Atlantic Council
CYBER STATECRAFT INITIATIVE
ai x ac
September 2025

Issue brief Securing data in the AI supply chain

Written by Justin Sherman

AI technologies involve a complex supply chain. Policymakers and technologists should better secure the AI supply chain's data components by taking a comprehensive view of the data in the AI supply chain and identifying which components can be more robustly secured and which risks require new mitigations.

Bottom lines up front

- Developers, users, maintainers, governors, and securers of AI technologies should map the data components of the AI supply chain to existing cybersecurity best practices—and use that mapping to identify where existing best practices fall short for AI-specific risks to the data components of the AI supply chain.
- Developers, users, maintainers, governors, and securers of AI technologies should "Know Your Supplier," using the supply chain-focused approach to mitigate both AI-specific and non-AI-specific risks to the data components of the AI supply chain.
- Policymakers should widen their lens on AI data to encompass all data components of the AI supply chain. This includes assessing whether sufficient attention is given to the diversity of data use cases that need protection (e.g., not just training data for chatbots but for transportation safety or drug discovery) and whether they have mapped existing security best practices to non-AI-specific and AI-specific risks.

Executive Summary

Underpinning AI technologies is a complex supply chain—organizations, people, activities, information, and resources that enable AI research, development, deployment, and more. The AI supply chain includes human talent, compute, and institutional and individual stakeholders. This report focuses on another element of the AI supply chain: data.

While a diversity of data types, structures, sources, and use cases exist in the AI supply chain, policymakers can easily fall into the trap of focusing on one AI data component at one moment (e.g., training data circa 2017), then switching focus to another AI data component next (e.g., model weights in current times), risking a lopsided policy that fails to take account of all the AI data components that are important for AI research and development (R&D). For example, overconfidence about which data element or attribute will most drive AI R&D can lead researchers and policymakers to skip past important, open questions (e.g., what factors might matter, in what combinations, and to what end), wrongly treating them as resolved. Put simply, a "one-size-fits-all" approach to AI-related data runs the risk of creating a regulatory, technological, or governance framework that overfocuses on one element of the data in the

AI supply chain while leaving other critical parts and questions unaddressed.

Managing the risks to the data components of the AI supply chain—from errors to data leakage to intentional model exploitation and theft—will require a set of different, tailored approaches aimed at achieving a comprehensive reduction in risk. As conceptualized in this report, the data in the AI supply chain includes the data describing an AI model's properties and behavior, as well as the data associated with building and using a model. It also includes AI models themselves and the different digital systems that facilitate the movement of data into and out of models. The report, therefore, spells out a framework to visualize the seven data components in the AI supply chain: training data, testing data, models (themselves), model architectures, model weights, Application Programming Interfaces (APIs), and Software Development Kits (SDKs).

It then uses the framework to map data components of the AI supply chain to three different ways that policymakers, technologists, and other stakeholders can potentially think about data risk: data at rest vs. in motion vs. in processing (focus on a data component within the supply chain and its *current state*); threat actor risk (focus on *threat actors* and risks to a data component within the supply chain); and supply chain due diligence and risk management (focus on a data component *supplier or source* within the supply chain and related actors).

In doing so, it finds that many risks to AI-related data are risks to data writ large that existing best practices could mitigate. These include National Institute of Standards and Technology (NIST) and International Organization for Standardization (ISO) specified data access controls, continuous monitoring systems, and robust encryption; the risks at hand in these cases do not require reinventing the wheel. Simultaneously, this report also finds that some security risks to AI data components do not map well to existing security best practices that would adequately mitigate the risk or even apply at all. At least two stand out immediately: bad actors' attempts to poison AI training data require data filtering mechanisms not well captured by existing measures, and which access controls or encryption would not appropriately mitigate; and emerging, malicious efforts to insert so-called neural backdoors into the behavior of neural networks require new security protections, too, beyond the realm of traditional IT data security. On top of implementing these two categories of mitigations, this report emphasizes that organizations can leverage "know your supplier" best practices to ensure all other entities in their AI supply chains have security best practices for both non-AI-specific and AI-specific data risks.

This report concludes with three recommendations.

1. Developers, users, maintainers, governors, and securers of AI technologies should map the data components of the AI supply chain to existing cybersecurity best practices—and use that mapping to identify where existing best practices fall short for AI-specific risks to the data components of the AI supply chain.
2. Developers, users, maintainers, governors, and securers of AI technologies should "Know Your Supplier," using the supply chain-focused approach to mitigate both AI-specific and non-AI-specific risks to the data components of the AI supply chain.
3. Policymakers should widen their lens on AI data to encompass all data components of the AI supply chain. This includes assessing whether sufficient attention is given to the diversity of data use cases that need protection (e.g., not just training data for chatbots but for transportation safety or drug discovery) and whether they have mapped existing security best practices to non-AI-specific and AI-specific risks.

Introduction

Recent advances in computing power have catalyzed an explosion of artificial intelligence (AI) and machine learning (ML) research and development (R&D). While many of the mathematical and statistical techniques behind contemporary AI and ML models have been around for decades,[1] these advancements in computing power have combined with larger datasets, energy sources, human labor, and other factors to bring AI and ML R&D to unforeseen heights.

This phrase, "artificial intelligence," is best understood not as a single, specific technology but as an umbrella term for a range of technologies and applications. Illustrating this point, companies, governments, academic institutions, civil society organizations, and individuals, among others, are designing, building, testing, and using AI and ML applications ranging from facial recognition systems in shopping malls, driving navigation systems in autonomous vehicles, and chatbots in academic research environments to highly tailored applications in drug discovery, climate change modeling, and military opera-

1. This is not true in every single case. See, for example, Ashish Vaswani et al., "Attention Is All You Need," arXiv, June 12, 2017 [last revision, August 2, 2023], https://doi.org/10.48550/arXiv.1706.03762.

tions.[2] Despite wide variations in design and function, all these software applications, as such, characterize "AI." Their variations capture the expansiveness of the "AI" term. They also underscore that research and policymaking on AI's impacts—to labor, the environment, workforce productivity, economic growth, privacy, civil rights, national security, and so forth—must reference and differentiate between specific application areas, because they may greatly vary.

Underpinning AI technologies is a complex supply chain—organizations, people, activities, information, and resources enabling research, development, deployment, and more.[3] The AI supply chain includes human talent: the people around the world contributing to university and nonprofit research, building and iterating on commercial products, hacking systems to boost their security, applying deployed AI technologies in innovative ways, and so forth. It includes compute: the dynamic provisioning, protection, and management of hardware and software systems across shared infrastructure, in this case to power AI training, refinement, and so on—the subject of a forthcoming companion report from the Cyber Statecraft Initiative.[4] It includes institutional and individual stakeholders, such as infrastructure providers, data providers, technology and service intermediaries, user-facing entities, and consumers.[5] And the AI supply chain includes data components, which are the focus of this report.

AI technologies are data-rich. That is, they both rely tremendously on data to function and produce large volumes of data as part of their operation. As explored in this report, this data richness entails a complex set of data elements in the AI supply chain that feed into, come out of, and underpin the research, development, deployment, use, maintenance, governance, and security of AI technologies. Corporate developers, researchers, and others building an AI application from the ground up may create an algorithm and run it on different kinds of "training data" before measuring its performance with "testing data." For instance, in training an image recognition model to identify whether a photo contains a cat, the training data may be full of pictures of cats, dogs, airplanes, coffee machines, and cats sitting on coffee machines (i.e., "yes," "no," and more complex "yes" options), and the testing data might consist of similar pictures the model has never trained on, to test how well the function it learned generalizes to the new data. Individuals using AI chatbots or AI facial recognition models, to give another example, may upload data (e.g., questions, face images) into the system as part of using it, after which the system may provide data back to the individual (e.g., answers, names associated with faces) as well as output some metadata into a system log (e.g., performance metrics). These data components are just some of those present in the AI supply chain.

Mapping and understanding this data in the AI supply chain matters greatly for companies, policymakers, and society to protect each data element against exploitation. Leaks, theft, exploitation, and adverse use of AI-related data could harm specific individuals or groups of people (e.g., extracting data from AI models to violate privacy); undermine specific national objectives like economic competitiveness (e.g., data theft to replicate proprietary applications) or national security (e.g., data theft to understand a model's behavior, and thereby attack it); and create other issues ranging from market consolidation (e.g., single points of failure in the entity supplying key AI-related data) to undermining trust in critical technology areas (e.g., between patients and healthcare institutions). The US National Security Agency (NSA) recently wrote, "as organizations continue to increase their reliance on AI-driven outcomes, ensuring data security becomes increasingly crucial for maintaining accuracy, reliability, and integrity."[6]

Conversely, each data element enables different aspects of AI research, development, deployment, use, maintenance, governance, and security—meaning developers, users, maintainers, governors, and securers of AI technologies should want to better safeguard them for positively framed reasons, too. Better protecting the data underpinning an expensive commercial AI advancement could enable the company to move

2. Such uses are not inherently positive for the rigor (e.g., result accuracy, reproducibility, etc.) of academic research. See: Miryam Naddaf, "AI Linked to Explosion of Low-Quality Biomedical Research Papers," *Nature* 641, no. 8065, (May 21, 2025): 1080–81, https://doi.org/10.1038/d41586-025-01592-0.
3. See: NIST's definition of "supply chain" (source: CNSSI 4009-2015). "Glossary: Supply Chain," US National Institute of Standards and Technology: Computer Security Resource Center, accessed August 26, 2025, https://csrc.nist.gov/glossary/term/supply_chain.
4. Thanks to Sara Ann Brackett for discussion of her forthcoming paper.
5. Aspen K. Hopkins et al., "Recourse, Repair, Reparation, and Prevention: A Stakeholder Analysis of AI Supply Chains," arXiv, July 3, 2025 [submit date], https://doi.org/10.48550/arXiv.2507.02648.
6. "NSA's AISC Releases Joint Guidance on the Risks and Best Practices in AI Data Security," US National Security Agency: Central Security Service, press release, May 22, 2025, https://www.nsa.gov/Press-Room/Press-Releases-Statements/Press-Release-View/Article/4192332/nsas-aisc-releases-joint-guidance-on-the-risks-and-best-practices-in-ai-data-se/.

faster without slowdowns due to leaks, breaches, and trade secret theft. Shielding training data related to individuals from inadvertent leaks and exposure could bolster public trust in responsibly executed AI deployments in healthcare or transportation. The list of benefits to mitigating leaks, theft, exploitation, and adverse use of AI-related data goes on for comparing specific data types and uses against relevant risk mitigations—optimizing the use of existing security best practices and identifying AI-specific gaps to fill.

Without an effective framing for how to think about all the data in the AI supply chain, policymakers and others looking at AI and data security may overfocus on a single data component in the AI supply chain without accounting for all the others in the picture. They may also conflate related but distinct data components in the AI supply chain together, failing to account for differences in data type, structure, source, and use case that may create distinct risks and require different, tailored mitigations in response.

Moreover, treating all AI-related data as part of a new, flashy set of AI technologies can perpetuate a sort of AI exceptionalism. This view suggests that AI technologies exist in isolation from cloud, telecommunications, and other systems—treating them as separate from, rather than interconnected with, other technologies that also matter for innovation, security, governance, and more. It can implicitly suggest the data is all new, raising fundamentally new questions and issues without good answers, instead of relating to data security discussions and best practices that have been around for decades, as well as a subset of data risks that demand AI-specific mitigations. None of these outcomes lend themselves to the most rigorous public policy, industry, research, and public discussions about AI R&D, security, and geopolitics.

At a high level, the concept of data in the AI supply chain, therefore, enables analysts to map out points of concentration, resilience, and security vulnerability in AI systems and the overall AI ecosystem that might vary based on AI data type, structure, source, and use case. (For example, does cybersecurity-focused training data come from too few companies? How does the security of open-source health testing data compare to the security of health model parameters?) The concept of data components in the AI supply chain can help policymakers, developers, and those impacted by AI technologies understand the broader supply chain of parts underpinning a commercially successful, data-secure (or -insecure) AI system. And it can help governments, companies, civil society groups, journalists, and individuals to more precisely, systemically evaluate AI risk mitigation methods against the security risks of the coming years.[7]

A mapping and understanding of the data in the AI supply chain can also inform better policy. To be sure, no regulations of technology (or anything, for that matter) will treat the technology (or other thing) in question perfectly symmetrically across every country or jurisdiction in the world. But "AI regulations" based on highly inconsistent formulations of "AI-related data" can unintentionally increase friction. If a legislature writes rules for "AI data" when picturing only one type of training data, and another country's legislature takes a more comprehensive view of all the data types and use cases in the AI supply chain, the widely varied approaches could make it harder to harmonize cross-border steps to curtail bad practices. The varied approaches could create cross-jurisdictional barriers to startup innovation that regulators never intended. And they could further confuse global discourse on governing "AI data." These are potentially unintended effects that a better policy formulation on how to think about risks to and protections for AI-related data would avoid.

