RESEARCH REPORT · ORG DESIGN × ARTIFICIAL INTELLIGENCE

AI hasn't eliminated management — it took over the relaying first

How eleven tech companies redesigned their organizations over three-plus years

Middle managers' daily work is partly leading people, making judgment calls and owning outcomes — and partly relaying information up and down, summarizing and filtering, and aligning across teams. AI compresses the second part first. The de-layering, layoffs and role redesign that follow have not made coordination disappear; they moved it into new roles and processes. The question ahead is not "how many more layers can we cut" but "how does the new organization actually run".

Companies studied **11 overseas tech firms**

Cited sources **100+ public links**

Video sources **19 YouTube links**

Evidence current through **2026-08**

The Full Picture First: Actions, Costs and How It Runs

FIGURE 1 · ONE PAGE, THE WHOLE PICTURE: ACTIONS, COSTS AND HOW IT RUNS

Three pillars answer three questions: what happened, what it cost, and how the organization keeps running.

> What is being replaced is the middle layer's "relaying and summarizing" work: passing information up and down, summarizing and filtering, aligning across teams — all three now being taken over by models and agents. Frontline management — leading people, making judgment calls, owning outcomes — has not been replaced (Google's move was precisely to turn managers of sub-3-report teams back into frontline ICs, individual contributors who build rather than manage; many stayed, rather than management being abolished)

ACTIONS

The 11 companies' public moves fall roughly into four types: structural reorganization, layoffs with de-layering, AI changing headcount, and long-standing flatness or deliberate leanness. Not every company is actively compressing layers; but from H2 2025, the layoffs clearly stopped being management-only.

COSTS

The day-to-day coordination managers carried has not disappeared — it changed address. Chief of Staff postings more than doubled in a year (a relayed figure, with no base or methodology disclosed), a signal worth watching though the sample and real effects still need observation. Meta's Project OT is the other reminder: AI can multiply code output without a matching rise in shipped value.

HOW IT RUNS

What is worth learning is not "how many layers to cut" but how things run afterwards: how the decision chain shortens, who makes the first draft, where the rules live, how roles are recombined. And the organization should be designed to be redone at any time — no "new structure" turns out to be a stable end state.

Numbers with different statistical bases are not compared side by side for scale; every fact carries a clickable source where it appears.

SEPARATE THE ACTION FROM THE MOTIVE

The core of this de-layering wave is **shortening the decision chain**: Google, Coinbase and Microsoft have all turned fewer layers or wider spans into concrete organizational moves. Meanwhile, Amazon CEO Jassy attributed the 2025 layoffs to culture and bureaucracy, not AI or financial pressure. Similar actions, differing public attributions — "de-layering the organization" and "the motive for layoffs" have to be judged separately.

FIGURE 2 · SEVEN KEY NUMBERS: MANAGEMENT CUTS AND FRONTLINE LAYOFFS IN PARALLEL

The first two items (Google) are the sample of "precisely cutting management roles"; the other five show this shift does not stop at management — the layoffs at Amazon, X, Block and Salesforce genuinely reached engineers and frontline roles. Nvidia's ~60 direct reports is "long-standing flatness", not "change".

Measured de-layering

Management-role measures

10%

Share of manager / director / VP roles Google has cut cumulatively (confirmed by Pichai at a December 2024 all-hands, as part of the efficiency drive started in September 2022)

Business Insider · Times of India

35%

Reduction in the number of Google managers with fewer than 3 direct reports (the precise measure as of 2025-08; many converted to IC roles and stayed, not a blanket "all managers" figure)

CNBC 2025-08-27

Workforce contraction

Percentages and headcounts use different scales

~80%

Share of X (Twitter) employees cut cumulatively after Musk's takeover, as the organization pivoted to engineer-led

The New York Times · The Verge

~40%

Share of Block employees cut in a single round on 2026-02-26 (roughly 4,000 people) — the immediate backdrop against which "From Hierarchy to Intelligence" (co-authored by Jack Dorsey, Block co-founder and Block Head, and Roelof Botha, Sequoia Capital partner and Block's lead independent director, 2026-03-31) was published — not a purely theoretical manifesto

Forbes 2026-03-31 · Fortune 2026-04-02

~30,000

Combined headcount of Amazon's two corporate layoff rounds in 2025.10—2026.01 (more than 1,800 of them engineers). This fact directly contradicts the older "flattening without layoffs" narrative

CNBC 2026-01-28 · CNBC 2025-11-21 · Seattle Times

Operating redesign

Span of control and role scale

~60

Number of direct reports Nvidia CEO Jensen Huang says he has, with no routine one-on-ones — media accounts put the figure anywhere between 36 and 60

Fortune 2024-11 · Lex Fridman Podcast #494

9,000→~5,000

Change in the size of Salesforce's customer-support workforce. Benioff said "I need less heads" on a podcast; Agentforce now handles roughly half of support cases

The Logan Bartlett Show · CNBC 2025-09-02

Sources: Business Insider · Times of India; CNBC 2025-08-27; CNBC 2026-01-28 · CNBC 2025-11-21 · Seattle Times; Fortune 2024-11 · Lex Fridman Podcast #494; The New York Times · The Verge; Forbes 2026-03-31 · Fortune 2026-04-02; The Logan Bartlett Show · CNBC 2025-09-02

CH.02 · DRIVERS

Three Drivers of the Shift

Why now? Why management?

KEY TAKEAWAYS

01 The traditional value of the middle layer is "information routing" — and routing is precisely the first job AI is good at. That is the core of the argument by Jack Dorsey (Block co-founder and Block Head) and Roelof Botha (Sequoia Capital partner, Block's lead independent director), and the starting point for understanding this shift.

02 Multiplied individual output shrinks headcount: Salesforce froze engineering hiring, Amazon signaled a leaner corporate workforce; with fewer people, the management layer contracts with them.

03 The competitive clock now ticks in months: Google's stated motive for cutting management roles was "fewer layers, faster decisions"; Coinbase wrote "no more than 5 layers below CEO/COO" into policy; Xbox compressed as many as 14 layers to 5 or fewer — and Nvidia's extreme co-design moves communication into collective settings rather than reducing it.

DRIVER 1

Coordination cost collapse

For decades, the core value of the middle layer has been "information routing" — summarizing upward, breaking work down downward, aligning sideways. Jack Dorsey (Block co-founder and Block Head) and Roelof Botha (Sequoia Capital partner, Block's lead independent director), in "From Hierarchy to Intelligence", co-published on March 31, 2026 (also carried on Sequoia's site), put it directly: **middle managers are essentially information routers, and AI can do the routing better**. The essay traces bureaucracy back two thousand years to the Roman legion — the eight-man contubernium as the smallest unit — then through the Prussian general staff and the railroad companies of the 1850s, to today's org chart.

DRIVER 2

Per-person output multiplies

AI coding assistants, agent-based support and automated workflows multiply what a single engineer or operator can produce. Salesforce, citing Agentforce, announced it would **add no new software engineers** in 2025; Amazon's June 2025 memo forecast "efficiency gains" from AI and a leaner corporate workforce. With fewer people, the management layer naturally contracts.

DRIVER 3

The competitive clock speeds up

Model capability now jumps month by month. Google reorganized twice within 2024 alone; Jensen Huang says Nvidia must "move at the speed of light", and his ~60 direct reports exist precisely to delete the middle of the decision chain — he calls the method "extreme co-design": not less communication, but all communication in collective settings. When the strategic window is measured in weeks, layers tend to mean delay.

CH.03 · TRENDS

Six Organizational Trends

Common threads distilled from the public moves of 11 companies; not every company shows every one of them.

KEY TAKEAWAYS

01 The Great Flattening is the shared move across companies — but from the second half of 2025 it evolved from "cutting only management roles" into layoffs that reach the front line.

02 Merging research and product teams is not an end state: Meta's MSL split just 7 weeks after it was formed, showing organizations must keep adjusting to match the model release cadence.

03 The need for coordination does not disappear: spans of control widen (Gallup average 12.1, Coinbase allowing 15+) while Chief of Staff postings more than double in a year — a signal that coordination work may be redistributing, not yet proof of cause and effect.

FIGURE 3 · THE SIX TRENDS ARE NOT SIX ISOLATED EVENTS

The six trends resolve into three simultaneous changes: structures get thinner, team boundaries get redrawn, and authority and coordination get reorganized. See the relationships first, then check the company evidence item by item.

COMMON DIRECTION	**From hierarchical routing to system-level coordination**	AI takes over part of the relaying, summarizing and status tracking; people move closer to the problem, the judgment call and the responsibility.
STRUCTURE	**Thinner layers, wider spans**	01 The Great Flattening · 05 Wider spans of control
DELIVERY	**Team boundaries keep getting redrawn**	02 Research and product repeatedly reorganized · 04 Human–AI hybrid teams
AUTHORITY	**Power moves to the front line; coordination is redistributed**	03 Engineer-led culture · 06 Coordination roles reorganized

01

The Great Flattening

Google, Amazon, Microsoft and Meta each cut into their management layers between 2023 and 2026: Google cut 10% of management roles cumulatively (the precise 2025-08 measure: 35% fewer managers with fewer than 3 direct reports); Amazon raised its IC ratio by 15%+, then added roughly 30,000 corporate layoffs in 2025.10—2026.01; Microsoft laid off more than 15,000 people in 2025 with the explicit goal of increasing span of control; Meta's 2023 "Year of Efficiency" turned many managers back into individual contributors, followed by three rounds of cuts in 2026. Business Insider called it a "crushing" of middle managers.

COMPANY EVIDENCE

Google Single report / dataset · CNBC (2025-08-27) ↗

Amazon Independently corroborated · CNBC (2026-01-28) ↗ · The Seattle Times ↗

Meta Official / firsthand · Meta · Year of Efficiency ↗

Microsoft Independently corroborated · CNBC (2025-05-13) ↗ · CNBC (2025-07-02) ↗

02

AI research and product merge — then split again

Google folded the Gemini App team into Google DeepMind; Meta merged model and product teams into Meta Superintelligence Labs in 2025.06 — then split it into four teams just 7 weeks later in 2025.08. The lesson is that the merger itself is not an end state: putting model people and product people in one org does not mean the org shape will hold still.

COMPANY EVIDENCE

Google `Official / firsthand · Google Blog · 2026-08 ↗`

Meta `Single report / dataset · Engadget (Reuters) · 2026-08-26 ↗`

03

Engineer-led culture returns

At X under Musk, "the people who write code" replaced the management layer as the center of power, and the organization grew even more engineer-centric after CEO Yaccarino's departure; Nvidia has long run an "engineer CEO + extremely flat" structure; on Anthropic's Claude Code team "everyone writes code" — product managers, engineering managers, designers, finance and data scientists all commit code directly (per team lead Boris Cherny). "Founder Mode" lives on as a Silicon Valley catchphrase.

COMPANY EVIDENCE

X `Independently corroborated · The Verge ↗ · CNBC (2025-03-28) ↗`

Nvidia `Official / firsthand · Lex Fridman Podcast #494 ↗ · HBR IdeaCast · Jensen Huang ↗`

Anthropic `Official / firsthand · Lenny's Podcast · interview transcript ↗`

04

Human-AI hybrid teams become the headcount unit

Inside Salesforce, roughly 30—50% of work is already done by AI agents; some Coinbase teams are experimenting with "one person + N agents" pods. The new org-design question is no longer "how many people report to one manager" but "how many humans plus how many agents report to one owner".

COMPANY EVIDENCE

Salesforce `Official / firsthand · Fortune · Marc Benioff interview ↗`

Coinbase `Single report / dataset · Fortune (2026-05-05) ↗`

Meta `Single report / dataset · Business Insider (2026-03) ↗`

05 Wider spans of control: the verifiable numbers

Amazon's internal guidance raised the minimum direct reports per manager from 6 to 8 (AWS internal guidance revealed by BI 2025.01, relayed by Fortune 2025.09); after its reorg Coinbase allows leaders to carry 15+ direct reports (TechCrunch / Fortune 2026.05); Gallup finds the average US manager's direct reports rose from 10.9 (2024) to 12.1 (2025); Google defined the same class of metric precisely with "35% fewer managers with fewer than 3 direct reports". Morgan Stanley, working from Jassy's 15% target, estimated Amazon could trim roughly 13,800 manager positions and free $2.1—3.6 billion a year. Wider spans mean managers cannot micromanage — status tracking and report-rollups shift to AI.

COMPANY EVIDENCE

Amazon Official / firsthand · Amazon · Andy Jassy memo ↗

Coinbase Single report / dataset · Fortune (2026-05-05) ↗

Google Single report / dataset · CNBC (2025-08-27) ↗

06 Coordination roles reorganized: rising Chief of Staff demand

With the middle layer cut away, executives' information bandwidth tightens — postings for "Chief of Staff" (a corporate aide-de-camp role, close to running the office of the CEO) more than doubled over the prior year (Bloomberg 2026.07 citing Revelio Labs; the report discloses no base or methodology), which organization-design commentators have called "the bill for the Great Flattening". Block's alternative is the "Directly Responsible Individual": named ownership with the ability to pull in resources ad hoc, but no standing management layer; the player-coach model has drawn named criticism (a Fortune 2026.05 column: it "misunderstands the value of good managers").

