T20 Policy Brief

G20 BRASIL 2024

T20 Brasil 2024 Let's rethink the world

Task Force 05

INCLUSIVE DIGITAL TRANSFORMATION

The Unseen Layers of AI: An Exploration of Poor Data Provenance in Model Training

Mohammad Hamdy, Chief Legal and Policy Officer, Almond FinTech (USA); Editor in Chief for Policy, This Week in Fintech (USA)

Mona Hamdy, Chief Executive Officer, Anomaly (Global)

TF05

Abstract

The opacity of AI model training presents a complex challenge with extensive implications. Training datasets are commonly compiled using methods like web scraping and data aggregation, often without explicit permission, attribution, or regard for intellectual property rights (IPRs). Tracing the origins of this data, known as data provenance, has become a focal point in recent lawsuits involving developers seeking to safeguard intellectual property rights (IPRs).

AI opacity not only threatens IPRs. With limited visibility into the training process, users' ability to assess output quality or biases is severely undermined, potentially leading to uninformed use, industry "monocultures," and systemic risks. Economically, poor data provenance may exacerbate inequalities, privileging data providers in the Global North over their Global South counterparts, who face greater challenges in asserting their data rights. Additionally, it poses regulatory challenges for national authorities tasked with protecting citizens' privacy, possibly triggering complex legal disputes and prompting risk-averse regulators to deny developers access to data. Consequently, these jurisdictions could be denied the opportunity to participate in AI development. Culturally, AI opacity hampers user assessment of model representativeness, which could threaten linguistic and cultural diversity and perpetuate the exclusion of certain cultures or groups.

This Policy Brief urges G20 countries to enhance data provenance in AI through regulatory and technological means. It provides an overview of various regulatory avenues for data provenance regulation and assesses their potential for success, highlighting the crucial role of the G20 in strengthening these standard-setting endeavors. The Policy Brief also explores the promise of emerging technologies in enhancing transparency within the AI sector and advocates for G20 support for these technologies as an additional means to promote transparency through market competition.

Diagnosis of the Issue

In 2023, during the Indian Presidency, the G20 discussed developing a governance framework for responsible technologies. This led to the unanimous approval of the global Digital Public Infrastructure (DPI) framework, which seeks to ensure equitable access to the advantages of digital technologies through secure and interoperable systems in the critical sectors of identification, payments, and data sharing. The framework emphasizes using open standards to build such systems with a view to advancing the Sustainable Development Goals (SDGs). The G20 also stressed the need for ongoing research in key areas of digital innovation, such as design, governance, and socioeconomic impacts.

Yet, despite the progress achieved through the DPI framework and the commitment to advacing research, the G20 countries have yet to agree on a robust AI framework. The 2019 G20 AI Principles, approved during Japan's Presidency, currently serve as the primary AI governance framework. These principles emphasize key objectives in the AI space, including inclusive growth, sustainable development, human-centered values, transparency, robustness, security, and accountability. However, translating these high-level objectives into practice requires the development of detailed standards. Absent such standards, the G20 countries, like most countries worldwide, are left without any widely accepted tools to tackle the substantial challenges posed by AI, including the systemic, economic, regulatory, and cultural risks arising from inadequate data provenance.

The systemic risk presented by opaque AI model training stems from users' inability to assess output quality or identify biases before using AI models. Users can only detect these issues after the fact, which is far less optimal than if users are aware of model limitations upfront and can better mitigate them. With limited prior knowledge, uninfomed users have less incentive to diversify among models, potentially giving rise to

"monocultures" in entire economic sectors. The lack of diversification could amplify model errors, increasing vulnerabilities and giving rise to systemic risk within and across sectors.

Moreover, poor data provenance in AI could deprive the Global South of some of the economic opportunities arising from AI, which in turn could exacerbate global inequality and exclusion. With business models based on untransapent AI training, data providers worldwide are hardly compensated. They face a collective action problem, as the low returns on their data may not always justify asserting their rights against developers. This issue is particularly acute in the Global South. While data providers in the Global North benefit from access to better governance systems and greater resources to assert their rights, those in the Global South lack access to both effective institutions and adequate means to monitor the usage of their data in AI training, let alone protecting their rights. These dynamics could become an independent source of global economic disparities, adding to the more commonly debated negative impacts of AI, such as job displacement.

