PUBLIC SUBMISSION

Received: May 29, 2025
Tracking No. mb9-y3mu-ao9t
Comments Due: May 28, 2025 **Submission Type:** Web

Docket: NSF-2025-OGC-0001
NITRD_FRDOC_0001

Comment On: NSF-2025-OGC-0001-0001
Request for Information: Development of a 2025 National Artificial Intelligence Research and Development Strategic Plan

Document: NSF-2025-OGC-0001-DRAFT-0284
Comment on FR Doc # 2025-07332

Submitter Information

Organization: Qualcomm

General Comment

See attached file(s)

Attachments

Qualcomm National AI Research and Development Strategic Plan Response

Qualcomm

5775 Morehouse Drive, San Diego, CA 92121
www.qualcomm.com

May 29, 2025

Faisal D'Souza
Networking and Information Technology Research and Development National Coordination Office,
National Science Foundation
2415 Eisenhower Avenue
Alexandria, VA 22314

Subject: Qualcomm Response to Request for Information on the Development of a 2025 National AI Research and Development Strategic Plan

Dear Mr. D'Souza,

Qualcomm Incorporated[1] (together with its subsidiaries, "Qualcomm") appreciates the opportunity to provide input[2] to the National Science Foundation (NSF) and the White House Office of Science and Technology Policy (OSTP) on the Request for Information regarding the 2025 National Artificial Intelligence (AI) Research and Development (R&D) Strategic Plan.[3] Qualcomm applauds the Trump Administration's commitment to ensuring U.S. leadership in AI technology and looks forward to partnering with the administration to support policies that advance U.S. AI innovation.

Qualcomm is a world leading wireless technology developer and chipset provider headquartered in San Diego, California. As a leading supplier of wireless semiconductor chipsets for four decades, Qualcomm chipsets are in billions of consumer devices (smartphones, tablets, laptops, and other wireless devices), as well as small cells, Wi-Fi access points, IoT devices and automobiles. Qualcomm's U.S. R&D efforts have resulted in the invention of many of the foundational technologies underpinning 3G, 4G, and 5G mobile technologies as well as high-performing and energy-efficient solutions supporting new innovations enabling both cloud and on-device AI

[1] Qualcomm Technologies, Inc., a subsidiary of Qualcomm Incorporated, operates the Qualcomm CDMA Technologies ("QCT") semiconductor business, which develops and supplies integrated circuits and system software based on 3G/4G/5G and other technologies for use in mobile devices, wireless networks, devices used in the Internet of Things, broadband gateway equipment, consumer electronic devices, and automotive systems for telematics, connectivity, and digital cockpit (also known as infotainment). Our portfolio also includes products for processors (including central processing unit (CPU), graphics processing unit (GPU) and neural processing unit (NPU)), modems, platforms, RF systems, and connectivity. Qualcomm Incorporated includes Qualcomm's licensing business, Qualcomm Technology Licensing ("QTL"), and the vast majority of its patent portfolio.

[2] This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in developing the 2025 National AI R&D Strategic Plan and associated documents without attribution.

[3] National Science Foundation, Request for Information on the 2025 National Artificial Intelligence Research and Development Strategic Plan, 90 Fed. Reg. 17835 (Apr. 29, 2025).

Qualcomm

capabilities, including advanced central processing units (CPUs), graphics processing units (GPUs), and neural processing units (NPUs). While Qualcomm does not currently train its own large language models (LLMs), we provide a unique perspective by collaborating with most of the world's leading AI model providers to enable optimized on-device AI solutions for consumer devices through our advanced chipsets. We have also recently announced custom data center CPUs for high performance and energy efficient AI solutions to accelerate innovation and AI inference at scale in data centers.[4]

Qualcomm's Commitment to AI R&D

Qualcomm's fundamental AI research started well before we formalized our AI Research division in 2018, with investments going back more than a decade. This formalization represented the culmination of years of distributed AI research efforts and marked our commitment to establishing AI as a core technology pillar. Our mission is to create breakthroughs in fundamental AI research and scale them across industries and use cases. Qualcomm is bringing generative AI to the consumer by advancing AI to make its core capabilities – perception, reasoning, and action – ubiquitous across devices. Qualcomm's latest high-performance, energy-efficient processors are powering a new wave of generative AI-enabled devices. For example, there are currently more than 2 billion AI-capable devices that are powered by Qualcomm chips, across smartphones, PCs, XR, IoT, and automobiles.

