Request for Information - NVIDIA Response

National Institute of Standards and Technology, Department of Commerce, Seeks Public Input on U.S. Chip Manufacturing

NVIDIA's technologies have a key role to play in an initiative to drive the U.S. semiconductor industry forward. The goal is to advance the U.S. semiconductor industry with benefits for the economy, the workforce and the environment. The CHIPS and Science Act holds gains for every industry, productivity and the fight against climate change because semiconductors are engines of the economy and human progress. In an effort to help landscape where the U.S. government should focus their efforts, we've outlined a dozen strategic opportunities where we'd like to help.

Background

Chips run software that modernizes and automates the world – chips drive human progress. The CHIPS and Science Act is a once-in-a-generation opportunity for the U.S. government to galvanize the semiconductor industry and ensure a strong foothold in its future. As one of the world's leaders in semiconductor design and artificial intelligence, NVIDIA looks forward to working with the current Administration to ensure that the CHIPS and Science Act gets a big bang for the taxpayer buck.

Based in Santa Clara, Calif., NVIDIA pioneered accelerated computing – the engine of computer graphics and artificial intelligence, and widely recognized as the path forward for sustainable computing. We are a full-stack computing company with platforms for scientific computing, AI, data science, autonomous vehicles, robotics, healthcare, and augmented and virtual reality. We have more than 25,000 employees in some 50 countries worldwide, with annual revenue of $26.9 billion in the latest fiscal year.

Innovation is at our core. Since our founding in 1993, we have invested over $29 billion in research and development, yielding more than 7,500 granted and pending patent applications globally, including inventions that are essential to modern computing. We are at the center of a vibrant ecosystem of partners, researchers, and customers, which includes 3.5 million developers and more than 12,000 global startups registered in our Inception program. The company is proud to offer long, rewarding careers to our people, and was voted the No. 1 Best Place to Work in 2022 in the U.S. by Glassdoor, based on employee surveys.

Over the past decade, we've been a driving force in artificial intelligence. We engineer the chips, systems, and software on which the world's AI infrastructure is built. And we've worked to democratize the power of AI by creating end-to-end platforms for the world's largest industries, from healthcare to transportation. Some 35,000 companies use NVIDIA AI technologies.

NVIDIA is also a foundational contributor to the metaverse – the 3D internet – through Omniverse, our real-time 3D collaboration platform that allows the creation of photorealistic virtual worlds in which physics and materials behave as they do in the physical world. This platform can be used to build and train robots and other AIs, as well as to create digital-twin simulations of massive structures like chip fabrication facilities and factories where manufacturing-floor design and mechanical processes can be optimized before getting rolled out in physical form.

While NVIDIA does not physically manufacture semiconductors, the company is a large consumer of them. We have thousands of talented engineers who design some of the

world's most advanced chips, which are manufactured by partners with whom we have long, deep relationships. Our founder and CEO, Jensen Huang, was proud to participate alongside President Biden and Secretary Raimondo at TSMC's recent ceremony to celebrate its $40 billion investment in a state-of-the-art fab in Arizona.

deep relationships. Our founder and CEO, Jensen Huang, was proud to participate alongside President Biden and Secretary Raimondo at TSMC's recent ceremony to celebrate its $40 billion investment in a state-of-the-art fab in Arizona.

NVIDIA Responses

Below are our responses to the Request for Information. To the extent a specific question is not addressed below, NVIDIA is generally supportive of the Semiconductor Industry Association's positions in its responses to this RFI. We have directly answered questions three, nine, seventeen, and twenty-one.

Question 3. Potential Technology Areas of Focus – US Semiconductor Manufacturing

Potential technology areas of focus that could be addressed by the Manufacturing USA semiconductor institutes to complement the National Advanced Packaging Manufacturing Program and the National Semiconductor Technology Research Center in Question 1 are listed below. What are your thoughts on the appropriateness of each for the scope of work for a Manufacturing USA semiconductor institute? What other topics should be included in the scope of an institute?

Chip-package architectures and codesign of integrated circuits and advanced packaging. (May include artificial intelligence, security, test methodologies, etc.).

1. **High-performance Computing**

   Future semiconductor manufacturing will require high-performance computing. The algorithms necessary to produce lithographic masks for emerging technology nodes are so computationally intensive that they will demand massively parallel supercomputers. The incredible volumes of data produced by semiconductor processor machines need supercomputers to process and extract information. Meeting this demand will require innovations up and down the computing stack, including circuits and architectures, parallel programming systems and compilers, and both simulation and machine learning algorithms.

