

The Military Needs Frontier Models

Capt. Zachary Szewczyk, U.S. Army

Not all artificial intelligence (AI) is created equal. While basic large language models can process and generate text, “frontier models” like OpenAI’s GPT-4.5, Anthropic’s Claude 4 Sonnet, Google’s Gemini 2.5 Pro, Meta’s LLaMA 4, and X’s Grok 4 are far more powerful. At the bleeding edge of the field, this more capable class of model draws from a deeper knowledge base, benefits from greater contextual understanding, and possesses enhanced reasoning abilities compared to their older, smaller, and less advanced counterparts. As the Army explores AI, it is crucial to choose capable models to handle the amorphous, ever-changing nature of modern warfare. In military applications where AI will have a large role

to play in consequential decisions, the level of sophistication present in frontier models is not a luxury but rather a necessity.

Why Frontier Models? The Limits of Small AI

Frontier models are powerful systems trained using enormous amounts of data. What counts as “frontier” changes over time, though—and it changes fast.

A key measure of a model’s capability is its number of “parameters”—think of these as internal settings the model adjusts as it learns from data. More parameters usually mean the model can learn more subtle patterns. For instance, GPT-3.5, with 175 billion parameters,



Examples of “frontier models” (from left): Grok, Claude, ChatGPT, and Gemini. (Composite graphic by Beth A. Warrington, *Military Review*)

was considered advanced in late 2022.¹ Just a few months later though, OpenAI released GPT-4 and the threshold for “frontier” changed. Released in early 2023, GPT-4 had over one trillion parameters and required over fourteen times the computing resources to build. GPT-4 also exhibited many unusual properties, what Microsoft famously called “emergent behaviors.”² The recently released Grok 4 reportedly has over 1.7 trillion parameters, an astonishing feat.³ Unlike their smaller, task-specific predecessors, the massive frontier models of today exhibit far greater abilities to reason, handle complexity, and understand context.⁴

Reasoning Ability and Complexity Handling

Larger models are better at thinking, especially when a task requires several logical steps, like drafting a complete document or performing a technical analysis. Their capacity to synthesize vast datasets allows them to draw logical inferences, connect disparate ideas, and maintain coherence across extended arguments. This is vital in areas like cybersecurity and intelligence, where good decisions depend on looking at many factors, finding hidden patterns, and understanding enemy actions. Their larger size seems to also help reduce common issues like making up information (often called “hallucinations”) or showing unfair tendencies (“biases”).⁵ Smaller models, on the other hand, often struggle in these areas. They find it hard to manage connected pieces of information or sort out conflicting details in complex problems.⁶ This can lead to disjointed thinking, overly simple answers, and an inability to deal with complicated tasks like analyzing varied threats or understanding complex operational situations. When accuracy and thoroughness are essential, a model’s size directly affects how well it can analyze information and support operations.

Context Retention and Generalization

Another key difference is that larger models can remember and use information from long documents or discussions. This is especially important for military uses where accuracy and consistency are critical. Whether processing intelligence reports, drafting strategic guidance, or synthesizing volumes of logs during cyber operations, frontier models maintain a level of coherence

that mitigates inconsistencies and reduces the cognitive burden on human analysts. Adapting to new situations is also a challenge for small models; they haven’t been trained using enough diverse data to handle things they haven’t seen before. Frontier models, trained on vast datasets and able to consider large amounts of information at once, are much better at handling general tasks and applying old knowledge to new problems without needing retraining. Smaller architectures require extensive fine-tuning to achieve similar levels of performance, but then only succeed in narrow areas, making them brittle in dynamic operational contexts where flexibility and rapid adaptation are essential.

Indeed, small models have shown impressive improvements over their predecessors and even approach the capabilities of models like GPT-4 according to some narrow measures.⁷ But GPT-4 has not been a state-of-the-art frontier model for quite some time. This comparison does not carry the weight many assume. Meanwhile, modern frontier models have grown so capable that they make their predecessors look like toys.

The Importance of Frontier Models for Commanders and Staffs

Senior leaders deal with complex, unclear situations and too much information. Frontier models can be a powerful aid for decision-making, quickly and accurately pulling together huge amounts of data and explaining key points.⁸ When staff work needs to follow specific policies, legal rules, and operational goals, these adaptable models could offer clear evaluations, making sure recommendations are logical and consistent. Unlike smaller models that find it hard to balance many factors in changing situations, larger systems can combine past examples, military doctrine, and current information to suggest

Capt. Zachary Szewczyk, U.S. Army, commissioned into the Cyber Corps in 2018 after graduating from Youngstown State University with an undergraduate degree in computer science and information systems. He has supported or led defensive cyberspace operations from the tactical to the strategic level, including several high-level incident responses. He has served in the Cyber Protection Brigade, the 3rd Multi-Domain Task Force, and currently serves in the Cyber Protection Brigade.

practical courses of action. This leads to a faster decision-making process. Senior leaders would get well-organized analyses that explain risks, benefits, and how choices could affect the mission—a single AI system could potentially do tasks that currently require many large staff teams. In a time when speed is as important as accuracy, using advanced AI to help make decisions isn't just helpful, it's essential.

