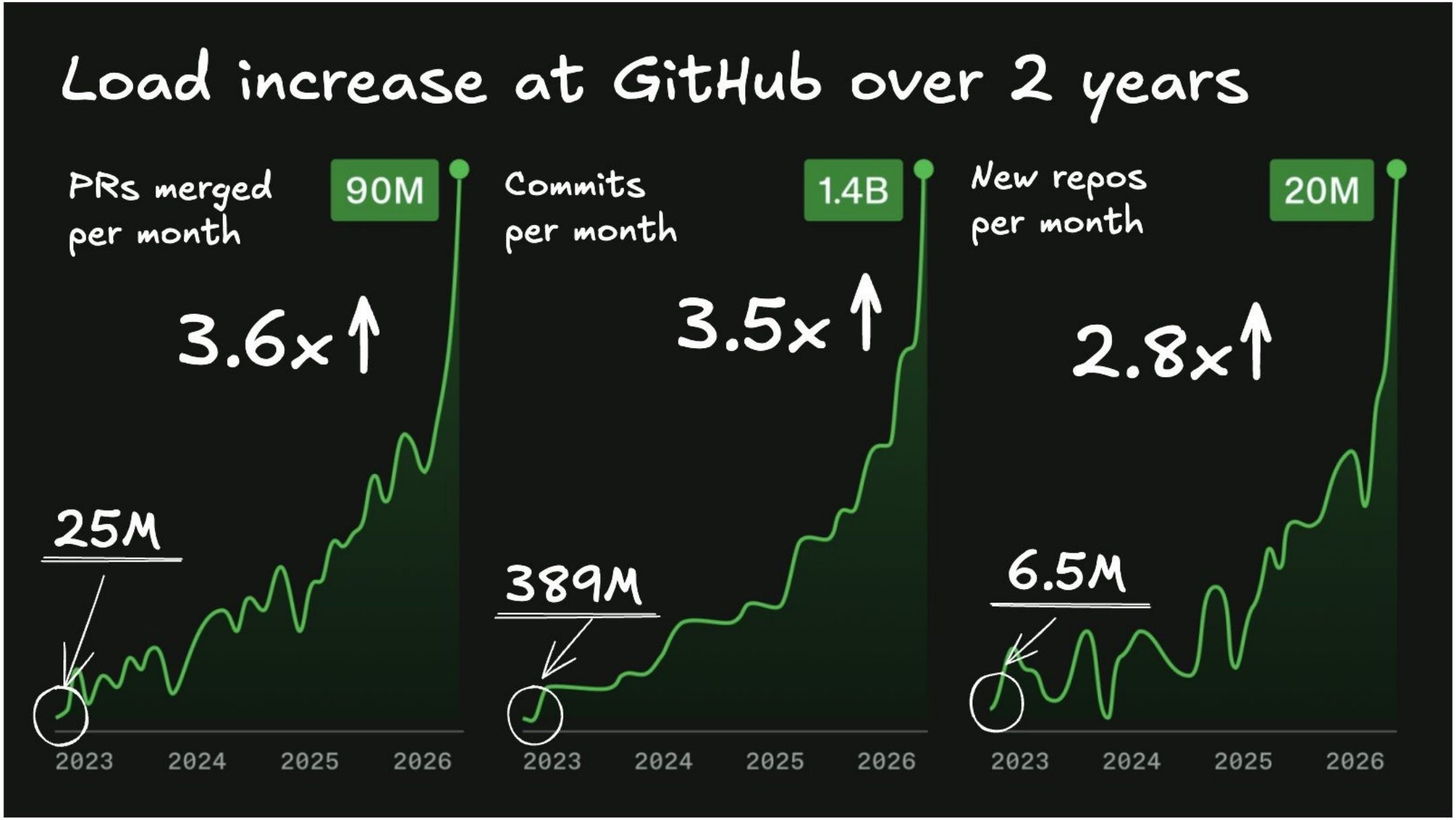
Here comes the normies

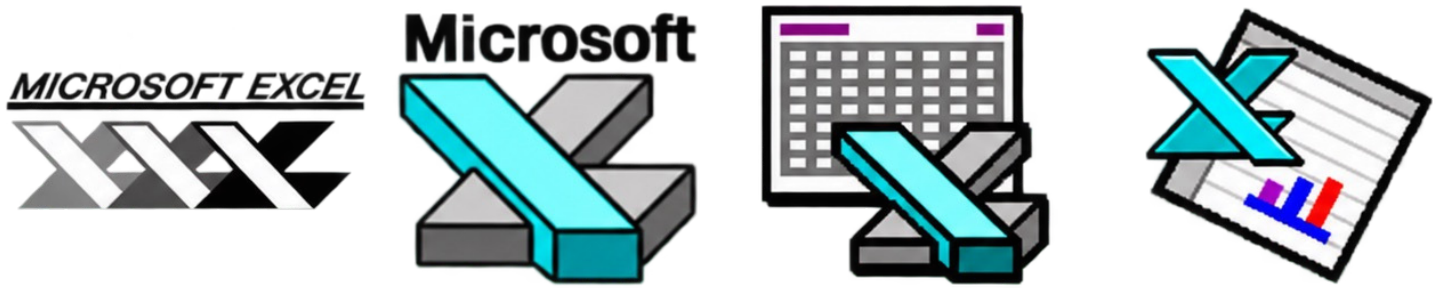
How DevOps can handle the flood of AI slop-apps

Coté, November 2nd, 2026 – SREday Amsterdam



Source: Gergely Orosz, The Pragmatic Engineer, May 2026, using data from GitHub.

THE PRAGMATIC ENGINEER



1985-90 | 90-94 | 94-95 | 95-99



99-2003 | 03-07 | 07-10 | 10-13



13-19 | 19-25 | 25-

“Vibe coding is letting Excel into the data center.”

Source: [”SRE at ING,”]() Daniele Tonella (CTO, ING), SREday Amsterdam, Nov 2, 2026.

“Accenture is trying to figure out how to stop non-technical workers from blowing through companies’ AI token budget on trivial tasks like **converting PDFs to presentation slides**.”

