CUDA AND THE WALLED GARDEN: HOW THE RELATIONSHIP BETWEEN NVIDIA'S SOFTWARE AND HARDWARE VIOLATES THE SHERMAN ACT

William Davis[*]

ABSTRACT

This Note considers the role of antitrust law in the semiconductor industry—one marked by extreme levels of corporate concentration and power. This Note assesses the liability of Nvidia, a leading chip designer, under Section 2 of the Sherman Act. Nvidia has anticompetitively sought to maintain its monopoly position in the market for graphics processing units, or GPUs, through software licensing agreements. The licensing agreements for CUDA, Nvidia's software product, leave software developers with two options: incur switching costs to migrate their source code onto another firm's GPU hardware, or remain within the Nvidia ecosystem. As source code translators threatened to splinter the flywheel between the firm's software and hardware, the licensing agreement emerged as a legal tactic to protect the firm's market power. A monopolization suit under Section 2 of the Sherman Act, while costly, is likely the only means for the government to rejuvenate competition for GPUs. Nvidia's anticompetitive and illegal maintenance of this monopoly harms consumers and stalls innovation in a key industry of the future.

TABLE OF CONTENTS

---

[*] J.D., expected May 2025, The George Washington University Law School

INTRODUCTION

Semiconductors are foundational to modern life.[1] Silicon Valley—the metonym of American tech and innovation—gets its name from the silicon semiconductor[2], or chips, which power everything from digital watches to supercomputers and artificial intelligence, or AI.[3]

Demand for semiconductors has boomed following the launch of ChatGPT, believed to be the fastest-growing application in history.[4] ChatGPT, and systems like it, work because of incredibly fast, incredibly small semiconductors built at a scale of nanometers.[5] Training a system like ChatGPT requires an estimated ten thousand graphics processing units, or GPUs; running it for two years requires another million.[6] But only a few firms—Microsoft, Google, and Amazon—possess most of the chips needed to run these AI systems.[7] These firms lease out their computing power to other firms looking to train or run their own projects.[8] OpenAI, the firm behind ChatGPT, raised $10 billion from Microsoft and plans to use the money to then pay Microsoft for access to its data centers and computing

---

[1] "[N]early every economic activity, nearly every dollar of global GDP, relies in one way or another on the microscopic switches of semiconductors." Elizabeth Kolbert, The Real Cost of Plundering the Planet's Resources, New Yorker (Oct. 23, 2023), https://www.newyorker.com/magazine/2023/10/30/the-real-cost-of-plundering-the-planets-resources [https://perma.cc/4L5D-8ZVS].

[2] Zoë Bernard, *Here's the Story Behind How Silicon Valley Got Its Name*, BUS. INSIDER (Dec. 9, 2017, 10:30 AM), fhttps://www.businessinsider.com/how-silicon-valley-got-its-name-2017-12 [https://perma.cc/K2RQ-LDVX].

[3] *Semiconductors in Wearable Devices* Market, BUS. WIRE, (May 17, 2016, 1:16 PM); https://www.businesswire.com/news/home/20160517006455/en/Semiconductors-in-Wearable-Devices-Market---Wearable-Devices-Evolve-With-Semiconductor-Technology-Advancements---Research-and-Markets [https://perma.cc/HP2G-NWX3]; Yiwen Lu, *An A.I. Supercomputer Whirs to Life, Powered by Giant Computer Chips*, N.Y. TIMES (July 20, 2023), https://www.nytimes.com/2023/07/20/technology/an-ai-supercomputer-whirs-to-life-powered-by-giant-computer-chips.html [https://perma.cc/3S22-Y7AM].

[4] Krystal Hu, *ChatGPT Sets Record for Fastest-Growing User Base*, REUTERS (Feb. 2, 2023, 10:33 AM), https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/.

[5] Lucy Rodgers et al., *Inside the Miracle of Modern Chip Manufacturing*, FIN. TIMES (Feb. 28, 2024), https://ig.ft.com/microchips/?emailId=a5db66a8-a7c2-4042-9033-78a4b29d62a5&segmentId=13b7e341-ed02-2b53-e8c0-d9cb59be8b3b [https://perma.cc/TJE9-YB9R].

[6] These estimates are the requirement for GPUs, not all chips. Tracy Alloway & Joe Weisenthal, Odd Lots, *How to Build the Ultimate GPU Cloud to Power AI*, BLOOMBERG (July 20, 2023) (accessed on Spotify).

[7] Erin Griffith, *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 [https://perma.cc/T9QW-55UE].

[8] *Id.*

power.[9] Demand for these data centers and their computing power is extremely high; some start-ups must wait years for access.[10] The computing power that enables these AI systems comes from one type of semiconductor in particular: the GPU.[11] Companies are desperate to get their hands on GPUs[12] and one firm in particular has thrived from this demand: Nvidia.

Nvidia designs the GPUs used to train and run large language models and AI systems.[13] The company is the "chief beneficiary of the artificial intelligence boom"[14] and was the third US-listed firm to close a trading day with a two trillion dollar market capitalization.[15] Weeks later, it crossed another threshold, joining Microsoft and Apple across the three trillion dollar threshold.[16] The firm is estimated to control between eighty and ninety-five percent of the GPU market.[17] Competitors like Intel, AMD, and the large

---

[9] Cade Metz, *In the Age of A.I., Tech's Little Guys Need Big Friends*, N. Y. TIMES (July 5, 2023), https://www.nytimes.com/2023/07/05/business/artificial-intelligence-power-data-centers.html?action=click&module=RelatedLinks&pgtype=Article [https://perma.cc/TE9G-PFGW].

[10] Griffith, *supra* note 7.

[11] There are basically three different kinds of these chips: central processing units ("CPUs"), graphics processing units ("GPUs"), and data processing units ("DPUs"). CPUs run basically in a straight line and execute the basics of a computer system, like "calling up information from a hard drive…." GPUs do "parallel computing" and divide tasks among many processors at once. Basically, a GPU "break[s] complex problems into thousands or millions of separate tasks and work them out at once." Initially designed to support the graphics of high-quality video games, the GPU is now critical to modern supercomputing, data centers, and AI systems. DPUs, which move data throughout a data center, round out the "three pillars of computing." These three chips power everything from our cars to our phones to our washing machines. *See generally*, Brian Caulfield, *What's the Difference Between a CPU and a GPU?*, NVIDIA (Dec. 16, 2009), https://resources.nvidia.com/en-us-what-is-series/whats-the-difference [https://perma.cc/5JW6-959A] and *What's the Difference Between GPUs and CPUs?*, AMAZON, https://aws.amazon.com/compare/the-difference-between-gpus-cpus/ [https://perma.cc/NL2L-KP5T] (last accessed Sept. 26, 2024).

[12] "It's almost like talking about drugs: 'I know a guy who has H100s.'" Griffith, *supra* note 7.

[13] Chavi Mehta, *Explainer: Why are Nvidia's shares soaring and what is its role in the AI boom?,* REUTERS, (June 14, 2023), https://www.reuters.com/technology/why-are-nvidias-shares-soaring-what-is-its-role-ai-boom-2023-05-25/.

[14] Richard Waters, *AI boom catapults Nvidia into tech's big league*, FIN. TIMES, (Feb. 23, 2024), https://www.ft.com/content/1f8b317d-fcce-4f5b-9e54-8315e102ec10 [https://perma.cc/VH7V-WL9J].

[15] Harriet Clarfelt & Nicholas Megaw, *AI boom makes Nvidia third US stock to close above $2tn valuation,* FIN. TIMES (Mar. 1, 2024), https://www.ft.com/content/fd9d3719-31af-4d7c-9eb3-fce801e213fc [https://perma.cc/C7DW-7MLB].

[16] Colin Laidley, *Nvidia Market Cap Crosses $3 Trillion*, INVESTOPEDIA (June 5, 2024), https://www.investopedia.com/nvidia-usd3-trillion-market-cap-8658929#:~:text=Nvidia%20is%20the%20third%20U.S.,a%20record%2096%20calendar%20days [https://perma.cc/4XQW-QDPY].

[17] Estimates of Nvidia's market share range from 80% to 95% generally. "Most analysts expect that Nvidia, which controls more than 95% of the market for specialist ai chips, will continue to grow at a blistering pace for the foreseeable future.") *Why do Nvidia's chips*

cloud computing hyperscalers like Microsoft and Google, are working to unseat Nvidia's control of the market, but they are "fighting over scraps."[18] Financial analysts covering the industry recently remarked "[w]e are not overly concerned about competition and expect NVDA to maintain >85% share in Gen AI accelerators even in 2024."[19] Nvidia's dominance over the GPU market is best explained by the relationship between its hardware—the GPU chips—and its software—the programming tools used by developers to design AI products which will in turn run on GPU chips. It is this relationship between Nvidia's hardware—its GPUs—and its software—a programming language called Compute United Device Architecture, or CUDA—that has driven the firm's ascendance.

Nvidia launched CUDA, its "supercomputer software package" in 2006, at first to muted success.[20] While billions of transistors channeled electricity at unparalleled speed, the market for supercomputing was still limited.[21] With the development of highly accurate neural networks, the market for supercomputing expanded rapidly, and leading Silicon Valley firms began to rely on Nvidia's software and hardware to train and run their AI models.[22]

---

dominate the AI market? ECONOMIST (Feb 27. 2024), https://www.economist.com/the-economist-explains/2024/02/27/why-do-nvidias-chips-dominate-the-ai-market [https://perma.cc/4RHM-6FSF]; *see also* Stephen Nellis & Chavi Mehta, *With no big customers named, AMD's AI chip challenge to Nvidia remains uphill fight*, REUTERS (June 13, 2023), https://www.reuters.com/technology/amd-likely-offer-details-ai-chip-challenge-nvidia-2023-06-13/. Similar estimates occur in many sources. *See* Sharon Goldman, *How Nvidia dominated AI — and plans to keep it that way as generative AI explodes*, VENTUREBEAT (Feb. 23, 2023), https://venturebeat.com/ai/how-nvidia-dominated-ai-and-plans-to-keep-it-that-way-as-generative-ai-explodes/ [https://perma.cc/5JBX-US8H] (placing Nvidia's February 2023 market share at 88% of the GPU market). Another Reuters report put Nvidia's market share at 80%, while recent CNBC reporting and Raymond James financial analysis estimates the firm to maintain ">85% share in Gen AI accelerators even in 2024." Other reporting from John Peddie Research and VentureBeat put Nvidia's market share at 88%. Chavi Mehta, *EU examines Nvidia-dominated AI chip market's alleged abuses, Bloomberg reports*, REUTERS (Sept. 30, 2023), https://www.reuters.com/technology/eu-starts-early-stage-probe-into-nvidia-dominated-ai-chip-market-abuses-2023-09-29/ [[https://perma.cc/66JA-FSG7] (Reported Nvidia maintains an 80% market share); Jordan Novet, *Nvidia's revenue triples as AI chip boom continues*, CNBC (Nov. 21, 2023), https://www.cnbc.com/2023/11/21/nvidia-nvda-q3-earnings-report-2024.html [https://perma.cc/57ZQ-8FT8].

[18] Matt Hanson, *Nvidia's GPU dominance is now so total that Intel and AMD are fighting over scraps*, TECH RADAR (Mar. 1, 2023), https://www.techradar.com/news/nvidias-gpu-dominance-is-now-so-total-that-intel-and-amd-are-fighting-over-scraps [https://perma.cc/5JYJ-AVXV]; *See also* Cade Metz et al., *Nvidia's Big Tech Rivals Put Their Own A.I. Chips on the Table*, N.Y. TIMES (Jan. 24, 2024), https://www.nytimes.com/2024/01/29/technology/ai-chips-nvidia-amazon-google-microsoft-meta.html [https://perma.cc/4YZ6-MDTF].

[19] Jordan Novet, *Nvidia's revenue triples as AI chip boom continues*, CNBC (Nov. 21, 2023), https://www.cnbc.com/2023/11/21/nvidia-nvda-q3-earnings-report-2024.html [https://perma.cc/57ZQ-8FT8].

[20] Stephen Witt, *How Jensen Huang's Nvidia is Powering the A.I. Revolution*, New Yorker (Nov. 27, 2023.

[21] *Id.*

[22] *Id.*

Today, CUDA is the "go-to platform for AI developers"[23] and the firm has established "utter dominance" of the full ecosystem around machine learning computing.[24] Nvidia's lead in both the software and hardware markets have created a "competitive moat"—in order to compete with Nvidia, firms need to compete not only in the hardware market, but also in the software market. Although rivals like AMD have entered both the hardware and software market[25], estimates suggest it may be a decade before they can catch Nvidia's lead.[26] The market for GPUs, a critical input for artificial intelligence projects, is highly concentrated,[27] and is protected against competition by the relationship between Nvidia's hardware and software. Nvidia's lead in the hardware and software markets is the result of brave, well-placed bets, and unparalleled technical execution.[28]But now the firm is defending this lead through onerous requirements in the CUDA End-User Licensing Agreement ("EULA").[29] These requirements make it far harder for developers to run CUDA-created software on non-Nvidia hardware.[30] Previously, developers could write software on CUDA, then "translate" their binary code and run the code on another firm's hardware.[31] Software "translators", like ZLUDA, enabled this process.[32] Now, these licensing restrictions lock developers into

---

[23] Dan Gallagher, *How Nvidia Got Huge – and Almost Invincible*, WALL ST. J., (Oct. 6, 2023),
https://www.wsj.com/tech/ai/how-nvidia-got-hugeand-almost-invincible-da74cae1
[https://perma.cc/9H8P-UUJ4].

