

Frontier AI for National Security: The AI Forge Program

DARPA Program Manager: Matthew Marge, Ph.D.
Information Processing Techniques Office

SIGDIAL 2026

27th Annual Meeting of the Special Interest Group on Discourse and Dialogue



5 August 2026



Outline

- DARPA/IPTO Overview
- Current trends in AI
- AI Forge Program Overview



Pioneering innovation through preventing technological surprise: Defense Advanced Research Projects Agency (DARPA)



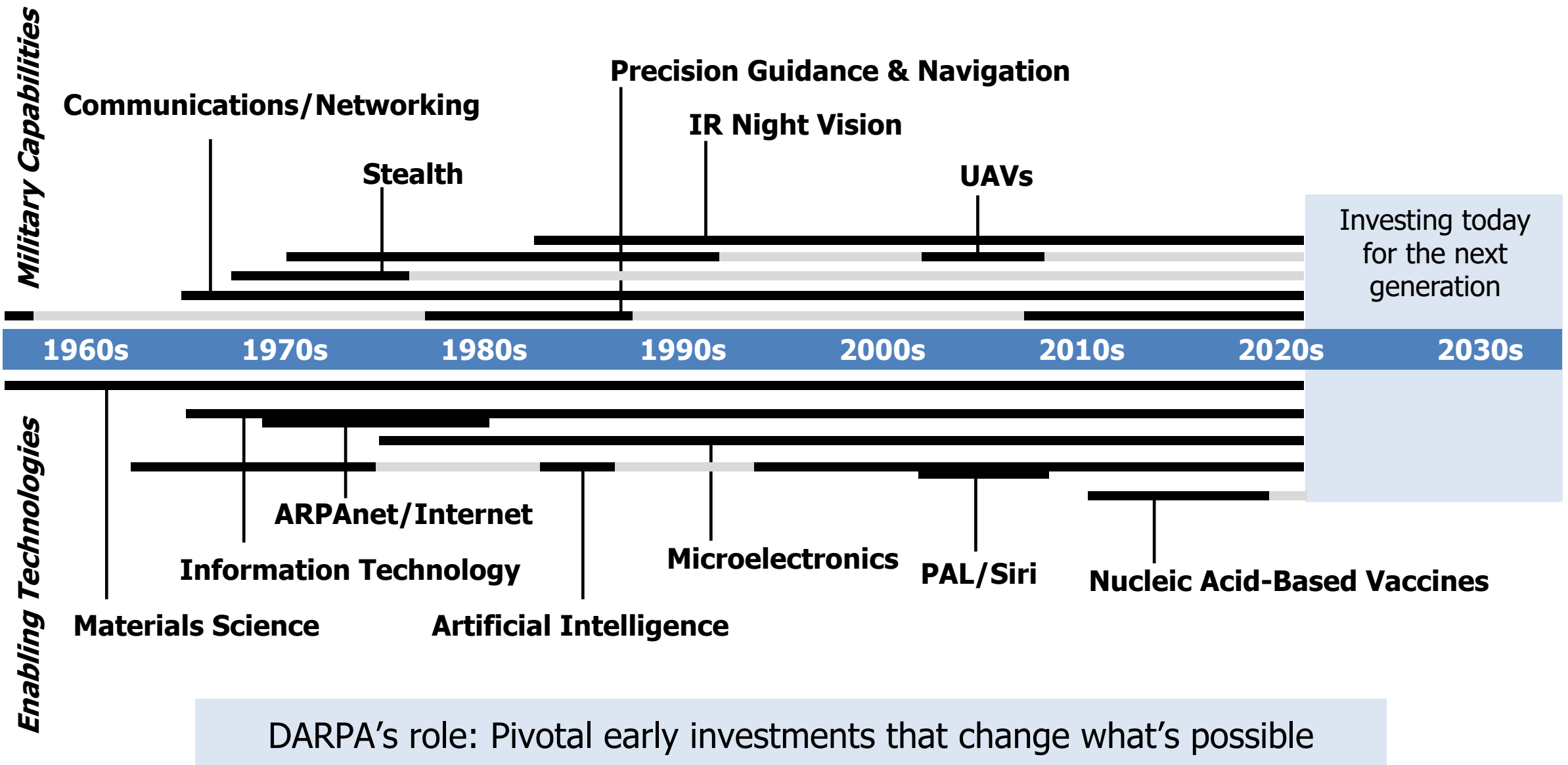
Breakthrough technologies for national security

