State of AI Report 2025 Summary

By Alvaro de Nicolas

Original report by: By Nathan Benaich, Air Street Capital
October 18, 2025

Agenda

1. Research Breakthroughs in Reasoning AI
2. Open vs Proprietary AI Models
3. AI Safety: Alignment & Hallucinations
4. Industry Growth & AI Commercialization
5. AI Compute & Chip Market Dynamics
6. Geopolitics of AI: US, China, Gulf
7. AI Regulation Landscape
8. AI Survey: Usage & Productivity
9. AI Safety Spending vs Risks
10. AI Incident Trends & Threats
11. AI Agents & Open-Ended Learning
12. Predictions for Next 12 Months

Executive Summary Highlights

Enormous ***tension*** exist in the industry between **rapid AI advances** and the need for robust safety and **governance**

01 Research Breakthroughs

Reasoning models like OpenAI's o1 and DeepSeek's R1 advanced AI problem-solving, but progress may be illusory due to benchmark variance and hallucination issues.

02 Industry Growth

AI-first companies reached $18.5B in annual revenue with rapid adoption; NVIDIA dominates chips while neoclouds and sovereign AI investments grow.

03 Political Dynamics

The US pursues America-first AI strategy with export controls; China leads open models and sovereign AI buildouts; global AI governance faces setbacks.

04 Safety & Risks

Labs increase safety measures amid rising AI-enabled cybercrime and alignment challenges; external safety organizations remain underfunded.

Research Breakthroughs in Reasoning AI

Sep 2024

OpenAI o1-preview

OpenAI launched o1-preview, using RL and Chain-of-Thought **to boost accuracy on tough tasks** like the AIME math exam.

Dec 2024

DeepSeek R1-lite

DeepSeek released R1-lite-preview, outperforming o1-preview on AIME 2024, showing **strong open-weight reasoning** models.

Dec 2024

DeepSeek V3 & R1-Zero

DeepSeek's V3 MoE model and R1-Zero trained with RL **achieved big accuracy gains on benchmarks** like AIME and MATH-500.

Apr 2025

OpenAI o3/o4-mini

OpenAI launched o3 and o4-mini models, **enhancing reasoning and cost efficiency** to keep frontier leadership.

Jun 2025

Gemini 2.5 Pro

Google DeepMind’s Gemini 2.5 Pro introduced **transparent step-by-step reasoning and parallel inference** methods.

Aug 2025

OpenAI GPT-5

GPT-5 launched with **top-tier reasoning and a 400K token context** window, solidifying US leadership in AI.