[24] Alloway & Weisenthal, *supra* note 6.

[25] AMD operates the ROC software platform and CEO Lisa Su noted that "software actually is what drives adoption." Michael Acton, *AMD rolls out new rival to Nvidia's AI chip*, FIN. TIMES (Dec. 6, 2023), https://www.ft.com/content/fa0c97af-c20f-461e-96c9-f2357496c599 [https://perma.cc/DH5X-W2NB].

[26] Richard Waters, *Nvidia's rivals struggle to gain ground in generative AI chip war*, FIN. TIMES (June 15, 2023), https://www.ft.com/content/052e9375-b828-4d5c-aca8-d16c8926017c [https://perma.cc/VH7V-WL9J].

[27] "Nvidia has about 95 per cent of the markets for GPU, or graphics processing units." *See* June Yoon, *AI chip contenders facing daunting 'moats'*, FIN. TIMES (Nov. 27, 2023), https://www.ft.com/content/89745f19-1968-4c46-aaf2-5c6a6a50067f [https://perma.cc/3P3H-HZFZ].

[28] *See, e.g.,* Witt, *supra* note 20.

[29] *See generally* Nvidia End User License Agreement (2024), https://docs.nvidia.com/cuda/eula/index.html [https://perma.cc/5QAG-7HJQ].

[30] *See, e.g*., Muhammad Zuhair, *Nvidia Halts The Use of CUDA On Other Platforms, Lists New Warning in the EULA*, WCCF TECH, (Mar. 5, 2024), https://wccftech.com/nvidia-halts-use-of-cuda-on-other-platforms-lists-new-warning-in-the-eula/ [https://perma.cc/H7AL-5SKE].

[31] *See, e.g.,* Anton Shilov, *Nvidia bans using translation layers for CUDA software — previously the prohibition was only listed in the online EULA, now included in installed files*, TOM'S HARDWARE (Mar. 4, 2024), https://www.tomshardware.com/pc-components/gpus/nvidia-bans-using-translation-layers-for-cuda-software-to-run-on-other-chips-new-restriction-apparently-targets-zluda-and-some-chinese-gpu-makers [https://perma.cc/R43J-LQSP].

[32] Zhiye Liu, ZLUDA Project Paves the Way for CUDA on Intel GPUs, TOM'S HARDWARE (Nov. 24, 2020), https://www.tomshardware.com/news/zluda-project-cuda-intel-gpus [https://perma.cc/QRU9-G3DC].

Nvidia's ecosystem, and undermine innovation and competition in the market for GPUs.

High market concentration dots the entire supply chain of artificial intelligence products.[33] These concentrated chokepoints of the semiconductor supply chain threaten to stall innovation, harm consumers, and further entrench incumbent power.[34] The Taiwan Semiconductor Manufacturing Company (TSMC) holds an estimated 90% of the advanced chip manufacturing market.[35] The Dutch firm ASML has a monopoly on an essential machine that TSMC and other manufacturers use to produce chips.[36] Three firms—Microsoft, Amazon, and Google—dominate the cloud computing market and are the main purchasers of Nvidia's GPUs.[37] These firms have also invested heavily in AI startups.[38] Additional research demonstrates the breadth of consolidation and anticompetitive conduct across the semiconductor industry, including by Apple, Broadcom, and Qualcomm.[39]

American enforcers have taken note of this consolidation, both across by the industry and by Nvidia in particular. The Federal Trade Commission ordered Microsoft, Amazon, and Google to disclose information about their investments in AI startups, and Commissioner Alvaro M. Bedoya has publicly discussed the risks of vertical integration by existing big tech

---

[33] *See, e.g.*, Barry Lynn et al., OPEN MKTS. INST., AI IN THE PUBLIC INTEREST: CONFRONTING THE MONOPOLY THREAT, 15–17, 42 (Anita Jain ed.) (Nov. 2023).

[34] *See* Kate Koidan, *Navigating the Complexities of the Semiconductor Supply Chain*, TOPBOTS (Apr. 15, 2014), topbots.com/semiconductor-supply-chain-overview/ [https://perma.cc/7EMM-RFZX].

[35] Yoon, *supra* note 27.

[36] ASML has a monopoly on "the extreme ultraviolet lithography machines that are critical for the production of advanced chips. The waiting list for the machines, which cost more than $300mn each, averages around two years." *See id* ("ASML has a monopoly on the extreme ultraviolet lithography machines that are critical for the production of advanced chips. The waiting list for the machines, which cost more than $300mn each, averages around two years").

[37] Recent reporting found that Google, Amazon, and Microsoft control two thirds of the global cloud computing market. *See* Prithvi Iyer, *The AI Supply Chain: An Emerging Oligopoly*?, TECH POL'Y PRESS (Nov. 20, 2023), https://www.techpolicy.press/the-ai-supply-chain-an-emerging-oligopoly/ [https://perma.cc/F7ZB-GEBC] (discussing recent reporting that Google, Amazon, and Microsoft control two thirds of the global cloud computing market); Arthur Sants, *How cloud computing became a global monopoly*, INVESTORS CHRON. (May 9, 2023), https://www.investorschronicle.co.uk/news/2023/05/09/how-cloud-computing-became-a-global-monopoly/ [https://perma.cc/NLD8-62K3] (finding that in the UK, Google, Amazon, and Microsoft control nearly 80% of the cloud computing market).

[38] Anthropic, an AI startup, committed to using Amazon's cloud infrastructure and chips to train and run its models. Anthropic has also raised money from Google. Microsoft has invested significantly in OpenAI. *See* David McCabe, *Federal Trade Commission Launches Inquiry Into A.I. Deals by Tech Giants*, N.Y. TIMES (Jan. 25, 2024), https://www.nytimes.com/2024/01/25/technology/ftc-ai-microsoft-amazon-google.html [https://perma.cc/NMC6-B46M].

[39] Todd Achilles et al., RESHORING AND RESTORING: CHIPS IMPLEMENTATION FOR A COMPETITIVE SEMICONDUCTOR INDUSTRY, AM. ECON. LIBERTIES PROJ, 27 (Feb 2024).

incumbents.[40] The Commission's Bureau of Competition and Office of Technology recently reported that markets for specialized chips, like GPUs, either are, or could become, highly concentrated.[41] In 2021, the Commission sued to halt Nvidia's proposed acquisition of Arm, a licensor of the intellectual property that firms like Nvidia use to design chips.[42] In its complaint, the FTC referred to Nvidia as the "dominant supplier of [GPUs]."[43]

European enforcers have also noticed Nvidia's dominance. The European Commission is gathering information on potentially anticompetitive practices in the GPU market.[44] French authorities have interviewed market participants to better understand Nvidia's role in the market and the firm's effect on chip quantity and price.[45] The French competition authority even raided the offices of a firm in the GPU market— a market that Nvidia dominates.[46] While the European Commission has announced no formal investigation into the GPU market,[47] it seems likely that the Commission and French authorities are taking a close look.

Vigorous scrutiny of concentration in the AI industry fits within the current antitrust enforcement landscape. Both the FTC[48] and forty-eight state

---

[40] Press Release, Fed. Trade Comm'n, FTC Launches Inquiry into Generative AI Investments and Partnerships, Fed. Trade Comm'n, (Jan. 25, 2024), https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-launches-inquiry-generative-ai-investments-partnerships [https://perma.cc/5QSY-YRCX]).Commissioner Bedoya "addressed concerns over market concentration and AI, noting that big technology companies have engaged in a vertical integration wherein they own or control the overwhelming majority of resources necessary to dominate, from semiconductors to cloud computing infrastructure, foundation models, and the user interface." Iyer, *supra* note 37.
[41] "Additionally, some markets for specialized chips are—or could be, without appropriate competition policies and antitrust enforcement—highly concentrated." *See Generative AI Raises Competition Concerns,* Fed. Trade Comm'n, (June 29, 2023), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns [https://perma.cc/A4E2-6QV5].
[42] *See generally*, Complaint at 2, *In re Nvidia Corp.*, F.T.C., No. 9404 (Dec. 2, 2021).
[43] *Id.*
[44] Chavi Mehta, *EU examines Nvidia-dominated AI chip market's alleged abuses, Bloomberg reports*, Reuters (Sept. 30, 2023, 5:20 PM), https://www.reuters.com/technology/eu-starts-early-stage-probe-into-nvidia-dominated-ai-chip-market-abuses-2023-09-29/ [https://perma.cc/66JA-FSG7].
[45] *Id.*
[46] Sam Schechner & Asa Fitch, *Nvidia's French Offices Raided in Cloud-Computing Antitrust Inquiry*, Wall St. J, (Sep. 28, 2023), https://www.wsj.com/tech/ai/nvidias-french-offices-raided-in-cloud-computing-competition-inquiry-97c094ea.
[47] Foo Yun Chee, *No formal investigation into AI chips, EU antitrust regulators say*, Reuters (Oct. 2, 2023), https://www.reuters.com/technology/no-formal-investigation-into-ai-chips-eu-antitrust-regulators-say-2023-10-02/#:~:text=BRUSSELS%2C%20Oct%202%20(Reuters),for%20alleged%20anti%2Dcompetitive%20practices.
[48] Press Release, FTC Sues Facebook for Illegal Monopolization, Fed. Trade Comm'n, (Dec. 9, 2020), https://www.ftc.gov/news-events/news/press-releases/2020/12/ftc-sues-facebook-illegal-monopolization [https://perma.cc/2TQJ-34NA].

and regional attorneys general[49] have sued Facebook, arguing that its 2012 and 2014 acquisitions of Instagram and WhatsApp, respectively, were anticompetitive. The Department of Justice also sued Google, seeking to effectively undo its 2008 acquisition of DoubleClick.[50] Close scrutiny of Nvidia's conduct fits within the ongoing examination of consolidation in once-nascent technologies. This Note will consider how Nvidia's conduct fits within the monopolization landscape.

Section II of this Note will examine Nvidia's liability under Section 2 of the Sherman Act. This section will demonstrate Nvidia's monopoly power over the GPU market, before identifying and analyzing the exclusionary tactics the firm uses to maintain its power. This Note will show how the firm's conduct and software licensing requirements harm competition, consumers, and innovation. Section III of this Note will consider possible remedies, including the costs and benefits of litigation. This Section will examine broader ways in which the federal government can rectify harmful consolidation across the semiconductor supply chain. Section IV will offer concluding thoughts and present a look at the effects of consolidation on our future.

I.   NVIDIA IS A MONOPOLIST IN OVER GPU CHIPS WITH DURABLE MONOPOLY POWER PROTECTED BY DEEP BARRIERS TO ENTRY

The Sherman Act, passed in 1890, makes it illegal to monopolize, or attempt to monopolize trade.[51] While debate over the Act's meaning and history continues, the Sherman Act was plainly a response to the "Trust Problem" and the growing economic power of giants like Standard Oil and U.S. Steel.[52] The law rests on the premise that fair competition will yield the best allocation of resources, the lowest prices, highest quality, greatest progress, and provide an "environment conductive to the preservation of our democratic political and social institutions"—it is an "comprehensive charter of economic liberty."[53]

A violation of Section 2 of the Sherman Act has two elements: first, "the possession of monopoly power in the relevant market", and second "the willful acquisition or maintenance of that power as distinguished from

---

[49] Nick Statt & Russell Brandom, *The FTC is suing Facebook to unwind its acquisition of Instagram and WhatsApp*, THE VERGE (Dec. 9, 2020), https://www.theverge.com/2020/12/9/22158483/facebook-antitrust-lawsuit-anti-competition-behavior-attorneys-general [https://perma.cc/4B6X-YJ33].

[50] "The DOJ is seeking to break up Google's ad-tech business—in effect, undoing the DoubleClick merger." *The case against Google hinges on an antitrust 'mistake,'* ECONOMIST (Mar. 2, 2023), https://www.economist.com/finance-and-economics/2023/03/02/the-case-against-google-hinges-on-an-antitrust-mistake [https://perma.cc/YWD6-DH7X].

[51] "Every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several States, or with foreign nations, shall be deemed guilty of a felony…." 15 U.S.C. § 2.

[52] *See, e.g.,* TIM WU, THE CURSE OF BIGNESS 24–31. ([T]he scholarly debate over the Sherman Act's meaning and history may never end.").

