Chair Coons, Ranking Member Sasse, and Members of the Subcommittee, thank you for the opportunity to testify today.

I am a Senior Fellow at Yale Law School's Paul Tsai China Center and a Cybersecurity Policy Fellow at New America. I have worked as an analyst of Chinese data and technology policies for the last decade, both in the U.S. national security community, and in the private sector. I also advise American corporate clients on China's technology regulations.

Today I will focus my testimony on risks posed by Beijing's access to Americans' data and offer recommendations to address those risks.

I. Economic and security risks associated with Beijing's access to Americans' data

U.S. policymakers face a new challenge in addressing data security risks in the U.S-China relationship. Beijing's history of state and quasi-state cyber hacking and illegal data exfiltration have been well documented. Policymakers now must grapple with how to address risks posed by Beijing's aggregation of illicitly acquired data with that gleaned from openly available commercial channels.

Beijing is already presumed to have sensitive national security information from the theft of personnel records of roughly 21 million individuals from the U.S. Office of Personnel Management; travel information from a cyber attack on Marriott hotels covering roughly 400 million records; and credit data from Equifax on roughly 145 million people. If the Chinese government were to combine those datasets with the legal purchase or scraping of publicly available large commercial data, or aggregate different individual datasets, it could threaten U.S. national security and economic competitiveness in a number of different ways:

- If additional sources of personal data such as location, social media, or pattern of life data were to be acquired legally from companies or bought openly through unregulated data brokers and combined with what Beijing has already acquired through cyber theft, Chinese security services could use it **to gain leverage over and to target Americans in sensitive government national security positions or military personnel for manipulation, blackmail, or other forms of coercion.** This is particularly concerning from a counterintelligence perspective for individuals with security clearances or those with access to critical infrastructure.

- Beijing's analysis of such lawful and unlawfully acquired datasets could also be used to **identify and monitor Chinese dissidents abroad**. Access to larger more, diverse datasets could also contribute to strengthening surveillance systems within China to **enable human rights abuses and political control of Uighurs and other minority populations**.

- Beijing could also use aggregated data sources in ways **that enable bulk electronic surveillance.** As Chinese online services and network infrastructure gain in prominence around the world, it is possible that the Chinese government could filter or monitor data processes abroad, just as the United States had done, as shown by Snowden, in utilizing data transmissions across U.S. networks for intelligence gathering. As we move toward a world in which people have online profiles built on aggregated data, we must ask: what are the implications of the CCP gaining effective control of information flows beyond China's closed internet system? What are the implications as the CCP takes even more drastic steps to close off the loopholes that to this day keep even the Great Firewall relatively porous and circumventable (e.g., stricter enforcement of restrictions on virtual private networks (VPNs) or shifting from a blacklist to a whitelist approach to permissible websites so technical controls can keep pace with online content deemed threatening)?

- Access to large datasets collected abroad **provides Chinese companies insight into population-level and individual consumer behavior, risk-tolerance, and other preferences—with distinct economic and security implications.** From an economic perspective, this helps to strengthen the economic competitiveness of Chinese firms by enabling them to develop AI applications that better serve diverse demographics in markets around the world. In terms of national security, if Beijing were to attempt to push out misinformation to audiences abroad, insights about individual preferences or psychology decision-making trends could help make that information more convincing and realistic. These data-driven insights could enable China to launch more effective cyber-attacks by helping create more appealing and realistic content to entice individuals to open malicious email attachments or link, for example.

II. Recommendations

Why Data Matters

To be effective, U.S. policy should be based on an accurate understanding of why data matters—both its potential as a source of value and vulnerability. The analogy of data as the new oil is false and leads to ineffective policies. Unlike oil, data is not a finite or zero-sum resource that is only valuable in large volumes. Matt Sheehan writes that five dimensions are crucial for machine learning data today: quantity, depth, quality, diversity, and access.[1] Inaccurate understandings of data can lead to unintended consequences that weaken national security and economic power, rather than increase it.

[1] Matt Sheehan, "Much Ado about Data," *MacroPolo,* https://macropolo.org/ai-data-us-china/, July 16, 2019.

Data poses unique challenges because it can be copied, stored, and analyzed later. Evaluating China's potential access to sensitive U.S. data should consider not just how data could be used today, but also the implications for the future economic and security interests of the United States. At the same time, many datasets lose value after a certain period depending on a range of factors including its type and how it is used in information systems.

Pathways Forward

U.S. lawmakers have an opportunity to address transnational security threats while also advancing a more secure, ethical, and democratic global internet. Although part of this challenge requires tools that are specific to risks posed by China, much of this challenge is bigger than China. Setting basic standards on what data can be collected and retained by all companies—regardless of country of origin or nationality—will help protect U.S. data, whether the risk comes from a state-sponsored hacker, a data broker, or a private company trading in consumer data without transparency or controls.

I offer a brief assessment of several different tools for U.S. lawmakers to consider in tandem:

(1) Federal Privacy Law

Setting basic standards on what data can be collected and retained by all companies will help protect U.S. personal data, regardless of where the risk originates. Developing a comprehensive federal privacy law is vital to this effort, along with the creation of strong enforcement mechanisms. Inaction by the United States means ceding leadership and influence in setting international standards to both Europe and China.