2. **Package Manufacturing: Particle Defect Inspection**

   Minimizing particle defects is mandatory for the economic viability of advanced silicon process nodes. Particle-inspection equipment is a critical piece of semiconductor manufacturing. The output of the inspection equipment currently requires a manual review process to interpret actual defects from noise or false readings.

   Automating this manual review process would improve the viability of manufacturing in higher-cost areas. Accelerating innovation in this field will require more fundamental research in imaging AI/machine learning (ML) algorithms and collaboration across fabrication, equipment vendors, and university research. NVIDIA has significant domain expertise in AI/ML imaging techniques.

3. **Package Manufacturing: 3D Packaging**

With the end of Moore's law, 2.5D and 3D stacked multi-die systems will be the primary mechanism for continuing to scale chip performance in a cost- and power-efficient way. However, critical research is needed at the intersection of circuits and packaging in areas including short reach electrical signaling, photonic integration, thermal management, and power delivery.

With chip stacking technologies such as die-to-die hybrid bonding, we can push nano-bump pitch below 10um, enabling more than 50Tbps/mm^2 bandwidth density. However, area- and power-density constraints can easily limit scaling of die-to-die bandwidth before reaching full-scaling potential, and at projected die-to-die bandwidth, signaling systems will reach hundreds of watts and consume significant die area. This requires new signaling approaches and close co-optimization with packaging, chip-level clocking, power delivery, and cooling.

The industry is also reaching the scaling limit for off-package electrical signaling around the 200Gbps generation in terms of edge bandwidth density, as well as reach. Tightly integrated package and die-level technologies such as dense wave-division multiplexed silicon photonics promise a way forward with much higher bandwidth density, lower power, and elimination of the reach issue. However multiple critical research areas remain, including photonic device design and reliability, packaging and fiber attach, and low-cost, high-reliability multiwavelength laser sources. These technologies provide a near-term path to 2.5D integrated optical signaling and eventually to 3D integrated optical systems, and they must be co-optimized with system, power delivery, and thermal management requirements.

Assembly of advanced packages presents many automation challenges. Each product package has unique geometries and requires different equipment types, making it challenging to construct and operate an assembly line. Further automation requires more investment in robotics, an area where NVIDIA has substantial domain expertise.

4. **Computational Lithography**

Computational lithography is used to make masks for advanced silicon process nodes. It typically takes seven days on a large compute farm to make a single silicon mask, and a typical semiconductor product has approximately 50 masks. With further silicon geometry scaling, the complexity will become more compute intensive, with a greater number of masks.

NVIDIA has invested in computational lithography on the GPU and has demonstrated approximately 30x acceleration over CPUs. We believe the advent of this technology will enable a new era in modeling and algorithms to further scale silicon geometries.

Accelerating innovation in this field will require collaboration across fabrication, equipment vendors, and university research. NVIDIA has significant domain expertise in computational lithography, as well as a design infrastructure to validate algorithms.

5. **Process-Informed Lithography**

   With advanced process nodes, critical dimension measurements can be less predictive of the delivered performance of devices, as measured by downstream statistical process control monitoring and test data. Capturing such downstream process control measurements and feeding them back to drive a cost function are promising areas to explore to further optimize mask patterning beyond aerial exposure intensity.

6. **AI for Lithography**

   In addition to direct GPU acceleration of traditional computational lithography models, AI and ML techniques offer the opportunity for additional speedups. AI has the potential to dramatically accelerate the computation for making masks in advanced silicon process nodes, helping to further silicon geometry scaling. New AI techniques – such as image translation neural networks, physics-informed neural networks, generative networks, and self-training techniques – are promising future approaches for lithography simulation and mask optimization. Early research in this field shows the potential for AI driving speedups of 1-2 orders of magnitude for computational lithography running on GPUs at <1% accuracy loss compared with traditional lithography simulators. Further research will be needed to improve these AI techniques, develop new models, and find ways to combine AI and traditional methods in computational lithography software.

