, <https://towardsdatascience.com/carbon-footprint-of-llm-fine-tuning-a-case-study-7703afc716a9>
28. Kai Huang, Hanyun Yin, Heng Huang and Wei Gao, "Towards Green AI in Fine-tuning Large Language Models via Adaptive Backpropagation", <https://arxiv.org/abs/2309.13192>
29. Oxford Economics, "How GenAI will change the world economy", <https://www.oxfordeconomics.com/resource/how-genai-will-change-the-world-economy/>
30. DeepMind, "Putting the power of AlphaFold into the world's hands", July 22nd 2022, <https://deepmind.com/blog/article/putting-the-power-of-alphafold-into-the-worlds-hands>
31. See: <https://github.com/google-deepmind/alphafold>
32. See: <https://alphafold.com/>
33. The Guardian, "Google DeepMind's 'leap forward' in AI could unlock secrets of biology", May 8th 2024, <https://www.theguardian.com/science/article/2024/may/08/google-deepmind-ai-biology-alphafold>
34. BioPharmaTrend, "Insilico Medicine Unveils AI Assistant for Medical Research Drafting", July 24th 2024, <https://www.biopharmatrend.com/post/873-insilico-medicine-unveils-ai-assistant-for-medical-research-drafting/>
35. Business Wire, "Profluent Successfully Edits Human Genome with OpenCRISPR-1, the World's First AI-Created and Open-Source Gene Editor", April 2nd 2024, <https://www.businesswire.com/news/home/20240422399482/en/Profluent-Successfully-Edits-Human-Genome-with-OpenCRISPR-1-the-World%E2%80%99s-First-AI-Created-and-Open-Source-Gene-Editor>
36. See: <https://github.com/Tesseract-MI/Tesseract-MI>
37. See: <https://bigscience.huggingface.co/blog/bloom>
38. See: <https://huggingface.co/blog/bloom>
39. Langenkamp, Max and Yue, Daniel, "How Open Source Machine Learning Software Shapes AI", 2022, https://maxlangenkamp.me/AIES_paper.pdf
40. European Parliament, "Challenges and limits of an open source approach to Artificial Intelligence", May 2021, [https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662908/IPOL_STU\(2021\)662908_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662908/IPOL_STU(2021)662908_EN.pdf)

