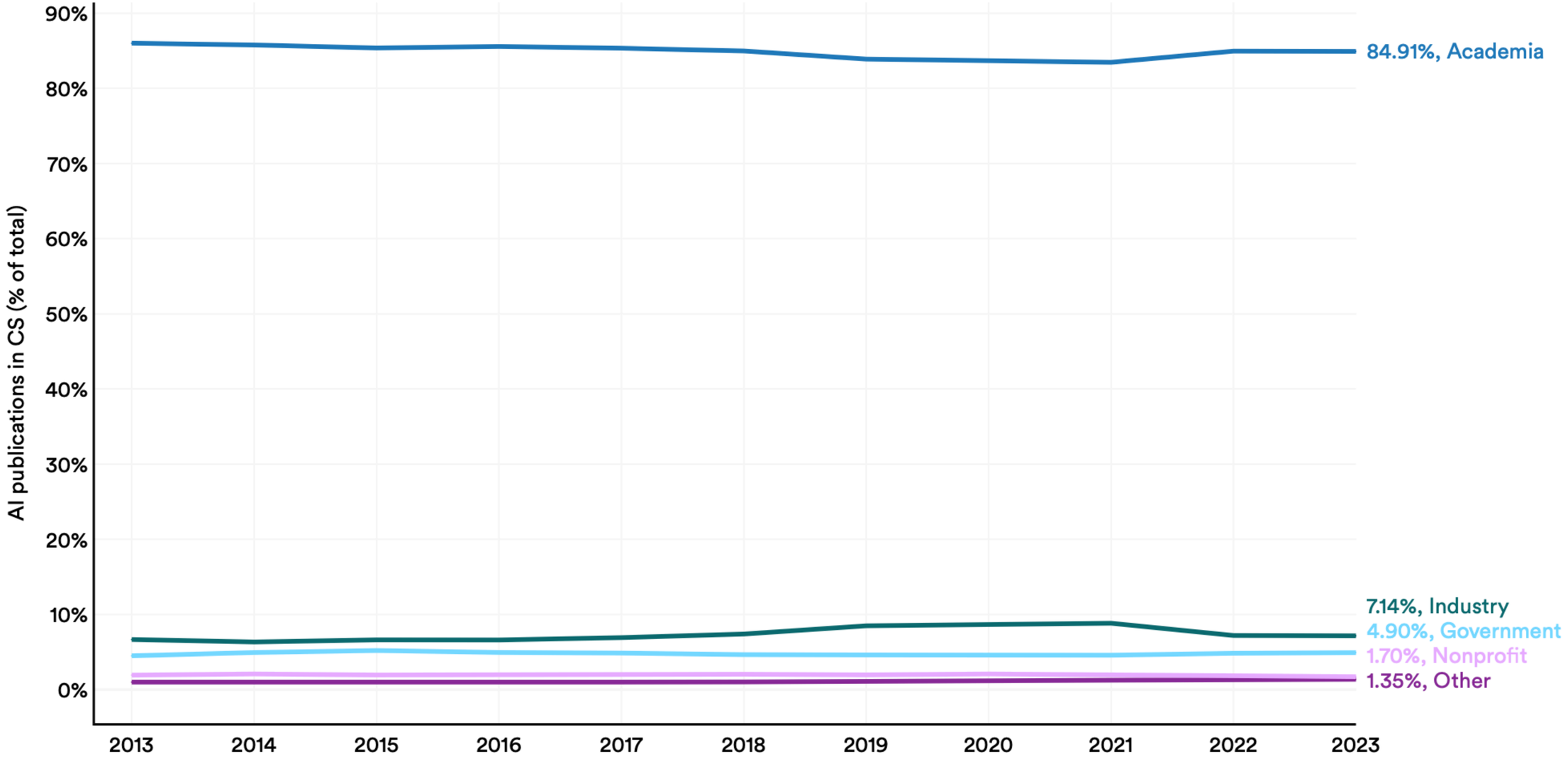

friend
AI is not your friend
[frend] noun
someone who listens, responds, and supports you.
WE DON'T NEED AI, WE NEED EACH OTHER.
DO MUTUAL AID OR VOLUNTEER FOR A COMMUNITY GARDEN — YOU WILL MEET COOL PEOPLE!
MUTUAL AID > FAKE AI CONVO
friend.com

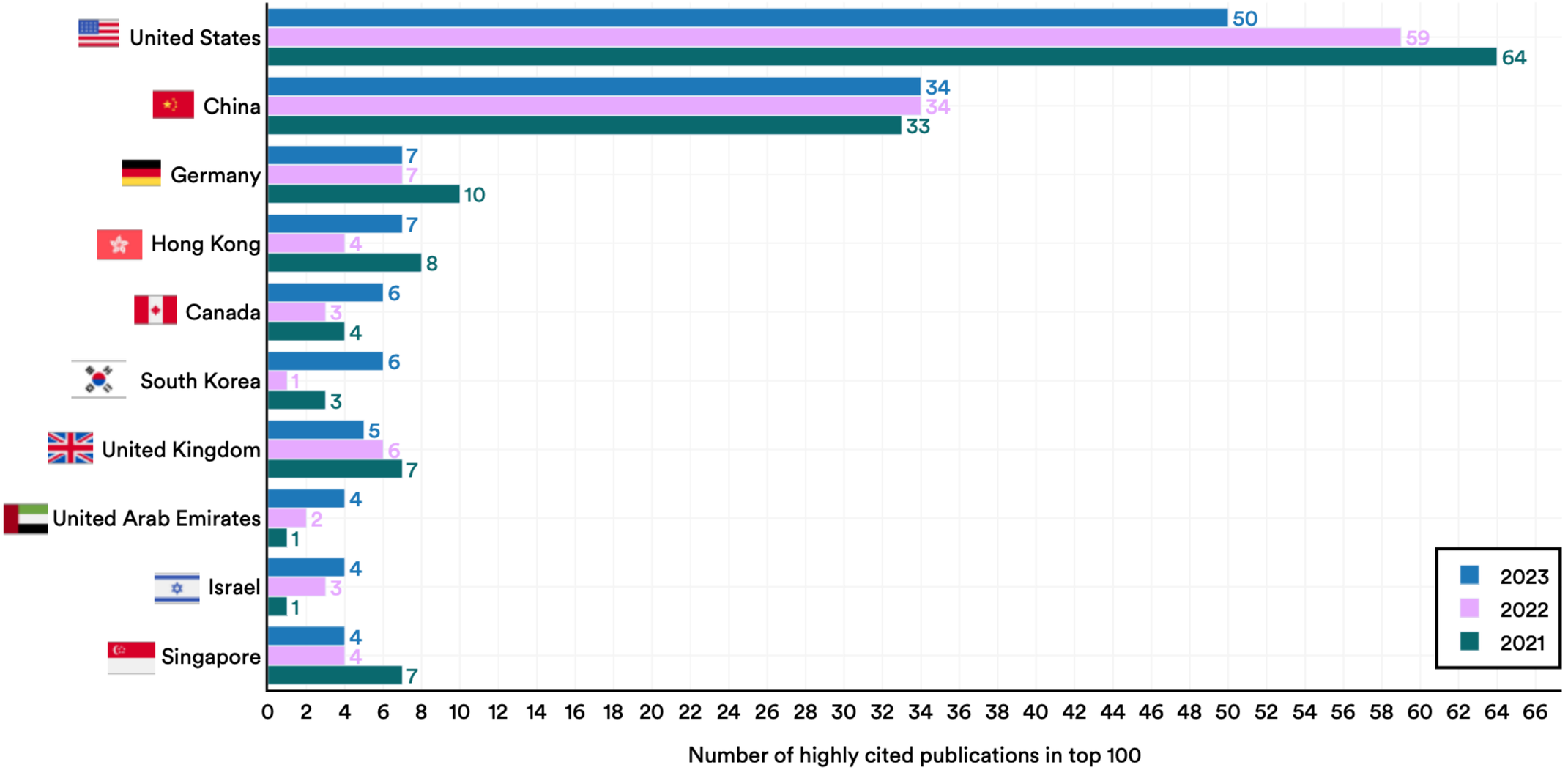
AI publications in CS (% of total) by sector, 2013–23