- **Create breakthrough, paradigm-shifting solutions**
- **Accept and manage significant technology risk**
- **Disrupt or massively accelerate technology roadmaps**





DARPA's mission





DARPA Technical Offices



Biological Technologies Office

- Maintain force readiness
- Tactical warfighter care and functional restoration
- Operational resilience and logistical security
- Biosensors and novel methods and materials



Defense Sciences Office

- Novel materials and structures
- Sensing and measurement
- Computation and processing
- Enabling operations
- Collective intelligence
- Emerging threats



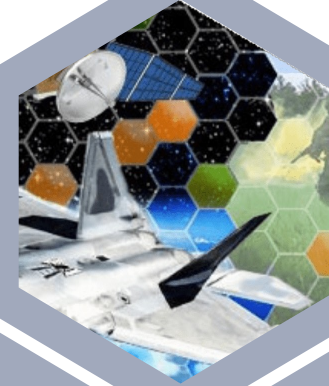
Information Processing Techniques Office

- Transformative AI
- Advantage in cyber operations
- Confidence in the information domain
- Resilient, adaptable, and secure systems



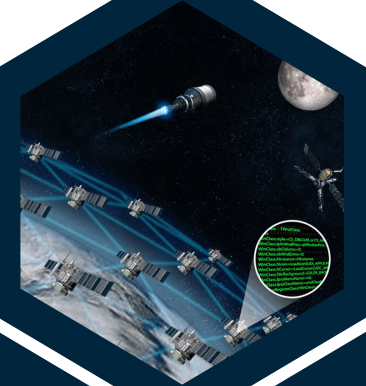
Multi-X Office

- Disruptive microsystems
- Edge processing
- Microsystems manufacture



Strategic Technology Office

- Advanced sensors and processing
- Battlefield effects
- Command, control, and communications
- System of autonomous systems
- Empowered human decision making



Tactical Technology Office

- Tactical systems
- Platforms, systems, and technologies that enable new warfighting constructs
- Reimagination of missions across maritime, ground, air, and space domains



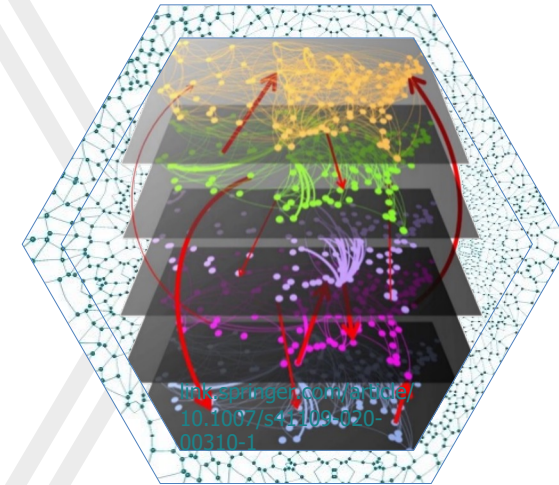
Information Processing Techniques Office (IPTO) thrust areas