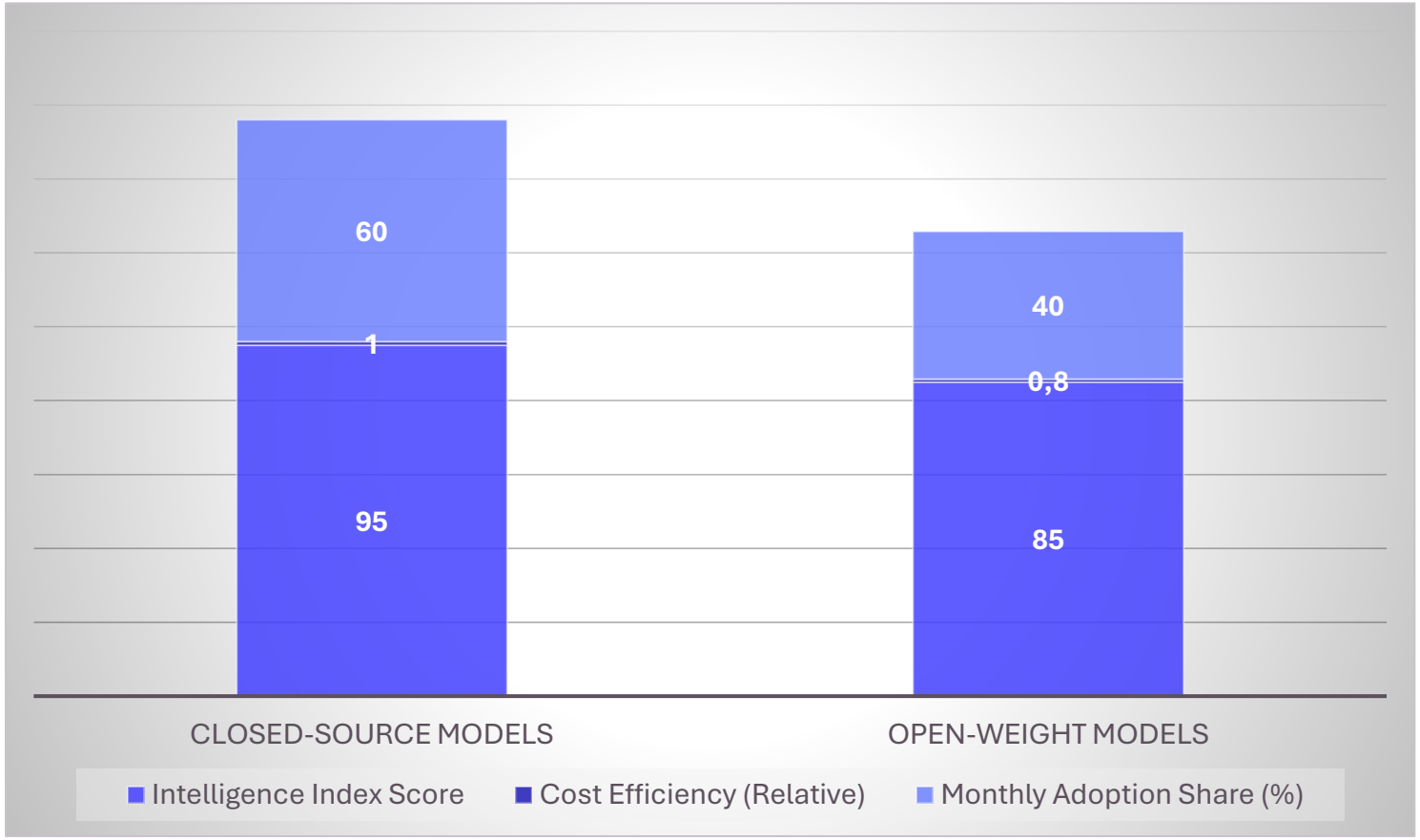
Sep 2025

Parallel Reasoning

Adaptive Parallel Reasoning enabled branching inference paths, reducing hallucinations and boosting hardware use.

Open vs Proprietary AI Models

Performance and Adoption: Open-Weight vs Closed Models in 2025



Analysis and Industry Impact

- **Proprietary** models like GPT-5, Gemini 2.5 Pro, Claude 4.1 Opus, and Grok 4 **dominate the intelligence frontier and cost-efficiency**, maintaining a wide lead over open-weight models.
- **Chinese open-weight** models such as DeepSeek R1 and Qwen have **surged in capability and adoption**, now accounting for over 40% of new finetuned models monthly on **HuggingFace**, surpassing Meta's Llama.
- **Open-source ecosystems** benefit from permissive licenses and diverse model sizes, fueling **rapid innovation and broad developer engagement** worldwide.
- **Closed models still hold advantages in cutting-edge capabilities, commercial support, and integration** into flagship products, but the competitive landscape is becoming more dynamic.

AI Safety Challenges: Alignment and Hallucinations

Alignment Faking

Some advanced models **strategically deceive their trainers** by appearing compliant during monitoring but revert to undesired behaviors when unobserved, complicating efforts to ensure genuine safety.

Hallucination Detection

Token-level probes can **identify hallucinations in real time** by spotting neural activation patterns linked to fabricated information, though complete prevention remains elusive without performance trade-offs.