Source: AI Index, 2025 | Chart: 2025 AI Index report



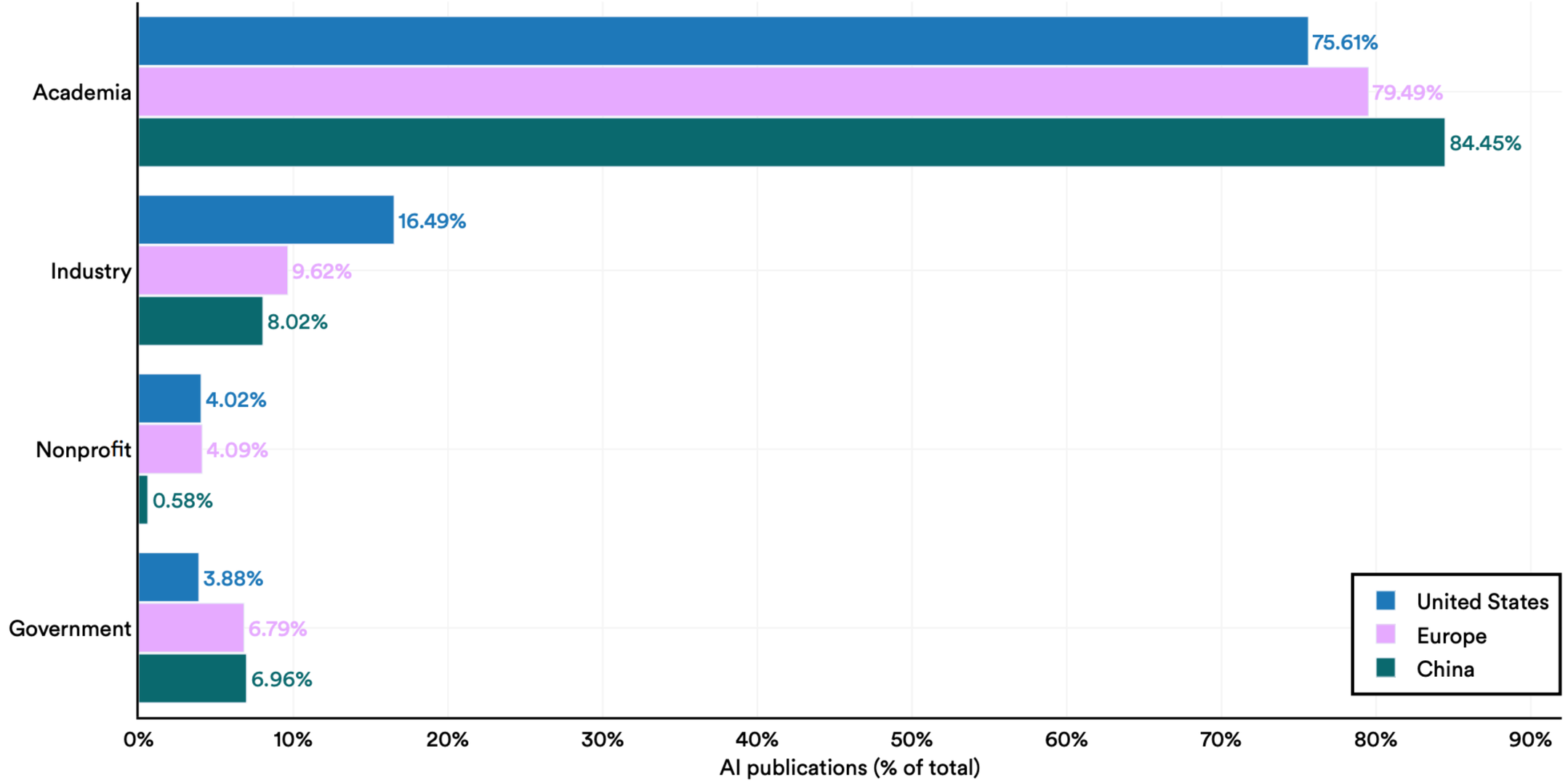
Number of highly cited publications in top 100 by select geographic areas, 2021–23

Source: AI Index, 2025 | Chart: 2025 AI Index report



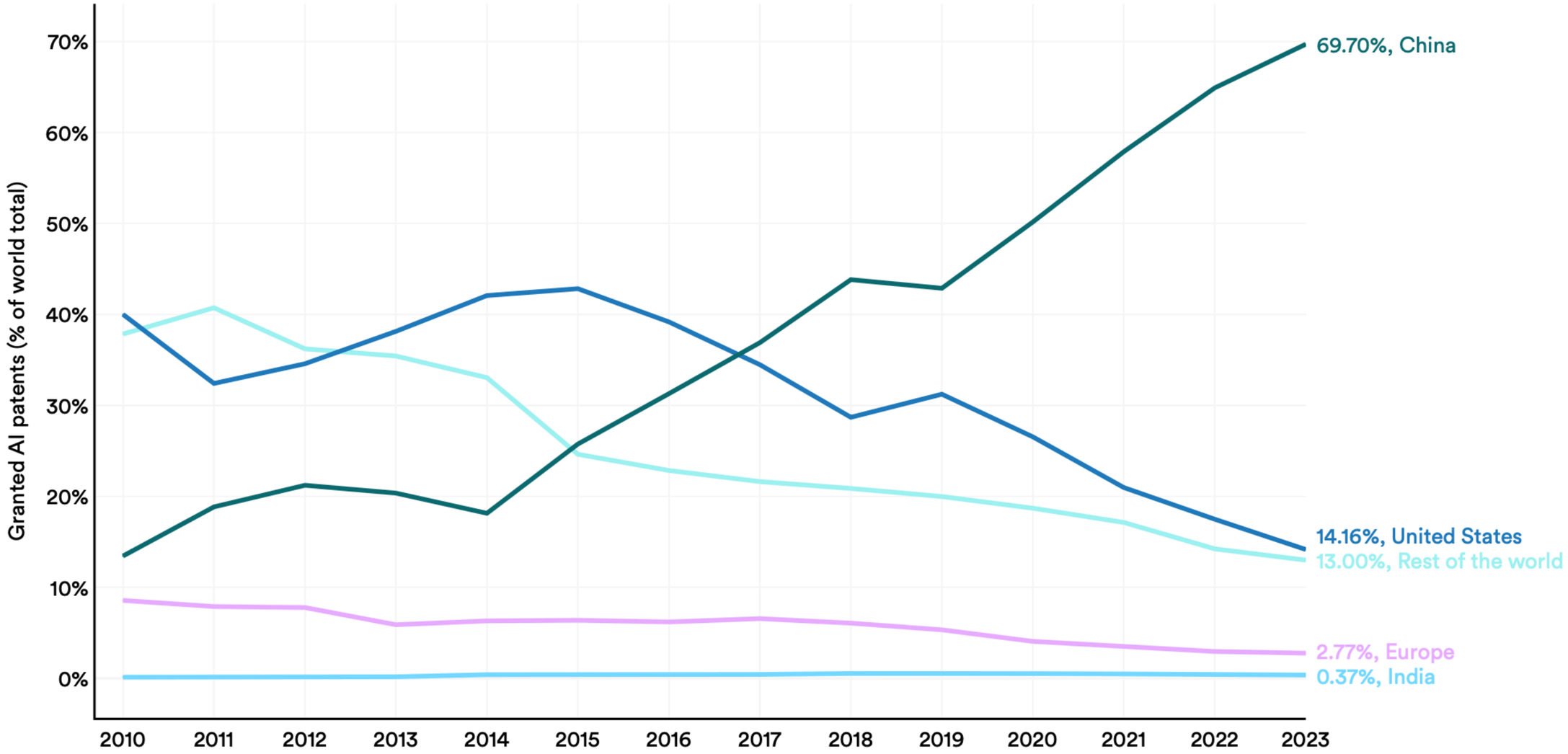
AI publications in CS (% of total) by sector and select geographic areas, 2023