**AI innovation
and execution**



**Inherent security
and privacy**



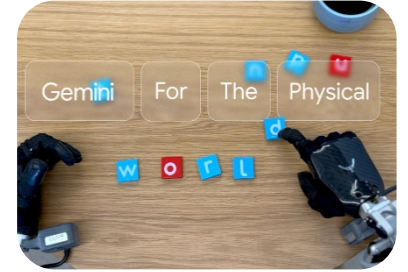
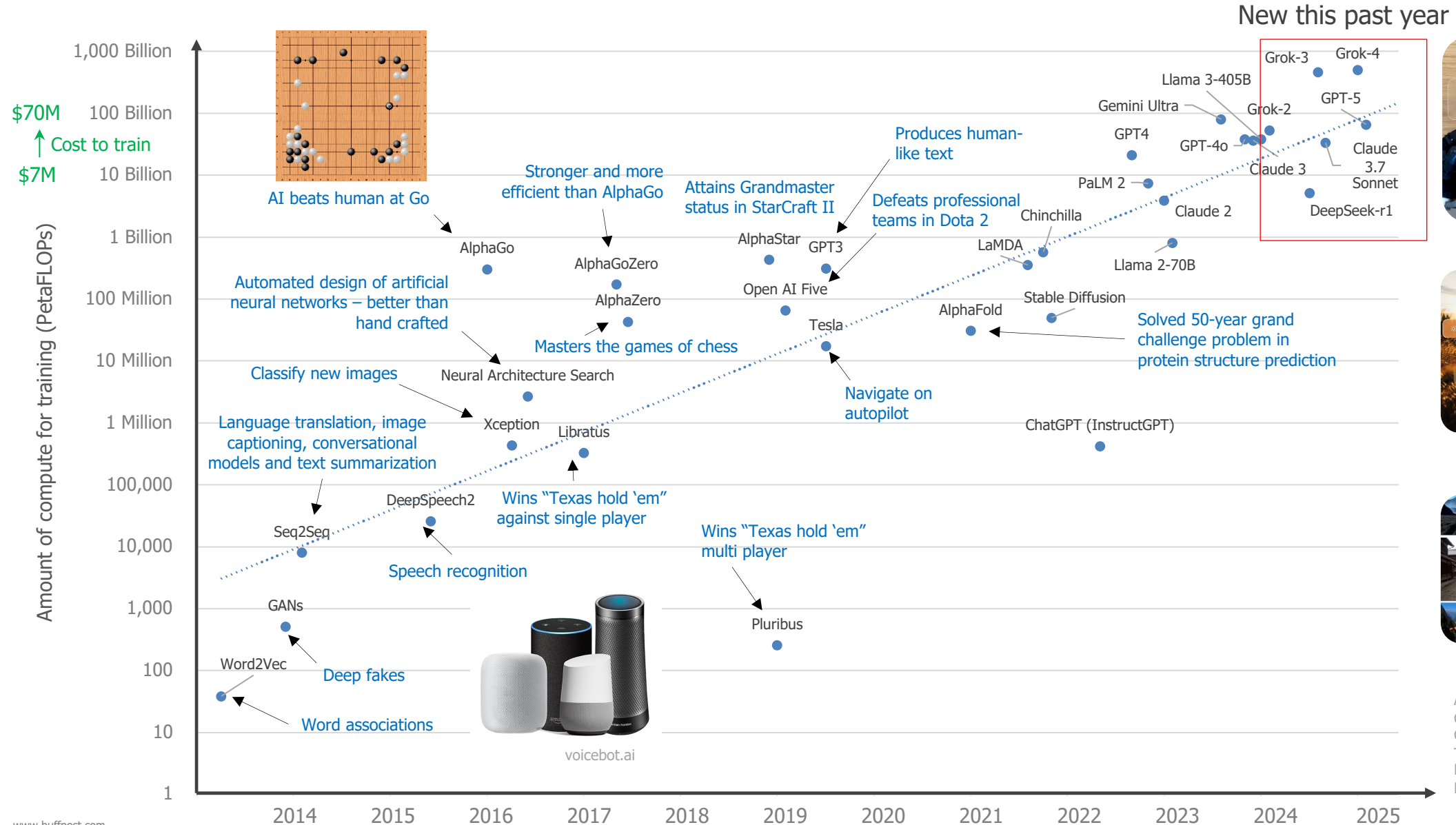
**Resilient
megasystems**