The Importance of Frontier Models in Defensive Cyber Operations

Perform preliminary analysis and accelerate investigations. In today's cybersecurity work, the sheer amount and speed of data make it hard to tell real threats apart from normal network activity quickly. Frontier models can play a key role by automating the first look at this data: sorting through security logs, connecting related warnings from different systems, and filtering out unimportant information before human analysts step in.⁹ This would free analysts from routine work, allowing them to focus on clear signs of a cyberattack. Beyond this initial sorting, these models can also speed up investigations by helping to create advanced ways to examine data, spot unusual activity, and provide background information that normally takes hours of manual work. Smaller models struggle with thinking through multiple steps and comparing information from different, especially large, datasets. Frontier models, however, are powerful enough to handle these very difficult tasks. As attackers operate faster and more often, this AI capability is vital for both efficiency and successfully defending our networks.

Enhancing training. Good cyber training needs more than fixed lesson plans and old examples; it requires realistic, flexible teaching that adapts to new threats. Even small language models have the capacity to supplement this training with dynamic datasets and responsive scenarios, but only frontier models have the capability to build the training itself.

My previous unit, the 3rd Multi-Domain Task Force, stood up without access to the training for its cyber formation to do its job. We estimated that the traditional approach, building an entire curriculum from zero, would take more than a year. Developing effective training is difficult and slow. Instead, we turned to large language models and found they could create everything we needed—from lesson plans and materials to

exercises and assessments—in just a few hours.¹⁰ This required the immensely capable GPT-4; even GPT-3.5, though new at the time, could not handle the complexity of this task.

Smaller models lack the deep understanding needed to create useful cyber training that goes beyond basic exercises. Using frontier models, organizations can create a cyber training system that constantly updates as enemy methods change. This ensures our cyber forces are ready for real-world challenges, not just textbook examples.

But Who Will Build Them?

The criticality of frontier models for military applications leads to important questions about how to resource and enable that capability. I've always felt that many soldiers, if given the chance, could do much more than routine daily tasks. This is the underlying assumption of the Army Software Factory, which takes service members and allows them to develop software that the government might otherwise have paid a contractor two, three, or ten times as much for.¹¹ But that is not the same as building, tuning, or deploying—or all three—state-of-the-art AI models at the bleeding edge of technical sophistication—the kind the military truly needs. I am hopeful that the Army could one day cultivate the talent necessary to push this field forward within its ranks, and a recent rumor of a new functional area focused on AI is an encouraging sign. But—for the same reasons the military struggles to recruit and retain talented cyber personnel—I am not sure the Army's culture is conducive to recruiting and retaining the deep expertise necessary for this to succeed.¹² We may have, unfortunately, contrived a situation where outsourcing is once again the only option.

Another challenge is the rapid pace of change. A year ago, I would have been thrilled to have access to modern models on an accredited platform. Today, I have that capability in the form of CamoGPT and others.¹³ However, the goalposts have moved. Over the last several months, game-changing innovations like dedicated reasoning engines for enhancing complex problem-solving, tools like NotebookLM for integrated research and writing, advanced semantic search capabilities like Deep Research, and human-quality text-to-speech models have drastically improved the reliability and utility of large language models.¹⁴ The military spent a year catching up only to

discover that the industry had rocketed even further ahead by the time it had. The Army's new generative AI platform, Army Enterprise Large Language Model Workspace, powered by Ask Sage, at least comes close but lacks many of those key features and more, and also fumbles the execution with a token-based subscription scheme that requires individual units to pay for access—a barrier few are likely to overcome.¹⁵ Maybe by 2026, government systems will have AI capabilities that can be accessed commercially today in 2025, but maybe not. Even if they do, a one-year-or-more delay between civilian and military technology is a significant gap.

Reliance on external innovation, however, brings significant legal and data security challenges to the forefront. The uncomfortable truth is that most commercial technologies, including the powerful frontier models the military needs, are not designed with the stringent requirements for handling government-owned, unclassified and classified data in mind. The prospect of commercial entities collecting, aggregating, and ultimately repurposing sensitive military data for their own training sets, model refinement, or commercial profit is not just a hypothetical concern but a tangible risk to data sovereignty and a serious concern for operational security.

