Task Force 05
INCLUSIVE DIGITAL TRANSFORMATION

Democratizing AI for the Public Good: Key Concepts and Recommendations

Lea Gimpel, Digital Public Goods Alliance (Norway)

Daniel Brumund, GIZ "FAIR Forward - Artificial Intelligence for All" (Germany)

Alek Tarkowski, Open Future Europe (Poland)

Maximilian Gahntz, Mozilla Foundation (Germany)

Urvashi Aneja, Digital Futures Lab (India)

Vukosi Marivate, University of Pretoria (South Africa)

Anita Gurumurthy, IT for Change (India)

TF05

Abstract

AI technology can be a powerful tool to advance the SDGs if designed, applied, and governed responsibly. Recent research shows that the AI infrastructure stack – from hardware to cloud infrastructure, data, models, and the application layer – is highly skewed towards a few companies, countries, and communities. At the same time, we see a move towards closed research and less transparency despite increased "open-washing". These developments result in economic, social, and political challenges: limited innovation and quality; risks of bias and disinformation; digital extractivism; and a lack of democratic control and digital sovereignty.

The policy brief outlines how the layers of AI democratization (use, development, benefit, governance) can be addressed. Specifically, the policy brief considers open-source AI as digital public goods instrumental for making progress toward AI democratization and accelerating the attainment of the SDGs. It focuses on:

- Providing public AI infrastructure and developing a shared strategy for developing public generative AI models

- Creating, curating, and governing high-value open data sets

- Developing industry standards for the safe disclosure of model weights

- Harmonizing AI governance for the public good

Diagnosis of the Issue

Recent progress in artificial intelligence (AI) capabilities demonstrates the potential of these technologies towards beneficial outcomes for individuals and societies. The G20 have identified AI technologies as a potential means to mitigate inequalities and foster sustainable development. In the 2023 Hiroshima process, the G7 committed to prioritizing AI for global benefit, tackling issues like climate change, education, and global health, aligned with the SDGs. Similarly, the 2024 UN resolution on AI for sustainable development solidifies member states' ambitions to use safe, secure, and trustworthy AI systems to tackle global challenges and inequalities, including open-source AI.

The present market concentration in AI impedes its potential for humanity at large and exacerbates existing inequalities within and across societies. The AI technology stack is heavily centralized and controlled by few economic actors and geographic regions (Vipra and Myers West 2023). Machine learning, the primary technique utilized in contemporary AI, displays natural monopoly characteristics (Narechania 2021). These include competition restraints due to concentration at the infrastructure layer, exclusive access to training data, and high costs for switching cloud providers (Besiroglu et al. 2024). Companies training advanced foundation models benefit from economies of scale, yielding troves of proprietary data due to a broad customer base, access to capital, and skilled workers (Küsters and Kullas 2024).

There is a trend towards less transparent research and increased "open-washing", despite a rise in openly available models on platforms like HuggingFace. Most high-quality foundation models lack transparency, making it hard to scrutinize their workings

or biases (Castelvecchi 2016). Model evaluations are often irreproducible due to isolated sharing of results without sufficient details on methods and data (McIntosh et al. 2024).

The term *open-source AI* lacks a precise definition and can apply to various components, including openly released model weights, training data, or code. Reproducibility and transparency require more than releasing model weights (White et al. 2024). AI owes much of its progress to researchers adhering to open-science principles and grassroots organizations like EleutherAI and BigScience sharing their models. This also contributed to advancements in AI safety (Mozilla 2024).

Market concentration and lacking access to AI development resources cause economic, social, and political challenges: limited innovation and quality; risks of discrimination and disinformation; a lack of democratic control; limited accountability and digital sovereignty. From a global majority perspective, the implications are even more acute: Advancements and applications of AI that exploit historically marginalized, equity-seeking communities through neo-colonialist data extraction practices worsen prevailing disparities (Birhane 2020; Muldoon and Wu 2023) and perpetuate dual objectification of people as marginalized and objectified (Kadiri 2021).

How do we ensure that communities disadvantaged by market structures are not left behind, are represented in their cultural, linguistic, and social diversity; and can develop AI tools that meet their own challenges and aspirations?