7. **Tools for Design Productivity**

   As Moore's law historically provided many orders of magnitude increases in transistor density per chip, the unique features now included in large chips are no longer predominantly limited by area constraints. Instead, new capabilities are increasingly limited by the engineering effort associated with digital design, verification, and VLSI implementation. Today's most complex system on chips (SoCs) contain billions of transistors and take thousands of engineer-years to implement. As applications demand more performance and energy efficiency in the form of specialization in the post-Moore's law era, we expect required complexity and design effort to increase. Research is needed in nearly all areas of a typical VLSI design flow to address this large, growing problem in chip-design costs.

   It is of critical importance in the front end of the design flow to reduce the effort in designing and especially verifying complex digital logic. Research in several areas could help reduce these costs dramatically. One promising approach being explored in the EDA research community focuses on designing hardware in higher-level languages with more advanced features (object-oriented design, type polymorphism, functional programming, etc.) and relying on trusted tools to map hardware designs into RTL and ultimately circuits for digital logic. New algorithms can also help speed up design verification throughput through GPU-accelerated RTL and logic simulation. Finally, AI and machine learning methods can help automate today's manual tasks, such as debugging and coverage closure. These approaches could benefit from additional research funding to help mitigate problems associated with chip design complexity and engineering design effort.

Once a design has been mapped to a digital circuit (or logic gates), it must be placed and routed on silicon. With ML transforming software in many domains, we expect this trend also to transform EDA software. However, research is still needed to understand how to best apply ML techniques to chip design tools. Neural network models developed for other tasks such as processing images or natural language are not directly suitable for VLSI design.

Instead, we must develop neural network models that process digital circuits as interconnected graphs laid out on a 2D canvas. Furthermore, although some recent research has applied ML to many EDA problems, it has been challenging to scale these ML-only approaches to solve real-world EDA problems. An emerging area of research proposes to integrate ML with conventional algorithms to solve EDA-specific problems more efficiently. GPU-accelerated computing combined with AI methods provides the most efficient approach for providing this capability, with early research demonstrating orders of magnitude speedups being possible.

We envision GPU-accelerated EDA algorithms for key steps in logic synthesis, VLSI placement, routing, and signoff closely integrated with AI-based methods for many common VLSI optimization tasks (transistor sizing, buffering, physical design space exploration, and others). AI also provides an opportunity to automate the layout of custom transistor designs, typically used in standard cells, SRAMs, and analog and mixed-signal blocks. Finally, advances in AI and GPU-accelerated computational lithography software could also enable EDA tools to become more manufacturing-aware during the design process. Since further scaling of EDA tools is critical to the success of the U.S. semiconductor industry, we feel it is important to invest in research in this area to improve the predictability, turnaround time, and quality of results with machine-learning-assisted EDA tools.

8. Virtual Fab: Tools for Manufacturing Productivity

A virtual fab is a digital model that accurately simulates complex semiconductor manufacturing processes such as etching, deposition, and chemical-mechanical polishing. Such a model is a digital tool to improve the yield, productivity, and utilization of expensive fabrication equipment that accounts for up to 75% of the cost of a modern fab, making such a fab fundamentally more economically competitive.

In addition to better equipment utilization, virtual process modeling can enable virtual testing of alternative process recipes without having to rely solely on expensive and time-consuming wafer-based physical testing. This can reduce time to market and improve yield in high-volume manufacturing, thereby accelerating time to revenue and return on invested capital costs.

Volumetric processing is critical for simulating manufacturing processes such as etching. Research and development on voxel-based “level-set” modeling (used in consumer gaming and media markets) is required for applying such modeling methods to high-fidelity manufacturing process simulations. Further, development of a rigorous simulation of the end-to-end fabrication workflow that exploits physics-ML based surrogates is critical for process control and optimization to accommodate the statistical variation of the fabrication process. Advances in such modeling techniques promise to yield better device power/performance/area – and ideally higher revenue per wafer produced in U.S. fabs.

Operational productivity can also benefit from modeling a virtual fab at the facility level. For example, overhead automated material handling systems (AMHS) route hundreds of payloads across complex intertwined networks. AI-assisted modeling and path optimization can maximize overall network throughput, reliability, and latency. Similarly, AI-guided path planning for autonomous mobile robot fleets trained in simulated factory environments can optimize material flow for maximum efficiency, typically in assembly.

Operational productivity must also account for worker safety and ergonomics. Modeling and analyzing worker movements throughout fabs or assembly can provide guidance for optimizing worker productivity and reducing the risk of potential safety hazards or adverse ergonomic impacts on the workforce. Worker efficiency and training can also be facilitated by training workers on operational and maintenance procedures in an immersive virtual environment, helping to boost domestic worker productivity in a shorter time.

