Countering AI Chip Smuggling Has Become a National Security Priority

An Updated Playbook for Preventing AI Chip Smuggling to the PRC

Erich Grunewald & Tim Fist

Center for a New American Security

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

About the Authors

Erich Grunewald is a researcher on the compute policy team at the Institute for AI Policy and Strategy (IAPS). He previously worked as a software engineer and earned a BSc in computer engineering and an MSc in interaction design from Chalmers University of Technology.

Tim Fist is an adjunct senior fellow with the Technology and National Security Program at the Center for a New American Security (CNAS) and the director of emerging technology policy at the Institute for Progress. Fist has an engineering background, having previously worked for five years on both AI chips and software. Fist holds a BA (honors) in aerospace engineering and a BA in political science from Monash University, and is a DPhil candidate in Engineering Science at the University of Oxford.

About the Technology & National Security Program

The CNAS Technology and National Security Program produces cutting-edge research and recommendations to help U.S. and allied policymakers responsibly win and manage the great power competition with China over critical and emerging technologies. The escalating U.S.-China AI, biotechnologies, next-generation information and communications technologies, digital infrastructure, and quantum information sciences will have far-reaching implications for U.S. foreign policy and national and economic security. The Technology and National Security Program focuses on high-impact technology areas with in-depth, evidence-based analysis to assess U.S. leadership vis-à-vis China, anticipate technology-related risks to security and democratic values, and outline bold but actionable steps for policymakers to lead the way in responsible technology development, adoption, and governance. A key focus of the Tech Program is to bring together the technology and policy communities to better understand these challenges and together develop solutions.

About the Artificial Intelligence Safety & Stability Project

The CNAS AI Safety & Stability Project is a multiyear, multiprogram effort that addresses the established and emerging risks associated with artificial intelligence. Its work is focused on anticipating and mitigating catastrophic AI failures, improving the U.S. Department of Defense's processes for AI testing and evaluation, understanding and shaping opportunities for compute governance, understanding Chinese decision-making on AI and stability, and understanding Russian decision-making on AI and stability.

Acknowledgements

The authors would like to acknowledge the CNAS Publications and Communications teams for their support, design, and editing. The authors are also grateful to Vivek Chilukuri, Paul Scharre, and others who provided valuable feedback and insights throughout the development of this report. This project is made possible with the generous support of Open Philanthropy.

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

activities will be limited to bona fide scholastic, academic, and research-related activities, consistent with applicable federal law. The Center publicly acknowledges on its website annually all donors who contribute.

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Table of Contents

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Executive Summary

Based on the available evidence, artificial intelligence (AI) chip smuggling has likely been occurring at a scale that significantly undermines U.S. attempts to restrict the People's Republic of China's (PRC's) access to advanced AI. This is indicated by four lines of argument:

1. **Smuggling should be expected based on historical precedent**. The PRC has a long history of smuggling U.S. technology despite export restrictions, which has rarely resulted in criminal or civil penalties.[1]

2. **Smuggling of U.S. AI chips is highly incentivized by their superior performance, higher supply, and more mature software ecosystem relative to chips legally available to Chinese AI labs**.[2] Based on publicly available data, of the 22 notable models that had been developed exclusively in the PRC by 2025, only two were trained with Chinese chips.[3]

3. **Six news outlets have independently reported evidence of large-scale AI chip smuggling, totaling tens to hundreds of thousands of chips smuggled in 2024**. One smuggler reportedly handled an order of for servers containing 2,400 NVIDIA H100s—worth $120 million—to a customer in the PRC.[4] Another facilitated an order worth $103 million.[5] Singapore authorities arrested three individuals suspected of diverting AI servers worth $390 million.[6] Within this reporting, multiple chip resellers and start-ups in the PRC have claimed that gaining access to export-controlled AI chips is straightforward, with one Chinese start-up founder estimating in 2024 that there were more than 100,000 NVIDIA H100s in the PRC.[7] Most of the Chinese chip sellers interviewed in these reports confirm that they work with multiple distributors, use shell companies based overseas, and employ simple tactics to avoid detection, such as relabeling shipments as tea or toys.[8]

4. **There are many online listings for export-controlled AI chips available for purchase in the PRC.** The authors conducted a non-exhaustive search of three Chinese online marketplaces and found 132 domestic listings for export-controlled chips, along with many photos of supposedly smuggled goods.[9] Where sellers provided information on stock, the average quantity of export-controlled graphics processing units (GPUs) per listing was around 1,200 for GPU server listings and 400 for GPU card listings. Though this data is patchy and likely unreliable, the total stock implied is around 100,000 H100 GPUs, as of December 2024.

Across these data sources, there is much uncertainty in either direction. Using publicly available data from this reporting, while accounting for uncertainty about its veracity, the authors estimate the total scale of AI chip smuggling to the PRC in 2024 could have ranged anywhere from 10,000 to several hundred thousand chips, with a median estimate of around 140,000. This estimate is based on a probabilistic model of chip smuggling as an extrapolation of known cases. It should be understood as a high-level distribution of possible outcomes, rather than a summation of confirmed smuggling cases.[10]

If the true number is within this distribution, **smuggled chips could make up a significant portion of AI compute acquired by the PRC in 2024— between 1 and 30 percent of its inference compute capacity (median 6 percent) or between 1 and 40 percent of its training compute capacity (median 10 percent).** Other sources of PRC AI compute acquisition considered are Huawei Ascend chips fabricated by China's Semiconductor Manufacturing International Corporation (SMIC) and Taiwan Semiconductor Manufacturing Company (TSMC), and NVIDIA H20 chips legally exported into the PRC. Because the export of the H20 was restricted in April 2025, and TSMC is under investigation for its role

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

in fabricating chips for Huawei, smuggled chips may make up a much larger portion of AI compute acquired by the PRC in 2025 and beyond.[11]

Given the potential scale of AI chip smuggling to the PRC, policymakers should prioritize gathering information to better understand its true extent. Large-scale chip smuggling and uncertainty about its true extent exist partly because the Bureau of Industry and Security (BIS)—the part of government tasked with administering and enforcing AI chip export controls—is under-resourced. The BIS's budget for enforcement has decreased in real terms over time, even after it was tasked with far more responsibilities in 2022 following Russia's invasion of Ukraine and the introduction of a wide range of new export controls on AI chips and related tooling. The profits likely netted by chip smugglers from just three reported smuggling cases from *The New York Times* and *The Information* are more than double the BIS's annual budget for export control enforcement.[12]

AI chip smuggling is a recent problem, and countereforts remain limited. Moreover, AI chip companies typically lack the deep experience in exporting sensitive goods that exists, for example, in the defense industry. As a result, there are many promising steps that government and industry could take to collect better information and tackle smuggling. The authors offer six recommendations:

1. AI chip companies and their distributors should leverage their technical capacities and resources to significantly strengthen their **due diligence processes,** make similar requirements of any distribution firms to which they sell, and report to the BIS on specific measures taken.

2. The BIS could use its existing authorities to **create a new notification requirement for controlled AI chip exports, reexports, and ownership transfers.** This would enable much better tracking of controlled chips exported outside the United States, allowing for more efficient enforcement.

3. AI chip designers should implement software-based **location verification** features, allowing the owner of a controlled chip to prove that the chip is not located in the PRC. This would both provide valuable data on the scale of chip smuggling and provide a useful mechanism for enforcement, enabling knowledge of which specific devices have been smuggled, and knowledge of which exporters and downstream entities were responsible. BIS could incentivize the development of these features through the normal export licensing process today, by requiring chip owners outside the United States to periodically prove that their chips are not in the PRC as a condition for importing chips.

4. The Department of Commerce, in **coordination with the intelligence community (IC)**, could establish processes to generate information on the scale, locations, and perpetrators of AI chip smuggling, and share it with the BIS to aid enforcement, inform policy, and share leads with relevant industry actors. This could be facilitated through the Director for National intelligence (DNI) updating the National Intelligence Priorities Framework to make intelligence collection related to export control enforcement a top priority, and the DNI and Secretary of Commerce establishing a joint analysis team (for example, in the Export Control Enforcement Center) to facilitate BIS-IC coordination.

5. Congress could authorize a **whistleblower incentive program and qui tam lawsuits** (lawsuits that allow a private person to prosecute on behalf of the government) to incentivize reports of possible export violations. These two measures could significantly boost the BIS's export enforcement while paying for themselves in revenue generated through additional penalties. A whistleblower program could be modeled on the highly successful Securities and Exchange Commission (SEC) Whistleblower Program, which tackles federal securities laws violations and has aided the collection of $7 billion to $22 billion in penalties since 2011.[13] In April 2025, a

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

bipartisan bill was introduced in the Senate to implement such a program.[14] Authorizing qui tam lawsuits against those who violate the export rules would additionally allow individuals to sue a violator and collect a portion of the resulting penalty, modeled on a similar law in the False Claims Act.

6. Finally, Congress could grant **the White House's requested budget of $313 million for the BIS** for fiscal year 2026. This 64 percent increase over the BIS's current budget of $191 million would help the BIS modernize its tooling and improve operational capacity. In addition to strengthening national security, funding to improve the enforcement of the AI chip controls would likely pay for itself through increased collection of fines on violators of the export rules. The multiple documented instances of AI chip smuggling from 2024 alone could each have resulted in penalties of over $200 million, more than the BIS's annual budget.

As controlling access to advanced AI capabilities becomes an increasing national security imperative, U.S. policymakers should prioritize actions to collect more reliable information on the extent of AI chip smuggling and close gaps wherever possible.

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Collecting Detailed Data on AI Chip Smuggling Is Difficult Under Existing Laws

Under the Export Control Reform Act (ECRA), the legal authority underpinning AI chip export controls, companies are incentivized to perform only the minimum supply chain due diligence required to avoid legal exposure, as an exporter is only subject to criminal penalty if it willfully commits a violation of the rules.[15] The Bureau of Industry and Security (BIS), the agency tasked with overseas export control enforcement for AI chips, can impose civil penalties without evidence that a company had positive "knowledge" of a violation, but it is reportedly reluctant to do so. This led a December 2024 report by the U.S. Senate's Permanent Subcommittee on Investigations Majority Staff to recommend that the BIS should "charge companies with 'knowing' violations when they fail to sufficiently investigate red flags or other strong indicia of potential diversion and violations occur."[16]

As a result, despite their large budgets and central position in the AI chip supply chain—which put them in a favorable position to collect *detailed data* on potential cases of chip smuggling—U.S. chip firms often remain unaware of the ultimate destination of their chips. For the purposes of this paper, "detailed data" refers to specific information about a suspected smuggling case that would directly lead to enforcement action, including the business identity of the exporter, the legal identities of persons involved in the transaction, and the serial numbers of the smuggled devices.

Another option for detailed data collection is via the U.S. government. However, the BIS has little spare capacity to proactively collect evidence about cases of smuggling. The BIS's budget for enforcement has decreased in real terms over the past few years despite a massive increase in its responsibilities.[17] The BIS has just six export control officers across all of East, South, and Southeast Asia (across China, India, Singapore, Taiwan, and Hong Kong), and reportedly only two in-house Mandarin speakers as of May 2024.[18] The BIS is also reportedly limited by Chinese law and a U.S.-China agreement from conducting investigations inside the PRC.[19] If smuggled AI chips are being used to support state-backed military and intelligence projects in the PRC—the intended targets of the U.S. chip export controls—this would be difficult for the BIS could uncover this with its current resources.

A final option for detailed data collection is open-source intelligence from nongovernmental research organizations. This approach has successfully uncovered large-scale chip smuggling to support Russia's invasion of Ukraine. However, this intelligence gathering relied on the collection of Russian military hardware by Ukrainian soldiers; there is no equivalent in the PRC incentivized to uncover evidence of smuggling.[20] Open-source intelligence gathering is further limited by a lack of access to private business data, and collecting data on the ground in the PRC, especially in coordination with Chinese citizens, risks attracting retribution from the Chinese Communist Party's national security dragnet.[21]

Taken together, these considerations highlight the difficulty of enforcing export controls on AI chips. They also explain why the lack of detailed data on large-scale smuggling of AI chips to the PRC does not necessarily prove its absence. Assessing the scale of AI chip smuggling to the PRC requires looking at the more scattered evidence that does exist.

Large Volumes of AI Chips Are Likely Being Smuggled into the PRC

In an October 2023 paper, the authors estimated the potential scale of AI chip smuggling to the PRC in the event of a concerted smuggling effort, estimating a 50 percent chance that more than 12,500 chips would be smuggled in 2024.[22] Since then, multiple sources of public reporting have indicated that a concerted smuggling effort is indeed underway. This reporting, combined with other sources of

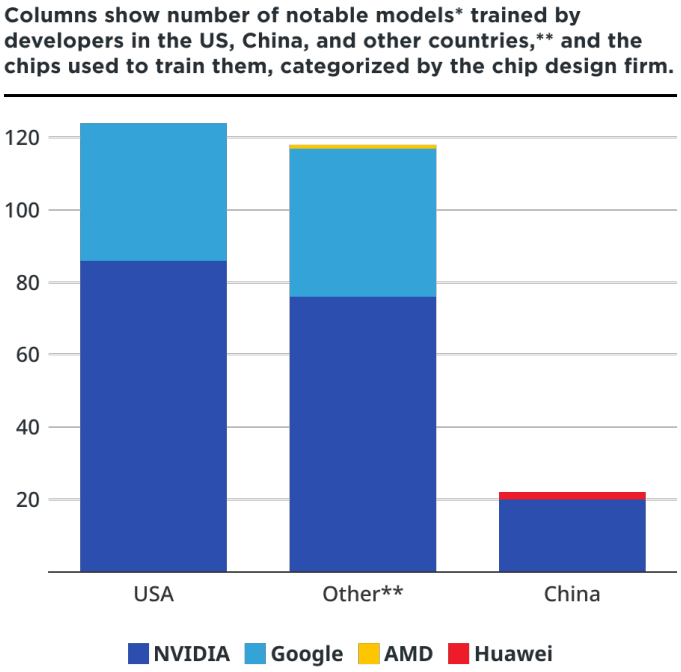
Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

evidence, strongly suggests that at least 100,000 AI chips were smuggled into the PRC in 2024. Four lines of argument support this conclusion:

First, there is a long record of PRC actors smuggling sensitive U.S. technology despite export controls. According to customs data analyzed by *Nikkei Asia*, over $570 million worth of U.S. chips were shipped to Russia from the PRC in 2022 despite sanctions, up from $51 million in 2021.[23] These illicit activities, largely carried out by actors outside U.S. jurisdiction, rarely result in criminal or civil penalties.

