

UNITED STATES  
SECURITIES AND EXCHANGE COMMISSION  
Washington, D.C. 20549

FORM 10-K

☒ ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934  
For the fiscal year ended January 28, 2024

OR

☐ TRANSITION REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934

Commission file number: 0-23985



NVIDIA CORPORATION

(Exact name of registrant as specified in its charter)

Delaware  
(State or other jurisdiction of  
incorporation or organization)

94-3177549  
(I.R.S. Employer  
Identification No.)

2788 San Tomas Expressway, Santa Clara, California  
(Address of principal executive offices)

95051  
(Zip Code)

Registrant's telephone number, including area code: (408) 486-2000

Securities registered pursuant to Section 12(b) of the Act:

Title of each class  
Common Stock, \$0.001 par value per share

Trading Symbol(s)  
NVDA

Name of each exchange on which registered  
The Nasdaq Global Select Market

Securities registered pursuant to Section 12(g) of the Act:  
None

Indicate by check mark if the registrant is a well-known seasoned issuer, as defined in Rule 405 of the Securities Act. Yes ☐ No ☒

Indicate by check mark if the registrant is not required to file reports pursuant to Section 13 or Section 15(d) of the Act. Yes ☐ No ☒

Indicate by check mark whether the registrant (1) has filed all reports required to be filed by Section 13 or 15(d) of the Securities Exchange Act of 1934 during the preceding 12 months (or for such shorter period that the registrant was required to file such reports), and (2) has been subject to such filing requirements for the past 90 days. Yes ☒ No ☐

Indicate by check mark whether the registrant has submitted electronically every Interactive Data File required to be submitted pursuant to Rule 405 of Regulation S-T (§232.405 of this chapter) during the preceding 12 months (or for such shorter period that the registrant was required to submit such files). Yes ☒ No ☐

Indicate by check mark whether the registrant is a large accelerated filer, an accelerated filer, a non-accelerated filer, a smaller reporting company, or an emerging growth company. See the definitions of "large accelerated filer," "accelerated filer," "smaller reporting company," and "emerging growth company" in Rule 12b-2 of the Exchange Act.

Large accelerated filer ☒ Accelerated filer ☐ Non-accelerated filer ☐ Smaller reporting company ☐ Emerging growth company ☐

If an emerging growth company, indicate by check mark if the registrant has elected not to use the extended transition period for complying with any new or revised financial accounting standards provided pursuant to Section 13(a) of the Exchange Act. ☐

Indicate by check mark whether the registrant has filed a report on and attestation to its management's assessment of the effectiveness of its internal control over financial reporting under Section 404(b) of the Sarbanes-Oxley Act (15 U.S.C. 7262(b)) by the registered public accounting firm that prepared or issued its audit report. ☒

If securities are registered pursuant to Section 12(b) of the Act, indicate by check mark whether the financial statements of the registrant included in the filing reflect the correction of an error to previously issued financial statements. ☐

Indicate by check mark whether any of those error corrections are restatements that required a recovery analysis of incentive-based compensation received by any of the registrant's executive officers during the relevant recovery period pursuant to §240.10D-1(b). ☐

Indicate by check mark whether the registrant is a shell company (as defined in Rule 12b-2 of the Act). Yes ☐ No ☒

The aggregate market value of the voting stock held by non-affiliates of the registrant as of July 28, 2023 was approximately \$1.1 trillion (based on the closing sales price of the registrant's common stock as reported by the Nasdaq Global Select Market on July 28, 2023). This calculation excludes 105 million shares held by directors and executive officers of the registrant. This calculation does not exclude shares held by such organizations whose ownership exceeds 5% of the registrant's outstanding common stock that have represented to the registrant that they are registered investment advisers or investment companies registered under section 8 of the Investment Company Act of 1940.

The number of shares of common stock outstanding as of February 16, 2024 was 2.5 billion.

DOCUMENTS INCORPORATED BY REFERENCE

Portions of the registrant's Proxy Statement for its 2024 Annual Meeting of Shareholders to be filed with the Securities and Exchange Commission pursuant to Regulation 14A not later than 120 days after the end of the fiscal year covered by this Annual Report on Form 10-K are incorporated by reference into Part III, Items 10-14 of this Annual Report on Form 10-K.