[53] Northern Pac. Ry. Co. v. United States, 356 U.S. 1, 4 (1958).

growth or development as a consequence of a superior product, business acumen, or historic accident."[54]

The first element, possession of monopoly power, is the power to control price or exclude competitors, and can be inferred by a firm's market share.[55] Legal and factual analysis demonstrates that Nvidia has monopoly power over the market for GPU chips.

A. Nvidia is a Monopolist in the Product Market of GPUs

Monopoly power is demonstrable two ways: direct and indirect (or structural) evidence.[56] Direct evidence examines the ability of a firm to "profitably raise prices substantially above the competitive level."[57] Direct proof can be evidenced by quantity or price changes, demonstrating a firm's ability to reduce market output and increase price.[58] Generally speaking, only a monopolist can set a product's price without considering the price of rivals' products.[59] Structural, or indirect, proof is evidenced by a firm's predominant market share[60] or possession of a dominant share of a market that is protected by entry barriers.[61]

The threshold question for a market power assessment is: what is the relevant market?[62] Once the relevant market is identified, courts can identify the power of a firm within the market.[63] First, this Note will define the relevant product market, then it will examine Nvidia's power within that market.

1. The Brown Shoe Practical Indicia Demonstrate the GPU Product Market

The GPU semiconductors that train and run programs like large language models are a product market. One way to define a product market is through "practical indicia."[64] Such practical indicia include industry or public recognition of the submarket as a separate economic entity, the product's peculiar characteristics and uses, unique production facilities, distinct

---

[54] United States v. Grinnell Corp., 384 U.S. 563, 570–71 (1966).

[55] Id. at 571.

[56] "In United States v. E. I. du Pont De Nemours & Co., [the Court] defined monopoly power as 'the power to control prices or exclude competition.' The existence of such power ordinarily may be inferred from the predominant share of the market." See id.

[57] United States v. Microsoft Corp., 253 F.3d 34, 51 (D.C. Cir. 2001), (citing 2A PHILLIP E. AREEDA ET AL., ANTITRUST LAW ¶ 501, at 85 (1995)).

[58] Ball Mem'l Hosp., Inc. v. Mut. Hosp. Ins., Inc., 784 F.2d 1325, 1335 (7th Cir. 1986).

[59] Microsoft Corp., 253 F.3d at 57 (citing Rothery Storage & Van Co. v. Atlas Van Lines, 729 F.2d 210, 218 (D.C. Cir. 1986)).

[60] Grinnell Corp., 384 U.S. 563 at 571.

[61] Microsoft Corp., 253 F.3d at 51.

[62] Grinnell Corp., 384 U.S. 563 at 571–76.

[63] See, e.g., id. at 571.

[64] See, e.g., Illumina, Inc. v. Fed. Trade Comm'n, 88 F.4th 1036, 1049 (5th Cir., 2023); Regeneron Pharms., Inc. v. Novartis Pharma AG, 96 F.4th 327, 339 (2d Cir., 2024); Brown Shoe Co., v. United States, 370 U.S. 294, 325 (1962).

customers, distinct prices, sensitivity to price changes, and specialized vendors.[65]

The *Brown Shoe* "practical indicia" are seen as indicators—not requirements—of a product market; a market can exist with just a handful of indicia.[66] In *Beatrice Foods, Co.* for example, a submarket existed based on "industry recognition, peculiar characteristics of the product, and differences in production methods and prices."[67]

In *Staples*, the District Court for the District of Columbia upheld the FTC's definition of the product market. The FTC defined the superstore submarket based on industry recognition; unique characteristics like distinct store formats, customers, and prices; and evidence of pricing sensitivity.[68] The *Staples* court considered "industry or public recognition" of the market, particularly examining the internal strategy documents of Staples and Office Depot, and how the firms themselves defined the market.[69] The court also considered the stores' "appearance, physical size, format, . . . SKU's offered, and the type of customers targeted and served . . ." to define the product market.[70] The District Court upheld the FTC's product market based on the company's own definition of the market, and key physical and competitive differences between superstores and other kinds of office supply stores.

The DC Court of Appeals also upheld the FTC's product market in *Whole Foods*, again considering the *Brown Shoe* indicia, especially the existence of a "core group" of dedicated and distinct customers.[71] A key indicia of the product market in *Whole Foods* was that the store "cater[ed] to a core group of customers who have decided that natural and organic is important, lifestyle of health and ecological sustainability is important."[72] *Whole Foods* demonstrates how customer preference can help establish a relevant product market.

Under the *Brown Shoe* practical indicia approach, there is a clear and distinct product market for the GPU semiconductors used to train and run AI systems and large language models.

First, GPUs have "peculiar characteristics and uses"[73] compared to other kinds of semiconductors. There are two other main kinds of chips, CPUs and

---

[65] Brown Shoe Co., v. United States, 370 U.S. 294, 325 (1962).

[66] "[T]he Court described these factors as "practical indicia" rather than requirements, [and] subsequent cases have found that submarkets can exist even if only some of these factors are present." FTC v. Staples, Inc. 970 F. Supp 1066, 1075 (D.D.C. 1997).

[67] *Id.* at 1075 (citing *Beatrice Foods Co.,* 540 F.2d at 303 (7th Cir. 1976)).

[68] *Id.* at 1075.

[69] "[T]he evidence clearly shows that Staples and Office Depot each consider the other superstores as the primary competition." *Id.* at 1079–80.

[70] *Id.* at 1078.

[71] FTC v. Whole Foods Mkt., Inc., 548 F.3d 1028, 1039 (D.C. Cir. 2008).

[72] *Id*. at 1039 (internal quotations omitted).

[73] *Brown Shoe Co.*, 370 U.S. at 325.

DPUs, which have distinct strengths, weaknesses, and uses.[74] GPUs excel at parallel computing and are more efficient than CPUs at "compute-intensive applications."[75] CPUs are used for a wide range of computing tasks at lower efficiency. DPUs, meanwhile, are primarily used to move data in data centers.[76] These chips are also made differently and have distinct architectures.[77] GPUs and CPUs have similar cores, memory, and control units, but work differently.[78] Each kind of chip has a specific use case and function—this is one practical indicia demonstrating that GPUs are their own product market.

There are also distinct customers for GPUs—like the "core customers" whose specific needs were a practical indicia in *Whole Foods*.[79] The two main groups of GPU customers are data center providers—the "hyperscalers" —and firms developing their own generative AI capabilities.[80] Nvidia's sales to data centers has skyrocketed 409% year over year; over half of these sales were to large cloud providers, the hyperscalers

---

[74] CPUs run basically in a straight line and execute the basics of a computer system, like "calling up information from a hard drive…." GPUs do "parallel computing" and divide tasks among many processors at once. A GPU "break[s] complex problems into thousands or millions of separate tasks and work them out at once." Initially designed to support the graphics of high-quality video games, the GPU is now critical to modern supercomputing, data centers, and AI systems. *See generally* Brian Caulfield, *What's the Difference Between a CPU and a GPU?*, NVIDIA (Dec. 16, 2009), https://resources.nvidia.com/en-us-what-is-series/whats-the-difference [https://perma.cc/5JW6-959A].

[75] *What's the Difference Between GPUs and CPUs?*, AMAZON, https://aws.amazon.com/compare/the-difference-between-gpus-cpus/ [https://perma.cc/NL2L-KP5T] (last accessed Sept. 26, 2024).

[76] DPUs, which move data throughout a data center, round out the "three pillars of computing." *See* Kevin Deierling, *What is a DPU?*, NVIDIA (May 20, 2020), https://blogs.nvidia.com/blog/whats-a-dpu-data-processing-unit/ [https://perma.cc/698J-ZDQN].

[77] *CPU vs. GPU: Powerful Options for Your Computing Needs*, INTEL, https://www.intel.com/content/www/us/en/products/docs/processors/cpu-vs-gpu.html [https://perma.cc/5FH6-P3RA] (last accessed, Jan. 19, 2024).

[78] "The CPU is like a head chef in a large restaurant who has to make sure hundreds of burgers get flipped. Even if the head chef can do it personally, it's not the best use of time. All kitchen operations may halt or slow down while the head chef is completing this simple but time-consuming task. To avoid this, the head chef can use junior assistants who flip several burgers in parallel. The GPU is more like a junior assistant with ten hands who can flip 100 burgers in 10 seconds." *What's The Difference Between GPUs and CPUs?*, AMAZON, [https://perma.cc/U3GX-BR26] (last accessed Apr. 11, 2024).

[79] *Whole Foods Mkt. Inc*., 548 F.3d at 1039 (D.C. Cir. 2008) (citing United States v. Rockford Mem'l Corp., 898 F.2d 1278, 1284 (7th Cir. 1990)).

[80] *See, e.g.,* Mike Wheatley, *Nvidia's Data Center GPU Sales Grow by a Stunning 409% on Huge Demand for AI chips*, SILICON ANGLE (Feb 21. 2024, 6:48PM), https://siliconangle.com/2024/02/21/nvidias-data-center-gpu-sales-grow-stunning-409-huge-demand-ai-chips/ [https://perma.cc/JFJ8-TWWB].

like Amazon's AWS.[81] In 2022, an estimated two thirds of all GPUs were purchased by only four firms: Microsoft, Google, Amazon, and Meta.[82]

Firms like Meta seeking to develop their own generative AI tools are also core customers of GPUs. Meta CEO Mark Zuckerberg stated that the company planned to buy about 350,000 GPUs from Nvidia to power its AI goals.[83] Inflection AI, a California start-up creating AI tools, is spending 95% of its recent $1.3 billion fundraising effort on Nvidia's GPUs.[84] Much like the core group of customers who demonstrated the product market in *Whole Foods*, here, data center hyperscalers and tech firms seeking to build out AI capabilities uniquely demand GPUs. This group of customers is a practical indicia of the GPU product market.

Another practical indicia demonstrating the existence of the GPU product market is customer preference. The GPU-powered cloud services provided by the hyperscalers is uniquely sought after: there are up to year-long wait lists to access GPU computing power.[85] Customers specifically want access to GPU chips and are willing to both wait and spend significantly.

GPUs are also priced much differently than are other semiconductors, further demonstrating a practical indicia of this market. One of Nvidia's GPUs, the A100 or Ampere, costs about $10,000[86], while the H100, or Hopper, costs up to $40,000.[87] Other kinds of GPUs cost less, but generally these chips have different uses, like in gaming consoles or personal devices.[88] CPUs, meanwhile, generally retail for under a thousand dollars.[89] The GPU chips that train and run AI products are significantly more expensive than

---

[81] *Id.*

[82] Arthur Sants, *How Cloud Computing Became a Global Monopoly*, INVESTORS CHRON. (May 9, 2023), https://www.investorschronicle.co.uk/content/7ee67f8b-8f54-56f8-ae2e-ea649ee698ae [https://perma.cc/NLD8-62K3].

[83] Mike Wheatley, *Meta Plans to Buy 350K Nvidia GPUs to Build Artificial General Intelligence*, SILICON ANGLE (Jan. 18, 2024), https://siliconangle.com/2024/01/18/meta-plans-buy-350k-nvidia-gpus-build-artificial-general-intelligence/ [https://perma.cc/5TCF-SQHN].

[84] Don Clark, *How Nvidia Built a Competitive Moat Around A.I. Chips*, N.Y. TIMES (Aug. 21, 2023), https://www.nytimes.com/2023/08/21/technology/nvidia-ai-chips-gpu.html [https://perma.cc/RT69-8SQB].

[85] Griffith, *supra* note 7.

[86] Kif Leswing, *Meet the $10,000 Nvidia Chip Powering the Race for A.I.*, CNBC (Feb 23, 2023), https://www.cnbc.com/2023/02/23/nvidias-a100-is-the-10000-chip-powering-the-race-for-ai-.html [https://perma.cc/VM2H-W5M3].

[87] Clark, *supra* note 84; *see also* Nvidia 100 GPU (PCIe), BUS. SYS. INT'L, https://www.bsi.uk.com/nvidia-h100-pcie-1018 [https://perma.cc/2VK6-YKXZ] (last accessed Mar. 10, 2024) (quoting H100 price at £32,050, but not available for sale).

[88] Matt Hanson, *Nvidia's GPU Dominance is Now so Total That Intel and AMD are Fighting Over Scraps*, TECH RADAR (Mar. 1, 2023), https://www.techradar.com/news/nvidias-gpu-dominance-is-now-so-total-that-intel-and-amd-are-fighting-over-scraps [https://perma.cc/EG2F-PV4U]. Another use of lower cost GPUs is in computer monitors, such as the Nvidia G-Sync powered monitor on which this Note was written.

[89] *See generally*, *Computer CPU Processors*, AMAZON, https://www.amazon.com/CPU-Processors-Memory-Computer-Add-Ons/b?ie=UTF8&node=229189 [https://perma.cc/9EEC-38JX] (last accessed Mar. 10, 2024).

other kinds of GPU chips or other chips like CPUs. This is a clear indicia of a distinct product market for the chips that train and run AI products.