Poor data provenance also presents significant challenges for national law enforcement agencies tasked with safeguarding citizens' privacy and enforcing data protection laws. With limited insight into developers' use of citizen data for AI model training, these agencies struggle to ensure compliance and protect privacy rights. Risk-averse regulators or those operating in jurisdictions with inadequate data protection frameworks may choose to restrict developers' access to protected data to mitigate risks. As a result, they may deny these jurisdictions the opportunity to participate in AI development. Alternatively, these agencies may engage in complex legal disputes with developers, even in the absence of actual privacy breaches. Ultimately, instead of accelerating AI advancement, these regulatory hurdles may hinder progress and widen the AI gap between leading and lagging countries.

Lastly, opaque model training, particularly of large models intended for global use, increases the risk of cultural exclusivity. Profit-driven developers may prioritize datasets from the Global North over those from the Global South due to factors such as accessibility, quality, or the use of more dominant languages, which could undermine linguistic and cultural diversity in the AI space. This could further perpetuate systemic underrepresentation or misrepresentation of certain constituencies, particularly those from the Global South and minority groups, and weaken their data providers' incentives to participate in AI development. The result is a less inclusive AI ecosystem that fails to cater to the diverse cultural needs of global users.

Recommendations

Collectively accounting for 80% of global GDP and serving as home to leading AI developers worldwide, the G20 countries have a moral obligation to promote AI transparency while fostering innovation. One way they can fulfill this obligation is by enhancing data provenance in AI, utilizing both regulatory and technological approaches. The G20 countries can lead the development of regulatory frameworks — whether international, national, or industry-specific — that establish clear transparency standards for the sourcing and use of data in AI training, thereby mitigating the various risks outlined in this Policy Brief. The standards should be supplemented by oversight bodies equipped with audit and sanction powers, which will help standardize practices, promote compliance, and ensure that the industry is effectively contributing to the achievement of the SDGs. Furthermore, the G20 countries are hosting nascent AI initiatives characterized by high levels of transparency, particularly open-source and decentralized, blockchain-based AI models. The G20 countries can not only guard against the risks presented by AI opacity but also steer AI innovation toward more transparency by supporting these new AI initiatives and reinforcing competition between them and traditional models. Through this dual strategy, the G20 countries can promote responsible AI while ensuring a robust AI economy estimated to contribute nearly $16 trillion to the global economy by 2030.

1. The Regulatory Approach: policymakers have three potential avenues for developing data provenance standards in AI: international, national, or industry-specific.

   a) Developing binding international standards for data provenance in AI presents significant challenges, particularly due to the divergent regulatory approaches among countries. Despite these hurdles, notable progress has been made,

primarily the unanimous approval of the UN General Assembly's comprehensive resolution on AI in March 2024. Although non-binding, this resolution represents one of the most comprehensive AI frameworks to date. It calls for the adoption of interoperable standards for data provenance, although without specifying detailed rules or institutional mechanisms for implementation. Nonetheless, given the present difficulties in achieving global consensus over detailed, binding AI standards, as well as the fact that the G20, although wielding significant influence in the AI space, is not a global forum, developing international standards for AI data provenance at the G20 would be impractical, at least in the forseable future.

b) The pursuit of national policies coordinated through frameworks like the G20 seems to be a more viable option for advancing data provenance in AI within the current landscape. Several G20 countries have taken significant strides in AI regulation, with a particular focus on improving data provenance. For instance, the US Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI mandates the development of effective labeling and content provenance mechanisms. Similarly, the EU AI Act stipulates documentation requirements for high-risk AI models, including details on data provenance. China's Gen AI Regulation addresses various issues, such as transparency, bias, and using inclusive data for training models, albeit with limied detail. While demonstrating a good understanding among policymakers of the importance of furthering AI transparency to combat immediate risks like deep fakes, these regulatory frameworks largely overlook the broader risks associated with poor data provenance, as outlined in this Policy Brief. Most frameworks only address data provenance requirements in passing, without clear guidance on crucial issues such as interoperability, operationalization, and enforcement. With its membership

comprising jurisdictions at the forefront of AI regulation globally, the G20 is uniquely positioned to serve as a forum for discussing and coordinating national policies on AI data provenance. By playing this role, the G20 can inform and support effective standard-setting efforts in member countries and beyond, in a bottom-up fashion, and even in the absence of binding international rules.