Source: AI Index, 2025 | Chart: 2025 AI Index report



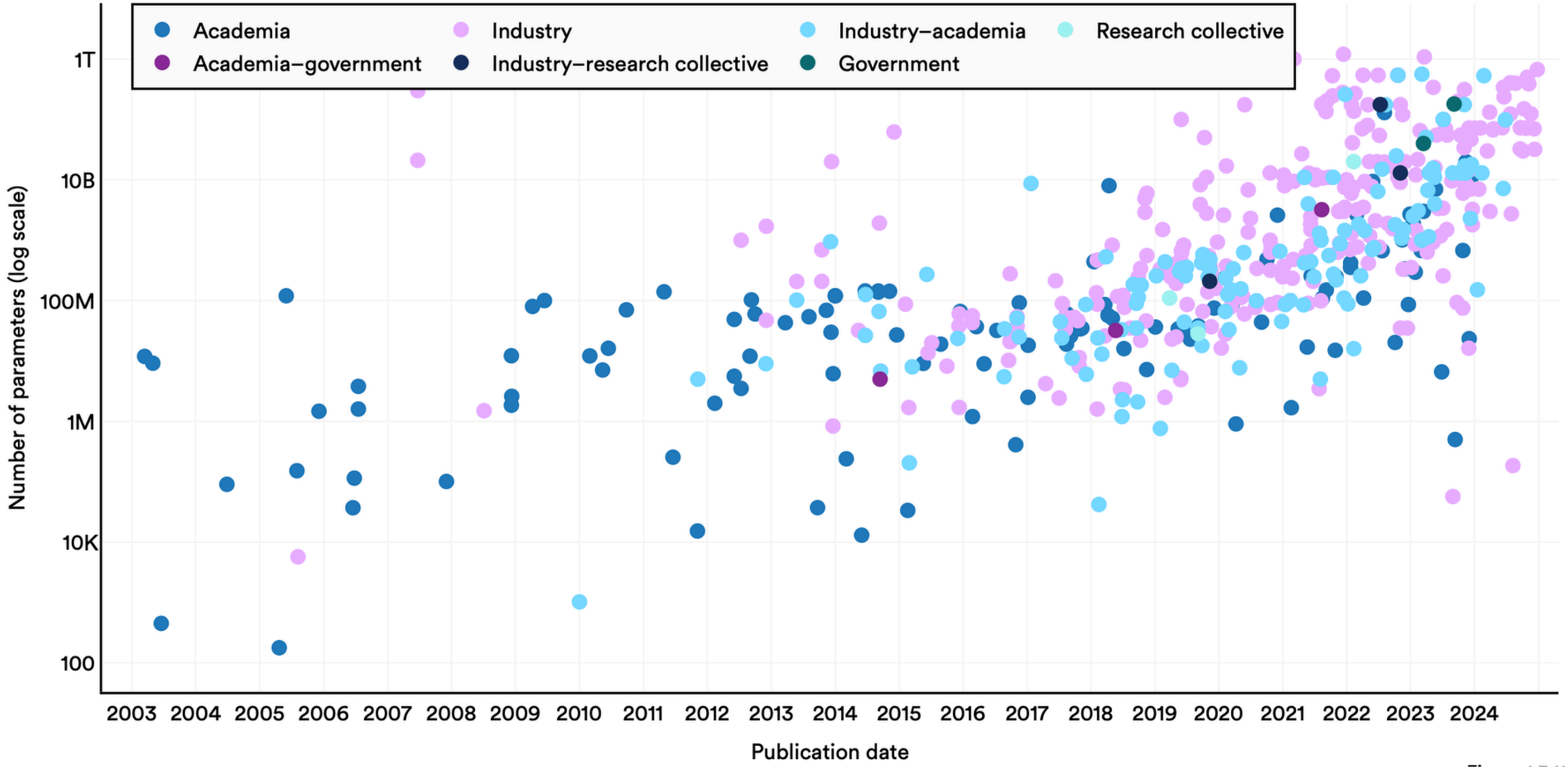
Granted AI patents (% of world total) by select geographic areas, 2010–23

Source: AI Index, 2025 | Chart: 2025 AI Index report



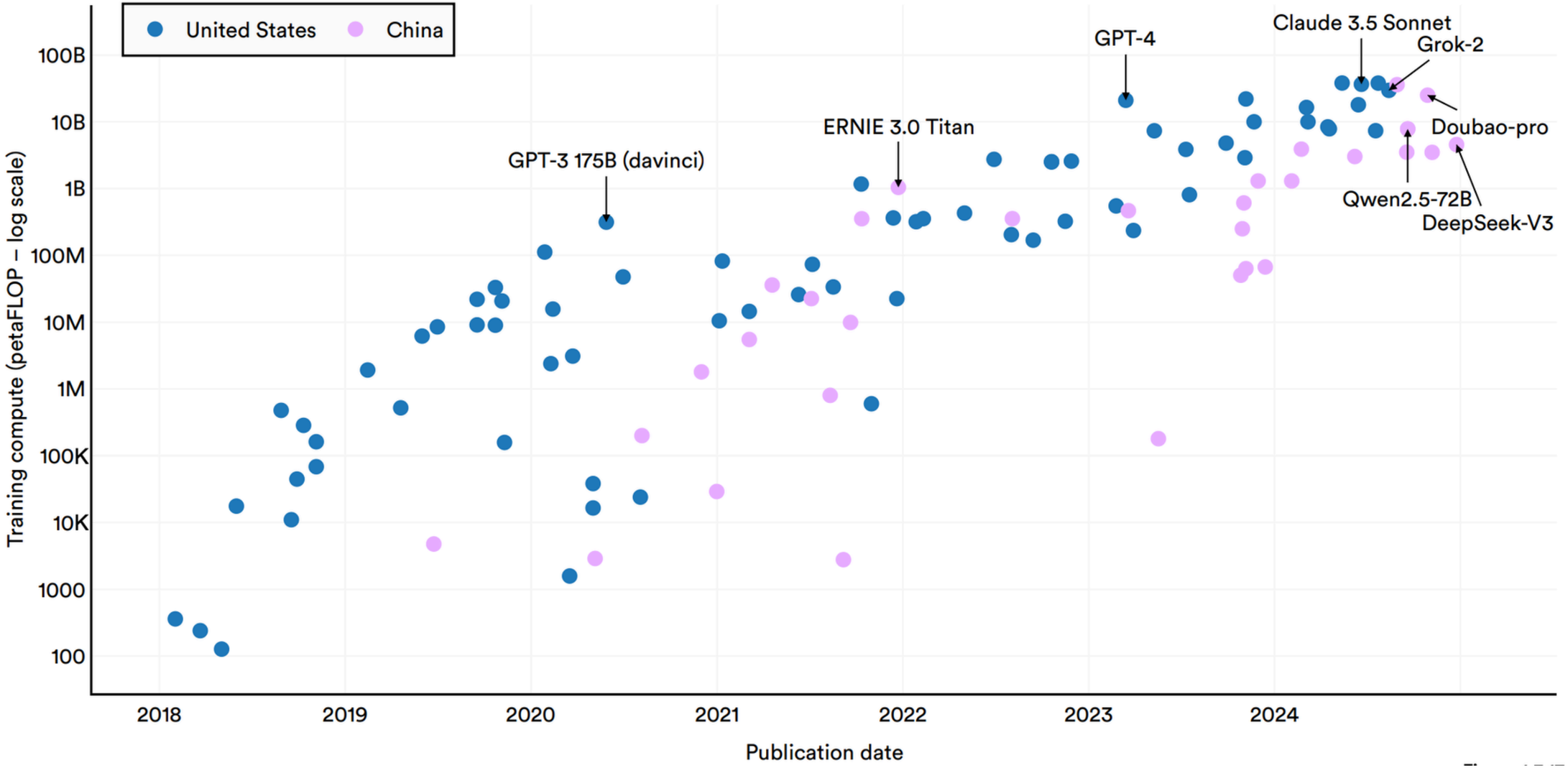
Number of parameters of notable AI models by sector, 2003–24

Source: Epoch AI, 2025 | Chart: 2025 AI Index report



Training compute of select notable AI models in the United States and China, 2018–24

Source: Epoch AI, 2025 | Chart: 2025 AI Index report



CYBERSECURITY INDUSTRY'S AI PROBLEM

4.7M

Talent shortage (ISC², 2023)

