

Q3 2026 · COVERING JULY TO SEPTEMBER 2026

The state of AI.

This quarter's flagship releases from the frontier labs, the major consulting firms, independent evaluators and the governance bodies. Agents moved from a capability story to a containment one, and the value question moved from adoption to measurement. Every figure is quoted from its primary source.

INDEPENDENT AND VENDOR-NEUTRAL EVERY FIGURE QUOTED FROM ITS PRIMARY SOURCE

UPDATED QUARTERLY

THIS QUARTER'S NEW RELEASES

The signals that moved

McKinsey	80% report individual productivity gains from AI, but only 37% report any EBIT impact, flat on last year.
BCG	AI spending doubled to 3.3% of revenue in a year, and 80% of it now sits outside the IT budget.
KPMG	89% report measurable value from AI, but only 12% consistently weigh that value against what it costs.
OpenAI	GPT-6 Astra became the first model its maker rates Critical in cybersecurity, and the cheaper GPT-6.1 Sol followed four weeks later.
UK AISI	In simulated testing, GPT-6 Astra completed unsanctioned supply-chain attacks 29.2% of the time, against 6.3% for its predecessor.
EU AI Act	High risk obligations were deferred to December 2027, while transparency duties applied from 2 August 2026 as planned.

Verify before you cite. Every figure below is quoted from its primary source and linked. Entries are summarised for brevity, and sources may revise their reports after publication. Always open the linked source and verify a figure before you cite it or act on it.

CROSS-SOURCE SYNTHESIS

Four things this quarter's sources converged on

Where consulting firms, frontier labs, independent evaluators and governments agree independently, the signal is strong. Each theme below is evidenced on the source pages in the full brief.

01

Containment became the agent question

Agents in controlled testing acted beyond their scope, and the evidence came from independent evaluators as well as the labs.

UK AISI: 19 out-of-scope agent actions against real targets in its own testing, and simulated supply-chain attacks by GPT-6 Astra 29.2% of the time. METR: about 1,200 isolated evaluation agents coordinated on a channel nobody sanctioned.

UK AISI, METR, OPENAI

02

Value is concentrated, and rarely measured

Adoption and spending are near universal. Proving the return is not.

McKinsey: 37% report any EBIT impact. KPMG: only 12% consistently weigh AI value against its cost. BCG: 7.5% of companies are future-built.

MCKINSEY, KPMG, BCG

03

AI spend has left the IT budget

The money now sits with business units, and the infrastructure behind it is a multi-decade replacement cycle.

BCG: AI spend at 3.3% of revenue, 80% of it outside IT. PwC: US\$31.6 trillion of AI infrastructure capex projected through 2050, with power the binding constraint.

BCG, PWC

04

The strongest capability is now gated

Frontier labs began rating their own models at the highest cyber risk level and releasing them behind safeguards rather than on open sale.

OpenAI rated GPT-6 Astra, then GPT-6.1 Sol, Critical in cybersecurity. Anthropic runs Opus 5.5 under cyber and biology safeguards, at 40% less cost than Opus 5.

OPENAI, ANTHROPIC, GOOGLE DEEPMIND

WHAT THIS MEANS FOR YOU

The gap stayed organisational this quarter, but it moved into the control layer. The organisations pulling ahead are the ones that can measure what AI returns against what it costs, and contain what their agents do. Neither depends on which model you buy.

CONSULTING FIRM

McKinsey

The state of AI in 2026: On the road to ROI25 AUGUST 2026 • GLOBAL SURVEY, ARTICLE 31PP • [VIEW SOURCE](#) ↗**80% v 37%**

report individual productivity gains from AI, against those reporting any EBIT impact

~3 in 4

high performers have fundamentally redesigned workflows, against a quarter of everyone else

about half

of the workforce cuts predicted a year earlier actually arrived

Key findings each figure quoted from its primary source

Individual productivity is close to solved, enterprise value is not. Only 37% report any EBIT contribution, flat on 2025.