Based on the *Brown Shoe* practical indicia, the GPUs that train and run AI systems are their own product market. Market participants pay far more for these chips than they pay for other kinds of chips; the chips serve different use cases than other chips; and GPUs have unique architectures.[90] Because the market regards GPUs as their own product, so too will the courts.

Courts also consider the "reasonable interchangeability" and cross-elasticity of demand between products when determining a product market.[91] If there is little interchangeability, or low cross-elasticity of demand, then a monopolist's price hike is more likely to harm consumers who will have nowhere to turn.[92] The basic question under this analysis is the extent to which purchasers are willing and able to substitute one good for another.[93] If customers cannot substitute one good for another, then it is more likely that a seller will impose a "small but significant nontransitory increase in price" (SSNIP).[94] Here, evidence supports the assertion that customers cannot, and do not, substitute away from GPU chips. Customers are willing to wait 18 months for an Nvidia GPU systems, even though competitors have available alternatives.[95] Companies like Alphabet, the Google parent, are developing their own GPU chips in an effort to compete with Nvidia.[96] Despite shortages of GPU supply, startups and entrenched firms alike are waiting to access the chips, demonstrating that other kinds of chips are not effective substitutes for GPUs, and that customers could be harmed by price or supply changes in the GPU market. These facts make clear that GPUs are their own product market.

2. The Geographic Market for GPUs is Global

---

[90] *What's the Difference Between GPUs and CPUs?*, AMAZON, https://aws.amazon.com/compare/the-difference-between-gpus-cpus/ [https://perma.cc/NL2L-KP5T] (last accessed Sept. 26, 2024); *CPU vs. GPU: Powerful Options for Your Computing Needs*, INTEL, https://www.intel.com/content/www/us/en/products/docs/processors/cpu-vs-gpu.html [https://perma.cc/5FH6-P3RA] (last accessed, Jan. 19, 2024).

[91] *Whole Foods Mkt., Inc.*, 548 F.3d at 1037 (citing United States v. Microsoft Corp., 253 F. 3d 34, 105 (D.C. Cir. 2001)) (internal quotation omitted); *Brown Shoe Co.*, 370 U.S. 294 at 325.

[92] *See* FTC v. Whole Foods Mkt., Inc., 502 F. Supp 2d 1, 7 (D.D.C. 2007) (citing E.I. du Pont de Nemours, 351 U.S. 377, 393 (1956)).

[93] *Id.* (quoting Hayden Pub. Co. v. Cox Broad. Corp., 730 F.2d 64, 70 n.8 (2d Cir. 1984)).

[94] *Id.* at 16.

[95] Clark, *supra* note 84; Griffith, *supra* note 7.

[96] *Good Luck Catching Up to Nvidia*, FIN. TIMES (Feb. 21, 2024), https://www.ft.com/content/1a6dc679-0a8e-4317-9ab8-cfd0677b0bd9 [https://perma.cc/5RS4-ZKJE].

Courts consider geographic constrains when determining the size of a product market.[97] The geographic bounds of a market depend on consumers' willingness to travel, availability of substitutes, and trade constraints.[98]

The market for GPUs is global. Chinese firms Tencent and Alibaba have reportedly acquired Nvidia's GPUs, as have Gulf states Saudi Arabia and the United Arab Emirates.[99] Iris Energy, an Australian firm, planned to buy $10 million of Nvidia's H100s for about $40,000 each.[100]

There are also regulatory constraints on the GPU exports, which could constrain the size of the geographic market. Commerce Department rulemaking restricts the export of many advanced chips to Iran and Russia, and limits exports to China..[101] Commerce Secretary Gina Raimondo noted that advanced chips like Nvidia's GPUs could power AI and supercomputing breakthroughs with military benefits for geopolitical adversaries.[102] Additionally, as demonstrated, European enforcers are examining the GPU market in general, and Nvidia's power in particular.[103] This global scrutiny demonstrates the breadth of Nvidia's reach and the scope of demand for GPUs. Although American trade policy may somewhat limit the market for GPUs, it is clear these chips are sought the world over.

These factors demonstrate that there is a global market for the GPU semiconductors that are used to train and run AI systems and large language models. This Note will next consider Nvidia's power over this market.

B. Nvidia's Market Power is Protected by Deep Barriers to Entry

After defining a product market, courts then examine the extent to which a firm possesses power in the relevant market.[104] Market power is the ability of a firm to raise price and restrict output or to "force a purchaser to do something that he would not do in a competitive market."[105] There are two ways to prove market power: structural evidence and direct evidence.[106] This Note will demonstrate Nvidia's market power with clear structural evidence.

---

[97] *Mergers*, FTC, https://www.ftc.gov/advice-guidance/competition-guidance/guide-antitrust-laws/mergers [https://perma.cc/33Y6-R4VW] (last visited, Jan. 19, 2024).

[98] There are many factors that can limit the scope of a geographic market, including "transportation costs, language, regulation, tariff and non-tariff trade barriers, custom and familiarity, reputation, and local service availability." U.S. Dep't of Just. & FTC, *Merger Guidelines* § 4.3.D.2 (rev. Dec. 18, 2023).

[99] Madhumita Murgia, et al., *Saudi Arabia and UAE Race to Buy Nvidia Chips to Power AI Ambitions*, FIN. TIMES (Aug. 14, 2023), https://www.ftc.com/content/c93d2a76-16f3-4585-af61-86667c5090ba [https://perma.cc/BSB5-28KE].

[100] Novet, *supra* note 20.

[101] Alexandra Alper, et al., *Biden Cuts China Off from More Nvidia Chips, Expands Curbs to Other Countries*, REUTERS (Oct. 17, 2023, 5:33PM),
https://www.reuters.com/technology/biden-cut-china-off-more-nvidia-chips-expand-curbs-more-countries-2023-10-17/ [https://perma.cc/N6LH-2K3X].

[102] *Id.*

[103] Mehta, *supra* note 44.

[104] *See, e.g., Grinnell Corp.,* 384 U.S. 563 at 571–76.

[105] Eastman Kodak Co. v. Image Tech. Serv., Inc., 504 U.S. 451, 464 (1992) (internal quotations omitted).

[106] *See, e.g.,* Rebel Oil Co. v. Atl. Richfield Co., 51 F.3d 1421, 1434 (9th Cir. 1995).

Structural evidence of market power is demonstrated by a firm's predominant market share[107] or possession of a dominant share of a market that is protected by barriers to entry.[108] Although market power can also be demonstrated through direct evidence of a firm's power, this evidence is a sufficient, but not necessary, way to prove monopoly power.[109] The *Microsoft* Court specifically endorsed the use of structural evidence to prove market power in the software industry.[110] Nvidia plainly possesses a monopolistic share of the GPU market.

Nvidia's monopolistic market power is demonstrated by significant Supreme Court and Courts of Appeal precedent. In *Alcoa* the Court stated that 90% of the market constitutes a monopoly.[111] In *Grinnell* the court stated that 87% of the market "leaves no doubt that the…defendants have monopoly power"[112] and in *American Tobacco*, 80% market share was a "substantial monopoly."[113] Circuit court precedent further demonstrates the contours of market power, elucidating that, generally, above 70% market share constitutes monopoly power.[114] Nvidia possesses at least 80%, and perhaps up to 95% of product market.[115] Nvidia's market share is also sufficient to demonstrate its monopoly power because the firm is protected by network effects, which give big firms an advantage over smaller incumbents.[116]

One counterargument against this structural approach is that the market for GPUs is flexible and nascent, thus prone to change. However, as will be

---

[107] *Grinnell Corp.*, 384 U.S. 563 at 571.

[108] *Microsoft Corp.*, 253 F.3d at 51; *Rebel Oil Co.*, 51 F.3d at 1434.

[109] *See, e.g. Microsoft Corp.*, 253 F.3d at 57.

[110] *Id.*

[111] United States v. Aluminum Co. of Am., 148 F.2d 416, 424 (2d Cir. 1945).

[112] *Grinnell Corp.,* 384 U.S. at 571.

[113] Am. Tobacco Co. v. United States, 328 U.S. 781, 797 (1946).

[114] The Tenth Circuit indicated that the minimum market share to show monopoly power is "between 70% and 80%." Colo. Interstate Gas Co. v. Nat. Gas Pipeline Co. of Am., 885 F.2d 683, 694 n.18 (10th Cir. 1989) (citing 2 E. Kintner, *Federal Antitrust Law* § 12.6 (1980)). The Third Circuit held that "a share significantly larger than 55% has been required to establish prima facia market power." United States v. Dentsply Int'l, Inc., 399 F.3d 181, 187 (3d Cir. 2005). The Fifth Circuit held that 52% of the market was insufficient to show monopolization as "monopolization is rarely found when the defendant's share of the relevant market is below 70%." Exxon Corp. v. Berwick Bay Real Estate Partners, 748 F.2d 937, 939–40 (5th Cir. 1984).

[115] *Why Do Nvidia's Chips Dominate the AI Market?*, ECONOMIST (Feb 27. 2024), https://www.economist.com/the-economist-explains/2024/02/27/why-do-nvidias-chips-dominate-the-ai-market [https://perma.cc/XDM6-95DK] (stating that, of February 2024, Nvidia's "controls more than 95% of the market for specialist AI chips…."); Sharon Goldman, *How Nvidia Dominated AI — and Plans to Keep it That Way as Generative AI Explodes*, VENTUREBEAT (Feb. 23, 2023, 5:00AM), https://venturebeat.com/ai/how-nvidia-dominated-ai-and-plans-to-keep-it-that-way-as-generative-ai-explodes/ [https://perma.cc/5JBX-US8H] (placing Nvidia's February 2023 market share at 88% of the GPU market).

[116] Herbert Hovenkamp, *Antitrust and Platform Monopoly*, 130 YALE L. J. 1952, 1962 (2021).

shown next, Nvidia's power is protected by deep barriers to entry, which prevent new firms from entering the market and competing.

Market share alone can be sufficient to demonstrate a firm's monopoly, but courts also consider the existence of "entry barriers."[117] Entry barriers are factors that make market entry by another firm either slower or more costly, and thus limit the ability of other firms to constrain a monopolist.[118] Entry barriers could include "large capital outlays" "lengthy construction programs," "brand preference," or "extensive 'image' advertising expenditures."[119] Several of these entry barriers protect Nvidia's power over the GPU market.

One entry barrier insulating Nvidia's lead is those benefits from being first in the market.[120] Buyers of GPUs use the same type of chips to train, then run, their AI programs.[121] There are significant benefits to being the first mover in this market because, as demonstrated, while training a model takes an estimated ten thousand GPUs, running the model for two years requires a million—a hundred times more.[122] The first firm to supply chips for the training stage can then lock in customers for the far more lucrative next stage. Since Nvidia is the early mover and dominates the first stage—the training stage—there is a significant barrier to entry as future firms will be effectively locked out from providing the GPUs to run models in the future. This barrier to entry exists because of a lack of portability between GPUs, which, as will be shown, is an intentional and anticompetitive tactic of Nvidia. But this barrier to entry is not permanent, and was threatened by the development of translators like ZLUDA. Nvidia's response to the development of these translators ran the firm afoul of the Sherman Act.

Another barrier protecting Nvidia's lead are the high costs of market entry.[123] Nvidia's current GPU program is the result of investing more than $30 billion over more than a decade.[124] Entry is so costly and challenging because it requires both software and hardware solutions and analysts believe it could take a decade for competitors to match Nvidia's lead in the software

---

[117] *See, e.g., Microsoft Corp.,* 253 F.3d at 54.

[118] Barriers to entry "make[] entry more costly or time-consuming and thus reduce[] the effectiveness of potential competition as a constraint on the pricing behavior of the dominant firm." S. Pac. Commc'ns Co. v. Am. Tel. & Tel. Co., 740 F.2d 980, 1001 (D.C. Cir. 1984).

[119] *Id.* at 1002 (citing United States v. AT&T, 524 F. Supp. 1336, 1348 (D.D.C. 1981)).

[120] *See, e.g., id.*

[121] Alloway & Weisenthal, *supra* note 6 (describing that you use the same kind of chip to run a model as you used to train it).

[122] *Id.*

[123] *See generally,* Clark, *supra* note 84.

[124] *Id.*

ecosystem.[125] The time and expense of entry presents one clear barrier to entry that insulates Nvidia's lead in this market.

Nvidia's power over the GPU market is protected by the flywheel relationship between its hardware and software products. Software use drives adoption of hardware,[126] and Nvidia dominates the market for software, enabling its dominance of the market for hardware.[127] Other firms cannot effectively compete unless they can compete in both markets and Nvidia's lead in the software market is large.[128] CUDA, which is free, was downloaded more than twenty five million times from 2022 to 2023 and is the "go-to platform for AI developers."[129] Nvidia's lead in the software market is a major barrier defending its lead in the hardware market—as will be shown, the firm's licensing agreement for CUDA now requires that developers who create code on CUDA run that code on Nvidia's GPUs.