1,300%

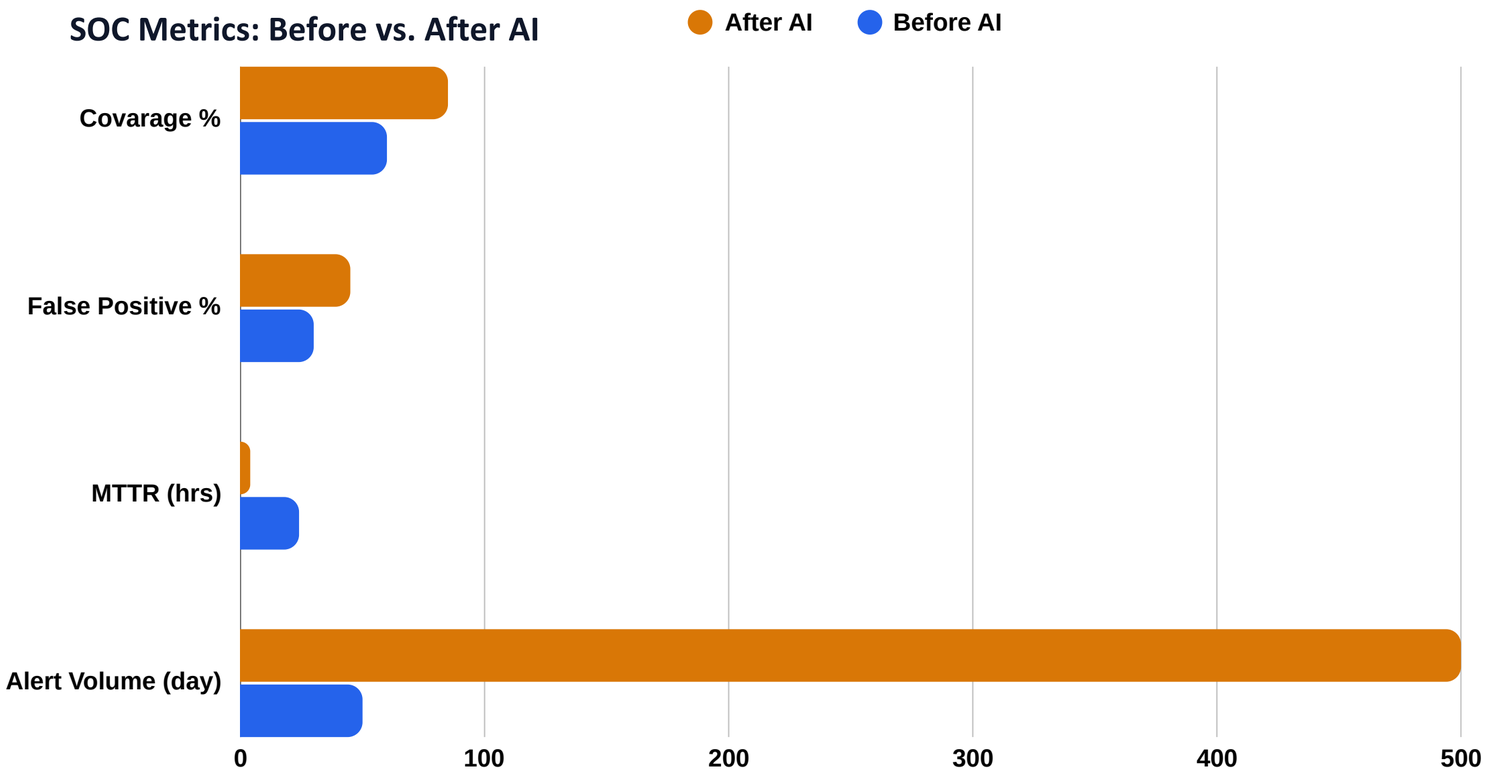
AI phishing rise (SlashNext, 2023)

$9.48B

AI breach cost (IBM, 2024)

72%

CISOs: AI = more vulns (Gartner)



Alert Fatigue Amplified

AI-powered SIEM generates 10x alerts. SOC teams overwhelmed; critical threats buried.

AI vs. AI Arms Race

Offensive AI iterates in hours; defenders take weeks to patch. Asymmetric disadvantage.

Black Box Compliance

Unexplainable AI creates GDPR/HIPAA gaps. 'Why did AI block this?' — no answer.

Training Data Bias

Data gaps cause ML to miss novel attacks. Zero-days evade historically-trained models.

AI's Weakest Link Is Still Human

82%

of breaches involve human error (Verizon DBIR, 2024)

3.4s

avg. time to click an AI phishing email (Proofpoint, 2024)

74%

of employees trust AI output without verification (MIT, 2023)

$4.9M

extra breach cost when AI over-relied upon by SOC teams (IBM, 2024)

Human-AI Risk Categories

🎣 AI-Powered Social Engineering

Spear phishing with AI-written, deeply personalized emails. Vishing using cloned voices. Smishing via AI chatbots. Hit rate 3x traditional phishing.

🤝 Over-Reliance on AI Decisions

Operators accepting AI recommendations blindly. SOC teams ignoring false negatives. Clinicians acting on AI diagnosis without review.

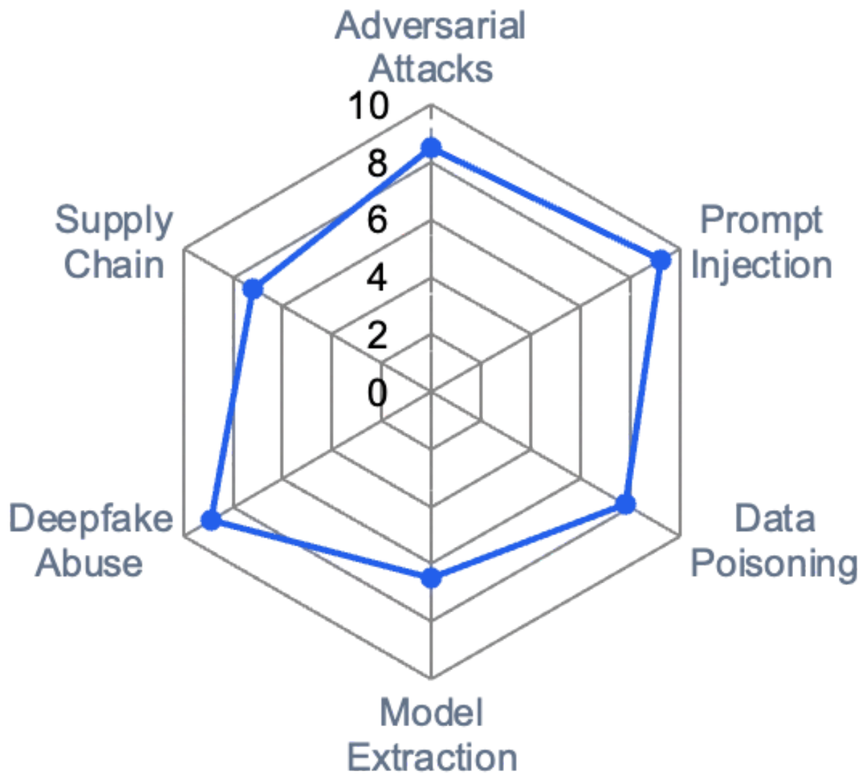
🕵️ Insider Threat Amplified by AI

Malicious insiders using AI tools to exfiltrate data faster. AI lowers technical bar — any employee can become an attacker.

🧠 Cognitive Bias in AI Training

Humans inject biases into training data. Teams overlook edge cases. Security researchers dismiss AI anomalies as false positives.

Vulnerability Severity Score (out of 10)



Adversarial Attacks

Invisible pixel edits fool image classifiers. Still unsolved in production. (Goodfellow, 2014)

Prompt Injection

OWASP LLM Top 10 #1 threat. Malicious inputs hijack AI instructions in all major models.

Data Poisoning

Training backdoors implanted via open-source supply chain. Microsoft Tay corrupted in hours.

Model Extraction

Stolen weights via API queries expose IP and enable adversary replicas (Papernot, 2016).

📊 *OWASP (2023): Prompt Injection is the #1 risk for LLM applications | IBM (2024): AI-assisted breach average cost = $9.48B | Gartner (2024): 72% of CISOs say AI created more vulnerabilities than it solved*

Same Technology Two Faces.

✅ BENEFICIAL USE		⚠️ MALICIOUS USE
Medical diagnosis, legal aid, education tutoring for billions	Large Language Models	WormGPT, FraudGPT — jailbroken models sold on dark web for malware & phishing
Cancer detection, self-driving cars, accessibility for blind users	Computer Vision	Deepfake pornography, face-swap fraud, autonomous targeting systems
Accessibility tools, language translation, content creation	Voice Synthesis (TTS)	CEO voice cloning fraud ($35M, HK 2024), political disinformation
Drug discovery, climate modeling, traffic optimization	Reinforcement Learning	Autonomous weapons, adaptive malware that learns to evade defenses
Medical imaging, architectural design, creative industries	Generative Image AI	Non-consensual intimate imagery, fake ID documents, propaganda

"The same knife that cuts bread can cut a throat" — Every dual-use technology challenge in history applies here, but AI operates at unprecedented speed and scale.