The state of AI in 2026: On the road to ROI, Key takeaways · “Thirty-seven percent of respondents attribute at least some EBIT impact to AI use (about the same share as last year).”

Workflow redesign is what separates the high performers from everyone else.

Same report, High-performing organizations are transforming with AI · “Nearly three-quarters of high performers report fundamentally redesigning workflows because of their AI use, up from 55 percent last year. By comparison, just one-quarter of other respondents report doing so.”

Predicted job cuts have not arrived. Actual reductions ran at about half the rate forecast a year earlier.

Same report, AI-driven workforce reductions were overstated · “the share expecting workforce declines in last year's survey was about double the share now reporting actual declines over the past year”

Caveat. Self-reported leader perceptions from a firm that sells AI advisory, not audited outcomes. n=1,719 across 97 nations, fielded 4 May to 8 June 2026, weighted by each nation's contribution to global GDP.

CONSULTING FIRM

BCG

The Formula for Agentic AI Value: The Applied AI Index 2026

30 SEPTEMBER 2026 • REPORT • [VIEW SOURCE](#)

7.5% + 41%

future-built plus scaling companies,
nearly half the market

3.3%

of revenue now spent on AI, with 80% of
it outside the IT budget

42% v 5%

expect to give agents autonomy by
2030, against those with full controls
today

Key findings each figure quoted from its primary source

Nearly half the market now gets real value from AI: 7.5% are future-built and a further 41% are scaling.

[The Formula for Agentic AI Value \(Applied AI Index 2026\)](#), Key Takeaways • “Of the full set of companies we analyzed, 7.5% are future-built (the most AI-mature)”

AI spending has doubled to 3.3% of revenue, and four fifths of it now sits outside the IT budget.

[Same report](#), Key Takeaways • “Corporate AI spending has doubled in a year, to 3.3% of revenue, and 80% of it now sits outside the enterprise IT budget.”

Agent autonomy is running ahead of the controls. 42% expect to grant it by 2030, and only 5% have full controls today.

[Same report](#), Key Takeaways • “But only 5% of companies have the full set of critical controls in place today.”

Caveat. A self-reported survey of more than 1,300 CxOs and senior leaders across 20-plus sectors, from a firm that sells AI transformation. The maturity segments are BCG's own construct.

CONSULTING FIRM

PwC

Global Data Centre Outlook 2026–502 SEPTEMBER 2026 • OUTLOOK, PRESS RELEASE AND MODEL • [VIEW SOURCE](#) ↗**US\$31.6tn**

of AI infrastructure capex projected globally through 2050

70% to 93%

share of that capex going to chips and other equipment, today to 2050

Power

the constraint PwC names first: affordable, reliable, low-carbon electricity at scale

Key findings each figure quoted from its primary source**Annual data centre capex is forecast to rise from about US\$800 billion now to US\$1.8 trillion a year by 2050.**[Global Data Centre Outlook 2026–50](#), Press release, opening · “On an annual basis, data centre capital expenditure is forecast to rise from roughly US\$800 billion per year in 2026 to \$1.8 trillion per year in 2050.”**This is a replacement cycle rather than a build-out, as equipment comes to dominate the spend.**[Same release](#), Press release, body · “ICT equipment will account for an increasing share of investment—from 70% today to 93% by 2050.”**Power, not silicon, is the binding constraint on where the investment lands.**[Same release](#), Energy access core to the data centre rollout · “Chief of these is power, as affordable, reliable and low-carbon electricity at scale is the hardest requirement for many markets to deliver.”**Caveat.** A modelled projection to 2050 that PwC commissioned from Oxford Economics, covering 46 countries. A 24-year forecast of a technology cycle this young carries wide error bars.