This report lays out a conception of the data components of the AI supply chain, which the research then maps to existing data security and supply chain security best practices—highlighting existing measures that work well and identifying, in the process, security gaps for issues more unique to AI data types and use cases. The framework once again focuses just on data, rather than all elements of the AI supply chain (e.g., compute). For simplicity's sake, it also excludes AI agents due to the complexity their permission-based, semi-autonomous functions introduce—instead, focusing on the wide range of non-agent models in use today.

First, this report discusses how policymakers can run the risk of overfocusing on one data component at the expense of the entirety of data types in the AI supply chain, which can contribute at best to lopsided policy and at worst to tendencies that undermine US AI competitiveness and leave critical parts of the data in the AI supply chain inadequately secured. Second, it introduces a concept of data in the AI supply chain with seven components, each defined and exemplified below: training data, testing data, models (themselves), model architectures, model weights, Application Programming Interfaces (APIs), and Software Development Kits (SDKs). It additionally discusses the interactions between data components, their

7. A single company, for instance, might find that maintaining better documentation for AI applications and data allows it to not just address vulnerabilities in those systems, but also the accidental failures, human errors, and bureaucratic issues (like the purchasing of new systems), too.

varied suppliers, and those suppliers' sometimes shifting or multiple roles vis-à-vis the data in the AI supply chain.

Finally, the report offers three different approaches to map data components in the AI supply chain to existing data security and supply chain security frameworks: data at rest vs. in motion vs. in processing (focus on a data component within the supply chain and its *current state*); threat actor risk (focus on *threat actors* and risks to a data component within the supply chain); and supply chain due diligence and risk management (focus on a data component *supplier or source* within the supply chain and related actors). These approaches can map concerns about training data theft, training data poisoning, API insecurity, and other data-related AI supply chain issues to established security controls and best practices from government agencies, standards bodies, cybersecurity literature, and areas like the financial sector and export control compliance. In doing so, it also begins to identify a few areas where existing security best practices may be insufficient for AI data risks—namely, confronting risks associated with the poisoning of data components in the AI supply chain and inserting neural "backdoors" into models through tampered training data or manipulation of model architectures. These risks, perhaps unique or relatively unique to AI models, require their own mitigations.

The report concludes by making three recommendations:

1. **Developers, users, maintainers, governors, and securers of AI technologies should map the data components of the AI supply chain to existing cybersecurity best practices—and use that mapping to identify where existing best practices fall short for AI-specific risks to the data components of the AI supply chain.** In the former case, they should use the framework of data at rest vs. in motion vs. in processing and the framework of analyzing threat actor capabilities to pair encryption, access controls, offline storage, and other measures (e.g., NIST SP 800-53, ISO/IEC 27001:2022) against specific data components in the AI supply chain depending on each data component's current state, the threat actor(s) pursuing it, and the traditional IT security controls the organization already has in place. In the latter case, developers, users, maintainers, governors, and securers of AI technologies should recognize how existing best practices will inadequately prevent the poisoning of AI training data and the insertion of behavioral backdoors into neural networks by manipulating a training dataset or a model architecture. They should instead look to emerging research on how to best evaluate training data to filter out poisoned data examples and how to robustly test network behavior and architectures to mitigate the risk of a bad actor inserting a neural backdoor, which they can activate after model deployment. And in both cases—of non-AI-specific and AI-specific risks to data—organizations can and should use the third listed approach of focusing on the data and supply chain itself to ensure their vendors, customers, and other partners are implementing the right controls to protect against risks of model weight theft, training data manipulation, neural network backdooring through model architecture manipulation, and everything in between, drawing on the two categories of mitigations they implement themselves.
2. **Developers, users, maintainers, governors, and securers of AI technologies should "Know Your Supplier," using the supply chain-focused approach to mitigate both AI-specific and non-AI-specific risks to the data components of the AI supply chain.** Those sourcing data for AI systems—whether training data, APIs, SDKs, or any of the other data supply chain components—should implement best practices and due diligence measures to ensure they understand the entities sourcing or behind the sources of different components. For example, if a university website has a public repository of testing datasets for image recognition, language translation, or autonomous vehicle sensing, did the university internally develop those testing datasets, or is it hosting those testing datasets on behalf of third parties? Can third parties upload whatever data they want to the public university website? What are the downstream controls on which entities can add data to the university repository—data which companies and other universities then download and use as part of their AI supply chains? Much like a company should want to understand the origins of a piece of software before installing it on the network (e.g., is it open-source, provided by a company, if so which company in which country, etc.), an organization accessing testing data to measure an AI model or using any other data component of the AI supply chain should understand the underlying source within the supply chain. Best practices in know-your-customer due diligence, such as in the financial sector and export control space, and in the supply chain risk management space, such as from cybersecurity and insurance companies, can provide AI-dependent organizations with checklists and other tools to make this happen. Avoiding entities potentially subject to adversarial foreign nation-state influence, data suppliers not sufficiently vetting the data they upload, and so forth will help developers, users, maintainers, governors, and securers of AI technologies to bring established security controls to the data in the AI supply chain itself. In the case of both non-AI-specific and AI-specific risks to data, organizations can and should use this supply chain due diligence approach to ensure their vendors, customers,

and other partners are implementing the right controls to protect against risks of model weight theft, training data manipulation, neural network backdooring through model architecture manipulation, and everything in between—drawing on the two categories of mitigations implemented as part of the first recommendation.

3. **Policymakers should widen their lens on AI data to encompass all data components of the AI supply chain. This includes assessing whether sufficient attention is given to the diversity of data use cases that need protection (e.g., not just training data for chatbots but for transportation safety or drug discovery) and whether they have mapped existing security best practices to non-AI-specific and AI-specific risks.** As multiple successive US administrations explore how they want to approach the R&D and governance of AI technologies, data continues to be a persistent focus of discussion. It comes up in everything from copyright litigation to national security strategy debates. The United States' previous policy focus on training data quantity, and little else, has already prompted policymakers to avoid discussing comprehensive data privacy and security measures, which now—in light of Chinese AI advancements and concern about AI model weight dissemination—are suddenly more relevant. To avoid these cycles in the future, where policy overfocuses on one AI data element when in fact many are relevant simultaneously, policymakers should take a comprehensive view of the data components of the AI supply chain. The framework offered in this paper, spanning seven data components, is one potential guide—though again, policymakers need not stick to necessarily one framework. What is most critical to avoid is developing data security policies that protect some data components of the AI supply chain (e.g., training data) while leaving others highly exposed (e.g., APIs). An expanded view of the different data components, the components' interaction, and the often multiple and shifting roles of suppliers should help inform better federal legislation, regulation, policy, and strategy—as well as engagements with other countries and US states. Right now, organizations such as the Congressional commerce committees, the Commerce Department (including because it implements export controls and the Information and Communications Services and Technologies supply chain program), the Defense Department (with all its current AI procurement), and the Federal Trade Commission (with responsibility for enforcing against unfair and deceptive business practices) should stress-test their assumptions about how to best protect AI data, and whether existing best practices achieve desired security outcomes, against this data component framework. This requires asking at least two questions. Do their existing security, governance, or regulatory approaches—e.g., in the security requirements used in Defense Department AI procurement, in how the Federal Trade Commission thinks about enforcing best practices for AI data security—apply well to a diversity of data use cases that need protection, such as with testing datasets for self-driving vehicle safety or training datasets for cutting-edge drug discovery? List out the use cases beyond chatbots that are not top of mind but are highly relevant from a security perspective, from defense to shipping and logistics to healthcare. And second, are they parsing out which risks they have concerns about, vis-à-vis AI-related data, that are specific to AI versus risks to data in general? For both categories, consider how the framework and some of the security mitigations cited in this report—for example, the NIST guidance, ISO practices, and new research on detecting neural backdoors, etc.—can serve as best practices to improve outcomes.

Moving Balls, Swinging Pendulums

Policymakers, researchers, and private-sector firms alike are in constant debate about what kinds of data, data analysis, and data characteristics (such as quantity or diversity) will lead to major breakthroughs in AI research and development. These debates span geopolitical and national security issues—like fights over whether a country's population and data collection reach may lend a strategic military advantage—and economic and social ones—like conversations about how best to maximize AI for medicinal innovations or minimize AI risks to worker privacy. Debates about AI and data implicate pressing and often broader issues such as tech innovation, responsible technology governance, cybersecurity, antitrust, and nation-state competition, too.

But the past few years alone have illustrated how simplistic this AI and data debate can become—and how quickly, and perhaps arbitrarily, the metaphorical ball can move. About seven or eight years ago, it became somewhat of a prevailing view in Washington, DC, that "data is the new oil" and that the volume of data to which a country had access would determine its

AI might.[8] Compelling perhaps because of its simplicity (data quantity is the key) and certainty (about the link between data quantity and AI leadership—and AI leadership and superpower status), the narrative quickly took hold, pushed by senior government officials and large Silicon Valley corporations alike.[9] Policy, industry, and media discourse focused highly on one element of the links between data, broadly defined, and AI R&D: training data quantity.

Now, though, the ball has moved. Policymakers talk far less about training data (even though it is still important) and much more about model weights—numerical parameters that specify how the model represents connections to leverage between pieces of data to achieve the desired output. (They also, rightfully, spent much time discussing compute, but that is once again outside the scope of this report). Discussions about new export controls and the Biden administration's last-minute,[10] multi-tiered framework for (on paper) limiting "AI diffusion" are chief among the recent policy efforts focused on this slice of AI R&D,[11] as are some of the Trump administration's efforts to deregulate AI with the stated objective of boosting AI development. (Lots of discussion also focuses, of late, on compute, which is beyond the scope of this particular report.) The heavy focus on model weights has hit industry stock prices and valuations as well. When Chinese firm DeepSeek released a new model that it claimed beat ChatGPT's performance, US AI firms lost about $1 trillion in valuation in 24 hours—amid the worry that other (Chinese) firms might easily replicate DeepSeek's use of open-source model weights.[12] Training data is still important for AI R&D—for instance, in how valuable curating the right, often proprietary datasets is for building AI drug discovery models[13]—but policy focus and debate have shifted greatly to legal, technical, innovation, tech governance, and national security issues surrounding model weights.[14]

At the height of the previous focus on AI training data, some scholars and analysts, certainly, pushed back against a myopic focus on one part of data and AI R&D. Matt Sheehan wrote a piece for the Macro Polo think tank in July 2019 arguing that strategic AI development depends not just on data quantity but on data depth (aspects of behavior or events captured), quality (accuracy, structure, storage), diversity (heterogeneity of users or events), and access (availability of data to relevant actors).[15] The industry-aligned Center for Data Innovation pu-