Recommendation 1: The G20 should serve as a platform for the coordination of members' data provenance policies, with a view to developing detailed standards and fostering consensus in a bottom-up fashion.

c) Developing industry-specific standards presents another avenue for establishing data provenance standards in AI. Initiatives like the Data & Trust Alliance and the Data Nutrition Project exemplify this approach by advocating for documenting and tracking the metadata of training datasets and setting standards for the types of data permissible in training. However, such initiatives face major challenges, prominent among which is the development of sufficiently detailed, enforceable standards, the avoidance of competition or fragmentation among frameworks developed by different initiatives, and the attainment of a high level of compliance despite their voluntary nature. As hosts of the most prominent industry-specific standard-setting initiatives in the world, the G20 countries can enhance transparency in AI at the industry level by collaborating with these initiatives and serving as a forum for coordination among them to mitigate the risks of competition or fragmentation.

Recommendation 2: The G20 should collaborate with and serve as a forum for coordination among industry-specific standard-setting efforts in member countries to enhance transparency in AI at the industry level.

2. The Technological Approach: the G20 can reinforce transparency and improve data provenance throughout the AI ecosystem by promoting market competition with alternative business models.

a) One option would be to support open-source AI initiatives, which offers full visibility into how models are trained and the potential sources of errors or bias in their outputs. Open-source AI can foster continuous improvement in AI model design and training, democratizing technology development and mitigating the risks associated with AI opacity. Industry groups supporting open-source AI should thus be supported by G20 countries as enablers of a more open and transparent AI ecosystem.

b) Another option is to promote the utilization of blockchain technology in AI development, which goes beyond open-source methodologies. By integrating blockchain into AI model design and training, the lineage of data is permanently preserved in a completely transparent way, facilitating attribution and fair compensation for data providers worldwide. Furthermore, the full decentralization of AI model development better enables all countries, cultures, and groups to participate in the development of the technology on an equal footing, all while equitably sharing the economic benefits from the technology.

Recommendation 3: The G20 countries should steer AI development by supporting alternative AI business models that could enhance data provenance in AI through market competition with traditional AI business models.

Scenario of Outcomes

Improving data provenance in AI holds the promise of addressing many of the myriad risks arising from AI opacity. Through increased transparency, the systemic, economic, regulatory, and cultural risks outlined in this Policy Brief could be mitigated or altogether prevented. For instance, with more information available to users about training datasets, users would be better able to differentiate between AI models, recognize inherent vulnerabilities ex ante, and diversify their usage, thus significantly reducing the potential for systemic risk. Additionally, informed users would more easily distinguish between "hallucinations" and biases, which would facilitate a constructive feedback loop for developers. This, in turn, would enhance model refinement and overall quality.

Furthering economic disparities—another risk exacerbated by poor data provenance—can be alleviated through the establishment of frameworks for licensing data and compensating data providers. Ultimately, the aim of increasing transparency in AI should not be to hinder technological advancement but to fairly distribute its economic benefits. While elaborating on the contours of these compensatory frameworks goes beyond the purview of this Policy Brief, it is crucial to highlight that creating such mechanisms would simply be infeasible with opaque AI models. This is because opaque models make it very hard, if not impossible, for data providers to monitor the use of their data in the first place. With more transparency, as well as proper licensing and compensatory frameworks, marginalized groups would be better incentivized to produce high-quality, licensable data, which would promote broader participation and more equitable distribution of the economic surplus derived from the new technology.

Transparent AI models would also address the regulatory risks pertaining to privacy by facilitating direct engagement between developers and regulators. With increased

visibility into the data used to train AI models, regulators can more effectively fulfill their mandates by ensuring developers' compliance with privacy regulations and implementing tailored safeguards to protect user privacy. This proactive approach not only enhances user trust but also alleviates the concerns of risk-averse regulators, who may be inclined to restrict or prohibit the new technology altogether as a first resort.

Furthermore, improved data provenance empowers users to address the cultural risks stemming from opaque AI. With more transparency, informed users can advocate for better inclusion of underrepresented groups and regions and request the development of models that serve the specific linguistic and cultural contexts where they are deployed. Developers, armed with a better understanding of consumer preferences and cultural peculiarities, can fine-tune their models to cater to the specific needs of minority groups or marginalized regions, fostering a more inclusive AI ecosystem.