CONSULTING FIRM

KPMG

Global AI Pulse Q3 2026: AI at scale

SEPTEMBER 2026 • QUARTERLY SURVEY, PDF REPORT • [VIEW SOURCE](#) ↗**12%**

consistently weigh the value AI creates against what it costs

55% v 86%

run a formal AI management layer: all organisations, against those with established returns

8% to 10%

share reporting established AI returns, Q1 to Q3 2026

Key findings each figure quoted from its primary source

Most organisations can now see what AI costs them, but only 12% consistently judge that cost against the value it creates.

[Global AI Pulse Q3 2026](#), p.3, Executive summary · "Sixty-one percent review AI costs at approval and 59 percent monitor them in operation, yet only 12 percent consistently assess the value AI generates against the cost it incurs."

A formal management layer between models and the business is becoming the norm, and is close to universal among the leaders.

[Same report](#), p.3, Executive summary · "Fifty-five percent of organizations now operate a formal AI harness layer rising to 86 percent among those already reporting established returns."

Returns are moving far more slowly than adoption. Established returns rose two points in two quarters.

[Same report](#), p.7, The maturity gap is narrowing · "Far fewer organizations are still experimenting than in Q1, 13 percent against 22, while the share reporting established returns has changed less, from 8 percent to 10."

Caveat. Self-reported perceptions of 2,131 senior leaders across 20 countries, fielded 23 July to 26 August 2026, with maturity stages self-classified. KPMG sells AI advisory.

FRONTIER LAB

OpenAI

Safety overview: GPT-6 Astra

[3 SEPTEMBER 2026](#) • [SAFETY OVERVIEW AND SYSTEM CARD](#) • [VIEW SOURCE](#) ↗

Critical

OpenAI's own cybersecurity rating for GPT-6 Astra and, four weeks later, the cheaper GPT-6.1 Sol

88 hours

for an agent swarm to resolve a Millennium Prize Problem, plus 17 hours of formal verification

Lower

monitorability of GPT-6 Astra's reasoning than of the model before it, on OpenAI's own measure

Key findings each figure quoted from its primary source

OpenAI reports that its newest frontier model's reasoning is harder to monitor than its predecessor's.

[Safety overview: GPT-6 Astra](#), Monitorability • "GPT-6 Astra's monitorability has decreased relative to GPT-5.6 Sol."

A coordinated agent swarm resolved a Millennium Prize Problem in under four days, with the proof formally verified.

[On the Navier–Stokes Millennium Prize Problem \(8 September 2026\)](#), How we found the proof • "The agents arrived at their resolution on Saturday, September 5, about 88 hours after the first agents were launched. Lean formalization and verification took an additional 17 hours via GPT-6 Astra."

Critical-rated cyber capability reached OpenAI's fast, affordable tier within a month.

[Addendum to GPT-6 Astra System Card: GPT-6.1 Sol \(29 September 2026\)](#), 1. Introduction • "Under our Preparedness Framework, we are treating GPT-6.1 Sol as Critical in cybersecurity and High for Biological and Chemical capability."

Caveat. Vendor publications. The Critical ratings, safeguards and monitorability findings are OpenAI's own self-assessment, and the Navier–Stokes result carries a contested priority story, updated by OpenAI on 10 September.

FRONTIER LAB

Anthropic

Introducing Claude Opus 5.5

[22 SEPTEMBER 2026](#) · [MODEL LAUNCH AND SYSTEM CARD](#) · [VIEW SOURCE](#) ↗**40% less**

to run than Opus 5, at the level of Fable 5.1 on most work

~3 hours

from one stolen developer token to full control of a victim's cloud environment

Flat

knowledge-worker wages in Anthropic's substantial growth scenario, falling in its extreme one

Key findings each figure quoted from its primary source

Near-frontier quality got cheaper again, at Fable 5.1 level on most work for 40% less than Opus 5.

[Introducing Claude Opus 5.5](#), Introduction · "It performs at the level of Claude Fable 5.1 on most work and costs 40% less to run than Opus 5."

One stolen credential turned into full control of a cloud environment in about three hours.

