

Creating a Public Cloud through the Defense Production Act

Joel Dodge



Vanderbilt
Policy Accelerator
for Political Economy & Regulation

September 2024

Introduction

Cloud computing is an integral component of modern digital infrastructure and the artificial intelligence tech stack.¹ AI relies on cloud infrastructure for its computing power, and cloud computing is powered by advanced semiconductor chips (GPUs). As Ganesh Sitaraman and Tejas Narechania have shown in a VPA white paper, these lower layers in the AI tech stack are increasingly concentrated among just a few firms.² They and others have argued that a public option for cloud computing would inject competition and supply chain resilience into the industry, improve national security, and support research and AI development for public purposes.³

To create its own public cloud infrastructure, the federal government would need to obtain GPUs, among other essential inputs.⁴ Congress could, of course, appropriate funding in the ordinary appropriations process. But if it does not, the President can still initiate the creation of a public cloud using authorities in the Defense Production Act of

¹ Ganesh Sitaraman & Natalie Foster, *How States Can Keep Big Tech from Dominating AI*, POLITICO (Mar. 13, 2024), <https://www.politico.com/news/magazine/2024/03/13/states-big-tech-ai-00146338>

² See Tejas Narechania & Ganesh Sitaraman, *An Antimonopoly Approach to Governing Artificial Intelligence*, VANDERBILT POLICY ACCELERATOR 31, 46-47 (2023) (forthcoming in YALE L. & POL'Y R.), <https://cdn.vanderbilt.edu/vu-sub/wp-content/uploads/sites/281/2023/12/19174621/Narechania-Sitaraman-Antimonopoly-AI-2023.10.6.pdf>; see also Ganesh Sitaraman & Tejas Narechania, *Antimonopoly Tools for Regulating Artificial Intelligence*, VANDERBILT POLICY ACCELERATOR (2023), <https://cdn.vanderbilt.edu/vu-sub/wp-content/uploads/sites/281/2023/12/19183408/Policy-Brief-2023.10.08-.pdf>.

³ See Narechania & Sitaraman, *supra* note 2, at 46-47. For proposals for a public option for AI, see Ben Gansky, Michael Martin & Ganesh Sitaraman, *Artificial Intelligence is Too Important to Leave to Google and Facebook Alone*, N.Y. TIMES (Nov. 10, 2019), <https://www.nytimes.com/2019/11/10/opinion/artificial-intelligence-facebook-google.html>; Bruce Schneier & Nathan E. Sanders, *Build AI by the People, for the People*, FOREIGN POL'Y (June 12, 2023), <https://foreignpolicy.com/2023/06/12/ai-regulation-technology-us-china-eu-governance/>.

⁴ This white paper focuses on government acquisition of GPUs. However, similar reasoning could apply to other cloud inputs, such as interconnect cabling and less-advanced chips. See Jai Vipra & Sarah Myers West, *Computational Power and AI*, AI NOW (Sept. 27, 2023), <https://ainowinstitute.org/publication/policy/compute-and-ai#h-what-is-compute-and-why-does-it-matter>.

1950, which gives the President broad powers to intervene in the economy to ensure the supply of strategically-important resources.

As detailed below, the DPA can be used to acquire and pay for GPUs to create a public cloud. GPUs are essential inputs for cloud computing and artificial intelligence research, which have useful applications for both national defense and renewable energy. A presidential administration can therefore use the DPA to both purchase GPUs and to assert its priority over other buyers, and use the DPA Fund to pay for GPUs. While the DPA provides authority for the administration to obtain GPUs to create a public cloud, Congress and the President can each take additional recommended actions to better facilitate this type of public project.

I. The DPA Can be Used to Acquire GPUs.

The DPA gives the President broad authority to secure the provision of strategically-important materials. Two of the Act's authorities can help create a public cloud: Title III's "other Presidential action" transaction authority, and Title I's "priorities and allocation" authority.