8. *AI Super-Powers*, first published in 2018. See: Andy Bast, Interview: "China's Greatest Natural Resource May Be Its Data," *60 Minutes*, CBS News, July 14, 2019, https://www.cbsnews.com/news/60-minutes-ai-chinas-greatest-natural-resource-may-be-its-data-2019-07-14/; Michael Chiu, Interview: "Kai-Fu Lee's Perspectives on Two Global Leaders in Artificial Intelligence: China and the United States," McKinsey Global Institute, June 14, 2018, https://www.mckinsey.com/featured-insights/artificial-intelligence/kai-fu-lees-perspectives-on-two-global-leaders-in-artificial-intelligence-china-and-the-united-states.
9. *WIRED*, January 14, 2020, https://www.wired.com/story/light-hand-ai-hard-line-china/; Justin Sherman, "Don't be Fooled by Big Tech's Anti-China Sideshow," *WIRED*, July 30, 2020, https://www.wired.com/story/opinion-dont-be-fooled-by-big-techs-anti-china-sideshow/.
10. "NTIA Solicits Comments on Open-Weight AI Models," US Department of Commerce, press release, February 21, 2024, https://www.commerce.gov/news/press-releases/2024/02/ntia-solicits-comments-open-weight-ai-models.
11. 90 FR 4544 (2025).
12. Dan Milmo et al., "'Sputnik Moment': $1tn Wiped off US Stocks after Chinese Firm Unveils AI Chatbot," *The Guardian*, January 27, 2025, https://www.theguardian.com/business/2025/jan/27/tech-shares-asia-europe-fall-china-ai-deepseek. Notably, the reaction described, of course, could have had more nuance in articulating the ways that a US company's research or advancement might benefit all kinds of companies, including other ones in the United States.
13. See: Milad Alucozai, Will Fondrie, and Megan Sperry, "From Data to Drugs: The Role of Artificial Intelligence in Drug Discovery," Wyss Institute, January 9, 2025, https://wyss.harvard.edu/news/from-data-to-drugs-the-role-of-artificial-intelligence-in-drug-discovery/; Chen Fu and Qiuchen Chen, "The Future of Pharmaceuticals: Artificial Intelligence in Drug Discovery and Development," *Journal of Pharmaceutical Analysis* 15, no. 8 (August 2025), https://doi.org/10.1016/j.jpha.2025.101248.
14. See: Sella Nevo et al., *Securing AI Model Weights: Preventing Theft and Misuse of Frontier Models*, *RAND*, May 2024, https://www.rand.org/pubs/research_reports/RRA2849-1.html; Janet Egan, Paul Scharre, and Vivek Chilukuri, *Promote and Protect America's AI Advantage*, *Center for a New American Security*, January 20, 2025, https://www.cnas.org/publications/commentary/promote-and-protect-americas-ai-advantage; Alan Z. Rozenshtein, "There Is No General First Amendment Right to Distribute Machine-Learning Model Weights," *Lawfare*, April 4, 2024, https://www.lawfaremedia.org/article/there-is-no-general-first-amendment-right-to-distribute-machine-learning-model-weights; Raffaele Huang, Stu Woo, and Asa Fitch, "Everyone's Rattled by the Rise of DeepSeek—Except Nvidia, Which Enabled It," *Wall Street Journal*, February 2, 2025, https://www.wsj.com/tech/ai/nvidia-jensen-huang-ai-china-deepseek-51217c40.
15. Matt Sheehan, "Much Ado About Data: How America and China Stack Up," Paulson Institute: MacroPolo, July 16, 2019, https://archivemacropolo.org/ai-data-us-china/?rp=e.

blished an article in January 2018 critiquing the flaws of making that economic comparison from an innovation standpoint.[16] Since that overfocus on AI training data, others have made points about the need for a broader view of AI's competitive data factors, too. For instance, Claudia Wilson and Emmie Hine recently cautioned against export controls on open-source models (and elements like model weights), which could trigger an "unfettered" AI "race"[17]—while scholars like Kenton Thibaut point out the drawbacks of hyper-fixating on a single, "silver bullet" for AI leadership in general.[18]

Still, many DC think tank roundtables and policy discussions center on model weights. This is not to say there is no reason for concern about how to best secure model weights, including the model weights of US AI companies, against theft by Chinese actors.[19] Trade secret theft is clearly a concern for US companies, as it is for the US government. Again, though, training data—and the importance of a range of types of training data (e.g., beyond just LLM training data to include training data used to power novel AI drug discovery models, etc.)—has taken somewhat of a backseat in recent policy conversations compared to model weights as well as other AI supply chain components beyond scope, notably compute.

However, the pendulum swing from focusing on training data quantity to focusing on model weights illustrates a few prevailing problems in policy and industry debates about AI and data.

- Overconfidence about which data element or attribute will most drive AI R&D can lead researchers and policymakers to skip past important, open questions (e.g., what factors might matter? in what combinations? to what end?), wrongly treating them as resolved.
- Oversimplified views of how data flows into and out of, constitutes, and powers AI models can lead policymakers to discuss "AI data" as one bucket of data to compete on, govern, and secure, rather than as many data types with different contexts.
- Over-fixation on a single, AI-related data component can guide policymakers and practitioners to treat the data component as a new, flashy "AI" phenomenon—overlooking existing security and risk mitigation frameworks and best practices, which many organizations may still not have implemented in the first place.

A continued challenge for plenty of US policymakers and industry leaders is taking, and having the intellectual and economic space for, a more comprehensive assessment of the different kinds of data and data components that enable AI R&D.[20] Focusing largely on training data quantity one moment and model weights the next can contribute to piecemeal, sometimes lopsided policy and often unfounded analytical assumptions. Take the example of protecting AI training data. When policymakers overfocused on AI training data and the idea that training data quantity matters most, many policy papers,[21] alongside much industry lobbying,[22] advocated for weak privacy laws so the United States could "beat" China—a country

16. Joshua New, "Why Do People Still Think Data Is the New Oil?" Center for Data Innovation, January 16, 2018, https://datainnovation.org/2018/01/why-do-people-still-think-data-is-the-new-oil/.
17. Claudia Wilson and Emmie Hine, "Export Controls on Open-Source Models Will Not Win the AI Race," *Just Security*, February 25, 2025, https://www.justsecurity.org/108144/blanket-bans-software-exports-not-solution-ai-arms-race/.
18. Kenton Thibaut, "What DeepSeek's Breakthrough Says (and Doesn't Say) About the 'AI race' with China," New Atlanticist (blog), January 28, 2025, https://www.atlanticcouncil.org/blogs/new-atlanticist/what-deepseeks-breakthrough-says-and-doesnt-say-about-the-ai-race-with-china/.
19. See, for instance, among the many recent articles and headlines: Jason Ross Arnold, "High-Risk AI Models Need Military-Grade Security," *War on the Rocks*, August 6, 2025, https://warontherocks.com/2025/08/high-risk-ai-models-need-military-grade-security/; Ryan Lovelace, "Congress Digs into China's Alleged Theft of America's AI Secrets," *Washington Times*, May 7, 2025, https://www.washingtontimes.com/news/2025/may/7/congress-digs-chinas-alleged-theft-americas-ai-secrets/
20. The frantic coverage of every new AI development by many media outlets does little to help resolve the data challenges surrounding AI R&D.
21. See the discussion by authors at the Belfer Center about whether "the United States has essentially conceded the [AI] race [with China] because of concerns over the average individual's privacy": Graham Allison and Eric Schmidt, *Is China Beating the U.S. to AI Supremacy?* (Cambridge: Harvard Kennedy School Belfer Center, August 2020), https://www.belfercenter.org/publication/china-beating-us-ai-supremacy.
22. Nitasha Tiku, "Big Tech: Breaking Us Up Will Only Help China," *WIRED*, May 23, 2019, https://www.wired.com/story/big-tech-breaking-will-only-help-china/; Josh Constine, "Facebook's Regulation Dodge: Let Us, or China Will," *TechCrunch*, July 17, 2019, https://techcrunch.com/2019/07/17/facebook-or-china/.

which, in this framing, has zero privacy restrictions or data limits whatsoever.[23] Now that the conversation has shifted to model weights, however, much policy discourse has focused on how China's restrictions on outbound data transfers lock down its technological advantages[24]—a sudden pivot in the conversation that now suggests the United States might benefit from some privacy laws in the first place.

This mall-moving, pendulum-swinging tendency can mean policymakers choose a single piece of the data in the AI supply chain to focus on myopically, which can cause policy to make more sudden lurches and miss opportunities to make longer-term investments in the security of all AI components. It can also cause policy narratives about AI to move in contradictory directions based on whichever slice of AI-related data is receiving the most attention at one given moment. When the policy focus centered on fueling training data quantity, some (as described above) talked about basic data privacy and security restrictions as harmful to technology development and the country. Yet these are precisely the kinds of policies that are helpful to protect against theft of and illicit access to model weights.

To provide another framework for researchers, industry leaders, and especially policymakers to approach important data and AI debates—from the nature of the data components most likely to drive AI R&D, to the economic and national security risks of ungoverned access to AI-involved data—the next section lays out a data-focused concept to help widen the lens.

Untangling the Data in the "AI Supply Chain"

The AI supply chain—organizations, people, activities, information, and resources enabling AI research, development, deployment, and more—is complex, shifting, and global. It involves several elements not covered in this report, such as human talent and compute, and it also includes the focus of this report: data.

Just as "AI" is not a single technology but an umbrella term for a suite of technologies, the data relevant to AI research, development, deployment, use, maintenance, governance, and security is no single data type, source, or format, either. Instead, there are several data components in the AI supply chain (described below). These data components fit into the AI supply chain because researchers, developers, deployers, users, maintainers, governors, securers, and attackers of AI systems depend upon and get access to different kinds of data—transmitted, stored, and analyzed in different ways—to make it all happen. They also fit into the AI supply chain because a wide range of entities around the world—from individuals who publish self-labeled datasets to corporations that analyze AI model outputs—supply, access, and use the underlying data, too. This idea echoes the concept of AI as a value chain (referring to the business activities that deliver value to customers),[25] though focused specifically on data components.

The data in the AI supply chain covers many data types, sources, and formats—all of which need secure safeguards to enable competition, boost public trust, and protect against the leaks, exploitation, and other risks delineated above. As conceptualized in this report, the data in the AI supply chain includes the data describing an AI model's properties and behavior, as well as the data associated with building and using a model. It also includes AI models themselves and the different systems that facilitate the movement of data into and out

23. While there are many differences between the US and Chinese environments vis-à-vis data, these notions are not entirely true. See: Samm Sacks and Lorand Laskai, "China's Privacy Conundrum," *Slate*, February 7, 2019, https://slate.com/technology/2019/02/china-consumer-data-protection-privacy-surveillance.html; Sam Bresnick, "The Obstacles to China's AI Power," *Foreign Affairs*, December 31, 2024, https://www.foreignaffairs.com/china/obstacles-china-ai-military-power.

24. Jessie Yeung, "China's Sitting on a Goldmine of Genetic Data – and It Doesn't Want to Share," CNN, August 12, 2023, https://www.cnn.com/2023/08/11/china/china-human-genetic-resources-regulations-intl-hnk-dst.

25. See: Beatriz Botero Arcila, *AI Liability Along the Value Chain* (San Francisco: Mozilla Foundation, April 2025), https://blog.mozilla.org/netpolicy/files/2025/03/AI-Liability-Along-the-Value-Chain_Beatriz-Arcila.pdf; Max von Thun and Daniel A. Hanley, *Stopping Big Tech from Becoming Big AI* (San Francisco: Mozilla Foundation, October 2024), https://blog.mozilla.org/wp-content/blogs.dir/278/files/2024/10/Stopping-Big-Tech-from-Becoming-Big-AI.pdf; SPEAR Invest, "Diving Deep into the AI Value Chain," NASDAQ, December 18, 2023, https://www.nasdaq.com/articles/diving-deep-into-the-ai-value-chain. See also: "The Value Chain," Harvard Business School: Institute for Strategy and Competitiveness, accessed August 26, 2025, https://www.isc.hbs.edu/strategy/business-strategy/Pages/the-value-chain.aspx.

of models.[26] Laid out in the visualized framework and tables below, this report conceptualizes seven parts of the data and core data systems in the AI supply chain:

- Training data
- Testing data
- Models (themselves)
- Model architectures
- Model weights
- Application Programming Interfaces (APIs)
- Software Development Kits (SDKs)

This paper draws some inspiration at a high level from the November 2024 paper by Qiang Hu and three other scholars on the large language model (LLM) supply chain, which envisioned a framework for thinking about the components and processes that go into LLMs.[27] Nonetheless, this paper differs in focusing more on the data components themselves rather than the activities to produce them (like dataset processing); delineates data components by their properties and functional differences (such as distinguishing between training data and testing data); and looking at the data supply chain for AI technologies broadly (rather than just LLMs). This paper's analysis also differs in that it focuses specifically on the need for security.