Second, Chinese AI companies are highly incentivized to obtain controlled U.S.-designed chips, given their substantially higher performance versus PRC-made alternatives, larger supply, and superior software stacks.[24] Epoch AI's publicly available data confirms the preferences of Chinese developers to use U.S. chips. Of the 22 notable models developed exclusively in the PRC by 2025, only two were trained with Chinese chips.

Figure 1: Which Chips Are AI Developers in Different Countries Using?



Third, several investigative reports by journalists point to large-scale AI chip smuggling into the PRC. In 2023, early reports identified cases of small-scale smuggling (tens of chips at a time), mostly related to NVIDIA A100 GPUs, which the PRC could have stockpiled prior to the October 2022 ban.[25] More recent reports suggest that a much larger smuggling ecosystem is now active, with brokers selling hundreds to thousands of AI chips at a time. These reported smuggling cases mostly involve NVIDIA H100 GPUs and servers, which were launched in March 2023 (after export controls came into force). More recent cases have also involved NVIDIA Blackwell GPUs, launched in December 2024.[26] Large-

scale AI chip smuggling of this kind has been independently reported by *The Wall Street Journal*, *The New York Times*, *The Information*, and the *Financial Times*.[27] One smuggler reportedly handled an order of 2,400 cutting-edge NVIDIA H100s, worth $120 million, to a customer in the PRC.[28] Another facilitated an order worth $103 million.[29] Outside of the PRC, *Bloomberg* reported that nearly 9,000 AI chips worth $300 million were smuggled into Russia via an Indian pharmaceutical company over a six month period, despite U.S. sanctions.[30]

According to interviews from these reports, it is accepted among buyers and sellers of AI chips in the PRC that large-scale smuggling of U.S. AI chips is occurring. One Chinese start-up founder estimated that there are more than 100,000 NVIDIA H100s in the PRC.[31] Most of the Chinese chip sellers interviewed in these reports confirm that they work with multiple distributors, use shell companies based overseas, and employ simple tactics to avoid detection, such as relabeling shipments as tea or toys.[32] As Figure 2 shows, reports also suggest that smuggled AI chips are not substantially more expensive than chips bought legally in the United States, and lead times for smuggled chips in 2024 were on the scale of weeks rather than months.

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Figure 2: A Summary of AI Chip Smuggling Cases Uncovered by Journalistic Investigations[33]

Source	Date	Summary	Price
Reuters Josh Ye, David Kirton, and Chen Lin, "Focus: Inside China's Underground Market for High-End Nvidia AI Chips"	Jun '23	• Interviewed 10 controlled chip vendors in China, and 1 AI lab who was procuring controlled chips from Chinese vendors • All 10 vendors were selling A100s, one was selling H100s • Vendors procured products as excess stock from large orders, or by importing from Indian, Taiwanese, Singaporean, and other companies	**A100 card:** $19,150 (1.9x MSRP)
IC Trends 辰壹 (Chen Yi), "美国出口禁令之下，20多万元的'天价芯片'流入黑市" (Under US Export Bans, "Sky-High Chips" Worth Over 200,000 Yuan Flow into Black Markets)	Nov '23	• Interviewed an unspecified number of controlled chip vendors and AI labs, one of which provided a photo of a shipment of Supermicro servers claimed to contain A100s and H100s • Monitored postings and prices for A100s on Chinese online marketplaces • 8 to 14 week lead time for A100s and H100s • Large companies use direct non-public channels for obtaining controlled chips • Smaller companies buy controlled chips on online marketplaces • Some products are actually counterfeit A100s/H100s	**A100 card:** $5,500 to $34,500 (0.5x to 3.4x MSRP) **H100 card:** $44,200 (1.5x MSRP)
Reuters Eduardo Baptista, "China's military and government acquire Nvidia chips despite US ban"	Jan '24	• Found over 100 tender documents where Chinese military organizations, state-run AI labs, and universities have purchased A100s, H100s, and A800s • Each individual purchase was small (<10 GPUs)	**A100 card:** $8,100 (0.8x MSRP)
Wall Street Journal Raffaele Huang, "The Underground Network Sneaking Nvidia Chips into China"	Jul '24	• Identified 70 distributors advertising A100s and H100s online, and communicated with 25 of them. Also interviewed a broker in Singapore who sells to Chinese AI labs • Verified purchases with Chinese buyers who bought from the distributors, and accessed transaction records, customs filings, and photos of chips up for sale • Reviewed procurement documents showing multiple Chinese entities are buying H100s from resellers in China • Vendors have dozens of chips in stock at a time, with larger orders able to be delivered in 1-2 weeks • Chips come in original wholesale packaging	**A100 card:** $22,500 (2.2x MSRP) **H100 card:** $32,400 (0.9x MSRP)
The New York Times Ana Swanson and Claire Fu, "With Smugglers and Front Companies, China Is Skirting American A.I. Bans" Ana Swanson, "Takeaways From Our Investigation Into Banned A.I. Chips in China"	Aug '24	• Interviewed 11 vendors in Shenzhen, China selling A100s and H100s • Identified an additional 100 stores that sell A100s and H100s • One vendor sells 200 to 300 chips at a time • Another vendor recently sold a large batch of 2,000 H100s to a Chinese company for $103 million. The NYT reviewed photos of the shipment, and a text conversation with the supplier • Acquired procurement documents showing over a dozen state-affiliated entities in China purchased controlled chips	**H100 card:** $51,000 (1.5x MSRP) **H100 server:** $380,000 (1.3x MSRP)
The Information Qianer Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry"	Aug '24	• Interviewed eight chip smugglers, NVIDIA employees, suppliers, and distributors, and reviewed procurement documents • One smuggler helped ship 300 H100 servers (2,400 H100 GPUs) to China through a shell company in Malaysia, for $130 million. The smuggler claimed to sell to Chinese state-owned companies • Another smuggler reported obtaining "thousands" of chips from Dell and Supermicro through relationships with those companies' sales representatives, and using shell companies abroad. The Information visited the smuggler's warehouse in China, and viewed 600 H100 servers (4,800 GPUs) from Dell and Supermicro. The servers were sold to Chinese buyers for $230 million • Another smuggler claimed to work with government buyers, smuggling chips for advanced AI data centers used by state-owned enterprises and government departments. The Information viewed associated business records	**H100 server:** $385,000 to $400,000 (1.2x to 1.3x MSRP)

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Fourth, online marketplace listings for export-controlled AI chips available in the PRC imply substantial smuggling operations. The authors conducted nonexhaustive automated searches of three online platforms for listings featuring potentially smuggled chips: Baidu Tieba (a forum often used as an informal marketplace), JD.com (a marketplace often used for electronics, operated by China's second largest e-commerce company), and Taobao (China's largest online shopping platform, owned by Alibaba Group). The authors then manually checked each retrieved listing to determine whether it claimed to sell export-controlled chips.[34] The results are in Figure 3. From this search, the authors found 57 listings for export-controlled servers (devices typically containing eight graphics processing units, or GPUs, for installation in data centers), and 75 listings for export-controlled cards (individual GPUs). On Baidu Tieba, information was sometimes available about the quantity of stock a seller possessed. Expressed as total GPUs, this averaged to around 1,200 GPUs for those selling servers and 400 GPUs for those selling cards. Many listings also contained photos of smuggled products.[35]

Figure 3: Export-Controlled AI Chip Listings on Chinese Online Marketplaces

A subset of local listings in December 2024, on three of the most popular Chinese online marketplaces: JD.com, Baidu, and Taobao.

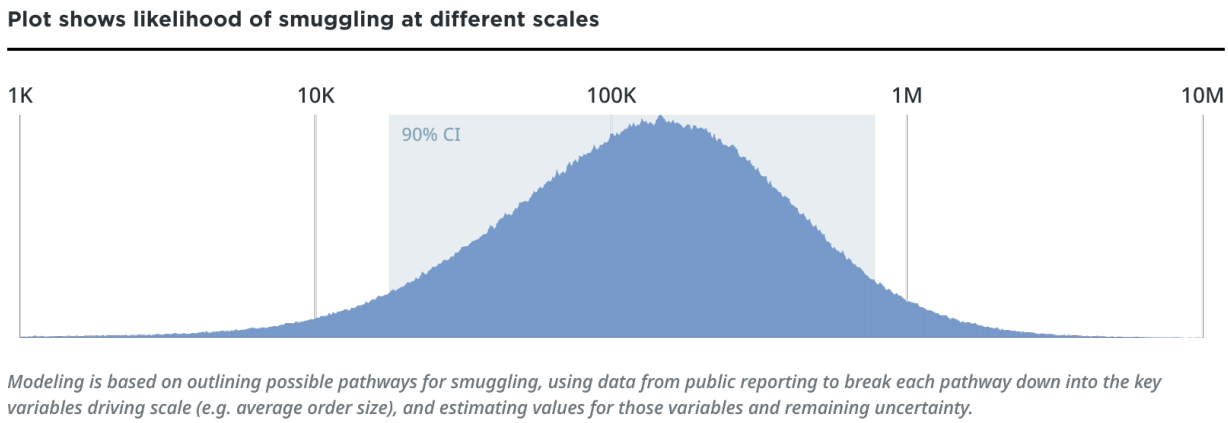
Product	Platform	Number of Listings	Average Stock per Listing (in # GPUs)
H100 server			
	Baidu Tieba	37	1,227
	JD.com	9	
	Taobao	7	
	Total	**53**	
H100 card			
	Baidu Tieba	3	400
	JD.com	54	
	Taobao	18	
	Total	**75**	
H200 server			
	Baidu Tieba	4	1,024
	Total	**4**	

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Despite the ready availability of this data, it is highly unlikely to reflect the true extent of AI chip smuggling. First, it is unlikely that this data would include larger sales, as direct negotiation with suppliers further up the chain likely yields lower prices and more reliable deliveries. Furthermore, it is likely that at least some of these listings are fake or are for counterfeit goods. This data should be taken mainly as evidence for the existence of an AI chip smuggling ecosystem, indicating that smuggled AI chips are readily available in the PRC in informal and secondhand markets.

Taken together, these four lines of argument strongly suggest a sizable pipeline of smuggled chips into the PRC, in turn suggesting that U.S. export controls on AI chips are not currently working as intended. The authors estimate that it is more likely than not that over 100,000 banned AI chips, worth several billion dollars, were likely smuggled into the PRC in 2024. The authors reached this estimate by breaking two separate smuggling pathways down into constituent factors (for example, the number of vendors selling banned chips and the number of chips sold per vendor and year), and anchoring those factors in the aforementioned data from public reporting.[36] The pathways are then weighed by probability and combined to yield an all-things-considered estimate. This estimate is based on probabilistic extrapolations of known reports. In other words, the authors estimate the potential scale of both reported and unreported smuggling. Importantly, the estimates should be understood as a high-level distribution of possible outcomes, rather than a summation of confirmed smuggling cases. The authors offer these estimates to help prioritize policy attention and resources in the absence of better data. Further details about the methodology are included in Appendix A.

Figure 4: Probabilistic Model of AI Chip Smuggling to the PRC in 2024



Modeling is based on outlining possible pathways for smuggling, using data from public reporting to break each pathway down into the key variables driving scale (e.g. average order size), and estimating values for those variables and remaining uncertainty.

If this distribution is accurate, smuggled chips could make up a significant portion of AI compute acquired by the PRC in 2024—between 1 and 30 percent of its inference compute capacity (median 6 percent) or between 1 and 40 percent of its training compute capacity (median 10 percent).[37] This comparison is based on:

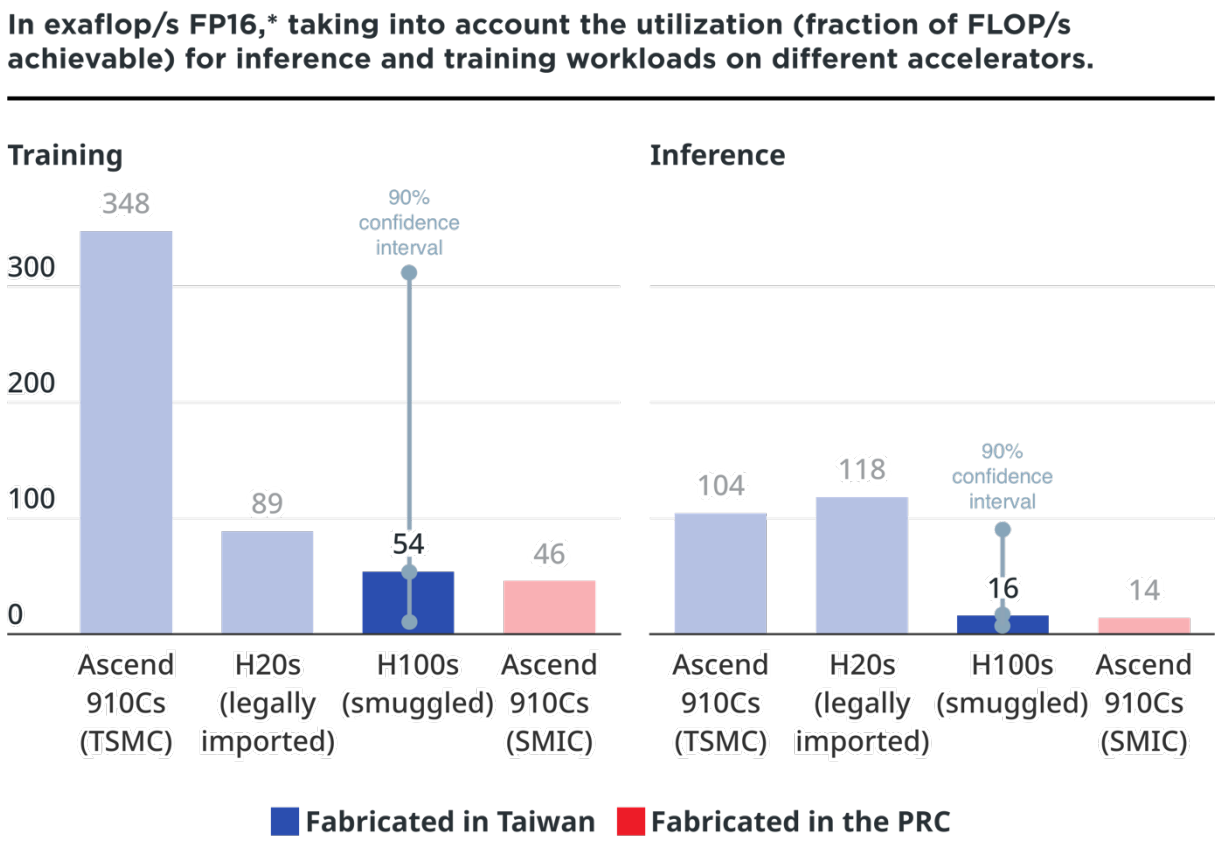
- Tear down analysis of Ascend chips and reporting from Reuters and the Center for Strategic and International Studies, stating that TSMC has fabricated between 2 and 3 million Ascend 910B dies for Huawei, which could be used to produce 750,000 to 1.125 million Ascend 910Cs, after taking into account a 75% success rate during packaging.[38]

- Analysis from the authors on the number of Ascend 910Cs manufactured by SMIC in 2024, based on reports of PRC wafer production capacity, fabrication yields, and the proportion of capacity used for AI chips.[39]

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

- A report from Reuters, claiming that NVIDIA shipped 1 million H20s in 2024.[40]

Figure 5: Estimates of AI Compute Acquired by the PRC in 2024



** One exaflop/s is 10^18 floating point operations per second. For training compute, 1 exaflop/s is around 2,500 H100s, assuming flop/s utilization of 40%. FP16 refers to the precision of the numbers used in the calculations. We specify it to provide a standardized comparison between different products.*

Why Does AI Chip Smuggling Occur?