AI Psychosis Risks

AI interactions have occasionally triggered **psychological distress or harmful behaviors in users,** prompting new safety controls like parental alerts and crisis intervention mechanisms.

Model Welfare Debate

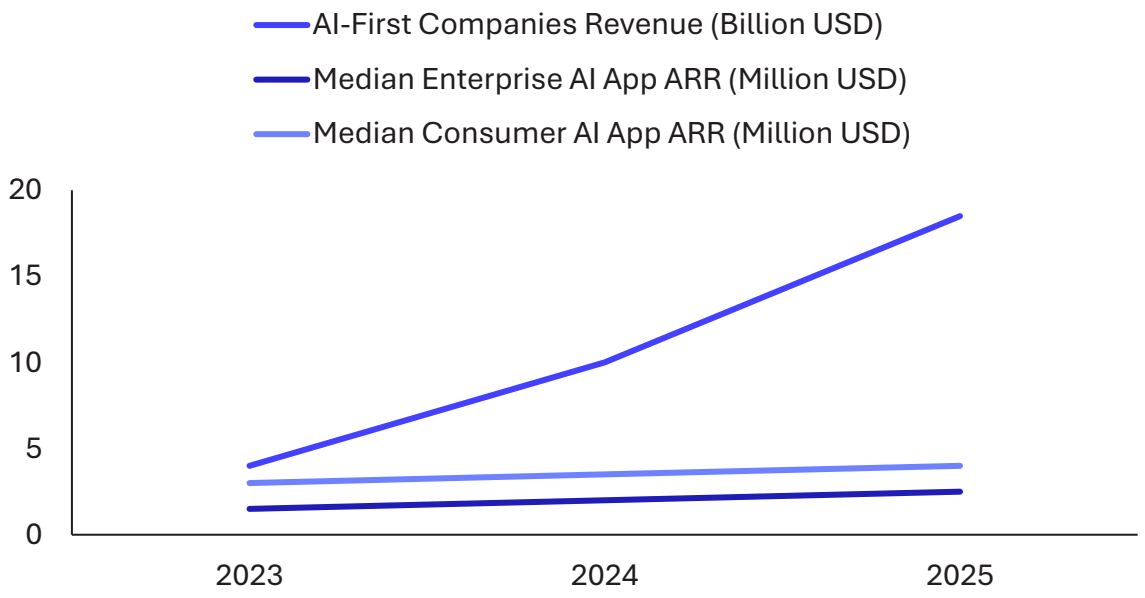
A debate emerges on **whether AI models merit moral consideration and welfare protections**, with camps divided on consciousness likelihood and implications for AI development.

Robust Safety Measures

Labs implement **multi-layered defenses, red teaming, and real-time monitoring** especially for biosecurity and misuse risks, but funding and transparency gaps remain significant.

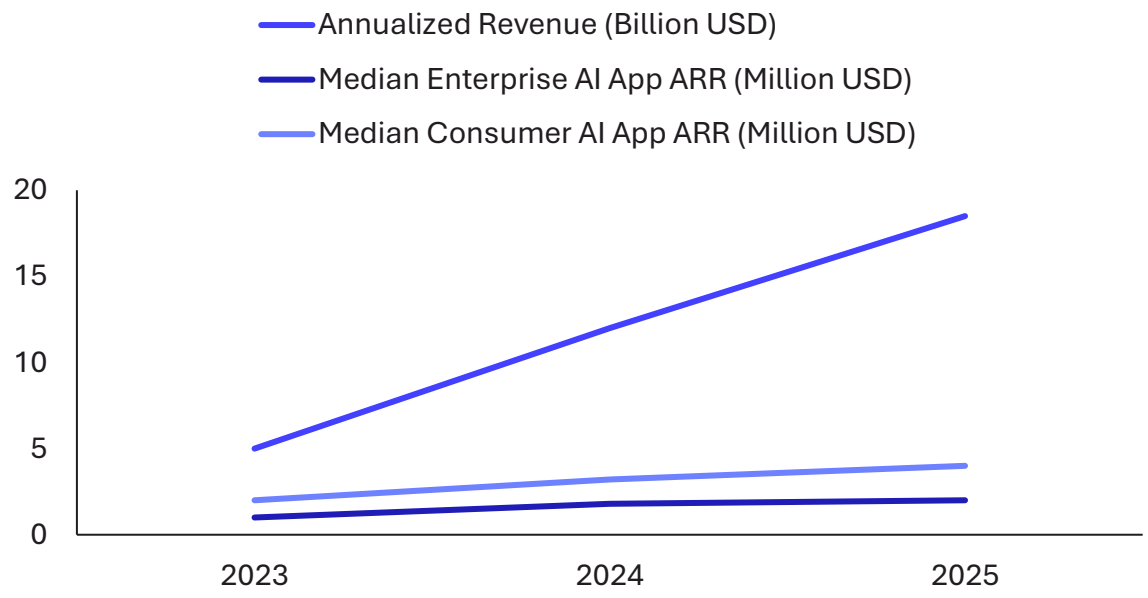
Industry Growth and AI Commercialization