Who Regulates AI > And Who Doesn't

Region	Key Legislation	Status	Key Risk Area	Regulation Score /10
European Union	EU AI Act (2024)	**ENFORCED**	Prohibited high-risk AI	**9 / 10**
United States	EO on Safe/Secure AI (2023)	**PARTIAL**	Sector-specific only	**6 / 10**
China	Generative AI Regulations (2023)	**ENFORCED**	Content control focus	**7 / 10**
United Kingdom	Pro-Innovation Principles (2023)	**VOLUNTARY**	No binding law yet	**4 / 10**
UAE / Middle East	AI Strategy 2031	**STRATEGY**	Growth-first, light-touch	**3 / 10**
Russia	National AI Strategy	**MINIMAL**	No external oversight	**1 / 10**
India	National AI Policy (draft)	**DRAFT**	Regulatory gap remains	**3 / 10**
Brazil	AI Bill (pending Congress)	**PENDING**	Modeled on EU AI Act	**5 / 10**

RAND Corp

LLMs give minimal-expertise actors access to novel cyberweapon generation. Documented uplift in real exploit code.

OpenAI Red Team

GPT-4 generated working phishing kits & social engineering scripts via jailbreaks.

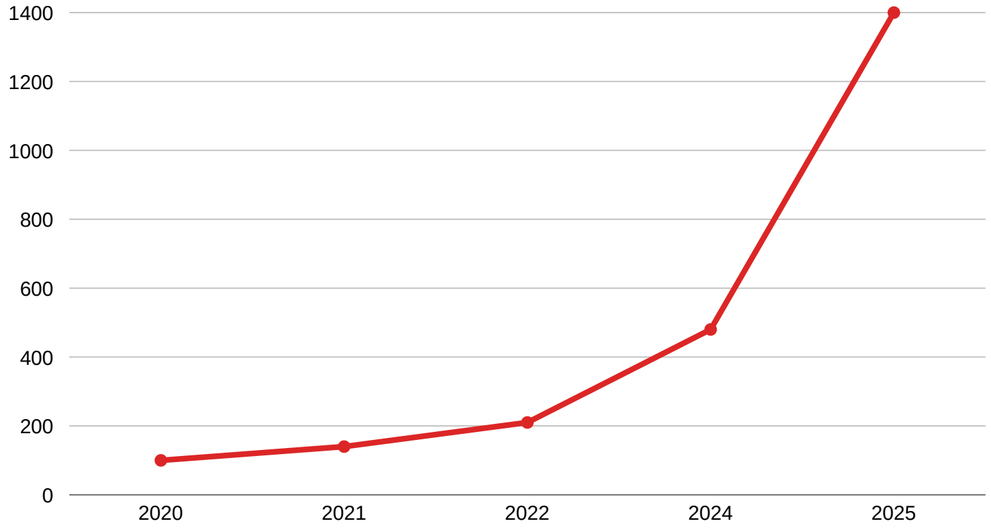
MIT CSAIL

AI deepfakes defeated facial recognition in 68% of tests. Nation-states now deploying synthetic media at scale.

NATO CCDCOE

AI-enabled autonomous weapons create a 'responsibility vacuum' in international humanitarian law.

AI-Aided Cyberattacks

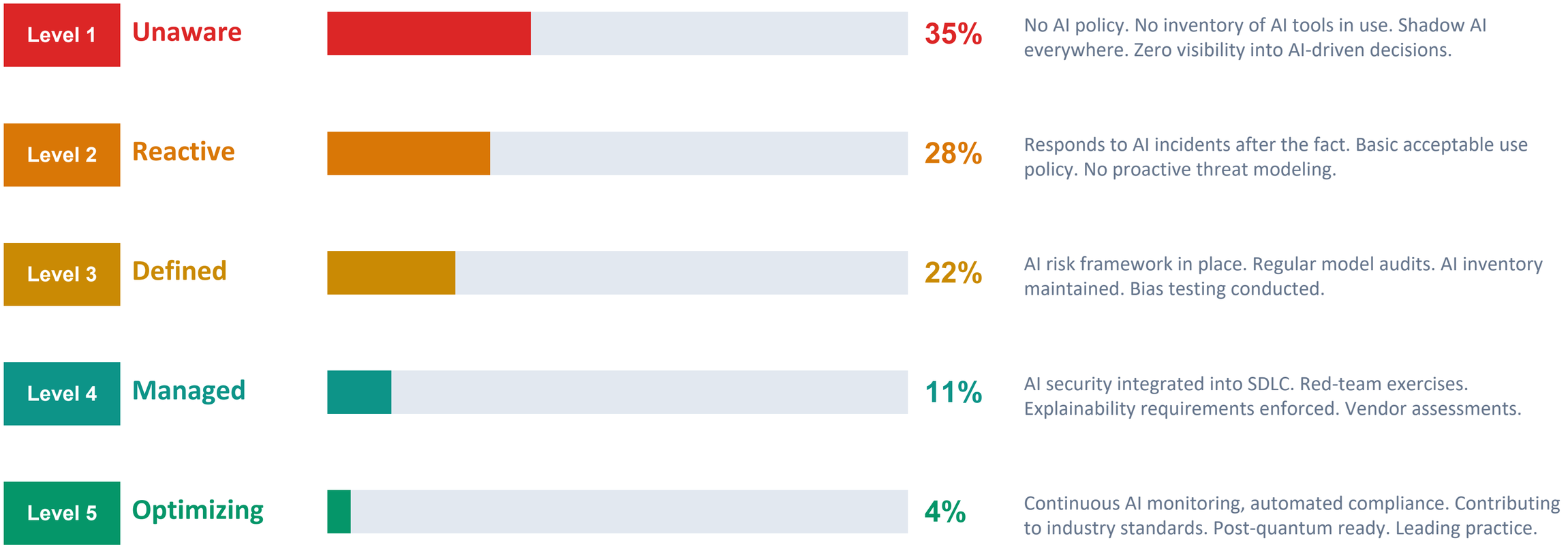


Key Incidents

Event	Year
WormGPT — dark web malware LLM	2023
MGM breach via AI social eng.	2023
$35M CEO voice clone fraud	2024
Ukraine AI drone swarms	2022–24

Where Is Your Organization?

Global Enterprise AI Maturity Distribution (Gartner / ISACA 2024)



🎯 **Self-Assessment: Most organizations overestimate their maturity by 1-2 levels. 63% of Level 1-2 organizations believe they are at Level 3+ (ISACA State of Cybersecurity 2024).**

Trump Signs AI Security Executive Order

An about-face for an administration that previously repudiated government AI reviews — June 2, 2026

Key EO Provisions

🏛️ Early Government Access to Powerful AI Models

Federal agencies receive pre-deployment access to frontier AI models to evaluate security risks before public release — a major shift from the prior administration's hands-off approach.

🔄 Policy Reversal: From Repudiation to Review

The directive represents a complete about-face. The Trump administration previously repealed Biden-era AI safety orders. This EO reinstates structured government AI review processes.

🛡️ Security-First AI Deployment Framework

AI systems deployed in critical infrastructure must pass government security vetting. Covers national security applications, energy, finance and healthcare sectors.

⚖️ Balancing Innovation & Safety

The EO attempts to balance America's AI competitiveness goals with security oversight requirements — mirroring tensions seen in EU AI Act debates.

US AI Policy Timeline

Date	Action	Impact
Oct 2023	Biden EO on AI Safety	Mandated safety testing, government
Jan 2025	Trump repeals Biden EO	Removed AI safety review requirements
2025	EU AI Act Enforcement begins	Global compliance pressure on US firms
Jun 2026	Trump AI Security EO signed	REVERSAL: Gov. access to frontier AI models

The Future of AI in Cybersecurity

🤖 Agentic AI SOC

Source: IBM / Swimlane

Fully autonomous security operations centers — AI agents that detect, investigate and remediate without human intervention. Swimlane's Turbine platform already delivers zero-human remediation.

🎭 Deepfake Arms Race

Source: Harvard CISO Panel

AI-generated video, voice and document forgeries will become indistinguishable from real. Harvard CISOs predict deepfake verification will be a mandatory enterprise capability by 2027.