A. *Title III of the DPA authorizes Presidential action to purchase GPUs.*

Title III of the DPA provides that: "[t]o create, maintain, protect, expand, or restore domestic industrial base capabilities essential for the national defense, the President may make provision ... for purchases of or commitments to purchase an industrial resource or a critical technology item, for Government use or resale."⁵ Under the Act, "national defense" encompasses, among other functions, "critical infrastructure protection and restoration."⁶ "Industrial resources" include "materials, services, processes, or manufacturing equipment ... needed to establish or maintain an efficient and modern national defense industrial base."⁷ A "critical technology item" includes any item related to "any technology designated by the President to be essential to the national defense."⁸ And "critical infrastructure" includes "any systems and assets, whether physical or cyber-based, so vital to the United States that the degradation or

⁵ 50 U.S.C. § 4533(a)(1).

⁶ *Id.* § 4552(14).

⁷ *Id.* § 4552(12).

⁸ *Id.* §§ 4552(3), (4).

destruction of such systems and assets would have a debilitating impact on national security[.]”⁹

GPUs likely qualify as a national defense industrial resource or a critical technology item. GPUs are essential inputs for artificial intelligence systems, including defense-related applications. Federal spending on artificial intelligence has soared in recent years to \$4.6 billion in 2023, nearly all of which was initiated by the Department of Defense across a range of national security-related applications.¹⁰ In a 2022 report, the President’s Council of Advisors on Science and Technology found that semiconductors are “critical to U.S. economic and national security.”¹¹ The information technology sector, including cloud computing, has also been deemed one of the sixteen “critical infrastructure” sectors in the economy “central to the nation’s security, economy, and public health and safety” by a 2013 presidential directive and the Cybersecurity and Infrastructure Security Agency.¹²

Deploying Title III to secure cyber infrastructure like cloud computing is also consistent with Congress’s 2003 amendments to the Act. That year Congress expanded “national defense” to include “critical infrastructure protection and restoration,” and defined “critical infrastructure” to include “cyber-based” systems and assets.¹³ The Report of the Senate Committee on Banking, Housing, and Urban Affairs stated that based in part on the “conclusion of the Commission on Critical Infrastructure Protection, the Committee was convinced that amendments to the Defense Production Act were warranted.”¹⁴ That Commission directly identified critical information and communications infrastructure as including “computing ... equipment, software, [and] processes ... that support[] the processing, storage, and transmission of data and

⁹ *Id.* § 4552(2). This includes national economic security and national public health or safety. *Id.*

¹⁰ JACOB LARSON & JAMES S. DENFORD ET AL., BROOKINGS INST., THE EVOLUTION OF ARTIFICIAL INTELLIGENCE (AI) SPENDING BY THE U.S. GOVERNMENT (2024), <https://www.brookings.edu/articles/the-evolution-of-artificial-intelligence-ai-spending-by-the-u-s-government/>.

¹¹ *Report to the President: Revitalizing the U.S. Semiconductor Ecosystem*, PRESIDENT’S COUNCIL OF ADVISORS ON SCI. & TECH. (Sept. 2022), https://www.whitehouse.gov/wp-content/uploads/2022/09/PCAST_Semiconductors-Report_Sep2022.pdf.

¹² THE WHITE HOUSE, Presidential Policy Directive/PPD-21, Critical Infrastructure Security and Resilience (Feb. 12, 2013), <https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/Presidential-policy-directive-critical-infrastructure-security-and-resil>; see also *Information Technology Sector*, CYBERSECURITY & INFRASTRUCTURE SECURITY AGENCY, <https://www.cisa.gov/topics/critical-infrastructure-security-and-resilience/critical-infrastructure-sectors/information-technology-sector>.

¹³ S. Rep. No. 108-156, 3 (2003); see 50 U.S.C. § 4552(2).

¹⁴ S. Rep. No. 108-156, 3 (2003).

information.”¹⁵ Thus, cloud computing and GPUs fall comfortably within the scope of the underlying recommendation Congress relied on when it expanded the meaning of “national defense” in 2003.