This predicament complicates the build-versus-buy calculus. While internal development of true frontier models presents a steep climb, simply plugging into commercial offerings without stringent data controls courts disaster. The careful construction of data governance frameworks, the establishment of separate, secured enclaves for model operation and fine-tuning for government-use cases, and the explicit definition and rigorous enforcement of intellectual property ownership and data usage rights—ensuring the government retains control over its data and any AI capabilities developed with it—are paramount. These are not mere bureaucratic hurdles but fundamental safeguards that must be engineered into the military's AI adoption strategy from the outset, lest the very tools meant to enhance our capabilities become vectors for compromise.

Conclusion

The term “military grade” is often a joke in the armed forces. Civilians think it means “high quality,”

but service members know it often means the cheapest product that met some vague standards. We risk the same thing happening with AI. To accelerate the adoption of AI across the Department of Defense, the Pentagon has established the AI Rapid Capabilities Cell, but the military's penchant for general requirements combined with poor evaluation methods for large language models could very well leave the military saddled with subpar chatbots because they cost less, not the far more capable force multipliers we desperately need.¹⁶ CamoGPT, the now-defunct NIPRGPT, and others are good but not great—the murky difference between the small models available through those platforms and today's frontier models is critically important, not something to be discarded because good is “good enough.”¹⁷ The difference between a small, open weight model running in CamoGPT and frontier models running in purpose-built data centers is not, in fact, minimal. AI should augment decision-making, streamline workflows, and enhance cyber defense. In these contexts, where AI will have a large role to play in consequential military decisions, the level of sophistication present in frontier models is not a luxury but rather a necessity.

By investing in and integrating frontier models, the military can realize the potential of AI not as a replacement for human expertise but as an indispensable tool that augments decision-making, streamlines workflows, and enhances cyber defense.¹⁸ The alternative—sticking on older, limited models due to procurement inertia or a failure to prioritize—ensures stagnation while adversaries rapidly adopt—or even develop themselves—superior AI technology.¹⁹ To avoid ceding the technological advantage, the military must not only invest in and integrate current frontier models but also cultivate the institutional agility to continuously adapt to the evolving AI landscape. We cannot afford to be left behind.²⁰ In an era of accelerating change, accepting “good enough” AI is a risk the Nation cannot afford. In the myriad areas where we ought to accept risk, I hope appropriately capable AI is at least one where we do not.²¹ ■

The views expressed in this work are those of the author and do not reflect the official policy or position of the Department of the Army or the Department of Defense.