404 Media, June 2026

BBVA

4,000+

GPTs in regular use, out of 20,000+ created

US Federal Government

1,818

Active or planned AI use cases

AIRFRANCE KLM

200,000

hours of manual work saved by ~170 bots

SIXT SE

1,000

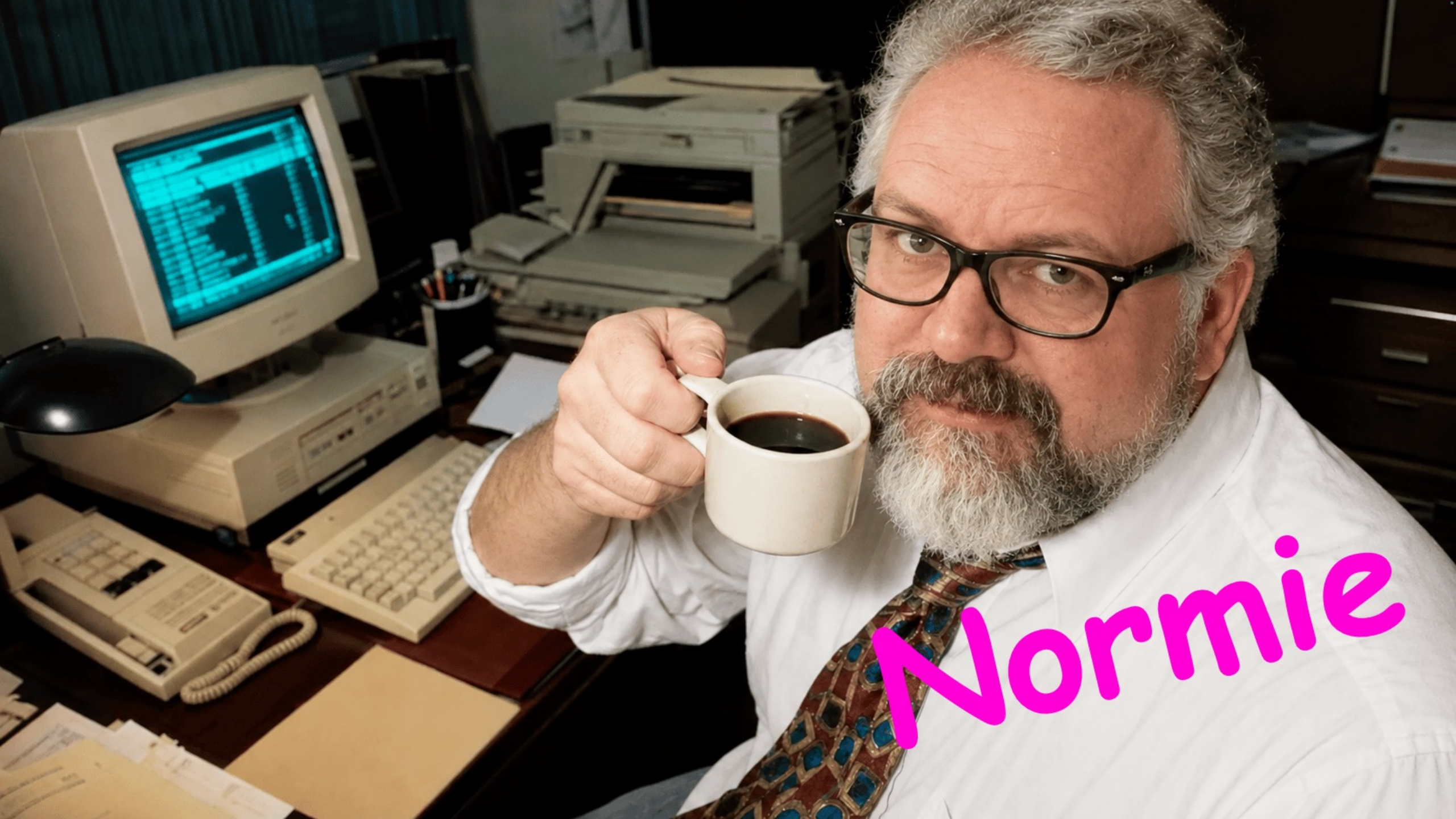
citizen-built apps in three months.

Sources: OpenAI, State of Enterprise AI 2025; 2025 Federal Agency Artificial Intelligence (AI) Use Case Inventory, as of 13 April 2026; Computer Weekly, July 2025; Boyan Dimitrov, PlatformCon 2026.

Table 1: Occupations and Tasks With Highest and Lowest genAI Adoption Rates

Top 10 occupations		**Top 10 tasks**	
Occupation	**Rate (%)**	**Task**	**Rate (%)**
Computer and information research scientists	87.3	Read documents to gather technical information.	61.3
Information security analysts	85.4	Prepare research reports.	60.7
Network and computer systems administrators	82.4	Design integrated computer systems.	59.5
Computer programmers	81.2	Analyze data to identify trends or relationships among variables.	57.5
Public relations specialists	80.4	Update knowledge about emerging industry or technology trends.	57.1
Personal financial advisors	79.7	Conduct quantitative failure analyses of operational data.	55.8
Chief executives	79.7	Prepare data for analysis.	55.4
Software developers	78.8	Develop marketing plans or strategies.	55.0
Market research analysts and marketing specialists	78.7	Collaborate with others to determine design specifications or details.	54.8
Computer and information systems managers	78.2	Design websites or web applications.	54.6
Bottom 10 occupations		**Bottom 10 tasks**	
Occupation	**Rate (%)**	**Task**	**Rate (%)**
Stockers and order fillers	14.8	Arrange delivery of goods or services.	0.0
Janitors and building cleaners	13.8	Assist practitioners to perform medical procedures.	0.0
Other installation, maintenance, and repair workers	13.1	Collect biological specimens from patients.	0.0
Maids and housekeeping cleaners	11.3	Drive trucks or other vehicles to or at work sites.	0.0
Fast food and counter workers	11.3	Inspect electrical or electronic systems for defects.	0.0
Maintenance and repair workers, general	11.2	Maintain medical equipment or instruments.	0.0
Painters and paperhangers	10.4	Measure the physical or physiological attributes of patients.	0.0
Licensed practical and licensed vocational nurses	10.4	Prepare patient treatment areas for use.	0.0
Receptionists and information clerks	7.6	Present food or beverage information or menus to customers.	0.0
Animal caretakers	5.3	Record service or repair activities.	0.0