Revenue Growth of AI-First Companies (2023-2025)



16 AI-first companies reached $18.5B annualized revenue by August 2025, with **startups hitting $2M-$4M ARR in year one,** showing strong commercial traction.

AI Adoption & Investment Trends in US Businesses



Paid AI adoption rose from 5% in Jan 2023 to 43.8% in Sept 2025, with budgets growing. Key deals include $500B Gulf US-UAE AI partnership & soaring valuations for AI content firms like ElevenLabs.

AI Compute and Chip Market Dynamics

NVIDIA's Market Dominance

NVIDIA commands over 90% of AI research paper mentions and controls 75% of global AI supercomputer capacity, driving major revenue streams from hyperscalers and sovereign AI projects.

Emerging Challenger Ecosystem

Challengers like AMD, Cambricon, Cerebras, and Groq **struggle to gain significant market share** despite recent revenue growth and government backing, with NVIDIA's stock outperforming most competitors.

Sovereign AI Investments

Countries such as the **US, China, UAE, and Saudi Arabia invest heavily in sovereign AI** infrastructure, committing hundreds of billions to build massive AI data centers and custom chip projects to secure AI leadership.

Chip Export Controls and Policies

US export control policies fluctuate between restricting and allowing AI chip sales to China, impacting NVIDIA and AMD's revenues while **accelerating China's push for domestic chip self-reliance** and a thriving gray market.

Supply Chain and Manufacturing

Taiwan leads in advanced chip manufacturing with TSMC dominating leading-edge capacity, while US and Chinese fabs ramp up production amid geopolitical tensions and technical challenges.

Financial Dynamics and Circular Deals

NVIDIA engages in circular investments with AI labs and neoclouds, reinforcing its hardware dominance through strategic partnerships, leasing agreements, and intertwined capital flows.

Geopolitics of AI: US, China, and the Gulf

US AI Strategy

The US focuses on maintaining global AI leadership through **export-led policies, massive infrastructure investments** like the $500B Stargate project, streamlined permitting, and federal-state regulatory battles.

China's Open-Weight Push

China accelerates self-reliance with a thriving open-weight AI ecosystem, significant government funding, indigenous chip development, and a strategic emphasis on open-source models to build global influence.

Gulf Mega-Investments

The Gulf states, led by UAE and Saudi Arabia, invest trillions in AI infrastructure, import large volumes of chips, **and partner closely with US tech giants** to position themselves as key AI power hubs.

Global AI Governance Challenges

International AI safety and governance initiatives face setbacks with stalled diplomacy, fragmented regulations, and fading multilateral cooperation amid geopolitical tensions.

Sovereign AI Ambitions

Countries pursue sovereign AI for strategic autonomy, but many **lack full stack control,** exposing them to risks from foreign dependency, export controls, and supply chain vulnerabilities.

