https://doi.org/10.1038/s43588-023-00540-0

Building open-source AI

Yash Raj Shrestha, Georg von Krogh & Stefan Feuerriegel

Artificial intelligence (AI) drives innovation across society, economies and science. We argue for the importance of building AI technology according to open-source principles to foster accessibility, collaboration, responsibility and interoperability.

The computer science community has a long tradition of embracing open-source principles. However, companies increasingly restrict access to AI innovations. An example is OpenAI, which was founded to make scientific research openly available but which eventually restricted access to research findings. Although such a strategy reflects a company's legitimate incentive to obtain financial returns, such protection increases concentration of power, restricting access to AI technology. Further down the road, concentrated power could lead to growing inequality in AI research, education and public use. Here we discuss why proprietary AI technology should be complemented by open-source AI across the essential components for building AI technology: datasets, source codes and models.

Why exclusive proprietary AI technology is a problem

AI is a key technology that drives innovation across society, economies and science. For example, large language models (LLMs) such as GPT-4 have recently become the backbone of text processing in many fields such as education, entertainment, media and management. Therefore, downstream innovations including novel business models, products and services may be at risk when widespread access to AI becomes restricted. Not surprisingly, the concentration of power over technology is known to hamper future innovation, fair competition, scientific progress, and hence human welfare and development at large[1].

Proprietary AI technology could also jeopardize inclusiveness and responsibility. When new AI technologies like LLMs are developed exclusively by a few companies, those companies may also arbitrarily decide which countries and languages to support in their systems and may thus exclude some users, such as those from small markets (for instance, the Global South and rare languages). A certain level of openness of AI technology is further necessary for researchers to determine the safety, security and fairness inherent in AI systems operations. Proprietary AI systems are difficult for members of the public to appraise, and thus to identify and fix errors within them.

The benefits of open-source principles in software development

The cardinal idea of open-source software (OSS) is that an organization relies not only on its internal knowledge sources and resources to innovate, but also draws on multiple external technical sources, such as software packages, bug reports, customer feedback, published patents or communities[2]. Depending on the chosen license, OSS may not preclude commercialization: companies can combine OSS with additional products and services to generate revenue (for instance, RedHat offers Linux for free but charges for premium support, Amazon contributes to Apache but charges for hosting services in the cloud, and so forth). Nowadays, the open-source model safeguards effective and efficient software development. Yet it has taken several decades for this model to mature and for companies to realize and utilize its full potential[3]. Also, an important lesson from OSS is that governments played an important part in boosting OSS adoption[4], suggesting that many ways to promote open-source AI will benefit from governments taking an active role.

OSS offers several benefits relative to proprietary software in terms of accessibility, collaboration, responsibility and interoperability[5]. First, proprietary software is mostly available under licensing fees. In contrast, OSS is free and comes with no or limited restrictions on use, inspection and modification. Second, OSS tends to be developed and maintained by a community. Diversity in an OSS community stimulates improved software quality, faster innovation and increased creativity of OSS development relative to proprietary alternatives[6]. Third, errors in OSS are detected and corrected with everyday use, and much faster than in closed-source software[7], thus making OSS applicable to critical and highly reliable technical systems. Fourth, OSS typically relies upon open standards and modularity, which decouples dependencies among software components and leads to greater reusability and interoperability[6].

In addition to cost savings, companies also benefit from OSS in multiple ways. Companies may establish low-cost OSS software as their backbone and then build business models around complementary goods and services[8]. In addition, companies can gain trust and reputation by contributing to OSS, which attracts top talent and fosters the dissemination of their own technologies and products. By participating in OSS activities, companies can also steer the direction of innovation and control the further development of the technology on which they depend (for instance, by introducing a standard that helps the company to compete more effectively). By contributing to OSS, companies also gain valuable feedback on their technologies and are able to identify potential issues with the products early on. Finally, the presence of OSS drives innovation and competitiveness of commercial solutions.

Promoting open-source AI technology

The development of open-source AI and that of OSS share several similarities. However, there are also some important differences that require a tailored approach to building open-source AI. Whereas conventional software is programmed with explicit rules to perform a task, AI is programmed to learn to perform a task. As a result, AI technology has three essential components: datasets for training, source codes for formalizing the training task, and models that eventually store the trained weights. In addition, training AI models requires substantial hardware resources and comes with high operating costs. Furthermore, the use of AI may expose society to large risks (for example, the malicious use of AI to create misinformation), which mandates a responsible societal approach to open-source AI technology. Below we discuss a tailored approach to open-source AI complementary to proprietary AI