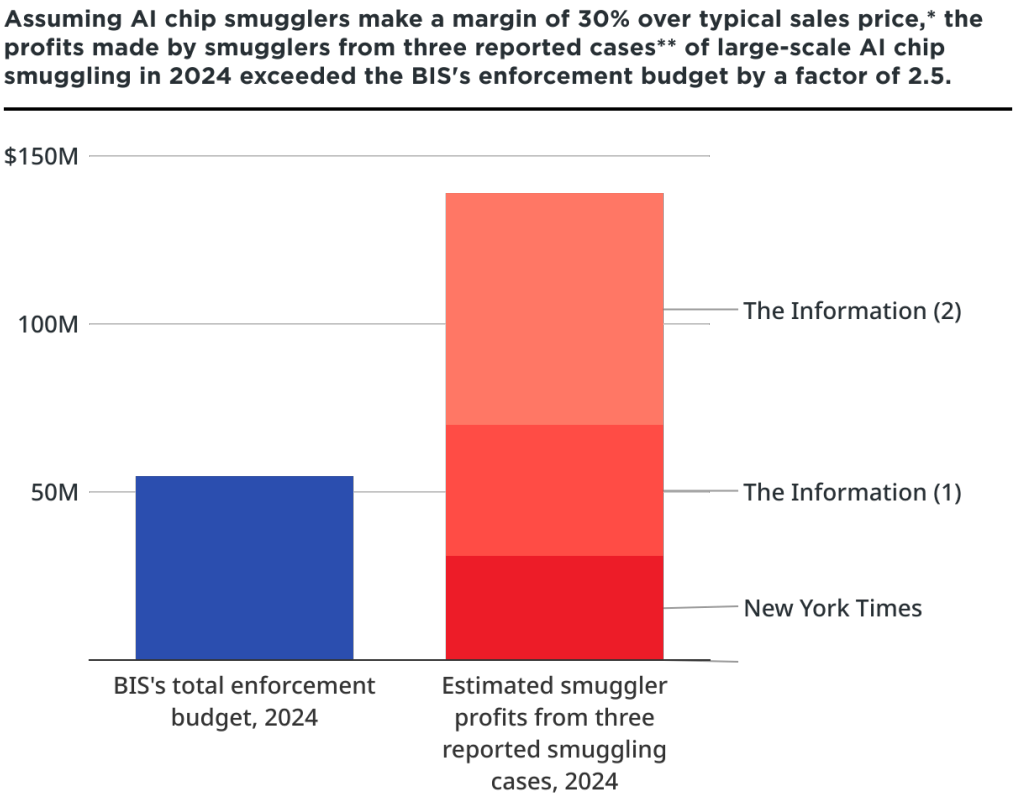
Widespread smuggling of U.S. AI chips occurs because PRC-linked actors are willing to disregard U.S. export law, because BIS's lack of resourcing creates opportunities for smuggling, and because U.S. chips are superior to domestic Chinese alternatives.

PRC-linked actors' willingness to disregard U.S. export law is well-documented.[41] For example, PRC entities have facilitated widespread smuggling of conventional chips to Russia—shipments from Hong Kong and China jumped tenfold after the invasion of Ukraine.[42] PRC firms have also circumvented end-use and end-user restrictions in the PRC: a 2022 analysis by the Center for Security and Emerging

Technology found that PRC military and state-owned defense enterprises were acquiring U.S.-made chips despite end-use and end-user controls.[43]

These PRC successes are enabled by resourcing issues at home. BIS is underresourced and ill-equipped to combat AI chip smuggling. For example, a single U.S. Export Control Officer is responsible for investigating violations, conducting inspections, and managing outreach across all of Southeast Asia and Australasia.[44] Overall, AI chip smugglers can likely spend more than fifteen times as much on smuggling operations as the BIS can spend on enforcing the AI chip controls.[45] This is because the BIS's budget for export enforcement is relatively small ($55 million in 2024), whereas AI chip smugglers likely generate significant profits, based on estimated volumes and reported premiums. Estimated profits from just three reported cases of large-scale AI chip smuggling in 2024 exceeded the BIS's entire enforcement budget for the year.

Figure 6: Resources Available to BIS vs AI Chip Smugglers



** 30% is the average margin over the recommended sales price from reports of AI chip smuggling aggregated by the authors. ** Ana Swanson and Claire Fu, "With Smugglers and Front Companies, China Is Skirting American A.I. Bans," The New York Times, August 4, 2024, https://www.nytimes.com/2024/08/04/technology/china-ai-microchips.html, Qianer Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry," The Information, August 12, 2024, https://www.theinformation.com/articles/nvidia-ai-chip-smuggling-to-china-becomes-an-industry.*

While the BIS can take certain steps immediately to reduce AI chip smuggling, it will ultimately need more resources to enforce the controls on AI chips and related tools effectively.

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Smuggling is additionally incentivized by the wide performance gap between export-controlled U.S. chips and domestically made Chinese chips. Access to large quantities of powerful AI chips will likely remain a key bottleneck in Chinese AI development, as it is for U.S. firms.[46] Before the October 2023 update to the Interim Final Rule published on October 7, 2023, Chinese companies could legally import NVIDIA A800s and H800s, the export control-compliant counterparts to the NVIDIA A100 and H100, respectively. DeepSeek's widely noted V3 and R1 language models, released in December 2024 and January 2025, respectively, were both developed using NVIDIA H800s.[47] These models achieved strong benchmark scores and demonstrated impressive algorithmic advances. Many commentators have thus interpreted DeepSeek's success as a sign that AI chip export controls have failed.[48] This is wrong for two reasons. First, effective AI chip export restrictions have not been in force long enough to have a large effect—the chips that DeepSeek bought before October 2023 are only 10 to 30 percent less computationally performant than their counterparts available outside the PRC, which are still the most widely used AI chips.[49] Second, DeepSeek's models would have been even more capable if they had been trained with better chips.[50] When the founder of DeepSeek was asked about financing plans in September 2024, he responded that money "has never been the problem for us; bans on shipments of advanced chips are the problem."[51]

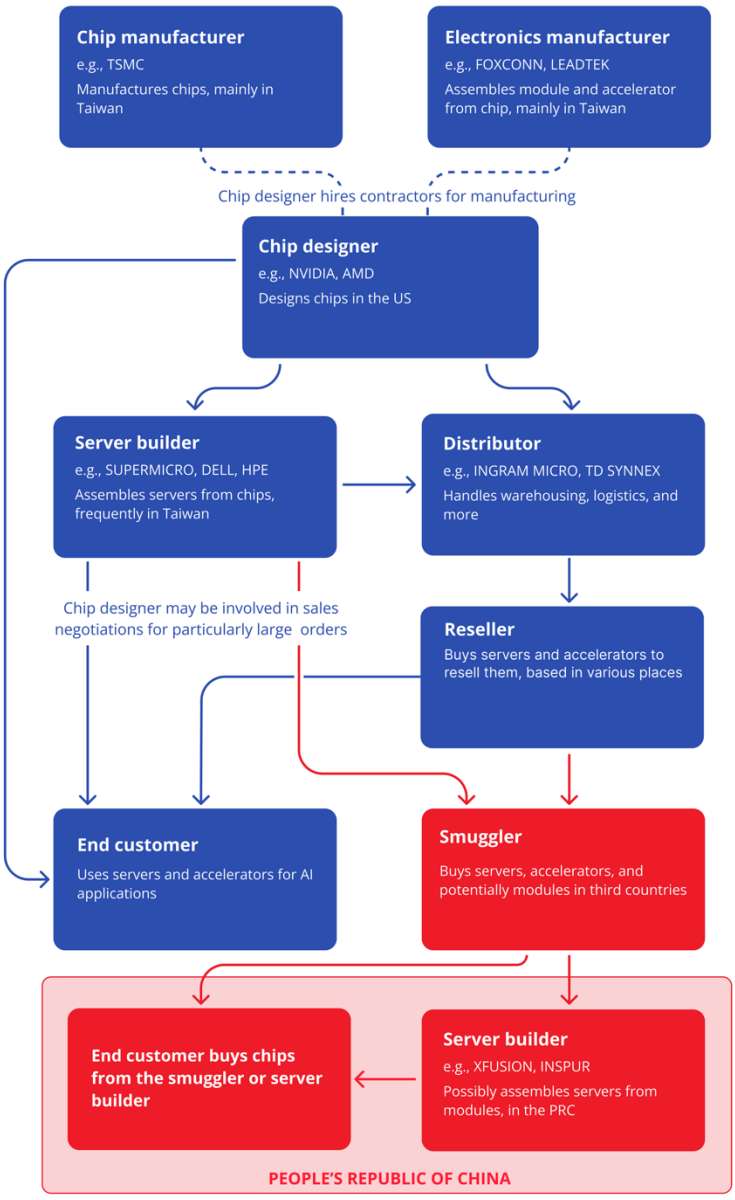
Now that the BIS has also banned the A800 and H800, the performance gap between the chips available for legal purchase in the PRC and outside the PRC has widened, and it will likely keep growing as NVIDIA's new Blackwell series of chips enters the market.[52] For context, the best banned chip is about seven times as performant for AI training as the best that can be exported legally to the PRC, and two to three times as performant as the best domestic Chinese AI chip.[53] It also has major software and other advantages over the best domestic Chinese alternative.[54] This dynamic likely creates strong incentives to smuggle.

Better Enforcement Can Curtail AI Chip Smuggling

AI chip smugglers follow a well-established playbook. Typically, they purchase chips via shell or front companies located in "third countries," generally in East and Southeast Asia. Smugglers then reexport the chips to the PRC, sometimes with intermediate steps to muddy the trail, such as by labeling them as another, non-restricted product. Shell companies can typically be set up online for a few thousand dollars in a matter of hours or days, whereas it can take years of investigation to uncover a shell company's illicit activities.[55] Bad actors use these methods to smuggle varying quantities of chips, ranging from small orders to massive shipments worth hundreds of millions of dollars.[56]

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Figure 7: The AI Chip Smuggling Chain



Smugglers used this playbook to supply Russia with sanctioned military goods. A report by the U.S. Senate's Permanent Subcommittee on Investigations Majority Staff found that countries such as Armenia, Georgia, and Kazakhstan saw enormous increases in imports from U.S. chip firms following Russia's invasion of Ukraine, suggesting that restricted items were being reexported to Russia through those countries.[57] Similar evidence now points to AI chips being smuggled into the PRC. In 2024, NVIDIA's exports to Singapore alone constituted 18 percent of its total revenue—a sharp increase shortly after the A800 and H800 ban.[58] It is unlikely that Singapore or neighboring countries required this many of NVIDIA's chips—between 2013 and 2024, only 300 AI companies were founded across the

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

entire Southeast Asian region, fewer than in Germany (394) or Japan (388)—suggesting a reasonable share of these chips were truly destined for the PRC.[59] In January 2025, Singapore authorities arrested three individuals for misrepresenting the ultimate destination of $390 million worth of AI servers containing NVIDIA chips shipped to Malaysia.[60] Chinese retailers are also advertising servers containing banned NVIDIA chips built by Chinese server makers—including NVIDIA partners like Inspur—suggesting there are also modules and/or graphics cards being smuggled into the PRC and assembled there.[61]

A Case Study of Large-Scale AI Chip Smuggling

Based on interviews with eight chip smugglers, several NVIDIA employees, and authorized distributors, along with a review of procurement documents, *The Information* tracked multiple instances of large-scale chip smuggling.[62]

In one case, the smuggling occurred as follows:

1. A company in eastern China placed a $120 million order for around 2,400 NVIDIA H100 graphics processing units (GPUs)—the most advanced AI GPU available from NVIDIA at the time.
2. The Chinese company placed the order with a broker in Malaysia, who assisted it in establishing a Malaysian shell company to conceal its national identity.
3. As part of this, the chip broker helped the Chinese company set up a Malaysian business website and email address. The broker also rented space in a Malaysian data center to temporarily house the GPUs once they arrived, in order to fool any NVIDIA staff conducting inspections.
4. After an unspecified number of weeks, the GPUs were shipped from the data center in Malaysia to the buyer in the PRC.