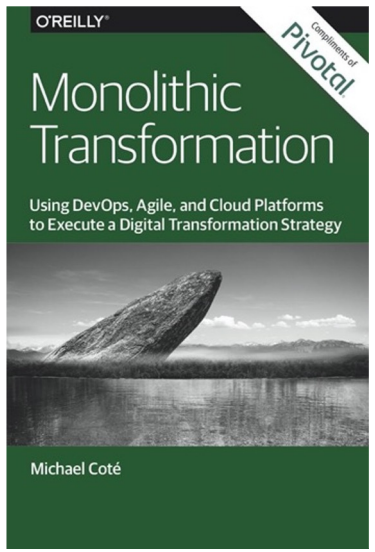
Source: Bick, Blandin, Deming & Schumacher, "What Work Does Generative AI Do?", Federal Reserve Bank of St. Louis Working Paper 2026-017A, August 2026, Table 1, page 16; n = 13,920 employed US respondents.

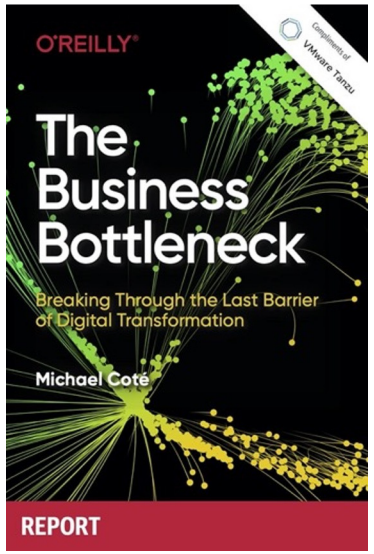

Normie

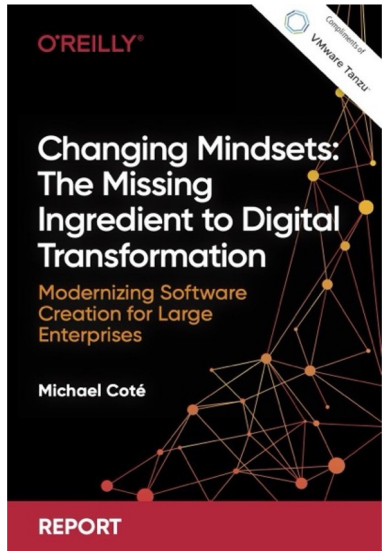


Coté

https://www.cote.io/ | michael.cote@broadcom.com











Tanzu Catsup

RedMonk DELL 451 Research S&P Global Market Intelligence Pivotal vmware Tanzu by Broadcom

read 600+ RSS feeds overnight

write the morning briefing

mirror the blog to the NAS

reindex 200,000 documents

prune old transcripts

sync the bank accounts

publish a blog post

tune Diane's own rules and memory

drain the summaries queue

post-produce a podcast episode

build or fix one of my own skills

prep a conference talk

file the paperwork

clean up files and free disk

summarize an article or paper

summarize a video or a talk

fix the overnight job that broke

track CFPs and submit talks

cut shorts and clips

write up a report

build a little web page

make an image or an icon

post a photo gallery

reconcile the finances

write the social posts

assemble the newsletter

work out what I'm paying for

sweep the task list

transcribe a recording

debug my own Mac

sweep Slack for links

file an expense claim

triage Gmail

write YouTube titles and descriptions

narrate something in my own voice

build a chart

tidy up the bookmarks

run the website analytics

Source: Claude analysis of Coté's usage from 2026-06-24 to 2026-10-01; n = 422 sessions over 78 days.

What we built

- feed, classify, summarize, publish
- assemble a recurring document
- search my own stuff
- an inventory I kept in a spreadsheet
- query & change a system of record
- jobs that run overnight

What that is at work

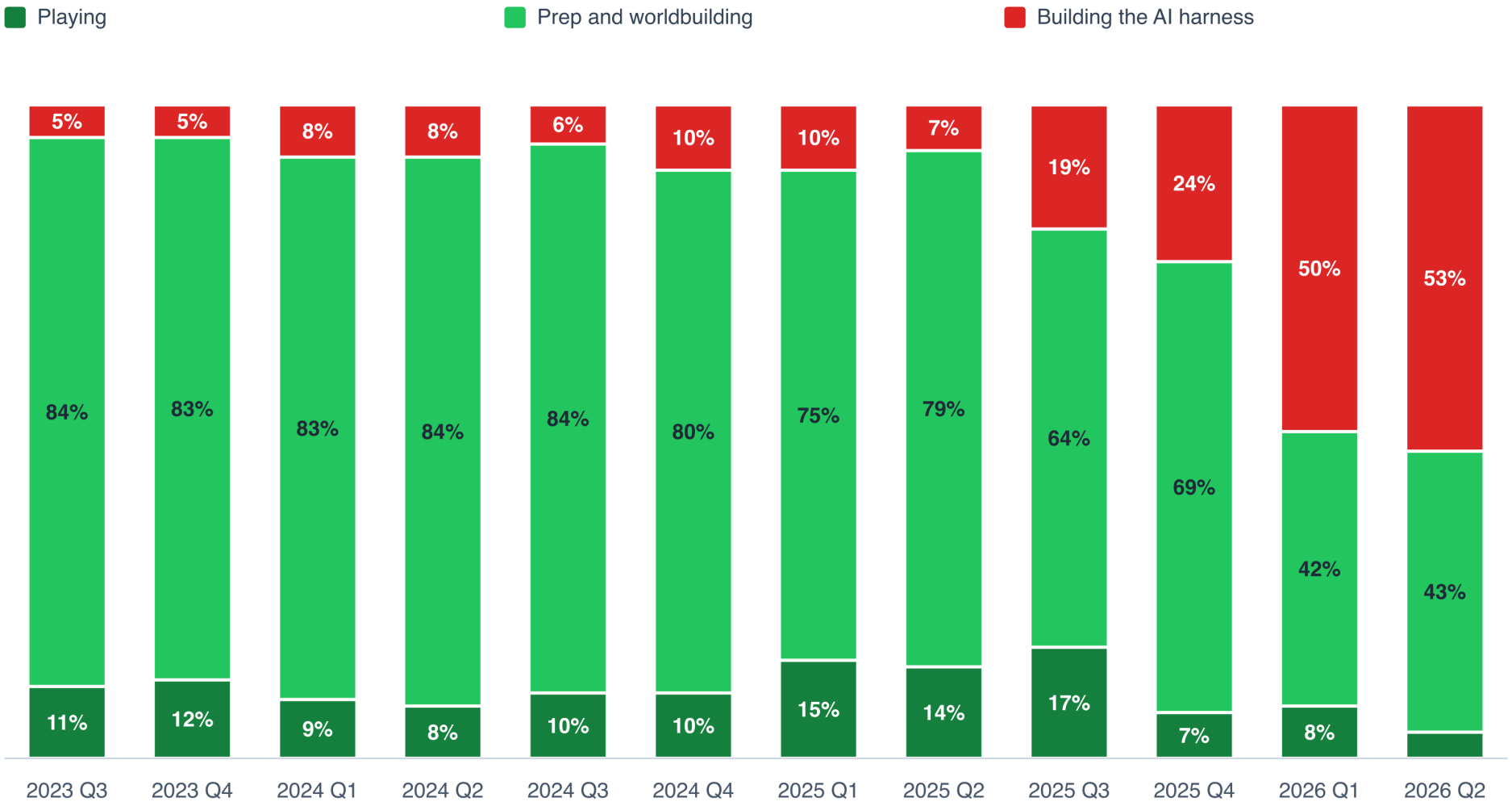
- the competitive-intel digest
- the weekly status report
- enterprise RAG over the shared drive
- the CMDB, also a spreadsheet
- ServiceNow, Workday, SAP, via MCP
- the batch job