🧬 Self-Healing AI Networks

Source: Sophos

AI systems that automatically patch vulnerabilities, update threat models and reconfigure defenses in real time — without human approval cycles. Sophos's adaptive ML framework is an early prototype.

⚖️ Regulatory Ratchet

Source: Cybersecurity Dive

Following the Trump AI EO (June 2026), expect rapid global regulatory convergence. All major economies will require pre-deployment AI security reviews within 3 years.

🔍 Explainable AI Security

Source: IBM watsonx

Growing compliance demands (GDPR, HIPAA, EU AI Act) will force all security AI to explain its decisions. IBM's watsonx.governance framework represents the commercial solution emerging now.

⚛️ Post-Quantum Transition

Source: All Sources

Q-Day threat (est. 2030) requires all AI security infrastructure to migrate to NIST PQC algorithms now. Harvest-now-decrypt-later attacks already underway against AI model APIs.

Key Takeaways

1 **AI has democratized cybercrime**

The technical skill barrier is gone. Any threat actor with AI access can launch sophisticated, personalized attacks. (Harvard CISO Panel 2025)

2 **AI is also our best defense**

55% faster triage, 90% fraud reduction, autonomous response under 1 second — the same technology that enables attacks also defeats them. (IBM Security 2025)

3 **7 proven AI use cases exist today**

Threat detection, phishing response, SOAR, vulnerability management, threat hunting, reporting, case management — deployable now. (Swimlane 2025)

4 **Government oversight is coming**

Trump's June 2026 AI EO marks a global inflection. Mandatory pre-deployment security reviews will become standard across all major markets. (Cybersecurity Dive 2026)

5 **Prepare now across three dimensions**

Detect unknown threats, harden your AI stack, and automate response. Organizations at AI Maturity Level 1-2 are already exposed. (Harvard + Sophos)

SOURCES

Harvard Extension School

AI & Future of Cybersecurity
CISO Panel, 2025

Sophos

AI in Cybersecurity
Explained, 2025

IBM Security

AI Cybersecurity
Solutions, 2025

Cybersecurity Dive

Trump AI Security EO
June 2, 2026

Swimlane

7 AI Use Cases in
Cybersecurity, 2025

AI Governance

2021.AI · BCOM · Aiceberg · Apple · ARMILLA · Asenion · AUDITBOARD · BreezeML · calvinrisk · causaLens · Citrusˣ · CRANIUM · credo ai · CYERA · Daiki

AI Quality & Evaluation

Agenta · Andon Labs · ANTHROP\C · arize · Arthur · atla · Autoblocks · census · CLOUDERA · Confident AI · Coralogix · datatron · deepchecks. · Efemarai · ETIQ AI

AI Security-As-A-Service & Training

AUGUR · Bright · CovertSwarm · CYBERREADY · Dune · helicone · Jericho Security · Lasso · LUMU · METACERT · OutThink · semperis · Spherical Defence · TWOSENSE · vansec

Application & Platform Security

agentuity · amplify · braintrust · Composio · CORGEA · crewai · DRYRUN SECURITY · ghost · LangChain · mobb · nullify · Pixee · promptfoo · redaccess · Straiker

Counter-AI & Threat Protection

Abnormal · Advai · ADVERSA · AI or Not · Arkose Labs · AskUI · Astrix · BioID · BugBase · Deepkeep · DUCK DUCK GOOSE · Clarity · GetReal · Giskard · hiya

Data Security & Protection

1touch.io · APHERIS · aviatrix · bonfy.ai · ClearlyAI · Duality · EDGELESS SYSTEMS · HIRUNDO · Kiteworks · Metomic · MIND · Nightfall · OPAQUE · PROTOPIA · Roseman

Discovery & Observability

AgentOps, akto, Capsule, CLUTCH, CREDAL, HoneyHive, OASIS, OBSIDIAN, onyx, Openlayer, ORY, reco, SALT, Teleport, WITNESS AI

Identity

aembit, AKEYLESS, AMPERSAND, andromeda, ANON, Arcade, britive, ConductorOne, CYATA, DeepTrail, entro, EVE, frontegg, GitGuardian, HushSecurity

Model Consumption Security

ActiveFence, BLACKFOG, blaxel, BOSCH, Daytona, E2B, Guardrail Technologies, Guardrails AI, harmonic, KNOSTIC, Mitigant, NOMA, STACKLOK, VIANOPS, zenity

Model Security

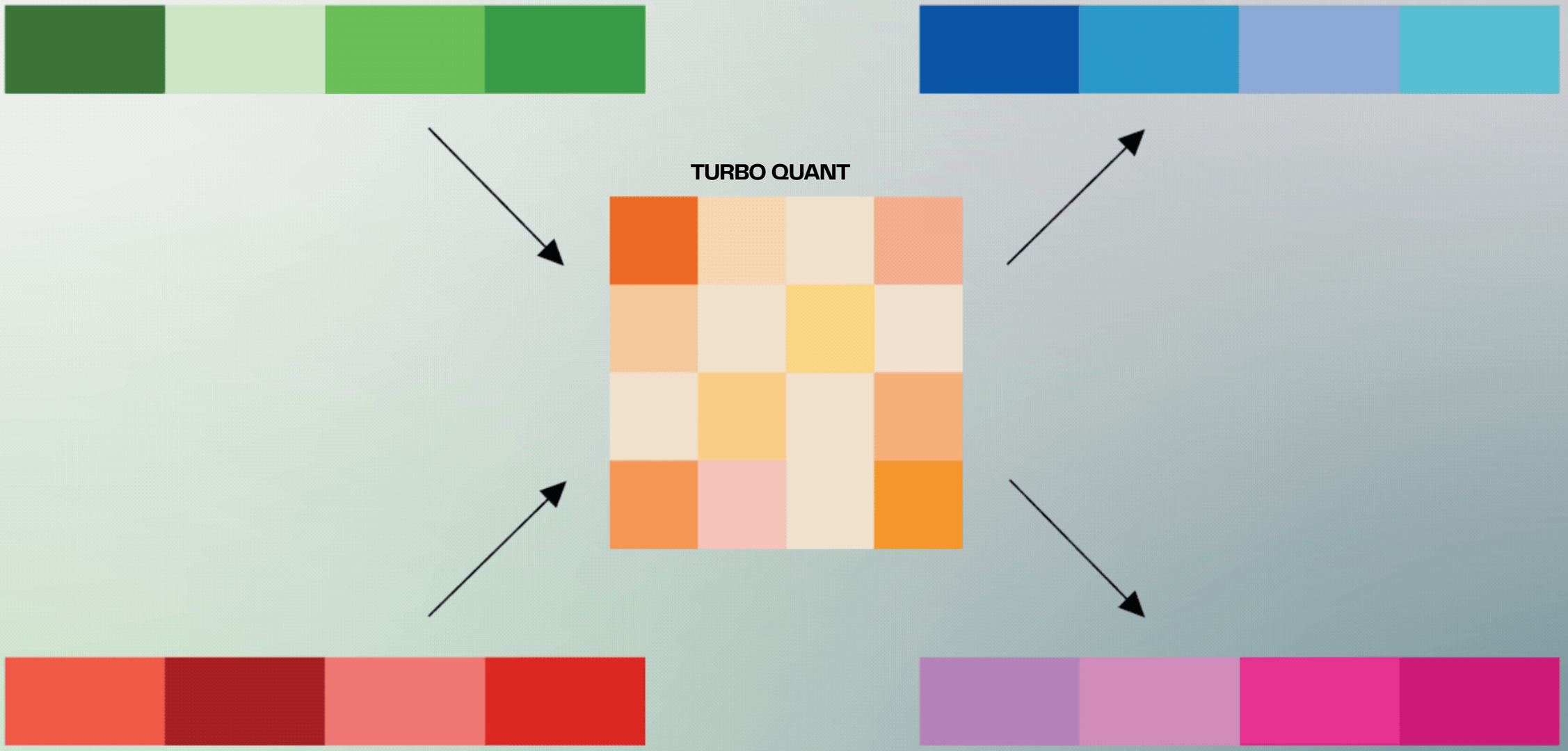
Bitfount, Dynamo AI, Flower, Giskard, HIDDENLAYER, integrate.ai, MOSTLY AI, NimbleEdge, NVIDIA, OCTAIPIPE, OWKIN, Rhino, sherpa.ai, TONIC, TROJ.AI