To enforce its controls, the BIS relies largely on exporters to screen buyers, look for red flags, report violations, and conduct other due diligence. However, these exporters have limited visibility into their distribution networks and little incentive to improve them. Within this ecosystem, NVIDIA is by far the most important player. It has an estimated 80 to 95 percent of the AI chip market, and all evidence of AI chip smuggling so far has concerned NVIDIA chips.[63] NVIDIA has responded to past reports of smuggling by stating that it insists its customers and partners "strictly adhere to all export control restrictions."[64] NVIDIA's AI chips are typically sold to partner companies that assemble them into servers. These servers are then sold to end-users, such as AI labs or cloud providers, sometimes via multiple layers of distributors and resellers. Large orders often involve a three-party negotiation between the buyer, the partner, and NVIDIA.[65]

Within the space of server maker companies, NVIDIA's largest revenue source is Supermicro.[66] According to the *Financial Times*, "People involved in the trade said merchants in Malaysia, Japan, and Indonesia often shipped Supermicro servers or NVIDIA processors to Hong Kong before bringing them across the border to Shenzhen."[67] *The Information* report cites a smuggler claiming to acquire thousands of chips from companies like Dell and Supermicro "thanks to what he called 'strong personal relationships' with sales representatives at these firms."[68] A report by analyst firm Hindenburg Research also documented multiple compliance failures by Supermicro, alleging, for example, that it has supplied

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

millions of dollars of products to a distributor in Russia through a Californian entity despite sanctions.[69] Supermicro servers have also been advertised on Chinese e-commerce sites.[70] Supermicro has responded to past reports of smuggling by stating that it follows "all US export control requirements on the sale, service, support, and export of GPU systems."[71]

These companies lack visibility into their distribution networks for several reasons. First, existing regulations create the incentive to know as little as possible about who ultimately owns exported chips, so long as companies do the limited due diligence required to avoid legal exposure. Under ECRA, an exporter is only subject to criminal penalty if it willfully commits (or tries, conspires, or helps someone else commit) a violation of the rules, and for civil penalties, factors like willfulness, negligence, and concealment are considered.[72] Second, while these companies are strongly motivated to ensure compliance with U.S. regulations, they must balance more aggressive anti-smuggling efforts against their financial obligations to shareholders. It is unclear how such above-and-beyond efforts would affect the market competitiveness of these companies. NVIDIA has stated that illicit trade "would be a burden on our business, not a benefit," and rampant smuggling could cause the BIS to extend the export ban to important reexport countries, possibly reducing NVIDIA's access to foreign markets.[73]

Despite these challenges, the U.S. government and industry can act to limit AI chip smuggling. AI chip smuggling is a relatively new phenomenon, and limited measures have been taken thus far to curb it. The incentives to smuggle AI chips increased substantially as the BIS tightened export controls in October 2023 and as AI companies began planning ever larger compute infrastructure buildouts. AI chip designers and resellers also lack deep experience in export control compliance, which is more common in the defense industry, suggesting substantial room for improvement.

With their significant technical capacity and resources, a few key companies could make a substantial difference in enforcing AI chip controls. As mentioned previously, NVIDIA holds an estimated 80 to 95 percent of the AI chip market, and it is likely that nearly all smuggled AI chips are NVIDIA chips.[74] Meanwhile, NVIDIA works closely with partners that assemble those chips into servers, such as Supermicro, and other partners that distribute servers to resellers. NVIDIA can encourage these partners to take additional measures to prevent smuggling—the authors recommend specific measures below—and even refuse to sell to them if they do not.

Finally, unlike other dual-use goods such as highly enriched uranium, AI chips are digital computers. This allows for a range of promising technical anti-smuggling measures enabled by functionality implemented directly on the chips' hardware or firmware. For example, AI chip designers could implement a software-based feature that allows chip owners to prove to some authority (like the chip designer or the BIS) that their chips are not located in the PRC, a proposal called "location verification."[75] Other hardware-enabled mechanisms could provide further tools to help the U.S. government enforce the AI chip controls.[76]

Recommendations

A key problem for enforcing the AI chip controls is that the BIS, AI chip designers like NVIDIA, and other AI chip exporters and reexporters currently lack a clear picture of the scale, geography, and key actors for AI chip smuggling. As a result, they cannot proactively block diversion attempts, take punitive action against smugglers to disincentivize future smuggling, or properly evaluate the effectiveness of the AI chip controls and related enforcement practices.

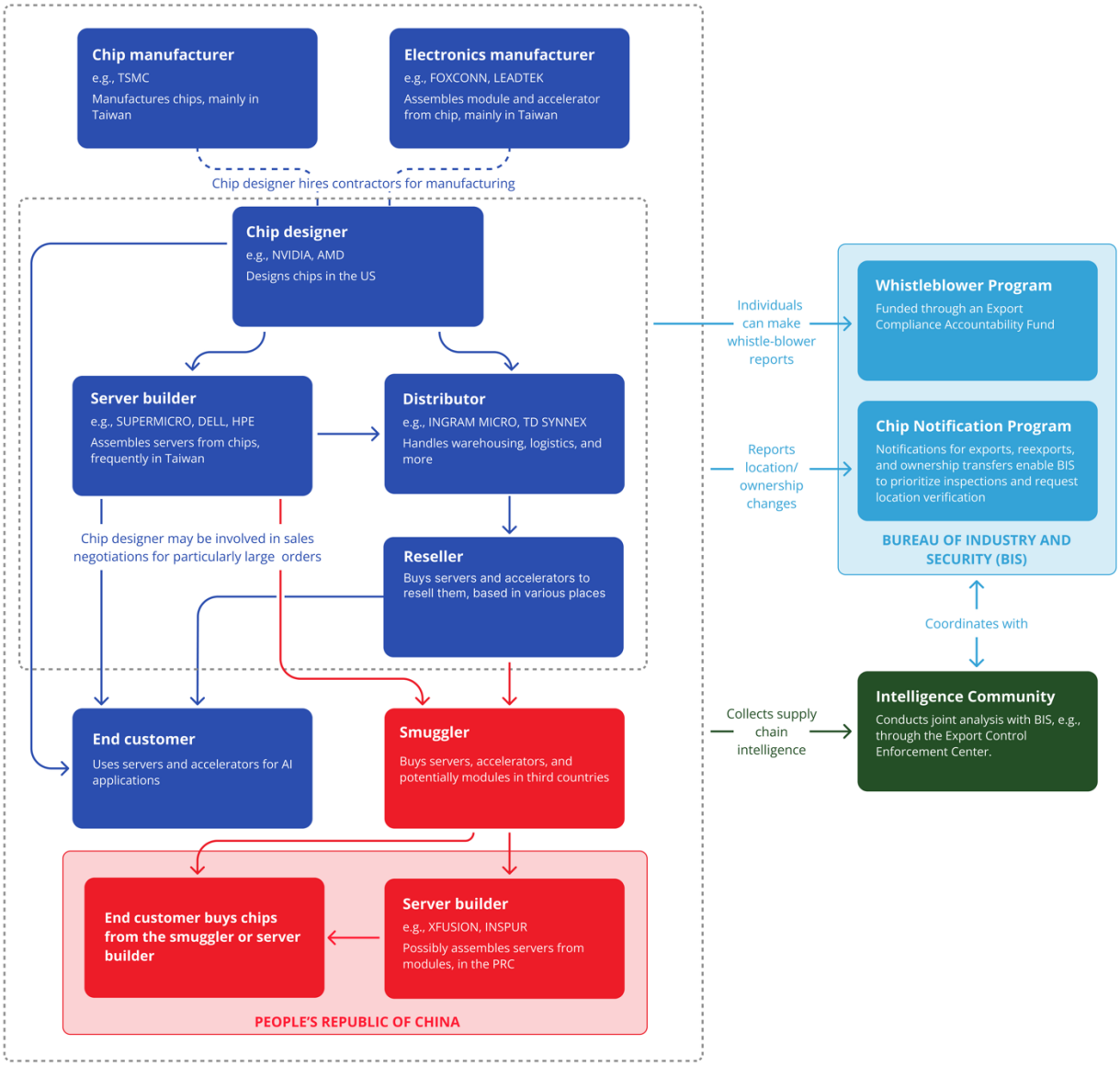
The BIS and chip exporters already have many capabilities to address AI chip smuggling, but they require a clearer picture of the smuggling ecosystem. For instance, the BIS can stop or penalize diversion attempts effectively when the violating entity has a presence in the United States. It is harder

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

for them to do this abroad, but even in this case, the BIS has some ability to interdict shipments and penalize offenders. It can also extend the AI chip ban to third countries commonly used for reexport. Meanwhile, AI chip designers and distributors can cut out buyers that are likely to engage in smuggling, or that do not take appropriate due diligence measures.

Below, the authors offer a potential playbook to address AI chip smuggling. These recommendations focus primarily on actions that the BIS and industry can take with their current authorities and resources. Though the authors have designed each recommendation to be useful on its own, their synergies would make them more effective taken together.

Figure 8: Overview of Proposed Recommendations



Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Strengthen Industry Due Diligence

Because of their substantial resources and interactions with potential buyers, AI chip companies and distributors have significant power to mitigate AI chip smuggling.[77] In particular, if NVIDIA were to strengthen its due diligence processes and cause companies in its distribution network to do the same, most smuggling cases would be covered by this strengthened regime.

AI chip companies are aware that smuggling happens and have taken some steps to address the problem.[78] The Semiconductor Industry Association (SIA) recently highlighted extra steps some chip companies have taken to curb smuggling into Russia.[79] These "compliance-plus" measures include deploying software to automatically scan for smuggling, verifying customer addresses, contractually prohibiting customers from reselling chips, requesting more information from customers about how and where they intend to use chips, regularly auditing distributors' sales data and practices, and establishing mechanisms for employees to anonymously report smuggling.[80] The SIA does not name specific companies or detail which measures each has implemented.

These measures are helpful but unlikely to dramatically reduce smuggling, given that most smuggling likely occurs further down the supply chain once chips have been sold by distributors. Figure 9 highlights actions firms could take to ensure a robust due diligence process for all substantial purchases of export-controlled chips.

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Figure 9: Enhanced Due Diligence Measures for AI Chip Exports

Enhanced Due Diligence Measures for AI Chip Exports

Stage	Measure
Initial customer risk assessment	Identify the customer and their beneficial owners, and then assess chip smuggling risk factors, such as: • Is the customer in a country where smuggling is more likely? • Is the customer a first-time buyer or otherwise unknown entity? • Does the customer have close ties to the People's Republic of China (PRC) or other countries where artificial intelligence (AI) chip exports are restricted? • Where will the chips will be kept and what they will be used for?
Further vetting of customers assessed as higher risk	Conduct more rigorous vetting, including: • If the customer is a reseller—who it has sold controlled items to in the past. • Identify key personnel employed by or associated with the customer, and check whether these appear on any relevant lists, or have significant ties to the PRC. • Examine the customer's operational and financial history. • See if there are other known parties that can vouch for the customer. • Visit the customer's offices. • Look for any of the red flags listed in Supplement No. 3 to Part 732 of the Export Administration Regulations. • Any other relevant measures used in the financial services industry for combating money laundering and terrorist financing. For customers that are expected to resell the chips, such as distributors and original equipment manufacturers: • Regularly audit the customer's sales records as well as their own Know Your Customer (KYC) investigations. • Ensure that the customer understands these KYC processes and has trained staff capable of carrying them out. • If the customer is a repeat buyer, check to whom they sold previous orders of chips and what KYC vetting they did for those sales.
Further risk mitigation (if there are significant unresolved concerns from previous stage)	Could include measures such as: • Place the order on hold pending further investigation. • Limit order quantities or impose staged delivery. • Consult with the Bureau of Industry and Security for guidance on complex cases. • Where risk cannot be adequately mitigated, consider order modification or cancellation.
Post-shipment verification	For customers that are not expected to resell the chips, such as cloud providers and AI labs: • Carry out post-shipment inspections to the data center where the chips are installed, first within a week of the installation and then again six months later.* • If the customer is a repeat buyer, ensure that the chips purchased previously are still being used as intended, for example, by inspecting the data center.

** NVIDIA (and possibly other firms) has started to make post-shipment inspections for at least some large orders to verify that the chips sold are installed in a data center. However, these measures likely need to be improved, as shown by a recent report of a smuggler temporarily installing the chips in a data center to fool NVIDIA's inspectors before shipping them to the PRC. Source: Qianer Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry," The Information, August 12, 2024, http://www.theinformation.com/articles/nvidia-ai-chip-smuggling-to-china-becomes-an-industry.*

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

These voluntary due diligence measures will only be effective if they are well implemented and have industry buy-in. To build confidence, NVIDIA and other exporters should form detailed plans for their intended due diligence measures and share them confidentially with the BIS and relevant executive branch officials and congressional committees. Meanwhile, the BIS should periodically verify implementation and, if necessary, check records of past Know Your Customer (KYC) evaluations.

These due diligence measures will incur costs on exporters and reexporters. AI chip exporters will bear most of the cost, but they have seen staggering revenue growth in recent years. For example, NVIDIA earned a net income of $30 billion in fiscal year 2024.[81] Additionally, extremely strong and growing U.S. demand for NVIDIA's AI chips means that it could redirect sales from potentially risky actors in Southeast and East Asia to more trustworthy domestic customers.[82] Though the BIS will need to play a consulting role with industry as these due diligence measures are implemented, it could likely do so without additional appropriations.

Create AI Chip Export Notification Requirements

The BIS could collect much more detailed information about the location and ownership of controlled AI chips by creating a new notification requirement. This would involve companies reporting ownership and location changes for restricted AI chips involving a foreign entity or location to the BIS. The BIS would then collect that data and store them in a database, with the BIS then using that data to improve its enforcement activities. This data would significantly enhance the BIS's visibility into advanced AI chip distribution networks, making it easier to detect diversion. A previous analysis conducted by the authors provides offers one potential form of implementation.[83] Here, the authors describe a leaner version that the BIS could implement without any change to its budget or existing authorities. Former Senator Mitt Romney mentioned a similar idea in a hearing on export control enforcement that took place in April 2024.[84]

The proposal requires that:

- The BIS adds a notification requirement to the Export Administration Regulations (EAR) for location or ownership changes involving a foreign location or entity of more than 100 items controlled by ECCN 4A090. The requirement is modeled after existing requirements in the EAR, and like those would exist separately from license requirements and requests. The authors provide draft text for these additions in Appendix B.

- Anyone who initiates a transaction of affected items reports it to the BIS within a month, using a standard form.

- The BIS sets up a database to collect, store, and use the reported information.

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Figure 10: Useful Data Attributes for AI Chip Notification Requirements

Attribute	Rationale
Serial number*	Used to keep track of items across transfers
Product name	Used to search for, filter, and group data
Export Control Classification Number	Used to search for, filter, and group data
Item status (whether it is declared lost)	Used to report items that have been legitimately lost or broken
Name, address, and telephone number of the owner after the transfer	Used to be able to contact the chip owner, as well as for searching, filtering, and grouping data
Name, address, and telephone number of the owner or operator of the location after the transfer	Used to be able to contact the operator to locate or inspect the chip physically, as well as for searching, filtering, and grouping data
Date that the transaction is expected to begin	Used to reconstruct a chip's location and ownership history
Date that the transaction is expected to be completed	Used to reconstruct a chip's location and ownership history

**This should be the most tamper-resistant unique identifier present and inspectable on the device.*

The authors expect the notification requirement to incur only a minor and manageable cost on exporters and reexporters, as larger AI chip sellers likely already track relevant information and could automate the reporting process, and smaller distributors sell low enough volumes that the reporting can likely be done manually. Smaller distributors located in third countries will be incentivized to report transactions because a failure to do so would violate the EAR and incur hefty penalties. Moreover, the BIS can add violating entities to the Entity List and advise NVIDIA and other industry actors to cut them out of their distribution networks.

Implement Software-Based Location Verification

Manual and one-off location checks are insufficient to address the growing challenge of AI chip smuggling, given the ease with which smugglers can fool manual inspections, and the fact that many smaller orders of servers can later be combined to build much larger AI clusters once in the PRC. However, it is possible to use well-established technical measures to automate the process of verifying the location of exported chips.