Training material, product strategy, go-to-market content, “research,” pptx and xls, expenses, competitive analysis, customer research, CRM, etc.

Source: Claude analysis of Coté’s usage from 2026-06-24 to 2026-10-01.

Harness building eats the fun

One person's AI-run D&D campaign, by quarter. Share of conversations spent building the machinery: 5% at the start, 53% by the end.



Note: 4,245 conversations across ChatGPT, Claude.ai, Claude Code and Gemini, of which 1,522 are about tabletop roleplaying. Classified by a local model, 87% accurate on a 30-case calibration set - the shape is reliable, individual counts less so. 2026 Q3 is omitted: it had not finished.

Source: librarian full-text index of Coté's own ChatGPT, Claude.ai, Claude Code and Gemini archives. Retrieved 2026-08-26.

Chart generated by AI on 2026-08-27. Likely not checked by a human, so it may have errors.

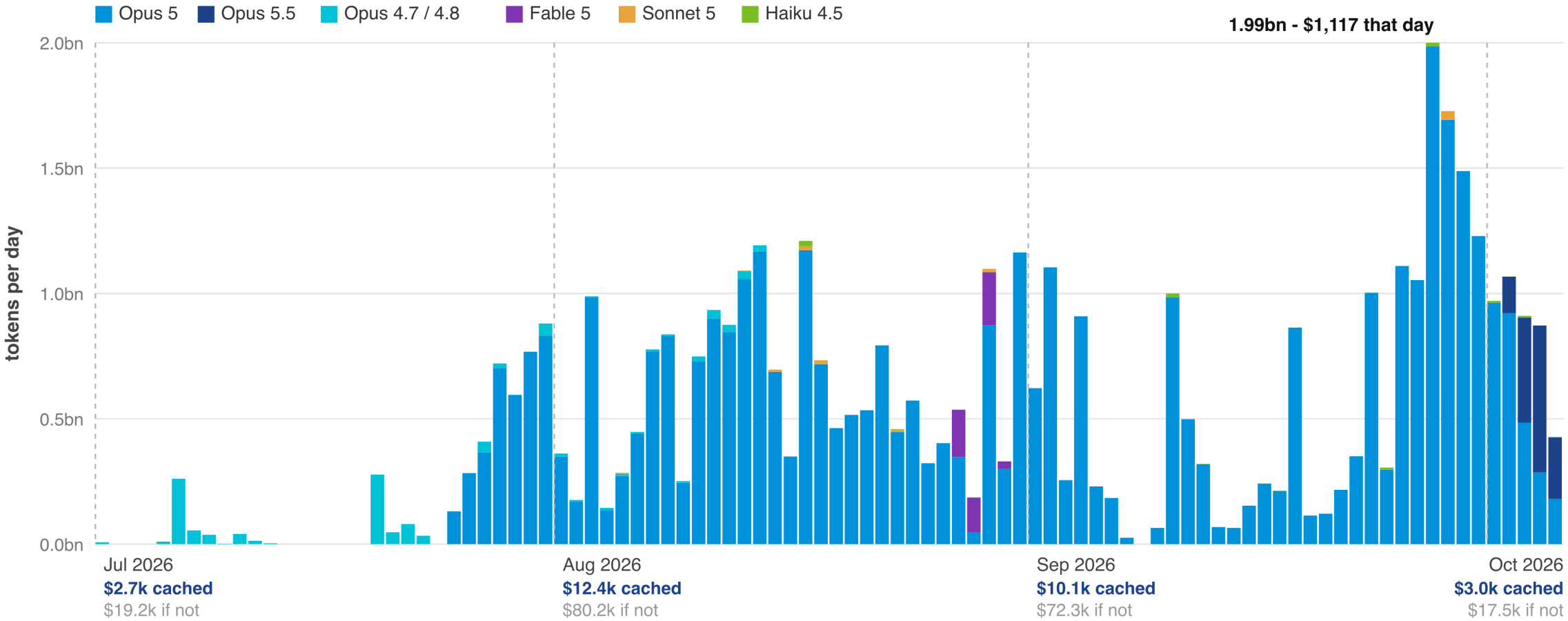
The one person “enterprise” needs:

- Authentication, authorization: credentials, keys, permissions.
- Governance: allowed, needs a human, never.
- Data policy: A list of what must never leave the ”enterprise.”
- Audit: a list of what counts as leaving, what did leave.
- Cost: tokens, used, and left.

Tokens per day, by model - and what they would cost at API rates

All Claude Code projects, 1 July to 5 October 2026. 86 active days, 173,973 requests, 45.6 billion tokens.

$28.2k as billed with caching. $189.2k with no cache at all.

