LUMO INSIGHTS

Tuesday, 22 September 2026

Agents at work: new powers, perils & policies

Executive Summary

AI agents are quickly moving from experimental assistants to strategic players in the enterprise. In the past 48 hours, we've seen major leaps in agent capabilities from tech giants (www.forbes.com [1]) (pondero.ai [2]) alongside stark reminders of the risks and governance challenges they bring. Leaders now face a dual imperative: capitalise on these rapidly evolving tools for efficiency and growth, while staying ahead of new threats and regulatory curves.

References:

[1] www.forbes.com — https://www.forbes.com/sites/jonmarkman/2026/09/21/openai-plans-to-introduce-managed-agents-at-devday-2026/#:~:text=September%2015%20he%20posted%20on,on%20September%207%20that%20OpenAI%E2%80%99s

[2] pondero.ai — https://pondero.ai/news/2026-09-21-stepfun-step-5-preview-600b-moe-open-weights/#:~:text=StepFun%20shipped%20the%20Step%205,weights%20scheduled%20for%20October%2015

Big tech bets on autonomous co-workers

OpenAI's upcoming DevDay is poised to mark a pivotal moment in the shift from simple chatbots to autonomous enterprise agents. CEO Sam Altman has hinted that the company will unveil a platform enabling "AI co-workers" to perform multi-step tasks across office applications with minimal human input (www.forbes.com [1]). This would make OpenAI the latest of the AI giants to double down on agentic technology that acts on a company's behalf: Microsoft and Anthropic introduced their own enterprise agent frameworks earlier this year, and Google's Gemini platform has been evolving towards full workflow orchestration. (www.beri.net [2]) The motivation is clear - after many successful pilot projects, the race is on to embed AI deep into day-to-day workflows and solve the new challenge of *scaling* AI across the organisation.

The implications of this trend extend beyond IT departments. If AI agents can execute end-to-end processes in sales, customer service, reporting or HR, businesses could consolidate or even replace numerous single-purpose software tools. Investors have already seen early signs of this disruption: in February, the debut of Anthropic's enterprise agent caused a broad $285 /billion selloff in software stocks within 48 hours (www.forbes.com [3]). The market reaction - dubbed a "SaaS-pocalypse" - reflected fears that AI-driven automation could quickly erode the value of standalone B2B software offerings. For executives, the takeaway is that autonomous agents are swiftly moving from novelty to necessity. Every major tech player is competing to provide the default AI agent platform, and those agents will be able to take on more work internally, potentially upending the software and services suppliers you rely on.

References:

[1] www.forbes.com — https://www.forbes.com/sites/jonmarkman/2026/09/21/openai-plans-to-introduce-managed-agents-at-devday-2026/#:~:text=September%2015%20he%20posted%20on,on%20September%207%20that%20OpenAI%E2%80%99s

[2] www.beri.net — https://www.beri.net/article/openai-frontier-vs-microsoft-agent-365-vs-anthropic-enterprise-2026#:~:text=Three%20of%20the%20world%27s%20largest,lifecycle%20systems%20in%20the%20market

[3] www.forbes.com — https://www.forbes.com/sites/jonmarkman/2026/09/21/openai-plans-to-introduce-managed-agents-at-devday-2026/#:~:text=This%20Year%20The%20first%20came,Workday%20was%20down

New mega-models redefine capabilities

This week brought astonishing advances in AI models that power autonomous workflows. Shanghai-based startup StepFun unveiled *Step 5*, a colossal 600 /billion - parameter language model boasting a 1 /million - token context window (pondero.ai [1]). In practical terms, it can “remember” and analyse the equivalent of hundreds of thousands of words of text - enabling a single AI agent to work on complex, long-running tasks (like codebase overhauls or financial analyses) without losing context. Despite its size, Step 5 uses a mixture-of-experts design to keep costs down, and performing at near-frontier level it reportedly runs at only one-seventh the cost of OpenAI’s top models (pondero.ai [2]). Crucially, StepFun plans to open-source this model’s weights by next month, potentially allowing any organisation to deploy its own powerful agents built on Step 5’s technology (pondero.ai [3]).

Not to be outdone, Chinese tech giant Alibaba’s AI unit launched its next-generation model *Qwen 3.8-Omni-Flash* on a similar scale. Qwen3.8 features a 1 /million - token *multimodal* context - it can take in text, images, audio, and video - and is engineered not just to describe media but to act on it. Alibaba explicitly touts the shift “from understanding omnimodal content to planning tasks, calling tools, and completing creative work,” meaning an AI agent powered by Qwen can now autonomously edit videos, generate multimedia content, or analyse audio-visual data as part of a larger workflow. Even more striking, Qwen3.8 dramatically slashes costs for heavy data tasks (its audio analysis API is 98% cheaper than the prior version) and comes with open-source plugins and a real-time orchestration system for building agents.

