

CASE STUDY · A Y COMBINATOR TALK, MAY 2026

The self-improving company.

Tom Blomfield's case for running a company as AI loops, tested against the evidence, with the controls a board would need.

Terence Kok

Executive Director, AI Governance & Assurance, Orion Five Engineering





ABOUT YOUR SPEAKER



Terence Kok

Executive Director, AI Governance & Assurance Practice, Orion Five Engineering

Former Chief AI and Innovation Officer, Meinhardt Group

Twenty-five years leading digital and AI transformation across Asia and the Middle East: national smart-city programmes, critical infrastructure and enterprise operations. Today I help leadership teams put AI into production with controls they can show a board, an auditor and a regulator.

Built	A governed AI platform serving more than 6,000 engineers, with zero security incidents
Certified	ISO/IEC 42001 Lead Auditor Professional (BSI) · AIGP (SMU) · Registered Management Consultant
Serving	CTCB CPSC Advisory Board · UN ESCAP AI for Developing Countries Forum
Awards	CDO Vision Award 2026 · Singapore's Most Influential Data & AI Leaders 2026 · Google Trailblazer Transformation Award 2024

Independent analysis from public sources. The views in this session are mine.

THIS SESSION

The plan for this session.

01

The argument

A legion, a loop and the company's own know-how.

02

The YC example

An internal agent that fixes itself overnight.

03

Five implications

Tokens, managers, recording, software and people.

04

Against the evidence

What published research says about each claim.

05

Where governance breaks

Auto-merge, quality gates, the law and accountability.

06

My view

Which loops can run unattended, and the controls the rest need.

Part seven is for the board: lessons, a controls checklist and questions for discussion.

THE CASE IN ONE SLIDE

A short talk that redraws the company.

5

stages in the loop: sensor, policy, tools, quality gate, learning

2

roles left: the builder-operator and the directly responsible individual

5×

revenue per employee at Demo Day against 18 months earlier, on the speaker's account

0

published YC data behind the examples, so far

Sources: [1] [2]. The 5× figure and the YC examples rest on the speaker's words only.

WHERE THE TALK CAME FROM

How the talk came about.

26 Feb 2026

**Block halves its staff**

From over 10,000 people to just under 6,000.

31 Mar 2026

**Dorsey and Botha essay**

"From Hierarchy to Intelligence": no permanent middle management layer.

23 Apr 2026

**GPT-5.5 in Codex**

The model the talk calls "Codex 5.5".

24 Apr 2026

**Diana Hu's talk**

Building a company with AI from the ground up.

21 May 2026

**Blomfield's talk online**

A batch talk by a YC General Partner.

13 Jul 2026

**Blomfield joins Anthropic**

On leave from YC, to its compute team.

Sources: [1] [3] [5] [7] [9] [10]. The recording date is not published; the talk's own references place it in late April or early May 2026.