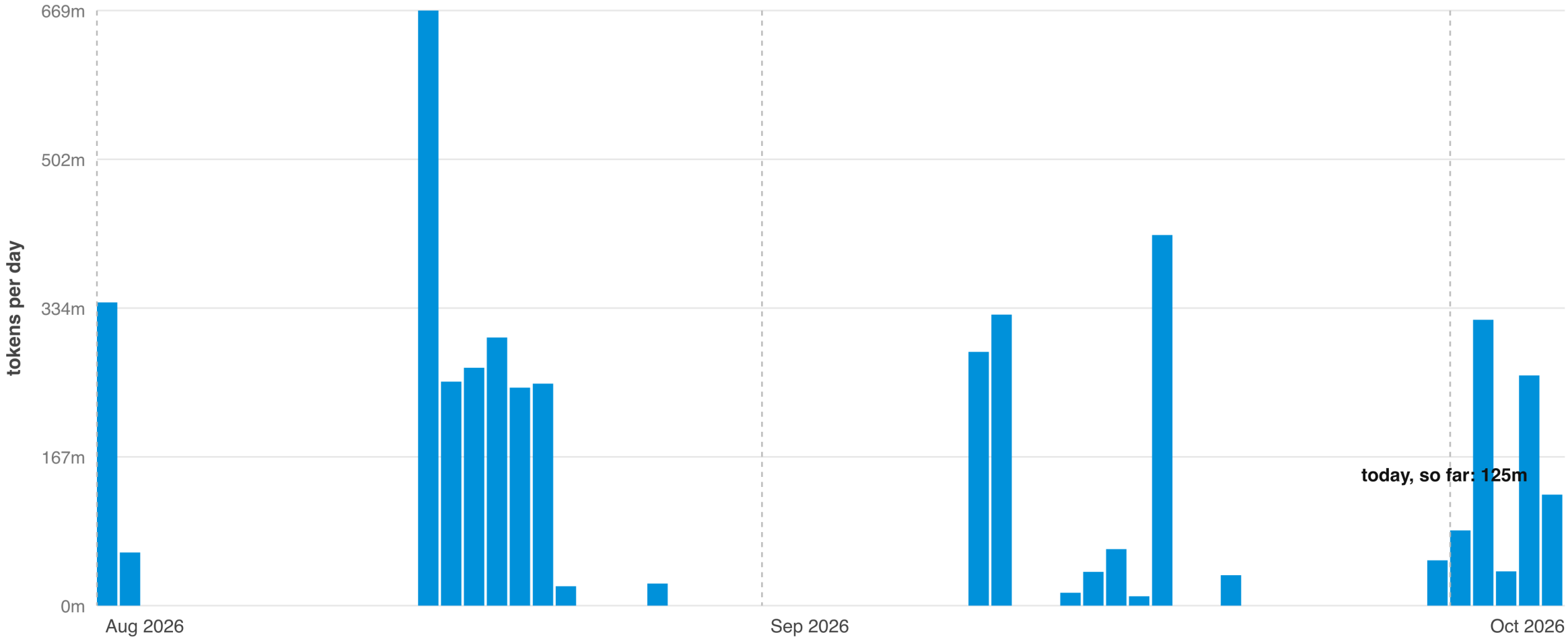


ALMOST NONE OF THIS IS NEW TEXT. Cache reads are 44.0bn of the 45.6bn, or 96.5% - context already sent, retained, billed at a tenth of the input rate.
Fresh input across all 86 days is 532,946 tokens. Output - the words actually written - is 133 million. Cache writes are 1455 million.
Cost is estimated at API rates: $4.10 per million input-equivalents, the rate implied by four days of metered telemetry. Cache read 0.1x, write 1.25x, output 5x.
Caching avoids $160.9k of the $189.2k, a factor of 6.7x. A high cache-read share is the sign it is working, not a warning.
Local models are not in this chart. Ollama logs requests, never tokens, so there is no local equivalent of any number here.

What this talk cost to make*

Tokens through Claude Code on the talks project, 3 August to 5 October 2026. 24 working days, 39 sessions, 10,211 requests.

4.50 billion tokens. About $2,432 at API rates - and about $18,598 if nothing had been cached.



98% of it is cache reads. Fresh input across the whole project is 20,515 tokens - about four pages of typing. Output is 8.9 million.
Cost is estimated at API rates: $4.10 per million input-equivalents, the rate implied by four days of metered telemetry. Cache read 0.1x, write 1.25x, output 5x.
Nothing before 3 August is in this project directory.

* Not including image generation, CFP work, etc.



Enterprise requirements for AI

Build

Coding & compiling.
Registry & CLI install.
Hidden SDLC.
Review/feedback.
Guardrails injection.

Run

Release management.
Runtime environment.
Sandboxed execution.

Security

Registry and CLI control.
CVEs and patching.
Credential rotation.
Auth & permissions
MCP gateway.

Connect

Service binding.
Data access.
Network access.

Observe

Performance.
Debugging, etc.
Audit records.
Evals and sampling.

Spend

Model access, public & private
Token tracking.
Rate limits and quotas.