At the facility level, overall site planning, design, and construction can also enjoy productivity and cost improvements using a virtual fab. High-fidelity 3D modeling and simulation can help planners design the building along with supporting critical infrastructure, including chemical/vapor/water delivery networks, electrical conduits, and HVAC control systems. These networks of pipes and conduits can be extremely extensive and complex, costing time and labor to optimize, build, and maintain. Such virtually designed sites can also better optimize energy efficiency, water consumption and maximize re-use, thereby reducing the environmental impact of fabs operating on domestic soil. These virtual site environments can also be used as an immersive AR/VR training platform to efficiently train and develop a world-class U.S.-based semiconductor manufacturing workforce.

The manufacturing equipment in a fab is extremely expensive and complex. Individual tools can be as much as $200M. Such equipment often accounts for up to 75% of the total cost of a fab. U.S. competitiveness can depend on how quickly and flawlessly such complex equipment can be installed and integrated into the production line, qualified, and ramped to high-volume production to start a return on the capital equipment cost. Virtual commissioning of this equipment requires an end-to-end platform to manage custom configuration requirements for each installation and integration of such wafer fabrication equipment. Such a platform can also be an automation tool for final qualification of the installed/integrated equipment on the line, as well as serving as a platform to enable virtual worker training for operations, service, and support throughout the lifetime of the equipment.

U.S. competitiveness depends on our ability to accelerate the adoption of new technologies broadly and at scale. Innovation to create new domestically originated technologies can be the key differentiators to accelerate the economic return on invested capital for a rejuvenated U.S.-based semiconductor manufacturing industry. NVIDIA has a proven track record of working with higher education and research institutions to support the adoption of new technologies such as accelerated computing, data science, AI/ML, and digital simulation. NVIDIA offers multiple tools and programs to support adoption at a scale through both our Deep Learning Institute and programs such as NVIDIA AI Technology Centers.

9. Architectures for Security

Computer security is increasingly critical for government entities, businesses, and private citizens. This growing need includes technical objectives, such as confidential computing, data integrity, privacy, and data attestation. As manufacturing becomes an increasingly digital and data-driven endeavor, protecting our manufacturing infrastructure from attacks, including both trojan horse and those that cause catastrophic failures, will be critical to national security.

Today's approaches for computer security are based on variations of encryption technologies coupled with isolation capabilities provided by traditional operating systems. Unfortunately, traditional approaches – such as secure enclaves, fully homomorphic encryption, and oblivious memory – can significantly compromise performance. While some applications may tolerate such slowdowns, many require high performance and will be reticent to employ such slower, expensive security technologies.

We anticipate opportunities for computer architecture research and development to enhance security across multiple dimensions: (1) hardware acceleration for confidential computing, data integrity and attestation, and privacy; (2) hardware support to defeat complex emerging side-channels including, but not limited to, row hammer and timing attacks; and (3) hardware acceleration for algorithms to process observed communication, computation, and memory access to facilitate anomaly detection for signs of intrusion attacks.

Coding and system software with novel computing paradigms and architectures, including chiplet compatibility with earlier generations

We consider coding and system software with novel computing paradigms and architectures to be an important area for the scope of work for a Manufacturing USA semiconductor institute. We would like to particularly highlight its significance in advancing 5G and successor wireless technologies.

The evolution of global wireless technology to fifth generation (5G) and beyond enables a new kind of network that not only enhances the broadband paradigm, but also aims to make “ubiquitous connectivity” a reality. To unleash the full potential of 5G, new trends are emerging toward making network deployments more dynamic, agile, and energy efficient. Radio access network (RAN) is arguably the most complex and compute-intensive component of the wireless network infrastructure. To meet the growing demand for network traffic, operators are looking into fundamentally new ways to optimize their RAN deployments.

At the heart of RAN transformation trends are two key aspects: scalability and compute. Traditional RAN deployments in the 4G era were predominantly brownfield, with tightly coupled network components (hardware and software) integrated through proprietary interfaces. This legacy “black box” approach, however, is largely inflexible in terms of reconfigurability and scalability, which are the essential traits for the next-generation 5G network. With the advent of advanced technologies like massive, multiple-input multiple-output (MIMO) antenna arrays, network dimensions are scaling in an unprecedented way. At the same time, 5G is aiming to support heterogeneous applications with vastly different requirements on the same network. As a result, the network dimensions and configurations must be flexible and manageable through network automation, as per the need for deployment scenarios and use cases.