COMPANY EVIDENCE

Chief of Staff Single report / dataset · Bloomberg (2026-07-16) ↗

Block Official / firsthand · Block · From Hierarchy to Intelligence ↗

CH.04 · COMPANY CASES

Deep Company Cases

This chapter walks through the 11 companies roughly by how prominent their org-change moves have been: Google, Amazon, X, Nvidia, Meta, Microsoft, Salesforce, Block, Coinbase, OpenAI and Anthropic. Each case includes a timeline, the key numbers and primary sources.

KEY TAKEAWAYS

01 The 11 companies' moves fall roughly into three types: structural reorganization (Google, Meta, OpenAI), layoff-driven (Amazon, Microsoft, X, Block, Coinbase), and long-standing flatness or deliberate leanness (Nvidia, Anthropic); Salesforce is the case of agents directly replacing headcount.

02 There is often tension between official rhetoric and actual moves: Amazon denies its layoffs were AI- or cost-driven, Block laid people off first and published the argument afterward — read each case's timeline together with its "tension" note.

03 No new structure has proven to be a stable end state: OpenAI's dual-track setup lasted about 14 months, Meta's MSL 7 weeks, and Google rewired its reporting lines again in 2026.08.

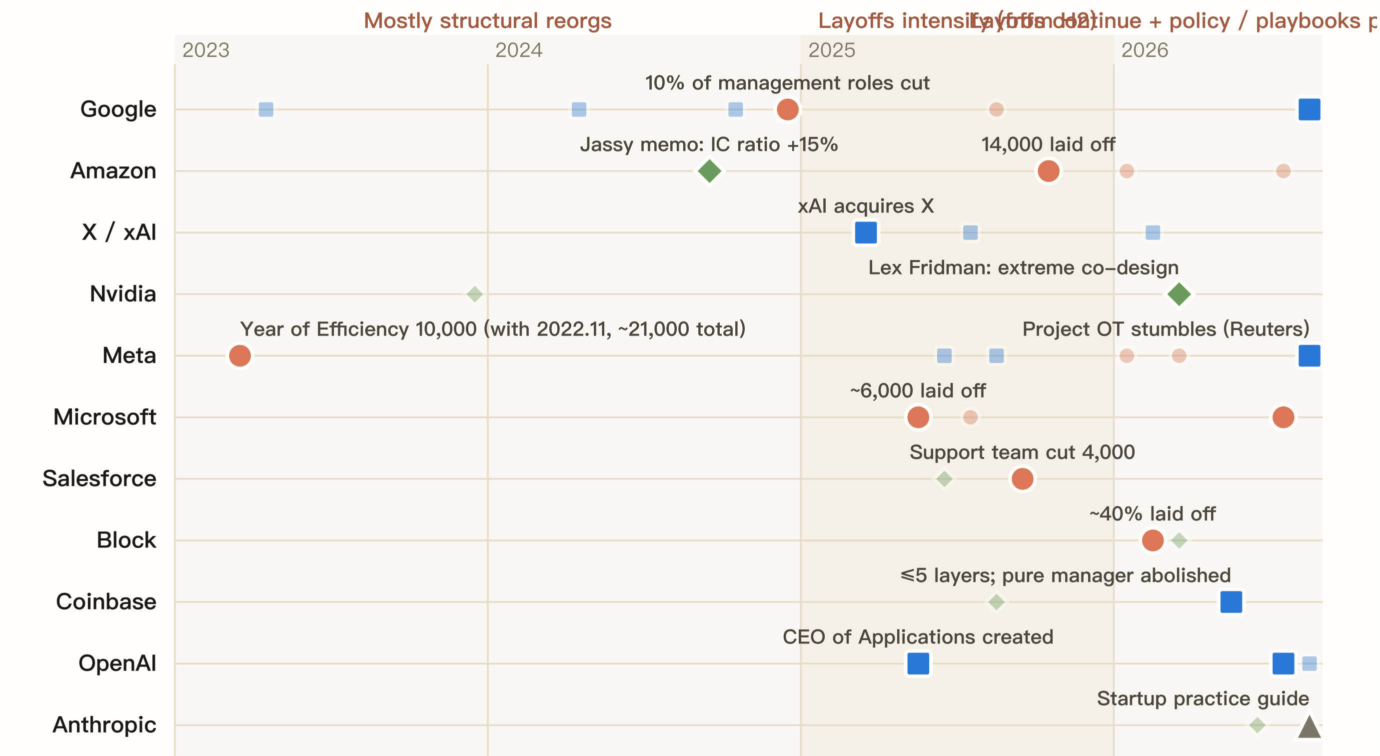
FIGURE 4 · THREE YEARS OF MILESTONES: REORGS, LAYOFFS AND INSTITUTIONALIZATION MOVES ACROSS 11 COMPANIES (2023.01–2026.08)

The moves are accelerating — and shifting gears: 2023—2024 was mostly structural reorganization, layoffs intensified from H2 2025, and by 2026 the practices are being written into policy and made public.

● Layoffs / management cuts ■ Reorg / reporting-line changes ◆ Policy and practice

▲ Publishing their own playbook

Solid large dots are key turning points (labeled); faint small dots are context events — hover or expand the table



Dates and original links for every event are in each company's timeline; key claims re-verified on 2026-08-30.

FIGURE 5 · IT IS ALL CALLED "FLATTENING" — IN PRACTICE THERE ARE FOUR DIFFERENT PLAYBOOKS

Nvidia is the "always been this way" template; Anthropic is "deliberately lean" (its reporting structure is single-source information). Learning from the current state of a company that never had to transform is a different exercise from learning from the process of one mid-transformation.

	STRUCTURAL REORG	LAYOFF-DRIVEN	AI SUBSTITUTION	LONG-STANDING FLAT / DELIBERATELY LEAN
GOOGLE	●	●	·	·
AMAZON	·	●	·	·
X / XAI	●	●	·	·
NVIDIA	·	·	·	●
META	●	●	·	·
MICROSOFT	·	●	·	·
SALESFORCE	·	●	●	·
BLOCK	·	●	·	·
COINBASE	●	●	·	·

	STRUCTURAL REORG	LAYOFF-DRIVEN	AI SUBSTITUTION	LONG-STANDING FLAT / DELIBERATELY LEAN
OPENAI	●	·	·	·
ANTHROPIC	·	·	·	●

● = a main practice verifiable in the company's 2023—2026 public moves; one company can occupy two types. The four-way split is our synthesis, not official company language. Anthropic's "long-standing flat / deliberately lean" placement rests on a single 2026.06 snapshot; whether it holds cannot be confirmed.

Basis for the classification: each company's public moves in 2023–2026; all 11 company timelines carry original links.

Google / Alphabet

Pichai: repeated reorgs + leaner management + the 2026.08 leadership reshuffle

Google is the big company that has reorganized most often in the AI era: two major restructures within 2024 alone, the core logic being to stitch the model team (DeepMind) directly to the product team (Gemini App); in August 2026 it completed its biggest rewiring of reporting lines in years.

—10%
manager / director / VP roles (cumulative, 2022.09—2024.12)

—35%
managers with fewer than 3 direct reports (2024.08—2025.08; many stayed on as ICs)

2
major reorgs within 2024

2023.04

Google Brain and DeepMind merged into Google DeepMind, reporting to Demis Hassabis — ending the era of two parallel, competing AI labs.

Why: competing with OpenAI/Microsoft required concentrating compute and talent, and avoiding internal wheel-reinventing.

2024.04–10

Two reorgs within one year: in April, parts of Research moved into Google DeepMind and a Platforms & Devices division was created; in October, the Gemini App team moved into Google DeepMind, the Assistant team into Platforms & Devices, with Search and Ads leadership adjusted in step.

Why: put application teams and model teams at zero distance, shortening the path from model capability to product feature.

2024.12.18

Pichai confirmed at an all-hands: **10% of manager / director / VP roles had been cut**, as part of the efficiency drive launched in September 2022.

Why: fewer layers, faster decisions — a response to the OpenAI threat.

2025.08.27

VP of People Analytics and Performance Brian Welle gave a more precise measure at an all-hands: over the past year the number of **managers with fewer than 3 direct reports fell 35%**, many of whom stayed on as ICs; ten product areas offered voluntary exit programs at the same time.

This supersedes the vaguer "a third of all managers cut" version that had circulated — the cuts targeted managers of teams below a certain size, not management roles across the board.

2026.05.27

On The Verge's Decoder podcast, Pichai described Google's post-ChatGPT response as more than layoffs — a redesign of "who owns how technology flows": AI infrastructure was consolidated under a new SVP position, and a "Chief AI Architect" role was created — held since 2025.06 by Koray Kavukcuoglu (SVP, reporting directly to Pichai), of a piece with his expanded remit in 2026.08.

2026.08.05

Google completed its most consequential rewiring of reporting lines in years: Demis Hassabis stepped down as Google DeepMind CEO to become GDM Chair and Alphabet Chief Scientist, continuing to lead Isomorphic Labs; Jeff Dean, the "godfather-tier" chief scientist with 27 years at Google, left to found Discovery Loop (with Google as investor and cloud partner); DeepMind CTO Koray Kavukcuoglu was promoted to Senior Vice President, **reporting directly to Pichai**, jointly overseeing Gemini model development, frontier research, the Gemini App and developer teams with Josh Woodward.

This round further tightened the reporting chain from AI models, applications and developer teams up to the CEO.

2026.08.06–10

Follow-up reporting sketched the backdrop: Time said DeepMind's communications, legal and marketing functions were folded into their Google counterparts; Fortune reported months of low morale at DeepMind before the reorg, several senior researchers leaving over July—August, Gemini 3.5 Pro repeatedly missing internal timelines, and the center of gravity shifting from London to Mountain View.

These accounts are each single-outlet reporting, unconfirmed by Google; recorded here as background only.

> The oft-quoted Google "20% more efficient" line traces, verifiably, to **September 2022** — Pichai's remarks at the Code Conference — not to the "2023 all-hands" some reports cite. This report uses the 2022.09 dating.
>
> — Sourcing note

SOURCES BI: Pichai confirms 10% management cut · Times of India: all-hands details · CNBC: the precise 35% measure (2025-08-27) · Android Central: the October 2024 reorg explained · IT Pro · blog.google (official): the 2026.08 reorg announcement · CNBC: Jeff Dean departs (2026-08-05) · Fortune: Hassabis steps down (2026-08-05) · Hassabis's X post (official) · Time: DeepMind functions folded into Google (2026-08-06) · Fortune: the DeepMind reorg backdrop (2026-08-10)

Amazon / AWS

`Jassy: anti-bureaucracy + IC ratio → 30,000 laid off across two 2025.10–2026.01 rounds`

Amazon was the company that made "flattening" the most engineered and most quantifiable: a hard numeric target for every S-Team organization, plus a "bureaucracy mailbox" and a culture reset. But from H2 2025, its own actions broke the "flattening without layoffs" narrative — two corporate layoff rounds totaling roughly 30,000 people, in direct conflict with the earlier line.

+15% IC : manager ratio-increase target (by Q1 2025)	**1,500 / 455** bureaucracy-mailbox complaints / process changes shipped
~30,000 combined corporate layoffs, 2025.10—2026.01	**1,800+** engineers among them (per WARN filings in four states, 2025.11)

2024.09

Jassy's internal letter: **every S-Team organization must raise its individual-contributor-to-manager ratio by at least 15% by Q1 2025**, alongside a return to five days in the office.

His words: "Fewer managers will remove layers and flatten organizations more than they are today."

2024

Created the **"bureaucracy mailbox"**: employees can report pointless process and meetings directly. By September 2025 it had received about **1,500 complaints and driven 455 process changes**.

Why: Jassy said he wants Amazon operating like "the world's largest startup".

2025.01

Leaked documents showed Amazon **hit the 15% target not through layoffs but by giving managers more reports** (per Entrepreneur).

That claim held in early 2025, but was broken by the two layoff rounds from 2025.10 — "flattening without layoffs" is no longer the whole story at this company.

2025.06–09

Published a culture memo tying performance reviews directly to the 16 Leadership Principles; in his June memo Jassy forecast AI-driven efficiency gains and a leaner corporate workforce.

2025.10.28 – 2026.01.28

Two rounds totaling roughly 30,000 corporate layoffs — among the largest in company history: about 14,000 roles on October 28; per CNBC's November 2025 analysis of WARN filings in four states (covering only about 4,700 people), **more than 1,800 engineers** were cut in that round, far beyond "middle managers only"; another ~16,000 followed on January 28, 2026. Notably, pressed by analysts on the October 30 Q3 earnings call, Jassy was explicit that the layoffs were **"not really financially driven, and it is not really AI driven right now... it is culture"** — attributing them to the organization being "too big, too many layers", not to "AI replacing people".

2026.07

AGI-org layoffs: on July 22 CNBC reported cuts to the Nova model team (headcount undisclosed); Business Insider then reported the AGI Lab (the Adept team absorbed in 2024) was shut down, the Nova flagship model moved to maintenance mode, and Pieter Abbeel now leads a new frontier-model team. The AGI org had already been folded into Peter DeSantis's SVP organization in 2025.12, with former head Rohit Prasad departing.