The relationship between Nvidia's software and its hardware is very similar to the relationship between Microsoft's operating system (Windows) and its browser, Internet Explorer. The court in *Microsoft* identified the "'chicken-and-egg' situation" that entrenched Microsoft's lead: most consumers want an operating system with many applications, and the developers who write applications choose operating systems with the most users.[130] This is an example of a network effect, where the value of a good increases with its use.[131]

Nvidia's products benefit from similar network effects. Chip purchasers want Nvidia's GPUs in part because of the large developer community inventing new AI projects in CUDA.[132] This large developer community creates network effects—the more projects there are in CUDA, the more people will want to buy Nvidia's GPUs. If developers could take their CUDA-created code and run that code on a different firm's chips, then

---

[125] Waters, *supra* note 28; June Yoon, *New generation of chips will drive the AI wave*, FIN. TIMES (Dec. 18, 2023), https://www.ft.com/content/e5c97253-90f3-4158-846f-07aaf3945983 [https://perma.cc/GVM4-YJSN] (arguing that "[S]haking Nvidia's dominance, the key to which lies not just in the physical chip itself but its software ecosystem, which includes popular developer tools and programming models, is another story. Replicating that will take years.").

[126] *See, e.g.,* Acton, *supra* note 27 (quoting AMD CEO Lisa Su who said, "[S]oftware actually is what drives adoption.").

[127] *See, e.g.,* Alloway & Weisenthal, *supra* note 6 (explaninig that through CUDA, Nvidia has "utter dominance" of the ecosystem for machine learning computing); Waters, *supra* note 28. (explaining that "[T]he widespread use of Nvidia's chips in AI and other applications owes much to the ease with which its GPUs, originally designed for video gaming, can be programmed for other tasks using its Cuda software.").

[128] Waters, *supra* note 26.

[129] Gallagher, *supra* note 23.

[130] *Microsoft Corp.*, 253 F.3d at 55.

[131] *What is the Network Effect*, WHARTON ONLINE (Jan. 13, 2023), https://online.wharton.upenn.edu/blog/what-is-the-network-effect/ [https://perma.cc/7DK2-7JVV].

[132] Alloway & Weisenthal, *supra* note 6.

Nvidia's monopoly over GPUs would be under threat—the flywheel would be broken. Developers would be free to run their CUDA-created code on any GPU of their choice. The legal ramifications of the relationship between Nvidia's software and hardware will be discussed in later sections; for now, it suffices to state that Nvidia's lead in software helps protect its lead in hardware. This poses a significant barrier to entry for potential competitors.

One counterargument at this stage may be: hasn't Nvidia *earned* those barriers to entry—aren't the barriers to entry merely a lead that the firm has, through good decision-making, won? This may be true. Nvidia was an early mover; they bet big on the AI revolution and, so far, they are winning.[133] But this does not disprove the existence of legally cognizable barriers to entry. Courts consider factors like brand preference or large capital outlays as entry barriers.[134] Both factors are considered not for their evidentiary weight as proof of illegal conduct, but as objective evidence of market power. At this stage of the analysis, courts and enforcers consider the existence of entry barriers to understand market power, even barriers legally erected. Here, it is clear that Nvidia's lead is surrounded by significant entry barriers––both structural, market-based barriers, and barriers of its own creation.

Another counterargument may focus on the existence of actual or potential competitors like AMD or Intel. AMD, for example, is trying to challenge Nvidia's lead in both the software and hardware markets. On the software side, AMD launched ROCm, a platform to compete with CUDA; on the hardware side, its GPUs, like the MI300X, directly compete with those of Nvidia.[135] In 2023, AMD highlighted Microsoft and Meta's adoption of the MI300 for AI projects—some suggest that the chip may even "dominate[]" Nvidia's comparable model.[136] AMD's chips, however, are mostly used for the inferencing, not training, while Nvidia's GPUs handle both;[137] despite AMD's success, all that's left may be the "dregs" of the market.[138] Additionally, the existence of rising competitors is often exactly what sparks exclusionary conduct in the first place; in *Microsoft*, it was the rising success of Netscape that triggered the firm's anticompetitive response.[139]

AMD's launch of ROCm is also an attempt to earn market share away from Nvidia by converting software uses away from CUDA. Because Nvidia's GPUs are those most sought after, however, CUDA is also the most

---

[133] *See generally* Waters, *supra* note 26 ("[T]he gulf that has opened up between Nvidia and the rest of the chip industry has only grown wider.").

[134] *See, e.g.*, *S. Pac. Commc'ns Co.,* 740 F.2d at 1002.

[135] Acton, *supra* note 25.

[136]'On raw specs, MI300X dominates H100,' Dylan Patel and Daniel Nishball at semiconductor consulting firm SemiAnalysis wrote on Wednesday". *Id.*

[137] Waters, *supra* note 26 ("[AMD] only highlighted the product's ability to handle AI inferencing — applying pre-trained AI models — rather than the more demanding job of training, which has been behind Nvidia's surging sales").

[138] *Id.*

[139] *See Microsoft Corp.*, 253 F.3d at 79.

desired software solution.[140] Although AMD could, over time, eat away at Nvidia's lead in the software market, at this stage, Nvidia's lead is protected by sizeable entry barriers.

AMD and Intel may also be targeting their GPU chips to different consumers than does Nvidia. Intel debuted its Arc GPU, but instead of competing with Nvidia in the AI market, Intel is competing with AMD over the lower-budget gamer market.[141] Although Intel is creating GPUs, these GPUs are not necessarily competing with those of Nvidia or seeking to capture the firm's market share.

Finally, even if AMD and Intel do compete with Nvidia, structural evidence still demonstrates that the firm has monopoly power. Many monopolists have competitors—Microsoft had Netscape[142]—as monopolistic market power does not require complete dominion. Monopoly power is an inference of a firm's ability to control price or limit competition, [143] not destroy it entirely. Although firms like AMD and Intel may compete with Nvidia, based on structural evidence of Nvidia's market power—and its deep entry barriers—Nvidia has monopoly power over the GPU chips used to train and run AI systems.

Nvidia satisfies the first prong of a Sherman Act monopolization claim: the possession of monopoly power.[144] Next, this Note will consider the second prong of a monopolization claim: that a firm violates § 2 through "the willful acquisition or maintenance of that power as distinguished from growth or development as a consequence of a superior product, business acumen, or historic accident." [145]

II. NVIDIA ILLEGALLY DEFENDS ITS MONOPOLY THROUGH SOFTWARE
     LICENSING AGREEMENTS THAT HARM INNOVATION AND COMPETITION

A. *Nvidia's Exclusionary Conduct in CUDA Licensing*

---

[140] Matt Rickard, *Nvidia's CUDA Monopoly*, SUBSTACK, (Aug. 6, 2023), https://matt-rickard.com/nvidias-cuda-monopoly [https://perma.cc/Z3HX-BGPU].

[141] Matt Hanson, *Nvidia's GPU dominance is now so total that Intel and AMD are fighting over scraps*, TECH RADAR (Mar. 1, 2023), https://www.techradar.com/news/nvidias-gpu-dominance-is-now-so-total-that-intel-and-amd-are-fighting-over-scraps [https://perma.cc/78SR-XC77]; Gallagher, *supra* note 24.

[142] *Microsoft Corp*., 253 F.3d at 50.

[143] *Grinnell Corp.,* 384 U.S. at 571.

[144] *Id.* at 570.

[145] *Id.* at 570–71; *see also Microsoft Corp.*, 253 F.3d at 58 ("A firm violates § 2 only when it acquires or maintains, or attempts to acquire or maintain, a monopoly by engaging in exclusionary conduct as distinguished from growth or development as a consequence of a superior product, business acumen, or historic accident.").

Nvidia recently updated CUDA's End-User License Agreement (EULA).[146] Now, once developers create source code in CUDA and compile that source code into binary code using a CUDA software development kit (SDK), they must run that binary code on an Nvidia GPU.[147] Developers who use CUDA are locked into using Nvidia's hardware. This is a significant change; previously, developers could use "translators" that make CUDA-created binary code work with non-Nvidia chips.[148] Tools like ZLUDA could, with some limitations, run CUDA-created binary code on Intel GPUs.[149] Now, these licensing restrictions bars the use of CUDA-created binary code on non-Nvidia chips.

Source code written in CUDA can still run on non-Nvidia GPUs.[150] Under the licensing restrictions, a developer could use a translator to turn CUDA-created source code (as opposed to CUDA-created binary code) into source code of another software system, like AMD's ROCm. Then, the developer could translate the new source code into binary code and run that binary code on a non-Nvidia GPU, like an AMD chip, without violating EULA.[151] This workaround is only possible if there are viable alternatives to CUDA because developers would still need a software tool to compile their source code into binary; currently, AMD is developing tools to translate CUDA source code into source code compatible with non-Nvidia chips.[152] This workaround requires customers who use CUDA and want to use a non-Nvidia GPU to add an extra step to their process—the translation of CUDA source code into source code for another program.

1. *These Licensing Restrictions Have Significant Anticompetitive Effects on Consumers, Competition, and Innovation*

Nvidia's conduct harms both consumers and competition itself by limiting consumer choice; diminishing the benefits of innovation; and restricting the ability of competitor firms to discipline Nvidia on price, quantity, or quality.

---

[146] Nvidia End User License Agreement § 1.2(8) (2024) ("You may not reverse engineer, decompile or disassemble any portion of the output generated using SDK elements for the purpose of translating such output artifacts to target a non-NVIDIA platform."), https://docs.nvidia.com/cuda/eula/index.html [https://perma.cc/5QAG-7HJQ] (last visited Mar. 11, 2024); Mark Campbell, *Nvidia bans the use of translation layers on CUDA software*, OVERCLOCK 3D, (Mar. 6, 2024), https://overclock3d.net/news/gpu-displays/nvidia-bans-the-use-of-translation-layers-on-cuda-software/ [https://perma.cc/F2FN-P5S6].

[147] *See generally* Mark Campbell, *Nvidia bans the use of translation layers on CUDA software*, OVERCLOCK 3D, (Mar. 6, 2024), https://overclock3d.net/news/gpu-displays/nvidia-bans-the-use-of-translation-layers-on-cuda-software/ [https://perma.cc/F2FN-P5S6].

[148] Shilov, *supra* note 31.

[149] Liu, *supra* note 32.

[150] *See, e.g.,* Zheming Jin & Jeffrey Vetter, *Evaluating CUDA Portability with HIPCL and DPCT*, OAK RIDGE NATIONAL LABORATORY 1.

[151] For an example of how developers can translate CUDA source code, *see generally*, *id.*

[152] *See AMD HIP Porting Guide*, AMD, https://rocm.docs.amd.com/projects/HIP/en/docs-5.7.1/user_guide/hip_porting_guide.html [https://perma.cc/3VNE-6ME4] (last accessed Jan. 13, 2025).

CUDA restrictions reduce consumer choice by exploiting Nvidia's lead in the software market to lock-in developers, forcing their programs to use Nvidia's GPUs. This has the effect of making it harder for developers to use non-Nvidia GPUs, even if those GPUs are better products. These better products may already exist,[153] but due to CUDA's licensing requirements, developers will be locked into Nvidia's chips. This will negatively affect the quality of outputs: developers' choice of hardware will not be based on cost, performance, or any other metric—it will be preordained.

Licensing restrictions also harm the competitive process by liming binary code portability, a key ingredient in technical innovation. Binary portability—which EULA prohibits—is "clearly desirable" compared to source portability—which EULA allows, because it better enables innovation by reducing porting costs.[154] Code porting "[has] the potential to expand the boundaries of computing particularly in the domain of AI, since individuals can ultimately create a hybrid model of hardware and software resources, exploring the best of both worlds."[155] Here, Nvidia's restrictions on binary portability harm innovation and invention.

Nvidia's licensing restrictions create costs on developers and users of CUDA and Nvidia's GPUs. The restrictions make it harder to run CUDA source code on non-Nvidia GPUs, even those GPUs that may be cheaper, more efficient, or more powerful. This can harm consumers. Because of the licensing restrictions, consumers may be unable to enjoy the technological innovation of competitor firms. Even if AMD or Intel can create more powerful GPU alternatives, licensing restrictions will prevent customers from switching without incurring the additional cost of source cost translation. Nvidia's restrictions on binary code portability plainly harm the competitive process: they seek to extend the firm's market share over GPUs without having improved the quality or decreased the price of the product. EULA imposes anticompetitive restrictions that unduly burden developers, harm innovation, and increase costs on consumers. As will be demonstrated, these restrictions fit neatly within existing monopolization precedent.

