

Frontier Science & Technology

# Securing Access to Foreign Data Flows for Al

Tim Hwang and Joshua Levine

## SUMMARY

For the United States to remain on the cutting edge of artificial intelligence (AI) development, model developers need access to novel, high-quality, and underused data sets. This is true for frontier model advancement, but even more so for ensuring that models can be effectively fine-tuned for accomplishing specific tasks such as industrial operations, drug development, and climate prediction that contribute to scientific discovery, economic dynamism, and national security. This policy should be advanced by directing the United States Trade Representative (USTR) within the Department of Commerce to prioritize working with foreign governments, particularly those the US already has strong relationships with, to establish a policy of licensing crucial data to be used to train AI models. The USTR should also work with Congress when crafting new trade agreements or treaties to include language that expressly calls for data sharing and unencumbered cross-border data flows for the purpose of training AI models.

### PROBLEM

Now that leading AI model developers have scraped the web and have incorporated most, if not all, publicly available data to train their models, they are increasingly seeking access to high-quality proprietary datasets to drive system improvements. The US government has an opportunity to secure the nation's technological advantage in AI by negotiating and securing access to key data flows on behalf of US industry. Specifically, access to data from allies in areas of strategic importance will support continued building and fine-tuning of AI models to support diffusion domestically, while also establishing mutually beneficial relationships globally around a key input for future AI model development. Where possible, the US may also aim to secure these arrangements on an exclusive basis, denying access to geopolitical rivals as they attempt to catch up in AI.

Exclusive agreements are particularly important in the context of competition with China. Presently, Chinese model developers enjoy the benefit of a domestic legal framework that facilitates firm access to information, as well as efforts by regional governments, such as those in Shanghai and Shenzhen, to collect and curate datasets to spur AI development. The federal government should promote these exclusive agreements to mitigate potential data shortages in the near term, while extending existing norms related to cross-border data flows and AI model training over the medium to long term. Such agreements could be a critical plank of Western collaboration around the sharing and use of key inputs for training AI models as a counter to the techno-authoritarian ecosystem being developed by the People's Republic of China and its collaborators.

Pursuing these agreements would establish a clear vision for US trade policy specifically as it relates to AI development, and broadly to other technologies. Ensuring that such data is accessible for training can build upon existing norms related to cross-border data flows for the AI era. Such a posture would strengthen collaboration on a critical technology with allies and support America's own domestic AI industry.

## SOLUTION

#### Executive

- The president should direct the USTR to include, when amending or initiating new trade agreements, specific language to promote and protect the sharing of data for domestic AI model training. Such a directive should instruct USTR to prioritize securing access to high-value, hard-to-access data sources surrounding scientific research, health data, and industrial tasks and operations. Datasets specifically focused on heavy industry and manufacturing, telecommunications network operations, geospatial and environmental areas, labor force participation, and transportation flows could all support US policy objectives domestically and around the world.
- The president should form a Presidential Commission on Data Acquisition (PCDA) that would bring together key model developers to advise the White

House and USTR on the evolving data needs within the AI industry and help prioritize targets for acquisition and data flow agreements.

- The president should direct the Director of the Office of Science and Technology Policy (OSTP) to produce a report identifying opportunities to create broader data alliances to facilitate and incentivize exchange of key types of data between the US and its allies to create mutual acceleration for each nation's respective AI industries. This would result in a report identifying the global landscape of "critical" data flows and offering potential trade agreement structures. The report should also explore opportunities to deny geopolitical rivals access to crucial datasets as a means of preserving US advantage.
- The president should direct the USTR and the Bureau of Industry and Security within the Department of Commerce to study and propose a framework for securing data access between the US and allies to mitigate cyberespionage threats that arise when data is acquired, transferred, and accessed by firms in the US, if deemed necessary. Such a framework could also provide an opportunity to update existing frameworks used for current agreements for cross-border data flows.

#### Congressional

The Senate Finance Committee and House Ways and Means Committee are the committees of jurisdiction for the USTR. As part of their oversight authority, they should conduct hearings with the USTR focused on improving relationships between the US and allies with regard to digital trade and pursuing agreements to support access to training data for AI models.

## JUSTIFICATION

The idea for facilitating access to data for the specific purpose of training AI models has not been tried, but it should be seen as a continuation of existing policies for cross-border flows present in multilateral organizations such as the Organization of Economic Cooperation and Development and the Indo-Pacific Economic Framework for Prosperity, and in trade agreements the US is party to, such as the United States-Mexico-Canada Agreement, the US-Japan Data Transfer Agreement, and an agreement between the US Department of Commerce and the Kenyan Ministry of Information, Communication, and the Digital Economy. What this new policy adds is specific language to protect the transfer and use of data from foreign countries to the United States for the express purpose of training AI models. This could include exceptions to local privacy and intellectual property rules that hinder private acquisition of key data flows. It could also include specific acquisition and licensing of key data sources by the government to accelerate the US AI industry. While it is likely that existing agreements that facilitate cross-border data flows have supported the construction of data sets for AI model training, adding explicit language would add an additional layer of protection for the domestic AI industry.

## FURTHER RESOURCES

- Frank Schweitzer and Ian Saccomanno, "The Rise of Artificial Intelligence, Big Data, and the Next Generation of International Rules Governing Cross-Border Data Flows and Digital Trade," White & Case, 2024
- Department of Commerce, Joint Statement on Harnessing Artificial Intelligence, Facilitating Data Flows and Empowering Digital Upskilling Between the United States Department of Commerce and the Kenyan Ministry of Information, Communication and the Digital Economy, 2024
- OECD, Fostering Cross-Border Data Flows with Trust, 2022
- GSMA, Cross-Border Data Flows: Realising Benefits and Removing Barriers, 2018

Tim Hwang is General Counsel and a Senior Fellow at the Foundation for American Innovation.

Joshua Levine is a Research Fellow at the Foundation for American Innovation.