AI chip designers should implement location verification features on controlled AI chips to allow chip owners to prove their chips are not in the PRC or other prohibited markets. The most promising method relies on measuring the network latency for a query (with a "ping") to trusted landmark servers in order to infer basic facts about the chip's physical location.[85] Importantly, this method would not allow the chip designer to access other data; it only gives the chip owner the ability to prove to the chip designer roughly where its chips are, if and when the user chooses to do so.[86] The chip designer—or the BIS—

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

could then require chip owners outside the United States to periodically prove that their chips are not in the PRC as a condition for importing chips.

Location verification features on AI chips would allow for significantly more reliable, scalable, and cost-effective post-shipment verification. It would reduce or even remove the need for the labor-intensive, in-person inspection that exporters are already doing for some sales. It could also provide chip firms, and by extension the BIS, with strategic information about how many chips go missing and where, and help direct and inform investigations into illicit trade in AI chips.

To implement location verification features, AI chip designers could implement the feature and deploy it to controlled chips through a firmware update. They could also install a set of trusted landmark servers in key locations outside the PRC.[87] Having done this, they can:

- Ask chip buyers to verify the chips' locations within a month of their ship date, instead of carrying out manual post-shipment data center inspections.
- Mandate that chip owners regularly verify the location of their chips, keep a running list of chips and whom they have been sold to, and, whenever a notable batch goes missing, investigate and inform the BIS.
- Support AI chip export notification requirements by allowing the BIS to ask chip owners to verify a chip's location and report if the verification was successful.

In May 2025, bipartisan legislation was introduced in the U.S. House and Senate that would require location verification on restricted AI chips within six months, after which exporters would need to report to the BIS if any chip stops confirming it is outside countries of concern.[88]

Increase the Intelligence Community's Coordination with the BIS

The intelligence community (IC) is well suited to analyze AI chip diversion, as it has substantial resources and prior experience with nonproliferation and smuggling, for instance, in combating the illicit drug trade, uncovering North Korea's nuclear weapons program, and hindering the proliferation of weapons of mass destruction via the National Counterproliferation and Biosecurity Center.[89] Meanwhile, the BIS's limited resources prevent it from conducting many potentially important analyses, which improved collaboration could address. As a result, the intelligence community and the BIS could likely see substantial benefits from increased coordination.

Concretely, the intelligence community could:

- Map AI chip smuggling networks in detail, including key personnel and companies and structures for financial transactions.
- Estimate how many chips are being smuggled and through which countries.
- Search for weak points in the AI chip distribution network.
- Provide real-time alerts of imminent diversion attempts.
- Investigate possible past instances of AI chip diversion to support legal cases.

Currently, coordination between the BIS and the intelligence community happens via the BIS's Information Triage Unit.[90] However, its main role is to inform license application reviews, and as such it

does not carry out the type of analyses mentioned above.[91] License reviews are largely irrelevant to AI chip diversion, since advanced AI chips are banned on a country-wide basis and hence likely see very few license applications. However, the BIS could use the Information Triage Unit to also solicit intelligence as described above, which it could then share with relevant parts of the BIS and declassify for industry where appropriate. Useful actions to establish this coordination include:

1. The Director for National Intelligence could make intelligence related to export control enforcement a top priority in the National Intelligence Priorities Framework (NIPF), enabling IC agencies to devote more resources to this kind of intelligence collection.

2. The Secretary of Commerce and Director for National Intelligence could agree to establish a new joint team, composed of BIS enforcement personnel and analysts from IC agencies. A natural home for such a team could be the existing Export Control Enforcement Center. This would facilitate coordination between BIS and the IC, and provide both BIS and the IC with a dedicated analysis capability to inform intelligence collection and enforcement priorities.

Incentivize Insiders to Report Violations

Congress should authorize a whistleblower incentive program and qui tam lawsuits to incentivize reports of possible export violations. These measures could significantly boost export enforcement for the BIS while paying for themselves in revenue generated through additional penalties.

A bipartisan bill introduced into the Senate in April 2025 would implement a whistleblower incentive program funded entirely through penalties and modeled on the Securities and Exchange Commission (SEC) Whistleblower Program, created in 2010 as part of the Dodd-Frank Act.[92] The SEC Whistleblower Program is considered highly successful.[93] Since its inception, the SEC Whistleblower Program has awarded more than $2.2 billion to 444 individuals, aiding $7 to $22 billion in penalties.[94]

Qui tam lawsuits allow individuals to file lawsuits on behalf of the government against violators of export law. Authorizing them would allow anyone to sue a violator and collect a portion of the resulting penalty, modeled on a similar law in the False Claims Act. The qui tam provision in the False Claims Act has been successful at uncovering fraud against the government. In fiscal year 2022, about 70 percent of cases brought under the False Claims Act were qui tam, and about 90 percent of the $2.2 billion recovered came through qui tam cases.[95] Penalties imposed in this way serve as a strong deterrent against violators and are revenue for the federal government.

AUTHORIZE A WHISTLEBLOWER INCENTIVE PROGRAM

The BIS offers financial rewards for some reports of sanctions violations.[96] However, these incentives are too weak to substantially reduce AI chip smuggling and other common violations. To be effective, a whistleblower program needs to ensure that potential whistleblowers are strongly incentivized to report leads by:

- Offering high monetary awards for successful tips, mirroring the SEC Whistleblower Program, which offers awards at between 10 and 30 percent of the penalty.[97]

- Providing robust protections against employer retaliation, with measures to protect the confidentiality of whistleblowers, making it illegal for U.S. employers to retaliate against whistleblowers, and allowing whistleblowers to report leads anonymously if represented by an attorney. Congress could also explore options such as increasing the BIS's ability to grant or recommend S visas for whistleblowers, or in extreme cases involving credible threats, consider protective measures similar to those used in witness protection programs.[98]

- Providing rapid review and action on whistleblower reports. Long turnaround times have reduced the effectiveness of other whistleblower programs.[99]

The bipartisan legislation introduced into the Senate in April 2025 aims to meet these goals. To ensure the whistleblower incentive program is adequately funded, the bill would also establish an Export Compliance Accountability Fund that would be financed by monetary penalties paid by export law violators. In addition to paying awards to whistleblowers, the fund would also finance activities like reviewing and investigating whistleblower reports, protecting whistleblowers from employer retaliation, educating businesses and individuals about the program, and record-keeping and reporting. Such a fund has a direct precedent in the SEC's Investor Protection Fund. The authors expect the Export Compliance Accountability Fund to pay for itself or even generate net income for the federal government, as whistleblowers would help uncover violations—and collect penalties—that would otherwise go undetected.

Authorize Qui Tam Lawsuits

A qui tam provision should be modeled on the qui tam provision in the False Claims Act, where it helps discover and penalize instances of fraud against the government. This provision should allow private individuals to file qui tam lawsuits if they have evidence of violations against the EAR (within a certain number of years of the violation). After a lawsuit has been filed, the Department of Justice (DOJ) should have the option of taking over the case if it wishes to. If it declines, the individual can proceed with the case on their own. If the case is won, the DOJ determines how much the individual is awarded following a similar methodology to that of the whistleblower incentive program described above.

Congress should empower the DOJ to prosecute qui tam lawsuits for export violations and determine the awards provided from those lawsuits. The DOJ already has a team within its National Security Division dedicated to prosecuting export law violations, including violations related to ECRA. The most natural alternative to the DOJ, the BIS, currently lacks the resources needed to carry out these responsibilities.[100]

Extend Incentives Globally

Most violations of the EAR likely occur abroad, in countries such as Malaysia, Singapore, and Taiwan. As a result, both U.S. and foreign citizens should be eligible to submit leads and receive awards via the whistleblower incentive program, as well as to bring qui tam lawsuits. The SEC program allows this and has received reports from at least 130 countries.[101] Many of the potential violators will also be foreign companies. As a result, the whistleblower incentive program and the qui tam provision should apply to any violation of the EAR, as described by ECRA, no matter if the violator is a U.S. or foreign entity.

Although the authors expect these measures to especially benefit efforts to prevent AI chip smuggling due to the large sums of money involved, the measures should not be written as specific to AI chips. Rather, they should allow for whistleblower reports on all violations of the EAR, thus also helping to prevent U.S. technology from reaching Russia, Iran, and other adversaries, as well as prevent advanced chip-making tooling from reaching the PRC. A floor of $1 million for any violation to result in an award could reduce the risk that the BIS gets overwhelmed by reports of insignificant violations.

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Increase Funding for the BIS

Congress should grant the White House's $313 million budget request for BIS for fiscal year 2026, a 64 percent increase over its current budget of $191 million.[102] This would improve national security by allowing substantial additional funding for export control enforcement, which has decreased in real terms over the past seven years. As Matt Borman, former principal deputy assistant secretary of commerce for export administration, said last year, "We spend 100 percent of our time on Russia sanctions, another 100 percent on China and the other 100 percent on everything else."[103] BIS's current budget for export control enforcement is around 55 million, half the cost of a single F-35 fighter jet.[104]

Figure 11: Bureau of Industry and Security: Total Budget and Budget for Export Enforcement

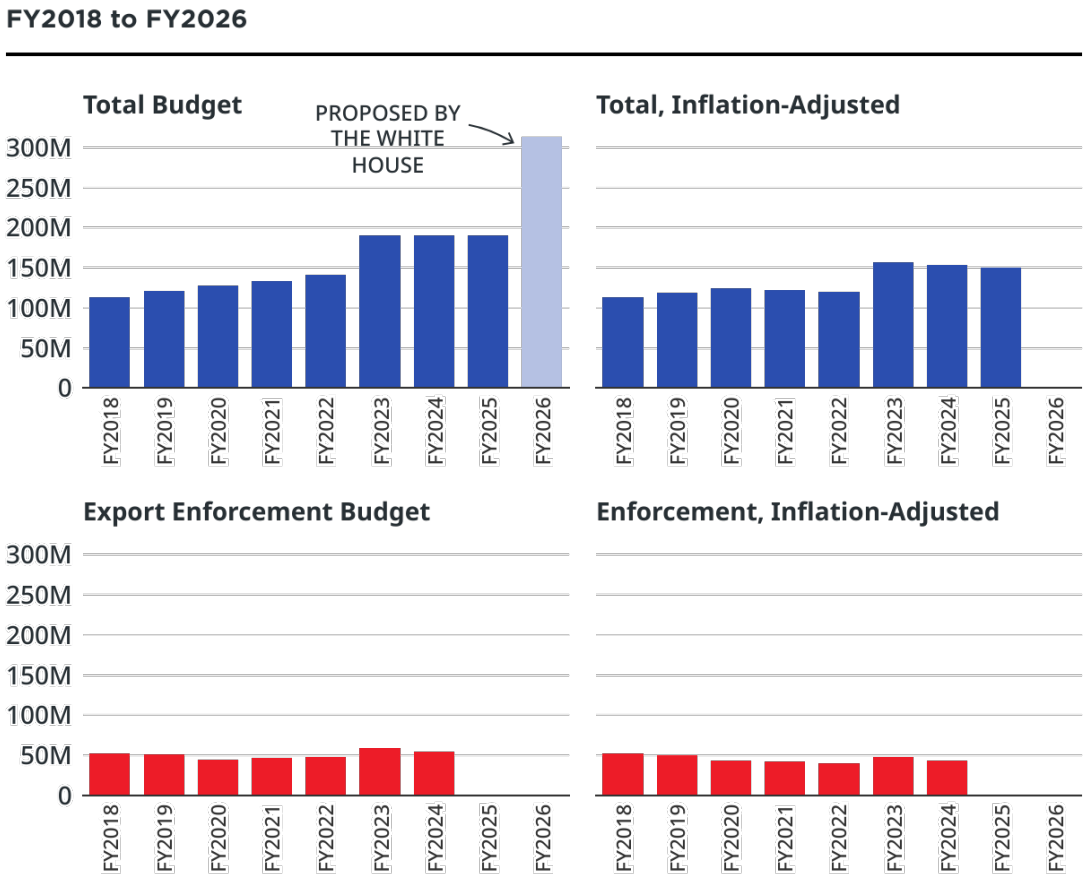


Chart: Center for a New American Security (CNAS) • Source: **U.S. Department of Commerce, U.S. Bureau of Labor Statistics, U.S. Office of Management and Budget.**

Investment to better enforce AI chip controls could likely pay for itself. The BIS's fines for export violations return to the federal government, which means that increased BIS enforcement increases government revenue. The high cost of AI chips means penalties for smuggling would also be high. There are multiple documented instances of AI chip smuggling from 2024 alone that each could have resulted in a penalty over $200 million, more than the BIS's annual budget.[105] These sums are not unprecedented: in 2023, the BIS imposed a $300 million penalty on Seagate for selling hard drives to Huawei.[106] If the aforementioned budget increase resulted in civil penalties for even 2,100 NVIDIA H100

chips per year—less than two percent of all estimated AI chips smuggled in 2024—it would pay for itself.[107]

Money spent to improve the enforcement of the BIS's AI chip controls is also a highly cost-effective way of strengthening national security. A budget increase of the scale proposed here would help the BIS modernize its antiquated tooling and improve its operational capacity, for example, by hiring more people. In a similar vein, a Center for Strategic and International Studies report has recommended an increase of $25 million to pay for BIS technological modernization, including procuring data sets, integrating those into a data analytics platform, and adding and training analysts.[108]

Conclusion

Despite U.S. export controls, tens or even hundreds of thousands of banned AI chips likely made it into the PRC in 2024. This is evidenced by historical precedent, listings of banned chips on Chinese online platforms, and credible investigative reports. While U.S. export controls still significantly constrain Chinese companies, enforcement gaps have reduced their effectiveness. To protect America's compute advantage over the PRC, the BIS, Congress, and AI chip companies must take decisive action. The challenge of AI chip smuggling is unlikely to go away on its own: as the capabilities of U.S.-designed AI chips grow, so will their black market demand.