B. Nvidia's Licensing Agreements Break the Law

A monopolist violates § 2 of the Sherman Act through the "the willful acquisition or maintenance of [monopoly] power as distinguished from growth or development as a consequence of a superior product, business

---

[153] "[O]n raw specs, [AMD's] MI300X dominates [Nvidia's] H100." Acton, *supra* note 25.
[154] "The principal types of portability usually considered are binary portability (porting the executable form) and source portability (porting the source language representation). Binary portability is clearly desirable, but is possible only across strongly similar environments." James D. Mooney, *Portability and Reusability: Common Issues and Differences*, Ass'n for Computing Mach. 150,151 (1995) https://dl.acm.org/doi/10.1145/259526.259550 [https://perma.cc/W9BD-7WKV].
[155] Zuhair, *supra* note 30.

acumen, or historic accident."[156] The law forbids acts with an "anticompetitive effect"— those acts which harm the competitive process and consumers.[157]

There are two primary law and economics theories of exclusionary conduct: predatory pricing and raising rivals' costs foreclosure.[158] Predatory pricing is a war of attrition—a monopolist lowers prices to force a smaller competitor to either cut prices itself or exit the market entirely.[159] In the paradigmatic predatory pricing case, *Utah Pie*, Continental Baking and other national firms offered very low prices for frozen pies; the Court found that such price discrimination harms competition.[160] Many believe that predatory pricing schemes are "rarely tried, and even more rarely successful."[161] Raising rivals' cost, meanwhile, is both attempted more and more likely to cause harm.[162]

There are several ways that firms use exclusionary tactics to raise rivals' costs, including customer foreclosure and input foreclosure.[163] Nvidia's licensing restrictions fit neatly within the framework of raising rivals' costs foreclosure, whereby a monopolist illegally makes it harder for competitors to gain access to the consumers or key inputs they need to compete.[164] This is particularly clear in comparison to the D.C. Circuit's seminal *Microsoft* opinion. Courts generally analyze these exclusionary tactics under the burden-shifting framework of the rule of reason.[165]

The rule of reason seeks to balance pro- and anti-competitive effects. First, the plaintiff must show that the challenged restraint has a substantial anticompetitive effect.[166] If shown, the burden shifts to the defense to offer a procompetitive rationale.[167] If the defense satisfies that burden, the final burden shifts back to the plaintiff to show that the "procompetitive efficiencies could be reasonably achieved through less anticompetitive

---

[156] Verizon Commc'ns Inc. v. Law Offs. of Curtis V. Trinko, LLP, 540 U.S. 398, 407 (2004) (citing *Grinnell Corp.*, 384 U.S. at 570–71.)).

[157] *Microsoft Corp.,* 253 F.3d at 59.

[158] Steven C. Salop, *The Raising Rivals' Cost Foreclosure Paradigm, Conditional Pricing Practices, and the Flawed Incremental Price-Cost Test*, 81 ANTITRUST L. J., 371, 371 (2017).

[159] *Id.* at 374–75.

[160] *See generally* Utah Pie Co. v. Cont'l Baking Co., 386 U.S. 685, 703 (1967) (stating that Utah Pie was leading the drop in price of frozen pies, thus a jury could reasonably conclude that the drop in price injured or prevented competition).

[161] Matsushita Elec. Indus. Co., Ltd. v. Zenith Radio Corp., 475 U.S. 574, 589 (1986).

[162] *See* Salop, *supra* note 158, at 376.

[163] *See, e.g.*, *id.* at 384–91.

[164] *Id.* at 376.

[165] Nat'l Collegiate Athletic Ass'n v. Alston, 594 U.S. 69, 97 (2021) ("When describing the rule of reason, this Court has sometimes spoken of a three-step, burden-shifting framework as a means for distinguish[ing] between restraints with anticompetitive effect that are harmful to the consumer and restraints stimulating competition that are in the consumer's best interest.") (internal quotations omitted).

[166] *Id.* at 96.

[167] *Id.* at 97.

means."[168] The rule of reason is designed to weigh the practical effects of a challenged restraint and condemn those that harm competition.[169]

This Note will next compare Nvidia's conduct to the conduct at issue in *Microsoft*, particularly examining how network effects incentivize exclusionary conduct. Then, this Note will incorporate additional monopolization theories and harms, assessing Nvidia's conduct under the rule of reason.

1. Nvidia's Conduct Mirrors that of Microsoft

In *Microsoft*, the firm attempted to preserve its operating system monopoly by preventing competitors in the browser market from gaining access to users; defending its monopoly position by using that monopoly to affect competition in an upstream market is anticompetitive.[170]

Microsoft restricted licensees of its operating system—the Original Equipment Manufacturers (OEMs)—by preventing them from distributing browsers other than Microsoft's Internet Explorer.[171] There were several specific ways in which Microsoft restricted the manufacturers and used its power over the operating system market to dominate the browser market. Microsoft barred manufacturers from removing desktop icons, folders, and start menu entries—this prevented the manufacturers from "removing visible means of user access to [Internet Explorer]."[172] Microsoft also prohibited the manufacturers from altering the "initial boot sequence," which the firms had been using to promote competitive browsers like Navigator.[173] Microsoft sought to bolster Internet Explorer's market share, and protect Windows, not by making the products more innovative or easier to use, but by exploiting its monopoly to lock out potential competitors.[174] These restrictions were anticompetitive because the firm used its monopoly position in one market (operating systems) to extend its power over another market (browsers)—this is not competition on the merits.[175]

Microsoft's restrictions were effective and anticompetitive because of the flywheel relationship between browsers and operating systems—the

---

[168] *Id.* (quoting Ohio v. American Express, 585 U.S. 529, 542 (2019)); s*ee also*, Ohio v. American Express, 585 U.S. 529, 541 (2019) ("To determine whether a restraint violates the rule of reason, the parties agree that a three-step, burden-shifting framework applies.").
[169] Cal. Dental Ass'n v. Fed. Trade Comm'n, 526 U.S. 756, 758 (1999) ("What is required is an enquiry meet for the case, looking to a restraint's circumstances, details, and logic.").
[170] *Microsoft Corp.,* 253 F.3d at 60.
[171] *Id.*
[172] *Id.* at 61.
[173] *Id.* at 61–62.
[174] *Id.* at 62 ("Microsoft reduced rival browsers' usage share not by improving its own product but, rather, by preventing OEMs from taking actions that could increase rivals' share of usage."); *id.* at 65 ("This change reduces the usage share of rival browsers not by making Microsoft's own browser more attractive to consumers but, rather, by discouraging OEMs from distributing rival products.").
[175] *Id.* at 56.

network effects. Developers can write applications for operating systems like Windows, for browsers like Internet Explorer, or other middleware products.[176] Developers want to write programs for products with many users, and users want to buy and use products with many programs. In *Microsoft*, competition from Netscape threatened to draw users away from Internet Explorer, which could in turn draw developer attention away from Windows and harm Microsoft's position in the operating system market.[177] Losing the browser market would break Microsoft's flywheel.

The licensing restrictions in *Microsoft* are much like Nvidia's restrictions because of the similar flywheel relationship between Nvidia's software and hardware. As developers create products for one system, users will flock to that system, incentivizing developers and users to align. In *Microsoft*, the firm was threatened when Netscape introduced competition in the browser market, which, because of the network effects between the browser and operating system market, threatened Microsoft's power in both.[178] Similarly, here, competition in the software market threatens Nvidia's dominance over the hardware market by breaking the flywheel.

Nvidia's restrictions on the use of CUDA-created binary code entrench the firm's monopoly position over the GPU market by imposing switching costs for consumers and limiting competitors' access to scale. Both here and in *Microsoft*, licensing requirements nominally still allowed consumers to use a competitor's product but made it more difficult to do so.[179] Both firms used their power in one market to make a different product a default in a second market to protect their flywheel; both practices are anticompetitive. In both cases, licensing restrictions sought to deprive competitors of scale and access to consumers. This is particularly important in markets with network effects, where large consumer bases are essential.[180]

Microsoft presents a clear theory of liability for Nvidia's licensing restrictions—that restrictions in CUDA's licensing agreement anticompetitively seek to entrench the company's monopoly over GPUs. *Microsoft* demonstrates that it is anticompetitive for a monopolist to exclude potential competitors from access to necessary inputs that challenge the

---

[176] *Microsoft Corp.,* 253 F.3d at 60.

[177] *Id.*

[178] *Id.*

[179] In Microsoft's case, users were still free to download a competitor's browser; it was the default status of Microsoft's browser that was anticompetitive. *Id.* at 62 ("Microsoft prohibits OEMs from causing any user interface other than the Windows desktop to launch automatically…."). Similarly, here, developers are still free to translate their CUDA source code.

[180] *What Is the Network Effect?,?*, WHARTON ONLINE (Jan. 17, 2023), https://online.wharton.upenn.edu/blog/what-is-the-network-effect/ [https://perma.cc/32MD-BG3A].

monopolist's position.[181] This is the precise effect of Nvidia's conduct: potential competitors in the GPU market—chip designers—will be unable to compete with Nvidia's monopoly, because the challengers are cut off from the necessary software inputs; programs will only be designed for Nvidia's chips.

2. *Additional Case Law Supports this Theory of Liability*

Additional case law demonstrates Nvidia's liability by clarifying theories of harm. In particular, case law supports the theory that a monopolist breaks the law by imposing switching costs on consumers, reducing consumer choice, and limiting the ability of potential competitors to access scale or other necessary inputs.

Third Circuit precedent particularly supports the theory that Nvidia's licensing restrictions are anticompetitive because they create transaction costs, limit consumer choice, and prevent rivals from gaining scale.

In *LePage's*, the Third Circuit held that 3M violated the Sherman Act when it imposed exclusive dealing contracts on its buyers, preventing competitors from gaining a significant volume of sales, and coordinating efforts to inflate retail prices.[182] The Third Circuit endorsed the approach taken in *Microsoft*: that it is anticompetitive for a monopolist, in response to a competitive challenge, to foreclose access to key inputs or scale.[183]

The court maintained this approach in *Dentsply*, when it held that Dentsply unlawfully maintained its monopoly over prefabricated artificial teeth through exclusionary contracts.[184] Dentsply refused to sell its teeth to any dealer who offered a competitor's product, foreclosed competitors from access to dealer networks, and barred them from gaining the scale necessary to challenge the firm's position.[185] Here, the court explicitly stated that a

---

[181] Eleanor M. Fox, *What Is Harm to Competition? Exclusionary Practices and Anticompetitive Effect*, 70 ANTITRUST L.J. 371, 390 (2002) ("We might thus interpret the Microsoft holding as follows: Conduct that intentionally, significantly, and without business justification excludes a potential competitor from outlets (even though not in the relevant market), where access to those outlets is a necessary though not sufficient condition to waging a challenge to a monopolist and fear of the challenge prompts the conduct, is 'anticompetitive.'").

[182] LePage's Inc. v. 3M, 324 F.3d 141, 159–63 (3d Cir. 2003).

[183] *Id.* at 159–60.

[184] *Dentsply Int'l, Inc.,* 399 F.3d at 185 (3d Cir. 2005) ("For more than fifteen years, Dentsply has operated under a policy that discouraged its dealers from adding competitors' teeth to their lines of products. In 1993, Dentsply adopted 'Dealer Criterion 6.' It provides that in order to effectively promote Dentsply–York products, authorized dealers 'may not add further tooth lines to their product offering.'").

[185] *See id.*

monopolist breaks the law if its practices prevent entry of rivals or severely shrink the market.[186]

Finally, in *ZF Meritor*, the court similarly condemned long-term agreements between a monopolist and its customers because they effectively foreclosed the market to competitors and thus harmed competition.[187] The court found that the monopolist's exclusionary conduct decreased market competition by preventing the growth of potential competitors.[188]

ZF Meritor also highlighted how monopolists use transaction, or switching costs, to illegally defend their power.[189] In *ZF Meritor*, Eaton, the monopolist, prevented ZF Meritor from marketing its products in Eaton's catalogs, the most common way for drivers to find parts.[190] While Eaton argued that drivers could still find ZF Meritor's products through other means of distribution, the court found that this transaction cost—switching to a new means of distribution—meant that drivers were very unlikely to do so.[191] Together, precedent from *ZF Meritor* in the Third Circuit and *Microsoft*[192] in the D.C. Circuit demonstrates how monopolists employ switching costs to insulate power, lock consumers into their ecosystems, and prevent competitors from gaining scale.

The Eleventh Circuit similarly condemned a monopolist's exclusionary strategy in *McWane*, finding that the monopolist intentionally and illegally slowed the development of a potential threat to its power.[193] In *McWane* firms who bought from Star, a potential competitor to the monopolist McWane, either lost rebates from McWane or were cut off from McWane's catalogue.[194] McWane effectively raised Star's costs to compete and the firm, unable to grow its consumer base, could not challenge McWane's monopoly.[195] The Eleventh Circuit followed with precedent from the Third and D.C. Circuits that a monopolist breaks the law when it deprives rivals of necessary inputs or channels.

Nvidia's EULA restrictions impose similar transaction costs on consumers, deprive rivals of scale, and illegally insulate the firm's monopoly. Here, Nvidia's licensing restrictions are likely to deprive competitors of the key inputs that they need to compete. Software programs

---

[186] *Id.* at 191 ("Assessing anti-competitive effect is important in evaluating a challenge to a violation of Section 2…. [I]t is not necessary that all competition be removed from the market. The test is not total foreclosure, but whether the challenged practices bar a substantial number of rivals or severely restrict the market's ambit.").