Beyond "compressing layers", Amazon's AI organization itself is being rapidly reshuffled.

AWS side

In 2024.06 Matt Garman succeeded Adam Selipsky as AWS CEO, followed by sweeping leadership changes across global sales, channels and the generative-AI organization.

TENSION BETWEEN NARRATIVE AND ACTION

Amazon was once the positive sample for "flattening by widening spans, without layoffs". The two rounds of 2025.10—2026.01 show the reality is messier: the company ultimately did turn to mass layoffs, and cut not just middle managers but more than 1,800 engineers. Amazon's official line decouples the layoffs from "AI" and "saving money", attributing them to an abstract "culture" — an explanation that still has to be read separately from the actual scope of the cuts.

SOURCES [About Amazon (official): Jassy's original memo] [CNBC: the endangered middle manager] [Entrepreneur: target met without layoffs (the early-2025 version)] [Fortune: the anti-bureaucracy scorecard] [About Amazon (official): AWS changes CEO] [CRN: Garman's leadership changes] [CNBC: 14,000 layoffs (2025-10-28)] [CNBC: 1,800+ engineers cut (2025-11-21)] [GeekWire: Jassy's "it is culture" quote (2025-10-30)] [CNBC: another 16,000 cut (2026-01-28)] [Seattle Times: a historic wave of layoffs] [CNBC: AGI-org layoffs (2026-07-22)]

X (formerly Twitter)

`Musk: an extreme-flattening testbed for "engineer sovereignty"`

X is the most radical sample of this cycle: first physically delete the organization (roughly 80% of employees gone), then rebuild it as an extreme-flat, engineer-driven structure; merged into xAI in 2025 and, with xAI, into SpaceX in 2026 — fully AI-native.

~80%
cumulative headcount reduction (2022.10—2023.04, Musk's own BBC figure)

7,500 → ~1,500
headcount change (cross-verified by multiple outlets)

2022.10-11

After the takeover Musk cut about 50% of staff, then issued the famous "extremely hardcore" ultimatum: accept "long hours at high intensity" or leave — hundreds left, and headcount fell from about 7,500 to about 1,500 (roughly 80% cumulative by 2023.04, Musk's own figure in a BBC interview).

The result: the management layer and middle processes were removed essentially wholesale.

2022-2023

Rebuilt as "Twitter 2.0 / X": multiple observers noted the new organization is run by "the people who write code", with engineers reporting directly to Musk and product decisions made in the open over Slack/email. Synthesizing The Verge's and Fortune's reporting, X in this period no longer had an org chart in the traditional sense — engineers, product teams and Musk himself shared one information channel, and whoever owned something simply claimed the task directly.

Why: Musk regards the management layer itself as information loss.

2025.03.28

X was **acquired by xAI in an all-stock deal (valuing it around $33 billion)**, folding the social platform and the AI lab into one.

2025.07.09

CEO **Linda Yaccarino resigned** — she had been the only executive buffer since the takeover; Musk then took direct control, pulling sales, product and engineering together. The announcement came **the day after** the controversy over the Grok chatbot's inflammatory outputs, but per people familiar speaking to NBC News, the decision "had been brewing for over a week" and was not directly caused by the Grok incident.

The consensus across outlets: X had cut roughly 80% of staff in the Musk era, and Yaccarino's exit marked the end of the "two-headed" structure.

2026.02–08

On February 2 SpaceX acquired xAI in an all-stock deal (X included); on February 11 Musk announced an xAI reorg, saying the first build "wasn't built right" and would be "rebuilt from the foundation" — 6 of the 12 co-founders had already left; in June the merged SpaceX completed its IPO; on August 5 X product head Nikita Bier departed with no successor named.

X is no longer an independent-company sample but a product line inside the SpaceX—xAI complex.

SOURCES [The Verge: the hardcore-Twitter deep dive] [The Verge: the hardcore email, in full] [CNBC: xAI acquires X (2025-03-28)] [NYT: Yaccarino departs] [NBC News: resignation predates the Grok furor] [Reuters: departures after the X/xAI merger] [CNBC: xAI reorg and co-founder exits (2026-02-11)] [TechCrunch: Nikita Bier departs (2026-08-05)]

Nvidia

`Jensen Huang: ~60 direct reports, almost no one-on-ones — "extreme co-design"`

Nvidia is the most valuable company of the AI era and the most thorough sample of flat structure — an organization that did not "transform" into this shape but has always run this way, now studied industry-wide as the template.

~60 direct reports to the CEO (his own figure; media accounts 36—60)	**≈0** routine one-on-ones (HBR's version: "rarely")
Top 5 employees' ad-hoc "top 5 things" emails straight to the CEO	**No long-term plan** rolling forecasts replace the annual plan

Current structure

Huang says he has **about 60 direct reports** and does almost no one-on-ones, preferring group discussions with everyone present and "top 5 things" emails. The figures are not consistent: Fortune (2024.11) used "60"; he himself said "60, maybe more" on the Lex Fridman podcast in 2026.03 and "as many as 55" in an earlier Stanford talk; a Business Insider report in 2025.10, citing an internal list, showed 36 direct reports — the number drifts with reorgs and counting method, and this report uses his own ~60.

His explanation: "The more direct reports a CEO has, the fewer layers are in the company." Information flows transparently within the group instead of being filtered layer by layer.

2026.03.23

On Lex Fridman podcast #494, Huang defined his method as **"extreme co-design"**: "No conversation has just one person in it. So I don't do one-on-ones — we pose a problem and everyone solves it together." Earlier, in a February 2025 Fortune interview, his version was: "They never hear anything I say to only one of them."

Management philosophy

"No long-term plans, no fixed job duties" — he asks everyone to "move at the speed of light", with the organization designed around information flow rather than rank.

Industry influence

Across 2024—2026, Huang's "no 1:1s, many direct reports" model became the reference point for AI-era CEOs — at Stripe Sessions in 2025.05, host John Collison unprompted compared Zuckerberg's "25—30 person core team" to the Huang model, while Zuckerberg stressed the two are not the same.

SOURCES [Fortune: the Jensen Huang leadership deep dive] [HBR IdeaCast (official): the 2023.12 interview] [Entrepreneur: why no 1:1s] [Tom's Hardware: no long-term plans] [Yahoo Finance: the 55 figure]

Meta

`MSL merged then split + three rounds of layoffs, ~10,000 people, in 2026`

Meta was the first mover of this Great Flattening — and the best evidence that merging organizations is never one-and-done: the top-level AI team it created in 2025 split again after just 7 weeks, followed by three rounds of contraction in H1 2026.

7 weeks
MSL from formation to split (2025.06.30 → 2025.08)

+220% / +36%
code changes vs user-visible new features during Project OT (year over year)

+40%
major technical / security incidents over the same period (year over year)

~21,000
cumulative layoffs 2022.11—2023.05 (10,000 in the Year of Efficiency + 11,000 before)

2023.03

Zuckerberg's open letter: **"Flatter is faster"** — the Year of Efficiency would "remove multiple layers of management" and ask many managers to become individual contributors (ICs); another 10,000 laid off and about 5,000 open roles closed.

2025.06.30

Zuckerberg's internal memo announced **Meta Superintelligence Labs (MSL)**, consolidating model development, product teams and fundamental research (FAIR); former Scale AI CEO Alexandr Wang (28) became Meta's first Chief AI Officer, and former OpenAI researcher and ChatGPT co-creator Shengjia Zhao was named MSL Chief Scientist on July 25 of the same year.

2025.08.19

Just 7 weeks later, MSL split again into four teams: TBD Lab (frontier models), FAIR (fundamental research), Products and Applied Research, and MSL Infra; two months after that, on October 22, Meta cut about 600 more AI roles across FAIR, product AI and infrastructure teams.

"Merge, then split" shows that merging by itself cannot guarantee collaborative efficiency — the split, equally, was about giving each team focus.

2026.01–05

Three consecutive rounds of contraction at Reality Labs: in January, about 10% of the division cut (multiple VR studios closed, budgets trimmed by up to 30%); in March, "hundreds" more across at least five units including Reality Labs, Facebook, recruiting and sales; on May 20, a company-wide round of about **8,000 people (roughly 10% of global staff)** — explicitly not aimed at low performers, but at reorganizing teams company-wide into AI-led "pods" (internal codename Project OT), with Chief People Officer Janelle Gale saying 7,000+ people were moved into newly created AI teams. A ~1,000-person Reality Labs developer-tools team piloted dropping traditional functional titles for three roles — **AI Builder / AI Pod Lead / AI Org Lead** — with AI systems assisting promotion evaluation.

2026.03.06

Created a new **Applied AI Engineering** organization reporting to CTO Andrew Bosworth, carving data and engineering functions out of MSL; the NYT reported friction between Alexandr Wang and Bosworth and Chris Cox, and Yann LeCun had left in 2025.11.

The third boundary redraw since MSL's merge-then-split.

2026.01.28

On the Q4 2025 earnings call, Zuckerberg told investors plainly: "Projects that used to take a big team can now be done by one very talented person," so the company is "elevating individual contributors and flattening teams".

2025.05

In conversation with Stripe president John Collison at Stripe Sessions, Zuckerberg said he directly runs a "core team" of about **25—30 people** (not a company-wide 60-person arrangement) and holds no regularly scheduled one-on-ones — adding "it is not that there are no one-on-ones, there are just no fixed-cadence ones; I am talking with these people all the time". The host, unprompted, compared this with the Jensen Huang model.

2026.08.26

A Reuters special report laid out **Project OT (Organization Transformation)** in full: drafted at a January 2026 leadership offsite, the plan envisioned AI taking over much of the work of thousands of employees, with smaller human teams supervising "virtual workers"; internal documents had assessed cutting some teams by up to 60%. Two waves were planned for May and November; Zuckerberg canceled the second late on May 19, so only the first round of about 8,000 was executed. Internal data showed AI-platform code changes up 220% year over year — but user-facing new features up only 36%, and major technical and security incidents up 40%.

This is currently the most important piece of counter-evidence on "AI-native reorganization": a surge in output volume did not convert proportionally into delivered value. The Reuters original is paywalled; cited here via Engadget / CTV and other secondary accounts.

> **"Flatter is faster. In our Year of Efficiency, we will make our organization flatter by removing multiple layers of management… ask many managers to become individual contributors."**
>
> — Mark Zuckerberg, open letter, 2023.03 (official)

SOURCES Economic Times: excerpts from the open letter | about.fb.com (official): Year of Efficiency | BI: the big-tech flattening overview | CNBC (with the official memo): MSL is formed | Axios: 600 AI roles cut in October | Built In: the MSL reorg, mapped | CNBC: Reality Labs January layoffs | CNBC: hundreds more cut in March | The Next Web: the May 8,000-person layoff | BI: the AI-native pod pilot | Meta IR (official): Q4 2025 earnings-call transcript | CFO Brew: the Stripe Sessions conversation | Fortune: the 25—30 person core team | Engadget (citing Reuters): Project OT stumbles (2026-08-26) | The Decoder: the new Applied AI Engineering org (2026-03)

Microsoft

`Nadella: 15,000+ roles cut in the name of "fewer layers, fewer managers"`

Microsoft's two big 2025 rounds (about 6,000 in May, about 9,000 in July — over 15,000 for the year) wrote org-theory language into the explanation: CFO Amy Hood's official phrasing was "fewer layers, fewer managers", while the internal target Business Insider cited was widening the span of control.

~15,300

2025 full-year layoffs (of ~228,000 employees, per the 2024.06 annual report)

3,200

Xbox layoffs, 2026.07 (~1,600 immediate)

14 → ≤5

Xbox management-layer compression target (GeekWire)

2025.05.13

About 6,000 laid off (under 3% of staff). BI, citing internal information, said the goal was precisely to **widen spans of control and cut management layers**; ironically, per the same report, the lists included several managers with only 1—2 reports — and later tallies showed a sizable share were frontline engineers.

2025.07.02

Another ~9,000 cut, taking the year past 15,000 (of roughly 228,000 employees). Xbox CEO Phil Spencer wrote in an internal memo that teams would "follow Microsoft's lead in removing layers of management" — the official line remained fewer layers and faster decisions, while acknowledging many frontline engineers were affected.

2025.10.01

Judson Althoff rose from EVP and Chief Commercial Officer to **"CEO of Microsoft's commercial business"**. Nadella wrote that this would let "our engineering leaders and me" focus on data-center buildout, systems architecture, AI science and product innovation. The promotion and the mid-year layoffs came from different documents and should be read as **concurrent but independent** adjustments, not cause and effect.

2026.02–07

On February 20 Xbox chief Phil Spencer retired and Asha Sharma became CEO of Microsoft Gaming; April brought the company's first-ever voluntary exit program (open to about 7% of US employees); on July 6 Xbox announced about 3,200 cuts within FY27 (about 1,600 that day) and the divestment of four studios, with Sharma saying "we spread ourselves too thin".