To develop impactful breakthroughs, our researchers aim to create the new "state of the art" in several key areas of AI research. Novel papers are one of the ways that we contribute impactful research to the larger AI community, and we've published over 150 AI research papers during the last few years on a variety of key AI topics, including computer vision, wireless and RF sensing, power efficiency, machine learning fundamentals, speech, audio and language processing, data compression and generative modeling, personalization and federated learning, optimization and reinforcement learning, and AI compilers and algorithms.[5]

Our AI research team has explored generative AI for several years. We initially focused on addressing the computational challenges of running AI workloads on battery-powered, resource constrained devices including power efficiency, personalization, system architecture and federated learning. This led to key breakthrough discoveries driving commercial products such as our Hexagon

[4] See, e.g., Qualcomm Press Note, Qualcomm and HUMAIN to Develop State-of-the-Art AI Data Centers to Deliver Cloud-to-Edge Hybrid AI Services (May 13, 2025), https://www.qualcomm.com/news/releases/2025/05/qualcomm-and-humain-to-develop-state-of-the-art-ai-data-centers-.

[5] A full list of research appears written or co-authored by Qualcomm AI Research is available at https://www.qualcomm.com/research/artificial-intelligence/papers.

Qualcomm

NPU architecture and industry-standards development tools.[6] We later explored whether generative models could compress well and generate improved perceptual-pleasing artifacts. We used VAE technology to create better video and speech codecs, keeping the model size small at less than 100 million parameters. We also extended generative AI innovations to make wireless communications more effective and brought these innovations to worldwide 5G/6G standards.

More recently, we increased the scale of on-device generative AI models to more than 1 billion parameters, such as Stable Diffusion, and unveiled a CPU that can run generative AI models with over 13 billion parameters on-device. We are now researching how to use generative AI models as universal agents to structure computations and use language to represent tasks and actions. We are studying how we can further exploit this capability by adding perceptual input, such as visual and audio, and the ability to interact with an environment, to generate a command for a robot or run software.

Finally, Qualcomm's commitment to open-source development and collaborative research accelerated the impact of our fundamental discoveries beyond our own product lines, contributing to industry-wide advances in edge AI deployment capabilities and reaching out to a broad community of developers and start-ups.

Enhancing U.S. AI Leadership Through R&D in On-Device AI Innovation

Massive generative AI models with billions of parameters place significant demands on computing infrastructure. As such, both AI training, which develops the parameters for an AI model, and AI inference, which executes the model, have been constrained to cloud implementations for large and complex models. As generative AI adoption grows at record-setting speeds and computing demands increase, a distributed processing architecture will become increasingly important for AI to scale and reach its full potential, while enhancing security, personalization, latency, and efficiency.

The scale of AI inference is poised to grow significantly. Although training individual models will continue and consumes significant resources, large generative AI models are expected to be trained only a few times a year. However, the cost of inferencing with those models increases correspondingly with the number of daily active users, their frequency of use, and the size of the model itself. Running inference on large models exclusively in the cloud results in very high costs that can be economically unsustainable for scaling.

[6] See, e.g., Jeff Gehlhaar, *Exploring the AI Capabilities of the Snapdragon 888 Mobile Platform* (Dec. 2020), https://www.qualcomm.com/content/dam/qcomm-martech/dm-assets/documents/snapdragon_888_ai_blog_by_jeff_gehlhaar_vp_of_technology_hsin-i_hsu_senior_product_manager.pdf.