Notably, the first five of these components are data per se, or the model itself. The last two, however—APIs and SDKs—are neither data nor models themselves; instead, they are code and software systems that enable data to pass into, extract from, and otherwise collect around AI models. For example, business users of an LLM may use an API to submit questions to the chatbot in batches; mobile consumers using an AI image recognition app may, whether they know it or not, depend on an SDK to take their snapshot of a bird and submit it to a cloud-hosted AI model, which then returns back through the SDK's code the species of the bird in question. While the concept of data components of the AI supply chain does not list out every software system that could interact with AI data components, it includes APIs and SDKs because of their prevalence and their security relevance in delivering AI data to and from cloud systems and mobile devices; after all, virtually all the major AI commercial companies offer access to APIs to use their models (to include submitting queries to chatbots and uploading images to recognition models).

Figure 1 lists each of the seven data components of the AI supply chain, defines them, and provides a few examples of the companies and other entities involved in that component or its sourcing. Again, this does not include several other elements of the AI supply chain (e.g., human talent, compute) and focuses mostly on traditional AI models (e.g., excludes AI agents).

26. While recognizing the necessity of evaluating AI in relation to the social, political, and economic systems that researchers, companies, and others operate within and use to build AI technologies—such as exploitative labor systems and the environmental system—this report focuses, for scope- and length-limitation purposes, on a typology of the digital and data elements themselves of relevance for AI R&D. For essential reading on other systems that generate data, move data into AI systems, and much more, see: Tamara Kneese, *Climate Justice and Labor Rights: Part I: AI Supply Chains and Workflows* (New York: AI Now Institute, August 2023), https://ainowinstitute.org/general/climate-justice-and-labor-rights-part-i-ai-supply-chains-and-workflows; Kashmir Hill, *Your Face Belongs to Us: A Tale of AI, a Secretive Startup, and the End of Privacy* (New York: Penguin Random House, 2023); Billy Perrigo, "Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic," *TIME*, January 18, 2023, https://time.com/6247678/openai-chatgpt-kenya-workers/; Adrienne Williams, Milagros Miceli, and Timnit Gebru, "The Exploited Labor Behind Artificial Intelligence," *Noema Magazine*, Berggruen Institute, October 13, 2022, https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence/.

27. Qiang Hu et al., "Large Language Model Supply Chain: Open Problems from the Security Perspective," arXiv, November 3, 2024, https://arxiv.org/abs/2411.01604 (see, in particular, page 2's LLM supply chain map).

Fig. 1: Data Components of the AI Supply Chain

Component	Definition	Examples of Involved Companies and Entities
Training data	Datasets used to shape a model's initial configuration to be able to have the model recognize patterns in data, make statistical predictions based on data, generate content by iteratively updating the model to increase its predictive accuracy, and more.[I] Example: Feeding an image recognition model images of cats and dogs (the training data) to learn the differences between them.	• The nonprofit Common Crawl, since 2007, maintains a free, open repository of web crawl data for researchers. Its current dataset: over 250 billion webpages scraped over 18 years.[II] • Labeled Faces in the Wild (LFW) from academic researchers, a dataset of 13,233 web-collected and labeled face images.[III]
Testing data	Datasets used to evaluate how a model performs on recognizing patterns in data, making statistical predictions based on data, and more, after the model has been initially trained—often used by many different developers, such that results can be compared and "benchmarked" across models. The data could be a subset of training data, though not always—its functional distinction singles it out (e.g., perhaps a startup or government agency saves its most proprietary, sensitive, and realistic data examples for its testing phase, so it can see how the model would perform in real-world environments). Example: Feeding an image recognition model images of cats and dogs it has not yet seen (the testing data) so the developer can see if the model has learned the differences between them.	• MMMLU from academic researchers, a dataset used to evaluate LLM performance on 57 subjects ranging from STEM to the humanities (which could also be training data).[IV] • ImageNet from academic researchers, a dataset of over 14 million annotated images used to test a model on whether it can (i) detect the items, objects, etc. that an image contains and (ii) where exactly those items, objects, etc. are in the images (i.e., pixel locations) (which could also be training data).[V]
Models (themselves)	The overall program with the data processing structure/architecture to take data as inputs; recognizes patterns in, makes predictions based off of, or perform other mathematical analysis of the data; and generate a resulting output.[VI] Models are made up of both a model architecture (below) and model weights (below that). Example: An anthropology student signs up for an account online to be able to upload text into a language translation model (the model), already trained, tested, and deployed, to help translate documents specific to the research project.	• YOLOv7 from academic researchers, a relatively high-performing object recognition model.[VII] • Codestral from French AI company Mistral AI, an LLM to assist with low-latency, high-frequency programming tasks such as code correction and test generation (e.g., where an internet user could access the model but not its underlying weights per se, for example).[VIII]

I See: Shigeyuki Takano, *Thinking Machines: Machine Learning and its Hardware Implementation* (Cambridge: Academic Press, 2021): 1–18, https://www.sciencedirect.com/science/article/abs/pii/B9780128182796000116; Nayna Jaen, "How AI Is Trained: The Critical Role of AI Training Data," RWS TrainAI, March 26, 2024, https://www.rws.com/artificial-intelligence/train-ai-data-services/blog/how-ai-is-trained-the-critical-role-of-ai-training-data/; Michael Chen, "What Is AI Model Training and Why Is It Important?" Oracle, December 6, 2023, https://www.oracle.com/uk/artificial-intelligence/ai-model-training/.

II Common Crawl (website), accessed April 15, 2025, https://commoncrawl.org.

III Labeled Faces in the Wild (LFW) Dataset, (website), accessed April 17, 2025, https://paperswithcode.com/dataset/lfw.

IV Dan Hendrycks et al., "Measuring Massive Multitask Language Understanding," arXiv, September 7, 2020, https://arxiv.org/abs/2009.03300; Dan Hendrycks et al., "Massive Multitask Language Understanding: Test Leaderboard," GitHub, accessed September 4, 2025, https://github.com/hendrycks/test?tab=readme-ov-file.

V ia Deng et al., "ImageNet: A Large-Scale Hierarchical Image Database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, https://doi.org/10.1109/CVPR.2009.5206848; "About ImageNet," ImageNet, accessed September 4, 2025, https://www.image-net.org/about.php.

VI See: "HPE Glossary: What are AI models?" HPE, accessed April 29, 2025, https://www.hpe.com/us/en/what-is/ai-models.html; Chen, "What Is AI Model Training."

VII Chien-Yao Wang, Alexey Bockhovskiy, and Hong-Yuan Mark Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Recognition," arXiv, July 6, 2022, https://arxiv.org/abs/2207.02696; Wong Kin Yiu et al., "Official YOLOv7," GitHub, accessed April 29, 2025, https://github.com/WongKinYiu/yolov7.

VIII "Models Overview," Mistral AL, accessed April 29, 2025, https://docs.mistral.ai/getting-started/models/models_overview/.

Fig. 1: Data Components of the AI Supply Chain (cont.)

Component	Definition	Examples of Involved Companies and Entities
Model architectures	The layout of how data flows and processes within an AI model, such as a neural network, including vis-à-vis the inputs, hidden layer or layers (which analyze the previous layer's output to create the next layer's input), and outputs.[IX] Example: A government research scientist downloads an open-source model architecture from the internet, which they duplicate. They then train each one on distinct training datasets, yielding two very different resulting models. (These differently resulting models therefore share a similar architecture but, as a result of the differences in training, have different model weights that impact their behavior—discussed below.)	• The ResNet architecture for deep neural networks laid out in 2015, where an image recognition model uses "residual blocks," or stacks of layers to take the output of one layer and add it to another layer deeper in the block.[X] • The encoder-decoder structure used in many commercial LLMs, where an "encoder" takes inputs and produces numerical representations of the inputs, passes those representations and the start of one input to the "decoder," and iterates to have the decoder generate the right sequences.[XI]
Model weights	Numerical parameters that specify the strength between variables as a model trains and fine-tunes—that is, how the model numerically represents connections between pieces of data to achieve the desired output or result.[XII] Example: As a self-driving vehicle AI model is trained, the model creates a set of numbers that represent how it connects inputs into the vehicle's sensors with analyzed outputs it can use to make driving decisions like steering and braking.	• The final learned parameters and biases of a trained neural network, shared online under an Open Source Initiative license, without the script or frameworks used to create and curate the training data—or the full dataset used for training the finished model.[XIII] • Model weights of closed-source models trained using more than 1026 computational operations, or those whose export from the United States was previously restricted under the Biden administration's now-rescinded AI diffusion rule.[XIV]
Application Programming Interfaces (APIs)	Software systems that allow individuals to write code to input data to and pull data from a system, which could include a database or an AI model—such as to download training data from an online repository or send queries or other data inputs to an externally hosted and deployed AI model for processing. Example: A medical student writes a script to upload batches of images they took in a clinical setting to a commercially available, externally hosted AI image recognition model and then download and capture the responses for their research.	• Perplexity API from Perplexity, which allows paying customers to integrate multiple of Perplexity's models into their own projects, such as having a model generate a response to a chat conversation.[XV] • Cloud Healthcare API from Google Cloud, which enables customers to ingest, transform, and store healthcare data—and integrate it with analytics and ML solutions such as BigQuery, AutoML, and Vertex AI.[XVI]
Software Development Kits (SDKs)	Prepackaged code that enables developers to easily add functionalities to their apps and programs—such as communicating with servers hosting an AI model. Example: A mobile app developer downloads an AI company's SDK from its website, which provides the code necessary to allow the developer's app users to interact with the AI company's language translation model through the app's interface.	• AI SDK by Vercel, the US cloud platform-as-a-service company, which helps developers integrate LLMs into software via common programming frameworks like React and Node.js.[XVII] • Stable Animation SDK by Stability AI, which helps developers integrate text-to-image generative models (like Stable Diffusion 2.0) into their software.[XVIII]

IX See: "What Is a Neural Network?," Amazon Web Services (AWS), accessed August 27, 2025, https://aws.amazon.com/what-is/neural-network/#:~:text=A%20neural%20network%20is%20a,that%20resembles%20the%20human%20brain; "What Is Neural Network Architecture?," H2O.ai, accessed August 27, 2025, https://h2o.ai/wiki/neural-network-architectures/; "The Essential Guide to Neural Network Architectures," V7, July 8, 2021, https://www.v7labs.com/blog/neural-network-architectures-guide.

X Kaiming He et al., "Deep Residual Learning for Image Recognition," arXiv, December 10, 2015, https://arxiv.org/abs/1512.03385; Prem Oommen, "ResNets – Residual Blocks & Deep Residual Learning," Towards Data Science, November 28, 2020, https://towardsdatascience.com/resnets-residual-blocks-deep-residual-learning-a231a0ee73d2/.

XI Hugging Face, "Transformer Models: Encoder–Decoders," YouTube (video), June 14, 2021, https://www.youtube.com/watch?v=0_4KEb08xrE&t=4s; Akshit Mehra, "Exploring Architectures and Configurations for Large Language Models (LLMs)," Labellerr, April 5, 2024, https://www.labellerr.com/blog/exploring-architectures-and-configurations-for-large-language-models-llms/; Patrick von Platen, "Transformers-Based Encoder-Decoder Models," Hugging Face, October 10, 2020, https://huggingface.co/blog/encoder-decoder.

XII See: Alisdair Broshar, "What Are LLMs? An Intro into AI, Models, Tokens, Parameters, Weights, Quantization and More," Koyeb, April 25, 2024, https://www.koyeb.com/blog/what-are-large-language-models; "Background: AI Model Weights," US National Telecommunications and Information Administration, accessed April 29, 2025, https://www.ntia.gov/programs-and-initiatives/artificial-intelligence/open-model-weights-report/background.

XIII "What Are Open Weights?" Open Source Initiative, accessed August 27, 2025, https://opensource.org/ai/open-weights.

XIV 90 FR 4544 (2025).