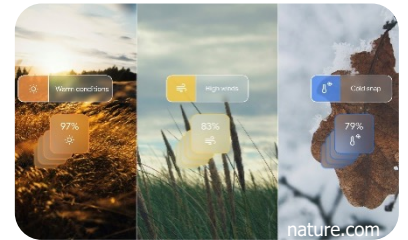
Frontier AI for National Security: AI Forge



Amazing progress in ML fueled by compute power (and data)



Model designed for robotics



Weather forecasting



Genie3: 3d world model

AI: Artificial Intelligence
GAN: Generative Adversarial Network
GPT: Generative Pre-Trained Transformer
ML: Machine Learning
PaLM: Pathway Language Model

https://deepmind.google/blog/genie-3-a-new-frontier-for-world-models/#:~:text=World%20models%20are%20also%20a,chevron_left



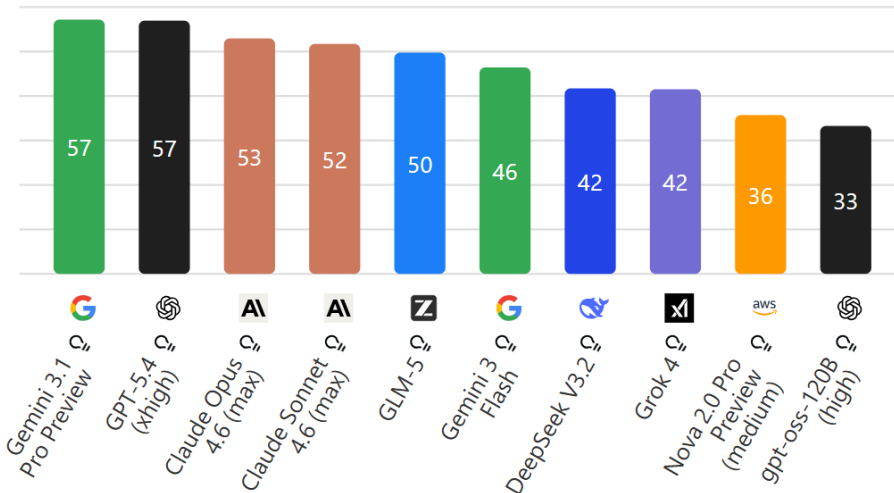
Newly released models in 2025 and 2026: highlights

Model releases are accelerating, not "once every couple of years" or even "once a year"

- DeepSeek R1 released Jan 2025: a powerful reasoning model, just a few months behind US GenAI
- Goedel agent May 2025: recursive self-improvement (also see 3-Lisp 1982, other reflective languages)
- Google's Gemini 3 Nov 2025: took the lead in most major metrics
- Claude Opus 4.5 Nov 2025: took the lead in code generation
- Claude 4.6 Feb 2026: recursive self-improvement
- OpenAI-5.3-codex Feb 2026: recursive self-improvement
- Qwen3.5 (Alibaba) Feb 2026: coming close to US GenAI offerings

