

Commentary

Regulatory barriers to US-China collaboration for generative AI development in genomic research

Zhangyu Wang,¹ Benjamin Gregg,² and Li Du^{3,*}¹PhD student in International Law, Faculty of Law, University of Macau, Avenida da Universidade, Taipa, Macau, China²School of Law, The University of Texas at Austin, 727 E. Dean Keeton Street, Austin, TX 78705, USA³Faculty of Law, University of Macau, Avenida da Universidade, Taipa, Macau, China*Correspondence: stephendu@um.edu.mo<https://doi.org/10.1016/j.xgen.2024.100564>

Here, we examine the challenges posed by laws in the United States and China for generative-AI-assisted genomic research collaboration. We recommend renewing the Agreement on Cooperation in Science and Technology to promote responsible principles for sharing human genomic data and to enhance transparency in research.

Introduction

Compared with the previous generation of artificial intelligence (AI) models trained to accomplish specific tasks, generative AI is more versatile and can be used across various sectors. By leveraging various combinations of data modalities, generative AI models can produce original and informed solutions based on the input data.¹ The application of generative AI has shown potential in advancing scientific knowledge, including genomic research, which relies on the effective processing and analysis of large genomic and health datasets.² The development of generative-AI-assisted genomic research is still in its early stages. It confronts challenges concerning access to highly sensitive personal data, such as genomic data and patients' health records, as well as to more diverse datasets under different legal and governmental jurisdictions.^{1,3} In particular, data diversity and quality play decisive roles in reducing AI bias and improving prediction accuracy. Consequently, access to and sharing of genomic data are crucial for the two leading players in genomic research—China and the United States, each possessing some of the largest active genomic databases.

Since 1979, the Agreement on Cooperation in Science and Technology (S&T Agreement) has promoted scientific cooperation between the United States and China in areas such as health, agriculture, and the environment by encouraging transborder personnel movement and access to research materials.⁴ The S&T

Agreement, renewed every 5 years, expired in August 2023. Subsequently, the United States sought a 6-month extension with China to negotiate further regarding potential revisions to the agreement.⁵ A report published in *Nature* on August 25, 2023, highlighted some of the Chinese American scientific community's concerns about the rising tension between the two countries and its potential impact on the agreement's renewal. Renewal is still pending.⁶ The obstacles to the extension of the S&T Agreement include increasing political distrust between the American and Chinese governments, especially against the backdrop of the two nations' fierce "strategic competition for technology and innovation capabilities." Beyond political and economic considerations that hinder international technological exchanges and collaboration between the United States and China, this commentary provides a legal perspective on the issue of AI and international collaboration with a focus on generative-AI-assisted genomic research. Our analysis indicates that in addition to significant friction between two profoundly different political, economic, and military systems quite beyond the scope of this commentary, there remains some legal and regulatory challenges to cross-border genomic data sharing in both countries. These obstacles will need to be handled in ways that are acceptable to both governments to continue the cooperation first envisioned in the S&T Agreement more than a quarter century ago.

The S&T Agreement and its concessions

The S&T Agreement seeks to foster collaborative research between the two countries. It urges the development of workable measures to facilitate international research cooperation between governmental and/or non-governmental institutions, including reasonable personnel transborder movement and access to research information and materials.⁴ The 2013 Protocol to the S&T Agreement stresses bilateral cooperation for health science research and health administration, including exchanges of scientific and technological information and of biological materials.

The S&T Agreement does not establish mandatory arrangements for regulating collaborative projects on scientific research. Instead, it steers cooperative research projects and investigators toward applicable national laws and appropriate regulatory authorities.⁴ Further, the 2013 Protocol includes two explicit provisions prohibiting genomic data transfers "identified as requiring protection for national security reasons or foreign relations" and "subject to export control regulations." These national security and export regulation conditions are determined by national authorities in accordance with respective national laws.