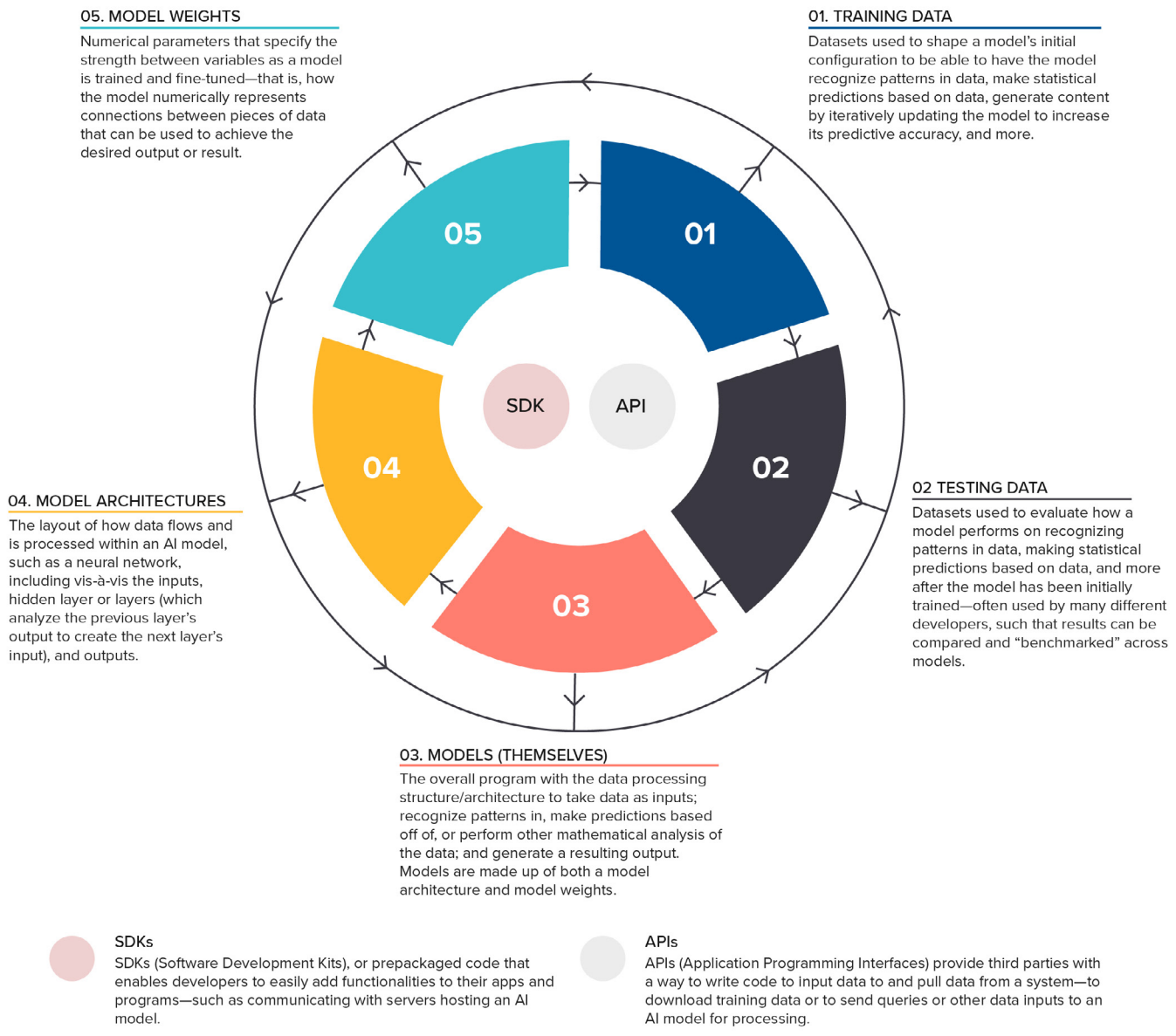
XV "What is the API?", Perplexity, accessed April 29, 2025, https://www.perplexity.ai/help-center/en/articles/10354842-what-is-the-api; "Sonar Developer Platform," Sonar by Perplexity, accessed April 29, 2025, https://docs.perplexity.ai/home; "Chat Completions," Sonar by Perplexity, accessed April 29, 2025, https://docs.perplexity.ai/api-reference/chat-completions-post.

XVI "Cloud Healthcare API," cloud.google.com, accessed April 29, 2025, https://cloud.google.com/healthcare-api?hl=en.

XVII "The AI Toolkit for TypeScript," accessed April 29, 2025, https://sdk.vercel.ai; "Why Use the AI SDK," AI SDK (Versel), accessed April 29, 2025, https://sdk.vercel.ai/docs/introduction.

XVIII "Stability AI Releases Stable Animation SDK, a Powerful Text-to-Animation Tool for Developers," stability.ai, May 11, 2023, https://stability.ai/news/stable-animation-sdk.

Graphic 1: Components of AI Data Supply Chain



Many of the above seven data components of the AI supply chain can stand on their own. A computer system can house thousands of training data images in one folder and separately store hundreds of testing image examples in another folder; the files themselves, in a literal sense, are distinct. Similarly, a company can build and deploy an AI model with a paid API, through which users can query the model without making any SDKs available for developers to more easily integrate the model into software. Some websites make training data publicly and freely available without ever supplying model weights to their visitors, and some companies will provide elements of their data supply chains to purchasers for auditing, but typically not all of their underlying training data or every model weight.

However, all seven data components have overlapping functional roles in the AI ecosystem. Academic researchers, government technologists, or startup developers looking to build a competitive healthcare image recognition model will need training data and testing data (including potentially rounds of training data, to fine-tune a model) to make it happen; without testing data, it is difficult to systemically evaluate a model's

performance so it can be tweaked, and without training data, there is no model to test and fine-tune. Companies that want to deploy their already built and tested models have many incentives to create both APIs and SDKs, so that different users working in different environments—whether a nontechnical lawyer looking to query a chatbot or a machine learning PhD looking to use the chatbot in their own app—can readily access the technology.

The seven data components have overlapping suppliers, which are also geographically dispersed. Companies like Amazon Web Services, for example, store and make publicly available countless training datasets, including those from other parties.[28] (AWS, in this specific example, also offers cloud services to government, companies, universities, civil society groups, and individuals to train, test, fine-tune, and deploy AI models.) Universities like Tsinghua University in China and the Indian Institute of Technology publish open-source AI models and the related data (e.g., training, testing data) as part of academic studies.[29] Community-maintained websites like Kaggle, popular in the AI R&D community, host many kinds of training and testing data, and open-source platforms like GitHub host various datasets as well as models themselves, too. Simultaneously, these and many other suppliers of data components in the AI supply chain are consumers of the data components. Amazon uses training and testing data to build AI products and services; universities, such as Tsinghua and the Indian Institutes of Technology (IIT), publish study-linked datasets just as they may procure AI data components and related technologies (e.g., cloud services) to conduct the research in the first place.

And the data components in the AI supply chain themselves may interact with each other. (Again, this report does not include coverage of AI agents as explained above.) When a developer initially trains a model (using a training dataset) and then iterates on the model by testing it (using testing data) and fine-tuning it further (using more training data), the resulting model and the model weights are in part the byproduct of the training and testing datasets used. The model architecture selected before sourcing the training data will likewise influence what the model weights and resulting model ultimately look like—as well as the resulting model's data inputs and outputs via an API or SDK. Similarly, when a company acquires a certain training dataset and uses it to train a model with specified parameters, it shapes the nature of which testing data and additional training data it will subsequently source from the supply chain. If the company wants a model to be completely open-source, for instance, then it will need to select or construct datasets of only open-source testing and training data from the data in the AI supply chain; if the company elects to go with Portuguese-language training data for building a voice-to-text AI chatbot, it will need Portuguese-language testing data, perhaps even sourced only from Brazil, to evaluate the initial model's behavior. These are additional ways in which interactions between data components of the AI supply chain can impact data sourcing decisions.

Even nation-states looking to secure their respective AI systems and potentially steal or compromise those in other countries may need to consider everything from safeguarding the models themselves (and all the data and weights within them) against exploitation to identifying sensitive testing datasets that need protection. These AI data components are distinct in the framework above. But their overlapping and interdependent roles in AI R&D make them collectively integral to understanding AI competitiveness and innovation—and how to ensure robust, effective governance across safety, security, privacy, trust, and much more. The concept of a supply chain, as in other areas like manufacturing, helps to drive analysis towards the interaction and interdependence of the various subcomponents and their suppliers. None truly can stand alone.

Instead of the policy and analytic pendulum swinging from one area (like training data) to another (like model weights) with underappreciation for the broader landscape, this framework and the functional overlaps between components make clear that strategic competition and governance over AI and data cannot myopically focus on one element. Doing so leads to the analytic issues laid out in the last section and detracts from the complex, entangled nature of the data supply chain components that are relevant to AI research, development, deployment, use, maintenance, governance, and security. Policymakers only increase the likelihood of missing major opportunities and risks. Hence, with this foundation, the next section uses the framework of data in the AI supply chain above to zoom in on the security risks facing the data components in the AI supply chain—to illuminate what organizations and policymakers might do about them.

Parsing the Risks—and Pursuing Better Security

Policymakers, technologists, and others working on AI (e.g., on governance) can use the framework from the last section

28. "Registry of Open Data on AWS," Amazon Web Services (AWS), accessed June 16, 2025, https://registry.opendata.aws.
29. See: Building AI for India! (website), accessed June 16, 2025, https://ai4bharat.iitm.ac.in; Tsinghua University: Institute for Artificial Intelligence Foundation Model Research Center (website), accessed June 16, 2025, https://fm.ai.tsinghua.edu.cn.

to map data components in the AI supply chain, in different states and contexts, to security controls and risk mitigations. This section describes and details how such a process would work across three different approaches. Using the framework to parse risks enables individuals and organizations to identify the best existing practices to leverage in protecting AI data components. In some cases, this may save organizations time and money if they already have the security controls and risk mitigations in place elsewhere—and even if organizations have not yet implemented the existing controls and mitigations for non-AI systems and data, they do not need to create the controls and mitigations from scratch. Related, the framework can also help individuals and organizations to identify gaps in existing best practices—and, as exemplified in the below discussion, think about how new security controls or risk mitigations could be developed and used to address AI-specific data risks.

As alluded to earlier, better security across the data components of the AI supply chain can mitigate risks of breaches and data interception, shield data and resulting AI technologies from theft (including by competitors), enhance protections for individuals' privacy, bolster public trust, limit organizations' liability risk, and strengthen US national security. Lapses in security across the data components of the AI supply chain, however, can contribute to universal problems such as data breaches and interceptions, intellectual property theft, privacy leaks and violations, and undermined public trust in AI technologies—as well as US-specific issues, such as better enabling governments adversarial to the US to hack data or infiltrate US technology supply chains. A methodological approach to this risk mapping can help organizations mitigate risk and help policymakers develop more rigorous, tailored policies on AI and data security.

Leveraging the last section's framework, this section evaluates three different approaches to mapping the data components of the AI supply chain to security controls and risk mitigations. The first looks at the state of a data component of the AI supply chain: is the data at rest, in motion, or in processing? The second looks at the threat actors with an interest in the AI supply chain and its data components: what are the threats, vulnerabilities, and consequences? And the third looks at the interaction of data components of the AI supply chain and the suppliers: who are the suppliers, and what are their security controls—or risks?

A recent paper on how to enhance third-party flaw disclosures for AI models argues that the AI sector has much to learn from software security.[30] This section follows in similar spirit. Instead of reinventing the wheel, these three approaches to data security in the AI supply chain help map complex questions about data in the AI supply chain to existing data security and supply chain security best practices. Then, where existing security controls and risk mitigations are insufficient for AI-specific risks to data—at least two of which are spotlighted below—these three approaches can help illuminate where new, AI-specific mitigations are needed. Figure 2 summarizes the three approaches, ahead of the more detailed discussion that follows.

30. Shayne Longpre et al., "In-House Evaluation Is Not Enough: Towards Robust Third-Party Disclosure for General-Purpose AI," arxiv.org, March 25, 2025, https://arxiv.org/abs/2503.16861 (an important point the authors make is to call the idea that general-purpose AI systems "are unique from existing software and require special disclosure rules" a "misconception").