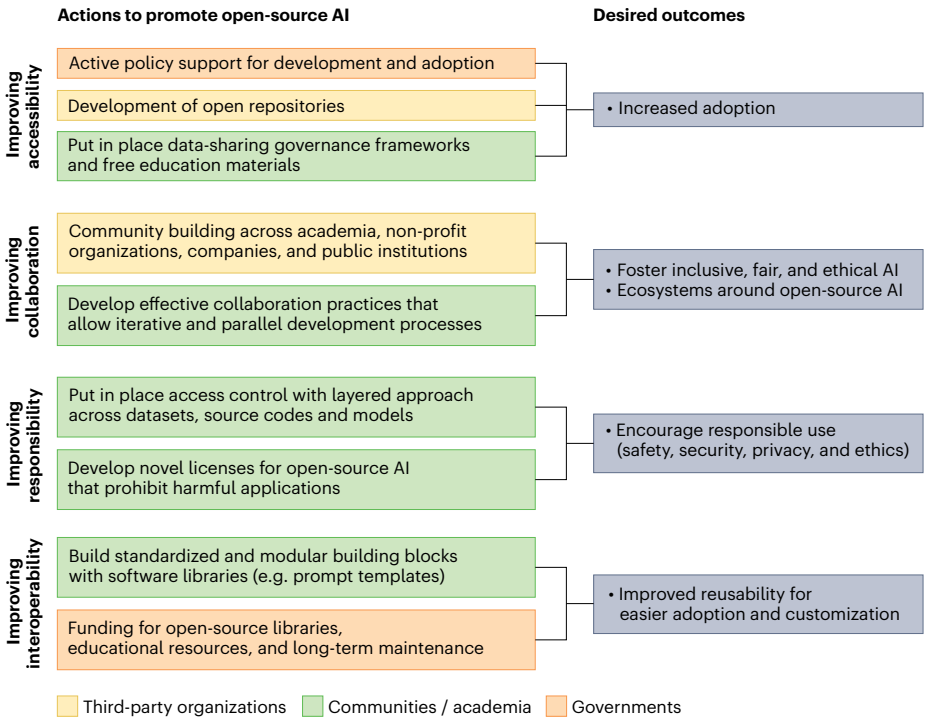

Fig. 1 | Key approaches to promote open-source AI technology. The suggested actions should foster accessibility, collaboration, responsibility and interoperability.

by fostering (1) accessibility, (2) collaboration, (3) responsibility and (4) interoperability (see Fig. 1).

Improving accessibility. To foster accessibility, policy-makers should proactively encourage the development and adoption of open-source AI. Since AI innovation is considerably more capital-intensive than regular software development, given the data and infrastructure needs of building contemporary AI models, additional resources (such as funding and access to large-scale infrastructure and data) are needed to kickstart and scale open-source AI technology. Importantly, existing computational resources are often not of sufficient magnitude to build state-of-the-art AI technology comparable to that of for-profit companies. For example, the development of a LLM is estimated to cost between 300 and 400 million euros. Another limiting factor is that, even if the resources are made available, they are often bound to academia and are thus inaccessible to other stakeholders such as non-profit organizations seeking opportunities where AI could be leveraged for social benefits. A promising counterexample is the US roadmap offering broader access to computational resources, including public–private partnerships. Scientists are currently often unable to replicate the AI technology obtained from companies owing to the lack of resources, so such roadmaps could help to facilitate reproducibility (for instance, via the ML Reproducibility Challenge).

To broaden access to data and models, policy-makers could support the development of open repositories for hosting both under a trustworthy and responsible governance model. Importantly, open datasets from public institutions are often large and originate from diverse sources, which is beneficial in practice. Furthermore, public institutions can actively incentivize data-sharing partnerships, which, in combination with federated learning, may promote AI across institutional boundaries while ensuring data privacy. For example, the German government recently launched a consortium called Mobility Data Space where different stakeholders in the mobility sector (such as public transport companies, private car-sharing providers and car manufacturers) are able to access shared data, even those of competitors.

However, data sharing comes with challenges. First, opening up datasets increases the likelihood of privacy breaches and raises ethical issues around confidentiality, data misrepresentation and informed consent. Second, to organize open data and to maintain fairness in terms of distribution rights and acknowledgments for its contributors is challenging. Fortunately, there has been recent progress with respect to the development of governance frameworks to tackle these challenges, such as the FOT-Net Data Sharing Framework, designed for connected automated driving under the General Data Protection Regulation in the European Union. Such frameworks could be useful starting points in improving accessibility while tackling the ethical, legal and organizational challenges.