Applying the DPA to GPUs is also consistent with prior use of the Act. In the 1980s, President Reagan used the DPA to support research into machine intelligence and microelectronics.¹⁶ Today the government continues to invest in microelectronics through the DPA.¹⁷

The President’s Title III authority can be subject to certain restrictions. First, the President must formally declare that (a) the resource is “essential to the national defense”; (b) absent DPA action, the resource would not be adequately supplied; and (c) Title III transactions are the most cost effective, expedient, and practical means of meeting the strategic need.¹⁸ Second, Title III actions exceeding \$50 million require congressional approval.¹⁹ However, these restrictions can be waived during a national emergency, or upon a presidential determination that a critical shortfall would “severely impair national defense capability.”²⁰ Indeed, as a practical matter, these requirements are frequently waived by presidential determination.²¹

¹⁵ CRITICAL FOUNDATIONS: PROTECTING AMERICA’S INFRASTRUCTURES, PRESIDENT’S COMM’N ON CRITICAL INFRASTRUCTURE PROT. 174 (1997), <https://nsarchive.gwu.edu/document/21584-document-02-President-s-commission-critical>.

¹⁶ DOUGLAS I. BELL, “A LITTLE-KNOWN BILL OF GREAT NATIONAL SIGNIFICANCE”: THE USES AND EVOLUTION OF THE DEFENSE PRODUCTION ACT, 1950-2020, U.S. ARMY HERITAGE & EDU. CTR. 23 (2020), https://ahec.armywarcollege.edu/documents/Defense_Production_Act_1950-2020.pdf.

¹⁷ “*Mission Critical: Restoring National Security as the Focus of Defense Production Act Reauthorization*,” *Hearing Before H. Comm. on Financial Servs.*, 118th Cong. 6 (Testimony of Luke A. Nicastro), <https://crsreports.congress.gov/product/pdf/TE/TE10092>.

¹⁸ 50 U.S.C. § 4533(a)(5).

¹⁹ *Id.* § 4533(a)(6)(C).

²⁰ *Id.* § 4533(a)(7).

²¹ *See, e.g.*, Memorandum on Presidential Determination Pursuant to Section 303 of the Defense Production Act of 1950, as amended, on Printed Circuit Boards and Advanced Packaging Production Capability, Presidential Determination No. 2023-06, 88 Fed. Reg. 19,545 (Mar. 31, 2023), <https://www.whitehouse.gov/briefing-room/Presidential-actions/2023/03/27/memorandum-on-presidential-determination-pursuant-to-section-303-of-the-defense-production-act-of-1950-as-amended-on-printed-circuit-boards-and-advanced-packaging-production-capability/>; Memorandum regarding the Presidential Determination and Waiver Pursuant to Section 303 of the Defense Production Act of 1950, as amended, on Essential Medicines, Medical Countermeasures, and Critical Inputs, Presidential Determination No. 2024-03, 89 Fed. Reg. 3 (Jan. 2, 2024), <https://www.whitehouse.gov/briefing-room/Presidential-actions/2023/12/27/memorandum->

Accordingly, the Administration could declare GPUs essential to national defense, and that security vulnerabilities (like cyber-attacks) among private cloud providers necessitate the creation of a public cloud. Due to persistent shortages, the Administration could determine that Title III action is necessary to secure the government's necessary supply of GPUs. (See Section I.B below.) Given these shortages, the Administration could conclude that Title III poses the most cost-effective and practical option to obtain GPUs. Alternatively, the President could simply determine that action to create a public cloud is necessary to avert an industrial resource or critical technology item shortfall that would severely impair national defense capability, and bypass these findings and the \$50 million limit on DPA transactions without congressional authorization.

Moreover, Administration action to create a public cloud would also advance the DPA's aims to support renewable energy. The DPA's legislative findings instruct the Administration that "to the maximum extent possible, domestic energy supplies should be augmented through reliance on renewable energy sources (including solar, geothermal, wind, and biomass sources)[.]"²² While the operation of large-scale AI systems imposes significant energy costs, AI applications also have the potential to help optimize the siting and operation of renewable energy projects, such as by identifying optimal locations for wind and solar farms, predicting wind speed and solar radiation to maximize energy generation, and assisting in geothermal production.²³ AI could also streamline renewable energy permitting.²⁴ Pursuant to a 2023 executive order, the Energy Department is studying how AI can improve planning and permitting.²⁵

regarding-the-Presidential-determination-and-waiver-pursuant-to-section-303-of-the-defense-production-act-of-1950-as-amended-on-essential-medicines-medical-countermeasures-and-critical/.
²² *Id.* § 4502(a)(6).