Fig. 2: Three Potential Approaches to Securing Data in the AI Supply Chain

Step 1: Identify Component and/or Source in Supply Chain	Step 2: Identify Security Approach	Step 3: Identify Mitigations from Existing Data and Supply Chain Security Best Practices
Approach one, two, and three: Training data, testing data, models (themselves), model architectures, model weights, APIs, or SDKs? Who is involved in sourcing the data, including the creation of the data, any subsequent processing of the data, and its dissemination?	**Approach one:** Is the data at rest, in motion, or in processing? (Focus on a data component within the supply chain and its current state.) **Approach two:** What are the threats, vulnerabilities, and consequences of security risks to specific components? (Focus on threat actors and risks to a data component within the supply chain.) **Approach three:** Who are the suppliers, and what are their security controls—or risks? (Focus on a data component supplier or source within the supply chain and related actors.)	**Approach one:** Identify relevant protections for data at rest, in motion, or in processing, such as NIST SP 800-53 SC-28 ("Protection of Information at Rest"). Tailor to the use case and, if applicable, the organization's supply chain role(s). **Approach two:** Identify relevant sources of information about the threat actor. Secure vulnerable systems (e.g., using NIST, ISO best practices) that match up against the threat actor's capabilities (and could be a way to access a specific data component in the AI supply chain). Tailor protections for the data component itself or components themselves (e.g., encryption, storage rules) based on the threat actor's targeting intent. **Approach three:** Conduct due diligence, to extent possible, on supply chain actors' ownership, maliciousness, and susceptibility to malicious influence—drawing on best practices from financial sector and other know-your-customer checklists. Conduct due diligence, to extent possible, on the data security, cybersecurity, and supply chain security measures taken by those suppliers, to attempt to mitigate risks through the supply chain—drawing on best practices from GDPR compliance, supply chain risk management, and so on.

Approach one: Understand the 'state' of data

First, five of the seven data components of the AI supply chain (excluding APIs and SDKs, as they are not data per se) can be in different data states at different times. Each of these may come with specific security risks, under three states, commonly described as: "data at rest" (e.g., model weights sitting on a server, though not in use), "data in motion" (e.g., training data downloading from a website to a local machine), and "data in processing" (e.g., testing data feeding into an initially trained model). Cybersecurity professionals, when building organizational policies, programs, and processes, often apply this framework—at rest vs. in motion vs. in processing—to understand risks to data and mitigate them. AI-related data at rest, for instance, can be siphoned from databases by a hacker or sit exposed on a public server with no password, ready for anyone to download, because it was not subject to proper encryption and protection. This could enable criminals to target people in the data with scams or sell the data on the dark web. AI-related data in motion, similarly, that is weakly encrypted or entirely unencrypted could be intercepted by a nation-state as it moves from a cloud system through an API, enabling intelligence-gathering or intellectual property theft.

Each of these data states may require different kinds of encryption, different levels of access controls for employees, and so

on. Perhaps one security team is responsible for protecting a stored training dataset, while a research team is the only one authorized to modify the training dataset; the same data thus requires different security measures, such as different kinds of encryption and access control rules, when stored compared to when undergoing modification. A state agency or company may choose to implement a certain kind of robust encryption on data at rest when access or modifications are unnecessary but leave it unencrypted while in processing, or only encrypt it in very specific ways that still enable computation (i.e., while in processing).[31] Focusing security measures only on the data component in question (e.g., is it testing data or training data?) will fail to account for the ways a piece of data's current state impacts the risks to the data in that moment and the security measures to apply to it.

This framework—at rest vs. in motion vs. in processing—is therefore an effective means of tying classes of risks to the data components of the AI supply chain to specific, existing risk mitigations. Rather than assuming that the data components of the AI supply chain need entirely different protections because they are "AI-related," leveraging this framework contextualizes risks to the data components within broader risks to data, AI-related or not. For example, one of the National Institute of Standards and Technology's many security best practices focuses on "Protection of Information at Rest." The security control, known as SP 800-53: SC-28, delineates three components: cryptographic protection for "information on system components or media" as well as "data structures, including files, records, or fields"; offline storage to eliminate the risk of individuals gaining unauthorized data access through a network; and using hardware-protected storage, such as a Trusted Platform Module (TPM), to store and protect the cryptographic keys used to encrypt data.[32] Universities attempting to secure health training datasets on a department computer, companies looking to prevent hackers from stealing model weights sitting on a cloud server, or government agencies hoping to protect testing data from spies can all use these techniques to protect the data components of the AI supply chain while they are at rest.

Specific data components in motion and in processing, as captured in Figure 3, can likewise be mapped to specific NIST and other security best practices. NIST SP 800-53: SC-08, "Transmission Confidentiality and Integrity," specifies cryptographic protection, pre- and post-transmission security measures, how to conceal or randomize communications, and other steps[33] that a civil society group could take to secure AI model weights it sends to a federal funder agency; the agency's Cybersecurity Framework: PR.DS-10 control focuses on the confidentiality, integrity, and availability of data in use (i.e., in processing) and has many related controls such as account management, access enforcement, monitoring for information disclosure, system backups, cryptographic protections, and process isolation,[34] that a corporation could implement for all its independent contractors building an LLM.[35] This approach enables entities to identify a data component in their AI supply chain, understand its state, and map that state to best practices from NIST, the Systems and Organization Controls (SOC) 2 framework,[36] and other data security compliance guidelines.

These controls could vary not just based on the data state (at rest vs. in motion vs. in processing) but on the type of data, its source, and its context. For instance, companies interested in protecting larger model weights could turn to security measures intended for larger datasets, such as tight access controls, two-party authorization for data access, and endpoint software controls.[37] Companies might use this framework to arrive at a stronger level of security controls for larger model

31. For example, see more on how homomorphic encryption can be used to encrypt data, including AI training data, while still enabling computation on it: "Combining Machine Learning and Homomorphic Encryption in the Apple Ecosystem," Machine Learning Research, Apple, October 24, 2024, https://machinelearning.apple.com/research/homomorphic-encryption.
32. SP 800-53 Rev. 5.1.1, SC-28: "Protection of Information at Rest," US National Institute of Standards and Technology, accessed June 27, 2025, https://csrc.nist.gov/projects/cprt/catalog#/cprt/framework/version/SP_800_53_5_1_1/home?element=SC-28.
33. SP 800-53 Rev. 5.1.1, SC-08: "Transmission Confidentiality and Integrity," US National Institute of Standards and Technology, accessed June 27, 2025, https://csrc.nist.gov/projects/cprt/catalog#/cprt/framework/version/SP_800_53_5_1_1/home?element=SC-08.
34. "The NIST Cybersecurity Framework (CSF) 2.0," US National Institute of Standards and Technology, February 26, 2024, https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf.
35. "PR.DS-10: The Confidentiality, Integrity, and Availability of data-in-Use Are Protected," CSF Tools, accessed June 27, 2025, https://csf.tools/reference/nist-cybersecurity-framework/v2-0/pr/pr-ds/pr-ds-10/.
36. See: MJ Raber, "SOC 2 Controls: Encryption of Data at Rest – An Updated Guide," Security Boulevard, Techstrong Group, December 6, 2022, https://securityboulevard.com/2022/12/soc-2-controls-encryption-of-data-at-rest-an-updated-guide/.
37. Anthropic, for example, talks about using more than 100 different security controls to protect model weights. See: "Activating AI Safety Level 3 Protections," Anthropic, May 22, 2025, https://www.anthropic.com/news/activating-asl3-protections.

weights, in all data states, than they would apply to smaller, less sensitive training datasets.

At the same time, this mapping demonstrates ways in which existing security controls and risk mitigations may not address all AI-related data risks. Take the poisoning of an AI model, where bad actors attempt to insert "bad" data into training data, such as data that could cause serious errors or vulnerabilities if used to train a specific AI model.[38] If an organization scrapes training data from the internet (i.e., data in motion), imposing confidentiality and integrity controls on the scraped data would only catch modifications to the data after collecting it—not detect whether the data uploaded in the first place was poisoned from the start. If an organization is trying to ensure the security of that training data after scraping (i.e., data at rest), to give another example, encryption and access control measures could help to mitigate the risk of post-scrape tampering of data stored on the organization's systems. These measures would again fail, however, to protect the organization from scraping data that was compromised from the outset. While this is an intentionally simplistic discussion of AI poisoning, it underscores that traditional IT security measures for protecting data at rest, in motion, and in processing may not fully mitigate all risks to data in the AI supply chain. In this case, policymakers and organizations can look to guidance from NIST that explains types of poisoning attacks and potential mitigations—such as differential privacy applied to datasets and data sanitization techniques to remove poisoned samples of data before using the dataset for AI model training.[39]

38. For example, a bad actor could generate training examples with incorrect or altered labels with the express purpose of causing someone to unintentionally train a harmful or erroneous model. Apostol Vassilev et al., *NIST Trustworthy and Responsible AI – NIST AI 100-2e2025*: *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations,* US National Institute of Standards and Technology (March 2025): 20, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-2e2025.pdf.
39. Vassilev et al., *NIST Trustworthy and Responsible AI* , 20–27.

Fig. 3: Approach one illustrated

Approach One Question: Is Data at Rest, in Motion, or in Processing?	Risks to Data in the AI Supply Chain in this State	Potential Controls
Data at rest	Include theft of data (e.g., by hackers inside a storage system, insider threats violating access control policies), leakage of data (e.g., inadvertent disclosure of data on servers), and poisoning of data (e.g., insertion of bad data points). Example—Training Data: A company stores its training data on an internal server while not in use but fails to implement access controls. Employees without an authorized need to view the data could access it, resulting in unauthorized downloads to personal devices (outside the organization's control).	• Encryption of data. • Encryption of related files, records, or fields. • Offline storage of data, or particularly sensitive data, as a backup. • Offline storage of data, or particularly sensitive data, while not storing online to mitigate breach risk. • Hardware-protected storage • Access controls for employees, contractors, vendors, and customers.[I]
Data in motion	Include interception of data (e.g., by hackers listening in on network traffic), transmission of data to incorrect recipient (e.g., due to poor verification mechanisms or malicious interference), and poisoning of data (e.g., insertion of bad data or bad queries). Example—Testing Data: A government agency is considering procuring a private-sector AI model and runs it, on a cloud instance, against a testing dataset that the agency developed to evaluate the model's performance. While querying the third-party model, an agency developer does not use sufficient encryption, and a foreign intelligence entity intercepts the government agency's custom testing dataset.	• Encryption of data. • Confidentiality and integrity measures when data is prepared for transmission and initially received. • Encryption of message headers and routing information. • Access controls for APIs. • Protections to make physical access to networks difficult.[II]
Data in processing	Include theft of data (e.g., by hackers with access to AI training software or hardware architecture), degradation of processing functionality or accuracy (e.g., interfering with completeness of tasks like model training or testing data structuring), and leakage of data (e.g., inadvertent disclosure). Example—Training Data: A PhD student is in the middle of downloading a large health data training dataset from a website, yet the computer is already infected with malware and therefore enables the cybercriminals who installed it to siphon out the proprietary, personal data and sell it on the dark web.	• Account management. • Access controls. • Information flows controls. • Protections for audit information (e.g., data on access control compliance). • Monitoring for information disclosure. • Information exchange management (e.g., with NDAs, user agreements, vendor agreements). • System backups. • Security and privacy engineering principles in design of systems to process data. • Encryption.[III]

I. US National Institute of Standards and Technology, NIST SP 800-53 Rev. 5.1.1, SC-28.

II. US National Institute of Standards and Technology, NIST SP 800-53 Rev. 5.1.1, SC-08.

III. US National Institute of Standards and Technology, NIST Cybersecurity Framework 2.0; see also CSF Tools: PR-DS.10; SP 800-53 Rev. 5.1.1, SC-12 "Cryptographic Key Establishment and Management," US National Institute of Standards and Technology, accessed September 5, 2025, https://csrc.nist.gov/projects/cprt/catalog#/cprt/framework/version/SP_800_53_5_1_1/home?element=SC-12; NIST SP-800 5.1.1, SC-13, "Cryptographic Protection," US National Institute of Standards and Technology, accessed September 5, 2025, https://csrc.nist.gov/projects/cprt/catalog#/cprt/framework/version/SP_800_53_5_1_1/home?element=SC-13.

Approach two: Assess threat actor profile

Second, different threat actors as well as unwitting individuals (e.g., employees deceived by a phishing note, users making weak API passwords, etc.) can take actions that undermine the security of data components of the AI supply chain—or the security of a specific data component. Instead of focusing on the state of data components at risk, the developers, users, maintainers, governors, and securers of AI technologies can focus on the threat actors and scenarios themselves. Established threat actor risk frameworks can enable those entities and individuals to identify risks to the data in the AI supply chain, map an adversary's capabilities against known mitigations, and prioritize the security measures that are the most urgent. These mitigations can be specific not just to a data component's current state, but to any threat actor in question.