Challenges posed by the US legal system

In November 2018, at the direction of then-President Trump, the US Department of Justice (DOJ) launched the China



Initiative with the stated aim of countering Chinese national security threats. Its goal was to identify and prosecute individuals involved in trade-secret theft, hacking, or economic espionage.⁷ Under President Biden, the DOJ ended the initiative in February 2022, largely due to “perceptions that it unfairly painted Chinese Americans and US residents of Chinese origin as disloyal,” and pivoted to a “reframing and recalibration—not an abandonment—of a muscular law enforcement response to the national security threat posed by the People’s Republic of China.”⁸ Against the backdrop of these and other significant tensions, this commentary seeks ways acceptable to both countries to improve the climate for continuing cross-border collaborative relationships in generative-AI-assisted genomic research.

The US’s National AI Initiative Act of 2020, which entered into force in 2021, aims to strengthen the American position as an international leader in AI development. It established the National Artificial Intelligence Initiative Office, responsible for implementing and overseeing national AI strategy (15 USC § 9411). Whereas the European Union has been proactive in proposing [regulations and guidelines for AI technologies](#), including generative AI, neither the United States nor China has enacted comprehensive national legislation specifically targeting AI. On the one hand, the [AI governance blueprint](#) released by the White House in October 2022 indicates that the Senate likely will pass AI laws on security, privacy, and algorithmic discrimination in the near future to tackle relevant challenges presented by generative AI. On the other hand, the United States has no federal law regarding data transfer and issues of data protection and informed consent. Instead, research institutions and enterprises are governed by state-level legislation (e.g., California’s Online Privacy Protection Act) or sector-specific regulations (e.g., HIPAA in the healthcare sector).⁹ According to a 2018 report, “US laws governing research and disclosure and use of data generated within the health care system do not impose different requirements on transfers to researchers and service providers based in third countries compared with US-based researchers or service providers.”¹⁰ Now, a [2024 Executive Order](#)

[announced by President Biden](#) directs the attorney general to enact regulations prohibiting or restricting US persons from engaging in transactions that would result in bulk access to US citizens’ sensitive personal data, such as biometric identifiers, human genomic data, and personal health data, by “countries of concern,” including China. It highlights potential national security risks, including misuse by countries of concern when accessing vast amounts of sensitive data from American citizens to develop AI capabilities.

The US export control regimes under the Export Administration Regulation (EAR) may impact the cross-border flow of genomic data and the communication of AI-related technologies. The export control regimes, while not directly constraining cross-border data transfer, applies to items subject to the EAR, including technology and software exported to other countries in transactions between the United States and foreign entities. The EAR defines the term “technology” broadly to encompass both tangible and intangible information crucial for the development, production, operation, and maintenance of an item (15 CFR 772.1). AI technology export controls thus implicate international genomic data-sharing and research collaborations.

The US Department of Commerce Bureau of Industry and Security (BIS) is responsible for regulating the export of items under the EAR. Two of its provisions to counter potential national security threats may affect US-China collaboration in AI development. First, the [Entity List](#) imposes specific license requirements on entities determined to be involved, or likely to become involved, in activities contrary to American national security or foreign policy interests (15 CFR § 744.16). The identified entity bears the onus of providing exculpatory evidence to facilitate the approval process for the proposed technology export (15 CFR 742.4). Second, the Unverified List controls export permission regarding the end-use and end-user of items subject to the EAR (15 CFR § 744.15) (see [Table 1](#)).

In recent years, the BIS has imposed export controls on many Chinese entities working in biological and medical sectors. For example, over the last three years, BIS has added five subsidiaries of BGI, China’s largest genetic testing service

provider, to the Entity List, such that they are now subject to licensing requirements for the export of all items under the EAR. The Entity List now applies to numerous Chinese universities and research institutions engaged in AI and genomic research, including Beihang University and Xi’an Jiaotong University.

Challenges posed by China’s legal regime

In August 2023, a new law on generative AI—Provisional Measures for the Administration of Generative Artificial Intelligence Services—entered into force (see [Table 1](#)). Formulated by several departments of the State Council of China, it is China’s first national administrative regulation of privacy and security issues related to the development and application of generative AI. It requires obtaining individual consent when collecting data and preventing discrimination in areas such as algorithmic programming and the selection of training data (Articles 4 and 7). It requires the respectful handling of intellectual property of databases, privacy, and data protection and encourages transparency in AI services (Articles 4 and 7). It endorses international dialogue concerning generative AI technology on equal and reciprocal terms and encourages proactive involvement in the development of international regulations (Article 6). It regulates the use of generative AI in creating texts, pictures, audios, videos, and other content for public consumption within China (Article 2). The Provisional Measures does not apply to entities that do not provide services to the general public, including public education and research institutions. Such entities must also adhere to other applicable laws on human genomic data, personal information protection, and healthcare data management.