²³ DAVID SANDALOW & COLIN McCORMICK ET AL., ARTIFICIAL INTELLIGENCE FOR CLIMATE CHANGE MITIGATION ROADMAP, ICEF INNOVATION ROADMAP PROF. 38-39 (2023), <https://icef.go.jp/wp-content/uploads/2024/02/AI-Climate-Roadmap-ICEF-Dec-1-2023.pdf>.

²⁴ As Energy Secretary Jennifer Granholm said in March, "I do think that AI can really be helpful in helping to streamline a lot of [permitting processes]." @Axios, TWITTER (Mar. 6, 2024), <https://twitter.com/axios/status/1765398909336039737>.

²⁵ Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 88 Fed. Reg. 75,191 (Nov. 1, 2023), <https://www.whitehouse.gov/briefing-room/Presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>.

Therefore, because advanced GPUs have both national defense and renewable energy applications, the administration can likely rely on the DPA to acquire them for a public cloud.²⁶

B. Title I of the DPA authorizes the President to compel chipmakers to prioritize GPUs for public use.

Title I of the DPA authorizes the President to “require that performance under contracts or orders ... necessary or appropriate to promote the national defense shall take priority over performance under any other contract or order[.]”²⁷ This ensures that the government gets priority to have its contracts for critical materials fulfilled by private producers. To invoke Title I, the President must find that (1) the material “is a scarce and critical material essential to the national defense,” and (2) the “requirements of the national defense for such material cannot otherwise be met without creating a significant dislocation of the normal distribution of such material in the civilian market to such a degree as to create appreciable hardship.”²⁸

Title I could be used to ensure that the government gets priority access to GPUs. The artificial intelligence explosion has driven up demand for GPUs, leading to shortages.²⁹ Just as it did to address baby formula shortages in 2022, the Administration could invoke Title I upon determining that GPU shortages “threaten[] the continued functioning of the [AI] supply chain, undermining critical infrastructure that is essential to the national defense[.]”³⁰ In the near term, the government’s purchase of scarce

²⁶ Because Title III authorizes DPA transactions to be made “without regard to the limitations of existing law,” the Administration could purchase GPUs while bypassing procurement rules and other procedural requirements. 50 U.S.C. § 4533(b); see JOEL DODGE, JOEL MICHAELS, LENORE PALLADINO, & TODD N. TUCKER, ROOSEVELT INST., PROGRESSIVE PREEMPTION: HOW THE DEFENSE PRODUCTION ACT CAN OVERRIDE CORPORATE EXTRACTION, BOOST WORKER POWER, AND EXPEDITE THE CLEAN ENERGY TRANSITION (2022), https://rooseveltinstitute.org/wp-content/uploads/2022/12/RI_ProgressivePreemption_Brief_202212.pdf.

²⁷ 50 U.S.C. § 4511(a).

²⁸ *Id.* § 4511(b).

²⁹ Erin Griffin, *The Desperate Hunt for the A.I. Boom’s Most Indispensable Prize*, N.Y. TIMES (Aug. 16, 2023), <https://www.nytimes.com/2023/08/16/technology/ai-gpu-chips-shortage.html>.

³⁰ Delegating Authority Under the Defense Production Act To Ensure an Adequate Supply of Infant Formula, 87 Fed. Reg. 31,357 (May 18, 2022), <https://www.federalregister.gov/documents/2022/05/24/2022-11273/delegating-authority-under-the-defense-production-act-to-ensure-an-adequate-supply-of-infant-formula>.

chips could displace other important private-sector uses.³¹ In the long term, supply pressure may ease as existing firms and new CHIPS Act-funded chipmakers ramp up production to meet growing demand.³²

II. The DPA Fund Can Be Used to Pay for GPUs.

The DPA Fund is available to “carry out the provisions and purposes” of the DPA, subject to certain restrictions in the Act and congressional appropriations bills.³³ The Fund cannot exceed \$750 million, excluding appropriations for the current fiscal year.³⁴ In fiscal year 2023, Congress appropriated \$373 million for the DPA Fund; in fiscal year 2024, it appropriated \$588 million.³⁵