**NVIDIA Corporation**  
**Table of Contents**

	<b>Page</b>
<b>Part I</b>	
<a href="#">Item 1. Business</a>	<a href="#">4</a>
<a href="#">Item 1A. Risk Factors</a>	<a href="#">13</a>
<a href="#">Item 1B. Unresolved Staff Comments</a>	<a href="#">31</a>
<a href="#">Item 1C. Cybersecurity</a>	<a href="#">31</a>
<a href="#">Item 2. Properties</a>	<a href="#">32</a>
<a href="#">Item 3. Legal Proceedings</a>	<a href="#">32</a>
<a href="#">Item 4. Mine Safety Disclosures</a>	<a href="#">32</a>
<b>Part II</b>	
<a href="#">Item 5. Market for Registrant's Common Equity, Related Stockholder Matters and Issuer Purchases of Equity Securities</a>	<a href="#">32</a>
<a href="#">Item 6. [Reserved]</a>	<a href="#">33</a>
<a href="#">Item 7. Management's Discussion and Analysis of Financial Condition and Results of Operations</a>	<a href="#">34</a>
<a href="#">Item 7A. Quantitative and Qualitative Disclosures About Market Risk</a>	<a href="#">43</a>
<a href="#">Item 8. Financial Statements and Supplementary Data</a>	<a href="#">44</a>
<a href="#">Item 9. Changes in and Disagreements with Accountants on Accounting and Financial Disclosure</a>	<a href="#">44</a>
<a href="#">Item 9A. Controls and Procedures</a>	<a href="#">44</a>
<a href="#">Item 9B. Other Information</a>	<a href="#">45</a>
<a href="#">Item 9C. Disclosure Regarding Foreign Jurisdictions that Prevent Inspections</a>	<a href="#">45</a>
<b>Part III</b>	
<a href="#">Item 10. Directors, Executive Officers and Corporate Governance</a>	<a href="#">45</a>
<a href="#">Item 11. Executive Compensation</a>	<a href="#">46</a>
<a href="#">Item 12. Security Ownership of Certain Beneficial Owners and Management and Related Stockholder Matters</a>	<a href="#">46</a>
<a href="#">Item 13. Certain Relationships and Related Transactions, and Director Independence</a>	<a href="#">46</a>
<a href="#">Item 14. Principal Accountant Fees and Services</a>	<a href="#">46</a>
<b>Part IV</b>	
<a href="#">Item 15. Exhibit and Financial Statement Schedules</a>	<a href="#">47</a>
<a href="#">Item 16. Form 10-K Summary</a>	<a href="#">83</a>
<a href="#">Signatures</a>	<a href="#">84</a>

## Where You Can Find More Information

Investors and others should note that we announce material financial information to our investors using our investor relations website, press releases, SEC filings and public conference calls and webcasts. We also use the following social media channels as a means of disclosing information about the company, our products, our planned financial and other announcements and attendance at upcoming investor and industry conferences, and other matters and for complying with our disclosure obligations under Regulation FD:

NVIDIA X Account (<https://twitter.com/nvidia>)

NVIDIA Corporate Blog (<http://blogs.nvidia.com>)

NVIDIA Facebook Page (<https://www.facebook.com/nvidia>)

NVIDIA LinkedIn Page (<http://www.linkedin.com/company/nvidia>)

NVIDIA Instagram Page (<https://www.instagram.com/nvidia>)

In addition, investors and others can view NVIDIA videos on YouTube (<https://www.YouTube.com/nvidia>).

The information we post through these social media channels may be deemed material. Accordingly, investors should monitor these accounts and the blog, in addition to following our press releases, SEC filings and public conference calls and webcasts. This list may be updated from time to time. The information we post through these channels is not a part of this Annual Report on Form 10-K. These channels may be updated from time to time on NVIDIA's investor relations website.

## Forward-Looking Statements

*This Annual Report on Form 10-K contains forward-looking statements which are based on our management's beliefs and assumptions and on information currently available to our management. In some cases, you can identify forward-looking statements by terms such as "may," "will," "should," "could," "goal," "would," "expect," "plan," "anticipate," "believe," "estimate," "project," "predict," "potential," and similar expressions intended to identify forward-looking statements. These statements involve known and unknown risks, uncertainties and other factors, which may cause our actual results, performance, time frames or achievements to be materially different from any future results, performance, time frames or achievements expressed or implied by the forward-looking statements. We discuss many of these risks, uncertainties, and other factors in this Annual Report on Form 10-K in greater detail under the heading "Risk Factors." Given these risks, uncertainties, and other factors, you should not place undue reliance on these forward-looking statements. Also, these forward-looking statements represent our estimates and assumptions only as of the date of this filing. You should read this Annual Report on Form 10-K completely and with the understanding that our actual future results may be materially different from what we expect. We hereby qualify our forward-looking statements by these cautionary statements. Except as required by law, we assume no obligation to update these forward-looking statements publicly, or to update the reasons actual results could differ materially from those anticipated in these forward-looking statements, even if new information becomes available in the future.*

*All references to "NVIDIA," "we," "us," "our," or the "Company" mean NVIDIA Corporation and its subsidiaries.*

*In addition, statements that "we believe" and similar statements reflect our beliefs and opinions on the relevant subject. These statements are based upon information available to us as of the filing date of this Annual Report on Form 10-K, and while we believe such information forms a reasonable basis for such statements, such information may be limited or incomplete, and our statements should not be read to indicate that we have conducted an exhaustive inquiry into, or review of, all potentially available relevant information. These statements are inherently uncertain and investors are cautioned not to unduly rely upon these statements.*

© 2024 NVIDIA Corporation. All rights reserved.