Addressing these concerns and reaping AI's benefits by the global majority are core issues for the G20. Yet, targeting market concentration is challenging. It requires concerted efforts towards democratizing AI assets. This necessitates leveling the playing field by breaking down barriers to entry and ownership, including access to and control over infrastructure, pre-trained models, and high-quality training data, and fostering diversity in AI research, development, and governance. Otherwise, the global majority

cannot harness AI's potential for socio-economic development, thereby exacerbating global economic and social disparities. Countries in the global majority will lack access to AI assets needed to develop fit-for-purpose solutions and value-adding services, while remaining dependent on costly commercial products that may perpetuate digital extractivism and neo-colonialism, undermine sovereignty, and neglect countries' contexts. Popular, predominantly English-trained generative AI models are a pertinent example of the lacking language diversity and cultural representation of the global majority in AI development[1].

As the G20 recognized, open-source AI is crucial to making progress in democratizing AI and gearing AI development toward the public good for the global majority. Open-source AI can support broad accessibility to AI technologies and grant the freedom to utilize, study, share, and modify underlying models. Yet, concentrated market and infrastructure power impede open-source AI potentials (Gray Widder, West, and Whittaker 2023). This policy brief recommends democratizing AI systems as digital public goods (DPGs) while outlining open-source AI challenges[2].

---

[1] Even with the advance of multilingual language models, key challenges regarding context, quality, equity and exploitation remain. See:

https://cdt.org/wp-content/uploads/2023/05/non-en-content-analysis-primer-051223-1203.pdf and https://blog.papareo.nz/whisper-is-another-case-study-in-colonisation/

[2] The UN Secretary General's *Roadmap for Digital Cooperation* defines digital public goods as open-source software, open standards, open data, open AI systems, and open content collections that adhere to privacy and other applicable laws and best practices, do no harm, and help attain the Sustainable Development Goals (SDGs). See also:

https://digitalpublicgoods.net/digital-public-goods/

Recommendations

Open-source technologies are foundational to the global digital economy by enabling everyone to use, study, modify, and distribute these artifacts. Today, these technologies undergird safety and robustness of the internet and create trillions in value (Hoffmann, Nagle, and Zhou 2024). Open-source can have a similar impact on AI development globally. To support the development of AI technologies as DPGs and empower global majority countries, the following recommendations are put forward:

Provide public AI infrastructure and develop a shared strategy for developing public generative AI models

Inequalities in accessing computing power are symptomatic of market concentration in AI. The scarcity of computing power causes geopolitical tensions: The "compute divide" impacts public interest in AI development and disadvantages companies and researchers in global majority countries (Besiroglu et al. 2024). To fully democratize AI, these disparities must be addressed through investment in publicly accessible computational infrastructure. While it is hard to match the scale of computing power by the largest commercial actors, it is possible to create computing infrastructure that sufficiently supports multitudes of AI research and development.

Recommendations:

● Establish an international institution to support AI research with publicly accessible computational resources for advancing public-interest AI, complementing proposals for international AI governance bodies. Several G20 members have already initiated programs to provide computational resources for AI development. Their national focus, however, limits bridging the computational gap. International collaboration is

essential to elevate the provision of competitive computational resources and other elements of the AI stack, such as developer and safety tools, for research and public interest purposes on a global scale (a "CERN for AI"). This would ensure access to critical computational resources, reduce switching costs if the public offering is competitive with market offers, and enable international AI research initiatives and the development of AI solutions serving the public good. This could potentially be funded by a Global Fund for AI, proposed in the Global Digital Compact at the time of writing.

● Ensure transparent, fair public procurement rules, including interoperability requirements. In cases where the provision of public computing infrastructure relies on commercial offerings, safeguarding fair competition and the users' ability to change providers without extra costs will be instrumental. Such procurement rules should consider local hosting requirements to strengthen national/regional digital economies where appropriate safeguards exist, e.g., comprehensive data protection regulation and cyber security capabilities. Competition authorities should move swiftly to counteract anti-competitive practices and partnerships and mitigate market concentration and unfair barriers to entry across the AI stack.

● Develop a coordinated strategy among G20 members to create public generative AI models, including large language models (LLMs); smaller, more specialized models; and fine-tuned models. While several national initiatives aim to recreate powerful models, evidence suggests other viable development approaches, some of which are less resource-intensive and deliver results on par with LLMs (Gunasekar et al. 2023). Developing a coordinated strategy, informed by established best practices, will help define the most effective approaches to building public AI systems and components.