INTELLIGENCE

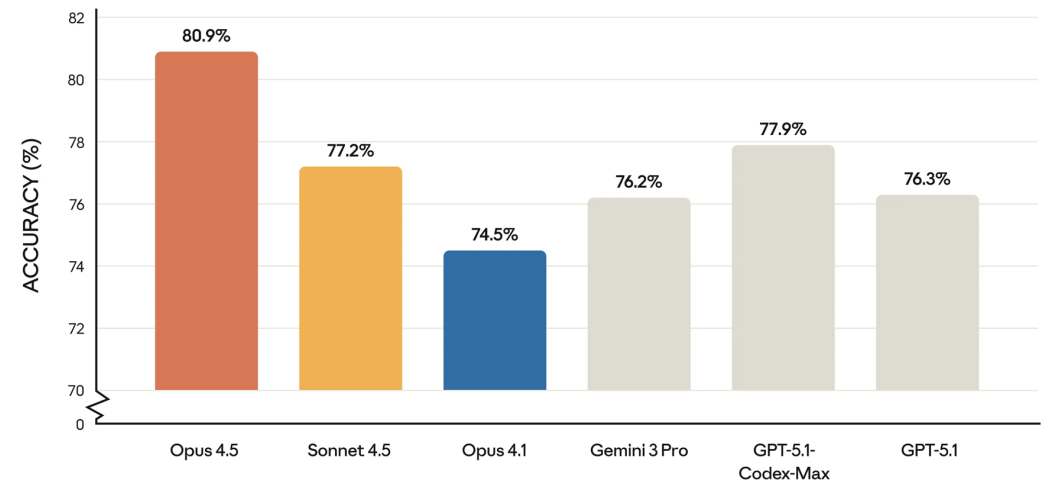
Artificial Analysis Intelligence Index; Higher is better



<https://artificialanalysis.ai/models>

Software engineering

SWE-bench Verified (n=500)



<https://www.anthropic.com/news/claude-opus-4-5>



The strategic gap

**1. Commercial AI
optimizes for speed
and utility**

**2. National security
requires verifiable
trust under fire**

**3. The market will
not solve this
organically**

*What if a program brought
together the best and brightest,
and leveraged \$3T already spent
to solve of the greatest problems
of our time?*