## Part I

### Item 1. Business

#### Our Company

NVIDIA pioneered accelerated computing to help solve the most challenging computational problems. NVIDIA is now a full-stack computing infrastructure company with data-center-scale offerings that are reshaping industry.

Our full-stack includes the foundational CUDA programming model that runs on all NVIDIA GPUs, as well as hundreds of domain-specific software libraries, software development kits, or SDKs, and Application Programming Interfaces, or APIs. This deep and broad software stack accelerates the performance and eases the deployment of NVIDIA accelerated computing for computationally intensive workloads such as artificial intelligence, or AI, model training and inference, data analytics, scientific computing, and 3D graphics, with vertical-specific optimizations to address industries ranging from healthcare and telecom to automotive and manufacturing.

Our data-center-scale offerings are comprised of compute and networking solutions that can scale to tens of thousands of GPU-accelerated servers interconnected to function as a single giant computer; this type of data center architecture and scale is needed for the development and deployment of modern AI applications.

The GPU was initially used to simulate human imagination, enabling the virtual worlds of video games and films. Today, it also simulates human intelligence, enabling a deeper understanding of the physical world. Its parallel processing capabilities, supported by thousands of computing cores, are essential for deep learning algorithms. This form of AI, in which software writes itself by learning from large amounts of data, can serve as the brain of computers, robots and self-driving cars that can perceive and understand the world. GPU-powered AI solutions are being developed by thousands of enterprises to deliver services and products that would have been immensely difficult or even impossible with traditional coding. Examples include generative AI, which can create new content such as text, code, images, audio, video, and molecule structures, and recommendation systems, which can recommend highly relevant content such as products, services, media or ads using deep neural networks trained on vast datasets that capture the user preferences.

NVIDIA has a platform strategy, bringing together hardware, systems, software, algorithms, libraries, and services to create unique value for the markets we serve. While the computing requirements of these end markets are diverse, we address them with a unified underlying architecture leveraging our GPUs and networking and software stacks. The programmable nature of our architecture allows us to support several multi-billion-dollar end markets with the same underlying technology by using a variety of software stacks developed either internally or by third-party developers and partners. The large and growing number of developers and installed base across our platforms strengthens our ecosystem and increases the value of our platform to our customers.

Innovation is at our core. We have invested over \$45.3 billion in research and development since our inception, yielding inventions that are essential to modern computing. Our invention of the GPU in 1999 sparked the growth of the PC gaming market and redefined computer graphics. With our introduction of the CUDA programming model in 2006, we opened the parallel processing capabilities of our GPU to a broad range of compute-intensive applications, paving the way for the emergence of modern AI. In 2012, the AlexNet neural network, trained on NVIDIA GPUs, won the ImageNet computer image recognition competition, marking the “Big Bang” moment of AI. We introduced our first Tensor Core GPU in 2017, built from the ground-up for the new era of AI, and our first autonomous driving system-on-chips, or SoC, in 2018. Our acquisition of Mellanox in 2020 expanded our innovation canvas to include networking and led to the introduction of a new processor class – the data processing unit, or DPU. Over the past 5 years, we have built full software stacks that run on top of our GPUs and CUDA to bring AI to the world’s largest industries, including NVIDIA DRIVE stack for autonomous driving, Clara for healthcare, and Omniverse for industrial digitalization; and introduced the NVIDIA AI Enterprise software – essentially an operating system for enterprise AI applications. In 2023, we introduced our first data center CPU, Grace, built for giant-scale AI and high-performance computing. With a strong engineering culture, we drive fast, yet harmonized, product and technology innovations in all dimensions of computing including silicon, systems, networking, software and algorithms. More than half of our engineers work on software.

The world’s leading cloud service providers, or CSPs, and consumer internet companies use our data center-scale accelerated computing platforms to enable, accelerate or enrich the services they deliver to billions of end users, including AI solutions and assistants, search, recommendations, social networking, online shopping, live video, and translation.

Enterprises and startups across a broad range of industries use our accelerated computing platforms to build new generative AI-enabled products and services, or to dramatically accelerate and reduce the costs of their workloads and workflows. The enterprise software industry uses them for new AI assistants and chatbots; the transportation industry for autonomous driving; the healthcare industry for accelerated and computer-aided drug discovery; and the financial services industry for customer support and fraud detection.

Researchers and developers use our computing solutions to accelerate a wide range of important applications, from simulating molecular dynamics to climate forecasting. With support for more than 3,500 applications, NVIDIA computing enables some of the most promising areas of discovery, from climate prediction to materials science and from wind tunnel simulation to genomics. Including GPUs and networking, NVIDIA powers over 75% of the supercomputers on the global TOP500 list, including 24 of the top 30 systems on the Green500 list.