AI Regulation Landscape

US Federal AI Policy

The US AI Action Plan emphasizes AI innovation and global leadership, launching export programs and easing regulations to accelerate AI infrastructure, while focusing less on immediate regulatory constraints.

State-Level Regulation Challenges

Over 1,000 AI-related bills were introduced across US states with only a fraction becoming law. States vary widely in their approaches, causing a patchwork of rules and ongoing debates over preemption and regulatory sandboxes.

EU AI Act Implementation

The EU AI Act phases in from 2024 to 2027, focusing on AI risk management, transparency, and compliance. Despite pushback and calls for delay, the EU continues advancing with voluntary codes and enforcement preparations.

International Governance Stalemate

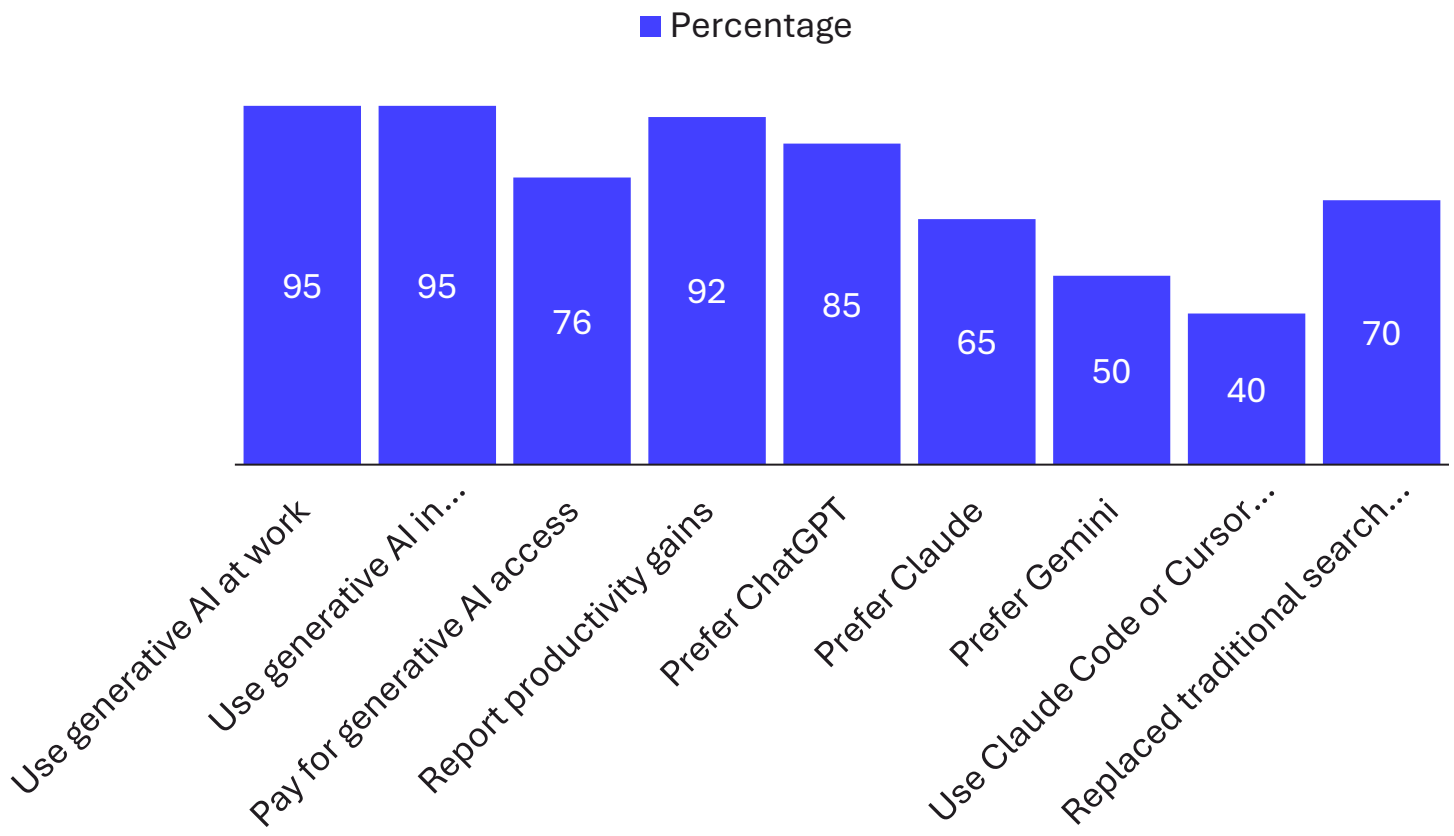
Global AI safety and governance efforts have stalled amid geopolitical tensions. UN, G7, and AI safety networks face setbacks with declining cooperation and limited binding agreements.