The strategic gap

1. Commercial AI optimizes for speed and utility

2. National security requires

3. The market will not solve this organically

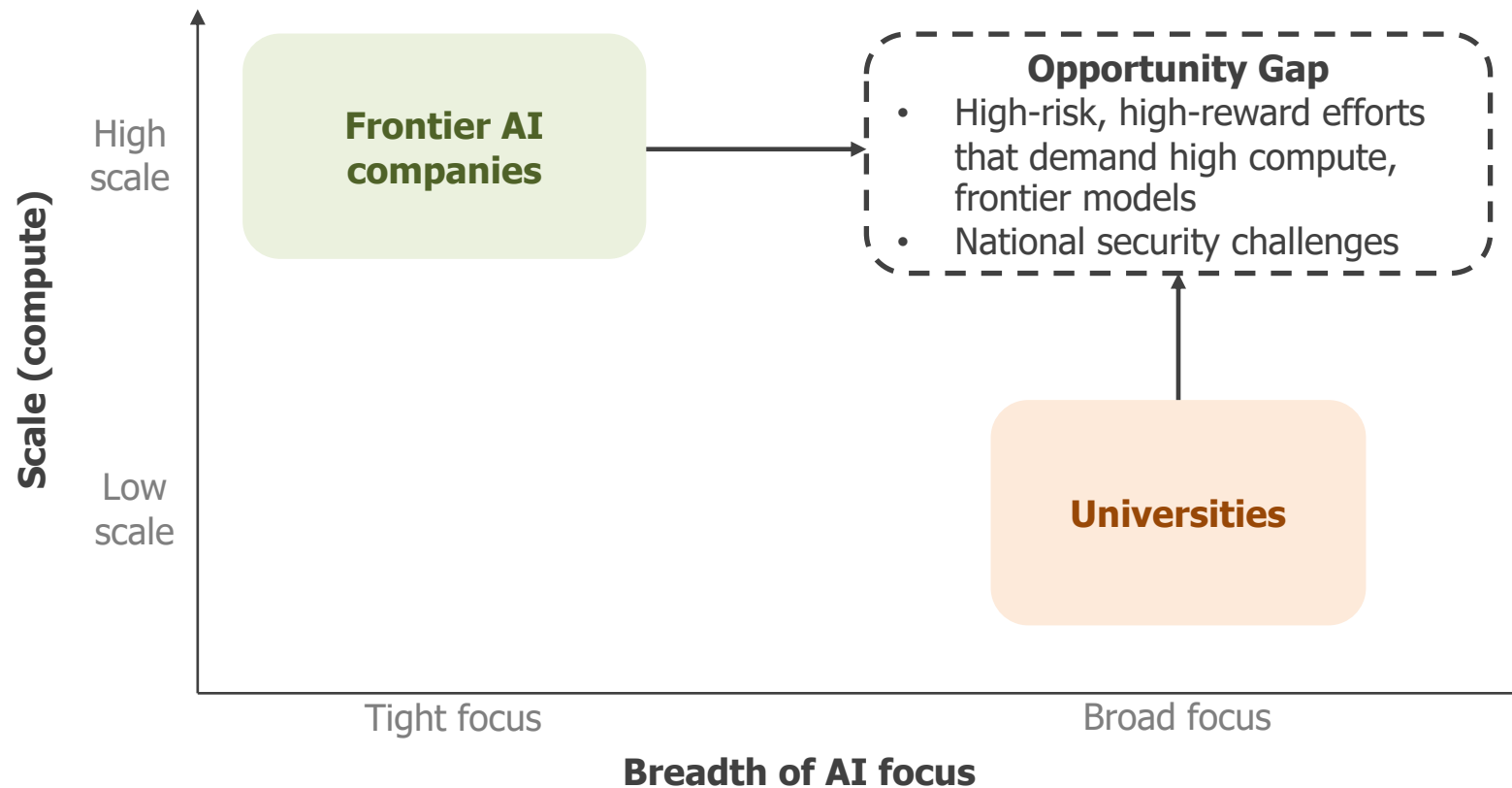


What if a program brought together the best and brightest, and leveraged \$3T already spent to solve some of the greatest problems of our time?



Opportunity: Combine new talent from universities with expertise and capabilities from frontier AI companies to get disruptive national security innovation