Gamers choose NVIDIA GPUs to enjoy immersive, increasingly cinematic virtual worlds. In addition to serving the growing number of gamers, the market for PC GPUs is expanding because of the burgeoning population of live streamers, broadcasters, artists, and creators. With the advent of generative AI, we expect a broader set of PC users to choose NVIDIA GPUs for running generative AI applications locally on their PC, which is critical for privacy, latency, and cost-sensitive AI applications.

Professional artists, architects and designers use NVIDIA partner products accelerated with our GPUs and software platform for a range of creative and design use cases, such as creating visual effects in movies or designing buildings and products. In addition, generative AI is expanding the market for our workstation-class GPUs, as more enterprise customers develop and deploy AI applications with their data on-premises.

Headquartered in Santa Clara, California, NVIDIA was incorporated in California in April 1993 and reincorporated in Delaware in April 1998.

## **Our Businesses**

We report our business results in two segments.

The Compute & Networking segment is comprised of our Data Center accelerated computing platforms and end-to-end networking platforms including Quantum for InfiniBand and Spectrum for Ethernet; our NVIDIA DRIVE automated-driving platform and automotive development agreements; Jetson robotics and other embedded platforms; NVIDIA AI Enterprise and other software; and DGX Cloud software and services.

The Graphics segment includes GeForce GPUs for gaming and PCs, the GeForce NOW game streaming service and related infrastructure; Quadro/NVIDIA RTX GPUs for enterprise workstation graphics; virtual GPU, or vGPU, software for cloud-based visual and virtual computing; automotive platforms for infotainment systems; and Omniverse Enterprise software for building and operating metaverse and 3D internet applications.

## **Our Markets**

We specialize in markets where our computing platforms can provide tremendous acceleration for applications. These platforms incorporate processors, interconnects, software, algorithms, systems, and services to deliver unique value. Our platforms address four large markets where our expertise is critical: Data Center, Gaming, Professional Visualization, and Automotive.

### **Data Center**

The NVIDIA Data Center platform is focused on accelerating the most compute-intensive workloads, such as AI, data analytics, graphics and scientific computing, delivering significantly better performance and power efficiency relative to conventional CPU-only approaches. It is deployed in cloud, hyperscale, on-premises and edge data centers. The platform consists of compute and networking offerings typically delivered to customers as systems, subsystems, or modules, along with software and services.

Our compute offerings include supercomputing platforms and servers, bringing together our energy efficient GPUs, DPUs, interconnects, and fully optimized AI and high-performance computing, or HPC, software stacks. In addition, they include NVIDIA AI Enterprise software; our DGX Cloud service; and a growing body of acceleration libraries, APIs, SDKs, and domain-specific application frameworks.

Our networking offerings include end-to-end platforms for InfiniBand and Ethernet, consisting of network adapters, cables, DPUs, and switch systems, as well as a full software stack. This has enabled us to architect data center-scale computing platforms that can interconnect thousands of compute nodes with high-performance networking. While historically the server was the unit of computing, as AI and HPC workloads have become extremely large spanning thousands of compute nodes, the data center has become the new unit of computing, with networking as an integral part.

Our end customers include the world's leading public cloud and consumer internet companies, thousands of enterprises and startups, and public sector entities. We work with industry leaders to help build or transform their applications and data center infrastructure. Our direct customers include original equipment manufacturers, or OEMs, original device manufacturers, or ODMs, system integrators and distributors which we partner with to help bring our products to market. We also have partnerships in automotive, healthcare, financial services, manufacturing, and retail among others, to accelerate the adoption of AI.

At the foundation of the NVIDIA accelerated computing platform are our GPUs, which excel at parallel workloads such as the training and inferencing of neural networks. They are available in the NVIDIA accelerated computing platform and in industry standard servers from every major cloud provider and server maker. Beyond GPUs, our data center platform expanded to include DPUs in fiscal year 2022 and CPUs in fiscal year 2024. We can optimize across the entire computing, networking and storage stack to deliver data center-scale computing solutions.

While our approach starts with powerful chips, what makes it a full-stack computing platform is our large body of software, including the CUDA parallel programming model, the CUDA-X collection of acceleration libraries, APIs, SDKs, and domain-specific application frameworks.

In addition to software delivered to customers as an integral part of our data center computing platform, we offer paid licenses to NVIDIA AI Enterprise, a comprehensive suite of enterprise-grade AI software and NVIDIA vGPU software for graphics-rich virtual desktops and workstations.

In fiscal year 2024, we launched the NVIDIA DGX Cloud, an AI-training-as-a-service platform which includes cloud-based infrastructure and software for AI, customizable pretrained AI models, and access to NVIDIA experts. We have partnered with leading cloud service providers to host this service in their data centers.

## **Gaming**

Gaming is the largest entertainment industry, with PC gaming as the predominant platform. Many factors propel its growth, including new high production value games and franchises, the continued rise of competitive gaming, or eSports, social connectivity and the increasing popularity of game streamers, modders, or gamers who remaster games, and creators.