Push to Harbor

Image Registry

VPC

Networking & Load Balancing

Git, Jenkins, Helm

K8s

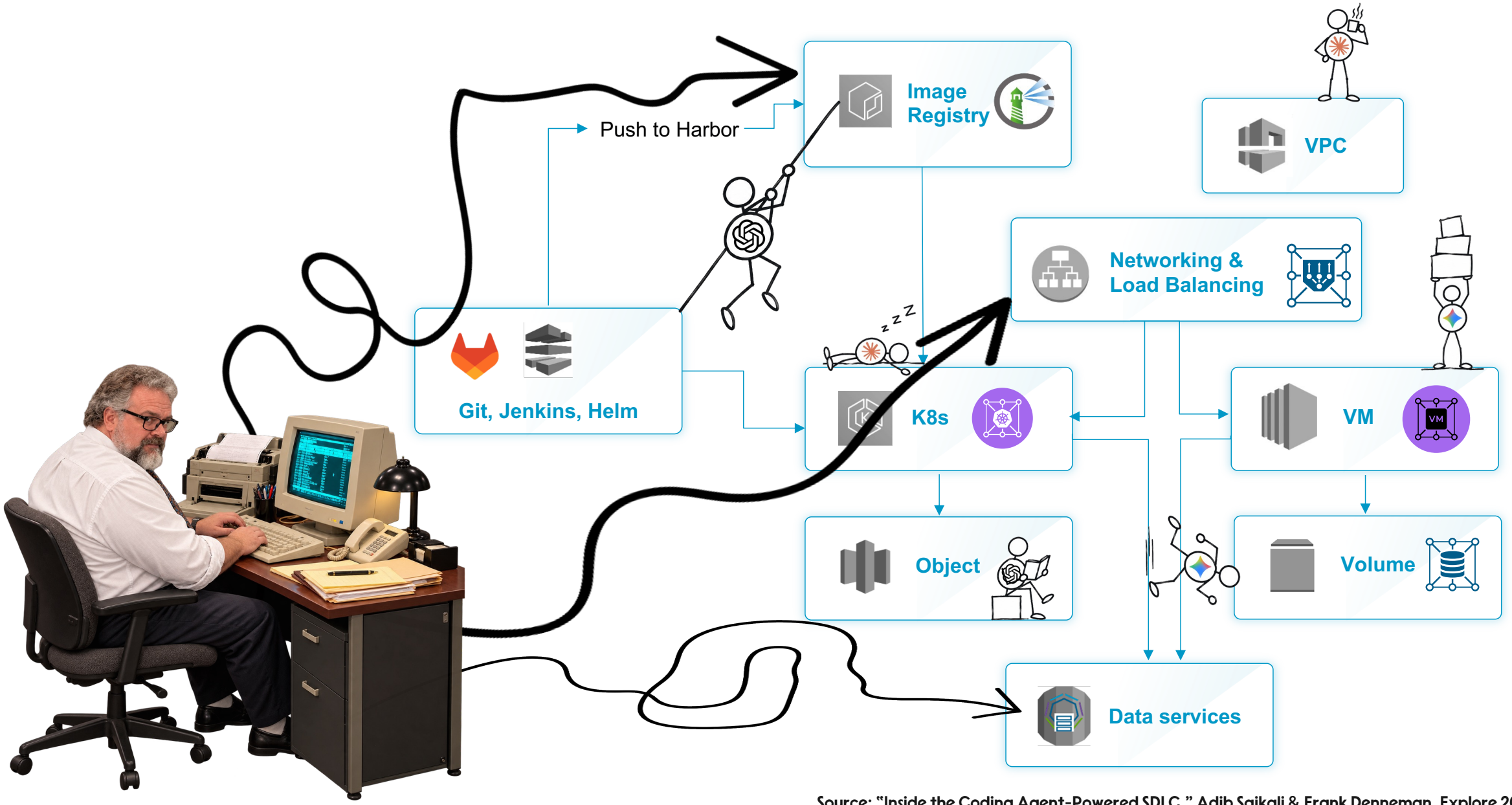
VM

Object

Volume

Data services

Source: "Inside the Coding Agent-Powered SDLC," Adib Saikali & Frank Denneman, Explore 2026.



Source: "Inside the Coding Agent-Powered SDLC," Adib Saikali & Frank Denneman, Explore 2026.

"Instead of developers 'shifting left,' they need to **shift down** and push more workloads down onto the platforms they're already using."

Richard Seroter, Google Cloud, June 2023

Guardrails at build time with a proper platform









cf push
cf bind-service
cf start
etc…



buildpack



Platform as a Service

build

app

app serving

service create

service bind

registry





Platform Engineer

VMware Tanzu® Platform

Sources: "Inside the Coding Agent-Powered SDLC," Adib Saikali & Frank Denneman, Explore 2026; "When everyone can build: The platform behind 1000 citizen-built app," Boyan Dimitrov, CTO of SIXT, PlatformCon 2026, published 22 June 2026.

Agent *Provisions* Infrastructure

- Agent must reason about infrastructure topology, sizing, and configuration
- Each agent can make different provisioning decisions
- Dedicated infrastructure leads to over-provisioning and sprawl
- More agent decisions create more opportunities for errors and varioation
- Agents control security



Agent *Consumes* Services



- Agent selects from platform-defined service plans
- Platform builds apps, not agents
- Platform controls sizing, configuration, policy, and lifecycle
- Services can be backed by shared, right-sized infrastructure
- Agents get repeatable, governed services
- Platform controls security

Source: "Inside the Coding Agent-Powered SDLC," Adib Saikali & Frank Denneman, Explore 2026.

"Treat every agent like it's a human. The identity controls, the access to information, the ability to take actions, the audit trail."

Stephen Elliott, IDC

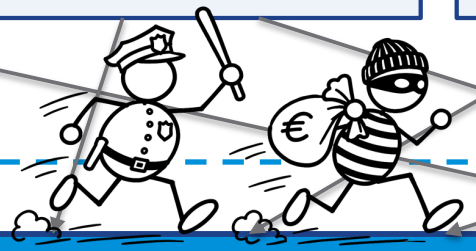
Source: "Conquering the Agentic AI Journey with a Private Cloud Platform," Stephen Elliott (IDC), Prashanth Shenoy and Shobhit Bhutani (Broadcom), September 24th, 2026.

