PUBLIC SUBMISSION

Received: May 28, 2025
Tracking No. mb8-t8ox-0zfz
Comments Due: May 28, 2025 **Submission Type:** API

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-0171
Comment on FR Doc # 2025-07332

Submitter Information

Organization: EdgeRunner AI

General Comment

See attached file(s)

Attachments

EdgeRunner AI OSTP RFI Response 05.28.25

May 28, 2025

Faisal D'Souza, NCO
Office of Science and Technology Policy
Executive Office of the President
2415 Eisenhower Avenue
Alexandria, VA 22314

Submitted by email to ostp-ai-rd-sp-rfi@nitrd.gov

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.

RE: Request for Information on the Development of a 2025 National Artificial Intelligence (AI) Research and Development (R&D) Strategic Plan

Introduction

EdgeRunner AI strongly supports OSTP's efforts to rewrite the National Artificial Intelligence Research and Development Strategic Plan (2023 Update) so that the U.S. can secure its position as the unrivaled world leader in artificial intelligence, enhance U.S. economic and national security, promote human flourishing, and maintain dominance in AI. In this submission, we focus on recommendations for leveraging AI specifically as it relates to the Department of Defense (DoD), the Department of Homeland Security, and other agencies relevant to U.S. national security.

Our recommendations encompass two categories: (1) prioritizing AI solutions for the warfighter that are capable of running air-gapped and offline; and (2) developing Large Language Models (LLMs) and AI technology for the DoD that are trained on fully transparent and auditable datasets.

About EdgeRunner AI

EdgeRunner AI is a US-based Series A dual-use defense technology company building domain-specific AI Agents for the warfighter. We develop state of the art (SOTA) Generative AI applications that enable the warfighter to accelerate workflows and make better decisions faster, increasing battle readiness and lethality.

We have been designated an "Awardable" vendor by the DoD's Chief Data and Artificial Intelligence Office (CDAO)[1], signed a Cooperative Research and Development Agreement (CRADA) with the Air Force Research Laboratory (AFRL), been named to CB Insights' AI 100[2], a list of the top emerging AI companies, and recently completed our Series A fundraise.[3]

Our first product is an AI assistant application that runs completely air-gapped, on-device (such as laptops, tablets, or ATAKs) and comes equipped with Military Occupational Speciality (MOS)-specific adapters. These adapters are trained on MOS-specific data, such as military doctrine, training materials, and equipment manuals, providing more accurate and personalized responses to specific questions and prompts. By running locally on-device, our product increases data privacy and security, eliminates hosting costs, and enables performance in Denied, Disrupted, Intermittent, and Low-Bandwidth (DDIL) environments.

The application can be deployed on a variety of devices and hardware, bringing the power of LLMs and domain-specific AI to the warfighter at the tactical edge. In terms of business model, we charge a monthly or annual license fee to access the application and provide users with regular software updates on a monthly or quarterly basis. We charge an extra professional services fee to develop the custom adapters for each customer. To deploy the application and adapters, users simply download and install the application as a single executable binary in a few minutes. The application can also be delivered as a container.

What we mean by "Air-gapped, on-device"

There are several definitions of "air-gapped." Often, it means having no direct connection to the internet or any other computer that is connected to the internet. However, when we discuss "air-gapped" AI we mean our application can operate completely disconnected from the internet, other networks, and other devices. Our

[1] https://www.edgerunnerai.com/news/edgerunner-ai-designated-awardable-vendor-for-department-of-defense-chief-digital-and-artificial-intelligence-offices-tradewinds-solutions-marketplace
[2] https://www.cbinsights.com/research/report/artificial-intelligence-top-startups-2025/
[3] https://www.edgerunnerai.com/news/edgerunner-raises-17-5m-to-develop-air-gapped-on-device-ai-for-the-warfighter

application runs completely on-device, such as a laptop or workstation, using the device's own computing hardware (e.g. CPU, NPU, GPU).

Enhancing U.S. economic and national security

Why on-device matters

It is critical to U.S. national security that certain AI solutions are air-gapped, especially those used by the warfighter, because they must be able to operate in ever increasing DDIL environments. A soldier operating in Ukraine, Taiwan, or the Indo-Pacific cannot rely on having constant network connectivity or access to the cloud. Electronic warfare, including the use of radio frequency (FR) jamming, is an increasingly common tactic on the battlefield. This means that cloud-hosted AI solutions, such as those built by OpenAI or Anthropic, which rely upon constant network connectivity to operate, are not suitable for many DoD use cases. These solutions could be easily disrupted by enemy forces or even perform unreliably when networks are constrained due to low bandwidth, excessive usage, or disrupted connectivity.

