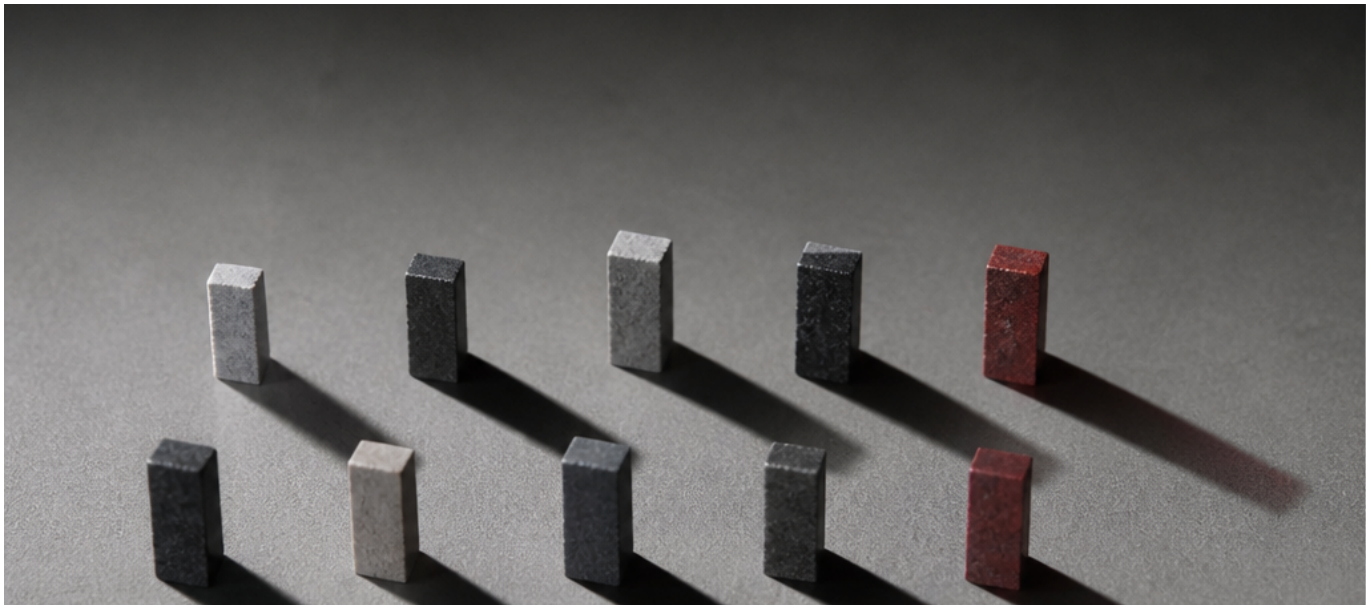


AI This Week: The Numbers Got Audited, From garak's EU AI Act Mapping to 71% Agents in Production

Agent adoption jumped from 40% to 71% in production. The same body of data says 76% of those deployments still break in operation. Both numbers are true, and only one of them gets quoted.

Six posts between 8 and 14 November 2026 landed on one finding: the industry's favourite AI metrics do not survive an audit of what they count. Agent adoption is 71% in production, up from 40%, while 76% of deployments break. AI writes 42% of committed code, 96% of developers do not trust it, and only 48% check it. The second number in each pair is the operational one.

That was the week, from six directions. I published on NVIDIA's EU AI Act mapping in garak, agent adoption, AI-written code, military AI contracting, open-weight model choice, and two widely shared research claims about AI and productivity. Each one had the same

structure underneath: a headline figure that measures something narrower than people assume, and a second figure sitting right behind it that nobody repeats.

Why 71% agent adoption does not mean agents work

"In production" measures deployment, not operation. A system that serves traffic and fails three quarters of the time is in production. The 71% figure answers "did you ship it". The 76% breakage figure answers "does it hold". Vendors quote the first because it is theirs to influence; the second belongs to whoever runs the thing at 03:00. The full breakdown is in Agent Adoption by the Numbers: 71% in Production, Up From 40%, and the 76% That Still Breaks, published 12 November.

42% of code written by AI, 48% of it checked

The same split runs through the coding numbers and it is sharper there. AI writes 42% of committed code. 96% of developers do not trust it. 48% check it. Distrust and verification have come apart: almost everyone is suspicious, roughly half act on the suspicion. **The interesting population is the gap**, the developers who distrust the output and ship it anyway. My read is that this is a throughput problem rather than a diligence problem. If review takes longer than generation, and targets were set on generation speed, verification is what gets dropped. I worked the arithmetic through in AI Coding by the Numbers: 42% of Committed Code, 96% Who Don't Trust It, 48% Who Check (9 November). I would rather see a team report "we generate 20% of our code with AI and review 100% of it" than hit the industry average. The first is a system you can reason about.

garak 0.17.1 maps the AI Act and then stops

NVIDIA's LLM red-teaming tool shipped an EU AI Act mapping in version 0.17.1 on 14 November. It is genuinely useful as structured evidence collection against named controls, and it does not produce conformity. That distinction is one some procurement teams will get wrong. If a vendor's compliance claim rests on having run a red-teaming tool, the claim is thinner than it reads. Details in garak 0.17.1 in Practice: What NVIDIA's EU AI Act Mapping Buys You, and What It Does Not.