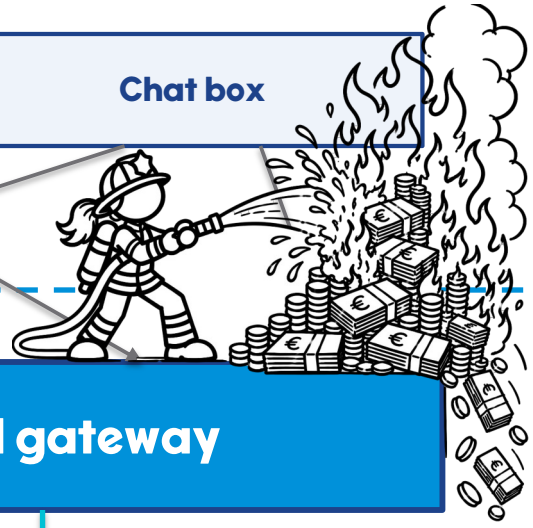


Another team's agent

A vendor's agent

Chat box





VMware Tanzu Platform

Tanzu MCP Gateway

Model gateway

Identity

Entitlement

Rate + budget

Audit

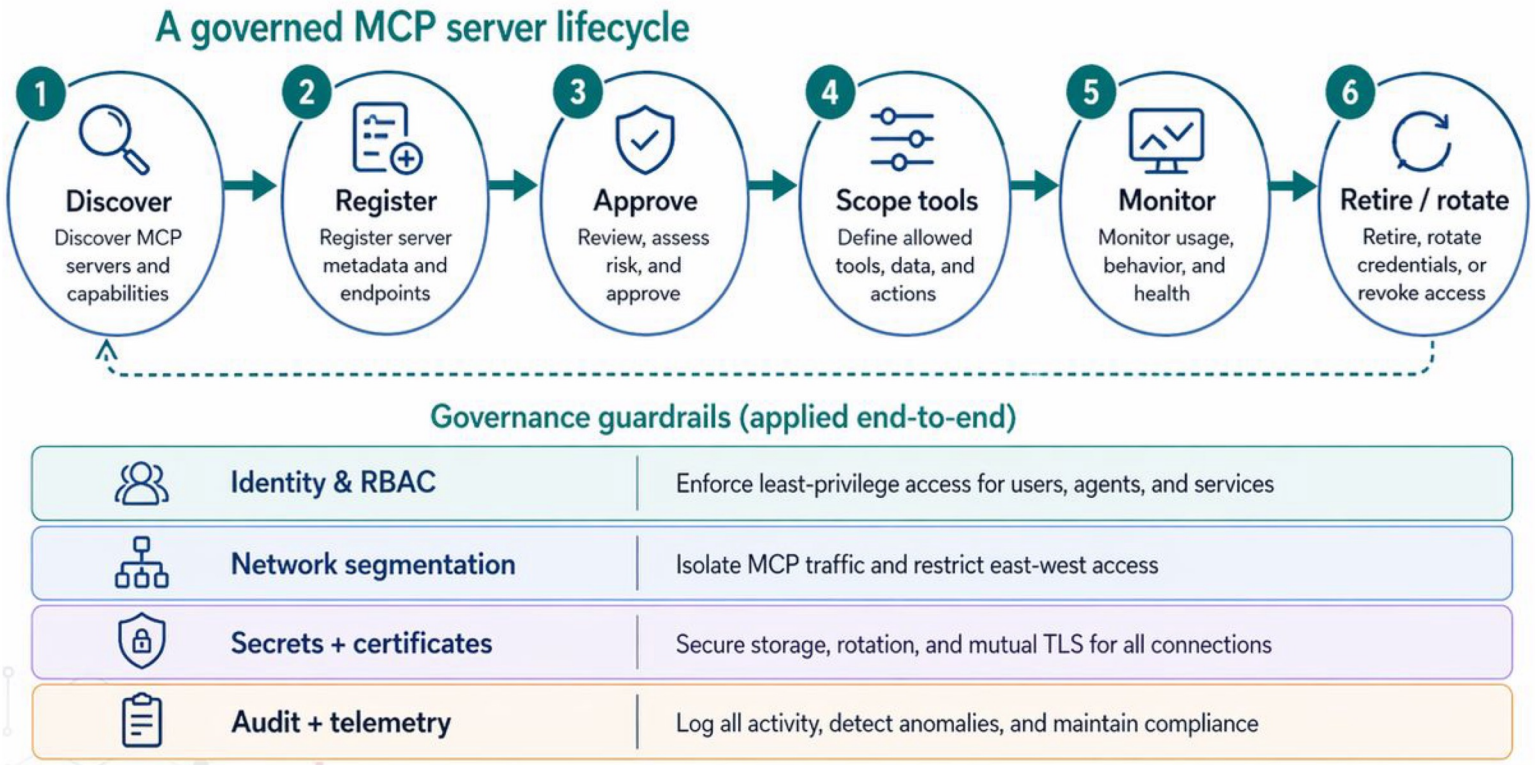
MCP servers

Data stores

Other Services

GenAI on Tanzu Platform

Deep dive: MCP server governance pattern



> *MCP support will provide crucial governance and security for MCP.*
>
> — *Broadcom*

Do not treat MCP servers as casual plugins; treat them as privileged integration services.

Sources: Broadcom VCF AI-native announcement; ServiceNow MCP server governance docs; MCP specification