Because the Administration is authorized to purchase GPUs under Title III, it may use the DPA Fund to finance those purchases. With several hundred million dollars in the Fund available for use, the Administration could purchase thousands of Nvidia’s H100 GPUs or equivalent chips from other suppliers (assuming the President determined such action is necessary to prevent a critical shortfall impairing the national defense, allowing it to bypass congressional approval, as discussed above).³⁶

³¹ U.S. GEN. ACCOUNTING OFFICE, GAO-22-105923, SEMICONDUCTOR SUPPLY CHAIN: POLICY CONSIDERATIONS FROM SELECTED EXPERTS FOR REDUCING RISKS AND MITIGATING SHORTAGES 17 (2022), <https://www.gao.gov/assets/gao-22-105923.pdf> (invoking the DPA for defense needs may “come[] at the expense of some other customer that is already dealing with short supply...for instance, a manufacturer of lifesaving medical equipment, [or] of critical infrastructure equipment[.]”).

³² Policymakers may additionally consider creating a “Strategic Semiconductor Reserve” to mitigate future shortages and build economic resilience against supply shocks. *Cf.* Daleep Singh & Arnab Datta, *Reimagining the SPR*, FIN. TIMES (Feb. 24, 2024), <https://www.ft.com/content/e948ae78-cfec-43c0-ad5e-2ff59d1555e9> (proposing to “repurpos[e] the [Strategic Petroleum Reserve] as a ‘Strategic Resilience Reserve’ ... with a flexible mandate to bolster resilience against future commodity shocks.”).

³³ 50 U.S.C. § 4534(c).

³⁴ *Id.* § 4534(e).

³⁵ See Further Consolidated Appropriations Act of 2024, H.R. 2882, 118th Cong. (2024); ALEXANDRA G. NEENAN & LUKE A. NICASTRO, CONG. RSCH. SERV., R43767, THE DEFENSE PRODUCTION ACT OF 1950: HISTORY, AUTHORITIES, AND CONSIDERATIONS FOR CONGRESS 17 (2023), <https://crsreports.congress.gov/product/pdf/R/R43767>.

³⁶ Narechania & Sitaraman, *supra* note 2, at 11. For comparison, Microsoft needs 20,000 eight-chip GPUs to integrate ChatGPT into Bing, and incorporating AI into Google search may cost \$80 billion. *Id.*

Absent congressional action, the Administration could try to maximize its DPA funding by negotiating a discount on GPUs from producers (or even attempting to produce them itself) under statutory authority authorizing public patent use.³⁷ Regardless, the Administration can use the DPA Fund to pay for GPUs to launch a public cloud.

III. Recommendations

While the DPA provides authority for the government to create a public cloud, there are a number of steps policymakers could take to improve its efficacy. Congress could increase the value and flexibility of the DPA Fund, or could grant a waiver for GPUs from the DPA's transaction limits. The President could issue an executive order launching a public cloud. Each of these recommendations is discussed in detail below.

Legislative – Increase the value and flexibility of the DPA Fund. Congress should strongly consider increasing the DPA Fund and relaxing its transaction limits to better meet current strategic imperatives, like securing cloud computing technology. The \$750 million cap on the DPA Fund has not been adjusted in fifteen years, and the \$50 million limit on DPA actions without congressional approval has not been adjusted in ten years. These thresholds should each be increased to at least keep up with inflation, and should automatically adjust for inflation in the future.

Legislative – Grant a GPU Waiver. Alternatively, Congress could approve a limited waiver for GPUs from the \$50 million DPA transaction limit. It has previously granted waivers for other critical materials, including high-purity beryllium metals (lifting the aggregate transaction cap to \$85 million) and radiation-hardened electronics (lifting the aggregate transaction cap to \$200 million).³⁸ In recognition of the high cost of

³⁷ See, e.g., 28 U.S.C. § 1498; Susannah Glickman, *Semi-Politics*, PHENOMENAL WORLD (June 24, 2024), <https://www.phenomenalworld.org/analysis/semi-politics/> (proposing that “the US government could run a nationally-owned [semiconductor] fab”); cf. Joel Dodge, *The Government Can Legally Commandeer Drug Patents*, PEOPLE'S POLY PROJ. (Oct. 2, 2017), <https://www.peoplespolicyproject.org/2017/10/02/the-government-can-legally-commandeer-drug-patents/>. For a comprehensive survey of federal laws authorizing government patent use, see Laura E. Dolbow, *Public Patent Powers*, 122 MICH. L. REV. (forthcoming 2024), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4739492.