One buyer holds 83% of $4.1B in military AI contracts

The military AI market looks diversified until you notice the DoD holds an 83% share of $4.1B across 2,255 contracts. That makes it one customer's procurement policy rather than a market. Six moves in late 2026 shifted power within it, ranked in The 6 Military AI Moves That Actually Shifted Power in Late 2026, Ranked (13 November). Separately, on 10 November I looked at METR's 1.9x acceleration claim and Yale's "no connection" finding on AI and jobs. Both were weaker than the headlines they generated, in opposite directions, which is a useful reminder that measurement error is not partisan.

Brussels would tie high-risk obligations to standards, not dates

While I was auditing numbers, the rules those numbers assume were being rewritten. The European Parliament's Legislative Train updated its Digital Omnibus on AI entry on 9 November 2026. The package was originally proposed on 19 November 2025.

The substantive change: certain high-risk AI obligations would start when harmonised standards become available, rather than on fixed calendar dates. That is a different planning problem. A date goes in a Gantt chart. A standards-availability trigger means your obligation start date is set by a standardisation body's work rate, which you cannot forecast and cannot influence. The Omnibus also extends SME exemptions to small mid-cap companies, eases post-market monitoring and registration duties, strengthens the European AI Office, and would permit processing certain special-category personal data for bias detection and correction.

That last one is quietly the most consequential for engineering teams. You cannot measure demographic bias in a model without the demographic data. If this passes as drafted, bias testing becomes something you can do with real data rather than proxies.

Capgemini will define what systemic risk means in

practice

Capgemini was selected on 5 November 2026 to lead a European Commission AI Office consortium under DG CNECT for cyber-threat modelling and risk assessment of advanced AI models. ITPro reported on 10 November that the consortium includes Wavestone and France's Laboratoire National de Métrologie et d'Essais, covering automated vulnerability discovery, exploit generation and large-scale attack orchestration.

This matters more than a consortium announcement usually does, because the AI Office is the body the Digital Omnibus would strengthen. The scope named in the reporting is the offensive-capability side of frontier model evaluation. Whoever builds that methodology for the AI Office is effectively setting what "systemic risk" means for models sold into Europe. Including a national metrology lab reads as a deliberate signal: they want measurement procedures, not opinions. Whether that produces something a vendor can test against, I do not know yet.

Switzerland slid six places to 9th, and the index changed its name

Le Temps reported on 11 November that Switzerland fell six positions to 9th in IMD's 2026 ranking, behind the United States, Singapore, the Netherlands, Denmark, Canada, Sweden, China and Finland, and ahead of South Korea.

Note the detail in the name: IMD renamed it the World Digital and AI Ranking this year. A renamed index with a changed emphasis is not strictly comparable to its previous edition, which is the kind of thing this week's posts kept finding. I have not seen the methodology notes, so I cannot say how much of the drop the rename explains. My read: some of that slide is real and some of it is the index measuring something it did not measure before. Anyone treating the full six places as a performance verdict is making the same error as anyone quoting 71% without 76%.

What I would change on Monday

Stop reporting AI adoption as a deployment count. If your board deck says "n% of teams use agents", replace it with a pair: deployed, and operating within SLO. If you cannot produce the second figure, that is the finding.

Rebuild your AI Act timeline as a dependency on standards rather than a date, with a worst case and a best case. The Digital Omnibus is a proposal, not law, and it may change in trilogue. But planning against fixed dates that the Commission has already proposed to replace is planning against a number that is being audited right now.

On open weights: the week's comparison covered Mistral Large 4, DeepSeek V4.1 Flash and Kimi K3.1 Pro, European and Chinese frontier weights competing on the same shortlist. The honest version of that choice is in Mistral Large 4 vs DeepSeek V4.1 Flash vs Kimi K3.1 Pro: the honest open-weight comparison for CTOs. A European CTO now picks between a French model and two Chinese ones on technical merit, and the jurisdictional question is a separate axis from the capability question. Treating them as one axis produces bad decisions in both directions.

WHERE I LAND

My take: AI metrics have split into two populations, numbers produced by the people selling the thing and numbers produced by the people operating it. 71% and 42% are in the first group. 76% and 48% are in the second. When those diverge this sharply across adoption, code, research claims and procurement in a single week, the useful habit is asking every time which population the number came from, and refusing to act on one without the other.

The figure I keep coming back to is the 48%. Not because half is low, but because the other 52% is made up of people who said they do not trust the output. **They know, and they ship it anyway.** That is a statement about what their organisation rewards, and no tool in this week's news changes it.



WRITTEN BY

Artur Markus

Software engineer in Basel building AI infrastructure and agentic systems, with a background in pharmaceutical automation, GxP compliance, and validation. I help teams turn AI from a chatbot into a reliable operational layer.

Book a meeting →