Our gaming platforms leverage our GPUs and sophisticated software to enhance the gaming experience with smoother, higher quality graphics. We developed NVIDIA RTX to bring next generation graphics and AI to games. NVIDIA RTX features ray tracing technology for real-time, cinematic-quality rendering. Ray tracing, which has long been used for special effects in the movie industry, is a computationally intensive technique that simulates the physical behavior of light to achieve greater realism in computer-generated scenes. NVIDIA RTX also features deep learning super sampling, or NVIDIA DLSS, our AI technology that boosts frame rates while generating beautiful, sharp images for games. RTX GPUs will also accelerate a new generation of AI applications. With an installed base of over 100 million AI capable PCs, more than 500 RTX AI-enabled applications and games, and a robust suite of development tools, RTX is already the AI PC leader.

Our products for the gaming market include GeForce RTX and GeForce GTX GPUs for gaming desktop and laptop PCs, GeForce NOW cloud gaming for playing PC games on underpowered devices, as well as SoCs and development services for game consoles.

## **Professional Visualization**

We serve the Professional Visualization market by working closely with independent software vendors, or ISVs, to optimize their offerings for NVIDIA GPUs. Our GPU computing platform enhances productivity and introduces new capabilities for critical workflows in many fields, such as design and manufacturing and digital content creation. Design and manufacturing encompass computer-aided design, architectural design, consumer-products manufacturing, medical instrumentation, and aerospace. Digital content creation includes professional video editing and post-production, special effects for films, and broadcast-television graphics.

The NVIDIA RTX platform makes it possible to render film-quality, photorealistic objects and environments with physically accurate shadows, reflections and refractions using ray tracing in real-time. Many leading 3D design and content creation applications developed by our ecosystem partners now support RTX, allowing professionals to accelerate and transform their workflows with NVIDIA RTX GPUs and software.

We offer NVIDIA Omniverse as a development platform and operating system for building virtual world simulation applications, available as a software subscription for enterprise use and free for individual use. Industrial enterprises are adopting Omniverse's 3D and simulation technologies to digitalize their complex physical assets, processes, and environments – building digital twins of factories, real time 3D product configurators, testing and validating autonomous robots and vehicles, powered by NVIDIA accelerated computing infrastructure on-premises and in the cloud.

## **Automotive**

Automotive market is comprised of platform solutions for automated driving and in-vehicle cockpit computing. Leveraging our technology leadership in AI and building on our long-standing automotive relationships, we are delivering a complete end-to-end solution for the AV market under the DRIVE Hyperion brand. We have demonstrated multiple applications of AI within the car: AI can drive the car itself as a pilot in fully autonomous mode or it can also be a co-pilot, assisting the human driver while creating a safer driving experience.

We are working with several hundred partners in the automotive ecosystem including automakers, truck makers, tier-one suppliers, sensor manufacturers, automotive research institutions, HD mapping companies, and startups to develop and deploy AI systems for self-driving vehicles. Our unified AI computing architecture starts with training deep neural networks using our Data Center computing solutions, and then running a full perception, fusion, planning, and control stack within the vehicle on the NVIDIA DRIVE Hyperion platform. DRIVE Hyperion consists of the high-performance, energy efficient DRIVE AGX computing hardware, a reference sensor set that supports full self-driving capability as well as an open, modular DRIVE software platform for autonomous driving, mapping, and parking services, and intelligent in-vehicle experiences.

In addition, we offer a scalable data center-based simulation solution, NVIDIA DRIVE Sim, based on NVIDIA Omniverse software, for digital cockpit development, as well as for testing and validating a self-driving platform. Our unique end-to-end, software-defined approach is designed for continuous innovation and continuous development, enabling cars to receive over-the-air updates to add new features and capabilities throughout the life of a vehicle.

### Business Strategies

NVIDIA's key strategies that shape our overall business approach include:

**Advancing the NVIDIA accelerated computing platform.** Our accelerated computing platform can solve complex problems in significantly less time and with lower power consumption than alternative computational approaches. Indeed, it can help solve problems that were previously deemed unsolvable. We work to deliver continued performance leaps that outpace Moore's Law by leveraging innovation across the architecture, chip design, system, interconnect, and software layers. This full-stack innovation approach allows us to deliver order-of-magnitude performance advantages relative to legacy approaches in our target markets, which include Data Center, Gaming, Professional Visualization, and Automotive. While the computing requirements of these end markets are diverse, we address them with a unified underlying architecture leveraging our GPUs, CUDA and networking technologies as the fundamental building blocks. The programmable nature of our architecture allows us to make leveraged investments in research and development: we can support several multi-billion-dollar end markets with shared underlying technology by using a variety of software stacks developed either internally or by third-party developers and partners. We utilize this platform approach in each of our target markets.