Finally, much educational material on state-of-the-art AI is managed by for-profit companies (such as Coursera and Udemy) and is often hidden behind paywalls. Hence, to promote the adoption of open-source AI, more effort is needed to improve access to high-quality educational materials. As a result of the above, the barriers to entry for contribution and access to AI applications will drop considerably.

Improving collaboration. AI technologies may be jointly developed and maintained by diverse and inclusive communities of developers, users and stakeholders. This collaborative approach may greatly reduce the cost of development and contribute to solving scaling

problems. This will result in broad participation by stakeholders who can make the future of AI more inclusive and fairer.

To promote collaboration in open-source AI technology, clear steps should be taken towards building communities across academia, non-profit organizations, companies and public institutions. Given that the development of AI models is less easily decomposable into smaller tasks and that task division is more difficult than in standard software development, further effort is needed to develop suitable collaboration practices that allow for more iterative and parallel development processes. Here, the lessons learned from the project BigScience[9], where over a thousand volunteer scientists have assembled to develop an LLM called BLOOM[10], should be valuable. Furthermore, policy-makers should fund large-scale initiatives to produce open-source LLMs as complements to proprietary LLMs.

Creating synergies and networks between universities, research centers, government and industry may establish new ecosystems around open-source AI and become a driver for future innovation. Building such ecosystems is especially relevant for start-up firms, and small- and medium-sized enterprises[11] because they often lack the dedicated infrastructure and capacity to boost AI technology.

Improving responsibility. It is important to establish clear barriers against the misuse of AI technology. To this end, access control, similar to existing norms for open data, is needed to enforce the responsible use of open-source AI in practice. Consider, for example, MIMIC-III, a large, freely available health-related dataset. Given the sensitive nature of medical data, MIMIC-III is open to researchers only after they undergo compulsory ethics training. Similarly, access control for open-source AI should consist of a layered approach that varies appropriately across datasets, source codes and models to ensure responsible use, taking into account safety, security and privacy.

In addition, novel licenses are required—inspired by those for OSS but carefully tailored to open-source AI[12]. Such licenses must ensure broad user access while enforcing guidelines that prohibit malicious practices (such as abusing LLMs by automatically generating propaganda campaigns) under legally enforceable premises. Furthermore, such licenses for open-source AI should include sub-clauses that define permissive and restrictive use and also how the technology can or cannot be repurposed. Prominent examples are the RAIL licenses, which prevent irresponsible and harmful applications of AI technologies by granting permission only for certain use cases. Over time, customized variants of licenses for open-source AI could be developed, so that high-risk applications of AI technology are more restricted.

Similar to OSS, the development and use of AI technology under open-source principles will be especially effective in addressing bias in AI systems and steering innovations in a fair, ethical and trustworthy direction. First, owing to the diversity of inputs from stakeholders from around the world, there will be a greater emphasis on removing bias. Addressing bias will be as important when curating datasets as when training models. Second, a common concern is that open-source AI may not have the same level of quality control and testing as proprietary solutions, leading to potential bugs accidentally introduced by its developers. To this end, collaboration is important because it naturally leads to extensive testing.

Further, the development of AI in open communities may introduce decentralized organizations (that is, without authority hierarchies based on employment contracts). Many open communities have developed effective organizational structures based on merit, effort and expertise that are effective at resolving both coordination and cooperation issues, including how to manage conflicts. For instance, the Debian community developed a constitution that determines the decision-making rights of contributors and a set of rules that the community can refer to in case of conflicts or accountability issues. Lessons from communities such as Debian could be incorporated into a functional organizational structure and effective governance for open-source AI communities. Likewise, given that designated bodies for maintaining adherence to legal frameworks are typically missing and questions around accountability are often unclear, there can be legal challenges that originate from regulatory compliance. Nevertheless, open-source AI technology brings important principles to the table that go beyond existing regulatory frameworks for responsible and trustworthy use of AI.

It is also worth noting that there are privacy and security threats associated with the use of open-source AI. For example, malicious actors could perform backdoor attacks in which they manipulate a small portion of the training data to make an AI model learn additional, hidden functionalities[13]. In general, vulnerabilities in open-source AI are often public knowledge, which can make attacks but also their identification easier. Furthermore, there are also risks for society when open-source AI is used for nefarious purposes. Examples are the use of open-source AI technologies for the development of weapons and AI-generated propaganda campaigns[14]. Nevertheless, the benefits are likely to outweigh the downsides of open-source AI, especially if a responsible open-source approach with clear barriers against misuse is pursued, as laid out above.