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

Appendices

Appendix A: Methodology

To generate estimates for the quantity of AI chips smuggled in 2024, the authors used Monte Carlo simulations to model a range of possible outcomes from the two highest likelihood large-scale smuggling pathways. In the first pathway, smugglers use shell companies to re-export relatively low volumes of AI chips purchased from resellers in various third countries. In the second pathway, smugglers set up real cloud service providers as front companies in third countries, use those to purchase relatively large volumes of AI chips directly from server builders and/or AI chip makers, and re-export a portion of the purchased chips. The model's input parameters, such as the number of cloud provider fronts and order volumes, are grounded in real-world data whenever possible, although with substantial uncertainty given the limited data available. The online methodology provides more detail and allows for parameter adjustments to explore different assumptions.[109]

The estimates suggest that smuggling of AI chips by PRC-linked actors more likely than not exceeded 100,000 GPUs in 2024, with a median estimate of around 140,000. These estimates have a large amount of uncertainty given a lack of firm data on chip smuggling cases across the ecosystem, and the complex dynamics involved in large-scale smuggling operations. Accordingly, these estimates rest on a series of assumptions based on a small set of public data; this is reflected in the authors' wide confidence intervals. They can be seen as an extrapolation of known reported cases. The authors present these estimates to help policymakers prioritize their attention in the absence of better data.

Smuggling Pathway One

In this pathway, which is highlighted in reporting from *The Wall Street Journal*, *The New York Times*, and *The Information*, each shell company buys a relatively small number of AI chips and disguises itself as a variety of different firms.[110] After procurement, the chips are relabeled as non-controlled chips and transported to front or trading companies in the PRC. Modeling this pathway uses two key parameters: the number of vendors selling controlled chips into the PRC and the average monthly sales volume per vendor.

Public reporting from July 2024 suggested that, at the time, somewhere around 70 sellers in the PRC were openly advertising export-controlled NVIDIA GPUs online.[111] This corresponds to the number of vendors found through the authors' own searches across Baidu, JD.com, and Taobao. The authors assume that the true number could plausibly be half or double; some of these vendors may not actually have real products for sale, some may be operating through multiple online storefronts, some may not be actively advertising online (especially those moving large numbers of GPUs), and some discoverable vendors may have been missed in the search process underlying public reporting. Of the 70 distributors identified by *The Wall Street Journal*, the reporters were able to verify 25 of them.[112] The authors model these dynamics as a simple probabilistic distribution, with a 90 percent confidence interval spanning from 38 to 150 vendors.

The authors' aggregation of news reports suggests that the average volume of export-controlled GPUs sold per vendor is approximately 200 per year (see Figure 2). However, this distribution will have a long tail, based on multiple reports of single transactions far exceeding this estimated average (2,000, 2,400, 4,800), and some reports of typical monthly order sizes numbering in the hundreds (see Figure 2). The

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

monthly average for the median vendor is likely lower than these large orders, as reporting is more likely to pick up larger-volume orders. However, there may also be larger, more professional smuggling operations that have evaded such reporting. The authors therefore assume an average monthly order volume for each vendor ranging from the bottom end of reporting (17 per month) up to the size implied by more regular, larger orders (1,000 per month).

Smuggling Pathway Two

In this pathway, rather than working through multiple smaller resellers, low- to mid-tier cloud providers located in countries in South and Southeast Asia place orders directly with chip exporters. After procurement, a portion of these chips could be relabeled as non-controlled goods and transported to companies in the PRC. This pathway makes sense to smuggle larger volumes of chips: there is growing demand for cloud computing services in the region, allowing local demand for AI computing to serve as a convenient cover story for illicit actors. The authors modelled this pathway using three parameters: the number of cloud operators involved in this strategy, the number of chips acquired by each operator, and the proportion of chips diverted.

The authors assume no more than a quarter of cloud providers operating in a country are engaging in smuggling, given the risk of attracting too much scrutiny. Based on a non-exhaustive search, the authors identified between five and ten mid-tier local cloud companies for each of Indonesia, Singapore, Thailand, and Vietnam, which together with a few Chinese providers operating there makes for a total of approximately 10 cloud providers per country. This yields between one and ten potential cloud service provider fronts across countries of concern.

The authors assume any cloud company engaged in smuggling will want to avoid news coverage of its orders, so it will stay below a certain quantity. Reporting from 2023 suggests ByteDance and Alibaba have placed orders of around 100,000 H800s and A100s, whereas large U.S. companies have attracted attention for orders reaching hundreds of thousands of GPUs in 2024.[113] Given this, the authors assume cloud companies engaged in smuggling will seek to stay below 100,000 GPUs ordered per year, with a modeled range between 10,000 and 100,000.

Too much diversion from front companies posing as cloud providers could raise red flags, due to a lack of maintenance requested from exporters, an unusually low power consumption footprint of the data center, or a lack of customers for its AI services. Depending on how readily observable these factors are, the authors roughly assume a cloud front would need to hold on to at most half of its chips. The authors model this as a fraction of chips diverted ranging from 40 percent to 99 percent.

Final Estimate

There are many reasons these scenarios and assumptions could be unrealistic. PRC-linked actors could aim for less or more ambitious, surreptitious, or inventive operations. The authors also expect that these numbers underestimate the difficulties involved in AI chip smuggling, given the authors' approach of focusing on the ways that large-scale smuggling could be successful, rather than the ways it could fail.

Therefore, the authors attempt to model their remaining uncertainty, building in uncertainty about whether one, both, or neither of the two pathways described above will be pursued. Given the overwhelming focus of public reporting, the authors assign a 95 percent chance to pathway one being pursued in 2024. This is decomposed into a 75 percent chance that pathway one is the only major pathway pursued, and a 20 percent chance that both pathway one and pathway two are pursued. An approximately 5 percent chance is assigned to this overall framework being completely off, which is modeled as noise across many possible outcome distributions.

Appendix B: Additions to the Export Administration Regulations for an AI Chip Notification Requirement

The authors offer draft text for inclusion in Part 743 (Special Reporting and Notification) of the EAR.[114] This text would implement the notification requirement and is based on an existing reporting requirement for thermal imaging cameras.[115]

§ 743.7 Advanced computing notification.

(a) **General requirement.** Changes in location or ownership must be reported to BIS as provided in this section.

(b) **Transactions to be reported.** Location or ownership changes as defined in § 772.1 of the EAR involving a foreign location or entity of more than 100 items controlled by ECCN 4A090 over any three-month period must be reported to BIS.

(c) **Party responsible for reporting.** The owner prior to the transfer must ensure the reports required by this section are submitted to BIS.

(d) **Information to be included in the reports.** For each export described in paragraph (b) of this section, the report must provide a purchase order if one exists, as well as a spreadsheet (following a template provided by BIS) with the following information for each transferred item: the serial number (the most tamper-resistant unique identifier present and inspectable on the device); the product name; the ECCN; the item's status (whether or not it is declared lost); the name, address, and telephone number of the owner after the transfer; the name, address, and telephone number of the owner or operator of the location after the transfer; the date that the transaction is expected to begin; and the date that the transaction is expected to be completed.

(e) **Where to submit reports.** Submit the reports via email with spreadsheets attached to AIChipNotifications@bis.doc.gov.

(f) **Reporting periods and due dates.** This reporting requirement applies to exports made on or after [insert date]. Transfers must be reported no later than within the last day of the month following the month in which the transfer took place.

Additionally, the terms "location change" and "ownership change" would need to be defined in Part 772 (Definitions of Terms) of the EAR.[116]

Location change. Any movement of an item subject to the EAR from one physical address to another. This includes, but is not limited to, transfers between different buildings of the same company, movements between separate facilities owned by the same entity, and relocations to a different city, state, or country. Temporary movements for the purpose of exhibition, demonstration, or testing are also considered location changes unless otherwise exempted.

Ownership change. Any transfer of ownership of an item subject to the EAR from one entity to another. This includes, but is not limited to, sales, gifts, and transfers between affiliated companies. However, it does not include temporary changes in possession such as loans and leases. An ownership change occurs even if the item remains in the same physical location. Changes in the ownership structure of the entity possessing the item (such as corporate mergers or acquisitions) may also constitute an ownership change if they result in a new entity gaining control over the item.

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

[1] Hugo Meijer, Trading with the Enemy: The Making of US Export Control Policy Toward the People's Republic of China (New York: Oxford University Press, 2016); Ryan Fedasiuk, Karson Elmgren, and Ellen Lu, Silicon Twist: Managing the Chinese Military's Access to AI Chips (Center for Security and Emerging Technology, June 2022), https://cset.georgetown.edu/publication/silicon-twist/; "Special Report: How U.S.-Made Chips Are Flowing into Russia," Nikkei Asia, April 11, 2023, https://asia.nikkei.com/Business/Tech/Semiconductors/Special-report-How-U.S.-made-chips-are-flowing-into-Russia.

[2] Fanny Potkin, "Exclusive: Huawei Aims to Mass-Produce Newest AI Chip in Early 2025, Despite US Curbs," Reuters, November 21, 2024, https://www.reuters.com/technology/artificial-intelligence/huawei-aims-mass-produce-newest-ai-chip-early-2025-despite-us-curbs-2024-11-21/; Eleanor Olcott, Ryan McMorrow, and Tina Hu, Huawei's Bug-Ridden Software Hampers China's Efforts to Replace Nvidia in AI," Financial Times, September 3, 2024, https://www.ft.com/content/3dab07d3-3d97-4f3b-941b-cc8a21a901d6.

[3] "Notable models" are defined as models that were state of the art at the time of release, highly cited, or otherwise historically significant. Only models that had information about the hardware used to train them were included. "Notable AI Models," Epoch AI, June 19, 2024, updated February 13, 2025, https://epoch.ai/data/notable-ai-models.

[4] Qianer Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry," The Information, August 12, 2024, https://www.theinformation.com/articles/nvidia-ai-chip-smuggling-to-china-becomes-an-industry.

[5] Ana Swanson and Claire Fu, "With Smugglers and Front Companies, China Is Skirting American A.I. Bans," *The New York Times*, August 4, 2024, https://www.nytimes.com/2024/08/04/technology/china-ai-microchips.html.

[6] Xinghui Kok, "Singapore Charges Three With Fraud That Media Link to Nvidia Chips," *Reuters*, February 28, 2025, https://www.reuters.com/technology/singapore-charges-three-with-fraud-that-media-link-nvidia-chips-2025-02-28/; Reuters, "Singapore Prosecutors Say US Server Fraud Case Involves $390 Million of Transactions," *Reuters*, March 13, 2025, https://www.reuters.com/world/asia-pacific/singapore-prosecutors-says-us-servers-fraud-case-involves-390-million-2025-03-13/.

[7] Ryan McMorrow and Eleanor Olcott, "Nvidia's AI Chips are Cheaper to Rent in China Than US," *Financial Times*, September 6, 2024, https://www.ft.com/content/10aacfa3-e966-4b50-bbee-66e13560deb4.

[8] Swanson and Fu, "With Smugglers and Front Companies, China Is Skirting American A.I. Bans"; Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry."

[9] See an overview of listings that contained images here: https://docs.google.com/spreadsheets/d/1eOYMu7va2TV1bi58EFivnUC__rVk9eYHJoCoowjildc

[10] Monte Carlo simulations are used to simulate a range of possible outcomes from the two distinct large-scale smuggling pathways that the authors see as most likely. In the first pathway, smugglers use shell companies to reexport relatively low volumes of AI chips purchased from resellers in various third countries. In the second pathway, smugglers set up real cloud service providers as front companies in third countries, use those to purchase relatively large volumes of AI chips directly from server builders and/or AI chip makers, and reexport a portion of the purchased chips. The model's input parameters, such as the number of cloud provider fronts and order volumes, are grounded in real-world data whenever possible, although with substantial uncertainty, given the limited data available. For full methodology and results, see Appendix A, as well as Timothy Fist & Erich Grunewald, "AI Chip Smuggling Estimates (for 2024)," Google Colab, https://colab.research.google.com/drive/1ZxiwsHjc_aYh2N3Fzy2x7fg_nkjoYh2J.

[11] Stephen Nellis and Karen Freifeld, "Nvidia Faces $5.5 Billion Charge as US Restricts Chip Sales to China," *Reuters*, April 17, 2025, https://www.reuters.com/technology/nvidia-expects-up-55-billion-charge-first-quarter-2025-04-15/; Reuters, "TSMC Says It Has Alerted US of Potential China AI Chip Curbs Violation," *Reuters*, October 22, 2024, https://www.reuters.com/technology/artificial-intelligence/tsmc-says-it-has-alerted-us-potential-china-ai-chip-curbs-violation-information-2024-10-22/; Karen Freifeld, "Exclusive: TSMC could face $1 billion or more fine from US probe, sources say," *Reuters*, April 8, 2025, https://www.reuters.com/technology/tsmc-could-face-1-billion-or-more-fine-us-probe-sources-say-2025-04-08.

[12] $55 million vs $140 million. See Figures 6 and 9. Data from "Fiscal Year 2025 President's Budget Request," March 31, 2024, https://www.commerce.gov/sites/default/files/2024-03/BIS-FY2025-Congressional-Budget-Submission.pdf; Swanson and Fu, "With Smugglers and Front Companies, China Is Skirting American A.I. Bans"; Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry."

[13] According to the annual SEC report, whistleblowers have been awarded over $2.2 billion since the program began in 2011. Because whistleblowers can receive between 10 and 30 percent of the overall penalty, the total penalties associated with these rewards range from at least $7 billion to at most $22 billion. "Securities and Exchange Commission Office of the Whistleblower Annual Report to Congress for Fiscal Year 2024," Securities and Exchange Commission Office of the Whistleblower, October 15, 2024, accessed May 17, 2025, https://www.sec.gov/files/fy24-annual-whistleblower-report.pdf.

[14] "Rounds Introduces Legislation to Prevent Smuggling of American AI Chips Into China," Mike Rounds, April 10, 2025, https://www.rounds.senate.gov/newsroom/press-releases/rounds-introduces-legislation-to-prevent-smuggling-of-american-ai-chips-into-china.

[15] "Supplement No. 1 to Part 766—Guidance on Charging and Penalty Determinations in Settlement of Administrative Enforcement Cases," Bureau of Industry and Security, https://www.bis.gov/ear/title-15/subtitle-b/chapter-vii/subchapter-c/part-766/supplement-no-1-part-766-guidance.

[16] Richard Blumenthal, *The U.S. Technology Fueling Russia's War in Ukraine: Examining the Bureau of Industry and Security's Enforcement of Semiconductor Export Controls*, Majority Staff report, December 18, 2024, https://www.hsgac.senate.gov/wp-content/uploads/The-U.S.-Technology-Fueling-Russias-War-in-Ukraine-Examing-BISs-Enforcement-of-Semiconductor-Export-Controls.pdf.

[17] See Figure 11.

[18] Samuel Hammond, "The Scramble for AI Computing Power," American Affairs, https://americanaffairsjournal.org/2024/05/the-scramble-for-ai-computing-power/; "Office of Enforcement Analysis (OEA)," Bureau of Industry and Security, https://www.bis.gov/about-bis/bis-leadership-and-offices/oea.