The primary legislation governing human genetic resources in China is the Administrative Regulations on Human Genetic Resources (HGR Regulation) (see [Table 1](#)). It prohibits foreign organizations or individuals from acquiring genomic data from China without involving Chinese research participants (Article 7). Article 28 specifies two exceptions: (1) the acquisition of genomic data through international research cooperation between China and foreign countries

Table 1. Key legal requirements exerting influence on collaborative AI-based genomic research between the United States and China

China's regulation of generative AI, export control law, and data protection laws	US export control rules
<p>Regulation of generative AI</p> <ul style="list-style-type: none"> ➤ The scope of application of the Provisional Measures (Article 2, Provisional Measures for the Administration of Generative Artificial Intelligence Services) ➤ The respectful handling of intellectual property of datasets, privacy, data protection, the prevention of discrimination in algorithmic programming and training data selection, and the need for improving transparency of AI service provision (Articles 4 and 7, Provisional Measures for the Administration of Generative Artificial Intelligence Services) ➤ The encouragement for participating international dialogue and proactive involvement of international rule making for AI development (Article 6, Provisional Measures for the Administration of Generative Artificial Intelligence Services) <p>Laws related to human genomic data</p> <ul style="list-style-type: none"> ➤ Restriction of foreign entities' direct access and use of China's human genetic resources including genetic data (Article 7, The Administrative Regulations on Human Genetic Resources) ➤ Prohibition of purchase and sale of human genetic resources (Article 10, The Administrative Regulations on Human Genetic Resources) ➤ Administrative approval procedures of international research collaborations using China's genetic resources (Article 27, The Administrative Regulations on Human Genetic Resources) ➤ Procedures of filing for record and submitting backup for providing open access to human genetic resources to foreign entities (Article 28, The Administrative Regulations on Human Genetic Resources) ➤ Legal liability for providing genetic data to foreign entities without fulfilling the procedures of filing for records or obtaining administrative approval (Articles 41 and 42, The Administrative Regulations on Human Genetic Resources) <p>Legal regimes of personal information protection</p> <ul style="list-style-type: none"> ➤ Three approaches to conduct the cross-border data transfer, i.e., passing the security evaluation, gaining the certification of data protection capacity, and concluding the Standard Contract (Article 38, Personal Information Protection Law) ➤ The mandatory adoption of security evaluation in cross-border data transfer activities (Article 4, Security Evaluation Measures for Data Provision Abroad) <p>Rules of healthcare data management</p> <ul style="list-style-type: none"> ➤ The data localization requirements and its wide applicability to medical institutions at all levels within China (Articles 2, 3 and 10, Notice of National Health and Family Planning Commission on Promulgation of the Measures for the Administration of Population Health Information) 	<ul style="list-style-type: none"> ● Definitions of "technology" as used in the Export Administration Regulations (EAR) (15 CFR § 772.1 Definitions of terms as used in the EAR) ● Application of Unverified List (UVL) to non-citizens <ul style="list-style-type: none"> ➤ UVL statement required of entities listed in the UVL to export, reexport, or transfer (in country) items subject to the EAR (15 CFR § 744.15 [a] and [b] Restrictions on exports, reexports and transfers (in-country) to persons listed on the UVL) ➤ Criteria for revising the UVL, including adding entities to and removing entities from the UVL (15 CFR § 744.15 [c] Restrictions on exports, reexports and transfers (in country) to persons listed on the UVL) ● Application of the Entity List to non-citizens <ul style="list-style-type: none"> ➤ Criteria for including entities into the Entity List and related license requirements for them (15 CFR § 744.16 [a] Entity List)

but only after approval by the Ministry of Science and Technology (MOST) regarding potential harm to national security, public health, and what the Ministry calls “social public interest” as determined by the MOST, and (2) accessing genomic data when Chinese data owners, such as private companies and government agencies that legally acquire the data, have passed a security review (安全审查) by the MOST. The HGR Regulation also prohibits the purchase and sale of human genomic data (Article 10). Violation results in the confiscation of illegally obtained genomic data and severe administrative penalties (Articles 41 and 42).