**Extending our technology and platform leadership in AI.** We provide a complete, end-to-end accelerated computing platform for AI, addressing both training and inferencing. This includes full-stack data center-scale compute and networking solutions across processing units, interconnects, systems, and software. Our compute solutions include all three major processing units in AI servers – GPUs, CPUs, and DPUs. GPUs are uniquely suited to AI, and we will continue to add AI-specific features to our GPU architecture to further extend our leadership position. In addition, we offer DGX Cloud, an AI-training-as-a-service platform, and NeMo – a complete solution for building enterprise-ready Large Language Models, or LLMs, using open source and proprietary LLMs created by NVIDIA and third parties. Our AI technology leadership is reinforced by our large and expanding ecosystem in a virtuous cycle. Our computing platforms are available from virtually every major server maker and CSP, as well as on our own AI supercomputers. There are over 4.7 million developers worldwide using CUDA and our other software tools to help deploy our technology in our target markets. We evangelize AI through partnerships with hundreds of universities and thousands of startups through our Inception program. Additionally, our Deep Learning Institute provides instruction on the latest techniques on how to design, train, and deploy neural networks in applications using our accelerated computing platform.

**Extending our technology and platform leadership in computer graphics.** We believe that computer graphics infused with AI is fundamental to the continued expansion and evolution of computing. We apply our research and development resources to enhance the user experience for consumer entertainment and professional visualization applications and create new virtual world and simulation capabilities. Our technologies are instrumental in driving the gaming, design, and creative industries forward, as developers leverage our libraries and algorithms to deliver an optimized experience on our GeForce and NVIDIA RTX platforms. Our computer graphics platforms leverage AI end-to-end, from the developer tools and cloud services to the Tensor Cores included in all RTX-class GPUs. For example, NVIDIA Avatar Cloud Engine, or ACE, is a suite of technologies that help developers bring digital avatars to life with generative AI, running in the cloud or locally on the PC. GeForce Experience enhances each gamer's experience by optimizing their PC's settings, as well as enabling the recording and sharing of gameplay. Our Studio drivers enhance and accelerate a number of popular creative applications. Omniverse is real-time 3D design collaboration and virtual world simulation software that empowers artists, designers, and creators to connect and collaborate in leading design applications. We also enable interactive graphics applications - such as games, movie and photo editing and design software - to be accessed by almost any device, almost anywhere, through our cloud platforms such as vGPU for enterprise and GeForce NOW for gaming.

**Advancing the leading autonomous vehicle platform.** We believe the advent of autonomous vehicles, or AV, and electric vehicles, or EV, is revolutionizing the transportation industry. The algorithms required for autonomous driving - such as perception, localization, and planning - are too complex for legacy hand-coded approaches and will use multiple neural networks instead. In addition, EV makers are looking for next-generation centralized car computers that integrate a wide range of intelligent functions into a single AI compute platform. Therefore, we provide an AI-based hardware and software solution, designed and implemented from the ground up based on automotive safety standards, for the AV and EV market under the DRIVE brand, which we are bringing to market through our partnerships with automotive OEMs,



tier-1 suppliers, and start-ups. Our AV solution also includes the GPU-based hardware required to train the neural networks before their in-vehicle deployment, as well as to re-simulate their operation prior to any over-the-air software updates. We believe our comprehensive, top-to-bottom and end-to-end approach will enable the transportation industry to solve the complex problems arising from the shift to autonomous driving.

**Leveraging our intellectual property, or IP.** We believe our IP is a valuable asset that can be accessed by our customers and partners through license and development agreements when they desire to build such capabilities directly into their own products or have us do so through a custom development. Such license and development arrangements can further enhance the reach of our technology.

### Sales and Marketing

Our worldwide sales and marketing strategy is key to achieving our objective of providing markets with our high-performance and efficient computing platforms and software. Our sales and marketing teams, located across our global markets, work closely with end customers and various industry ecosystems through our partner network. Our partner network incorporates global, regional and specialized CSPs, OEMs, ODMs, system integrators, independent software vendors, or ISVs, add-in board manufacturers, or AIBs, distributors, automotive manufacturers and tier-1 automotive suppliers, and other ecosystem participants.

Members of our sales team have technical expertise and product and industry knowledge. We also employ a team of application engineers and solution architects to provide pre-sales assistance to our partner network in designing, testing, and qualifying system designs that incorporate our platforms. For example, our solution architects work with CSPs to provide pre-sales assistance to optimize their hardware and software infrastructure for generative AI and LLM training and deployment. They also work with foundation model and enterprise software developers to optimize the training and fine-tuning of their models and services, and with enterprise end-users, often in collaboration with their global system integrator of choice, to fine-tune models and build AI applications. We believe that the depth and quality of our design support are key to improving our partner network's time-to-market, maintaining a high level of customer satisfaction, and fostering relationships that encourage our end customers and partner network to use the next generation of our products within each platform.

To encourage the development of applications optimized for our platforms and software, we seek to establish and maintain strong relationships in the software development community. Engineering and marketing personnel engage with key software developers to promote and discuss our platforms, as well as to ascertain individual product requirements and solve technical problems. Our developer program supports the development of AI frameworks, SDKs, and APIs for software applications and game titles that are optimized for our platforms. Our Deep Learning Institute provides in-person and online training for developers in industries and organizations around the world to build AI and accelerated computing applications that leverage our platforms.