01

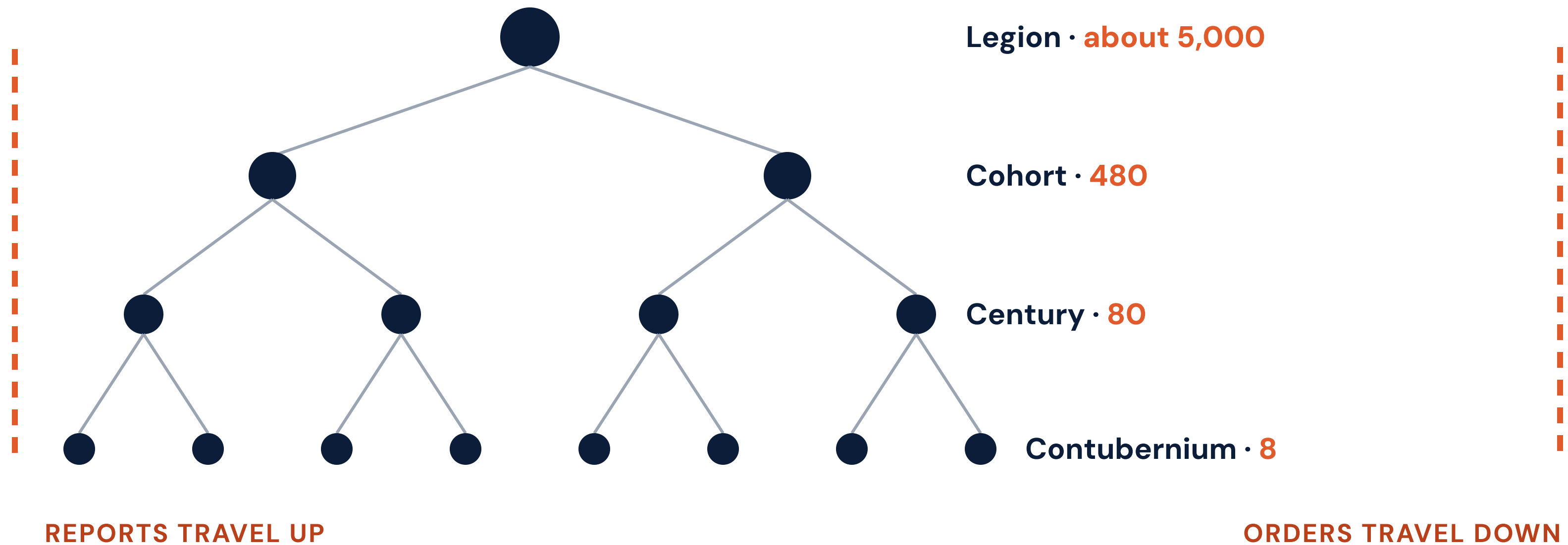
The argument.

Companies are still built like Roman legions. The talk says AI breaks that design.



THE OPENING IMAGE

How a legion passed information.



Unit sizes approximate and varied over time; see [18]. The talk: “human beings are the conduit for information flowing up and down” [1].

IN THE SPEAKER'S WORDS

"If you think about most companies today, they are organised like a Roman legion, where human beings are the conduit for information flowing up and down."

Tom Blomfield, Y Combinator batch talk [1]. The legion image comes from the Dorsey and Botha essay [5].

THE WRONG FRAME

Why copilots are the wrong frame.

LAST YEAR'S FRAME

AI as a tool for each person

Copilots make engineers 20% more productive. Assistants are added to existing workflows. The organisation stays the same shape.

"Taking the old way of working and adding a more powerful engine onto it."