The scalability challenge is compounded by a commensurate increase in the demand for 5G RAN compute, especially for physical layers, as network dimension and channel bandwidth go up. Existing RAN design approaches, including use of special-purpose hardware to partially accelerate physical layer processing, often turns out to be inadequate for meeting throughput and latency requirements for various 5G applications. Advanced accelerated computing is a key solution addressing the processing demand. Besides, new signal processing algorithms powered by AI and ML will become increasingly prominent in next generation 5G/6G networks.

NVIDIA tackles the scalability and compute challenges of 5G in a power-efficient, pure software-defined way by keeping all physical layer processing within the GPU's high-performance memory and exploiting the GPU's massive compute parallelism of the array of streaming multiprocessors. NVIDIA has demonstrated the feasibility of a fully software-based approach for 5G base stations by utilizing the compute capacity of recent generation GPUs like those based on the NVIDIA Ampere architecture. Beyond the silicon, the NVIDIA CUDA application programming interface provides developers with a high-level abstraction for programming GPUs using a C/C++ syntax, putting the massive compute capability of the GPU within easy reach of application programmers. GPU-accelerated 5G RAN makes next-generation networks both scalable and compute efficient.

We anticipate opportunities for coding and system software with novel computing paradigms and architectures to advance 5G and successor wireless technologies across multiple dimensions: (1) programing models that enable developers to efficiently utilize the compute capability of the underlying silicon technology with ease for 5G/6G workloads; (2) software development stacks that facilitate native support of AI and ML across all 5G/6G layers; and (3) cloud-native software that is designed to be open and automated for deployment to support 5G/6G on cloud infrastructure.

Environmental Sustainability for Semiconductor Manufacturing

1. **Energy Efficiency: Domain-Specific Processors**

   The demise of Moore's law and Dennard scaling has placed great pressure on the energy efficiency of computer systems. No longer do successive generations of chips benefit from reduced voltage that automatically lowers the transistors' power consumption. At the same time, the dramatic increase in demand for computation power, driven in part by the proliferation of AI algorithms, has increased the overall power consumption of computing platforms. The combination of these factors motivates future semiconductor systems to provide increased performance at constant or reduced power consumption.

   These factors require an investment in energy-efficient technologies at all levels of the stack, including: (1) advances in semiconductor device technologies that have less switching energy and less leakage power; (2) advances in signaling circuits that enable high-speed data transmission at less energy; (3) architectures that match the demands of the data-intensive applications running on them; and (4) programming systems that reduce software bloat while also enabling productive development of high-performance applications.

Specific opportunities in computer design include tailoring the architecture to specific applications, also known as domain-specific processors. In the domain of AI and neural networks, domain-specific architectures include tensor processing accelerators either embedded in general programmable systems (NVIDIA's Tensor Cores) or in dedicated systems (Google's TPU). There are promising opportunities in eliminating unnecessary data movement and computation through innovative methods of exploiting locality or exploiting characteristics of the data or computations, including data sparsity. Further potential rests in the co-design of architectures, packaging, and circuits to improve both performance and energy efficiency. We foresee the opportunity to apply domain-specific acceleration for both high performance and energy efficiency in such areas as autonomous robotic systems, imaging and diagnosis in health care, large-scale data analytics, and virtual/augmented reality.

Question 9. The authorizing statute for Manufacturing USA requires at least an equal non-federal co-investment in Manufacturing USA institutes to match the federal investment.

b. What is the anticipated impact of the new Investment Tax Credit (ITC) for industry established in the CHIPS Act on the level of investment in the new Manufacturing USA semiconductor institute(s), in facilities, including for manufacturing equipment and construction?

Commerce can best encourage co-investment for a Manufacturing USA semiconductor institute by ensuring that participating companies have flexibility in the type of support they provide. Companies seeking to support the institute are more likely to do so by being able to provide different types of support at their discretion, including in-kind contributions of people or equipment, in addition to cash contributions.

Adopting a broader definition of “manufacturing” for purposes of § 48D(b)(3) – i.e., what expenditures can qualify for the tax credit – to accurately capture the full range of activities integral to semiconductor manufacturing supply chain, including design, assembly, and test, would be most beneficial for the enhancement of the U.S. semiconductor industry and likely create greater co-investment and participation.