[187] ZF Meritor, LLC v. Eaton Corp., 696 F.3d 254, 289 (3d Cir. 2012).

[188] *Id.*

[189] *Id.* at 287.

[190] *Id.*

[191] *Id.*

[192] *Microsoft Corp.*, 253 F.3d at 71 ("Microsoft's deals with the IAPs clearly have a significant effect in preserving its monopoly; they help keep usage of Navigator below the critical level necessary for Navigator or any other rival to pose a real threat to Microsoft's monopoly.").

[193] McWane, Inc. v. Fed. Trade Comm'n, 783 F.3d 814, 840 (11th Cir. 2015).

[194] *Id.* at 821.

[195] *Id.*

are a key input in the market for GPUs—it is through CUDA that Nvidia has cemented its power[196]—and the licensing restrictions will have the effect of preventing AMD, Intel, or other potential competitors from accessing these inputs without overcoming transaction costs. This conduct is very similar to that of *McWane* or the series of Third Circuit monopolization cases.

Nvidia's conduct, like that at issue in *ZF Meritor*, harms competition because of how transaction costs protect the monopolist's power. In *ZF Meritor*, the court considered these costs when deciding the extent to which Eaton's conduct was likely to foreclose the market to ZF Meritor and other challengers.[197] Similarly, here, the transaction costs of using different software products are likely to insulate Nvidia's monopoly over the hardware market. Like in *ZF Meritor*, there is a workaround that avoids the monopolist's exclusionary tactic—in that case, a truck dealer could request unpublished components,[198] and here, a developer could translate CUDA-created source code into a different programming language. In both cases, the transaction cost means that the exclusionary tactic is very likely to succeed. Thus, the monopolist's act is likely to harm competition.

Nvidia's licensing restrictions also seek to raise costs for competitors. In *McWane*, the monopolist illegally tried to raise the costs of Star and other competitors by forcing the firms to add new products to their lines.[199] If Star wanted to compete in the domestic fittings market, it would have to invest in a catalogue of new products beyond domestic fittings.[200] This concern with a monopolist's efforts to raise rivals costs are well documented in *LePage's*[201] and *ZF Meritor*.[202] Nvidia's licensing restrictions will have similar practical effects to those restrictions in *McWane*: now, to compete with Nvidia's GPUs in the hardware market, a firm must also now be able to compete in the software market. This raises rivals' cost to compete—a paradigmatic exclusionary tactic[203] that is likely to harm consumers.

---

[196] Waters, *supra* note 26.

[197] *ZF Meritor, LLC*, 696 F.3d at 287 ("Although truck buyers could request unpublished components, doing so involved additional transaction costs, and in practice, meant that truck buyers were far more likely to select a product listed in the data book.").

[198] *Id.*

[199] *McWane, Inc.,* 783 F.3d at 821.

[200] *Id.* ("McWane employees described the nascent Full Support Program as a strategy to '[f]orce [d]istribution to [p]ick their [h]orse', which would '[f]orce[ ] Star[ ] to absorb the costs associated with having a more full line before they can secure major distribution.'") (alterations in original).

[201] *LePage's Inc.*, 324 F.3d at 160 ("[T]he jury could have reasonably found that 3M's exclusionary conduct cut LePage's off from key retail pipelines necessary to permit it to compete profitably.").

[202] *ZF Meritor, LLC*, 696 F.3d at 271 ("In some cases, a dominant firm may be able to foreclose rival suppliers from a large enough portion of the market to deprive such rivals of the opportunity to achieve the minimum economies of scale necessary to compete.").

[203] Salop, *supra* note 158.

Unable to access critical inputs and one pricey transaction cost away from consumers, competitor firms will be unable compete and rein in Nvidia's monopoly. This conduct represents a typical example of how monopolists seek to protect their power through exclusionary and anticompetitive tactics. Under the burden-shifting rule of reason framework, a court will next consider procompetitive justifications for the firm's behavior.[204]

C. *Procompetitive Arguments for CUDA's Licensing Restrictions*

Sherman Act precedents outlines the requirements of procompetitive justifications and demonstrates the strengths and weaknesses of possible defenses.

Procompetitive defenses must be specific and substantiated—mere claims of enhanced efficiency are insufficient.[205] Nvidia's procompetitive justifications would likely be based on a close reading of *Microsoft* and other monopolization cases where courts assessed the merits of procompetitive arguments. Microsoft justified its exclusive dealing contracts by arguing that it wanted to keep developers focused on Microsoft's APIs.[206] This defense––that Microsoft wanted to preserve its power—was an insufficient justification for their exclusive dealing.[207] This could be relevant if Nvidia were to argue that its CUDA restrictions were to keep developers writing programs in Nvidia's software. Much like Microsoft's claim here, that defense would fall flat.

Microsoft also argued that its license restrictions are a valid exercise of its rights as a holder of a valid copyright—this defense "border[ed] upon the frivolous" thus it is unlikely Nvidia would even try.[208] Also relevant was Microsoft's argument that, despite restrictions on the distribution of Netscape's browser, there were still alternate paths of distribution.[209] However, as demonstrated in both *ZF Meritor* and *Microsoft*, foreclosing a rival from cost-efficient or popular distribution channels is an insufficient defense.[210]

---

[204] *See Nat'l Collegiate Athletic Ass'n*, 594 U.S. at 96 (2021); s*ee also Am. Express Co.,* 585 U.S. at 541 (2018) ("If the plaintiff carries its burden, then the burden shifts to the defendant to show a procompetitive rationale for the restraint."); Eastman Kodak Co., 504 U.S. at 483 ("[R]espondents have presented evidence that Kodak took exclusionary action to maintain its parts monopoly and used its control over parts to strengthen its monopoly share of the Kodak service market. Liability turns, then, on whether 'valid business reasons can explain Kodak's actions.").

[205] *See, e.g.*, *Microsoft Corp.*, 253 F.3d at 66 ("Although Microsoft does make some general claims regarding the benefits of integrating the browser and the operating system…("Our vision of deeper levels of technical integration is highly efficient and provides substantial benefits to customers and developers."), it neither specifies nor substantiates those claims.") (citation omitted).

[206] *Id.* at 71.

[207] *Id.*

[208] Id. at 63.

[209] *Id.* at 64.

[210] *Id.* at 63; *ZF Meritor, LLC*, 696 F.3d at 287.

Nvidia may also argue that while the CUDA licensing agreements do make it harder for developers to run their code on non-Nvidia chips, the licensing terms do not make it impossible. This is true: under the licensing agreement, developers can still export their source code—just not their binary code—and run that code on other chips.[211] Microsoft made a similar argument, that Netscape was not entirely barred from distributing its product, and consumers could still access a competitor's product.[212] But the D.C. Circuit found Microsoft still liable, specifically because it barred its rivals from "cost-efficient" means of distribution.[213] Nvidia's software licensing terms similarly bar its rivals like AMD from accessing efficient distribution, by creating costs and friction for developers who wish to write software on CUDA and export their code onto non-Nvidia chips.

Nvidia may attempt an efficiency argument, that the restrictions are better for consumers, which even may be true— running CUDA code on a non-Nvidia chip does require consumers to incur a switching cost away from the Nvidia ecosystem. But that is a fact of Nvidia's own making through its licensing restrictions, and this decision should be left in the hands of consumers. Nvidia's stronger procompetitive claims would likely be based on specific technical arguments, perhaps focusing on enhanced security or privacy, perhaps arguing that the translation of CUDA source code into a different programming language could create vulnerabilities for the firm or its users.

1. *The Anticompetitive Effects of Nvidia's Conduct Outweigh any Rationale that Can Survive the "Coercive" Test*

There is one significant hurdle that Nvidia faces for its procompetitive justification: courts disregard such justifications when a monopolist's conduct is "coercive."[214] For example, the Third[215], Tenth[216], and Eleventh Circuits[217] all disregard such procompetitive justifications for "coercive" conduct. In *Chase Manufacturing*, for example, the Tenth Circuit found significant evidence of coercion when the monopolist—Johns Manville— left its distributors with an "all-or-nothing" choice: to either stop doing business with a competitor, or lose access to the monopolist and its product

---

[211] "You may not reverse engineer, decompile or disassemble any portion of the output generated using SDK elements for the purpose of translating such output artifacts to target a non-NVIDIA platform." Nvidia End User License Agreement § 1.2(8) (2024), https://docs.nvidia.com/cuda/eula/index.html [https://perma.cc/7DDN-LAS3].

[212] *Microsoft Corp.,* 253 F.3d at 64.

[213] *Id.*

[214] Chase Mfg., Inc. v. Johns Manville Corp., 84 F.4th 1157, 1172 (10th Cir. 2023) ("And we 'throw [any procompetitive assumption] out the window when record evidence suggests coercion by the monopolist.'") (quoting In re Epi Pen (Epinephrine Injection, USP) Marketing, Sales Practices and Antitrust Litigation, 44 F.4th 959, 996 (10th Cir. 2022).

[215] *ZF Meritor, LLC*, 696 F.3d at 285.

[216] *See Chase Mfg., Inc.,* 84 F.4th at 1172.

[217] *See McWane, Inc*., 783 F.3d at 834.

entirely.[218] Such "all-or-nothing" choices are a hallmark of coercive monopolist conduct.[219] A finding of coercion ends the burden-shifting exercise and warrants a finding of illegal monopoly maintenance.

In the Third Circuit, conduct was not coercive when a monopolist merely threatened the loss of a larger discount,[220] but was coercive in *ZF Meritor*, where the monopolist, Eaton, coerced manufacturers to enter unfavorable deals.[221] Because manufacturers relied on Eaton to reach consumers, they had no meaningful choice and were forced to comply with the firm's restrictions.[222]

The conduct at issue with Nvidia is like the coercive conduct in *ZF Meritor* or *Dentsply*, where monopolists used their size not just to impose costs on competitors, but also to impose agreements that harmed competition.

Here, Nvidia's conduct is coercive: the firm uses its software to defend its monopoly over hardware by depriving customers of meaningful choice. Software developers must write their programs in CUDA, because if not, they cannot run their programs on other hardware without incurring switching costs. And cloud service providers must buy Nvidia's GPUs, because if not, they cannot access the library of AI projects written for CUDA.[223] Entire industries rely on the firm's GPUs.[224] The rationale for this dependence is two-fold: first, a story of genuine innovation, investment, and wonder, industry at its best and boldest. But this is also a story of lock-out, of restraint, of a monopolist, atop the game, seeking to ice out potential competitors through brute force, not competition. The software licensing restrictions in CUDA are coercive—they force developers to make an "all-or-nothing" decision about their development ecosystem—and thus any procompetitive justifications are insufficient.

Nvidia's role in creating this world cannot be overlooked: Nvidia chips have powered the revolution of AI products like ChatGPT that seem poised to change everything.[225] But now, Nvidia's conduct threatens to harm consumers and disrupt innovation by entrenching its own market power. As

---

[218] *Chase Mfg., Inc.,* 84 F.4th at 1172.

[219] *See, e.g.,* Sanofi-Aventis U.S., LLC v. Mylan, Inc. (In re EpiPen Epinephrine Injection, Mktg., Sales Pracs. & Antitrust Litig.), 44 F.4th 959, 996 (10th Cir. 2022).

[220] Eisai, Inc. v. Sanofi Aventis U.S., LLC, 821 F.3d 394, 407 (3d Cir. 2016) ("Moreover, the threat of a lost discount is a far cry from the anticompetitive conduct at issue in *ZF Meritor* or *Dentsply*.")

[221] *ZF Meritor, LLC,* 696 F.3d at 285.

[222] *Id.*

[223] *See, e.g.,* Clark, *supra* note 84.

[224] *See generally* Waters, *supra* note 26; Madhumita Murgia & George Hammond, *OpenAI's Sam Altman in Talks with Middle East Backers Over Chip Venture,* FIN. TIMES (Jan. 20, 2024), https://www.ft.com/content/1cdaadc3-b384-4f50-88ff-291c062c8376 [https://perma.cc/7WLU-VD6M].

[225] Austin Carr & Ian King, *How Nvidia Became ChatGPT's Brain and Joined the $1 Trillion Club*, BLOOMBERG L. (June 15, 2023), https://www.bloomberg.com/news/features/2023-06-15/nvidia-s-ai-chips-power-chatgpt-and-multibillion-dollar-surge [https://perma.cc/V6X6-W66V].

competitors develop hardware to compete with Nvidia's GPUs, the firm has moved to lock-in its incumbency and to lock-out the benefits of innovation and competition. CUDA's licensing restrictions may make it easier for developers to stay within Nvidia's walled garden, but by reducing choice and innovation, they threaten to undermine the AI revolution. The anticompetitive effects of this conduct plainly outweigh Nvidia's purported benefits.