SOURCES BI: reading the flatten-management layoffs | BI: the second round in July | Directions on Microsoft: reading the Althoff promotion | CNBC: the May layoffs (2025-05-13) | CNBC: the July layoffs + Spencer's words (2025-07-02)

Microsoft (official): Althoff named commercial-business CEO (2025-10-01)

Fortune: Xbox CEO on the 3,200 cuts (2026-07-06)

Salesforce

`Benioff: the extreme case of an "AI agent workforce"`

Salesforce is the bluntest case of "AI directly changing headcount": the CEO personally quantified the share of work AI performs and cut the support team; engineering hiring had been frozen even earlier (2024.12).

30—50%
share of internal work done by AI agents (Benioff's figure)

9,000→~5,000
customer-support workforce (Benioff's own account)

4,000
roles confirmed cut in 2025.09

2025.02

Benioff announced **no new software engineers in 2025** — because Agentforce and related tools had lifted engineering productivity, with "engineering velocity incredible".

2025.06-07

Bloomberg/Fortune interviews: AI agents already perform **30%—50% of the work** inside Salesforce; sales roles, by contrast, are growing.

2025.08.29

On The Logan Bartlett Show, Benioff first said **"I need less heads"**: "I've reduced it from 9,000 heads to about 5,000, because I need less heads" — cutting the customer-support workforce from about 9,000 to about 5,000.

2025.09.02

CNBC confirmed **4,000 roles** cut, quoting the podcast line above; support tickets had fallen sharply thanks to agents, and support engineers were not backfilled.

Note: this widely circulated quote originates from the 08.29 podcast interview; CNBC's version is secondhand, not the primary source.

2026.02-08

Three smaller rounds: under 1,000 in February (marketing, product, analytics and the Agentforce team), then dozens-to-hundreds each in June and August (Agentforce, MuleSoft, Marketing Cloud); the filings did not cite AI as a reason. Agentforce annualized revenue passed $1 billion in May.

SOURCES [Fortune: the Benioff exclusive] [SF Standard: no engineering hires] [CNBC: the 4,000 layoffs confirmed (2025–09–02)] [The Next Web: the third 2026 round (2026–08–07)]

Block (Jack Dorsey)

`Behind the theory paper, a near–halving layoff`

Where the giants have been "doing", Jack Dorsey (Block co–founder and Block Head) and Roelof Botha (Sequoia Capital partner, Block's lead independent director) tried to turn it into theory — but the theoretical essay itself is best read as an **after–the–fact justification** of a near–halving reorganization, not a pure thought experiment.

~40%
share of staff cut in one round on 2026.02.26 (~4,000 people)

~18%
next–day stock move after the announcement

DRI
the DRI mechanism: named ownership, no standing management layer

2026.02.26

About a month before the essay, Block had already answered with action: a single round of roughly **4,000 layoffs, about 40% of all staff**, taking headcount from over 10,000 to under 6,000. Dorsey wrote to shareholders: "A much smaller team, using the tools we are building, can do more and do it better." The stock jumped over 20% after hours on announcement day and closed up about **18%** the next day.

2026.03.31

Dorsey and Roelof Botha published their co–authored essay **"From Hierarchy to Intelligence"**, carried simultaneously on Block's site and Sequoia's site. It traces bureaucracy back two thousand years to the **Roman legion** — the eight–man contubernium as the smallest unit — then through the Prussian general staff and the railroad companies of the 1850s to today's org chart; the core claim is that the middle layer's core function is information routing, AI routes better, and therefore middle management in the traditional sense can be abolished.

Block practice

The organization was rebuilt on the **"Directly Responsible Individual (DRI)"** model: three roles — IC / DRI / player–coach — where a DRI takes named ownership of cross–team problems and can pull in resources ad hoc, with no standing management layer; still rolling out as of 2026.08 (Q2 earnings call: "the reorganization is on track"). Some organization–design commentators worry the risk is "responsibility without authority" — a criticism with no named primary source aimed at Block specifically, recorded here as observation only.

2026.04–08

On April 2, on Sequoia's podcast, Dorsey said the ideal would be "no layers at all — all 6,000 people reporting to me", with a plan to compress management from 5 layers to 2—3 within a year; on August 5 the Q2 earnings call said the "intelligence–centric reorganization is proceeding to plan", per–engineer code changes were up 150% since the start of the year, Square shipped 130 features in H1 (more than 3x the same period last year), and full–year guidance was raised.

The mirror image of Meta's Project OT: what Block has published so far is positive data — but the measures differ (code–change volume vs features shipped) and cannot be compared directly.

Ripple effects

Forbes / Fortune read the essay alongside Amazon's and Meta's de–layering moves; but set against the 02.26 layoff timeline, the more accurate reading is that the essay is an **after–the–fact justification** of a large reorganization that had already happened, not a purely theoretical manifesto.

SOURCES
- Block (official): the essay
- Sequoia (official): the simultaneous publication
- Forbes: AI ends middle management (2026–03–31)
- Forbes: the three DRI roles, mapped (2026–04–01)
- Fortune: the layoff backdrop confirmed (2026–04–02)
- Fast Company: Dorsey's "6,000 direct reports" (2026–04)
- Yahoo Finance: Block Q2 2026 earnings–call highlights (2026–08–05)

Coinbase

`Armstrong: at most 5 layers below the CEO, "pure manager" role abolished`

Coinbase is the most aggressive layer–slimming case in the sample: its May 2026 reorg compressed the company to at most 5 layers below the CEO and abolished the "pure manager" job category outright.

≤5 layers

maximum layers below CEO/COO (from 2026.05, written into policy)

~14%

share laid off in 2026.05 (~700 people)

15+

direct reports a leader may carry (permitted, not required)

40%

share of day-to-day code that is AI-generated (Armstrong's own figure, 2025.09)

2025.08

Armstrong gave engineers one week to adopt AI coding assistants (GitHub Copilot / Cursor); those who did not were let go.

2025.09.03

Armstrong said on X that AI-generated code makes up about **40%** of Coinbase's day-to-day code, targeting over 50% by October 2025.

2026.05

Announced layoffs of about **14%** alongside a restructuring: the "pure manager" position was abolished in favor of the **"player-coach"** (managing a team while remaining an active IC); the hierarchy was compressed to **at most 5 layers** below CEO/COO, with leaders allowed 15+ direct reports (permitted, not required); some teams are piloting "AI-native pods" — one person directing multiple AI agents to do work that used to take an engineer, a designer and a product manager.

SOURCES Fortune: the full 2026.05 reorg (2026-05-05) | Fortune: the AI coding-assistant mandate (2025-08-25) | Armstrong's X post (official) | Fortune column: named criticism of the player-coach model (2026-05-07)

OpenAI

`A research/applications dual track that lasted only ~14 months`

OpenAI answered scale with a new CEO title while keeping a "dual-head" structure on the research side — but that architecture itself is iterating fast, with visible adjustments by 2026.

~14 months

how long the "CEO of Applications" structure lasted

13

executives departed in 2026 (TechCrunch roundup)

2025.03

The research side formalized "dual heads": Mark Chen became Chief Research Officer, co-leading with Chief Scientist Jakub Pachocki.

2025.05.07

Former Instacart CEO Fidji Simo took the newly created role of **CEO of Applications**, reporting directly to Sam Altman; COO Brad Lightcap, CFO Sarah Friar and CPO Kevin Weil, who had reported to Altman, now report to Simo. Altman used the change to concentrate on research, compute and safety, writing on X that the new setup would let him "increase focus on research, compute and safety".

2025.09

The ~14-person Model Behavior team (responsible since GPT-4 for shaping model "personality" and reducing sycophancy) was folded into the larger Post Training team; its lead, Joanne Jang, left to found **OAI Labs**, a new team focused on human-AI interfaces.

2025.09–2026.04

Kevin Weil moved from CPO to VP of AI for Science (announced 2025.09), driving the creation of the **OpenAI for Science** team; on 2026.04.17 Weil departed and the team was dissolved into the research groups.

2026.07.09

Fidji Simo, managing a chronic condition (POTS), stepped back from full-time "Applications CEO" to part-time advisor — the "research/applications dual-CEO" structure **lasted about 14 months.** OpenAI named no single successor, splitting the duties among Greg Brockman (product strategy), CFO Sarah Friar and Jason Kwon; COO Brad Lightcap had already moved to special projects in 2026.04 and left to found a company on 08.11.

SOURCES
- OpenAI (official): the Simo appointment
- CNBC: the Simo appointment covered (2025-05-07)
- Altman's X post (official)
- MIT Tech Review: research dual heads (2025-07-31)
- The AI Insider: Model Behavior folded into Post Training
- Cryptopolitan: supplementary coverage of the same event
- MIT Tech Review: OpenAI for Science (2026-01-26)

[CNBC: Simo moves to part-time advisor (2026-07-09)] [TechCrunch: same-day coverage]

[TechCrunch: Weil departs, OpenAI for Science dissolved (2026-04-17)]

[TechCrunch: Lightcap departs (2026-08-11)]

Anthropic

`No official headcount figure — a deliberately small "research-meets-engineering" team`

Anthropic publishes no official headcount and third-party estimates diverge widely; but the consensus across reports is that, relative to its revenue and model capability, the team is kept lean deliberately — not under cost pressure.

1 direct reports to the CEO (Fortune, single source)	**2,300—5,600** third-party headcount estimate range (no official figure)

Headcount

Sources disagree: Revelio Labs puts it between about 3,173 (2025) and 4,020 (2026.03); Fortune reported about 2,300 at end-2025; Tracxn gave about 5,600 in 2026.07. Anthropic publishes no official number, so this report presents the range "2,300—5,600".

2026.02.19

Claude Code lead Boris Cherny (who started Claude Code as a side project on the Anthropic Labs team in 2024.09) told Lenny's Podcast that Claude Code had grown to **about 4% of public GitHub commits**, with daily active users doubling in the month before the interview; the team deliberately thins functional boundaries — in his words: "On the Claude Code team, everyone writes code. Product managers write code, engineering managers write code, designers write code, folks in finance write code, data scientists write code."

2025.05

Judging that it could not sufficiently rule out elevated risk, Anthropic formally activated **ASL-3** protections.

2025.10.02

Rahul Patil was appointed CTO, with co-founder Sam McCandlish becoming Chief Architect; product engineering was folded into the infrastructure and inference teams.

2026.02.24

Published the fully rewritten **v3.0** of its Responsible Scaling Policy (RSP), adding two mechanisms — a "frontier safety roadmap" and "risk reports" — approved by the Responsible Scaling Officer (RSO) and CEO, overseen by the board and the Long-Term Benefit Trust; the dedicated team breakdown (Frontier Red Team, Alignment Science, Trust & Safety, RSP Team) is in the 2024.10 v2.0 announcement.

A rare case of an AI company adding process to itself — "safety" built out as a concrete division of teams, not a slogan.

2026.06.18

Fortune reported Anthropic's unconventional reporting structure: CEO Dario Amodei has **exactly 1 direct report (a Chief of Staff)**, with every other executive reporting to President Daniela Amodei.

The opposite end of the spectrum from Nvidia's "60 direct reports" — showing there is no single answer to "flat" even inside AI companies. Fortune, single source.

SOURCES

- Revelio Labs: headcount estimates
- SaaStr: "deliberately lean" — revenue per head 6–10x Google's (2026-04)
- Lenny's Newsletter (industry interview publication): transcript of the Boris Cherny interview
- Anthropic (official): RSP v3.0
- Anthropic (official): the RSP v2.0 announcement (2024-10, with team breakdown)
- Fortune: Dario Amodei's single direct report (2026-06-18)
- TechCrunch: new CTO Rahul Patil (2025-10-02)

CH.05 · COMPARISON

Side-by-Side Comparison

All called "flattening" — but the path, the rhetoric and the cost differ company to company.

KEY TAKEAWAYS

01 Under the same "flattening" label, the companies' quantified targets are mutually incomparable: Google counts "managers with fewer than 3 direct reports", Amazon counts the IC ratio, Coinbase counts the number of layers.

02 The "typical rhetoric" column shows the different explanations each company gives for the same move — efficiency, culture, speed, AI — the rhetoric itself is information that needs decoding.

03 The evidence-strength column at the end is a filter to apply before quoting: official/firsthand statements can be checked directly; single reports and analytical inference are directional at best.