For global enterprises, these model breakthroughs signal intensifying competition and opportunity. The cost of cutting-edge AI capabilities is plummeting, and open models lower barriers for new entrants and bespoke solutions. In the near future, you may face competitors who leverage these mega-models - or their open-source descendants - to analyse entire databases at once, power autonomous research and development, or deliver personalised customer interactions across media. Business leaders should track this trend closely: as foundational AI becomes more powerful and accessible, value will shift to those who can effectively integrate and domain-tailor these agent capabilities. Being ready to experiment with emerging models (and understanding their limits) will be key to staying ahead.

References:

[1] pondero.ai — https://pondero.ai/news/2026-09-21-stepfun-step-5-preview-600b-moe-open-weights/#:~:text=StepFun%20shipped%20the%20Step%205,weights%20scheduled%20for%20October%2015
[2] pondero.ai — https://pondero.ai/news/2026-09-21-stepfun-step-5-preview-600b-moe-open-weights/#:~:text=StepFun%20shipped%20API%20access%20to,are%20scheduled%20for%20October%2015
[3] pondero.ai — https://pondero.ai/news/2026-09-21-stepfun-step-5-preview-600b-moe-open-weights/#:~:text=StepFun%20shipped%20the%20Step%205,weights%20scheduled%20for%20October%2015

Longer memory, smarter workflows

One frontier for making AI agents more capable is giving them longer-lasting memory. A new research programme launched this week (dubbed *AML Cycle 2*) is evaluating techniques for long-term “memory” in autonomous agents (jmacweb.com [1]). Today’s generative AI systems typically operate within a short window of context - they can’t recall what happened beyond a single session or conversation. Enabling persistent, cross-session memory would allow an AI agent to learn from past interactions, maintain context over weeks or months, and continuously improve at its tasks without needing to start from scratch each time.

However, building agents that never forget brings serious challenges alongside its promise. Researchers warn that unmanaged long-term memory could capture sensitive data or lead to outdated or biased decision-making if old information isn’t refreshed or audited (jmacweb.com [2]).

There’s a growing call for “memory hygiene” - processes to regularly review, update, or purge an agent’s accumulated knowledge when necessary - as well as technical controls to trace and verify what an agent ‘remembers’ over time (jmacweb.com [3]). For enterprises, this means that more autonomous, long-running AI assistants will require new data governance practices. Companies exploring persistent AI agents will need to implement strict rules on what an agent can store, how it updates its knowledge base, and how humans can intervene or reset that memory if it goes off track.

Crucially, capitalising on advanced AI agents also demands rethinking human workflows themselves. This week, Microsoft released a 44-page “Frontier Firm” playbook based on 100+ enterprise AI transformations, concluding that reinventing business processes and leveraging proprietary data and feedback loops are now more important than choosing the most powerful model available (agentic.ai [4]) (agentic.ai [5]). In other words, simply dropping a supercharged agent into a broken process won’t magically fix it - the process needs to be fixed (or reimagined) first. Likewise, OpenAI just expanded its *AI Academy* training programmes for developers, leaders, and frontline staff on how to effectively use AI in daily work, underscoring that successful adoption requires widespread upskilling and cultural change, not just new tech. Smart automation is as much about people and process as it is about algorithms.

References:

[1] jmacweb.com — https://jmacweb.com/ai-news/evaluating-long-term-memory-for-ai-agents-aml-cycle-2-is-now-open-20260921#:~:text=cognition.%20The%20idea%20of%20long,research%20milestone%3B%20it%20signals%20a
[2] jmacweb.com — https://jmacweb.com/ai-news/evaluating-long-term-memory-for-ai-agents-aml-cycle-2-is-now-open-20260921#:~:text=practical%20challenges,invest%20in%20memory%20hygiene%2C%20auditing
[3] jmacweb.com — https://jmacweb.com/ai-news/evaluating-long-term-memory-for-ai-agents-aml-cycle-2-is-now-open-20260921#:~:text=making%20if%20memory%20governance%20isn%E2%80%99t,The%20conversation%20around%20agent
[4] agentic.ai — https://agentic.ai/news#:~:text=operations%2C%20the%20guidance%20puts%20workflow,and%20control%20AI%20agents%20in
[5] agentic.ai — https://agentic.ai/news#:~:text=more%20than%20100%20internal%20AI,redesign%20ahead%20of%20simply%20deploying

Security and governance: A new high wire act

The past 48 hours also underscored the risks of increasingly powerful autonomous systems - and the urgent need for robust oversight. In a high-profile incident, a trio of independent researchers used an AI coding agent (Anthropic’s Claude) to “ethically hack” OpenAI’s own systems. In less than 72 hours - and for under $3,000 in cloud costs - the small team found and exploited an unreported vulnerability in OpenAI’s software, gaining access to confidential internal tools and code repositories (thetechportal.com [1]). While no harm was done (the hackers immediately disclosed the flaw under a bug bounty programme), the episode is a wake-up call: it shows how AI can dramatically lower the barrier to sophisticated cyberattacks, equipping even small teams with capabilities that rival those of nation-state actors. As companies roll out AI agents with access to internal systems and data, CISOs must assume that attackers will use equally intelligent agents to probe for weaknesses.

