CASE STUDY

U.S. Navy Charters Secure AI Course with Confidential Computing

AMERICA'S NAVY

Headquarters

Washington, DC

Use Cases

Confidential large language models (LLMs) on ships

Challenges

Comparing the performance of confidential vs. conventional GPUs, while ensuring data privacy and security

Solution

Llama3 LLM deployed on NVIDIA H100 GPUs using Anjuna Seaglass

Results

- 1-hour deployment
- Maximum GPU performance
- Verified trustworthiness of AI models
- Complete privacy in all data states

Background

The U.S. Navy, one of the world's most advanced maritime forces, plays a crucial role in maintaining global security and ensuring freedom of the seas. With a history dating back to 1775, the Navy has evolved into a force equipped with cutting-edge technology. Central to its success is the Navy's commitment to innovation, particularly in areas like cybersecurity, artificial intelligence, and secure communications, ensuring that it remains a dominant force in an increasingly complex and interconnected world.

Challenge

The U.S. Navy aimed to evaluate the efficacy, speed, and performance of Confidential GPUs compared to conventional GPUs for running large language models (LLMs). A key goal was to determine whether the high-performance capabilities of NVIDIA's H100 GPUs could enable the Navy to execute LLM workloads directly on ships, where local access to data is crucial for maximizing responsiveness. Ensuring data security and privacy throughout the process was a primary concern, given the sensitive nature of military operations.

Solution

In a groundbreaking initiative, the U.S. Navy collaborated with Anjuna and NVIDIA to assist their exploration of advanced Confidential AI capabilities.

The Navy successfully deployed Llama3 LLM models on confidential NVIDIA H100 Tensor Core GPUs within the NVIDIA LaunchPad environment. To simplify and accelerate the deployment, the Navy selected the Anjuna Seaglass platform. The Seaglass platform, known for its Universal Confidential Computing capabilities, provided a virtualized, secure environment that was easy to deploy and operate.

The deployment demonstrated the combined power of NVIDIA's cutting-edge GPUs and Anjuna's flexible Confidential Computing platform. The U.S. Navy was able to test and prove the high level of performance achievable on LLM workloads using NVIDIA GPUs, while ensuring that data remained protected in all three states: in-use, in-motion, and at-rest.

Results

- **1-hour deployment:** The Anjuna Seaglass platform enabled a swift, 1-hour deployment, significantly accelerating the Navy's ability to harness Confidential Computing capabilities.
- **Security without compromise:** The deployment proved that the high processing speed of NVIDIA's Confidential GPUs could be fully utilized, delivering maximum performance without sacrificing data privacy.
- **Trust fully verified:** Through attestation reports, the integrity and trustworthiness of the LLMs were validated, ensuring that the AI models were operating securely and as intended.
- **Complete data privacy:** The deployment confirmed that data was fully protected against malicious entities at all stages – in-use, in-motion, and at-rest –, reinforcing the Navy's confidence in the security of its AI operations.

Conclusion

The successful deployment demonstrated the U.S. Navy's ability to securely and efficiently deploy AI workloads using NVIDIA's Confidential H100 GPUs and the Anjuna Seaglass platform. The project validated the potential for the Navy to run critical AI models, with the assurance of complete data privacy and high performance. This collaboration marks a significant step forward in the Navy's pursuit of secure, AI-driven operations, setting the stage for future advancements in military technology.

Get Started

REV-0924