Promote creation, curation, and governance of trusted high-value open datasets

Lacking availability and access to high-quality data impedes the development of AI-based tools that help foster sustainable development globally. Major commercial generative AI systems are trained on either proprietary data or data scraped from the public web, often in unsustainable ways (Baack 2024). Few efforts exist to develop and share datasets as DPGs. The fact that Wikipedia is a core asset in almost all LLMs proves the value of such datasets. Most such initiatives are driven by non-profit actors with limited resources[3]. There is a need to support work to repurpose available collections[4] and create new datasets. Data governance models that balance sharing with respect for data subject rights must complement, especially datasets incorporating sensitive data, such as health data.

Recommendations:

- Invest in and facilitate the creation of high-value open datasets that benefit societies, the environment, and economic development. This could build on the EU's efforts to identify, describe the characteristics of, and unlock such datasets. These efforts should focus on open government data, additional data sources, and creating new datasets, e.g., for under-resourced languages. The authors recommend the creation of a funding

---

[3] Examples worth mentioning include Mozilla's Common Voice platform – a platform for collecting and making available voice data for around the world, or Language Model Evaluation Harness, a set of benchmarking datasets built by Eleuther.ai.

[4] Examples of existing collections include, for instance, Open Access and Open Data repositories, or digital library resources.

vehicle for such datasets, e.g., modeled on the Lacuna Fund[5], and a unified data space for high-value datasets, which incorporates the DPG Standard[6] for datasets and harmonizes technical standards across jurisdictions to enable widespread use and impact. This also necessitates increased standardization and/or codes of practice for responsible data collection; transparent documentation; and diversity and equity through inclusive governance.

- Boost research and innovation by enshrining the right to access data as a mandatory requirement (e.g., the right to access platform data stipulated in the EU Digital Services Act); and adopting licensing agreements in the public sector that foster data commons while lowering the cost of transactions and respecting data subjects' right to privacy, e.g. the Open Data Commons Licenses framework (Benhamou and Dulong de Rosnay 2023). It should include a commitment to utilize public procurement as an investment vehicle for supporting the open-source AI ecosystem by preferring open

---

[5] The Lacuna Fund is a private initiative funded by the Rockefeller Foundation, Google.org, IDRC, the German development cooperation and several other philanthropies to support the development of datasets to solve social challenges through machine-learning in low- and middle-income countries. The fund has reached between 2020-2022 54 low- and middle-income countries for data collection and funded datasets were downloaded more than 407k times. See Lacuna Fund's 2023 evaluation report for more information and lessons learned

[6] For more information on the Digital Public Good Standard by the DPGA, see: https://github.com/DPGAlliance/DPG-Standard/blob/main/standard.md

solutions over proprietary AI, and the G20 could serve as a forum for exchanging best practices and facilitating learning in this regard.

Develop industry standards for the safe disclosure of model weights and related AI assets

There is no defined standard for safe release practices of foundational AI models (Solaiman 2023). Policymakers must develop detailed standards that balance the safe dissemination of model components with the benefits of an open-source approach. What constitutes meaningful openness (DPGA 2023) must be defined against the backdrop of potential substantial risks and how to achieve this tradeoff. The G20 must ensure harnessing the public benefits of powerful AI models by enabling access, transparency, and reuse, while safeguarding against harm.

Recommendations:

● Develop a multidimensional framework to govern the release of foundational AI models, including open release practices, informed among other initiatives by the public consultation process of the US National Telecommunications and Information Administration and the EU's AI Act to ensure coherence across instruments. Based on national/regional frameworks, this effort should also include working towards "global technical standards for interoperable and trustworthy AI," (MoFa Japan 2019), for instance, through GPAI. To ensure legitimacy and broad support, the G20 should take a multi-stakeholder approach, including the open-source community, industry partners, standard-setting bodies, civil society, and policymakers. Where appropriate, policy instruments ranging from incentives and voluntary commitments to legal and regulatory options must be explored and defined to ensure compliance. Regulatory safeguards and

obligations should further consider who develops open-source technology, the compliance burden, and its risk profile. This supports appropriate oversight and independent audits of foundational models, and their planned release practices. It needs to be complemented by investments in AI safety norms and practices to harness the benefits of open-source AI while minimizing its risks.