Improving interoperability. Over time, AI technology will need to build upon more standardized and modular building blocks within software libraries (such as prompt templates and standardized prompt optimizers in the case of LLMs) that allow for easier adoption and customization in downstream applications. Interoperability of pre-trained models across platforms should also drastically reduce the need to retrain large models. The result will be a greater reusability of AI technologies, thus reducing the need to 'reinvent the wheel' and promoting faster iterations during development. Interoperability is not only important for rapidly building AI applications but also so that high-quality source codes and models designed in a responsible and robust manner can be reused.

In terms of standardization, various regulatory bodies such as the International Organization for Standardization have several standards under draft that aim at the harmonization of AI technology. The current initiatives cover various aspects including life cycle management, data quality, risk management and auditing. Such standardization roadmaps are helpful for developing trustworthy AI systems in high-risk applications (for example, through standardized conformity checks). Crucially, standardization must be brought to life through software libraries for developing AI technology. In this regard, public funding to support the development of open-source libraries could be necessary, as well as corresponding educational resources and long-term maintenance.

As a result of growing harmonization, dependence on a specific AI technology will diminish, so that end-users can avoid 'lock-in' effects and benefit from reduced switching costs (for instance, when changing from the LLM of company A to that of company B). For developers, interoperability can eventually help to counteract growing inequality in the development of, access to and use of AI technology, while also promoting effective competition. In this regard, a concern from a corporate perspective may be that, if AI research is forced to be open,

then companies may not see value in investing as much in research and development as they would otherwise do. For example, the motivation of companies to develop new AI technologies may be reduced in the presence of open-source alternatives, which may hamper innovation more broadly and could eventually also lead to gatekeeping behavior in established companies. However, we argue that the presence of open-source AI complementary to proprietary alternatives may increase healthy competition, which can also make commercial products better.

A call for open-source AI technology

We have argued that companies and society can benefit enormously from fostering open-source AI technology to complement proprietary alternatives. The broad adoption of open-source principles across datasets, models and source codes will foster accessibility, collaboration, responsibility and interoperability in AI technology and will help to reverse the growing inequality in AI research and thereby lower barriers to future innovation.

Yash Raj Shrestha[1], Georg von Krogh[2,3] & Stefan Feuerriegel[4,5]
[1]Faculty of Business and Economics (HEC), University of Lausanne, Lausanne, Switzerland. [2]Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland. [3]ETH AI Center, ETH Zurich, Zurich, Switzerland. [4]Institute of AI in Management, LMU Munich, Munich, Germany. [5]Munich Center for Machine Learning, Munich, Germany.
e-mail: feuerriegel@lmu.de

Published online: 26 October 2023

References

1. Aghion, P., Harris, C., Howitt, P. & Vickers, J. *Rev. Econ. Stud.* **68**, 467–492 (2001).
2. Chesbrough, H. W. *Open Innovation: the New Imperative for Creating and Profiting from Technology* (Harvard Business Press, 2003).
3. Fitzgerald, B. *MIS Q.* **30**, 587–598 (2006).
4. Lerner, J. & Schankerman, M. *The Comingled Code: Open Source and Economic Development* (MIT Press, 2013).
5. Haefliger, S., Von Krogh, G. & Spaeth, S. *Manage. Sci.* **54**, 180–193 (2008).
6. Von Krogh, G. & Von Hippel, E. *Manage. Sci.* **52**, 975–983 (2006).
7. Paulson, J. W., Succi, G. & Eberlein, A. *IEEE Trans. Softw. Eng.* **30**, 246–256 (2004).
8. Golden, B. *Succeeding with Open-Source* (Addison-Wesley Professional, 2005).
9. Akiki, C. et al. BigScience: a case study in the social construction of a multilingual large language model. In *NeurIPS Worksh. on Broadening Research Collaborations in ML* (NeurIPS, 2022).
10. BigScience Workshop (Scao, T. L. et al.) BLOOM: A 176B-parameter open-access multilingual language model. Preprint at https://arxiv.org/abs/2211.05100 (2022).
11. Jacobides, M. G., Brusoni, S. & Candelon, F. *Strat. Sci.* **6**, 412–435 (2021).
12. Contractor, D. et al. Behavioral use licensing for responsible AI. In *ACM Conf. on Fairness, Accountability, and Transparency (FAccT)* (ACM, 2022).
13. Saha, A., Subramanya, A. & Pirsiavash, H. Hidden trigger backdoor attacks. In *AAAI Conf. on Artificial Intelligence* (AAAI, 2020).
14. Feuerriegel, S. et al. *Nat. Human Behav.* (in the press).

Author contributions

All authors wrote, edited and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature Computational Science* thanks Giada Pistilli and the other, anonymous reviewer(s) for their contribution to the peer review of this work.