### Seasonality

Our computing platforms serve a diverse set of markets such as data centers, gaming, professional visualization, and automotive. Our desktop gaming products typically see stronger revenue in the second half of our fiscal year. Historical seasonality trends may not repeat.

### Manufacturing

We utilize a fabless and contracting manufacturing strategy, whereby we employ and partner with key suppliers for all phases of the manufacturing process, including wafer fabrication, assembly, testing, and packaging. We use the expertise of industry-leading suppliers that are certified by the International Organization for Standardization in such areas as fabrication, assembly, quality control and assurance, reliability, and testing. Additionally, we can avoid many of the significant costs and risks associated with owning and operating manufacturing operations. While we may directly procure certain raw materials used in the production of our products, such as memory, substrates, and a variety of components, our suppliers are responsible for procurement of most raw materials used in the production of our products. As a result, we can focus our resources on product design, quality assurance, marketing, and customer support. In periods of growth, we may place non-cancellable inventory orders for certain product components in advance of our historical lead times, pay premiums, or provide deposits to secure future supply and capacity and may need to continue to do so.

We have expanded our supplier relationships to build redundancy and resilience in our operations to provide long-term manufacturing capacity aligned with growing customer demand. Our supply chain is concentrated in the Asia-Pacific region. We utilize foundries, such as Taiwan Semiconductor Manufacturing Company Limited, or TSMC, and Samsung Electronics Co., Ltd., or Samsung, to produce our semiconductor wafers. We purchase memory from Micron Technology, Inc., SK Hynix Inc., and Samsung. We utilize CoWoS technology for semiconductor packaging. We engage with independent subcontractors and contract manufacturers such as Hon Hai Precision Industry Co., Ltd., Wistron Corporation, and Fabrinet to perform assembly, testing and packaging of our final products.



## Competition

The market for our products is intensely competitive and is characterized by rapid technological change and evolving industry standards. We believe that the principal competitive factors in this market are performance, breadth of product offerings, access to customers and partners and distribution channels, software support, conformity to industry standard APIs, manufacturing capabilities, processor pricing, and total system costs. We believe that our ability to remain competitive will depend on how well we are able to anticipate the features and functions that customers and partners will demand and whether we are able to deliver consistent volumes of our products at acceptable levels of quality and at competitive prices. We expect competition to increase from both existing competitors and new market entrants with products that may be lower priced than ours or may provide better performance or additional features not provided by our products. In addition, it is possible that new competitors or alliances among competitors could emerge and acquire significant market share.

A significant source of competition comes from companies that provide or intend to provide GPUs, CPUs, DPUs, embedded SoCs, and other accelerated, AI computing processor products, and providers of semiconductor-based high-performance interconnect products based on InfiniBand, Ethernet, Fibre Channel, and proprietary technologies. Some of our competitors may have greater marketing, financial, distribution and manufacturing resources than we do and may be more able to adapt to customers or technological changes. We expect an increasingly competitive environment in the future.

Our current competitors include:

- suppliers and licensors of hardware and software for discrete and integrated GPUs, custom chips and other accelerated computing solutions, including solutions offered for AI, such as Advanced Micro Devices, Inc., or AMD, Huawei Technologies Co. Ltd., or Huawei, and Intel Corporation, or Intel;
- large cloud services companies with internal teams designing hardware and software that incorporate accelerated or AI computing functionality as part of their internal solutions or platforms, such as Alibaba Group, Alphabet Inc., Amazon, Inc., or Amazon, Baidu, Inc., Huawei, and Microsoft Corporation, or Microsoft;
- suppliers of Arm-based CPUs and companies that incorporate hardware and software for CPUs as part of their internal solutions or platforms, such as Amazon, Huawei, and Microsoft;
- suppliers of hardware and software for SoC products that are used in servers or embedded into automobiles, autonomous machines, and gaming devices, such as Ambarella, Inc., AMD, Broadcom Inc., or Broadcom, Intel, Qualcomm Incorporated, Renesas Electronics Corporation, and Samsung, or companies with internal teams designing SoC products for their own products and services, such as Tesla, Inc.; and
- networking products consisting of switches, network adapters (including DPUs), and cable solutions (including optical modules) include such as AMD, Arista Networks, Broadcom, Cisco Systems, Inc., Hewlett Packard Enterprise Company, Huawei, Intel, Lumentum Holdings, and Marvell Technology Group as well as internal teams of system vendors and large cloud services companies.