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

[19] Michael McCaul, "Bureau of Industry & Security: 90-Day Review Report," Foreign Affairs Committee, https://foreignaffairs.house.gov/wp-content/uploads/2024/01/1.2.24-BIS-Report.pdf.

[20] James Byrne et al., "Silicon Lifeline," Royal United Services Institute, August 2022, https://static.rusi.org/RUSI-Silicon-Lifeline-final-web.pdf.

[21] Jeremy Daum, "What China's National Intelligence Law Says, and Why It Doesn't Matter," China Law Translate, February 22, 2024, https://www.chinalawtranslate.com/en/what-the-national-intelligence-law-says-and-why-it-doesnt-matter/.

[22] Tim Fist and Erich Grunewald, Preventing AI Chip Smuggling to China (Center for a New American Security, October 25, 2023), https://www.cnas.org/publications/reports/preventing-ai-chip-smuggling-to-china.

[23] "Special report: How U.S.-made chips are flowing into Russia."

[24] Potkin, "Exclusive: Huawei Aims to Mass-Produce Newest AI Chip in Early 2025, Despite US Curbs"; Elizabeth Olcott, Ryan McMorrow, and Tina Hu, "Huawei's Bug-Ridden Software Hampers China's Efforts to Replace Nvidia in AI," *Financial Times*, September 3, 2024, https://www.ft.com/content/3dab07d3-3d97-4f3b-941b-cc8a21a901d6; Chris Miller, "Why China Can't Export AI Chips," American Enterprise Institute, December 24, 2024, https://www.aei.org/foreign-and-defense-policy/why-china-cant-export-ai-chips/.

[25] Josh Ye, David Kirton, and Chen Lin, "Focus: Inside China's Underground Market for High-End Nvidia AI Chips," Reuters, June 20, 2023, https://www.reuters.com/technology/inside-chinas-underground-market-high-end-nvidia-ai-chips-2023-06-19/.

[26] Raffaele Huang and Liza Lin, "Chinese Buyers Are Ordering Nvidia's Newest AI Chips, Defying U.S. Curbs," WSJ, March 3, 2025, https://www.wsj.com/tech/china-nvidia-blackwell-chips-ai-531fed0c.

[27] Raffaele Huang, "The Underground Network Sneaking Nvidia Chips into China," Reuters, July 2, 2024, https://www.wsj.com/tech/the-underground-network-sneaking-nvidia-chips-into-china-f733aaa6; Swanson and Fu, "With Smugglers and Front Companies, China Is Skirting American A.I. Bans"; Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry"; McMorrow and Olcott, "Nvidia's AI Chips are Cheaper to Rent in China Than US."

[28] Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry."

[29] Swanson and Fu, "With Smugglers and Front Companies, China Is Skirting American A.I. Bans."

[30] Andy Lin et al., "Russia Is Getting Nvidia AI Chips from an Indian Pharma Company," Bloomberg, October 27, 2024, https://www.bloomberg.com/news/features/2024-10-27/russia-is-getting-nvidia-ai-chips-from-an-indian-pharma-company.

[31] McMorrow and Olcott, "Nvidia's AI Chips are Cheaper to Rent in China Than US."

[32] Swanson and Fu, "With Smugglers and Front Companies, China Is Skirting American A.I. Bans"; Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry."

[33] Sources: Josh Ye, David Kirton, and Chen Lin, "Focus: Inside China's Underground Market for High-End Nvidia AI Chips," Reuters, June 20, 2023, https://www.reuters.com/technology/inside-chinas-underground-market-high-end-nvidia-ai-chips-2023-06-19/; 辰壹 (Chen Yi), "美国出口禁令之下，20多万元的'天价芯片'流入黑市" (Under US Export Bans, "Sky-High Chips" Worth Over 200,000 Yuan Flow into Black Markets), 芯潮IC/芯观察 (IC Trends/Chip Observer), November 9, 2023, https://new.qq.com/rain/a/20231108A07I8P00; Eduardo Baptista, "China's military and government acquire Nvidia chips despite US ban," Reuters, January 15, 2024, https://www.reuters.com/technology/chinas-military-government-acquire-nvidia-chips-despite-us-ban-2024-01-14/; Raffaele Huang, "The Underground Network Sneaking Nvidia Chips into China," Reuters, July 2, 2024, https://www.wsj.com/tech/the-underground-network-sneaking-nvidia-chips-into-china-f733aaa6; Ana Swanson and Claire Fu, "With Smugglers and Front Companies, China Is Skirting American A.I. Bans," The New York Times, August 4, 2024, https://www.nytimes.com/2024/08/04/technology/china-ai-microchips.html; Ana Swanson, "Takeaways From Our Investigation Into Banned A.I. Chips in China," The New York Times, August 4, 2024, https://www.nytimes.com/2024/08/04/technology/china-ai-microchips-takeaways.html; Qianer Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry," The Information, August 12, 2024, https://www.theinformation.com/articles/nvidia-ai-chip-smuggling-to-china-becomes-an-industry.

[34] For the purposes of this analysis, we restricted searches to NVIDIA H100 and H200 products, which were the most popular AI data center products released after U.S. export controls were in place.

[35] See an overview of listings that contained images here: https://docs.google.com/spreadsheets/d/1eOYMu7va2TV1bi58EFivnUC__rVk9eYHJoCoowjiIdc

[36] Timothy Fist & Erich Grunewald, "AI Chip Smuggling Estimates (for 2024)," Google Colab, https://colab.research.google.com/drive/1ZxiwsHjc_aYh2N3Fzy2x7fg_nkjoYh2J.

[37] These figures use the 90 percent confidence interval of the distribution. The estimates assume model FLOP/s utilization (MFU) figures of 12% for inference for the H100 and Ascend 910C, 80% for inference for the H20, 40% for training for the H100 and Ascend 910C, and 60% for training for the H20. We assume developers can achieve similar MFU figures for the Ascend 910C to the H100, given their similar ratios of compute to bandwidth. Chao Jin et al., "MegaScale-MoE: Large-Scale Communication-Efficient Training of Mixture-of-Experts Models in Production," ByteDance, May 19, 2025, https://www.arxiv.org/pdf/2505.11432v2; Dylan Patel and Daniel Nishball, "100,000 H100 Clusters: Power, Network Topology, Ethernet vs InfiniBand, Reliability, Failures, Checkpointing," SemiAnalysis, June 17, 2024, https://semianalysis.com/2024/06/17/100000-h100-clusters-power-network.

[38] Allen, "DeepSeek, Huawei, Export Controls, and the Future of the U.S.-China AI Race"; Fanny Potkin & Che Pan, "Exclusive: Nvidia's H20 chip orders jump as Chinese firms adopt DeepSeek's AI models, sources say," Reuters, February 24, 2025, https://www.reuters.com/technology/artificial-intelligence/nvidias-h20-chip-orders-jump-chinese-firms-adopt-deepseeks-ai-models-sources-say-2025-02-25/; Lennart Heim, "Huawei's next AI Accelerator: Ascend 910C," March 12, 2025, https://blog.heim.xyz/huawei-ascend-910c/

[39] Fist & Grunewald, "AI Chip Smuggling Estimates (for 2024)"

[40] Potkin & Pan, "Exclusive: Nvidia's H20 chip orders jump as Chinese firms adopt DeepSeek's AI models, sources say"

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

[41] Hugo Meijer, Trading with the Enemy: The Making of US Export Control Policy Toward the People's Republic of China (New York: Oxford University Press, 2016)

[42] Special Report: How U.S.-Made Chips Are Flowing into Russia," Nikkei Asia, April 11, 2023, https://asia.nikkei.com/Business/Tech/Semiconductors/Special-report-How-U.S.-made-chips-are-flowing-into-Russia.

[43] Ryan Fedasiuk, Karson Elmgren, and Ellen Lu, "Silicon Twist: Managing the Chinese Military's Access to AI Chips" (Center for Security and Emerging Technology, 2022), https://cset.georgetown.edu/publication/silicon-twist/.

[44] "Export Control Officer Program Home," Bureau for Industry and Security, https://www.bis.doc.gov/index.php/2012-06-25-20-44-31/2012-06-25-20-45-50/export-control-officer-program-home; "Export Control Officer Areas of Responsibility," Bureau for Industry and Security, https://www.bis.doc.gov/index.php/2012-06-25-20-44-31/2012-06-25-20-45-50/export-control-officer-program-home?layout=edit&id=2047.

[45] If smugglers sell chips for 1.3 times more than the initial selling price (median price relative to manufacturer's suggested retail price [MSRP] from known reports outlined in Figure 2), with 100,000 smuggled chips per year, smugglers collectively have a gross profit of $900 million annually for operations. This is more than 16 times the BIS's entire annual budget for export control enforcement ($55 million in 2024). A single shipment of 800 servers (each containing eight chips) would be sufficient for a smuggler's budget to match the BIS's annual export enforcement budget. "Fiscal Year 2025 President's Budget Request," March 31, 2024, https://www.commerce.gov/sites/default/files/2024-03/BIS-FY2025-Congressional-Budget-Submission.pdf.

[46] Jaime Sevilla et al., *Can AI Scaling Continue Through 2030?* (Epoch AI, 2024), https://epochai.org/blog/can-ai-scaling-continue-through-2030.

[47] Chenggang Zhao et al., "Insights Into DeepSeek-V3: Scaling Challenges and Reflections on Hardware for AI Architectures," arXiv.org, May 14, 2025, https://arxiv.org/abs/2505.09343; DeepSeek-Ai et al., "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning," arXiv.org, January 22, 2025, https://arxiv.org/abs/2501.12948.

[48] Ege Erdil. 2025. "How Has DeepSeek Improved the Transformer Architecture?" *Gradient AI* (blog), Epoch AI, January 17, 2025, https://epoch.ai/gradient-updates/how-has-deepseek-improved-the-transformer-architecture; Lizzi C. Lee, "DeepSeek and the Strategic Limits of U.S. Sanctions," The Wire China, January 26, 2025, https://www.thewirechina.com/2025/01/26/deepseek-and-the-strategic-limits-of-u-s-sanctions/; Angela Zhang, "US Export Controls Have Forced Chinese Tech Companies to Be More Innovative," *Financial Times*, January 23, 2025, https://www.ft.com/content/c99d86f0-2d17-49d0-8dc6-9662ed34c831.

[49] "NVIDIA A800 China-Tailored GPU Performance Within 70% of A100," TechPowerUp, https://www.techpowerup.com/308324/nvidia-a800-china-tailored-gpu-performance-within-70-of-a100; Stephen Nellis, Josh Ye and Jane Lee, "Focus: China's AI Industry Barely Slowed by US Chip Export Rules," Reuters, May 3, 2023 https://www.reuters.com/technology/chinas-ai-industry-barely-slowed-by-us-chip-export-rules-2023-05-03/.

[50] Lennart Heim and Sihao Huang, "The Rise of DeepSeek: What the Headlines Miss," *Lennart Heim* (blog), January 25, 2025, https://blog.heim.xyz/deepseek-what-the-headlines-miss/.

[51] Jordan Schneider et al., "Deepseek: The Quiet Giant Leading China's AI Race," *ChinaTalk* (blog), November 27, 2024, https://www.chinatalk.media/p/deepseek-ceo-interview-with-chinas.

[52] Jarred Walton, "Nvidia's Next-Gen AI GPU Is 4X Faster Than Hopper: Blackwell B200 GPU Delivers up to 20 Petaflops of Compute and Other Massive Improvements," Tom's Hardware, March 18, 2024, https://www.tomshardware.com/pc-components/gpus/nvidias-next-gen-ai-gpu-revealed-blackwell-b200-gpu-delivers-up-to-20-petaflops-of-compute-and-massive-improvements-over-hopper-h100.

[53] "NVIDIA H100 Tensor Core GPU Architecture Overview," NVIDIA, 2023, https://resources.nvidia.com/en-us-tensor-core/gtc22-whitepaper-hopper; Dylan Patel, Daniel Nishball, and Myron Xie, "Nvidia's New China AI Chips Circumvent US Restrictions | H20 Faster Than H100 | Huawei Ascend 910B," SemiAnalysis, November 5, 2024, https://semianalysis.com/2023/11/09/nvidias-new-china-ai-chips-circumvent/; Jacob Feldgoise and Hanna Dohmen, "Pushing the Limits: Huawei's AI Chip Tests U.S. Export Controls," Center for Security and Emerging Technology, June 28, 2024, https://cset.georgetown.edu/publication/pushing-the-limits-huaweis-ai-chip-tests-u-s-export-controls/.

[54] Eleanor Olcott, Ryan McMorrow, and Tina Hu, "Huawei's Bug-Ridden Software Hampers China's Efforts to Replace Nvidia in AI", *Financial Times*, September 3, 2024 https://www.ft.com/content/3dab07d3-3d97-4f3b-941b-cc8a21a901d6.

[55] J. C. Sharman, "What Are Anonymous Shell Companies?" in *The Money Laundry* (Cornell University Press, 2011), posted on Cornell University Press, October 16, 2019, https://www.cornellpress.cornell.edu/what-are-anonymous-shell-companies/; Gregory C. Allen, Emily Benson, and William Alan Reinsch, "Improved Export Controls Enforcement Technology Needed for U.S. National Security," November 30, 2022, https://www.csis.org/analysis/improved-export-controls-enforcement-technology-needed-us-national-security.

[56] Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry"; Swanson and Fu, "With Smugglers and Front Companies, China Is Skirting American A.I. Bans."

[57] Committee on Homeland Security and Governmental Affairs Permanent Subcommittee on Investigations Majority Staff, "February 27, 2024 Hearing on 'The U.S. Technology Fueling Russia's War in Ukraine: How and Why,'" memorandum, February 21, 2024, https://www.hsgac.senate.gov/wp-content/uploads/2024.2.21-PSI-Staff-Memo-to-Members-on-Sanctions-Hearing.pdf.

[58] NVIDIA, *Form 10-K* (annual report for the fiscal year ended January 26 2025), filed February 26 2025, U.S. Securities and Exchange Commission, https://www.sec.gov/Archives/edgar/data/1045810/000104581025000023/nvda-20250126.htm.

[59] NVIDIA has stated that Singapore appeared as a big revenue source in its records because customers used it to handle billing, but its actual chip shipments to Singapore were less than 2% of that year's total revenue. NVIDIA, *Form 10-K*, fiscal year ended January 26, 2025. According to the Stanford 2025 AI Index Report, Singapore had 239 AI companies founded between 2013 and 2024, Indonesia had 19, Malaysia and Vietnam had 15 each, Thailand had 7, and the Philippines had 5. Nestor Maslej et al., *The AI Index 2025 Annual Report* (AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, 2025), April 2025, https://hai.stanford.edu/ai-index/2025-ai-index-report.