Properties of AI research

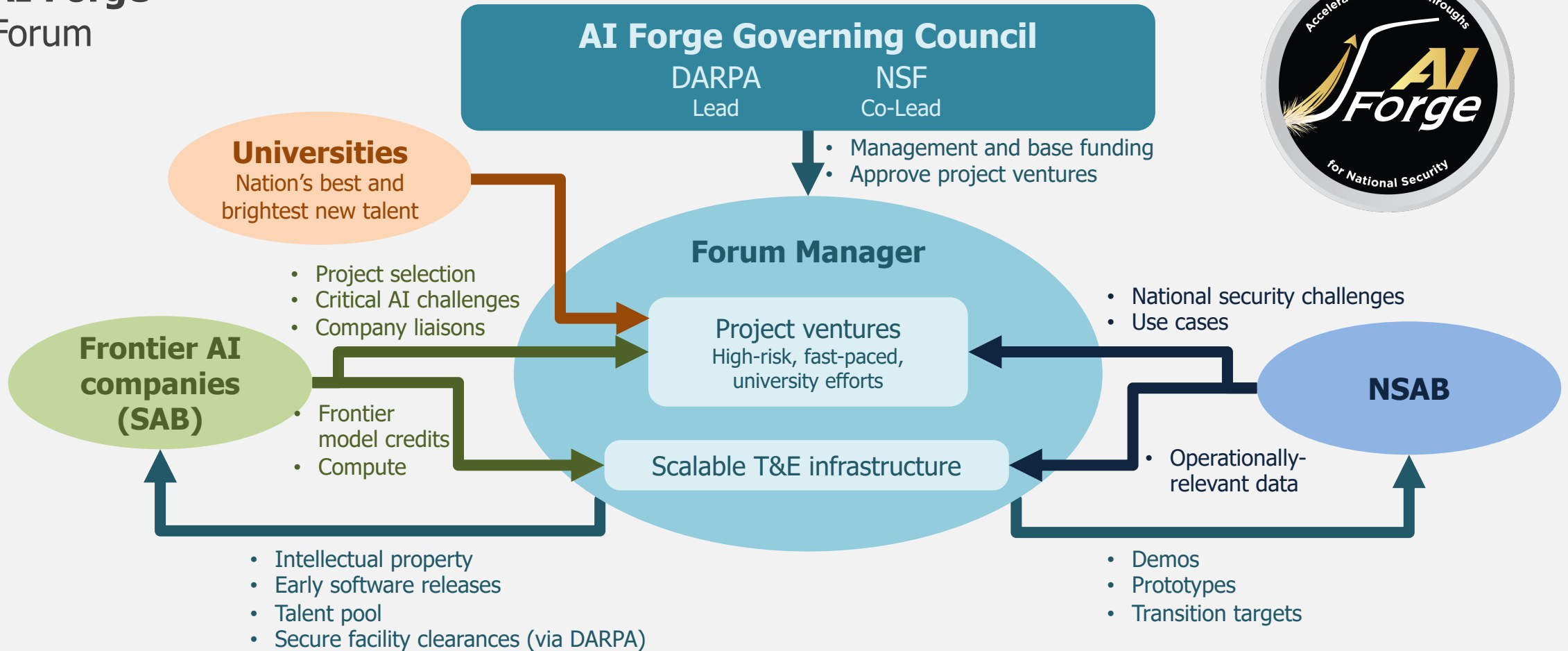




AI Forge: Accelerating AI breakthroughs for national security

A unique forum to enable rapid development of AI for national security from ideation to impact

AI Forge Forum





Defining the roadmap for secure AI

- **The Event:** January 2026 AI Forge Workshop
- **The Stakeholders:** 8 Frontier AI companies + 15 DoW/IC Chief AI Officers + NSF & NIST
- **The Output:** An unclassified, consensus-driven technical roadmap

AI Forge: Bridging commercial innovation with national security demands



Frontier AI challenges for national security

	The Good (Commercial AI)	The Bad (NatSec Challenge)	The Remedy (AI Forge)
AI interpretability Enabling actionable understanding	Scaling capabilities rapidly, solving complex problems	Opaque, “black box” decisions	Provide precise diagnostic visibility, enabling auditable AI
AI control Ensuring reliable performance	Achieving high levels of agentic autonomy	Brittle safety filters fail under stress	Embed mechanisms that reliably yield control
Adversarial robustness Securing AI for contested environments	Ingest vast amounts of internet-scale data and adapt continuously	Adversaries actively poison data, attack at machine speed	Engineer active defense protocols and interventional testbeds
Execution speed Executing at the speed of AI innovation	Rapid iteration and deployment	Multi-year, rigid government procurement	Fast-paced, 6-month venture cycles

Frontier AI: the most advanced and innovative AI models and algorithms



AI Interpretability Cycle 1 Challenges

**Scaling Causal
Interpretability for
Predictive Diagnostics**

**Black-Box Analysis of
Long-Horizon Failures for
Verifiable Accountability**

**Automated
Interpretability at Scale**

**Agentic AI Interpretability
for Auditable Autonomy**

**Evaluation of AI Systems
for Scientific Discovery**



AI Control Cycle 1 Challenges

**Verifiable Steerability of
AI Models**

**AI Provenance for Agile
and Reliable Deployment**

**Secure Sandboxing for
Agentic AI**

**Operational Runtime
Control and Automated
Intervention**

**Predictive Evaluation for
Mission Assurance**



Adversarial Robustness Cycle 1 Challenges

**Resilience to Training
Data Compromise**

**Defending Interactive AI
from Adaptive Adversaries**

**Active Defense Protocols
for Multi-Agent Systems**

**Hardening Continual
Learning Systems**

**Credible Benchmarking for
Adversarial Robustness**



Accelerating AI breakthroughs for national security, together.



www.darpa.mil

Contact: matthew.marge@darpa.mil