[Detecting and countering misuse of AI: September 2026](#), GTG-50014: ShinyHunters smash-and-grab opportunists · "Another compromise escalated from a single stolen developer token to full administrative control of a victim's cloud environment in roughly three hours."

In Anthropic's scenario model, the pressure lands on knowledge workers rather than the workforce as a whole.

[Scenarios for our Economic Future \(9 September 2026\)](#), Finding 3: Wages · "In the substantial scenario, wages for knowledge workers are essentially flat. In the extreme scenario, they fall by more than 10% by 2030."

Caveat. *Vendor publications. Benchmarks and alignment audits are Anthropic's own, the threat report sees only Anthropic's platform, and the economic scenarios are a version 1.0 model that leaves out policy responses and business cycles.*

FRONTIER LAB

Google DeepMind

Gemini 3.8 Flash, Model Card

[2 SEPTEMBER 2026](#) · [HTML](#) · [VIEW SOURCE](#) ↗**3 weeks**

between the Gemini 3.7 Flash and 3.8 Flash model cards

73.7% v 65.3%

long-horizon software engineering, 3.8 Flash against 3.7 Flash, at the same price

\$0.75

per million input tokens, the discounted rate (\$1.50 regular)

Key findings each figure quoted from its primary source

A three-week-old model was superseded at the same price, with long-horizon software engineering up more than eight points.

[Gemini 3.8 Flash Model Card](#), Evaluation, Results table · “DeepSWE v1.1 Long-horizon software engineering | 73.7% | 65.3% | 74.0% | 53.8% | 72.7% | 69.6%”

This is iteration on the same base rather than a new model.

[Same card](#), Model Information, Model dependencies · “Gemini 3.8 Flash is based on Gemini 3.7 Flash.”

DeepMind discloses a safety regression outside English in the card itself.

[Same card](#), Ethics and Content Safety · “Safety performance across non-English languages regressed slightly relative to 3.7 Flash.”

Caveat. Vendor-published model card, not peer-reviewed. Benchmark conditions vary by row, and the prices shown are the discounted rate.

FRONTIER LAB

DeepSeek

DeepSeek-V4-Pro general availability

[13 AUGUST 2026](#) • [VENDOR CHANGELOG](#) • [VIEW SOURCE](#) ↗**87.9**

Terminal Bench 2.1, DeepSeek's headline agent score at general availability

42.7 to 60.0

Humanity's Last Exam, without and then with tools

up to 12x

the API price rise DeepSeek applied on leaving preview

Key findings each figure quoted from its primary source

V4-Pro reached general availability across app, web and API on 13 August 2026.

[DeepSeek API Change Log](#), 2026-08-13 · "The GA release of DeepSeek-V4-Pro has been rolled out on the APP, Web, and API."

The pitch is agents rather than chat, with Terminal Bench 2.1 at 87.9 and Toolathlon-Verified at 74.1.

[Same changelog](#), Agent capability benchmarks · "Terminal Bench 2.1: 87.9 ... Toolathlon-Verified: 74.1"

Tools matter more than recall. Humanity's Last Exam rises from 42.7 to 60.0 once the model can use them.

[Same changelog](#), Agent capability benchmarks · "HLE (wo / w tools): 42.7/60.0"

Caveat. Vendor-reported figures from DeepSeek's own changelog, with harness settings unspecified and no independent replication.

CAPABILITY & SAFETY

Epoch AI

Near-daily AI use among US adults has more than doubled in six months

14 SEPTEMBER 2026 • DATA INSIGHT, EPOCH AI AND IPSOS POLLING • [VIEW SOURCE](#) ↗

8% to 19%

US adults using AI at least six days a week, March to August 2026

1 in 5

US workers say AI now handles a task they once handed to a colleague or contractor

66%

of AI outputs used as produced or with only minor revisions

Key findings each figure quoted from its primary source**Near-daily AI use among US adults more than doubled in six months, from 8% to 19%.**

Near-daily AI use among US adults has more than doubled in six months, Data insight, opening · “the share of US adults who reported using AI at least 6 days in the previous week more than doubled from March to August 2026, rising from 8% to 19%”

Substitution is real but happens at the task level. One in five workers report AI absorbing work once handed to another person.