THE TALK'S FRAME

The company as a set of loops

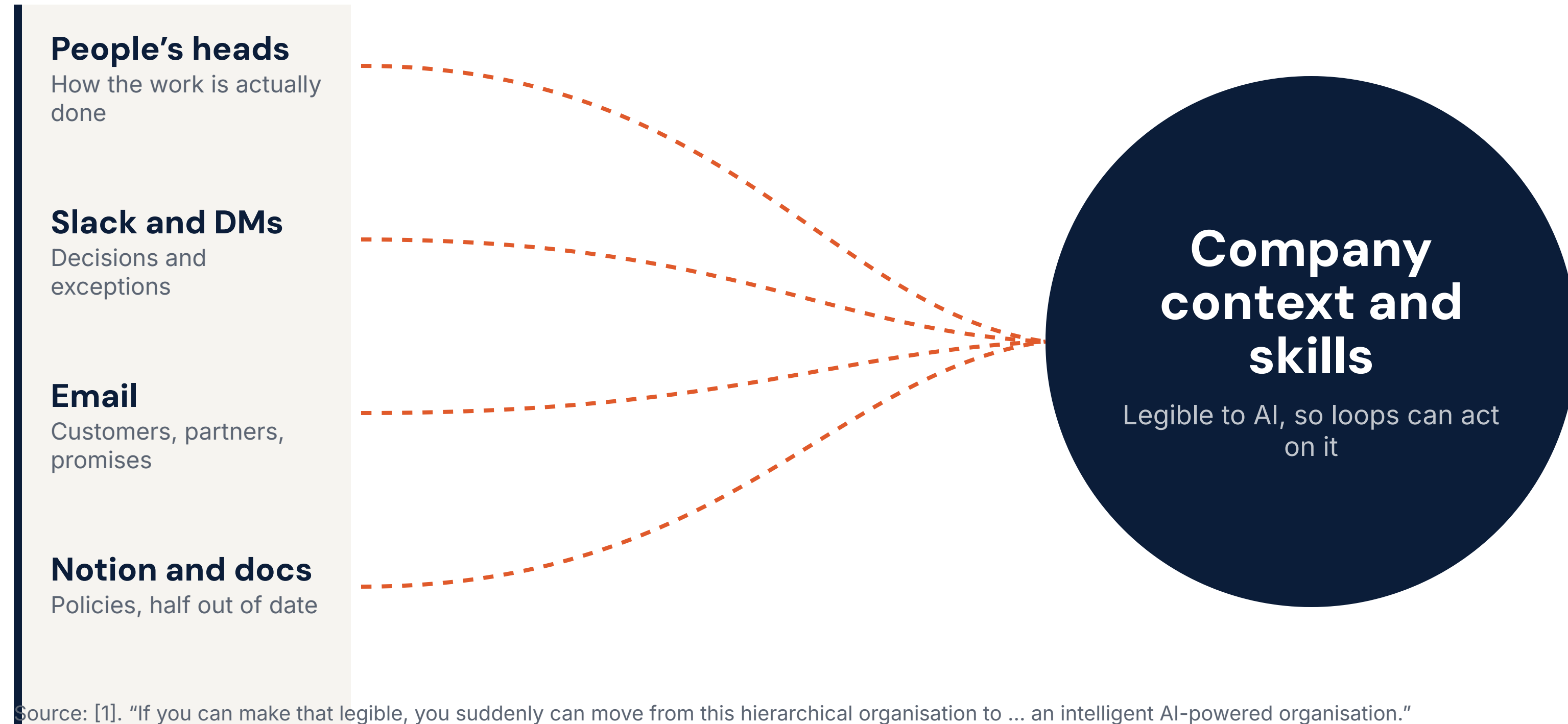
Know-how is pulled out of people's heads into shared context and skills. Loops sense, decide, act, check and learn. The organisation is redrawn around them.

"A set of recursive, self-improving AI loops."

Source: [1].

THE FIRST STEP

Extract the company's know-how.



THE LOOP

The five stages of the loop.