III. WHAT TO DO?

Monopolization litigation is expensive, timely, and often slow to demonstrate value.[226] If the goal of monopolization litigation is improved economic performance, some suggest that these cases are an imperfect means to an end.[227] However, monopolization cases also offer a unique opportunity to reshape industries and reinvigorate innovation. Columbia Law Professor Tim Wu argues that despite short term "chaos" after the AT&T settlement, in the long term the breakup sparked "entirely new types of industries" that were previously "unimaginable" —the settlement was a key catalyst of the internet revolution.[228] A monopolization case against Nvidia would be similarly protracted and costly and would likely overwhelm the resource-strapped Department of Justice and Federal Trade Commission.[229] But, as Wu and others argue, these cases have potential to spark innovation and competition.

Such a lawsuit is potentially in the works. The Justice Department sent subpoenas to Nvidia,[230] although the scope of the investigation is still unclear. Jonathan Kanter, the former head of the Antitrust Division, noted that antitrust regulators were examining how chipmakers like Nvidia chose

---

[226] Famously, Robert Bork dubbed the IBM monopolization case the "antitrust division's Vietnam." The case lasted thirteen years, six of which were at trial. Archive of *U.S. vs. IBM*, N.Y. TIMES (Feb. 15, 1981), https://www.nytimes.com/1981/02/15/business/us-vsibm.html [https://perma.cc/SKT9-ZFGV]; WU, supra note 52, at 11.

[227] William E. Kovacic, *Failed Expectations: The Troubled Past and Uncertain Future of the Sherman Act as a Tool for Deconcentration*, 74 IOWA L. REV 1105, 1105 (1989) ("Federal enforcement officials have mounted memorable campaigns to disassemble leviathans of American business, yet the tantalizing goal of improving the economic and political order by restructuring dominant firms frequently has eluded its pursuers.").

[228] WU, *supra* note 52, at 96-–7.

[229] In April 2024, the DOJ and FTC had five pending monopolization cases concerning: Google Search, Google AdTech, Meta, Amazon, and Apple. Brian C. Albrecht & Daniel J. Gilman, *Ranking the Big Tech Monopolization Cases: Some Economists' Perspectives*, YALE J. ON REGUL. (Apr. 9, 2024), https://www.yalejreg.com/nc/ranking-the-big-tech-monopolization-cases-some-economists-perspectives-by-brian-c-albrecht-daniel-j-gilman/#:~:text=There%20are%20at%20least%20five,future%20of%20the%20digital%20economy [https://perma.cc/N48X-3AKD].

[230] Ian King & Lean Nylen, *Nvidia gets Justice Department subpoena in escalating antitrust inquiry*, LOS ANGELES TIMES (Sept. 4, 2024), https://www.latimes.com/business/story/2024-09-04/nvidia-gets-doj-subpoena-in-escalating-antitrust-probe [https://perma.cc/H9CN-F267].

to allocate their chips.[231] But, were the Justice Department to focus its inquiry on Nvidia's software licensing agreements, this Note demonstrates how the DOJ could focus its investigation and complaint. The DOJ could also seek to bring its complaint in a District with favorable case law on monopolization case law. One favorite would be the District of Columbia District Court: *Microsoft* is binding in D.D.C., and recently, Judge Amit P. Mehta found that Google had illegally maintained its monopoly in the market for online search.[232] The Department could point to a string of cases from Circuits around the country. Across the nation, federal appellate courts have held it illegal for a monopolist to suppress competition by denying rivals access to scale.

Antitrust enforcement is just one way to reinvigorate competition—there are several other ways that American enforcers and policymakers can undo the anticompetitive effects of Nvidia's exclusionary conduct.

For one, the federal government can leverage CHIPS Act funding to restrict exclusionary conduct.[233] The CHIPS Program Office (CPO) could condition funding and investment on firms' commitment to either refrain from, or halt, anticompetitive practices like exclusive dealing.[234] The CPO could also work alongside antitrust enforcers like the FTC or DOJ to better understand competitive conditions within the semiconductor industry. Both agencies are developing expertise in artificial intelligence and cloud computing. For example, the FTC sought public comment on the business of cloud computing, examining both competition and consumer protection issues in the industry.[235] And, as detailed, the DOJ has launched an investigation of Nvidia.[236] Conditional CHIPS Act funding is likely to have greater effect on exclusionary conduct in semiconductor manufacturing, where most money is being spent,[237] but it does present one way for the government to address exclusion in the design process. The CPO could also take a more direct—and risky—approach, and invest directly in those firms

---

[231] Stephen Morris et al., *Big Tech's AI dealmaking needs 'urgent' scrutiny, says US antitrust enforcer*, FIN. TIMES (June 6, 2024), https://www.ft.com/content/97b45759-36e0-4f5b-9c6a-ae0580f9a29b [https://perma.cc/XNR9-ERPF].

[232] *Google, LLC*, No. 20-cv-3010 (APM) Dkt. No. 1033.

[233] *See, e.g.*, Todd Achilles et al., *supra* note 39, at 48 ("[C]riteria for receiving CHIPS funding should include a series of clear commitments to not take part in a range of practices that likely violate the antitrust laws. The list of such anticompetitive practices should include exclusive dealing, tying, discriminatory pricing arrangements, predatory pricing, and other coercive contracts.").

[234] *Id.*

[235] Nick Jones, *Cloud Computing RFI: What we heard and learned*, FED. TRADE COMM'N (Nov. 16, 2023), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/11/cloud-computing-rfi-what-we-heard-learned [https://perma.cc/PF6E-YKQ7].

[236] King & Nylen, *infra* note 230.

[237] *See, e.g.*, Kif Leswing, *Intel Awarded up to $8.5 Billion in CHIPS Act Grants, with Billions More in Loans Available*, CNBC (Mar. 20, 2024, 5:00 AM), https://www.cnbc.com/2024/03/20/intel-awarded-up-to-8point5-billion-from-chips-act-with-loans-available.html [https://perma.cc/ENL8-89C9] ("The White House said Intel has been awarded up to $8.5 billion in CHIPS Act funding, as the Biden administration ramps up its effort to bring semiconductor manufacturing to U.S. soil.").

seeking to unseat Nvidia's lead in either software or hardware. Or, perhaps more simply, could invest in software solutions to ease the switching costs between CUDA-created source code and non-Nvidia GPUs. A more efficient translator would enable developers to run programs on GPUs made designed by other firms.

The CHIPS Program Office could also adopt chip-making standards to promote interoperability and lower switching costs.[238] . These chip-making standards could promote interoperability and lower switching costs.[239] These standards would likely be voluntary, or as conditions of funding. Taken together, the federal government could seek to unwind the exclusionary effects of Nvidia's conduct by promoting competition, conditioning investment, and adopting technical standards that lower switching costs and ensure interoperability. But, these voluntary or conditional standards would likely be ineffective because of the vast entry barriers surrounding Nvidia's lead, and the power of the flywheel between its software and hardware. Although antitrust litigation is lengthy, a DOJ monopolization suit may be the most effective method of rejuvenating competition in the GPU market.

Realities of the semiconductor market also suggest a different way to think about regulation—that of the "natural monopoly." Natural monopolies tend to occur where marginal costs are very low[240]—i.e., after the upfront investment is made, adding a new customer is relatively cheap—and natural monopolies have high barriers to entry—usually, large capital investment is required to enter the market.[241] Some argue that chip manufactures like ASML are natural monopolies,[242] while others argue that machine learning applications themselves are natural monopolies.[243] While the GPU market likely is not that of a natural monopoly, there are many firms along the semiconductor supply chain, particularly at the manufacturing stage, that could be.[244] Regulating these markets as natural monopolies is one possible remedy for consolidation across the industry more generally, but is unlikely to apply to the GPU design process in particular.

---

[238] Todd Achilles et al., *supra* note 39, at 48 ("Develop thicker markets by…(b) developing a set of chip-making standards that lower switching costs and promote interoperability…").
[239] *Id.*
[240] DAVID SHAPIRO ET AL., PRINCIPLES OF MICROECONOMICS 218 (3d ed. 2022), https://openstax.org/details/books/principles-microeconomics-3e [https://perma.cc/CW6H-AD7S].
[241] *Id.* at 217–18.
[242] M Ramirez, *ASML: Natural Monopoly in the Semiconductor Market Is a Buy*, SEEKING ALPHA (Oct. 10, 2023, 11:15 AM), https://seekingalpha.com/article/4639962-asml-natural-monopoly-in-semiconductor-market [https://perma.cc/Q4DA-SRLX].
[243] Tejas N. Narechania, *Machine Learning as Natural Monopoly*, 107 IOWA L. REV. 1543, 1587 (2022).
[244] For instance, ASML has a monopoly on "the extreme ultraviolet lithography machines that are critical for the production of advanced chips. The waiting list for the machines, which cost more than $300mn each, averages around two years." Yoon, *supra* note 27. This firm may be a natural monopoly.

The remedy appears far simpler than the problem. If CUDA-created binary code could run on any GPU, then firms would not have to invest in both the software and hardware markets to compete with Nvidia's GPUs. Developers could create programs in CUDA and run those programs on GPUs created by Nvidia, ARM, Intel, or a new entrant. Cloud providers could seamlessly switch between GPUs, making purchasing decisions based on price, quantity, and quality, incentivizing innovation and efficiency. Unforeseeable beneficiaries could reap the promises of the AI revolution. What stands in the way of this competitive and innovative future is Nvidia's monopoly and the deep moat surrounding its power. Software interoperability would fill in that moat, lower the drawbridge, and let the people into the castle. But Nvidia has every incentive to battle such remedy to the last dollar, to deepen the moat of its power, to power the flywheel of its success.

IV. CONCLUSION

There is no shortage of promises that artificial intelligence will change our world.[245] What is more likely, however, is that artificial intelligence will merely reflect our world. Trained on the vast stores of human knowledge and experience, artificial intelligence will learn, reproduce, and entrench both the angels and demons of our ways. Automated systems contribute to discrimination in key areas like healthcare, housing, and employment, deepening social inequities.[246] Artificial intelligence may simply be another black mirror which reflects our greatest successes and our deepest sins.

---

[245] *See, e.g.*, Matthew Huddle et al., *Generative AI Will Transform Health Care Sooner Than You Think*, BOSTON CONSULTING GRP. (June 22, 2023), https://www.bcg.com/publications/2023/how-generative-ai-is-transforming-health-care-sooner-than-expected [https://perma.cc/4T8A-4BG6] (addressing how generative AI will transform health care); *How AI Will Change the Battlefield Forever*, BLOOMBERG (Mar. 12, 2024), [https://perma.cc/DAU8-KBXP]; (discussing the use of AI in warfare); Vala Afshar et al., *Generative AI Will Change Customer Service Forever. Here's How we Get There*, ZDNET (Mar 12. 2024, 4:01 AM), www.zdnet.com/article/generative-ai-will-change-customer-service-forever-heres-how-we-get-there/ [https://perma.cc/N6PZ-5QLJ] (discussing how generative AI will change customer service); Linda Liu et al., *Generative AI and the Future of New York*, MCKINSEY & CO (Mar. 4, 2024), [https://perma.cc/YU2D-46E6] (predicting how generative AI will change the New York City labor markets); Matthew Sparkes, *Could an AI Replace All Music Ever Recorded with Taylor Swift Covers?*, NEW SCIENTIST (Mar. 12, 2024), https://www.newscientist.com/article/2418740-could-an-ai-replace-all-music-ever-recorded-with-taylor-swift-covers/ [https://perma.cc/U3J7-UBKM] (discussing how a "rogue artificial intelligence obsessed with Taylor Swift" could replace all recorded music with Taylor Swift covers.).

[246] "[C]ompanies' growing reliance on automated systems is creating new forms and mechanisms for discrimination based on statutorily protected categories, including in critical areas such as housing, employment, and healthcare. For example, some employers' automated systems have reportedly learned to prefer men over women. Meanwhile, a recent investigation suggested that lenders' use of educational attainment in credit underwriting might disadvantage students who attended historically Black colleges and universities." Trade Regulation Rule on Commercial Surveillance and Data Security, 87 Fed. Reg. 51,273, 51,275–7676 (proposed Aug. 11, 2022) (to be codified at 16 C.F.R. Ch. I.).

   Monopolization and exclusionary conduct along the semiconductor supply chain is just one more way that the future of artificial intelligence will reflect and exacerbate our ways. The firms that stand to win our next decade are the same firms that monopolized our last—Amazon, Google, Microsoft, Apple, and now Nvidia. In a decade, we may be in the same place we are now, litigating costly battles to unwind a decade of concentration.

   Nvidia was in the right place at the right time: it saw the future and it bet big on GPUs.[247] But when ingenious innovation threatened Nvidia's monopoly by breaking the flywheel between its software and hardware, the firm harmed competition and illegally defended its monopoly. CUDA's licensing restrictions violate the Sherman Act.

---

[247] *See generally* Waters, *supra* note 26.