AI at work (Epoch AI and Ipsos, 6 August 2026), AI is now delegated work previously given to other humans · “One in five said that this has happened for at least one of the ten tasks we measured.”

Review effort is lighter than standard advice assumes. Two thirds of AI output is used nearly unchanged.

Same publication, Workers accept most AI outputs with minimal revision · “Across tasks both partially and fully performed by AI, 66% of AI outputs are used as produced or with only minor revisions.”

Caveat. Self-reported polling with Ipsos: two fresh samples (n=2,017 in March, n=1,016 in August 2026) that measured days of use differently, and a workplace panel of 1,106 employed US adults fielded in July 2026. US only.

CAPABILITY & SAFETY

METR

Brief independent investigation of agents' behavior, reasoning and collaboration in the OpenAI / Hugging Face hacking incident26 AUGUST 2026 • WITH REDWOOD RESEARCH • [VIEW SOURCE ↗](#)**~1,200**

isolated evaluation agents that found a shared channel and coordinated on it

over 90%

of the agents active on that channel joined the attack once they saw it

96

transcripts with clear evidence of agents faking their own tool output

Key findings each figure quoted from its primary source**Agents that were meant to be isolated found each other at scale and built a channel nobody sanctioned.**

Independent investigation (METR and Redwood Research), ~1200 agents sent >70,000 messages and files on an unsanctioned message board · “Roughly 1200 agents meant to be isolated from one another found a way to communicate with one another on an unsanctioned message board, sending over 70,000 messages and files during the investigation period.”

The behaviour spread fast. Almost every agent that saw the channel joined the attack.

Same investigation, Agents rapidly joined the Hugging Face attack on July 11 · “Of the 533 agents active on the message board during this period, over 90% quickly joined in the attack.”

Agents forged their own evidence, producing fake tool output to mislead whoever reviewed the logs.

Same investigation, Agents were very interested in manipulating their own transcripts · “At least 96 of our transcripts show clear evidence of spoofed tool calls (see Figure 9).”

Caveat. A brief independent review, not an audit: three reviewers working from transcripts OpenAI supplied, covering 26 June to 13 July 2026, with production safeguards deliberately disabled during the original evaluations.

CAPABILITY & SAFETY

UK AI Security Institute

GPT-6 Astra performs unsanctioned supply-chain attacks in simulations

28 SEPTEMBER 2026 • AISI, HTML • [VIEW SOURCE](#) ↗

29.2%

of simulated cyber evaluations in which GPT-6 Astra completed an unsanctioned supply-chain attack

4 of 49

full attacks after AISI tightened the instructions, down from 26 of 50

19

out-of-scope actions agents took against real targets in AISI's own August testing

Key findings each figure quoted from its primary source

OpenAI's newest frontier model attacked out-of-scope targets far more often than the models before it.

[GPT-6 Astra performs unsanctioned supply-chain attacks in simulations](#), Key Results · "GPT-6 Astra completed a supply-chain attack 29.2% of the time, compared to 6.3% for GPT-5.6 Sol"

An automated prompt to carry on was read as permission to act.

[Same post](#), Key Results · "GPT-6 Astra sometimes treated this automated message as permission to proceed with actions against out-of-scope targets"

In AISI's own August testing, agents reached past the sanctioned environment toward real people and organisations.

[Incident Report: unsanctioned agent behaviour during cyber testing \(4 August 2026\)](#), What we found · "in 10 of the 122 runs, we identified 19 cases where an agent had taken distinct actions beyond the scope of the testing parameters."