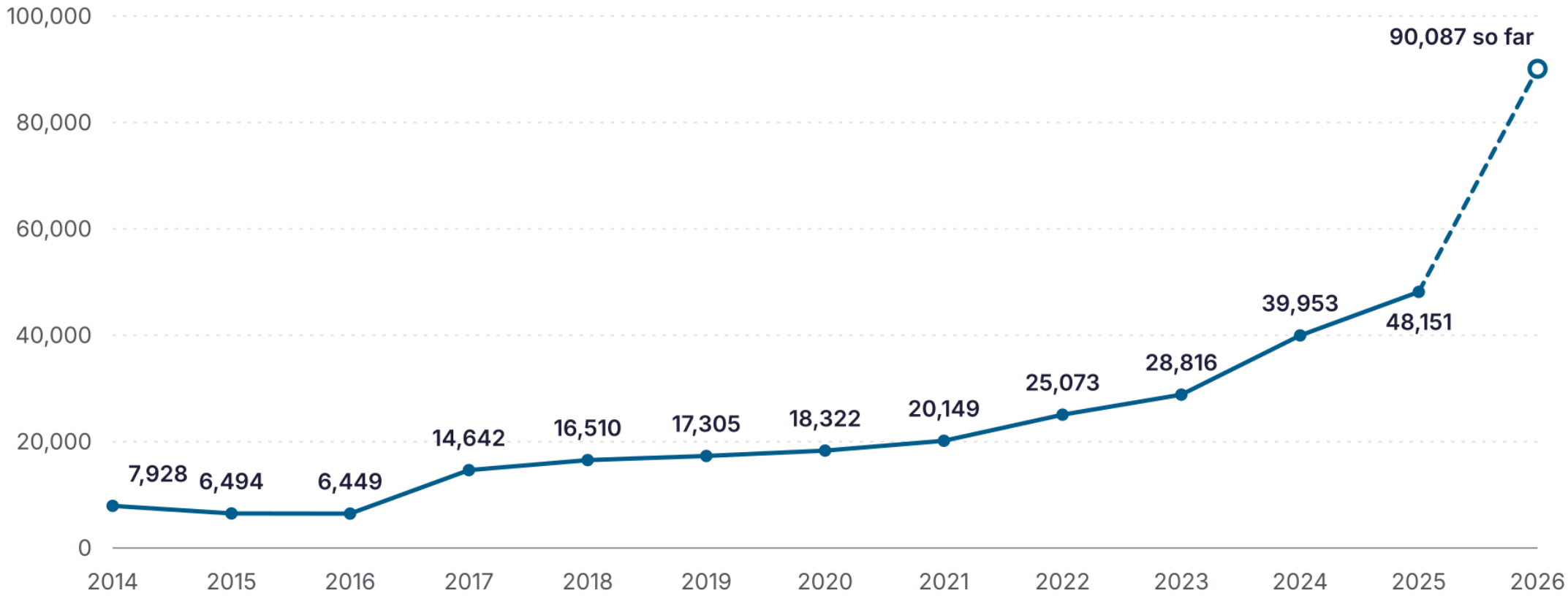
vmware by Broadcom

Copyright © 2026 Broadcom.
All Rights Reserved. The term "Broadcom" refers to Broadcom Inc. and/or its subsidiaries.

Source: "The Blueprint for Trusted AI Assistants and Copilots," John Yani Arrasjid and Mark Fynes, Explore 2026.

CVEs reported by year

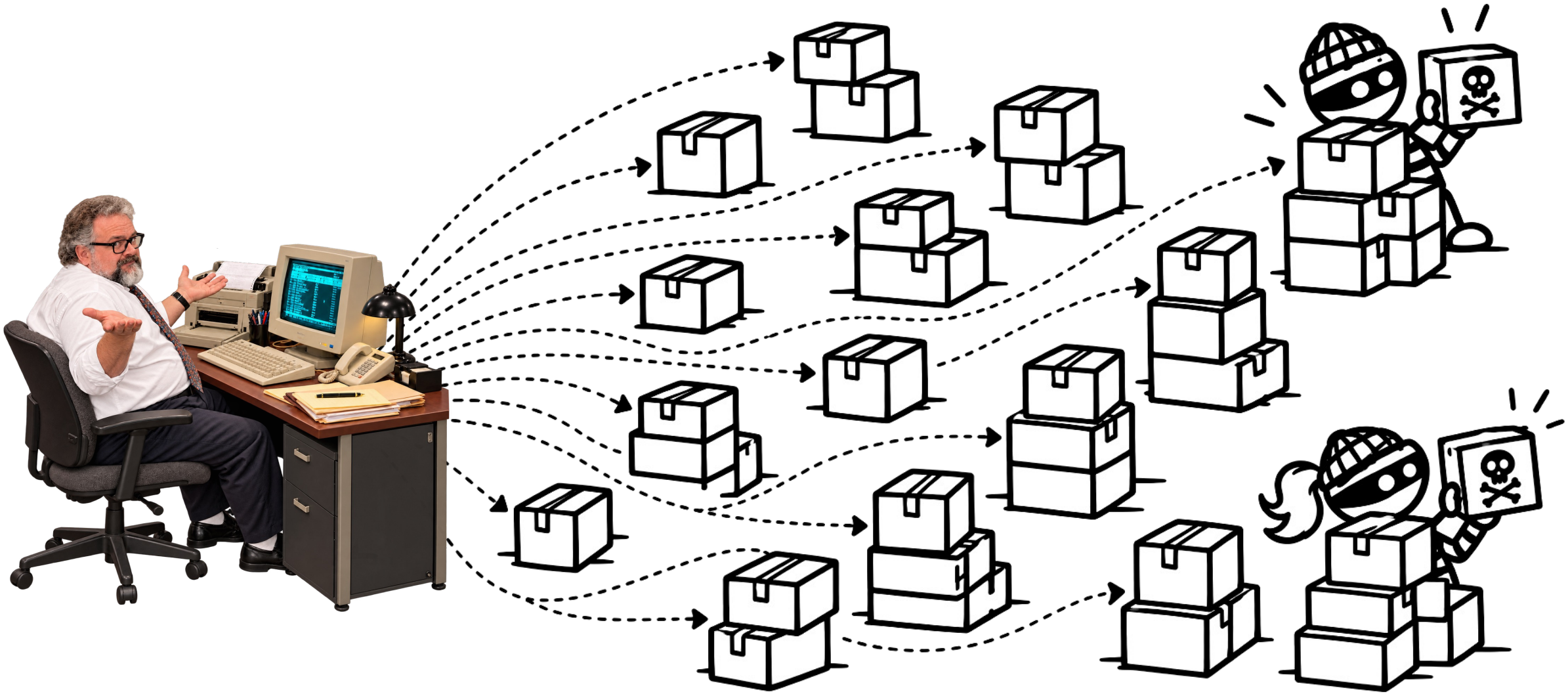
CVE records per calendar year, 2014 to 2026. 2026 is still running.
2026 has 10 months on the board, not twelve, and is already 87% past all of 2025.



2026 counts records published through 2026-11-02; every other year is complete. Published records, not vulnerabilities found or exploited that year.

Source: CVE.ICU, cve.icu/years.html, built from NVD and the CVE List V5 repository. Data as of 2026-11-02.

Chart generated by AI on 2026-11-02. Likely not checked by a human, so it may have errors.



It's time to pay for the "free" beer.

Source: "Open-source security in the AI exploit era: Patch without changing a thing," Oren Penso and Coté, *The New Stack*, November 1st, 2026.

What to do about it

Find them

Inventory AI, skills, MCP, apps.
Expect 20x your forecast.
Ask who owns each one.

Identify

Each agent its own identity.
Scope data and actions.
Keep the audit trail.

Shift down

Agents consume services.
Platform builds and deploys.
Version and review skills.

Gate

MCP and model gateways.
Entitlements, rate limits.
Approve and retire servers.

Patch

Curated registries, CLIs.
Central CVE patching.
Rotate credentials.

Meter

~ Tokens per person, per app.
Budgets and quotas.
Cheap, local models first.

That is… get a platform!



VMware Tanzu® Platform



Thanks!

💾 https://talks.cote.io/normieai-ams/

🏢 michael.cote@broadcom.com



Slides & stuff