## Patents and Proprietary Rights

We rely primarily on a combination of patents, trademarks, trade secrets, employee and third-party nondisclosure agreements, and licensing arrangements to protect our IP in the United States and internationally. Our currently issued patents have expiration dates from February 2024 to August 2043. We have numerous patents issued, allowed, and pending in the United States and in foreign jurisdictions. Our patents and pending patent applications primarily relate to our products and the technology used in connection with our products. We also rely on international treaties, organizations, and foreign laws to protect our IP. The laws of certain foreign countries in which our products are or may be manufactured or sold, including various countries in Asia, may not protect our products or IP rights to the same extent as the laws of the United States. This decreased protection makes the possibility of piracy of our technology and products more likely. We continuously assess whether and where to seek formal protection for innovations and technologies based on such factors as:

- the location in which our products are manufactured;
- our strategic technology or product directions in different countries;
- the degree to which IP laws exist and are meaningfully enforced in different jurisdictions; and
- the commercial significance of our operations and our competitors' operations in particular countries and regions.

We have licensed technology from third parties and expect to continue entering such license agreements.

## Government Regulations

Our worldwide business activities are subject to various laws, rules, and regulations of the United States as well as of foreign governments.

During the third quarter of fiscal year 2023, the U.S. government, or the USG, announced licensing requirements that, with certain exceptions, impact exports to China (including Hong Kong and Macau) and Russia of our A100 and H100 integrated circuits, DGX or any other systems or boards which incorporate A100 or H100 integrated circuits.

In July 2023, the USG informed us of an additional licensing requirement for a subset of A100 and H100 products destined to certain customers and other regions, including some countries in the Middle East.

In October 2023, the USG announced new and updated licensing requirements that became effective in our fourth quarter of fiscal year 2024 for exports to China and Country Groups D1, D4, and D5 (including but not limited to Saudi Arabia, the United Arab Emirates, and Vietnam, but excluding Israel) of our products exceeding certain performance thresholds, including A100, A800, H100, H800, L4, L40, L40S and RTX 4090. The licensing requirements also apply to the export of products exceeding certain performance thresholds to a party headquartered in, or with an ultimate parent headquartered in, Country Group D5, including China. On October 23, 2023, the USG informed us the licensing requirements were effective immediately for shipments of our A100, A800, H100, H800, and L40S products.

Our competitive position has been harmed, and our competitive position and future results may be further harmed in the long term, if there are further changes in the USG's export controls. Given the increasing strategic importance of AI and rising geopolitical tensions, the USG has changed and may again change the export control rules at any time and further subject a wider range of our products to export restrictions and licensing requirements, negatively impacting our business and financial results. In the event of such change, we may be unable to sell our inventory of such products and may be unable to develop replacement products not subject to the licensing requirements, effectively excluding us from all or part of the China market, as well as other impacted markets, including the Middle East.

While we work to enhance the resiliency and redundancy of our supply chain, which is currently concentrated in the Asia-Pacific region, new and existing export controls or changes to existing export controls could limit alternative manufacturing locations and negatively impact our business. Refer to "Item 1A. Risk Factors – Risks Related to Regulatory, Legal, Our Stock and Other Matters" for a discussion of this potential impact.

Compliance with laws, rules, and regulations has not otherwise had a material effect upon our capital expenditures, results of operations, or competitive position and we do not currently anticipate material capital expenditures for environmental control facilities. Compliance with existing or future governmental regulations, including, but not limited to, those pertaining to IP ownership and infringement, taxes, import and export requirements and tariffs, anti-corruption, business acquisitions, foreign exchange controls and cash repatriation restrictions, data privacy requirements, competition and antitrust, advertising, employment, product regulations, cybersecurity, environmental, health and safety requirements, the responsible use of AI, climate change, cryptocurrency, and consumer laws, could increase our costs, impact our competitive position, and otherwise may have a material adverse impact on our business, financial condition and results of operations in subsequent periods. Refer to "Item 1A. Risk Factors" for a discussion of these potential impacts.

## Sustainability and Governance

NVIDIA invents computing technologies that improve lives and address global challenges. Our goal is to integrate sound environmental, social, and corporate governance principles and practices into every aspect of the Company. The Nominating and Corporate Governance Committee of our Board of Directors is responsible for reviewing and discussing with management our practices related to sustainability and corporate governance. We assess our programs annually in consideration of stakeholder expectations, market trends, and business risks and opportunities. These issues are important for our continued business success and reflect the topics of highest concern to NVIDIA and our stakeholders.

The following section and the Human Capital Management Section below provide an overview of our principles and practices. More information can be found on our website and in our annual Sustainability Report. Information contained on our website or in our annual Sustainability Report is not incorporated by reference into this or any other report we file with the Securities and Exchange Commission, or the SEC. Refer to "Item 1A. Risk Factors" for a discussion of risks and uncertainties we face related to sustainability.

### Climate Change

In the area of environmental sustainability, we address our climate impacts across our product lifecycle and assess risks, including current and emerging regulations and market impacts.

In May 2023, we published metrics related to our environmental impact for fiscal year 2023. Fiscal year 2024 metrics are expected to be published in the first half of fiscal year 2025. There has been no material impact to our capital expenditures, results of operations or competitive position associated with global environmental sustainability regulations, compliance, or costs from sourcing renewable energy. By the end of fiscal year 2025, our goal is to purchase