On-device AI also provides increased data privacy and data security. By running locally, the user's prompts and outputs do not leave the user's device. Unlike cloud or network-hosted solutions, which require sending data over the internet externally for processing, on-device AI limits the ability for adversaries to intercept user data. This also reduces the chances of detection by limiting the signals emitted by users, which enhances force protection and operational security.

Additionally, on-device AI has cost and efficiency advantages. The generative AI boom has resulted in a massive spike in cloud computing costs, as well as energy costs, as businesses race to build data centers and compute infrastructure. According to McKinsey, $7 trillion in capital outlays are needed by 2030 to support demand for compute power related to AI and traditional IT applications.[4]

These costs will eventually be passed onto users in the form of higher usage costs or constrained capacity. Most AI applications charge based on usage or consumption on a per token basis, which means users pay more as usage increases and adoption scales across the organization. This can result in unpredictable and unforeseen costs. The benefit of on-device AI is that there are no incremental hosting costs associated with increased usage. Because the application is running fully on-device, the user has already acquired all the necessary hardware to run the application. This also means the

[4] https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-cost-of-compute-a-7-trillion-dollar-race-to-scale-data-centers

energy consumption and infrastructure build (e.g. construction costs) is significantly decreased when leveraging on-device AI solutions.

Finally, on-device AI generally results in easier and faster compliance. Because no data is leaving your device or being processed in the cloud, FedRAMP and other compliance requirements either do not apply or are easier to comply with, which means an organization can adopt and scale on-device AI solutions faster and more safely.

Why transparent, auditable training data matters

In addition to on-device AI, we believe the transparency and auditability of AI solutions is critical to U.S. national security. LLMs and the applications built on top of them are in large part functions of the training data and methods used to train the base model. LLMs can be made to think certain ways and give certain answers based on how they are trained. As seen in Google's release of Gemini in early 2024[5] and more recently in models released by DeepSeek in early 2025[6] models can have both political and policy biases.

While the Google and Gemini examples were particularly egregious and easy to spot, it's clear how more subtle biases could be inserted into LLMs. These biases could become difficult to detect and users may not even realize they are getting answers that are either incorrect or skewed in certain ways. As more mission-critical systems become integrated with LLMs and users become more comfortable making critical decisions based on recommendations or responses from AI-based solutions, it's imperative that we guard against potential bias and contamination.

Even "open source" models, such as Meta's Llama models, cannot be considered fully transparent. While Meta, and other open source model providers, open parts of their model code and weights, they do not make their training data publicly available. At EdgeRunner AI, we are committed to developing AI solutions that are maximally transparent. For DoD customers, we make our model code, weights, and training data available for auditing and inspection. For our MOS-specific adapters, we make our training data completely open and transparent. We also provide explanations about our training methods and techniques.

[5] https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical
[6] https://www.csis.org/analysis/hawkish-ai-uncovering-deepseeks-foreign-policy-biases

Conclusion

Extremely powerful AI technology will be developed during this Administration that will transform how we fight war. President Trump, Secretary Hegseth, and this Administration have an opportunity to dramatically enhance U.S. national security by bringing these capabilities to the DoD and getting them into the hands of warfighters at the front lines.

However, it will be critical that AI solutions for specific use cases are capable of running completely on-device for increased data privacy, security, performance, and cost effectiveness. Furthermore, it will be critical that the DoD leverages AI solutions that are fully transparent with auditable datasets to migrate potential biases.

It is essential that this Administration prioritize AI capabilities that can operate securely and safely in the Age of AI. Our adversaries are actively developing new technology to disrupt, deny, and degrade our capabilities on the battlefield. For the warfighter operating at the front lines, we must ensure their AI capabilities are both resilient and reliable. We recommend the following actions:

- The DoD should prioritize researching, developing, and acquiring AI capabilities that can run fully on-device without requiring connectivity to the cloud or network
- The DoD should encourage the service branches to leverage AI capabilities that can operate air-gapped, reducing dependency on the cloud and lowering hosting costs
- The CDAO should study and require that AI model providers selling to the DoD make their training data and techniques public for auditability and safety

We look forward to engaging with this Administration to ensure that AI enhances U.S. national security while safely increasing lethality of the warfighter.