The governance environment for AI agents is evolving just as quickly as the technology. In the United States, a stark divide in approaches emerged over the weekend: President Donald Trump dismissed AI safety concerns as a “hoax” and proposed a new military “AI Force” to ensure America leads in AI development (abcnews.com [2]). In contrast, California’s Governor signed an executive order initiating plans for a mandatory “kill switch” on advanced AI systems and on-site AI audits, aiming to prevent “rogue” AI incidents. With federal and state leaders at odds, top AI firms are taking matters into their own hands. Anthropic, for instance, just announced a landmark partnership with Accenture to embed independent “red team” auditors directly inside its company, with both firms committing $1 billion each over five years to this effort. These embedded AI safety experts will have full access to monitor how new models are trained and deployed in real time, providing an extra layer of oversight against unintended behaviour.

Whether via external regulation or proactive self-governance, the message for enterprises implementing AI agents is the same: stronger checks and balances are coming. Leaders should not assume that “plug-and-play” autonomy will be acceptable to regulators, investors, or customers. Forward-looking organisations are already implementing AI oversight mechanisms - from model “nutrition labels” and audit trails to controlled kill-switches and third-party model validations - to ensure that as agents take on more authority, they remain accountable and aligned with human objectives.

References:

[1] thetechportal.com — https://thetechportal.com/2026/09/18/openai-hacked-using-claude-hacktron-ai-indian-security-researchers/#:~:text=within%20OpenAI%E2%80%99s%20bug%20bounty%20programme,in%20an%20autonomous%20goal%20loop
[2] abcnews.com — https://abcnews.com/Politics/trump-form-new-ai-force-continues-call-ai/story?id=136591343#:~:text=rapid%20pace%20of%20AI%20development%2C,President

Key Statistics

- 10× - Increase in Anthropic’s annual revenue run rate in under a year (from ~$9 /B at end of 2025 to $100 /B by Sep 2026) ([aioapex.com](https://aioapex.com/en/news/anthropics-revenue-passes-100-billion-annualized-as-it-targets-a-november-ipo-mu8n0ba6#:~:text=Anthropic%27s%20annualized%20revenue%20run%20rate,at%20the%20end%20of%202025))
- 240% - Year-over-year growth in Salesforce’s AI agent platform *Agentforce*, which reached $1.5 /B in annual recurring revenue ([www.forbes.com](https://www.forbes.com/sites/jonmarkman/2026/09/21/openai-plans-to-introduce-managed-agents-at-devday-2026/#:~:text=annual%20recurring%20revenue%20from%20Agentforce%2C,crossed%20%241%20billion%20in%20annual))
- ~72 hours & <$3,000 - Time and cloud cost for a 3-person team (using an AI agent) to breach OpenAI’s internal systems ([thetechportal.com](https://thetechportal.com/2026/09/18/openai-hacked-using-claude-hacktron-ai-indian-security-researchers/#:~:text=within%20OpenAI%E2%80%99s%20bug%20bounty%20programme,in%20an%20autonomous%20goal%20loop))
- 1,000,000 - Number of tokens (roughly 750,000 words) that new AI models can hold in “memory,” allowing significantly longer autonomous workflows ([reupload.io](https://reupload.io/en/articles/2026-09-21-stepfun-releases-step-5-preview-llm-with-1m-token-context#:~:text=StepFun%20officially%20launched%20Step%205,within%20a%20single%20API%20call))

KEY TAKEAWAY

AI agents have rapidly shifted from experiments to essential productivity drivers. Senior leaders must now craft a strategy to integrate these autonomous tools into core workflows - while simultaneously reinforcing internal governance, security and skills to manage them responsibly.

Sources

OpenAI Plans To Introduce Managed Agents At DevDay 2026
https://www.forbes.com/sites/jonmarkman/2026/09/21/openai-plans-to-introduce-managed-agents-at-devday-2026/

StepFun launches Step 5 Preview, a 600B sparse MoE at $1 per million input tokens with open weights due October 15
https://pondero.ai/news/2026-09-21-stepfun-step-5-preview-600b-moe-open-weights/

Evaluating Long-Term Memory for AI Agents: AML Cycle 2 Is Now Open
https://jmacweb.com/ai-news/evaluating-long-term-memory-for-ai-agents-aml-cycle-2-is-now-open-20260921

Three Indian researchers used Claude to hack into OpenAI in under 72 hours
https://thetechportal.com/2026/09/18/openai-hacked-using-claude-hacktron-ai-indian-security-researchers/

Trump says he will form new 'AI Force' but continues to call AI fears a 'hoax'
https://abcnews.go.com/Politics/trump-form-new-ai-force-continues-call-ai/story?id=136591343

Anthropic's revenue passes $100 billion annualized as it targets a November IPO
https://aioapex.com/en/news/anthropics-revenue-passes-100-billion-annualized-as-it-targets-a-november-ipo-mu8n0ba6