The initiation of a security review by the MOST usually triggers a concurrent security evaluation (安全评估) by the Cyberspace Administration of China (CAC) under the Personal Information Protection Law (PIPL). Article 38 of the PIPL mandates that personal information, if internationally transferred, must undergo a compulsory evaluation when the volume of data exceeds limitations stipulated in relevant laws (see Table 1). The successful completion of one procedure does not ensure the success of another. For example, sharing genomic data internationally for the purpose of developing a commercial AI program may not pose an obstacle in the CAC security evaluation, yet the MOST might block transfer because the HGR Regulation prohibits “trading human genetic resources.” These two procedures run parallel to one another, imposing substantial legal compliance obligations on data processors. No official legal documents specify whether and how these procedures should be coordinated to expedite cross-border genomic data flow.

To protect China’s human genetic resources, national security, and public interest, China restricts the cross-border flow of genomic data and administratively reviews all data-transfer activities. The stringency of these regulations significantly hinders US-China cooperation on generative-AI-assisted genomic research under the S&T Agreement. Observers point to the lack of clarity in the regulation of sharing of academic and health data from China.¹¹ Prof. Shuhua Xu, an influential Chinese genetics scholar, expressed in an interview with the journal *Nature* that “he supports the regulation of human ge-

netic resources in principle, but thinks some of the requirements under the latest guidelines are too restrictive and will deter scientists from doing genetics work.”¹¹

Further, genomic data generated in the medical sectors can only be stored on servers within China for the purposes of safeguarding patients’ privacy and information security, which greatly restricts the international transmission of the data.¹² This requirement is derived from Article 10 of the Measures for the Administration of Population Health Information (PHI Measures), promulgated by the National Health and Medical Commission (NHMC) in 2014. It strictly prohibits storing PHI on a server located outside China. This requirement applies to nearly all Chinese healthcare and medical institutions.¹² According to the Reformation Plan of State Council Institutions 2023, the Chinese government recently transferred the supervision of China’s Biotechnology Development Center from the MOST to the NHMC—that is, to the Center currently responsible for managing human genetic resources. It is uncertain whether placing the Biotechnology Development Center under NHMC supervision will tighten regulations on the sharing of genomic data in the future, given that the in-country data storage requirement of the PHI Measure enacted by the NHMC has been in force for nearly a decade. Because the PHI Measure was promulgated by the NHMC, which proposed the data localization requirement, and because this requirement has continued for nearly a decade, the NHMC may impose greater restrictions on cross-border genomic data sharing.

Next steps

The extension of the agreement by the United States and China in August last year indicates some hesitation on the part of both countries to substantially abandon bilateral communication and joint scientific research. Since the extension has now expired, the countries once again stand at a crossroads. We have identified several of the numerous legal and regulatory challenges to cross-border genomic data sharing. While privacy-preserving federated learning and secure homomorphic encryption can facilitate cooperative AI model training without necessarily exporting genomic

data internationally, it and other technological solutions do not fully address the legal challenges we have identified, and any such solutions must, of course, comply with the legal frameworks related to data privacy and protection in both countries.

Given the various regulations on cross-border transfer of genomic data in the United States and China, both countries would do well to maintain their tradition of scientific dialogue and collaboration. Renewing the agreement offers an opportunity for such dialogue, toward mutual understanding on responsible ethical principles for the use of human genomic data. Legal mechanisms that enhance transparency in the use and sharing of human genomic data for research purposes and scientific and technological exchanges are in the interest of both countries. The establishment of the US-China [Clean Energy Research Center \(CERC\)](#) under the S&T Agreement constituted a promising approach toward alleviating at least some legal obstacles while contributing to increased transparency. The CERC, a bilateral agreement designed to run from 2011 to 2020, successfully engaged both governments, key policymakers, researchers, and industry in sustainable joint research and policymaking in the clean energy domain. Such mechanisms reassure the international community that both nations adhere to responsible research collaboration practices.