41. GitHub blog, "Research: quantifying GitHub Copilot's impact on developer productivity and happiness", May 21st 2024, <https://github.blog/news-insights/research/research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness>
42. Langenkamp, Max and Yue, Daniel, "How Open Source Machine Learning Software Shapes AI", 2022, https://maxlangenkamp.me/AIES_paper.pdf
43. GoOpenAI, "Unsloth: Unleashing the Speed of Large Language Model Fine-Tuning", April 20th 2024, <https://blog.gopenai.com/unsloth-unleashing-the-speed-of-large-language-model-fine-tuning-986ae7040711>
44. TechCrunch, "Landing AI brings in \$57M for its machine learning operations tools", November 8th 2021, <https://techcrunch.com/2021/11/08/landing-ai-machine-learning-operations-tools/>
45. LandingAI, "LandingAI Launches Docker Deployment for LandingLens Models", November 7th 2023, <https://landing.ai/news/landing-ai-launches-docker-deployment-for-landinglens-models>
46. Rasa, "ROI, Cost Savings, and Business Benefits of Rasa's Conversational AI", <https://info.rasa.com/forrester-tei-report-rasa-conversational-ai>
47. CIFAR, "Towards Measuring and Mitigating the Environmental Impacts of Large Language Models", September 2023, <https://cifar.ca/wp-content/uploads/2023/09/Towards-Measuring-and-Mitigating-the-Environmental-Impacts-of-Large-Language-Models.pdf>
48. Hugging Face, "Synthetic data: save money, time and carbon with open source", February 16th 2024, <https://huggingface.co/blog/synthetic-data-save-costs>
49. Reuters, "AI startup Hugging Face valued at \$4.5 bln in latest round of funding" August 24th 2023, <https://www.reuters.com/technology/ai-startup-hugging-face-valued-45-bln-latest-round-funding-2023-08-24/>
50. Yahoo Finance, "Databricks predicts 60% sales growth, anticipates \$2.4 billion annualized revenue", June 13th 2024, <https://finance.yahoo.com/news/databricks-predicts-60-sales-growth-161810048.html>
51. See: <https://jacarandahealth.org/prompts/>
52. TechCrunch, "Edtech Foondamte eyes speedy takeoff after \$2M funding", <https://techcrunch.com/2022/05/24/south-africa-edtech-foondamate-eyes-speedy-takeoff-after-2m-funding/>
53. Harvard Business School, "Open Source Software: The \$9 Trillion Resource Companies Take for Granted", March 22nd 2024, <https://hbswk.hbs.edu/item/-software-the-nine-trillion-resource-companies-take-for-granted>
54. HAI, Stanford University, Rishi Bommasani, Sayash Kapoor, Kevin Klyman, et al, "Considerations for Governing Open Foundation Models", <https://hai.stanford.edu/sites/default/files/2023-12/Governing-Open-Foundation-Models.pdf>
55. MSN "U.S. Commerce Department report endorses 'open' AI models", <https://www.msn.com/en-us/news/technology/us-commerce-department-report-endorses-open-ai-models/ar-BB1qRml3?ocid=BingNewsSerp>
56. Gov.uk, "Joint statement on competition in generative AI foundation models and AI products", July 23rd 2024, <https://www.gov.uk/government/publications/joint-statement-on-competition-in-generative-ai-foundation-models-and-ai-products/joint-statement-on-competition-in-generative-ai-foundation-models-and-ai-products>
57. Sequoia Capital, "AI's \$600B Question", https://www.sequoiacap.com/article/ais-600b-question/?utm_source=substack&utm_medium=email
58. European Parliament, "Challenges and limits of an open source approach to Artificial Intelligence", May 2021, [https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662908/IPOLE_STU\(2021\)662908_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662908/IPOLE_STU(2021)662908_EN.pdf)
59. Ibid
60. Allianz, "Cyber attacks on critical infrastructure", June 2016, <https://commercial.allianz.com/news-and-insights/expert-risk-articles/cyber-attacks-on-critical-infrastructure.html>
61. Wired, "Criminals Have Created Their Own ChatGPT Clones", August 7th 2023, <https://www.wired.com/story/chatgpt-scams-fraudgpt-wormgpt-crime/>
62. Security Intelligence, "Open source, open risks: The growing dangers of unregulated generative AI", <https://securityintelligence.com/articles/unregulated-generative-ai-dangers-open-source/>
63. CameraForensics, "The dark reality of Stable Diffusion", February 8th 2024,
64. Guembe, B., Azeta, A., Misra, S., et al, "The Emerging Threat of Ai-driven Cyber Attacks: A Review", 2022, Applied Artificial Intelligence, 36(1). <https://doi.org/10.1080/08839514.2022.2037254>
65. HAI, Stanford University, Rishi Bommasani, Sayash Kapoor, Kevin Klyman, et al, "Considerations for Governing Open Foundation Models", <https://hai.stanford.edu/sites/default/files/2023-12/Governing-Open-Foundation-Models.pdf>
66. Red Hat, "The State of Enterprise Open Source: A Red Hat report", <https://www.redhat.com/en/resources/state-of-enterprise-open-source-report-2022>
67. The Linux Foundation, "2023 Open Source Generative AI Survey Report", December 2023, https://www.linuxfoundation.org/hubfs/LF%20Research/GenAI_Report_2023_020524.pdf?hsLang=en
68. Axios, "1 big thing: Public trust in AI is sinking", March 5th 2024, https://www.axios.com/newsletters/axios-ai-plus-5c84b9b3-eaf4-48b1-98ce-0b5fe099d03c.html?chunk=0&utm_term=emshare#story0
69. Elsevier, "Insights 2024: Attitudes toward AI", <https://www.elsevier.com/insights/attitudes-toward-ai>
70. Alli Gee via Medium, "Reducing Bias in AI: The Power of Open-Source Data Science", <https://medium.com/@alligee/reducing-bias-in-ai-the-power-of--data-science-cdde5c32ac1c>
71. The Register, "What is Model Collapse and how to avoid it", January 26th 2024, https://www.theregister.com/2024/01/26/what_is_model_collapse/
72. IEEE Spectrum, "The Internet Isn't Completely Weird Yet; AI Can Fix That", 23rd June 2023, <https://spectrum.ieee.org/ai-collapse>
73. Kruschwitz, Udo and Schmidhuber, Maximilian, "LLM-Based Synthetic Datasets: Applications and Limitations in Toxicity Detection", May 2024, <https://aclanthology.org/2024.trac-1.6/>
74. Widder, David Gray and West, Sarah and Whittaker, Meredith, "Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI", August 17th 2023, <https://ssrn.com/abstract=4543807> or <http://dx.doi.org/10.2139/ssrn.4543807>
75. Open Future, "A Frankenstein-like approach: open source in the AI act", December 14th 2023, <https://openfuture.eu/blog/a-frankenstein-like-approach-open-source-in-the-ai-act/>
76. Ibid
77. The New York Times, "Some A.I. Companies Face a New Accusation: 'Open Washing'", <https://www.nytimes.com/2024/05/17/business/what-is-openwashing-ai.html>
78. Aruna Kolluru via Medium, "Exploring the World of Open Source and Open Weights AI", <https://medium.com/@aruna.kolluru/exploring-the-world-of-open-source-and-open-weights-ai-aa09707b69fc>
79. Matt White et al, "The Model Openness Framework: Promoting Completeness And Openness For Reproducibility, Transparency, And Usability In Artificial Intelligence", <https://arxiv.org/pdf/2403.13784>
80. Open Source Initiative, "The Open Source AI Definition – draft v. 0.0.9 – Open Source Initiative", <https://opensource.org/deepdive/drafts/open-source-ai-definition-draft-v-0-0->