³⁸ See National Defense Authorization Act for Fiscal Year 2010, H.R. 2647, 111th Cong. (2009); Defense Production Act Reauthorization of 2003, S. 1680, 108th Cong. (2003).

advanced GPUs—Nvidia’s H100 GPU, for instance, can cost up to \$30,000 per chip³⁹—Congress could lift the transaction cap for GPU transactions to several hundred million dollars, and accordingly appropriate additional funding specifically for GPU transactions under the DPA.

Executive – Issue an executive order launching a public cloud pilot program. The President could issue an “Executive Order Securing American Cloud Computing Infrastructure” that instructs a federal agency to (a) use the DPA to purchase and pay for GPUs, and (b) create a pilot public cloud pursuant to laws such as the National Artificial Intelligence Initiative Act that authorize steps by the President to promote AI access.⁴⁰ Several different agencies could be plausible hosts for this initiative, including the Department of Commerce, Department of Defense, the Department of Energy, and the Department of Homeland Security’s Cybersecurity & Infrastructure Security Agency.

The Administration could also build upon its existing National AI Research Resource (“NAIRR”) pilot program, which has the potential to lead to a public option for cloud. NAIRR was first proposed through bipartisan legislation in 2023, and the legislative text leaves open the ability of the NAIRR to work through either a public cloud infrastructure or contracts with private firms.⁴¹ As is conventional with programmatic legislation, the NAIRR legislation merely authorizes creating the NAIRR, it does not appropriate funds for the federal government to acquire the GPUs needed for building a public cloud.

The Administration subsequently created a NAIRR pilot program by executive order.⁴² Operated by the National Science Foundation to broaden access to AI tools, the NAIRR pilot currently appears to be operating through private company cloud services.⁴³

³⁹ Alfonso Maruccia, *Nvidia generates up to 1,000% profit for each H100 GPU sold*, TECHSPOT (Aug. 18, 2023), <https://www.techspot.com/news/99839-nvidia-garners-reported-1000-profit-each-h100-gpu.html>.

⁴⁰ See 15 U.S.C. § 9411(b)(1) (instructing the President to ensure “[s]ustained and consistent support for artificial intelligence research and development through ... access to data and computing resources”).

⁴¹ CREATE AI Act of 2023, S. 2714, 118th Cong. (2023), <https://www.congress.gov/bill/118th-congress/senate-bill/2714/text>.

⁴² Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 88 Fed. Reg. 75,191 (Nov. 1, 2023), <https://www.whitehouse.gov/briefing-room/Presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>.

⁴³ See *National Artificial Intelligence Resource Pilot*, NAT’L SCI. FOUND., <https://new.nsf.gov/focus-areas/artificial-intelligence/nairr> (listing private sector partners including Amazon Web Services,

Therefore, the President could issue a new executive order expanding the NAIRR pilot program to create a public cloud, and invoking the DPA to provide the program with the GPUs necessary to create a cloud.

Conclusion

Artificial intelligence has the potential to transform our economy. With a few dominant firms controlling its infrastructure, both the profits and massively consequential decisionmaking around AI threaten to consolidate among a small number of private actors who have the power to charge high (or even cost-prohibitive) rents to AI start-ups and researchers. Providing public alternatives for AI infrastructure would help counteract this market power. The Defense Production Act provides the tools the government needs to acquire the hardware to launch a public cloud. Such a cloud may initially be primarily for government use, but could grow to become a full public option for cloud computing, available to government users and citizen researchers alike. That would help break the hold of dominant market actors on this critical infrastructure, and would democratize our AI future.

Google, and Microsoft); *see also* Amba Kak & Sarah Myers West, *The Problem With Public-Private Partnerships in AI*, FOREIGN POL'Y (Feb. 12, 2024), <https://foreignpolicy.com/2024/02/12/ai-public-private-partnerships-task-force-nairr/> (observing that “the largest tech companies have remained central beneficiaries of” NAIRR).