FIGURE 6 · FOUR VERIFIABLE METRICS OF MANAGEMENT-ROLE CUTS

The four companies count four different things — the numbers cannot be added or ranked; Microsoft has only the "fewer layers" language and no published management-cut figure, so the chart shows the four metrics the other three have published. Together they still point the same way: flattening is turning from slogan into verifiable metric.

manager / director / VP roles

−10%

Google · cumulative 2022.09—2024.12

Pichai all-hands, via BI

Managers with fewer than 3 reports

−35%

Google · 2024.08—2025.08

Brian Welle all-hands, CNBC

IC : manager ratio

+15%

Amazon · target by Q1 2025

Jassy memo (official)

Max layers below CEO/COO

≤5 layers

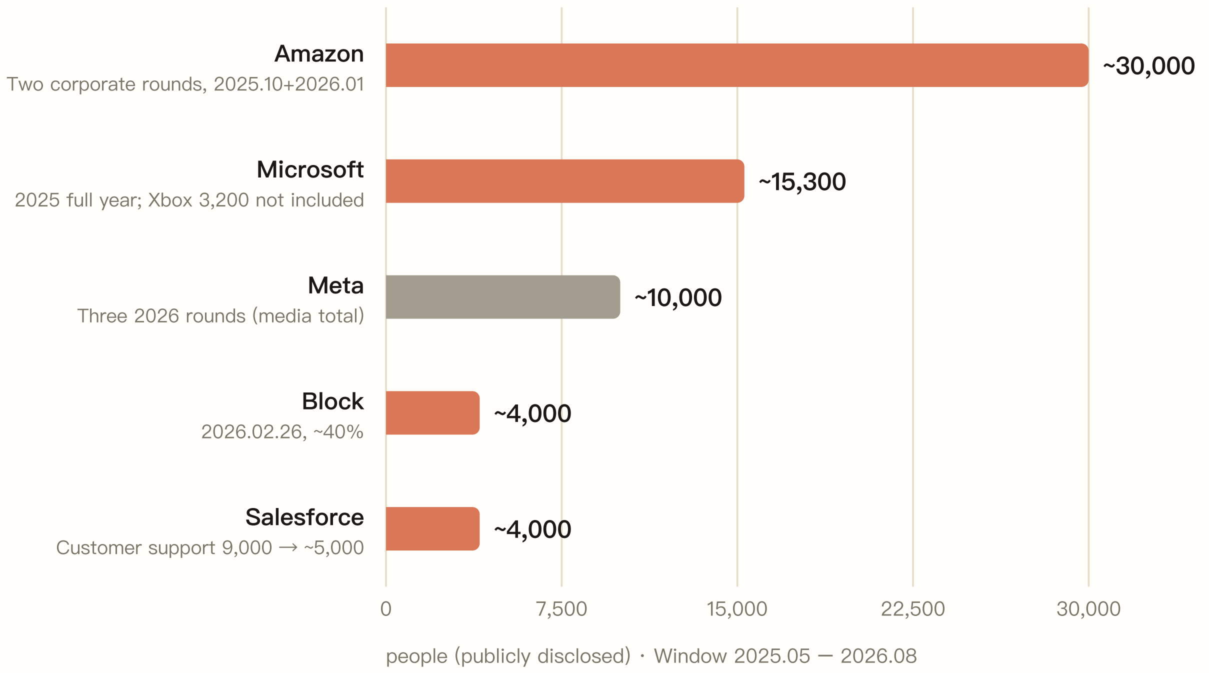
Coinbase · below CEO/COO, from 2026.05

Fortune

Pichai / Welle all-hands (via BI, CNBC) · Jassy memo (official) · Fortune 2026-05-05.

FIGURE 7 · THE "MANAGEMENT ROLES ONLY" LINE IS BROKEN: PUBLISHED LAYOFF SIZES

The chart lists only publicly disclosed absolute headcounts; total workforces differ widely, so bar length does not indicate "aggressiveness".

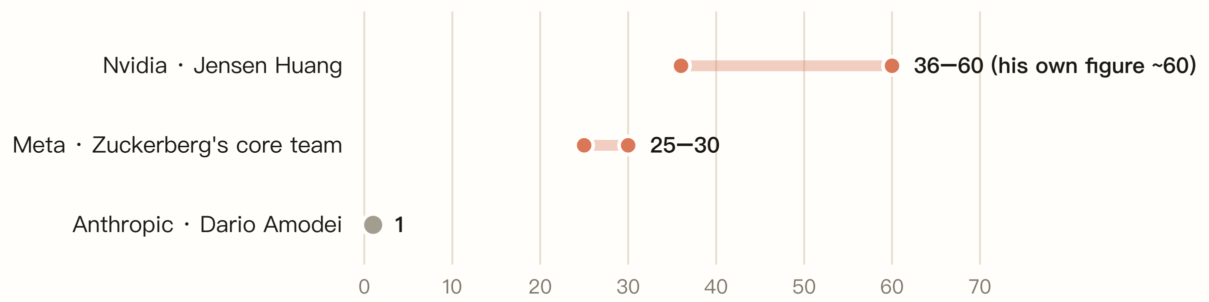


CNBC / Fortune / Forbes et al.; window 2025.05–2026.08.

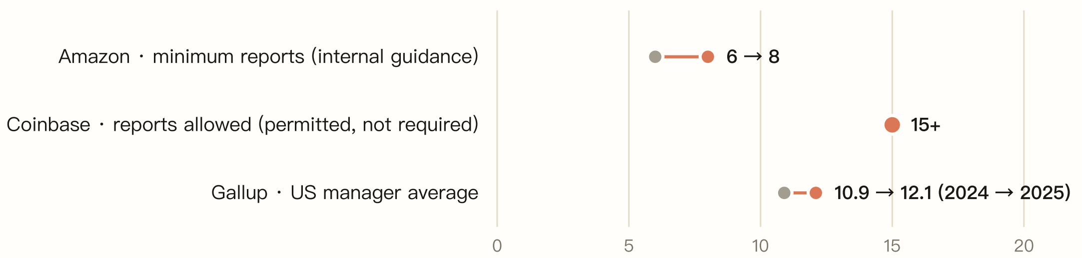
FIGURE 8 · EVERY MANAGER RUNS MORE PEOPLE: BOTH THE AVERAGES AND THE POLICY FLOORS ARE RISING

Dozens of direct reports to the CEO has become a template, and minimum direct-report counts are being written into policy; as spans widen, coordination costs shift onto new roles. Note Gallup's 12.1 is a mean pulled up by a few very large teams — the median is still 5–6.

Direct reports to the CEO



Managers' direct reports: policy floors, permitted ceilings and the industry average



Lex Fridman #494 · Fortune · Entrepreneur (BI's account of the AWS guidance) · Gallup 2026-01-13.

FIGURE 9 · 11 COMPANIES SIDE BY SIDE: PATHS, RHETORIC, COSTS AND EVIDENCE STRENGTH

Under the one label "flattening", the actual paths split four ways: structural-reorganization (Google, Meta, OpenAI), layoff-driven (Amazon, Microsoft, X, Block, Coinbase), AI-substitution (Salesforce), and always-flat (Nvidia, Anthropic). The last column is not decoration: every entry opens its source directly, and it flags which conclusions are cross-verified and which still rest on a single report.

COMPANY	CORE ACTION	QUANTIFIED TARGET / RESULT	SIGNATURE LINE	METHOD & COST	EVIDENCE STRENGTH
Google	AI teams merged (Brain+DeepMind, Gemini App → GDM) + management-role cuts + the 2026.08 leadership shake-up	manager/director/VP down **10%** (two years); the precise measure: managers with fewer than 3 reports down **35%**	"Fewer layers, team reorganization"	Mostly reorgs, with limited layoffs; risk: frequent reorgs breed morale swings	**Official / firsthand** ·Google Blog · 2026-08 ↗ management roles and AI-org changes; **Single report / dataset** ·CNBC (2025-08-27) ↗ managers with fewer than 3 reports down 35%
Amazon/AWS	IC:manager ratio +15%, bureaucracy mailbox → then two more layoff rounds 2025.10—2026.01	1,500 complaints → 455 process changes; still followed by ~**30,000** corporate layoffs	"Fewer managers = fewer layers"; later attributed to "culture", not AI	First flattened by widening spans, then laid off at scale anyway; cost: visible tension between narrative and action	**Official / firsthand** ·Amazon · Andy Jassy memo ↗ manager ratio and the anti-bureaucracy program; **Independently corroborated** ·CNBC (2026-01-28) ↗ · The Seattle Times ↗ layoff scale

COMPANY	CORE ACTION	QUANTIFIED TARGET / RESULT	SIGNATURE LINE	METHOD & COST	EVIDENCE STRENGTH
X / xAI	Physically deleted the organization: ~80% of employees gone, engineers reporting straight to Musk	7,500 → about 1,500 people	"extremely hardcore"	Extreme compression; costs: ad-revenue decline, brand and compliance risk	**Independently corroborated** ·The Verge ↗ · CNBC (2025-03-28) ↗ people and org changes; no official org-structure account
Nvidia	Always flat: ~60 direct reports, almost no 1:1s, no long-term plans	**~60** direct reports to the CEO (his own figure; media accounts 36—60)	"Move at the speed of light" / "extreme co-design"	Culture-borne flatness; cost: high intensity, attrition of those who cannot adapt	**Official / firsthand** ·Lex Fridman Podcast #494 ↗ · HBR IdeaCast · Jensen Huang ↗ Huang's own interviews; direct-report counts vary by source (36—60)
Meta	MSL merged then split, three Reality Labs rounds, the AI-native pod pilot	~**21,000** laid off in 2023; ~**10,000** more across three rounds in 2026	"Flatter is faster" / "one very talented person is enough"	Mass layoffs + repeated reorgs; structural stability in doubt	**Official / firsthand** ·Meta · Year of Efficiency ↗ · Meta IR · Q4 2025 transcript ↗ layoffs and org changes; **Single report / dataset** ·Engadget (Reuters) · 2026-08-26 ↗ Project OT (secondary accounts of a single Reuters report)

COMPANY	CORE ACTION	QUANTIFIED TARGET / RESULT	SIGNATURE LINE	METHOD & COST	EVIDENCE STRENGTH
Microsoft	15,000+ laid off across the year in the name of "span of control"	2025 layoffs **15,000+** / 228,000 employees	"Fewer layers, fewer managers"	Real layoffs reaching IC engineers; accused of cost-cutting under the flattening banner	**Independently corroborated** ·CNBC (2025-05-13) ↗ · CNBC (2025-07-02) ↗ layoff scale; **Single report / dataset** ·Business Insider (2025-05) ↗ the internal "span of control" target
Salesforce	Froze engineering hiring; agents replacing support roles	AI does **30—50%** of internal work; customer support down from 9,000 to about 5,000	"I need less heads"	Headcount tracks agent capacity directly; sales roles actually growing	**Official / firsthand** ·Fortune · Marc Benioff interview ↗ the CEO's own interviews; the work-share figure is his own account
Block	One big layoff in 2026.02 → the theory essay in 2026.03	~**40%** laid off (~4,000 people); stock closed up ~18% the next day	"From Hierarchy to Intelligence"	Layoffs first, theory after; the DRI "responsibility without authority" worry has no named source yet	**Official / firsthand** ·Block · From Hierarchy to Intelligence ↗ the company essay; **Independently corroborated** ·Forbes (2026-04-01) ↗ · Fortune (2026-04-02) ↗ the layoff and reorg backdrop

COMPANY	CORE ACTION	QUANTIFIED TARGET / RESULT	SIGNATURE LINE	METHOD & COST	EVIDENCE STRENGTH
Coinbase	Abolished the "pure manager" role in favor of player–coach	~**14%** laid off; hierarchy compressed to **≤5 layers** below the CEO	"player–coach" + AI–native pods	The most aggressive layer–slimming in the sample; cost: dense compression of frontline staff	**Official / firsthand** ·Brian Armstrong · X ↗ the CEO's own statements; **Single report / dataset** ·Fortune (2026–05–05) ↗ reorg details
OpenAI	Created a CEO of Applications; dual heads on the research side	COO/CFO/CPO moved to report to the new CEO	"Increase focus on research, compute and safety"	The structure lasted only ~**14 months** before changing; no single successor on the applications side	**Official / firsthand** ·OpenAI · Fidji Simo announcement ↗ · Sam Altman · X ↗ the appointment and reporting lines; **Independently corroborated** ·CNBC (2026–07–09) ↗ · TechCrunch (2026–07–09) ↗ the later adjustments

COMPANY	CORE ACTION	QUANTIFIED TARGET / RESULT	SIGNATURE LINE	METHOD & COST	EVIDENCE STRENGTH
Anthropic	Research–engineering integration + RSP–institutionalized safety process	headcount range **2,300—5,600** (no consistent measure); the CEO has exactly 1 direct report	—	Deliberately small teams; cost: doubts about whether scale can keep growing	**Official / firsthand** ·Anthropic · RSP v3.0 ↗ · Lenny's Podcast · interview transcript ↗ safety process and team practices; **Independently corroborated** ·Revelio Labs ↗ · SaaStr (2026-04) ↗ headcount estimates; **Single report / dataset** ·Fortune (2026-06-18) ↗ the CEO reporting structure

CH.06 · OPERATING PRACTICES

The Real Practice: How AI-Era Companies Actually Run

Removing layers does not automatically make an organization faster. **How decisions travel, who gets to make the first draft, where the rules live, and who owns verification and release** all have to be redone together. "Shipping" here does not mean bypassing professional review: it means the person who understands the problem best uses agents to produce a working version first, with product, design and engineering then reviewing, verifying and releasing.

KEY TAKEAWAYS

01 Decision chains get shorter by moving communication into collective settings (Nvidia, Meta's core team) — not by communicating less.

02 The right to make the first draft is shifting to whoever understands the problem best. After interviewing a dozen-plus fast-growing startups, Anthropic distilled the practice as *Everyone ships*; professional review and controlled release still follow.

03 Rules move out of managers' heads into shared environments (convention files, Skills, design systems); with the "pure manager" abolished, those who remain must build.