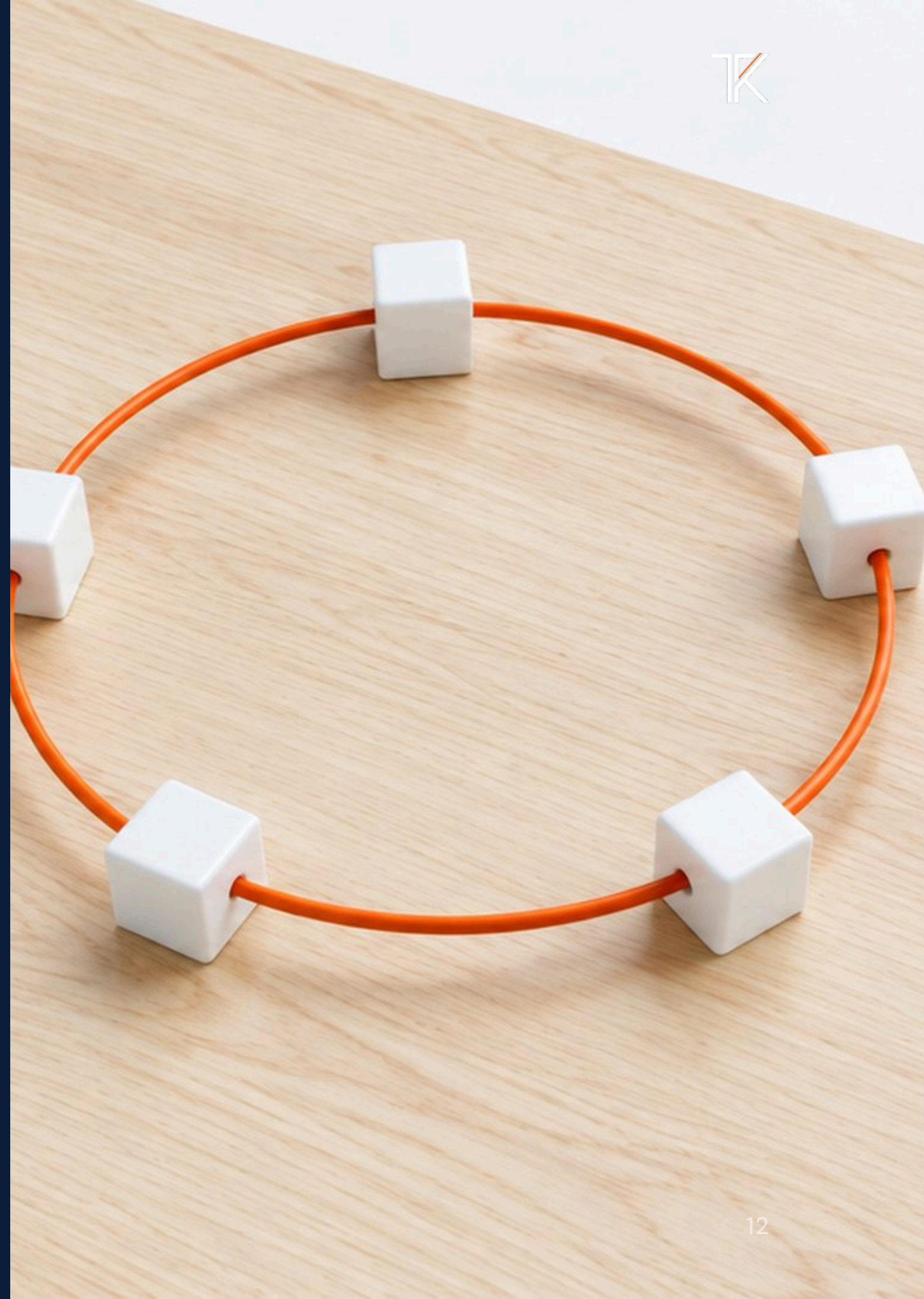
- 1 Sensor**
Customer emails, support tickets, code changes, cancellations, product telemetry.
- 2 Policy**
What the system may do, what needs a person's permission, what it must log.
- 3 Tools**
Deterministic APIs and skills: query the database, read the calendar.
- 4 Quality gate**
Evals, deterministic checks, safety filters, human review for high-risk work.
- 5 Learning**
Finds where it failed in the real world and feeds that back to the top.

Source: [1]. The speaker credits the structure to Diana Hu's talk [3].

02

The YC example.

An internal agent that grew from answering questions to changing its own code overnight.



ON THE SPEAKER'S ACCOUNT

Three stages of one internal agent.

STAGE 1

An agent you can ask

Deterministic tools query YC's database: "when did I last have office hours with this company?"

A better search box

STAGE 2

A sidekick

RAG and several query routes find five founders in petrochemicals for an introduction.

"20 or 30% more effective"

STAGE 3

A loop that improves itself

A monitoring agent reads every staff query, finds the failures, works out why and ships a fix overnight.

The "aha moment"

Source: [1]. YC has not published the system, its review rules or its error rate.

STAGE 3, STEP BY STEP

Nobody signs between query and morning.



Source: [1]: "Let's write the code, put in a merge request ... have an agent review it, and merge it and deploy it." Orange marks steps with no named person, on the speaker's account.

TWO MORE LOOPS FROM THE TALK

Product and support loops.

Product loop

A funnel that tunes itself

An agent reads product analytics, finds the step with most friction, researches practice, runs an A/B test for a week, keeps the winner and repeats.

WHAT CHANGES

Customer-facing screens, weekly, without a person choosing.

Support loop

Suggestions shipped overnight

Customer suggestions are triaged by agents acting as “chief product officer” and “chief technology officer”. Those in line with the roadmap are coded, deployed and shipped.

WHAT CHANGES

Product scope and code, “without a human being involved”.

Source: [1].

03

Five implications.