Notes

1. Winnie Nwanne, "Comparing GPT-3.5 & GPT-4: A Thought Framework on When to Use Each Model," *AI—Azure AI Services Blog*, Microsoft Tech Community, 18 March 2024, <https://techcommunity.microsoft.com/blog/azure-ai-services-blog/comparing-gpt-3-5-gpt-4-a-thought-framework-on-when-to-use-each-model/4088645>.
2. Nwanne, "Comparing GPT-3.5 & GPT-4"; "Data on Notable AI Models," Epoch AI, updated 9 July 2025, <https://epoch.ai/data/notable-ai-models>; Sébastien Bubeck et al., "Sparks of Artificial General Intelligence: Early Experiments with GPT-4," arXiv, updated 13 April 2023, <https://doi.org/10.48550/arXiv.2303.12712>.
3. Michael Spencer, "How Good Is xAI's Grok 4?," *AI Supremacy*, 11 July 2025, <https://www.ai-supremacy.com/p/how-good-is-xai-grok-4-elon-musk>.
4. Ethan Mollick, "A New Generation of AIs: Claude 3.7 and Grok 3," *One Useful Thing*, 24 February 2025, <https://www.oneusefulthing.org/p/a-new-generation-of-ais-claude-3-7>.
5. Sandeep, "Grok-3: The Next Evolution in AI by xAI," *OpenCV*, 19 February 2025, <https://opencv.org/blog/grok-3/>.
6. Wenjie Wu et al., "Graph-Augmented Reasoning: Evolving Step-by-Step Knowledge Graph Retrieval for LLM Reasoning," arXiv, 3 March 2025, <https://doi.org/10.48550/arXiv.2503.01642>.
7. Kourosh Hakhamaneshi and Rehaan Ahmad, "Fine-Tuning Llama-2: A Comprehensive Case Study for Tailoring Models to Unique Applications," *Anyscale*, 11 August 2023, <https://www.anyscale.com/blog/fine-tuning-llama-2-a-comprehensive-case-study-for-tailoring-models-to-unique-applications>.
8. Jairo Eduardo Márquez-Díaz, "Benefits and Challenges of Military Artificial Intelligence in the Field of Defense," *Computación y Sistemas [Computing and systems]* 28, no. 2 (April-June 2024): 309–26, <https://doi.org/10.13053/cys-28-2-4684>.
9. Zachary Szewczyk, "Advantage Defense: Artificial Intelligence at the Tactical Cyber Edge," *Modern War Institute*, 17 May 2024, <https://mwi.westpoint.edu/advantage-defense-artificial-intelligence-at-the-tactical-cyber-edge/>.
10. Zachary Szewczyk, "Artificial Intelligence-Enabled Cyber Education," *Gray Space*, 1 July 2025, <https://www.lineofdeparture.army.mil/Journals/Gray-Space/Archive/Summer-2025/AI-Cyber-Ed/>.
11. Abdul Subhani, "Is the Army Ready to Think Like a Tech Company? Why the Service Needs to Value Coding the Same Way as Shooting," *Modern War Institute*, 17 August 2023, <https://mwi.westpoint.edu/is-the-army-ready-to-think-like-a-tech-company-why-the-service-needs-to-value-coding-the-same-way-as-shooting/>.
12. Zachary Szewczyk, "A Cyber Force Is Not the Only Solution," *War on the Rocks*, 25 July 2024, <https://warontherocks.com/2024/07/a-cyber-force-is-not-the-only-solution/>.
13. Mikayla Easley, "DISA Launching Experimental Cloud-Based Chatbot for Indo-Pacific Command," *DefenseScoop*, 25 March 2025, <https://defensescoop.com/2025/03/25/disa-siprgpt-chatbot-indopacom-joint-operational-edge-cloud/>.
14. Schreiner, Vaccariello, Moschetto, hosts, *The Phoenix Cast*, podcast, "Deep Research: Prompting for Garcia," 9 March 2025, <https://overcast.fm/+AAaFCdVFjF8>; "OpenAI.fm," OpenAI, accessed 25 April 2025, <https://www.openai.fm/>.
15. U.S. Army Public Affairs, "Army Launches Army Enterprise LLM Workspace, the Revolutionary AI Platform That Wrote This Article," *U.S. Army*, 15 May 2025, <https://www.army.mil/article/285537/army-launches-army-enterprise-llm-workspace-the-revolutionary-ai-platform-that-wrote-this-article>.
16. Shaun Waterman, "For Pentagon's AI Programs, It's Time for Boots on the Ground," *Signal Media*, 1 March 2025, <https://www.afcea.org/signal-media/pentagons-ai-programs-its-time-boots-ground>; Timothy R. McIntosh et al., "Inadequacies of Large Language Model Benchmarks in the Era of Generative Artificial Intelligence," arXiv, updated 14 October 2024, <https://doi.org/10.48550/arXiv.2402.09880>.
17. Easley, "DISA Launching Experimental Cloud-Based Chatbot for Indo-Pacific Command."
18. Richard Farnell and Kira Coffey, "AI's New Frontier in War Planning: How AI Agents Can Revolutionize Military Decision-Making," *Belfer Center for Science and International Affairs*, 11 October 2024, <https://www.belfercenter.org/research-analysis/ais-new-frontier-war-planning-how-ai-agents-can-revolutionize-military-decision>; Jim Gallagher and Edward J. Oughton, "Transforming the Multidomain Battlefield with AI: Object Detection, Predictive Analysis, and Autonomous Systems," *Military Review Online Exclusive*, 18 September 2024, <https://www.armyupress.army.mil/Journals/Military-Review/Online-Exclusive/2024-OLE/Multidomain-Battlefield-AI/>; Szewczyk, "Advantage Defense."
19. DeepSeek-AI et al., "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning," arXiv, 22 January 2025, <https://doi.org/10.48550/arXiv.2501.12948>.
20. Michael R. Weimer, "Combat Doesn't Care: How Ready Are You?," *Muddy Boots*, NCO Journal, 14 October 2024, <https://www.armyupress.army.mil/Journals/NCO-Journal/Muddy-Boots/Combat-Doesnt-Care-Weimer/>.
21. Scott Dawe, "Cut It All, Then Build Back: Shrinking the Mountainous Burden of Administrative Requirements That Hinder Army Units' Effectiveness," *Modern War Institute*, 4 March 2025, <https://mwi.westpoint.edu/cut-it-all-then-build-back-shrinking-the-mountainous-burden-of-administrative-requirements-that-hinder-army-units-effectiveness/>.