QUESTION 1 · HOW THE DECISION CHAIN GETS SHORTER

Put communication in collective settings instead of filtering it layer by layer

Nvidia: Huang describes ~60 direct reports and no routine one-on-ones — "no conversation has just one person in it; we pose a problem and everyone solves it together — that is extreme co-design". Official / firsthand · Lex Fridman Podcast #494 ↗

Meta: Zuckerberg directly runs a core team of about 25—30, with "no fixed-cadence one-on-ones; I am talking with these people all the time". Official / firsthand · Stripe Sessions 2025 ↗

Google: from 2026.08, Gemini models, frontier research, the Gemini App and developer teams sit under one SVP reporting directly to Pichai. Official / firsthand · Google Blog · 2026-08 ↗

Coinbase: at most 5 layers below CEO/COO, with leaders allowed 15+ direct reports (permitted, not required). `Single report / dataset · Fortune (2026-05-05) ↗`

Block: Dorsey calls the ideal "no layers at all — all 6,000 people reporting to me", planning to compress from 5 layers to 2—3 within a year.

`Official / firsthand · Sequoia Podcast · transcript · 2026-04 ↗`

Anthropic (the counterexample): CEO Dario Amodei has exactly 1 direct report (a Chief of Staff), with all other executives reporting to President Daniela Amodei.

`Single report / dataset · Fortune (2026-06-18) ↗` The common thread: what gets shortened is not headcount but the number of hand-offs between the front line and the decision-maker; "flat" by itself is not the only answer.

QUESTION 2 · WHO GETS TO MAKE THE FIRST DRAFT

Give 0 → 1 prototyping to whoever understands the problem best

AI-native startups: an official guide on Claude's site, written by Michael Segner from interviews with a dozen-plus fast-growing startups, distills "the person closest to the problem ships the first version" as working principle number one: *Everyone ships*. Crosby's lawyers edit the product directly, Parahelp's non-technical staff submit UI changes, and Heidi's CEO frames it as solving the "broken telephone" problem. `Official / firsthand · Claude by Anthropic · 2026-08-20 ↗`

Anthropic's Claude Code team: lead Boris Cherny's words — "everyone on the team writes code: product managers, engineering managers, designers, finance, data scientists".

`Official / firsthand · Lenny's Podcast · interview transcript ↗`

Coinbase: some teams pilot AI-native pods — a few people plus multiple agents.

`Single report / dataset · Fortune (2026-05-05) ↗`

Meta Reality Labs: the ~1,000-person developer-tools team is also reported to be piloting AI-native pods. `Single report / dataset · Business Insider (2026-03) ↗`

Meta Project OT: per Reuters, the AI-native reorg had assessed cutting some teams by up to 60% and its second wave was canceled when output fell short; internal data showed code changes +220% year over year, but shipped features only +36%, and major incidents +40%.

`Single report / dataset · Engadget (Reuters) · 2026-08-26 ↗` Writing more code is not delivering more value — which is exactly why "trust, but verify" exists.

QUESTION 3 · WHERE THE RULES LIVE

Move the norms out of managers' heads into a shared environment

AI-native startups: invariant architecture rules and safety boundaries go into CLAUDE.md (a project convention file the agent reads at every start); team practices are shared as Skills (reusable process instructions); MCP (an open protocol connecting tools and data) or CLIs plug agents into real systems. Official / firsthand · Claude by Anthropic · 2026-08-20 ↗

Amazon: the 16 Leadership Principles are tied directly to performance reviews; the "bureaucracy mailbox" logged ~1,500 complaints and drove 455 process changes.

Independently corroborated · Fortune (2025-09-17) ↗ · CNBC (2025-09-16) ↗

Anthropic: Responsible Scaling Policy (RSP) v3.0 writes the safety process into institution — RSO and CEO approval, board and Long-Term Benefit Trust oversight.

Official / firsthand · Anthropic · RSP v3.0 ↗

Coinbase: CEO Brian Armstrong said the company gave engineers one week to adopt AI coding tools. Official / firsthand · Brian Armstrong · X ↗

QUESTION 4 · WHERE THE MIDDLE LAYER GOES

The "pure manager" is abolished; those who stay must build

Google: managers with fewer than 3 direct reports fell 35% in a year, many moving to non-managing specialist roles (Individual Contributor, IC).

Single report / dataset · CNBC (2025-08-27) ↗

Block: deleted the word "manager" from the hierarchy in favor of three roles — IC, Directly Responsible Individual (DRI), and the ship-while-leading player-coach; a DRI takes named ownership and can pull in resources ad hoc, with no standing management layer.

Official / firsthand · Block · From Hierarchy to Intelligence ↗

Single report / dataset · Block Q2 call / Yahoo Finance · 2026-08-05 ↗

Coinbase: abolished the "pure manager" role; managers must also carry real delivery work.

Single report / dataset · Fortune (2026-05-05) ↗

Meta Reality Labs: the ~1,000-person developer-tools team is reported to be piloting AI Builder / AI Pod Lead / AI Org Lead. Single report / dataset · Business Insider (2026-03) ↗

Across companies: Chief of Staff postings more than doubled in a year, with salaries reaching up to $400K — though Ladders puts the more common high end at $250–300K, with only occasional cases above. `Single report / dataset · Bloomberg (2026-07-16) ↗`

FIGURE 10 · HOW "THE PERSON CLOSEST TO THE PROBLEM SHIPS FIRST" CHANGES THE DELIVERY PATH

The traditional delivery chain transcribes the problem from role to role; the AI-native path lets the person who understands the problem best produce a working version first, which professional teams then review, complete and release. What shrinks is the distance from idea to verifiable value — not the professional division of labor.

TRADITIONAL DELIVERY CHAIN

01 **Problem owner** States the problem and context

02 **Product / design** Turned into a PRD and prototype

03 **Engineering team** Scheduling, build and integration

04 **User feedback** Understanding verified after launch

AI-NATIVE PATH

01 **Person who knows the problem** Ships a working first version via agents

02 **Shared execution environment** Convention files, Skills, tools and the design system

03 **Professional review** Product, design and engineering fill the gaps together

04 **Controlled release** Tests, review and permission gates

> How to read this figure: left to right there are still four gates — the change is only at the start: the first draft goes from "transcribed requirements" to "a working version, directly". Professional review and controlled release remain.

Source: Anthropic, "The Claude Code guide for startups" (2026-08-20), based on interviews with a dozen-plus fast-growing startups; the diagram and the PMF analogy are our analytical synthesis, not an official figure or Anthropic's words.

PRINCIPLE 1 · THE PERSON CLOSEST TO THE PROBLEM SHIPS FIRST

Everyone ships: opening up 0 → 1, not abolishing specialization

After interviewing a dozen-plus fast-growing startups, Anthropic distilled the practice as *Everyone ships*: whoever understands the problem best can use agents to produce a working version first; product, design and engineering review still follows.

PRINCIPLE 2 · MECHANICAL WORK GOES TO AGENTS

Hand the mechanical 80% to agents

"Agents take over the mechanical 80% of the lifecycle, and engineers spend their time on the cases that genuinely need judgment." First-pass code review, bug triage and routine migrations belong in this bucket.

PRINCIPLE 3 · TRUST, BUT VERIFY

Trust, but verify

"You cannot automate a process unless you have reliable means to monitor and verify the outcome." Tests, code review and permission gates are not abolished — they become the precondition for decentralized delivery.

PRINCIPLE 4 · DESIGN FOR REBUILDING AT ANY TIME

Build for rebuilding

"Model capability keeps shifting under these teams' feet, so almost nothing is treated as permanent." Processes and tools are designed to be redone at any time, not to be done right once.

PRINCIPLE 5 · PROTOTYPE → INTERNAL USE → PRODUCTIZATION

Prototype → dogfood → productize

"Building with AI is how these startups make disruptive AI products — the flywheel at the core of their process." Get the idea running, use it yourselves first, then decide whether to productize.

⚠ SCOPE: A FRONTIER SAMPLE, NOT A UNIVERSAL PLAYBOOK

These five principles come from Anthropic's interviews with a dozen-plus fast-growing AI-native startups — and Anthropic is itself the tool vendor. They are worth studying as frontier practice, **not adopting wholesale as policy for large enterprises**; among the 11 companies compared here, verifiable day-to-day operating cases remain limited, and the setback of Meta's Project OT shows that faster delivery must ship together with verification and governance.

CH.07 · WHERE DID THEY GO

Where Did Middle Managers Go: Five Paths

Once the layers are gone, where did the people go? Combining official accounts from Google / Meta / Amazon / Coinbase with reporting from BI / CNBC / Guardian / Fortune, we group the publicly traceable destinations of former middle managers into five paths. One distinction first: whether a capability still has value, and whether a role or a person survives a reorganization, are not the same question.

KEY TAKEAWAYS

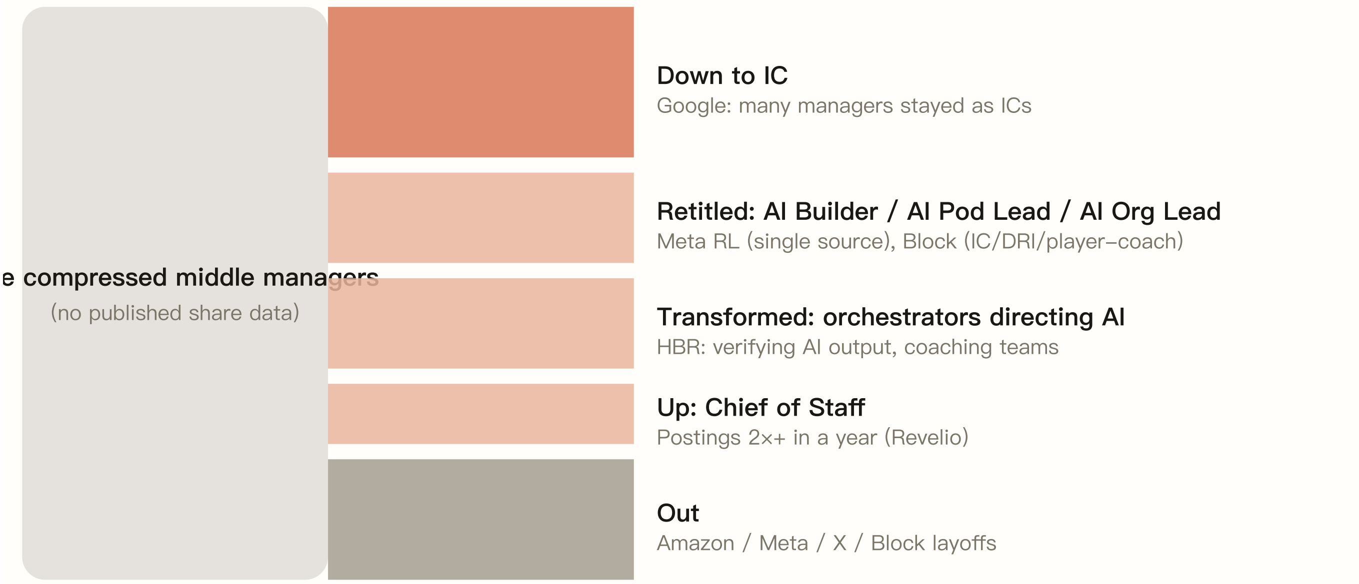
01 What gets compressed first is the middle layer's "information routing": pure relaying is the easiest to replace, while frontline managers who understand the business and its customers, make judgment calls and execute personally remain scarce.

02 The five paths are not mutually exclusive exits but a sorting funnel: the same managers may go IC first, then double as player–coach, with a few entering the Chief of Staff track.

03 But who stays is not a pure test of ability: cost targets, business priorities, reporting relationships and internal politics all shape who survives a reorganization.

FIGURE 11 · THE COMPRESSED MIDDLE: FIVE DESTINATIONS FORM A SORTING FUNNEL

The five destinations are not mutually exclusive exits but a sorting funnel; for the individual, the test reduces to one question — is your core work "relaying" or "judging"?



None of the five paths has published share data; band width reflects our judgment of relative prominence, not statistics

CNBC 2025-08-27 · BI (Meta RL, single source) · Bloomberg 2026-07-16.

FIGURE 12 · FOUR QUANTITATIVE ANCHORS FOR WHERE THE MIDDLE LAYER WENT

35% and ≤5 layers describe the force with which layers were removed; $400K and 2× suggest the need for coordination did not disappear and may be concentrating into a few highly paid roles — but with no base or methodology disclosed, they stand only as a signal of rising demand. What was removed is the layer — the need to coordinate remains.

35%

Reduction in Google managers with fewer than 3 direct reports, many of whom stayed on as ICs (2025.08)

CNBC 2025-08-27

$400K

Ceiling of reported Chief of Staff salaries (Bloomberg; Ladders puts the common high end at $250—300K), with postings more than doubling in a year

Bloomberg 2026-07-16 · Revelio Labs

2×

Year-over-year growth in Chief of Staff postings — called "the bill for the Great Flattening" by organization-design commentators

Bloomberg 2026-07-16 · Revelio Labs

≤5 layers

Maximum layers below the CEO after Coinbase's 2026.05 reorg — with the pure-manager role abolished

Fortune 2026-05-05

Sources: CNBC 2025-08-27; Bloomberg 2026-07-16 · Revelio Labs; Fortune 2026-05-05

FIGURE 13 · FLATTENING IS NOT FREE: THREE NEW BILLS AND WHO PAYS THEM

Cutting layers is meant to save management cost — but a company that removes its middle layer must fill the coordination gap with tools and named-ownership mechanisms, or it is simply shifting the cost onto whoever remains.