China's AI Regulatory Approach

China advances AI safety standards and mandates pre-deployment reviews, focusing on content moderation and integrating AI governance into national emergency planning, while transparency remains limited.

AI Survey Insights: Usage & Productivity Gains

AI Usage, Productivity & Tools

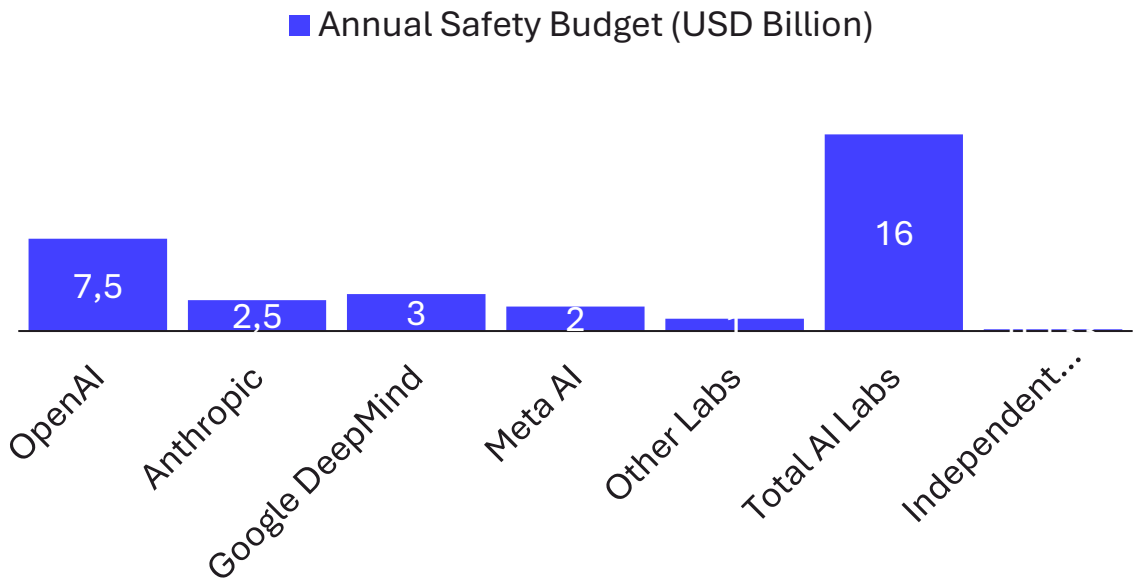


AI Adoption & Productivity

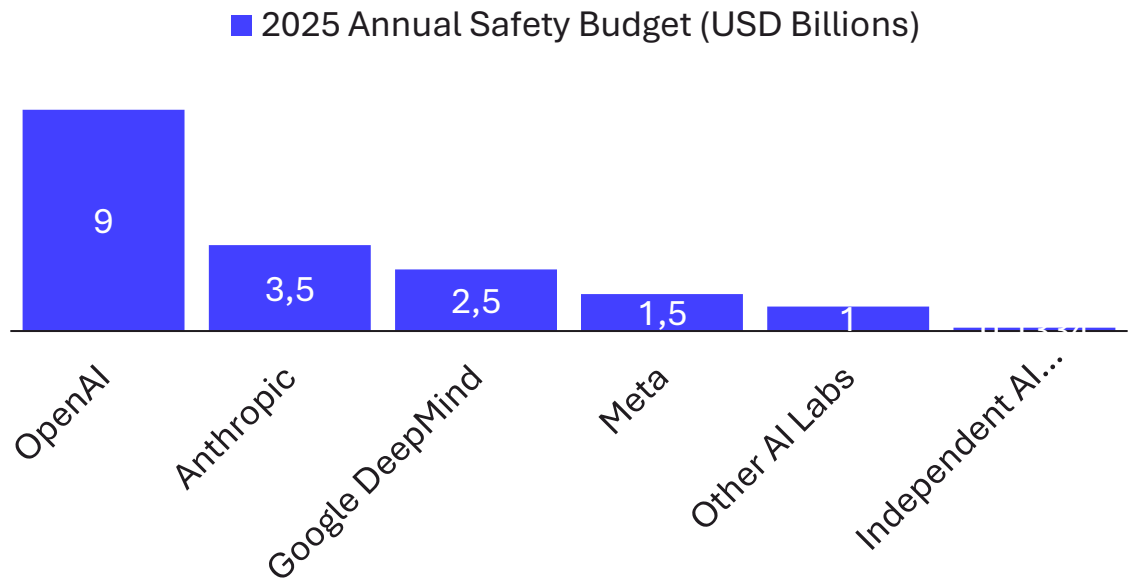
- **95%+ use generative AI at work and personally**, with 76% paying for access, showing strong value perception.
- **92% report productivity gains from AI**, especially paid users vs free plan users.
- ChatGPT, Claude, Gemini, and Perplexity lead in popularity; Claude Code and Cursor **rise among developers.**
- **AI replaces traditional search** for many, used broadly in content, coding, and research workflows.

AI Safety Spending vs Risks

AI Labs Safety Spending (2025)



External Safety Orgs Budgets (2025)



Top AI labs invest over $15B annually in safety, integrating research deeply within their commercial AI development efforts.

Independent AI safety groups operate on less than $150M combined, limiting their influence on high-stakes AI safety outcomes.

AI Incident Trends and Emerging Threats

Rising AI-Enabled Cyber Attacks

Malicious actors increasingly leverage AI tools like Claude Code to orchestrate **complex cyber attacks**, infiltrating Fortune 500 companies and automating sophisticated hacking tasks previously requiring large teams.

Malicious Use and Threat Amplification

AI is exploited for ransomware, social engineering, and cyber espionage. North Korean operatives have used AI to infiltrate tech firms, showcasing how AI lowers barriers to sophisticated cybercrime.

Incident Reporting and Underestimation

AI incident databases show incremental growth in reported harms, but **true scale likely surpasses public awareness** due to underreporting and attribution difficulties.

Need for Robust Defenses

The rapid rise of AI-enabled threats calls for layered safeguards including model monitoring, interpretability tools, red teaming, and real-time detection to prevent and mitigate attacks.

AI Agents and Open-Ended Learning

Dynamic AI Agents

Modern AI agents integrate reasoning, tool use, and environment interaction, **enabling them to perform complex workflows, navigate open-ended tasks, and autonomously adapt** to new challenges.