Yet despite the preceding benefits, policymakers must grapple with the trade-off between the desire for more transparent AI models and the urgency to lead the AI innovation race. As nations compete to attract top-tier AI ventures, they run the risk of compromising regulatory standards to lure these projects, sparking a detrimental race to the bottom. To address these challenges and ensure fair competition among nations, international coordination within the G20 is crucial. The G0 can foster collaboration and consensus-building among member countries and, hence, mitigate the temptation to weaken regulatory standards while promoting responsible AI development.

Another formidable challenge lies in the development of detailed, interoperable standards and enforcement mechanisms that can enhance data provenance in the AI space without hindering technological innovation. Achieving this delicate balance requires harmonizing diverse regulatory frameworks to establish consensus on minimum standards that are coherent and effective. In this context, the G20 emerges as the ideal

platform for leading AI nations to convene, share insights, and establish common regulatory grounds.

To meet these challenges, the G20 should embrace the dual strategy outlined in this Policy Brief. By leveraging collective expertise and fostering collaboration among its members, the G20 can spearhead the development of interoperable, inclusive, and forward-thinking regulatory standards for data provenance in AI, a necessity not only for the G20 countries but also for the global community at large.

References

Computer Security Resource Center. "Glossary: Data Provenance." *National Institute of Standards and Technology*. Accessed April 22, 2024. https://csrc.nist.gov/glossary/term/data_provenance.

"Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence." *The White House*. Last modified October 30, 2023. Accessed April 22, 2024. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

GDPIR. "Global DPI Repository." ICD, *Ministry of Electronics & Information Technology, Government of India*. Accessed April 22, 2024. https://www.dpi.global/home/aboutus.

General Assembly. "Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development." *The United Nations*. Last modified March 11, 2024. Accessed April 22, 2024. https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf?token=y3BzDdQQTeLJysR7Rk&fe=true.

Heath, Ryan. "Decentralizers Look to Break Giants' Hold over AI." *Axios*. Last modified April 2, 2024. Accessed April 22, 2024. https://www.axios.com/2024/04/02/ai-decentralized-big-tech-blockchain.

Office of the Central Cyberspace Affairs Commission. "Interim Measures for Generative Artificial Intelligence Service Management." *Cyberspace Administration of China*. Last modified July 10, 2023. Accessed April 22, 2024. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm.

OECD AI Policy Observatory Portal. “SEC Chief Warns AI 'Monoculture' Could Create 'Pretty Fragile Financial System.’” *Oecd.ai*. Accessed April 22, 2024. https://oecd.ai/en/incidents/62229.

PricewaterhouseCoopers. “PwC’s Global Artificial Intelligence Study: Sizing the Prize.” *PwC*. Last modified 2017. Accessed April 22, 2024. https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf.

“Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts.” *Council of the European Union*. Last modified January 26, 2024. Accessed April 22, 2024. https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf.

Roberts, Megan. “The G20 Gathers: A Digital Consensus amid Wider Divisions.” *German Marshall Fund of the United States*. Last modified September 6, 2023. Accessed April 22, 2024. https://www.gmfus.org/news/g20-gathers-digital-consensus-amid-wider-divisions.

“The AI Alliance.” *Thealliance.ai*. Accessed April 22, 2024. https://thealliance.ai/.

“The Data & Trust Alliance.” *Data & Trust Alliance*. Accessed April 22, 2024. https://dataandtrustalliance.org/.

“The Data Nutrition Project.” *Datanutrition.org*. Accessed April 22, 2024. https://datanutrition.org/.

“TF-2: Our Common Digital Future: Affordable, Accessible and Inclusive Digital Public Infrastructure.” *ThinkTwenty (T20) India 2023 - Official Engagement Group of G20*. Accessed April 22, 2024. https://t20ind.org/taskforce/our-common-digital-future-affordable-accessible-and-inclusive-digital-public-infrastructure/.

"G20 AI Principles." *The G20*. Last modified June 9, 2019. Accessed April 22, 2024. https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/pdf/documents/en/annex_08.pdf.

G20 BRASIL 2024
BUILDING A JUST WORLD AND A SUSTAINABLE PLANET

T20 Brasil 2024
Let's rethink the world

Let's rethink the world

CEBRI Rethink Tank

FUNDAÇÃO ALEXANDRE DE GUSMÃO

ipea
Instituto de Pesquisa Econômica Aplicada