What was cut: the middle layer's coordination and gatekeeping

Relaying up and down, summarizing and filtering, frontline coaching, layered gatekeeping — this work does not vanish with the role

→

BILL 1 · COORDINATION UNOWNED

Paid by high-salary coordinators

Cross-team coordination may concentrate in a few hands: Chief of Staff postings doubled in a year (2×+), salaries up to $400K (common high end $250—300K); managers run more people, frontline coaching shrinks

Revelio / Bloomberg 2026.07 · BI 2025.05

→

BILL 2 · RESPONSIBILITY WITHOUT AUTHORITY

Paid by the leaders who remain

The player-coach builds while leading — criticized by name in a Fortune column as "misunderstanding the value of good managers"; the Block DRI "responsibility without authority" worry has no named source yet, recorded as observation

Fortune 2026-05-07

→

BILL 3 · OUTPUT ≠ VALUE

Paid by delivery quality

In Meta's experiment code changes +220%, user-visible features only +36%, major incidents +40% — surging output bought no shipped value

Reuters 2026-08-26

"Who pays" is our synthesis, not official company language; the numbers here are shorthand for relative change, not same-basis comparison.

Revelio / Bloomberg 2026-07-16 · Fortune 2026-05-07 · Reuters 2026-08-26.

01

Demoted sideways into an IC

Google: when 10% of management roles were cut in 2024.12, a spokesperson was explicit that some "transitioned to individual contributor roles"; by 2025.8 the precise figure was out — managers with fewer than 3 reports down 35%, many staying on as ICs. **Meta**: Zuckerberg's Year of Efficiency language was literally to "ask many managers to become individual contributors". **Amazon**: internal guidance shows the transition often comes with more direct reports, fewer senior openings, even pay cuts — going IC is not a lateral move; it carries a real price.

COMPANY EVIDENCE

Google Single report / dataset · CNBC (2025-08-27) ↗

Meta Official / firsthand · Meta · Year of Efficiency ↗

Amazon Official / firsthand · Amazon · Andy Jassy memo ↗

02

New title, redefined duties

2025—2026 brought outright rewrites of the manager role: parts of **Meta Reality Labs** retitled to AI Builder / AI Pod Lead / AI Org Lead, organized into small pods by outcome rather than function; **Block** (Dorsey) deleted the word "manager" in favor of IC / DRI / player-coach (still rolling out as of 2026.08); **Coinbase** (Armstrong) ran the fullest version in May 2026 — about 14% laid off, the pure-manager role abolished, everyone moved to "player-coach", the hierarchy compressed to at most 5 layers below the CEO, and some teams piloting AI-native pods. The common thread: organizations no longer keep management roles that only coordinate and never deliver; managers must also build.

COMPANY EVIDENCE

Meta Single report / dataset · Business Insider (2026-03) ↗

Block Official / firsthand · Block · From Hierarchy to Intelligence ↗

Coinbase Single report / dataset · Fortune (2026-05-05) ↗

03

Becoming a Chief of Staff

Part of the middle layer did not disappear — it was moved next to the decision-makers after the slimming. Bloomberg (2026.07), citing Revelio Labs, reports Chief of Staff postings more than doubled over the prior year; salaries reach up to $400K, though Ladders puts the more common high end at $250—300K, with only occasional cases above. Organization-design commentators call the boom "the bill for the Great Flattening" — but the public reporting discloses no base, geography or precise growth rate: it shows demand for the role is rising, not that the rise was caused by AI or flattening, nor that these roles have effectively absorbed the coordination work that was removed.

COMPANY EVIDENCE

Chief of Staff Single report / dataset · Bloomberg (2026-07-16) ↗

04

From "coordinator" to "orchestrator"

For those who stay in management, the job itself changed: less relaying, summarizing and progress-chasing (now AI's work), more coaching, strategy translation and building the conditions for human-AI collaboration — an upgrade from coordinator to orchestrator. The ironic reality: the managers who were just cut are now the ones tasked with driving AI adoption — they are precisely the execution handle AI rollout depends on.

05

The role disappears; they leave

Google's "role eliminations", Meta's Year of Efficiency and its three 2026 rounds, X's ~80% reduction, Amazon's ~30,000 layoffs in 2025.10—2026.01 — all show that some roles had no transition exit and simply left the organization.

COMPANY EVIDENCE

Google Single report / dataset · CNBC (2025-08-27) ↗

Meta Official / firsthand · Meta · Year of Efficiency ↗

Amazon Independently corroborated · CNBC (2026-01-28) ↗ · The Seattle Times ↗

ANALYSIS

What's eliminated is the function, not the person

The five paths show that what gets compressed first is the middle layer's "information routing": relaying up and down, summarizing and filtering, chasing process.

Capabilities still scarce: understanding the business and its customers, cross-domain judgment, leading a team while executing personally.

Two things to keep separate: a capability having value does not make a role safe. Reorganizations are not pure tests of ability — cost targets, business priorities, shifts in power and reporting lines, and internal politics all shape who stays.

Being laid off cannot be read backwards as proof the capability had no value — nor the reverse.

THE PATHS ARE NOT MUTUALLY EXCLUSIVE EXITS

The same cut managers may go IC first, then double as player-coach, with a few eventually entering the Chief of Staff track; the five paths are better seen as one sorting funnel than five separate doors.

CH.08 · INDUSTRY VIEWS

How the Industry Reads This Shift

Beyond what companies themselves are doing: how analysts and academia read this shift.

KEY TAKEAWAYS

01 Gartner forecasts that through 2026, 20% of organizations will use AI to eliminate more than half of their current middle-management positions; Morgan Stanley treats organizational flatness as a modelable financial variable.

02 The shared judgment of management scholarship (HBR, Gallup) is that the middle role is being reshaped, not eliminated — and is now carrying the extra burden of verifying AI output and coaching teams.

03 The criticism is explicit: flattening taken too far stunts talent growth and weakens leadership development; the player-coach model has drawn named criticism.

FIGURE 14 · ONE ORGANIZATIONAL SHIFT, FOUR SCALES OF OBSERVATION

The four lenses answer different questions and do not form one causal chain: forecasts ask whether structure will change, capital markets price what the change is worth, management research tracks who catches the work, and the critics test whether flattening has gone too far.

OBSERVATION FRAMEWORK

One shift, four scales of observation

Do not blend forecasts, financial models, management research and criticism into one conclusion; first see what question each is answering.

01 · Macro forecast

Will org hierarchies change?

Gartner's forecast: through 2026, 20% of organizations will use AI to eliminate more than half of current middle-management positions.

02 · Capital markets

What is the change worth?

Morgan Stanley, using Amazon as the sample, modeled management compression as $2.1—3.6 billion of annual cost headroom.

03 · Management practice

Who catches the work that was removed?

HBR, Gallup and MIT TR Insights track manager load, coaching quality, and the new processes human–AI collaboration requires.

04 · The opposing check

Will flattening go too far?

BI, Fast Company, Fortune and the Guardian point to talent growth, leadership development and the limits of player–coach.

Gartner 2024.10 · Morgan Stanley / CNBC 2024.10 · HBR / Gallup / MIT Technology Review Insights · BI / Fast Company / Fortune / Guardian.

GARTNER

The structural forecast

"Through 2026, 20% of organizations will use AI to flatten their organizational structure, eliminating more than half of current middle management positions." (annual prediction, IT Symposium/Xpo, 2024.10) Widely cited by the Guardian, Forbes, Fast Company and others, it has become the industry's benchmark call.

MORGAN STANLEY

The financial lens

Taking Jassy's "IC ratio +15%" as the assumption, it modeled roughly 13,800 trimmable Amazon manager positions, freeing $2.1–3.6 billion a year (2024.10 research note, via CNBC) — the first time Wall Street treated "organizational flatness" as a modelable financial variable.

HBR / GALLUP / MIT TR INSIGHTS

Management scholarship and industry research

HBR's "How AI Is Redefining Managerial Roles" (2025.07) and "AI Adoption Is Overloading Your Middle Managers" (2026.06) share one judgment: the middle role is being reshaped, not eliminated, and now carries the extra burden of verifying AI output and coaching teams. Gallup's 2026.01 data shows the average US manager's direct reports rose from 10.9 to 12.1, while manager quality explains about 70% of the variance in employee engagement. MIT Technology Review Insights (commissioned research with Ema) argues that when the "actor" is an AI agent operating across systems at machine speed, both the organization and the technology stack designed for human workflows need rebuilding.

THE CRITICS

"The Great Flattening may have gone too far"

Business Insider's May 2025 overview "Big Tech is crushing middle managers. Some fear the great flattening has gone too far" warns that badly managed flattening stunts talent growth and weakens leadership development; Fast Company (2026.06) cites a Korn Ferry survey in which 41% of employees say their company has already cut layers, and marshals Gallup data to argue "the data says don't"; a Fortune column (2026.05) criticizes the player-coach model by name as "misunderstanding the value of good managers"; the Guardian (2026.05) records the view from managers who were cut. The rapid growth in Chief of Staff postings is what the org-design blog Business Model Analyst calls "the bill for the Great Flattening".

Conclusions & Outlook

KEY TAKEAWAYS

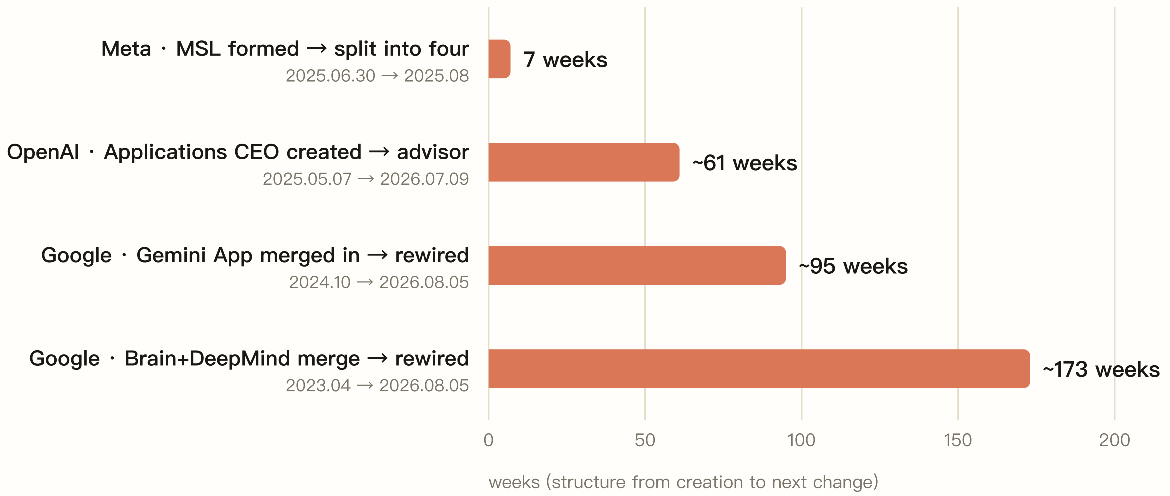
01 Routing-type management roles are being replaced by models and agents — a structural, not cyclical, change; but the new structures themselves keep iterating, with no stable end state.

02 The new architecture = fewer layers × wider spans × human–AI mixed teams; the costs are coordination vacuums and "responsibility without authority", and rising Chief of Staff demand is a signal that coordination cost may be relocating rather than disappearing.

03 For Chinese companies: build the information infrastructure before cutting layers; transform management roles rather than simply cutting them; use verifiable quantitative metrics to make flattening measurable and auditable.

FIGURE 15 · THE "NEW STRUCTURES" ITERATE FAST TOO: TIME FROM CREATION TO THE NEXT CHANGE, FOUR CASES

From creation to the next adjustment: 7 weeks at the shortest, under four years at the longest — all four "new structures" in the chart were later changed again.



CNBC (the MSL memo / Simo) · blog.google (official).

TAKEAWAY 1

Middle management is AI's first structural casualty

Google cutting 10% of management roles over two years and 35% of sub-3-report managers in one; Amazon's IC ratio +15%; Microsoft's 15,000+ layoffs; Gartner's "20% of organizations will cut over half their middle layer" — the evidence converges on one conclusion: routing-type management roles are being replaced directly by models and agents. This is not cyclical layoffs; it is a paradigm switch in organizational technology.

TAKEAWAY 2

New architecture = fewer layers × wider spans × human–AI blend

The emerging standard shape: 3—5 layers; dozens of direct reports to the CEO (the Nvidia / Meta core-team pattern); managers carrying 12—15+ reports (Gallup mean 12.1, Coinbase allowing 15+); AI agents embedded in teams (Salesforce's 30—50%). The minimal unit of "organization" shifts from "person/manager" to "person + agent cluster".