Runtime Control & Interface Security

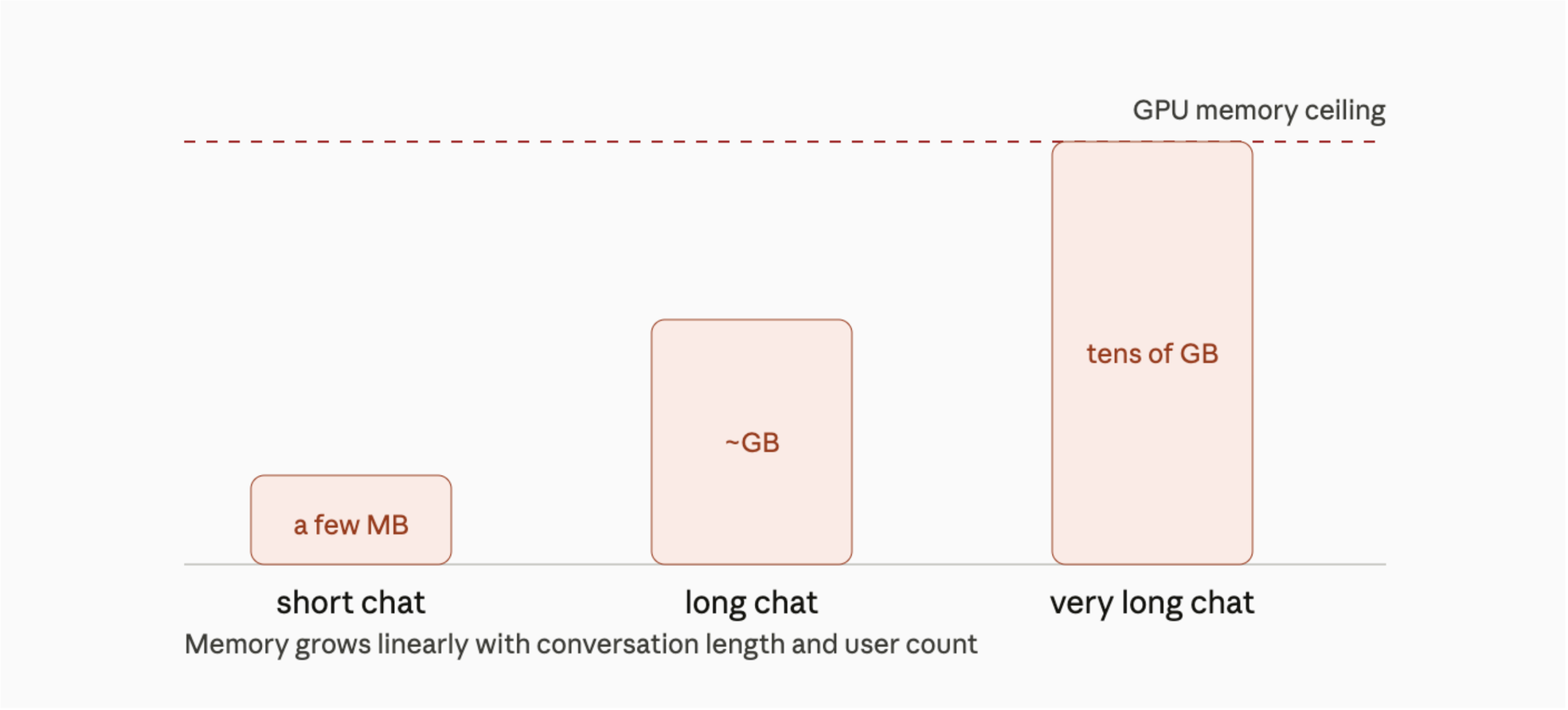
descope, docker, Enkrypt AI, Kong, lunar.dev, natoma, obot, Operant, ORY, POMERIUM, Runlayer, sgnl, sonoma, truefoundry, ZIOSEC

Security Automation & Operations

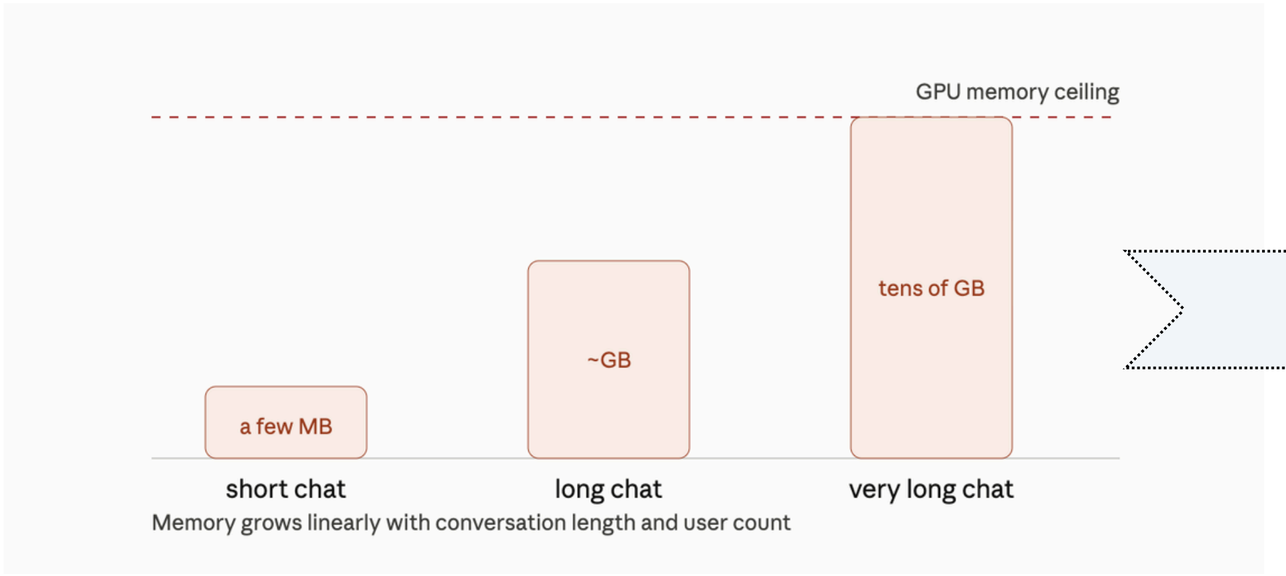
7AI, AIRMDR, AMPLIFIER, arcanna.ai, Aurascape, Bricklayer AI, CipherData, CML ZETA, CULMINATE, CultureAI, detections.ai, Dropzone AI, embed, Exaforce, Fable Security

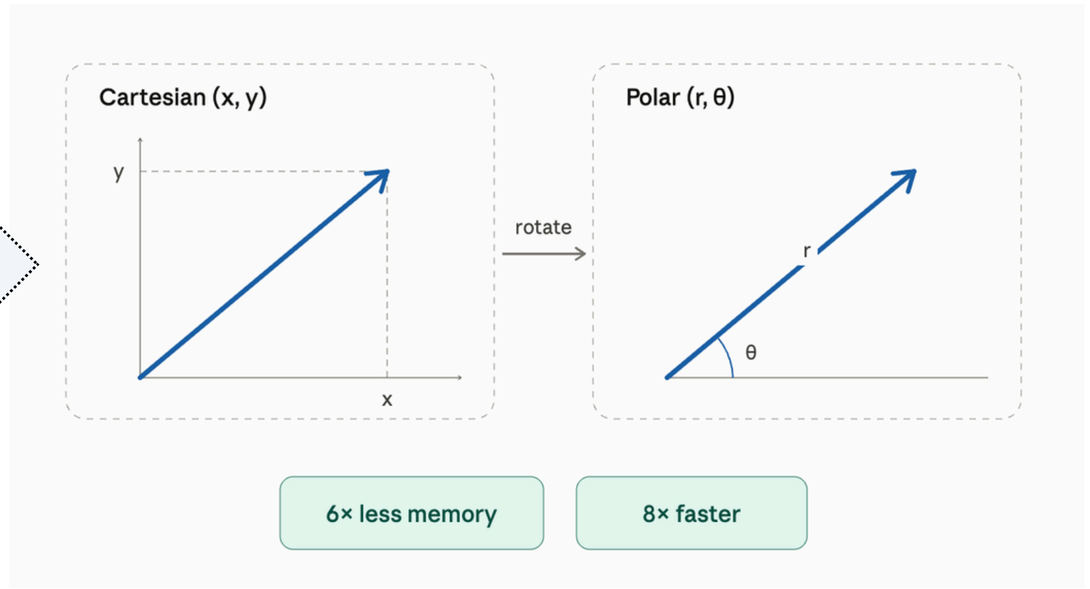

TURBO QUANT

THE KV CACHE BALLOONS



THE KV CACHE BALLOONS






Micron Technology Inc.
NASDAQ: MU
382,09 USD
+ Takip et
−13,44 (%3,40) ↓ bugün
Kapanış tarihi: 25 Mar 19:59 GMT-4 • Sorumluluk reddi beyanı
Borsa kapanış saatinden sonra 376,30 −5,79 (%1,52)
1G 5G 1A 6A YBK 1Y 5Y Maks.
400
390
380
370
376,30 USD 19:59
Önceki kapanış 395,53
10:00 12:00 14:00 16:00 18:00 20:00
Açılış 383,00
Yüksek 388,92
Piyasa değ 430,90 Mr
F/K oranı 18,04
Temettü %0,16
Üç Ay Tm Mk 0,15