Harmonize AI governance for the public good

AI governance globally is fragmented, with over 1000 policy initiatives implemented across some 70 countries. This lack of multi-stakeholder cooperation results in disparate regulatory frameworks and inconsistent ethical standards. The absence of a unified global governance framework causes governance interoperability issues and divergent standards for AI data labeling, testing, data protection, and exchange protocols, risking privacy and algorithmic biases. The UN advocates for a unified global governance framework to mitigate deployment risks and bridge regional and sectoral governance divides. The Hiroshima Process Guiding Principles underscore the need for international technical standards and datasets to promote responsible AI use.

Recommendations:

● Establish a taskforce to increase interoperability and harmonization of national and international AI policy initiatives, support implementation and promote international learning, thereby avoiding governance silos and increasing the strength of AI governance instruments globally. This helps bridge the capacity gap between countries in crafting policy and regulatory frameworks for AI and strengthen existing efforts towards building a unified global AI governance framework. These efforts must be co-designed with global majority countries, often excluded from, or underrepresented in such initiatives.

● Develop and test a model for multi-stakeholder, participatory governance for foundational AI technologies, including advisory bodies, citizen assemblies and participatory design. Collectively, such bodies should define public benefit and ensure foundational AI models are rooted in human rights and promote planetary well-being. This should include capacity development efforts for stakeholders in global majority countries to ensure meaningful participation and consistently implemented public audits and oversight.

● Establish a code of conduct for developing training datasets to build AI solutions that demand implementation safeguards and check for biases and risks early and throughout AI lifecycles. This helps avoid risks from non-transparent AI solutions, especially in global majority countries.

Scenario of Outcomes

There is widespread debate about open-source foundational AI models' risks and benefits. Advantages include transparency, adaptability, customization to fit novel contexts, domains, and languages, and the ability to scrutinize models with the right expertise. Open-source AI assets also accelerate research, including in fields such as AI safety. Yet, generative AI models are dual-use in nature, with the risk of being misused by malicious actors. Several studies posit substantial risks of open-source AI models, including undermining democracies by spreading dis- and misinformation and generating hazardous bio- and chemical weapons (Seger et al. 2023). Most of these assertions lack validation[7].

Concerns about AI safety extend beyond open models and are also valid for closed models. This policy discourse requires greater precision and rigor. One promising avenue is the concept of marginal risk: evaluating whether open models intensify risk compared to closed models and scrutinizing how open models aid in risk mitigation (Kapoor et al. 2024). AI safety research is optimally advanced by sharing model components and

---

[7] One noteworthy exception pertains to harmful images generated by AI, such as non-consensual intimate imagery (NCII). Researchers have already initiated the development of mitigation strategies, including the curation of training data, regulation of prompts for output generation, and implementation of filter options. For further details, refer to: Qu, Y., Shen, X., He, X. et al. (2023). "Unsafe Diffusion: On the Generation of Unsafe Images And Hateful Memes From Text-to-Image Models." Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security. Available at: https://doi.org/10.48550/arXiv.2305.13873.

associated knowledge artifacts (Mozilla 2023). The current swift advancement in AI capabilities and safeguards would be impossible without it.

By developing policy instruments that promote AI systems and their components as DPGs, the world community can counter an increasingly present AI monoculture, and support global majority countries in seizing AI's potential for economic prosperity and people's and planetary well-being.

References

Baack, Stefan. 2024. 'Training Data for the Price of a Sandwich: Common Crawl's Impact on Generative AI'. Mozilla. https://assets.mofoprod.net/network/documents/Common_Crawl_Mozilla_Foundation_2024.pdf.

Benhamou, Yaniv, and Melanie Dulong de Rosnay. 2023. 'Open Data Commons Licenses (ODCL): Licensing Personal and Non-Personal Data Supporting the Commons and Privacy'. SSRN Scholarly Paper. https://doi.org/10.2139/ssrn.4662511.

Besiroglu, Tamay, Sage Andrus Bergerson, Amelia Michael, Lennart Heim, Xueyun Luo, and Neil Thompson. 2024. 'The Compute Divide in Machine Learning: A Threat to Academic Contribution and Scrutiny?' arXiv. https://doi.org/10.48550/arXiv.2401.02452.

Birhane, Abeba. 2020. 'Algorithmic Colonization of Africa'. *SCRIPT-Ed* 17 (2): 389–409. https://doi.org/10.2966/scrip.170220.389.