TAKEAWAY 3

Risk: a coordination vacuum and "responsible but powerless"

Flattening is not free: wider spans mean less frontline coaching (BI's "gone too far" warning); rising Chief of Staff demand signals coordination cost relocating rather than disappearing (base and methodology undisclosed, so not proof of causation on its own); player-coach / DRI models have drawn named criticism (the Fortune column: "misunderstands the value of good managers"), and Meta's Project OT shows surging code output does not necessarily convert into shipped value. A company that cuts its middle layer must fill the coordination gap with AI tools plus named-ownership mechanisms — or it is just shifting the cost onto individuals.

TAKEAWAY 4

The new structures are themselves iterating fast

No "new organizational form" is a stable end state. OpenAI's research/applications dual-CEO structure lasted about 14 months; Meta's MSL was split 7 weeks after formation, followed by three rounds of contraction in H1 2026; Amazon slid from "flattening without layoffs" to "two rounds, 30,000 people". Flattening is a continuous organizational experiment, not a destination reached once.

TAKEAWAY 5

Implications for Chinese enterprises

① Build the "information infrastructure" before cutting layers — Amazon backstopped with the bureaucracy mailbox plus process governance; ② transform management roles rather than simply cutting them: managers either move up into strategy/resource orchestration or down into coaching/cross-team coordination; ③ establish verifiable, auditable metrics of the management-role-ratio kind (see Google's precise "fewer than 3 direct reports" definition) so flattening is measurable and acceptance-testable rather than sloganized layoffs; ④ keep "motive" and "action" consistent in external communication — Amazon's official denial that its layoffs relate to AI or cost is a cautionary example: muddled attribution erodes public trust.

FOUR SELF-CHECK QUESTIONS FOR MANAGERS

From "cutting layers" to "changing how it runs"

① In your organization, how many roles have "relaying" rather than "judging" as their core work? ② Does the person who best understands the customer's problem have the authority and the tools, today, to produce the first working version? ③ Where are your rules written — in managers' heads, or in a shared environment everyone can invoke (convention files, the design system, the toolchain)? ④ Each time you remove a layer, do you install verifiable gates (tests, review, permissions) to replace the checking that layer used to do?

Video sources

20 publicly viewable videos/interviews, first-hand or close secondary material on each company's org changes — 12 carried over from the first edition (title/channel/date re-verified), 8 new in this revision.

(A) Firsthand executive interviews and official video · 11 items

Leading the teams behind the world's fastest AI — Jensen Huang

Sana (Strange Loop 2023) · 2023-06-29 Primary

Nvidia's management philosophy

Open on YouTube ↗

Jack Dorsey: Every Company Can Now Be a Mini-AGI

Sequoia Capital · — Primary

Block's DRI thinking

Open on YouTube ↗

Amazon CEO Andy Jassy on Agility, AI Strategy, and the Changing Role of Managers

Harvard Business Review (HBR IdeaCast) · 2025-05 Primary

Amazon's IC ratio and anti-bureaucracy reform

Open on YouTube ↗

How Sundar Pichai is rethinking Google for the AI era

Decoder with Nilay Patel · 2026-05-27 Primary

The rebuild of decision rights before Google's 2026.08 reorg

Open on YouTube ↗

The history and future of AI at Google, with Sundar Pichai

Stripe (Cheeky Pint, host John Collison) · 2026-04-07 Primary

Background on the evolution of Google's AI organization

Open on YouTube ↗

Google CEO Sundar Pichai on Gemini, Self-improving AI, and World Models

Forward Future Interviews (Matthew Berman) · 2025-05-23 Primary

The technical-organization logic after the DeepMind merger

Open on YouTube ↗

Sundar Pichai Reveals What AI Will Do Next

TIME · 2026-04-30 Primary

Google's positioning of AI's role

Open on YouTube ↗

AI will change the workforce, says Amazon CEO Andy Jassy

CNBC (Mad Money) · 2025-06-30 Primary

Amazon's position on AI and workforce structure

Open on YouTube ↗

The Leadership Principles Explained by Andy Jassy (full film)

Amazon official · — Primary

Background on the 16 Leadership Principles entering performance reviews

Open on YouTube ↗

Scott Answers: Is AI Replacing Middle Management? | Office Hours

Scott Galloway / Prof G Media · 2026-06-01 Opinion

The industry-critic perspective

Open on YouTube ↗

AI is changing the rules of business, see how to adapt from a Harvard Business School professor

Microsoft (guest: Prof. Karim Lakhani, HBS) · 2025-06-18 Interview

The academic perspective

Open on YouTube ↗

(B) Supplementary interviews and reporting · 8 items

Jensen Huang: NVIDIA — The $4 Trillion Company & the AI Revolution (#494)

Lex Fridman Podcast · 2026-03-23

Huang details the "extreme co-design" philosophy: no 1:1s, ~60 direct reports

Open on YouTube ↗

[Head of Claude Code: What happens after coding is solved | Boris Cherny]

Lenny's Podcast · 2026-02-19

How Anthropic's Claude Code team runs research-engineering as one

Open on YouTube ↗

[A conversation with Mark Zuckerberg]

Stripe Sessions 2025 (Stripe official) · 2025-05-07

Zuckerberg's 25—30 person core-team philosophy, compared with the Huang model

Open on YouTube ↗

[Dario Amodei — "We are near the end of the exponential"]

Dwarkesh Patel · 2026-02-13

Anthropic's scaling logic: compute investment vs team size

Open on YouTube ↗

[Amazon targets middle managers in mass layoffs, memo suggests more cuts coming as AI thins Big Tech]

CNBC · 2025-10-28

How Amazon's 2025.10 layoffs aimed squarely at middle management

Open on YouTube ↗

[Meta AI Chief Wang on Winning the Race in AI]

Bloomberg Live · 2026-06-05

The post-split state of MSL — firsthand material from Alexandr Wang

Open on YouTube ↗

[The Man Behind Google's AI Machine]

CNBC International · 2026-01-15

The baseline before Hassabis stepped down as GDM CEO

Open on YouTube ↗

[Marc Benioff Predicts Half of Conversations Will be With AI Agents Next Year]

The Logan Bartlett Show · 2025-08-29

The primary source for the "I need less heads" quote

Open on YouTube ↗

NOTE

The titles and channel attributions of the 19 items above were verified via YouTube oEmbed (2026-08-30); one secondhand Block explainer originally listed is no longer publicly accessible and was removed. For Google DeepMind's 2026.08 reorg, only AI-generated commentary or re-edited videos exist so far — not firsthand channels — so none is listed; for that event, prefer the CNBC / Fortune written reporting, or the 2026-01-15 Hassabis interview above as the pre-reorg baseline. Firsthand audio sources: HBR IdeaCast — NVIDIA's CEO on Leading Through the A.I. Revolution, and the Sequoia podcast transcript.

References

All are publicly accessible news reports, official company statements or first-hand interviews (YouTube sources are listed separately in Chapter 10 and not repeated here); items marked "primary" are official company/individual primary sources. Grouped by company/topic for cross-checking against Chapter 4.

GOOGLE / ALPHABET

SOURCES
- BI: the 10% management cut
- Times of India: all-hands details
- CNBC: the precise 35% measure
- Android Central: the October 2024 reorg explained
- IT Pro: the efficiency drive
- blog.google (official): the 2026.08 reorg announcement
- CNBC: Jeff Dean departs
- Fortune: Hassabis steps down
- Hassabis's X post (official)
- Time: DeepMind functions folded into Google (2026-08-06)
- Fortune: the DeepMind reorg backdrop (2026-08-10)

AMAZON / AWS

SOURCES
- About Amazon (official): the Jassy memo
- About Amazon (official): AWS changes CEO
- CNBC: the endangered middle manager
- CNBC: cutting bureaucracy like a startup
- CNBC: what comes after the layoffs
- Entrepreneur: target met without layoffs (the early-2025 version)
- Fortune: the anti-bureaucracy scorecard
- CRN: Garman's leadership changes
- BI: the internal manager-cutting guidance
- CNBC: the 14,000 layoffs
- CNBC: 1,800+ engineers cut
- GeekWire: Jassy's "it is culture"
- CNBC: another 16,000 cut
- Seattle Times: a historic wave of layoffs
- HBR: the Jassy podcast transcript (2025.05)
- CNBC: AGI-org layoffs (2026-07-22)

X (FORMERLY TWITTER)

SOURCES
- The Verge: the hardcore-Twitter deep dive
- The Verge: the hardcore email, in full
- CNBC: xAI acquires X
- NYT: Yaccarino departs
- NBC News: resignation predates the Grok furor
- Reuters: departures after the X/xAI merger
- CNBC: xAI reorg and co-founder exits (2026-02-11)
- TechCrunch: Nikita Bier departs (2026-08-05)

NVIDIA

SOURCES [Fortune: the Jensen Huang leadership deep dive] [HBR IdeaCast (official): the 2023.12 interview] [Entrepreneur: why no 1:1s] [Tom's Hardware: no long-term plans] [Yahoo Finance: the 55 figure]

META

SOURCES [Economic Times: excerpts from the open letter] [about.fb.com (official): Year of Efficiency] [BI: the big-tech flattening overview] [CNBC (with the official memo): MSL is formed] [Axios: the TBD Lab split] [Built In: the MSL reorg, mapped] [CNBC: Reality Labs January layoffs] [CNBC: hundreds more cut in March] [The Next Web: the May 8,000-person layoff] [BI: the AI-native pod pilot] [BI: player-coach / org leads at Meta and Block] [Meta IR (official): Q4 2025 earnings-call transcript] [CFO Brew: the Stripe Sessions conversation] [Fortune: the 25—30 person core team] [Engadget (citing Reuters): Project OT stumbles (2026-08-26)] [The Decoder: the new Applied AI Engineering org (2026-03)]

MICROSOFT

SOURCES [BI: flattening the management layer] [BI: the second round in July] [Directions on Microsoft: reading the Althoff move] [CNBC: the May layoffs] [CNBC: the July layoffs + Spencer's words] [Microsoft (official): Althoff named commercial-business CEO (2025-10-01)] [Fortune: Xbox CEO on the 3,200 cuts (2026-07-06)]

SALESFORCE

SOURCES [Fortune: the Benioff exclusive] [SF Standard: no engineering hires] [CNBC: the 4,000 layoffs confirmed] [The Next Web: the third 2026 round (2026-08-07)]

BLOCK

SOURCES [Block (official): the essay] [Sequoia (official): the simultaneous publication] [Sequoia (official): the podcast transcript] [Forbes: AI ends middle management] [Forbes: the three DRI roles, mapped] [Fortune: the layoff backdrop confirmed] [Fast Company: Dorsey's "6,000 direct reports" (2026-04)] [Yahoo Finance: Block Q2 2026 earnings-call highlights (2026-08-05)]

COINBASE

SOURCES Fortune: the full 2026.05 reorg | Fortune: the AI coding-assistant mandate | Armstrong's X post (official) | Ramp: the pure-manager layoff trend | Fortune column: named criticism of the player-coach model (2026-05-07)

OPENAI

SOURCES OpenAI (official): the Simo appointment | CNBC: the Simo appointment covered | Altman's X post (official) | MIT Tech Review: research dual heads | The AI Insider: Model Behavior folded into Post Training | Cryptopolitan: supplementary coverage | MIT Tech Review: OpenAI for Science | CNBC: Simo moves to part-time advisor | TechCrunch: same-day coverage | TechCrunch: Weil departs, OpenAI for Science dissolved (2026-04-17) | TechCrunch: Lightcap departs (2026-08-11)

ANTHROPIC

SOURCES Revelio Labs: headcount estimates | SaaStr: "deliberately lean" — revenue per head 6—10x Google's (2026-04) | Lenny's Newsletter (industry interview publication): transcript of the Boris Cherny interview | Anthropic (official): RSP v3.0 | Anthropic (official): the RSP v2.0 announcement (2024-10, with team breakdown) | Anthropic (official, primary): The Claude Code guide for startups (2026-08-20) | Fortune: Dario Amodei's single direct report (2026-06-18) | TechCrunch: new CTO Rahul Patil (2025-10-02)

INDUSTRY VIEWS / WHERE THE MIDDLE LAYER WENT

SOURCES Gartner (official): the 2026 prediction | Economic Times: the Gartner prediction relayed | MIT Tech Review: org design in the agentic-AI era | The Guardian: the AI middle-management purge | Advisor Perspectives: the $400K Chief of Staff | Business Model Analyst: the Chief of Staff boom, analyzed | Lepaya: coordinator → orchestrator | BI: managers tasked with driving AI adoption | Gallup: spans of control and optimal team size (2026-01-13) | Bloomberg: the $400K Chief of Staff of the AI age (2026-07-16) | HBR: AI adoption is overloading your middle managers (2026-06-26) | BI: has the Great Flattening gone too far? (2025-05-19) | Fast Company: don't rush to cut the middle — the data disagrees (2026-06-01) | Ramp: the "pure manager" layoff trend (2026-05-12)

AI-era corporate org-structure research report · compiled from 100+ public sources and 19 YouTube videos · evidence current through 2026-08