[60] Kok, "Singapore Charges Three With Fraud That Media Link to Nvidia Chips"; Reuters, "Singapore Prosecutors Say US Server Fraud Case Involves $390 Million of Transactions."

[61] "浪潮 [Inspur] NF5688A7HGX H100 8 GPU 80GB," ICQQG, September 9, 2024, https://archive.ph/7yCHP.

[62] Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry."

[63] Sastry et al., *Computing Power and the Governance of AI* (Center for the Governance of AI, February 14, 2024), https://www.governance.ai/post/computing-power-and-the-governance-of-ai.

[64] Liu, Nvidia AI Chip Smuggling to China Becomes an Industry."

[65] Erich Grunewald and Michael Aird, *AI Chip Smuggling into China: Potential Paths, Quantities, and Countermeasures* (Institute for AI Policy and Strategy, October 4, 2023), https://www.iaps.ai/research/ai-chip-smuggling-into-china.

[66] Bloomberg L.P. (2024). Retrieved October 10, 2024 from Bloomberg terminal.

[67] McMorrow and Olcott, "Nvidia's AI Chips Are Cheaper to Rent in China Than US."

[68] Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry."

[69] "Super Micro: Fresh Evidence of Accounting Manipulation, Sibling Self-Dealing and Sanctions Evasion at This AI High Flyer," Hindenburg Research, August 27, 2024, https://hindenburgresearch.com/smci/.

[70] Tim Fist, "Searching 'H100' on Baidu leads to interesting results," X (formerly Twitter), April 19, 2024, https://x.com/fiiiiiist/status/1781437325513203987.

[71] Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry."

[72] Penalties, U.S. Code, 50 U.S. Code § 4819, accessed September 9, 2024, https://www.law.cornell.edu/uscode/text/50/4819; "Supplement No. 1 to Part 766—Guidance on Charging and Penalty Determinations in Settlement of Administrative Enforcement Cases," Title 15, Subtitle B, Chapter VII, Subchapter C, Electronic Code of Federal Regulations, accessed September 9, 2024, https://www.ecfr.gov/current/title-15/subtitle-B/chapter-VII/subchapter-C/part-766/appendix-Supplement%20No.%201%20to%20Part%20766.

[73] Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry."

[74] Lennart Heim, et al., "Computing Power and the Governance of AI," GovAI (blog), February 14, 2024, https://www.governance.ai/post/computing-power-and-the-governance-of-ai.

[75] Asher Brass and Onni Aarne, *Location Verification for AI Chips* (Institute for AI Policy and Strategy, 2024), https://www.iaps.ai/research/location-verification-for-ai-chips.

[76] Onni Aarne, Tim Fist, and Caleb Withers, *Secure, Governable Chips: Using On-Chip Mechanisms to Manage National Security Risks from AI & Advanced Computing* (Center for a New American Security, 2024), https://www.cnas.org/publications/reports/secure-governable-chips; Gabriel Kulp et al., *Hardware-Enabled Governance Mechanisms: Developing Technical Solutions to Exempt Items Otherwise Classified Under Export Control Classification Numbers 3A090 and 4A090* (RAND Corporation, 2024), https://www.rand.org/pubs/working_papers/WRA3056-1.html; Tim Fist, Tao Burga, and Vivek Chilukuri, *Technology to Secure the AI Chip Supply Chain: A Working Paper* (Center for a New American Security, December 11, 2024), https://www.cnas.org/publications/reports/technology-to-secure-the-ai-chip-supply-chain-a-primer.

[77] For example, NVIDIA's operating expenses for sales and administrative functions amounted to $4.4 billion in fiscal year 2024, about 20 times the BIS's entire budget. Its net income surged by nearly 600 percent in 2024, from $4 billion in 2023 to $30 billion that year. "2024 NVIDIA Corporation Annual Review," NVIDIA, May 14, 2024, https://s201.q4cdn.com/141608511/files/doc_financials/2024/ar/NVIDIA-2024-Annual-Report.pdf.

[78] Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry."

[79] Mary Thornton, "The Critical Effort to Combat Illicit Chip Diversion," Semiconductor Industry Association, October 14, 2024, https://www.semiconductors.org/the-critical-effort-to-combat-illicit-chip-diversion/.

[80] Thornton, "The Critical Effort to Combat Illicit Chip Diversion."

[81] "2024 NVIDIA Corporation Annual Review."

[82] "NVIDIA's Data Center Business Fuels Explosive Growth in FY2Q25 Revenue; H200 Set to Dominate AI Server Market from 2H24, Says TrendForce," TrendForce, September 3, 2024, https://www.trendforce.com/presscenter/news/20240903-12283.html; "NVIDIA Corp (NVDA) Stock Analysis and Forecast for 2024," RoboForex, accessed September 10, 2024, https://roboforex.com/beginners/analytics/forex-forecast/stocks/stocks-forecast-nvda-2024/.

[83] Fist and Grunewald, "Preventing AI Chip Smuggling to China."

[84] U.S. Senate Committee on Homeland Security & Governmental Affairs, "Improving Export Controls Enforcement," Subcommittee on Emerging Threats and Spending Oversight, April 10, 2024, https://www.hsgac.senate.gov/subcommittees/etso/hearings/improving-export-controls-enforcement/.

[85] Brass and Aarne, "Location Verification for AI Chips."

[86] Brass and Aarne, "Location Verification for AI Chips."

[87] Ideally, AI chip designers would collaborate on a standard protocol for location verification, so that the same landmark servers can be used for AI chips by different companies. But as a start, NVIDIA could implement these measures alone.

[88] "Cotton Introduces Bill to Prevent Diversion of Advanced Chips to America's Adversaries and Protect U.S. Product Integrity," Tom Cotton, May 9, 2025, https://www.cotton.senate.gov/news/press-releases/cotton-introduces-bill-to-prevent-diversion-of-advanced-chips-

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc

to-americas-adversaries-and-protect-us-product-integrity; "Chairman Moolenaar, Bipartisan Lawmakers Unveil Bill to Stop AI Chip Smuggling to China," Select Committee on the CCP, May 15, 2025, https://selectcommitteeontheccp.house.gov/media/press-releases/chairman-moolenaar-bipartisan-lawmakers-unveil-bill-stop-ai-chip-smuggling..

[89] U.S. Government Accountability Office. *Counternarcotics: DOD Should Improve Coordination and Assessment of Its Activities.* GAO-24-106281. Washington, DC: Government Accountability Office, 2024. https://www.gao.gov/assets/gao-24-106281.pdf; George Bunn, "The Nuclear Nonproliferation Treaty: History and Current Problems," Arms Control Today, December 2003, https://www.armscontrol.org/act/2003-12/features/nuclear-nonproliferation-treaty-history-and-current-problems; and "How We Work," Office of the Director of National Intelligence, National Counterproliferation and Biosecurity Center, accessed September 10, 2024, https://www.dni.gov/index.php/ncbc-how-we-work.

[90] "Office of Enforcement Analysis (OEA)," U.S. Department of Commerce, Bureau of Industry and Security, accessed September 10, 2024, https://www.bis.gov/OEA.

[91] "Office of Enforcement Analysis (OEA)."

[92] Other similar whistleblower programs include the Anti-Money Laundering Whistleblower Program administered by the Financial Crimes Enforcement Network, a bureau of the Treasury Department; the Commodity Futures Trading Commission Whistleblower Program, aimed at preventing violations of federal securities laws; and the Internal Revenue Service Whistleblower Program, aimed at reducing tax fraud. "Rounds Introduces Legislation to Prevent Smuggling of American AI Chips Into China."

[93] Allison Herren Lee, "A Proven Success: The SEC Whistleblower Regime Provides a Roadmap for DOJ's New Program," Harvard Law School Forum on Corporate Governance, April 25, 2024, https://corpgov.law.harvard.edu/2024/04/25/a-proven-success-the-sec-whistleblower-regime-provides-a-roadmap-for-dojs-new-program/.

[94] Securities and Exchange Commission Office of the Whistleblower, "Securities and Exchange Commission Office of the Whistleblower Annual Report to Congress for Fiscal Year 2024."

[95] Megan Mocho and Sean Belanger, "DOJ's Annual False Claims Act Statistics: Relators Forging Ahead with Success," February 14, 2023, https://www.hklaw.com/en/insights/publications/2023/02/dojs-annual-false-claims-act-statistics.

[96] Matthew S. Axelrod, "Memorandum for All Export Enforcement Employees: Clarifying Our Policy Regarding Voluntary Self-Disclosures Concerning Others," U.S. Department of Commerce, Bureau of Industry and Security, April 18, 2023, https://www.bis.doc.gov/index.php/documents/enforcement/3262-vsd-policy-memo-04-18-2023/file.

[97] "Whistleblower Frequently Asked Questions," U.S. Securities and Exchange Commission, accessed September 10, 2024, https://www.sec.gov/enforcement-litigation/whistleblower-program/whistleblower-frequently-asked-questions#faq-16.

[98] Industry whistleblowers historically have been subjected to income loss and even death threats: "Whistleblower Retaliation Cases and Settlements," The Employment Law Group, accessed September 10, 2024, https://www.employmentlawgroup.com/resources/notable-cases/whistleblower-retaliation-cases-settlements/; James Bikales, "'I Was Told, Frankly, to Shut up,' Boeing Whistleblower Tells Senate," *POLITICO*, April 17, 2024, https://www.politico.com/live-updates/2024/04/17/congress/boeing-whistleblower-senate-shut-up-safety-planes-00152805; Marie Brenner, "The Man Who Knew Too Much," Vanity Fair, May 1996, https://www.vanityfair.com/magazine/1996/05/wigand199605.

[99] Ashley Lin, "Data Brief: Whistleblower Programs for AI", unpublished report; Grace Schepis, "IRS Whistleblower Program's Annual Report: Long Waits, Low Rewards," Whistleblower Network News, July 5, 2023, https://whistleblowersblog.org/corporate-whistleblowers/tax-whistleblowers/irs-whistleblower-programs-annual-report-long-waits-low-rewards/.

[100] U.S. Department of Justice, "Export Control and Sanctions," National Security Division, accessed September 10, 2024, https://www.justice.gov/nsd/export-control-and-sanctions.

[101] "Rewards for Non-U.S. Whistleblowers," National Whistleblower Center, accessed September 10, 2024, https://www.whistleblowers.org/rewards-for-non-u-s-whistleblowers/.

[102] "Fiscal Year 2025 President's Budget Request," U.S. Department of Commerce, Bureau of Industry and Security, March 31, 2024, https://www.commerce.gov/sites/default/files/2024-03/BIS-FY2025-Congressional-Budget-Submission.pdf.

[103] Alex W. Palmer, "'An Act of War': Inside America's Silicon Blockade Against China," *The New York Times Magazine*, July 12, 2023, https://www.nytimes.com/2023/07/12/magazine/semiconductor-chips-us-china.html.

[104] Justin Hayward, "How Much Does an F-35 Cost?" Simple Flying, December 20, 2023, https://simpleflying.com/how-much-does-an-f-35-cost/.

[105] The Export Control Reform Act allows civil penalties of up to two times the value of the transaction involved. As mentioned above, there have been multiple reports of smuggling of AI chips worth over $100 million.

[106] In a statement about the penalty, Seagate's chief executive officer said: "While we believed we complied with all relevant export control laws at the time we made the hard disk drive sales at issue, we determined that engaging with BIS and settling this matter was the best course of action." Karen Freifeld, "Seagate to Pay $300 Million Penalty for Shipping Huawei 7 Million Hard Drives," Reuters, April 20, 2023, https://www.reuters.com/legal/seagate-settles-with-us-shipping-11-bln-hard-drives-huawei-2023-04-19/; "Seagate Reaches Resolution with the U.S. Department of Commerce's Bureau of Industry and Security," Seagate, April 19, 2023, https://investors.seagate.com/news/news-details/2023/Seagate-Reaches-Resolution-With-the-U.S.-Department-of-Commerces-Bureau-of-Industry-and-Security/default.aspx.

[107] The proposed budget increase is $122 million. An NVIDIA H100 chip costs about $30,000, and the BIS can impose penalties of up to two times the value of the transaction involved. That means a transaction of 2,100 NVIDIA H100s could result in a penalty of $126 million.

[108] Allen, Benson, and Reinsch, "Improved Export Controls Enforcement Technology Needed for U.S. National Security."

[109] Timothy Fist & Erich Grunewald, "AI Chip Smuggling Estimates (for 2024)," Google Colab, https://colab.research.google.com/drive/1ZxiwsHjc_aYh2N3Fzy2x7fg_nkjoYh2J.

[110] Huang, "The Underground Network Sneaking Nvidia Chips into China"; Swanson and Fu, "With Smugglers and Front Companies, China Is Skirting American A.I. Bans"; Liu, "Nvidia AI Chip Smuggling to China Becomes an Industry."

[111] Huang, "The Underground Network Sneaking Nvidia Chips into China."

[112] Huang, "The Underground Network Sneaking Nvidia Chips into China."

[113] Karen Freifeld, "Seagate to Pay $300 Million Penalty for Shipping Huawei 7 Million Hard Drives," Reuters, April 20, 2023, https://www.reuters.com/legal/seagate-settles-with-us-shipping-11-bln-hard-drives-huawei-2023-04-19/; "ByteDance and Alibaba Place Massive GPU Orders with NVIDIA, Fueling the AI Race," Pandaily, Jun 14, 2023, https://web.archive.org/web/20250131203723/https://pandaily.com/bytedance-and-alibaba-place-massive-gpu-orders-with-nvidia-fueling-the-ai-race/.

[114] "Part 743—Special Reporting and Notification," Title 15, Subtitle B, Chapter VII, Subchapter C, Electronic Code of Federal Regulations, accessed October 6, 2024. https://www.ecfr.gov/current/title-15/subtitle-B/chapter-VII/subchapter-C/part-743.

[115] "§ 743.3—Thermal Imaging Camera Reporting," Title 15, Subtitle B, Chapter VII, Subchapter C, Electronic Code of Federal Regulations, accessed October 6, 2024, https://www.ecfr.gov/current/title-15/subtitle-B/chapter-VII/subchapter-C/part-743/section-743.3.

[116] "Part 772—Definitions of Terms," Title 15, Subtitle B, Chapter VII, Subchapter C, Electronic Code of Federal Regulations, accessed October 6, 2024, https://www.ecfr.gov/current/title-15/subtitle-B/chapter-VII/subchapter-C/part-772.

Center for a New American Security
1701 Pennsylvania Ave NW, Suite 700, Washington, DC 20006
T: 202.457.9400 | F: 202.457.9401 | CNAS.org | @CNASdc