While every effort has been taken to verify the accuracy of this information, Economist Impact cannot accept any responsibility or liability for reliance by any person on this report or any of the information, opinions or conclusions set out in this report. The findings and views expressed in the report do not necessarily reflect the views of the sponsor.



LONDON

The Adelphi
1-11 John Adam Street
London WC2N 6HT
United Kingdom
Tel: (44) 20 7830 7000
Email: london@economist.com

GENEVA

Rue de l'Athénée 32
1206 Geneva
Switzerland
Tel: (41) 22 566 2470
Fax: (41) 22 346 93 47
Email: geneva@economist.com

SÃO PAULO

Rua Joaquim Floriano,
1052, Conjunto 81
Itaim Bibi, São Paulo,
SP, 04534-004, Brasil
Tel: +5511 3073-1186
Email: americas@economist.com

NEW YORK

900 Third Avenue
16th Floor
New York, NY 10022
United States
Tel: (1.212) 554 0600
Fax: (1.212) 586 1181/2
Email: americas@economist.com

DUBAI

Office 1301a
Aurora Tower
Dubai Media City
Dubai
Tel: (971) 4 433 4202
Fax: (971) 4 438 0224
Email: dubai@economist.com

WASHINGTON DC

1920 L street NW Suite 500
Washington DC
20002
Email: americas@economist.com

HONG KONG

1301
12 Taikoo Wan Road
Taikoo Shing
Hong Kong
Tel: (852) 2585 3888
Fax: (852) 2802 7638
Email: asia@economist.com

SINGAPORE

8 Cross Street
#23-01 Manulife Tower
Singapore
048424
Tel: (65) 6534 5177
Fax: (65) 6534 5077
Email: asia@economist.com