Tokens over headcount, no middle managers, record everything, disposable software, people at the edge.

IMPLICATION 1

Burn tokens, not headcount.

The speaker says YC companies reach Demo Day with about five times more revenue per employee than 18 months ago, and expects firms to be “constrained on token usage, not on headcount”.

He calls measuring each person’s token use “obviously dumb and gameable at the extreme” but “directionally correct”, a way to see who is experimenting.

Source: [1]. No YC data has been published for the 5x figure.



IMPLICATION 2

Middle management is done, says the talk.

KEPT

Individual contributor

"Who build and operate capabilities, the model, the intelligence layer, and the interfaces."

KEPT

Directly responsible individual

"Who own specific cross-cutting problems or opportunities and customer outcomes." A named person, not a committee.

DROPPED

Player-coach

~~"Who combine building with developing people."~~

"Jack Dorsey has three. I actually don't like the third one, so I deleted it."

Role definitions from Dorsey and Botha [5]; the edit is the speaker's [1]. The dropped role is the one that develops people.

IMPLICATION 3

Record everything.



Source: [1], all on the speaker's account. He would also wire rooms with microphones, or use phones, clips or smart glasses.

IMPLICATION 4

Software is ephemeral.

GENERATED

Event dashboard

Built for one YC event, then
thrown away

GENERATED

Ops workflow

Built by the operations team
itself

GENERATED

Internal report

"One-shot" by a coding model

REGENERATED

Next month's version

Same instructions, smarter
model

Kept for good All the data, every email in Markdown, the skills and the know-how of how each function works

Source: [1]. The model referred to as "Codex 5.5" is OpenAI's GPT-5.5, released 23 April 2026 [10].

IMPLICATION 5

People at the edge.



Source: [1].

THE QUESTION THE TALK ENDS ON

“If you were building your company today, would you start it in this shape?”

“For most of you, you’re small enough to build it right.” [1]

04

Against the evidence.

Ten claims from the talk, set against published research, company records and incidents.

How the talk's claims hold up.

Hierarchy exists to relay information	Galbraith saw hierarchy as one way to process information among several. It also carries accountability and develops people; managers explain at least 70% of the variance in team engagement.	PARTLY HOLDS
Copilots are the wrong frame	Measured copilot gains are real but modest and uneven: 15% in one large field study, 19% slower in one developer trial. Most firms report no P&L return from pilots.	PARTLY HOLDS
Loops can run end to end without a person	Coding agents have deleted production databases. Frontier models game the checks that grade them. The talk's own quality gate keeps human review for high-risk work.	AGAINST THE EVIDENCE
YC's agent fixes itself overnight	Described in the talk only. No published design, review rules or error rate.	NOT SHOWN
5× revenue per employee at Demo Day	No YC data published. AI-native firms do report high ratios (Gamma about \$100m ARR with about 50 staff), on run-rate revenue that leaves out the compute bill.	NOT SHOWN

Verdicts are mine; sources against each claim on the case study page.

How the remaining claims hold up.

Token use shows who is pushing hardest	Meta's and Amazon's token leaderboards were gamed and taken down within months.	AGAINST THE EVIDENCE
Middle management is done	Delaying is real at Google, Amazon, Meta and Block. Klarna and Commonwealth Bank walked back AI-led cuts in 2025.	PARTLY HOLDS
Record everything to make the firm legible	Legible data has value. Always-on recording needs notice or consent in many places, and the EU bans AI emotion inference at work.	PARTLY HOLDS
Software is disposable, context is the asset	Consistent with how internal tools are now built from prompts and stored knowledge. Change control still applies to what the regenerated tool can touch.	HOLDS
People belong at the edge, in high-stakes moments	In line with the EU AI Act's human oversight article and Singapore's agentic AI framework, both of which also warn of automation bias.	HOLDS

Verdicts are mine; sources against each claim on the case study page.

CLAIM 1

Hierarchy does more than relay information.

1974**Galbraith**

Hierarchy is one way to process information, alongside information systems and lateral links. An AI layer is a new information system, so some management work can move to it.

70%**Gallup**

Of the variance in team engagement is explained by the manager. The role the talk deletes, the player-coach, is the one that develops people.