The advanced manufacturing investment credit included in the CHIPS and Science Act could assist in attracting contributions to the institute, depending on factors such as the institute’s structure, the type and use of the contributions to the institute, and how the Treasury Department will interpret certain terms in the credit. The credit is applicable to a “qualified investment” in an “advanced manufacturing facility.” IRC § 48D(b)(1). A qualified investment is defined as certain “qualified property” that includes tangible property either constructed by the taxpayer or “acquired by the taxpayer if the original use of such property commences with the taxpayer” and is “integral to the operation of the advanced manufacturing facility.” § 48D(b)(2). An advanced manufacturing facility means a facility “for which the primary purpose is the manufacturing of semiconductors or semiconductor manufacturing equipment.” § 48D(b)(3).

The interpretation of each of these terms will likely impact whether the credit will help drive contributions to the semiconductor institute. In addition, the institute itself must have the "manufacturing of semiconductors or semiconductor manufacturing equipment" as its "primary purpose." This could include research activities, but ultimately the primary purpose must be the manufacturing of semiconductors or equipment, including chip design. The semiconductor institute could best leverage the tax credit by having as its "primary purpose" the manufacture of semiconductors or equipment.

Question 17. How could a Manufacturing USA semiconductor institute best engage and leverage the diversity of educational and vocational training organizations (e.g., universities, community colleges, trade schools, etc.)?

The skills needed to support a manufacturing semiconductor institute will span many AI technology areas, including robotics, digital twins, and simulation capabilities. NVIDIA offers skills development and education for each of these technology areas at scale through our Deep Learning Institute and other programs. Many of these programs are currently in use in universities, community colleges, and trade schools.

Our Deep Learning Institute offers on-demand, self-paced fundamental level courses as well as advanced, instructor-led workshops in deep learning, accelerated computing, data science, graphics and simulation, plus additional industry- and workload-specific courses. Furthering intensive skills development, NVIDIA also supports NVIDIA AI Technology Centers, where teams of NVIDIA AI and accelerated computing experts collaborate with customers, industry partners, and researchers to accelerate applied research. These programs provide a strong foundation for getting an institute productive quickly. Many universities are already leveraging these programs and using them as the basis of building a diverse workforce. Two examples include initiatives at the University of Florida and Houston Community College.

In 2020, the University of Florida embarked on a project to become one of the nation's first AI universities. The curriculum's foundation rests on NVIDIA's Deep Learning Institute and a University of Florida NVIDIA AI Technology Center, which are available to every student across all disciplines, from undergraduate to PhD level, as well as for faculty training. The NVIDIA AI Technology Center has strategically enhanced researcher skills. Its "SynGatorTron," for example, the largest clinical language generator, is used to develop better AI for rare disease research and clinical trials. Additionally, University of Florida is taking multiple steps to uplevel AI and data science skills outside its walls, opening its curriculum and supercomputer access to schools across the state, including Florida A&M University, the nation's third largest HBCU.

Houston Community College, one of the nation's largest, most diverse two-year public colleges, is another example where NVIDIA programs drive critical technology adoption. With a goal of getting skilled workers into the Houston job market, HCC's two-year AI program prepares students for jobs as support specialists who help enterprise teams put the latest AI technologies to work. Teaching kits from NVIDIA's Deep Learning Institute are included in some courses. In the past few years, the program has included a virtual trip to NVIDIA's GTC conference, calling it "the greatest field trip of all time," which it uses to bring students up to speed on AI fast, to understand the trends in the field better, and to have a broader perspective on the growing number of industries AI is reshaping.

Question 21. How might a Manufacturing USA semiconductor institute integrate research and development activities and education to best prepare the current and future workforce?

> The US needs to rapidly train a world-class workforce in order to become competitive in advanced semiconductor manufacturing. Vast technical and practical experience is uniquely available from domain experts from other countries and training that scales across continents could facilitate effective technology transfer. But such a platform would need to have high-quality and reliable real-time language translation and transcription. Modern state-of-the-art language translation and transcription services heavily use AI-enabled models. New language models will need to be developed to accommodate the technical vocabulary and domain-specific concepts of the semiconductor industry.

We are grateful for the opportunity to provide input and look forward to working with the Department as the legislation is implemented.

If you have any questions or would like additional information, contact NVIDIA's Government Affairs team at **govaffairs@exchange.nvidia.com**.

NVIDIA