Having a threat actor-driven risk approach is essential for companies, universities, nonprofits, government agencies, and other organizations and groups involved with developing, using, maintaining, governing, and securing AI technologies and data. Focusing on technical mitigations, such as encrypting data at risk, can help organizations prioritize their biggest technological or process vulnerabilities internally, but they do little to help the organization understand which threat actors have an interest in which of their datasets. Using the first approach described above can help an organization to shore up its own defenses, but knowing which actors are the biggest threat to an organization—and what capabilities they bring to bear—might shift which security controls and risk mitigations are the biggest priority; threat actors could focus on stealing unexpected datasets, for instance, or have far better ability to poison training data than a university or corporate research lab might appreciate. While it is again not the only approach, centering threat actors and their capabilities is another lens through which to approach securing the data components of the AI supply chain.

Among many other risk assessment frameworks in the world, the US government often uses the framework of risk as a function of threat, vulnerability, and consequence.[40] Threats are composed of an adversary's intentions and capabilities.[41] Vulnerabilities are weaknesses inherent to a system (e.g., due to poor coding practices, interactions between components, or simply the inevitability of human error in a complex software system) or that have been introduced by an outside actor.[42] And consequences, in this framework, are outcomes that are either fixable or "fatal".[43]

Developers, users, maintainers, governors, and securers of AI technologies can use this approach to understand how different data components of the AI supply chain may be at risk. Because many of the threat actors targeting data components of the AI supply chain—from nation-states to cybercriminals—are often already on the radar of large and boutique cybersecurity firms, organizations can use existing threat data to render their assessments. From there, they can look to industry best practice guides such as International Organization for Standardization (ISO) controls and standards from organizations like NIST to mitigate risks most appropriately[44]—rather than taking a one-size-fits-all approach to a diversified threat and security landscape.

For example, a medium-sized research university might worry about an industrial cyber espionage firm targeting its large AI health training data repository. If the university knows that the group has strong financial motive (intent), that it is highly sophisticated at penetrating network edge devices, insecure routers, and mobile devices but does not have the ability to decrypt large datasets (capabilities), and that the university has far too many connected devices and routers to achieve adequate security (vulnerabilities), the university may avoid a high-impact theft of the data (consequence) by choosing to encrypt the data, store it offline whenever possible, and securely isolate the encryption keys. The robust encryption and the shift towards offline storage could minimize the likelihood that the firm is able to steal the data—and minimize the likelihood they could make use of the data even if they did manage to steal it.

If a leading US AI startup, to give another example, worries about a Chinese military hacker stealing its image recogni-

40. "Framework for Assessing Risks," US Office of the Director of National Intelligence (ODNI), April 2021, https://www.dni.gov/files/NCSC/documents/supplychain/Framework_for_Assessing_Risks_-_FINAL_Doc.pdf.
41. As the noted in the cited ODNI publication, "Key to this is using the latest threat information to determine if specific and credible evidence suggests an item or service might be targeted by adversaries. But, it must be noted, that while adversaries wish to do harm, they can only be successful if systems, processes, services, etc. are vulnerable to attacks."
42. These are examples provided by the author of how weaknesses can be "inherent" to a system in this context; they are not examples listed in the cited ODNI publication.
43. US Office of the Director of National Intelligence, "Framework for Assessing Risks."
44. See: ISO/IEC 27001:2022, "Information Security, Cybersecurity and Privacy Protection — Information Security Management Systems — Requirements," International Organization for Standardization, 2022, https://www.iso.org/standard/27001.

tion model weights, it could also apply this threat actor risk framework. The startup might suspect that the task is to steal its technology (intent), know that the hacker is highly sophisticated at network penetration and decryption of data (capabilities), and feel it has locked down its user accounts with strong passwords and multifactor authentication, but that its wide vendor and contractor base introduces many points of entry into its technological supply chain (vulnerability). As the firm is pre-revenue, its executives are of the view that a Chinese competitor getting a copy of its proprietary model weights and beating it to market could put the company out of business (consequence). Well aware that standard cybersecurity measures may not be enough to stop the military hacker's capabilities, the startup may choose to invest in even more advanced mechanisms—partitioning systems; storing numerous false copies of model weights that purport to be the real thing; moving training datasets and testing datasets already used into offline storage[45]—to ensure that its model weights are as well-protected as possible.

These are insights that would not be as obvious were the hypothetical medium-sized research university or the hypothetical US AI startup to focus only on the current state of the training data or model weights in question (i.e., at rest vs. in motion vs. in processing). Understanding the threat actor itself was necessary to identify the most appropriate mitigations based on the varied capabilities brought to bear, data targets of interest to the adversary, and consequences of a security incident. Figure 4 lays out this approach.

As with the prior section, some risks to data components of the AI supply chain are unlikely to be adequately addressed with existing security controls and risk mitigations. Poisoning of AI models may already require unique or relatively unique security requirements, such as filtering mechanisms to screen for poisoned data once a dataset is scraped from the internet or otherwise assembled. That is especially the case—applying this framework—when dealing with threat actors that are well-resourced, sophisticated, and persistent, such as the Chinese government. The amount of resources potentially put into attempting to poison specific datasets may require enhanced planning for the threat at hand and the consequences of the threat unfolding.

Similarly, a sophisticated threat actor may have the capability and time to focus not just on poisoning a training dataset broadly (as discussed in the first approach subsection), but on creating what some call a neural backdoor: tampering with training data to embed a vulnerability in a deep neural network, so that the trained model does not behave erroneously or harmfully in response to standard events, but hides its learned, malicious behavior until it encounters a highly specific trigger.[46] Ongoing research looks at how to tailor protections to training data under very specific assumptions about the bad actor's approach;[47] hence, a threat actor framework may provide more useful information to an organization attempting to defend against sophisticated attempts at neural backdoors. (Like with defending against poisoning attempts, encryption measures or access controls on training data at rest, in motion, or in processing are not going to mitigate these kinds of highly specific risk scenarios.) Still, more research is needed to understand generalized defenses against attempts to backdoor neural networks—including in areas that get relatively less attention than others (i.e., images getting more attention than video).[48] More advanced mechanisms to filter training data are one promising set of approaches to address this type of risk.[49]

Recent work shows that bad actors can tamper not just with training data in the AI supply chain, to create de facto behavioral backdoors in neural networks, but can do so by manipulating model architectures in the AI supply chain as well.[50] Again, taking a threat actor-focused view of the risks enables policymakers and organizational security experts to game out the risk scenarios—and how exactly attempts at architectural backdoors might unfold, offering insight hints on how to plan for and potentially prevent them in advance.

45. See also some of the security controls and risk mitigations laid out in: Sella Nevo et al., *Securing AI Model Weights*.
46. See: Hossein Souri et al., "Sleeper Agent: Scalable Hidden Trigger Backdoors for Neural Networks Trained from Scratch," arXiv, June 16, 2021, https://arxiv.org/abs/2106.08970.
47. See: Wei Guo, Benedetta Tondi, and Mauro Barni, "An Overview of Backdoor Attacks Against Deep Neural Networks and Possible Defences," arXiv, November 16, 2021, https://arxiv.org/abs/2111.08429.
48. Guo, Tondi, and Barni, "An Overview of Backdoor Attacks, 20.
49. Anqing Zhang et al., "Defending Against Backdoor Attack on Deep Neural Networks Based on Multi-Scale Inactivation," *Information Sciences* 690, no. 121562 (February 2025), https://www.sciencedirect.com/science/article/abs/pii/S0020025524014762.
50. Harry Langford et al., "Architectural Neural Backdoors from First Principles," arXiv, February 10, 2024, https://arxiv.org/abs/2402.06957.

Fig. 4: Approach two illustrated

Approach Two Question: What are the threats, vulnerabilities, and consequences of security risks to specific components?	Risks to Data in the AI Supply Chain from this Threat Actor	Potential Mitigations
Identify threat actor and identify specific data component in the AI supply chain or source entity at risk.	Spell out threat actor's intentions and capabilities, the vulnerabilities in the target data component or source entity, and the potential outcomes from threat actor success. Threats could span nation-states, competitor companies and universities in nation-states known for intellectual property theft, cybercriminals, unwitting insiders, and unwitting users. Vulnerabilities could span unencrypted data, lack of access controls on networks and data, poorly secured APIs and SDKs, untrained employees and vendors, and code flaws in third-party cloud systems. Consequences could span theft, leakage, interception, and poisoning of data as well as supply chain infiltration.	Prioritize risk mitigations based on most important risks to address; for some organizations, that may be risks with low likelihood but high consequence, while for others it might be the opposite (or something else entirely). These factors of which threats to prioritize could include: • Likelihood of threat scenario manifesting. • Consequences of threat scenario manifesting. • One-time cost of mitigating threat. • Ongoing cost of mitigating threat. • Speed of threat actor evolution and adaptation. • Feasibility of patching relevant vulnerabilities.

Approach three: Map suppliers to the supply chain

Third, the large number of suppliers of data in the AI supply chain means that security risks can come from suppliers themselves. These risks can take at least two forms: actors within the AI supply chain—such as groups of researchers uploading training data or finished models to websites—deliberately poisoning data or inserting malicious code to compromise others who use those components downstream; and actors within the AI supply chain having poor security practices that enable security risks to spill over to others. The latter of these risk categories could range from an AI service provider pushing API code to hundreds of customers, while not requiring employees to use multifactor authentication, to a volunteer research group unknowingly scraping inaccurate websites to build a flawed training dataset for a chatbot intended for security questions and login verification. Unlike focusing on the current state of a data component of the AI supply chain or the threat actor targeting a specific data component, this potential approach focuses on data security across an organization's AI supply chain and relevant suppliers.

Developers, users, maintainers, governors, and securers of AI technologies can draw on existing supply chain security best practices used in everything from export control regulations to compliance with the EU's General Data Protection Regulation (GDPR). These broadly fall under the bucket of "Know Your Supplier." First is knowing one's suppliers to identify ownership, country of incorporation, and any other signals of potential illegality (e.g., a criminal front set up to scam others) or potential susceptibility to nation-state influence (e.g., ownership by a foreign government). Organizations such as government agencies, companies, and universities can look up the suppliers of their training and testing datasets, ensure the software engineers hired to build their APIs and SDKs are reputable, and check with AI developers about the sources of their training data. (As discussed below, the last of these may be practically difficult but is worth mentioning as a potential security practice.) This could help to identify risks such as unknowingly sourcing data or models from a Chinese military-linked university or hiring freelance data labelers previously involved with illicit hacking.

Second, as compelled by vendor security requirements under GDPR,[51] is ensuring an organization's vendors and supply chain do not have weak security that voids the organization's security efforts. This is likewise a concern for AI-dependent organizations that are typically dependent on a highly complex, often globalized, and frequently shifting data supply chain. AI developers, users, maintainers, governors, and securers can therefore implement contractual requirements for data security wherever possible. They can ensure vendors and customers receive adequate training on how to handle different data components that the organization might pass their way. And they can conduct or request independent audits from cloud companies deploying their AI models, data cleaning firms formatting their unstructured training data, and other entities in their supply chains. There are plentiful resources to draw on for such efforts. Financial sector know-your-customer guidelines,[52] US government advisories on supply chain vetting,[53] and cybersecurity and insurance readings on mitigating third-party cybersecurity risks,[54] among others, can help give AI organizations tools to understand the data-involved actors in their AI supply chains and what risks might result. Such measures can help developers, users, and others to avoid relying on a cloud vendor whose security posture is wholly insufficient to match up against a threat actor the organization has worries about, or help them steer away from hiring API developers at a software support firm that has suffered repeated, simple security breaches. As some have suggested, this could include asking vendors and others in an organization's supply chain for an AI bill of materials, or "AI-BOM" (modeled after a software bill of materials, or SBOM).[55]

Seriously complicating many of these efforts, of course, is just how much of the data in the AI supply chain either touches highly opaque corporate components—or depends on datasets uploaded to freely available websites with often unclear chains of custody and potentially obscured origins. Again, however, the solution is not to immediately treat these situations as unique without examining the applicability of existing best practices to specific risks. Some of the same could apply to vendors from which companies buy their software, which may not wish to make source code available to customers or to provide potential buyers with comprehensive lists of all the contractors, vendors, and dependent software packages on which their products and services were built. Much of the same also pertains to open-source software, too, where companies and developers must come up with ways to manage the risks that arise from a subset of less-well-maintained software or from otherwise well-maintained software that a threat actor compromises.[56] In both cases, companies can use audits, continuous monitoring, and other measures to still mitigate risk from complex supply chains.