Without higher standards for data security and privacy, U.S. citizen data held by unregulated private companies are more vulnerable to breaches by hackers from China or from being sold to third-parties openly buying, aggregating, and selling consumer data. Equifax's many security issues are well-documented, such as the company's failure to patch known vulnerabilities that ultimately left exposed the data of 145 million Americans. But the hack was also conducted by a foreign government entity with sophisticated hacking capabilities and access to considerable state resources. Companies should not have access to such a volume of personal or sensitive data that it creates a target to be hacked or transmitted to China at all.

As U.S. policymakers weigh a federal privacy, there are a variety of models and best practices from around the world for further study, particularly regarding international transfers to third countries.One possible approach deserving further attention is the United Kingdom's Data Protection and Digital Information Bill. While the UK plans to maintain controls over data flows to countries that do not provide equivalent levels of protection to the UK's own privacy laws, the government is retaining the ability to use other determinations—such as political or economic considerations—to allow for data transfers. What the UK has done with this reform is to offer a bridge to countries with emerging privacy laws to partner with and invest in the UK. The EU is bound to a much stricter standard for data flows that creates practical challenges for the US to implement. However, this UK approach could offer a way for the United States to maintain its geopolitical partnerships, ensuring data flows to trusted countries, and not walling off the U.S. digital economy from the world. Efforts to address data security for Americans are less likely to be effective without passage of a federal privacy law that includes provisions governing where and how data can be transferred.

Limits on international transfers and access to data should take a risk-based approach evaluating a range of factors. The Biden administration's Executive Order on Protecting Sensitive Americans' Data from Foreign Adversaries[2] established a framework for evaluating risks that included the idea that not all data has the same level of sensitivity. The mere fact that a foreign company handles U.S. citizen data or transfers it to China may or may not warrant a restriction. U.S. national security risks should be evaluated to determine (a) what kind of U.S. citizen data is being accessed (e.g., metadata, images, geographic data, critical infrastructure data, etc.); (b) how that data is being used and what data protection measures are in place to protect the rights and interests of U.S. consumers; and (c) with whom that data is being shared and through what mechanisms. If based on the outcomes of such an evaluation, we cannot verify that the interests and rights of U.S. consumers will be protected, then certain transfer limits or restrictions should be put in place.

(2) Data Sharing Among Allies

United States should work with like-minded governments to develop a common set of standards that would allow data to flow—building off of the concept of "data free flows with trust" put forward by Japan in the Osaka Track of the G-20.[3] A multilateral approach should be based on creating a system of incentives for compliance. The objective would be to establish an interoperable data framework encouraging other countries in the world to set similar but not identical standards. The United States could lead the way in setting up a certification system that would extend benefits to countries whose data regimes and companies meet certain clear criteria for data protection. The Organization for Economic Cooperation and Development (OECD) privacy guidelines, for example, could serve as a reference in creating a baseline for commercial data flows.[4] The OECD initiative on creating principles for government access to data is also

[2] https://www.whitehouse.gov/briefing-room/presidential-actions/2021/06/09/executive-order-on-protecting-americans-sensitive-data-from-foreign-adversaries/.

[3] Data Free Flow with Trust (DFFT): Paths towards Free and Trusted Data Flows, World Economic Forum, https://www.weforum.org/whitepapers/data-free-flow-with-trust-dfft-paths-towards-free-and-trusted-data-flows/.

[4] "The OECD Privacy Framework," Organisation for Economic Co-operation and Development, https://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm.

important step in this regard deserving more attention and study.[5] I'd also like to note the potential for the Global Cross Border Privacy Rules Forum, a data transfer alliance that requires companies to certify to common standards for privacy protection and enables cross-border transfers for those certified companies. This U.S.-backed system offers the potential to address many facets of protecting American data, as it not only takes into consideration the country where data is traveling, but also requires that companies certify to a government-backed standard to be able to transfer data.

Such an approach creates a coalition of allies sharing data with the United States. The ability of U.S. firms to maintain a high rate of innovation depends upon access to global markets, talent, and international datasets. If U.S. firms cannot send data out of countries in which they operate overseas, this directly impacts economic growth and AI innovation that are core to building applications that work across a variety of different geographies, languages, cultures, and demographics. As the technology competition between the United States and China increasingly plays out less in each other's countries than in third countries around the world, data will fuel U.S. technological leadership that is vital to playing offense.

(3) Technical Measures

Technical measures to safeguard data should complement policy. If effectively encrypted, data can be transferred to and shared with even countries deemed adversaries since data can only be read as cipher text. One approach could focus on creating incentives for companies to better encrypt data stored at rest and in transit. If the data is well encrypted and the recipient does not have the key, then limits to third parties would not apply, for example. Congress and the Administration should work with American companies to develop privacy preserving technologies— which allow for computing on data without seeing it—and implement them in an open source way available to everyone. For example, homomorphic encryption protects data in transit and during computation without needing to decrypt it. American companies are right now leading the global conversation on privacy preserving technologies and investment in these technologies by the government will enable companies of any size to implement these technological innovations.

Conclusions

We need to address national security risks where they exist as part of a broader U.S. initiative for comprehensive data privacy and higher cybersecurity standards for all companies —whether domestic or foreign. Failure to offer an affirmative vision for U.S. data governance will make the United States less secure, less prosperous, and less powerful, and allow more space around the world for companies controlled by the CCP to gain ground across the world.

[5] https://www.lawfareblog.com/towards-oecd-principles-government-access-data.