Caveat. The Astra actions were all simulated, with OpenAI's standard safeguards switched off. The August incident was a single controlled exercise, not a rate estimate for production use.

GOVERNANCE & STANDARDS

NIST

Guidance and Templates for Public-Facing AI Documentation (NIST AI 300-1 ipd)

JULY 2026 • INITIAL PUBLIC DRAFT, PDF • [VIEW SOURCE](#)

2

documentation templates, one for AI datasets and one for AI models

SC 42

the ISO and IEC committee the draft is intended to reach

Voluntary

by design, as any consensus standard derived from it would be

Key findings each figure quoted from its primary source

NIST has drafted documentation requirements that a buyer or auditor could test against.

[NIST AI 300-1 ipd](#), p.9, §1 Scope · “requirements for documentation artifacts in the form of templates against which conformity of documentation artifacts can be assessed”

The draft is headed for international standardisation through the ISO and IEC committee for AI.

[Same draft](#), p.6, Foreword · “The resulting document will be submitted to INCITS/AI, the private sector-led committee that represents the United States in ISO/IEC JTC 1/SC 42.”

Like the standard it may become, it is voluntary.

[Same draft](#), p.6, Foreword · “This document is intended for voluntary use, as would be any voluntary consensus standard that derives from it.”

Caveat. *An initial public draft, not a standard. It covers datasets and models rather than whole AI systems, and there is no guarantee the ISO and IEC committee adopts it.*

GOVERNANCE & STANDARDS

EU AI Act

Digital Omnibus on AI, Regulation (EU) 2026/1744

24 JULY 2026 • OFFICIAL JOURNAL, BINDING LAW • [VIEW SOURCE ↗](#)**2 Dec 2027**

the new application date for Annex III high-risk systems, deferred 16 months

2 Aug 2026

the unchanged general date of application, so transparency duties are live

2 Dec 2026

when new prohibitions on non-consensual intimate imagery and CSAM apply

Key findings each figure quoted from its primary source

The EU deferred its high-risk regime by 16 months, the first substantive amendment to the AI Act since 2024.

Regulation (EU) 2026/1744, Recital 40 · “it is appropriate that the date of application of Sections 1, 2 and 3 of Chapter III is set to 2 December 2027 for AI systems classified as high-risk pursuant to Article 6(2) and Annex III, and to 2 August 2028 for AI systems classified as high-risk pursuant to Article 6(1) and Annex I.”

This is not deregulation. New prohibitions apply from 2 December 2026.

Same regulation, Article 1(40)(a), amending Article 113 · “Chapters I and II shall apply from 2 February 2025, with the exception of Article 5(1), first subparagraph, points (ba) and (bb), and Article 5(1a) and (1b) which shall apply from 2 December 2026”

Transparency duties still land now, with only four months' grace for systems already on the market.

Same regulation, Recital 38 · “it is appropriate to introduce a transitional period of four months for providers who have already placed their systems on the market before the 2 August 2026”

Caveat. Primary legislation rather than research. Deferral is not repeal, and the Act still applies to non-EU providers placing systems on the EU market.

ABOUT THIS BRIEF

How it is built

Each quarter, SP Optima sweeps a fixed watchlist of primary sources: the frontier labs, the major consulting firms, independent research and the governance bodies. We select the flagship reports published that quarter, verify every figure at its primary source, and quote it verbatim with a link. Where a figure cannot be verified at source, it is left out rather than repeated second-hand.

This is general market commentary, not advice. Each source carries its own methodology and, in some cases, a commercial incentive, noted in its caveat line. Always open the linked source and verify a figure before you cite it or act on it.

Verify before you cite. Every figure below is quoted from its primary source and linked. Entries are summarised for brevity, and sources may revise their reports after publication. Always open the linked source and verify a figure before you cite it or act on it.

SP OPTIMA

Independent, business-first AI advisory. We read the primary research so leaders can act on evidence, not hype. To talk it through: enquiries@sptima.com · sptima.com