Castelvecchi, Davide. 2016. 'The Black Box of AI'. *Nature* 538 (October): 20–23.

DPGA. 2023. 'Exploring a Gradient Approach to the Openness of AI System Components'. *Digital Public Goods Alliance*. 27 October 2023. https://digitalpublicgoods.net/blog/exploring-a-gradient-approach-to-the-openness-of-ai-system-components/.

Gray Widder, David, Sarah West, and Meredith Whittaker. 2023. 'Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI'. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4543807.

Gunasekar, Suriya, Yi Zhang, Jyoti Aneja, Caio César Teodoro Mendes, Allie Del Giorno, Sivakanth Gopi, Mojan Javaheripi, et al. 2023. 'Textbooks Are All You Need'. arXiv. https://doi.org/10.48550/arXiv.2306.11644.

Hoffmann, Manuel, Frank Nagle, and Yanuo Zhou. 2024. 'The Value of Open-Source Software'. Working Paper 24–038. Harvard Business School. https://www.ssrn.com/abstract=4693148.

Kadiri, Aisha P.L. 2021. 'Data and Afrofuturism: An Emancipated Subject?' *Internet Policy Review* 10 (4). https://doi.org/10.14763/2021.4.1597.

Kapoor, Sayash, Rishi Bommasani, Kevin Klyman, Shayne Longpre, Ashwin Ramaswami, Peter Cihon, Aspen Hopkins, et al. 2024. 'On the Societal Impact of Open Foundation Models'. arXiv. https://doi.org/10.48550/arXiv.2403.07918.

Küsters, Anselm, and Matthias Kullas. 2024. 'Competition in Generative Artificial Intelligence'. 6. cepInput. Centrum für Europäische Politik. https://www.cep.eu/fileadmin/user_upload/cep.eu/Studien/cepInput_Generative_Artificial_Intelligence/cepInput_Competition_in_Generative_Artificial_Intelligence.pdf.

McIntosh, Timothy R., Teo Susnjak, Tong Liu, Paul Watters, and Malka N. Halgamuge. 2024. 'Inadequacies of Large Language Model Benchmarks in the Era of Generative Artificial Intelligence'. arXiv. https://doi.org/10.48550/arXiv.2402.09880.

MoFa Japan. 2019. 'G20 AI Principles'. G20 Summit - Outcome Documents. Ministry of Foreign Affairs Japan. https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/pdf/documents/en/annex_08.pdf.

Mozilla. 2023. 'Joint Statement on AI Safety and Openness'. Mozilla. 31 October 2023. https://open.mozilla.org/letter/.

———. 2024. 'Accelerating Progress Toward Trustworthy AI'. Version 0.9. Mozilla. https://assets.mofoprod.net/network/documents/Mozilla_Accelerating_Progress_Toward_Trustworthy_AI_v2.pdf.

Muldoon, James, and Boxi A Wu. 2023. 'Artificial Intelligence in the Colonial Matrix of Power'. *Philosophy & Technology* 36 (4): 80. https://doi.org/10.1007/s13347-023-00687-8.

Narechania, Tejas N. 2021. 'Machine Learning as Natural Monopoly'. SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.3810366.

Seger, Elizabeth, Noemi Dreksler, Richard Moulange, Emily Dardaman, Jonas Schuett, K. Wei, Christoph Winter, et al. 2023. 'Open-Sourcing Highly Capable Foundation Models: An Evaluation of Risks, Benefits, and Alternative Methods for Pursuing Open-Source Objectives'. *LPP Working Paper* 2023 (4). https://doi.org/10.2139/ssrn.4596436.

Solaiman, Irene. 2023. 'The Gradient of Generative AI Release: Methods and Considerations'. arXiv. https://doi.org/10.48550/arXiv.2302.04844.

Vipra, Jai, and Sarah Myers West. 2023. 'Computational Power and AI'. *AI Now Institute*, 27 September 2023. https://ainowinstitute.org/publication/policy/compute-and-ai.

White, Matt, Ibrahim Haddad, Cailean Osborne, Xiao-Yang Liu, Ahmed Abdelmonsef, and Sachin Varghese. 2024. 'The Model Openness Framework: Promoting Completeness and Openness for Reproducibility, Transparency and Usability in AI'. arXiv. https://doi.org/10.48550/arXiv.2403.13784.