Just as heavily software-dependent companies should conduct due diligence on and assess their prospective vendors' cybersecurity (and their vendors' cybersecurity, and so on) to avoid major security lapses and vulnerabilities, AI organizations can do some degree of risk assessment and mitigation for the data suppliers in the AI supply chain (Figure 5).

These supply chain security measures can be applied to AI-specific data risks, too, where existing best practices fail to properly secure data components of the AI supply chain. Companies can require that new partners and vendors attest to the measures they take to mitigate risks of poisoning of training data, such as by carrying out appropriate data filtering.[57] Universities could evaluate their data dependencies in their AI supply chain and, in doing so, catalogue instances where organizations do not mention poisoning in their data security

51. See: Article 32, GDPR (General Data Protection Regulation): Security of Processing, Intersoft Consulting, accessed September 4, 2025, https://gdpr-info.eu/art-32-gdpr/.
52. See: Gus Tomlinson, "KYC Process: Ask the Right Questions," GBG, accessed June 27, 2025, https://www.gbg.com/en/blog/kyc-process-ask-the-right-questions/.
53. See: "Guidance on End-User and End-Use Controls and U.S. Person Controls," US Department of Commerce, Bureau of Industry and Security, accessed June 27, 2025, https://www.bis.gov/licensing/guidance-on-end-user-and-end-use-controls-and-us-person-controls#"KnowYourCustomerGuidance"andRedFlags.
54. See: "Best Practices in Cyber Supply Chain Risk Management," US National Institute of Standards and Technology, accessed June 27, 2025, https://csrc.nist.gov/CSRC/media/Projects/Supply-Chain-Risk-Management/documents/briefings/Workshop-Brief-on-Cyber-SCRM-Vendor-Selection-and-Management.pdf; "Third-Party Cyber Risks Impact All Organizations," Marsh, April 22, 2025, https://www.marsh.com/en/services/cyber-risk/insights/defining-uncovering-cyber-risks-digital-supply-chain.html.
55. Shaked Rotlevi, "AI-BOM: Building an AI Bill of Materials," WIZ, July 20, 2025, https://www.wiz.io/academy/ai-bom-ai-bill-of-materials.
56. See: Andy Greenberg and Matt Burgess, "The Mystery of 'Jia Tan,' the XZ Backdoor Mastermind," *WIRED*, April 3, 2024, https://www.wired.com/story/jia-tan-xz-backdoor/.
57. In addition to the above discussion of poisoning, see: Nicholas Carlini et al., "Poisoning Web-Scale Training Datasets is Practical," arXiv, February 20, 2023, https://arxiv.org/abs/2302.10149.

plans. Government agencies could ensure that any company pitching them on a contract for an externally hosted or already trained, to-be-internally-migrated AI model spell out the steps they have taken to test for and mitigate potential threats of neural backdoors (either through training data or model architectures)—which may be exactly the kind of flaw a nation-state would want moved down the supply chain into a government computer system. Here, the concept of focusing on the supply chain itself can be effective in helping shield against both widely held and AI-unique risks to data components across the AI supply chain.

Fig. 5: Approach three illustrated

Approach Three Question: Who are the suppliers, and what are their security controls—or risks?	Risks to Data in the AI Supply Chain from this Supplier	Potential Mitigations
Who is the supplier in question? What is their role in the data components of the AI supply chain (e.g., from providing public-facing cloud storage for other entities' uploaded testing datasets to creating open-source models and releasing the weights)? What are their security controls or risks?	Could span deliberately poisoning data, deliberately inserting malicious code, unintentionally sourcing poisoned data, and having poor security that impacts partners, vendors, customers, and other third parties elsewhere with a touchpoint to data components of the AI supply chain. Evaluations should look at both indicators of potential maliciousness or susceptibility to malicious influence (e.g., ownership, country of incorporation) as well as their security controls (e.g., via independent audits).	• Know-your-customer questions. • Due diligence investigations of downstream suppliers and their partners. • Standard security terms and conditions for all contracts. • Test-and-assessment period for all new suppliers. • Quarterly reviews of suppliers for risk changes. • Mentoring and training programs for certain vendors and suppliers on risks to data in the AI supply chain. • Policy requiring the organization to manually approve exceptions to security guidelines, evaluated against potential business and risk impacts.[I]

I. Drawing in part on US National Institute of Standards and Technology "Best Practices in Cyber Supply Chain Risk Management."

Conclusion and Recommendations

The fact is governments, companies, universities, individuals, and others pursuing AI R&D will not stop collecting, labeling, disseminating, using, and producing tremendous volumes of data. Therefore, security of the data in the AI supply chain remains evermore paramount to mitigating risks of breaches and data interception, shielding data and resulting AI technologies from theft (including by competitors), enhancing protections for individuals' privacy, bolstering public trust, limiting organizations' liability risk, and strengthening US national security.

Breaking down the seven data components of the AI supply chain can enable developers, users, maintainers, governors, and securers of AI technologies to understand the data components they depend upon, how they interact with and relate to one another, and the varied sources and entangled suppliers. It can then empower organizations to take one of many different existing data security and supply chain security approaches—including data at rest vs. in motion vs. in processing, threat actor risk, and supply chain due diligence and risk management—to map their concerns about data in the AI supply chain to specific, established best practices. More broadly, however, the concept of data in the AI supply chain promises something else for policymakers: the ability to see the whole data supply chain picture at once, leading to more cohesive policymaking.

This paper makes the following three recommendations:

1. **Developers, users, maintainers, governors, and securers of AI technologies should map the data components of the AI supply chain to existing cybersecurity best practices—and use that mapping to identify where existing best practices fall short for AI-specific risks to the data components of the AI supply chain.** In the former case, they should use the framework of data at rest vs. in motion vs. in processing and the framework of analyzing threat actor capabilities to pair encryption, access controls, offline storage, and other measures (e.g., NIST SP 800-53, ISO/IEC 27001:2022) against specific data components in the AI supply chain depending on each data component's current state, the threat actor(s) pursuing it, and the traditional IT security controls the organization already has in place. In the latter case, developers, users, maintainers, governors, and securers of AI technologies should recognize how existing best practices will inadequately prevent the poisoning of AI training data and the insertion of behavioral backdoors into neural networks by manipulating a training dataset or a model architecture. They should instead look to emerging research on how to best evaluate training data to filter out poisoned data examples and how to robustly test network behavior and architectures to mitigate the risk of a bad actor inserting a neural backdoor, which they can activate after model deployment.
2. **Developers, users, maintainers, governors, and securers of AI technologies should "Know Your Supplier," using the supply chain-focused approach to mitigate both AI-specific and non-AI-specific risks to the data components of the AI supply chain.** Those sourcing data for AI systems—whether training data, APIs, SDKs, or any of the other data supply chain components—should implement best practices and due diligence measures to ensure they understand the entities sourcing or behind the sources of different components. For example, if a university website has a public repository of testing datasets for image recognition, language translation, or autonomous vehicle sensing, did the university internally develop those testing datasets, or is it hosting those testing datasets on behalf of third parties? Can third parties upload whatever data they want to the public university website? What are the downstream controls on which entities can add data to the university repository—data which companies and other universities then download and use as part of their AI supply chains? Much like a company should want to understand the origins of a piece of software before installing it on the network (e.g., is it open-source, provided by a company, if so which company in which country, etc.), an organization accessing testing data to measure an AI model or using any other data component of the AI supply chain should understand the underlying source within the supply chain. Best practices in know-your-customer due diligence, such as in the financial sector and export control space, and in the supply chain risk management space, such as from cybersecurity and insurance companies, can provide AI-dependent organizations with checklists and other tools to make this happen. Avoiding entities potentially subject to adversarial foreign nation-state influence, data suppliers not sufficiently vetting the data they upload, and so forth will help developers, users, maintainers, governors, and securers of AI technologies to bring established security controls to the data in the AI supply chain itself. In the case of both non-AI-specific and AI-specific risks to data, organizations can and should use this supply chain due diligence approach to ensure their vendors, customers, and other partners are implementing the right controls to protect against risks of model weight theft, training data manipulation, neural network backdooring through model architecture manipulation, and everything in between—drawing on the two categories of mitigations implemented as part of the first recommendation.

3. **Policymakers should widen their lens on AI data to encompass all data components of the AI supply chain. This includes assessing whether sufficient attention is given to the diversity of data use cases that need protection (e.g., not just training data for chatbots but for transportation safety or drug discovery) and whether they have mapped existing security best practices to non-AI-specific and AI-specific risks.** As multiple successive US administrations explore how they want to approach the R&D and governance of AI technologies, data continues to be a persistent focus of discussion. It comes up in everything from copyright litigation to national security strategy debates. The United States' previous policy focus on training data quantity, and little else, has already prompted policymakers to avoid discussing comprehensive data privacy and security measures, which now—in light of Chinese AI advancements and concern about AI model weight dissemination—are suddenly more relevant. To avoid these cycles in the future, where policy overfocuses on one AI data element when in fact many are relevant simultaneously, policymakers should take a comprehensive view of the data components of the AI supply chain. The framework offered in this paper, spanning seven data components, is one potential guide—though again, policymakers need not stick to necessarily one framework. What is most critical to avoid is developing data security policies that protect some data components of the AI supply chain (e.g., training data) while leaving others highly exposed (e.g., APIs). An expanded view of the different data components, the components' interaction, and the often multiple and shifting roles of suppliers should help inform better federal legislation, regulation, policy, and strategy—as well as engagements with other countries and US states. Right now, organizations such as the Congressional commerce committees, the Commerce Department (including because it implements export controls and the Information and Communications Services and Technologies supply chain program), the Defense Department (with all its current AI procurement), and the Federal Trade Commission (with responsibility for enforcing against unfair and deceptive business practices) should stress-test their assumptions about how to best protect AI data, and whether existing best practices achieve desired security outcomes, against this data component framework. This requires asking at least two questions. Do their existing security, governance, or regulatory approaches—e.g., in the security requirements used in Defense Department AI procurement, in how the Federal Trade Commission thinks about enforcing best practices for AI data security—apply well to a diversity of data use cases that need protection, such as with testing datasets for self-driving vehicle safety or training datasets for cutting-edge drug discovery? List out the use cases beyond chatbots that are not top of mind but are highly relevant from a security perspective, from defense to shipping and logistics to healthcare. And second, are they parsing out which risks they have concerns about, vis-à-vis AI-related data, that are specific to AI versus risks to data in general? For both categories, consider how the framework and some of the security mitigations cited in this report (e.g., NIST, ISO, new research on detecting neural backdoors, etc.) can serve as best practices to improve outcomes.

The more governments, companies, civil society groups, individuals, and others move AI technologies into areas ranging from e-commerce, social media, and business administration to manufacturing, healthcare, transportation, and defense, the more important it becomes to secure all data related to AI technologies. A complex set of data components in the AI supply chain demands policy and security practices that account for the entire supply chain and all its complexity at once, or at least through piecemeal efforts that add up to the whole—making existing data security and supply chain security best practices, paired with newer responses to AI-specific data risks, an optimal place to start.

About the author

Justin Sherman is a nonresident senior fellow at the Atlantic Council's Cyber Statecraft Initiative. He is also the founder and CEO of Global Cyber Strategies, a Washington, DC-based research and advisory firm; an adjunct professor at Georgetown University's School of Foreign Service; a contributing editor at Lawfare; and a columnist at Barron's.

Acknowledgments

The author thanks Trey Herr, Nitansha Bansal, Kemba Walden, Devin Lynch, Harriet Farlow, Ben Goldsmith, Kenton Thibaut, and other individuals who could not be thanked by name for their comments on earlier drafts of this report, as well as all the individuals who participated in the background and Chatham House Rule discussions about issues related to data, AI applications, and the concept of an AI supply chain.

About the center

The **Cyber Statecraft Initiative** works at the nexus of geopolitics, technology, and security to craft strategies to help shape the conduct of statecraft and to better inform and secure users. This work extends through the competition of state and non-state actors, the security of the internet and computing systems, the safety of operational technology and physical systems, and the communities of cyberspace. The Initiative convenes a diverse network of passionate and knowledgeable contributors, bridging the gap among technical, policy, and user communities.