International collaboration brings together experts with different perspectives and skills while allowing access to genomic data from various populations around the world. Collaboration is crucial for understanding genetic variations and their implications for health and diseases, in part because many genetic diseases and health-related challenges affect all people. Collaboration facilitates the development of solutions with worldwide applicability, leading to research outcomes of a quality higher than otherwise possible. By validating and replicating research findings across different populations and environments, international collaboration enhances the credibility and robustness of genomic discoveries. It reduces redundancy in research efforts and promotes the sharing of such resources as computing power, data, and expertise,

making research more efficient. In these respects, among others, scientific cooperation between the United States and China is important for advancing AI-assisted genomic research. The S&T Agreement continues to provide an important framework for ensuring continuous commitment by both nations to foster collaborative scientific research.

Finally, we encourage both countries to invest in training programs for both regulators and industry professionals to ensure that they have the necessary knowledge and skills to navigate the evolving regulatory landscape. We encourage both countries to support initiatives for educating domestic researchers, legal experts, and policymakers on the benefits and best practices of international collaboration in generative-AI-assisted genomic research.

ACKNOWLEDGMENTS

This research is funded by the University of Macau under grant no. MYRG2020-00096-FLL.

AUTHOR CONTRIBUTIONS

Z.W., B.G., and L.D. designed the study. Z.W. drafted the original manuscript. B.G. and L.D. contributed extensively to editing and revising the

manuscript. All authors read and approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Moor, M., Banerjee, O., Abad, Z.S.H., Krumholz, H.M., Leskovec, J., Topol, E.J., and Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. *Nature* 616, 259–265. <https://doi.org/10.1038/s41586-023-05881-4>.
- Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., and Telenti, A. (2019). A primer on deep learning in genomics. *Nat. Genet.* 51, 12–18. <https://doi.org/10.1038/s41588-018-0295-5>.
- Acosta, J.N., Falcone, G.J., Rajpurkar, P., and Topol, E.J. (2022). Multimodal biomedical AI. *Nat. Med.* 28, 1773–1784. <https://doi.org/10.1038/s41591-022-01981-2>.
- Government of the United States of America; Government of the People's Republic of China (1979). Agreement between the Government of the United States of America and the Government of the People's Republic of China on Cooperation in Science and Technology. <https://www.energy.gov/sites/default/files/pniec/098b7ef980003cff.pdf>.
- Martina, M. (2023). US seeks extension to controversial science agreement with China. *Reuters*. <https://www.reuters.com/world/us-says-it-seeks-six-month-extension-science-agreement-with-china-2023-08-23/>.
- Gilbert, N., and Mallapaty, S. (2024). US and China likely to delay renewal of key science pact again. *Nature*. Published online February 8. <https://doi.org/10.1038/d41586-024-00380-6>.
- Gilbert, N. (2023). China Initiative's shadow looms large for US scientists. *Nature* 615, 198–199. <https://doi.org/10.1038/d41586-023-00543-x>.
- Gerstein, J. (2022). DOJ shuts down China-focused anti-espionage program. *POLITICO*. <https://www.politico.com/news/2022/02/23/doj-shuts-down-china-focused-anti-espionage-program-00011065>.
- Boyne, S.M. (2018). Data protection in the United States. *Am. J. Comp. Law* 66, 299–343. <https://doi.org/10.1093/ajcl/avy016>.
- Majumder, M.A. (2018). United States: Law and policy concerning transfer of genomic data to third countries. *Hum. Genet.* 137, 647–655. <https://doi.org/10.1007/s00439-018-1917-9>.
- Mallapaty, S. (2022). China expands control over genetic data used in scientific research. *Nature* 605, 405. <https://doi.org/10.1038/d41586-022-01230-z>.
- Chen, Y., and Song, L. (2018). China: Concurring regulation of cross-border genomic data sharing for statist control and individual protection. *Hum. Genet.* 137, 605–615. <https://doi.org/10.1007/s00439-018-1903-2>.