



## AT A GLANCE

### WHAT IS IT?

NVIDIA DGX GB300 is a rack-scale “AI Supercomputer” system built on the Grace Blackwell Ultra platform, engineered for large-scale AI training and high-throughput inference (including reasoning and long-context workloads).

### OBJECTIVES

- Deliver a campus on-prem “AI factory” for secure, scalable model development, fine-tuning, and inference experimentation aligned to Fleet needs.
- Accelerate applied research and prototyping by providing a shared high-end compute backbone for interdisciplinary teams and mission-focused sprints.
- Enable next-generation simulation and digital twins (wargaming, autonomy, multi-domain modeling) with high-throughput AI pipelines.



### POINTS OF CONTACT:

Shared mailbox: [AI@nps.edu](mailto:AI@nps.edu)  
Trenton Hancock, NPS CIO  
Jeffrey Haferman, ITACS  
Michael Hernandez, ITACS

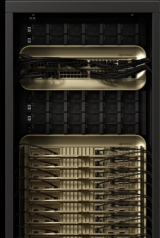
# NPS AI FACTORY


## NVIDIA DGX GB300 AI SUPERCOMPUTER

The GB300 provides Navy, Marine Corps, and other services access to a level of AI compute power that was previously available only in large commercial or national-scale facilities or via the cloud. This capability allows military teams to train and operate large, domain-specific models that can support maritime operations, expeditionary warfare, and special operations missions, including AI-enabled planning, ISR fusion, autonomous behaviors, and operational decision support.

The system’s extreme compute density and memory capacity enable workloads that would otherwise be infeasible, such as training long-context reasoning models, running multi-agent simulations at scale, and integrating diverse operational data streams into a single analytic environment. This power is critical for exploring complex, highly-contested scenarios where speed, scale, and fidelity directly affect the operational relevance of the output or application.

By placing this level of capability within a military research university environment, the GB300 allows NPS and its partners to stress, evaluate, and mature advanced AI concepts at operational scale. This ensures that AI systems intended for Naval, expeditionary, and special operations missions are developed with sound science, operational realism, engineering rigor, and a clear path toward transition and adoption.

<b>System:</b>	Rack-scale AI factory (Grace Blackwell Ultra)	
<b>Compute:</b>	72× Blackwell Ultra GPUs + 36× Grace CPUs (2,592 Arm cores)	
<b>Performance:</b>	Up to 1.44 EFLOPS FP4 (sparse) 1.08 EFLOPS (dense)	
<b>Memory:</b>	~20 TB HBM; ~37 TB total fast system memory	
<b>Interconnect:</b>	NVLink fabric + 800 Gb/s InfiniBand (ConnectX-8)	
<b>Storage (NPS):</b>	~0.5 PB high-performance shared AI storage	



## USE CASES

- Large-scale foundation model training on operational and defense-specific datasets
- Fine-tuning and adaptation of LLMs for Naval planning, intelligence, and operations
- Multimodal model training (text, imagery, FMV, signals, cyber) for Joint warfighting tasks
- Reinforcement learning training for autonomous surface, subsurface, and aerial platforms
- Multi-agent AI training for fleet tactics, swarming, and adversarial competition
- Synthetic data-driven model training to overcome classified, sparse, or denied datasets
- Long-context model training for MDA and global tracking problems
- Reasoning and planning model training for AI-enabled CSISR-T
- AI-assisted wargaming training using self-play and red & blue force learning loops
- Continual learning and model retraining pipelines using operational feedback data
- AI safety, robustness, and adversarial resilience training for contested environments
- Rapid experimentation and assessment of emerging AI capabilities prior to Fleet transition
- High-resolution modeling of operational environments
- Physics-based digital twins for ships, ports, facilities, and energy systems