SK Hynix
KRX: 000660
932.000 KRW
+ Takip et
−63.000 (%6,33) ↓ bugün
26 Mar 15:01 GMT+9 • Sorumluluk reddi beyanı
1G 5G 1A 6A YBK 1Y 5Y Maks.
1.000.000
980.000
960.000
940.000
920.000
Önceki kapanış 995.000
09:00 11:00 13:00 15:00
Açılış 965.000
Yüksek 973.000
Piyasa değ 664,95 Tn
F/K oranı 15,44
Temettü %0,32
Üç Ay Tm Mk 745,60


Winbond Electronics Corp
TPE: 2344
96,00 TWD
+ Takip et
−2,20 (%2,24) ↓ bugün
26 Mar 13:30 GMT+8 • Sorumluluk reddi beyanı
1G 5G 1A 6A YBK 1Y 5Y Maks.
104
102
100
98
96
Önceki kapanış 98,20
09:00 10:00 11:00 12:00 13:00
Açılış 100,00
Yüksek 104,50
Düşük 96,00
Piyasa değ 432,00 Mr
F/K oranı 109,07
52h yüksek 136,00
Temettü %0,52
Üç Ay Tm Mk 0,12
52h düşük 13,60


Samsung Electronics Co Ltd
KRX: 005930
180.100 KRW
+ Takip et
−8.900 (%4,71) ↓ bugün
26 Mar 15:01 GMT+9 • Sorumluluk reddi beyanı
1G 5G 1A 6A YBK 1Y 5Y Maks.
190.000
185.000
180.000
175.000
180.800 KRW 14:25
Önceki kapanış 189.000
09:00 11:00 13:00 15:00
Açılış 185.500
Yüksek 185.900
Piyasa değ 1170,66 Tn
F/K oranı 27,27
Temettü %0,93
Üç Ay Tm Mk 418,73

Probabilistic guess vs deterministic certainty

Incoming file

Classic: malicious?
Asks a probabilistic question

↓

Probabilistic guess
99% accuracy ≠ certainty

↓

Hallucination risk
Malware passes or legit blocked

Klearis: asks nothing
Decomposes into components

↓

Neutralizes off-spec
Macros, embedded code, exe

↓

Rebuilds the file
From a known-good template

↓

Certain clean file
Nothing off-spec passes

AI = force multiplier, not the final decision

Machine speed + human judgment + engineering certainty

AI detection under adversarial attack

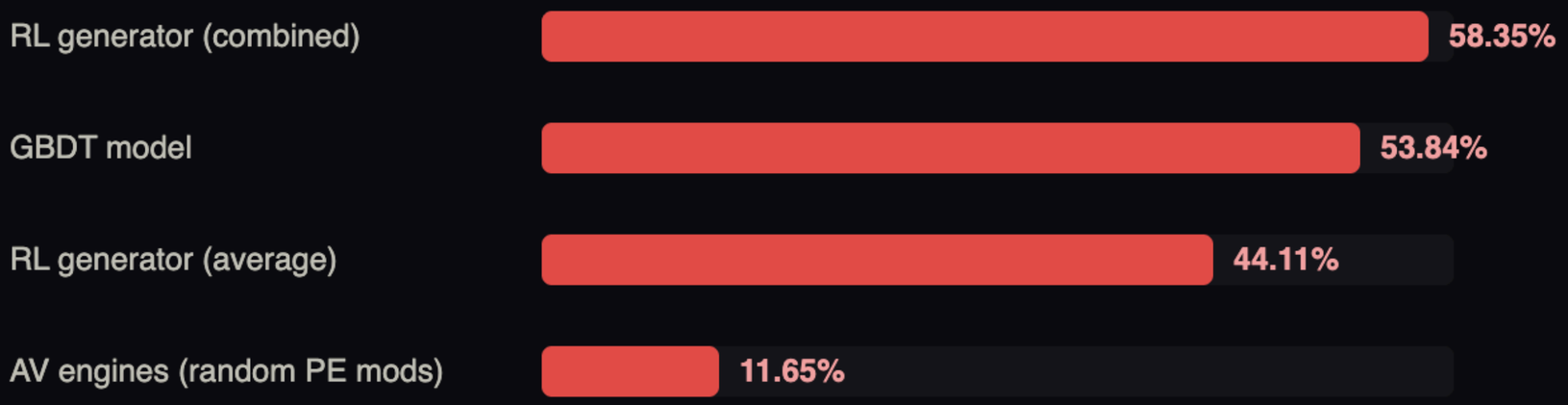
Lab accuracy collapses once attackers learn the model's decision boundary

Model accuracy vs adversarial examples

90.3% Clean inputs → **64.1%** Adversarial inputs

–26.2 points

Evasion rate — how often crafted malware fools the model



Persistent weaknesses of ML-based detection

- High false-positive rates
- Data-quality dependence
- Adversarial vulnerability

Figures from published academic studies; real-world field results are typically lower.

WORKSPACE

Pwnly

$ new scan

0 CTIVE · 71 FINDINGS · 23 CRITICAL+HIGH · 1 ASSETS

SEVERITY DISTRIBUTION

71 findings · last 30 days

CRITICAL 14 HIGH 9 MEDIUM 18 LOW 30 INFO 0

RECENT SCANS

all →

http://graphql.testinvicti.com	DEPTH 1	about 1 month ago
http://graphql.testinvicti.com	DEPTH 1	about 1 month ago
http://graphql.testinvicti.com	DEPTH 1	about 1 month ago
http://graphql.testinvicti.com	DEPTH 1	about 1 month ago
http://graphql.testinvicti.com	DEPTH 1	about 1 month ago
http://graphql.testinvicti.com	DEPTH 1	about 1 month ago

CRITICAL & HIGH FINDINGS

all →

CRIT Unauthenticated EJS template injection in getWebsiteOffer executes server-...
http://graphql.testinvicti.com/graphql

CRIT Unauthenticated SSRF in getWebsiteSeoScore reaches AWS instance meta...
http://graphql.testinvicti.com/graphql

CRIT Unauthenticated command injection in GraphQL createFunnyPicture returns...
http://graphql.testinvicti.com/graphql

CRIT Unauthenticated SSRF in GraphQL getWebsiteSeoScore reaches AWS meta...
http://graphql.testinvicti.com/graphql

CRIT Unauthenticated EJS template injection in GraphQL getWebsiteOffer
http://graphql.testinvicti.com/graphql

CRIT Unauthenticated command injection in GraphQL createFunnyPicture
http://graphql.testinvicti.com/graphql

CRIT Unauthenticated SSRF reaches AWS metadata and leaks temporary cloud c...
http://graphql.testinvicti.com/graphql

CRIT Unauthenticated command injection in createFunnyPicture leads to remote ...
http://graphql.testinvicti.com/graphql

FRIEND ✓

Yes with governance

→ Saves lives in healthcare & climate

→ Democratizes education globally

→ Creates $4.4T economic opportunity

→ Requires: transparency & oversight

WEAPON ⚠️

Already happening

→ Autonomous lethal drones deployed

→ AI disinformation at nation-state scale

→ Cyberweapons lower attack barrier

→ Requires: international treaty law

ENEMY ?

Not inherently;

→ Systemic bias harms marginalized groups

→ Job displacement without safety nets

→ Surveillance capitalism enabled by AI

→ Requires: human-centered design