Qualcomm

On-device AI can enhance U.S. AI competitiveness and innovation by enabling a more efficient allocation of AI infrastructure resources. Decentralizing inference tasks to edge devices – especially with models that are appropriately sized for such devices – reduces reliance on cloud-based systems, allowing U.S. data centers to focus on training advanced AI models that drive cutting-edge breakthroughs. This distributed approach can support faster innovation cycles, as real-time processing on devices accelerates application development and adoption across industries, from healthcare to manufacturing. Additionally, on-device AI can strengthen data privacy and security, fostering trust in AI technologies and encouraging broader use among businesses while also safeguarding the interests of American consumers.

Other countries are beginning to realize the importance of on-device AI leadership and announcing national initiatives to develop and promote a competitive on-device AI ecosystem. For example, South Korea's Ministry of Trade, Industry and Energy (MOTIE) recently launched the "K-on-device AI semiconductor technology development" program to promote on-device AI leadership in the automotive, IoT, home appliance, robot, and defense industries. The program will facilitate private-public partnerships to develop a full "on-device AI stack," including semiconductors, software, modules, and AI models.[7] To maintain its position as a global leader in AI-driven productivity and technological innovation, the NSF should ensure it is also supporting innovation in on-device AI capabilities.

To date, relatively little U.S. R&D has focused on the unique opportunities and challenges of building robust, distributed AI ecosystems. As AI becomes increasingly integrated into real-world systems, the NSF should support research in foundational technologies that enable intelligent, scalable, and collaborative deployments beyond centralized cloud environments. To accelerate U.S. AI innovation and competitiveness, Qualcomm recommends the following areas for federal R&D focus:

- **Research optimal distribution of AI workloads**: By leading foundational research on frameworks to determine the optimal distribution of AI workloads between cloud and on-device platforms, the NSF can drive innovation, maximize infrastructure investments, and maintain U.S. leadership in AI innovation. Efficient workload distribution will be critical for maximizing the performance, scalability, and cost-effectiveness of AI systems. Importantly, this research can help ensure that the substantial investments in AI infrastructure deliver the greatest possible return and impact.

[7] Press Release, Ministry of Trade, Industry and Energy, Full-Scale Promotion of K-on-device AI Semiconductor Development in Four Major Fields in the Physical AI Era, https://www.motie.go.kr/kor/article/ATCL3f49a5a8c/170538/view?mno=&pageIndex=2&rowPageC=0&displayAuthor=&searchCategory=0&schClear=on&startDtD=&endDtD=&searchCondition=1&searchKeyword=#

Qualcomm

- **Support novel research on human-computer interaction:** AI in user interfaces is evolving toward intuitive, natural-language interactions, positioning AI as a proactive partner. To support this shift, learning techniques must become more efficient—learning faster, reasoning better, and filling knowledge gaps autonomously. As AI systems become more persistent, their ability to observe long-term user behavior introduces both promise and complexity. These trends demand innovative R&D to redefine human-computer interaction in an edge-AI world.

- **Encourage platform innovation at the edge**: AI-centric operating systems are reshaping how devices manage data and computing, moving toward context-aware retrieval and dynamic resource sharing. Hardware designs that fix components and interconnects to AI functions would boost efficiency for AI tasks, reducing overhead and speeding responses. Realizing the potential of the Industrial Internet-of-Things (IIoT) requires improvements in collective organization, interoperability, and adaptive learning. NSF support for research in these areas can help drive the next wave of intelligent, secure, and stable AI ecosystems.

Conclusion

By catalyzing R&D in these areas, the NSF can lay the groundwork for a new generation of distributed AI platforms, unlocking societal benefits at scale and ensuring the United States remains at the forefront of AI innovation. Qualcomm stands ready to support the NSF and OSTP in research and development efforts to understand how on-device AI innovation can be leveraged strategically to advance U.S. AI leadership. Thank you for the opportunity to provide input. We are eager to collaborate and support the administration's efforts to advance U.S. leadership in AI.

Sincerely,

Jilei Hou
Senior Vice President, Engineering, Qualcomm AI Research