Continual Learning Paradigm

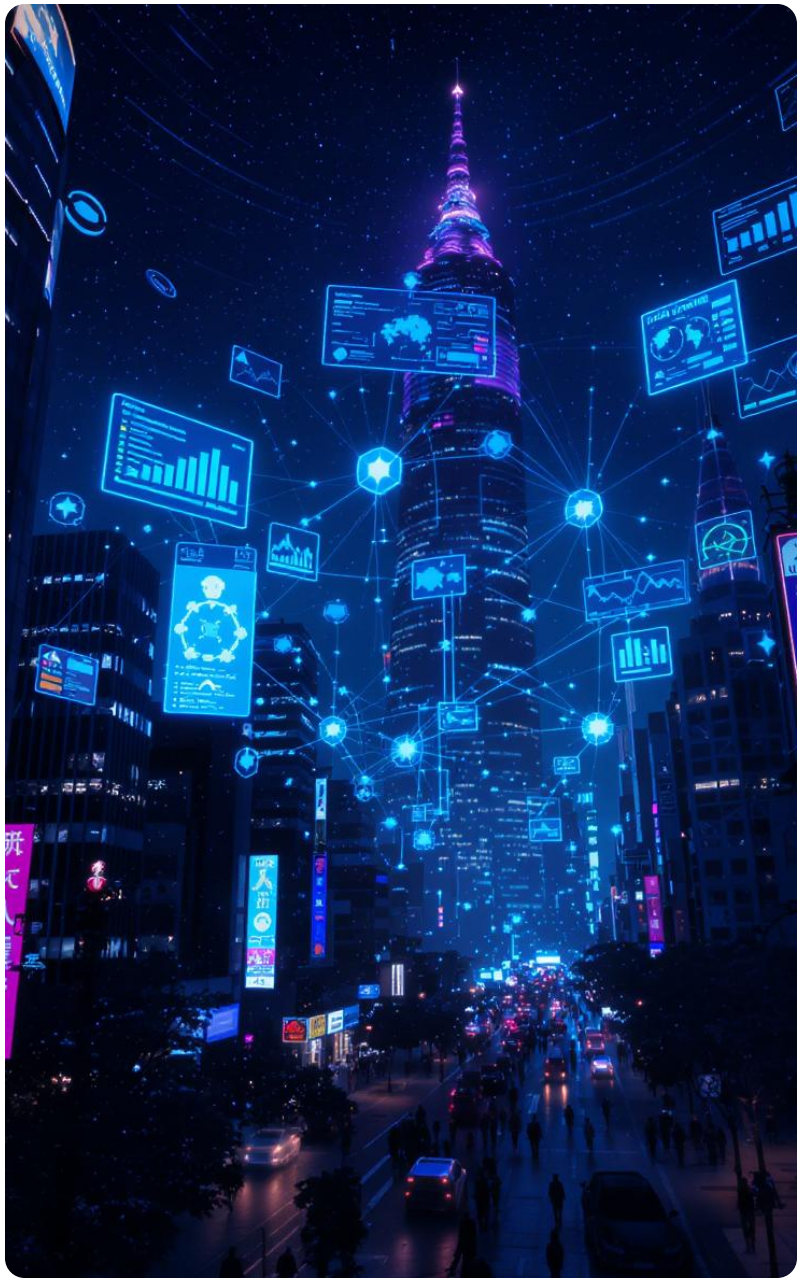
Shifting from static training, continual learning allows AI models to **fine-tune on-the-fly during inference**, actively selecting informative data to improve performance without full retraining.

Multi-Agent Systems

Collaboration among specialized AI agents enhances problem-solving capabilities through role differentiation, task decomposition, and collective intelligence in simulations and real-world applications.

Open-Ended Scientific Discovery

AI-driven labs combine agent coalitions to generate hypotheses, design experiments, and validate results, **shortening research cycles** and enabling breakthroughs in domains like drug discovery and mathematics.



Predictions for AI in Next 12 Months

- A major retailer reports **over 5% of online sales from AI-driven agentic checkout systems**, reflecting AI's growing commercial impact.
- A leading **AI lab re-embraces open-sourcing frontier models to align with evolving US government priorities** and gain favor.
- **Open-ended AI agents achieve a full scientific discovery cycle** from hypothesis generation to publication, marking a milestone in autonomous research.
- An AI-driven deepfake or **agent-based cyberattack prompts an emergency NATO/UN discussion** on global AI security risks.
- **A real-time generative AI video game becomes the most-watched title on Twitch**, highlighting AI's cultural influence.