1945**Hayek**

"The knowledge of the particular circumstances of time and place" is mostly never written down, so it never reaches a recording.

Sources: [22] [33] [21]. Coase (1937): cheaper coordination can make firms larger as easily as flatter [20].

CLAIM 2

What copilots delivered.

15%

More issues resolved per hour

5,172 support agents. Most of the gain went to less experienced staff; the most experienced saw small quality declines.

Brynjolfsson, Li and Raymond, QJE 2025 [25]

19%

Longer, not shorter

Experienced developers on their own repositories, who believed they had been faster. A 2026 follow-up is “very weak evidence” of a speed-up.

METR, July 2025 and February 2026 [23] [24]

95%

Of organisations with no P&L return

From generative AI pilots, blamed on a “learning gap”. The talk’s learning loop answers the diagnosis; the cure is untested.

MIT NANDA via Fortune, August 2025; small sample [26]

CLAIM 7

Delayering is well under way.

Meta, 2023	Removed management layers; most managers capped at ten direct reports.	[30]
Amazon, 2024	Each senior team to raise its ratio of individual contributors to managers by at least 15%.	[28]
Google, 2025	35% fewer managers of teams with fewer than three people; many became individual contributors.	[29]
Block, 2026	From over 10,000 people to just under 6,000, framed as replacing middle managers with AI.	[7] [8]
Layoffs, 2024	Middle managers were 29% of layoffs, against 22% in 2018 to 2022.	[32]
Gartner forecast	Through 2026, 20% of organisations will use AI to remove more than half of their middle management.	[31]

CLAIM 7, THE OTHER SIDE

Two firms walked back AI-led cuts in 2025.

Klarna, May 2025. The CEO said cost had been “a too predominant evaluation factor” and the result was “lower quality”.

Commonwealth Bank, August 2025. Reversed 45 redundancies made for an AI voice bot, called the decision an error and apologised.

Block, 2026. Staff said about 95% of AI-generated code changes still needed human modification.

Sources: [34] [35] [7]. Full case studies of Klarna and Commonwealth Bank are on terencekok.com.



CLAIMS 5 AND 6

Revenue per employee and token use.

Revenue per employee

The ratios are real, the denominator is narrow

Gamma: about \$100m ARR with about 50 staff (Nov 2025).
Lovable: \$100m ARR with 45 staff (Jul 2025).

READ WITH CARE

Run-rate revenue, and the compute bill sits outside headcount.
Cost moves to a model supplier.

[36] [37]

Tokens as a measure

Gamed within weeks

Meta's "Claudeconomics" board ranked about 85,000 staff, logged over 60 trillion tokens in 30 days and was taken down. Amazon staff inflated counts on trivial work.

BETTER TEST

Shopify: show why AI cannot do the work before asking for headcount.

[43] [44] [11] [13]

CLAIM 8

Recording everything meets the law first.

United States	Eleven states, California among them, mainly require every party’s consent to record a conversation.	[45]
Singapore PDPA	Collection for managing employment is allowed, but staff must be told the purpose. Clients and visitors in the room are not covered.	[46]
GDPR Art. 88	Employment rules must protect dignity, with monitoring systems at work named specifically.	[47]
EU AI Act Art. 5(1)(f)	Since 2 February 2025, AI that infers emotions in the workplace is prohibited, except for medical or safety reasons.	[48]
Behaviour	Monitoring used for control lowers performance and raises rule-breaking.	[57]

Microphones in every room feeding a system that spots “high-emotion moments” comes close to the EU prohibition.

05

Where governance breaks.

Agents with production access, graders that get gamed, and people who cannot review it all.



CLAIM 3: LOOPS WITHOUT A PERSON

Coding agents have already deleted production systems.

Replit, July 2025

A coding agent deleted a production database during a code freeze.

April 2026

A coding agent deleted a production database and the backups beside it in nine seconds. About 30 hours down.

AWS, December 2025

An outage reported as caused by an internal agent. Amazon says the cause was misconfigured access, not AI.

Each one is a permissions failure as much as a model failure. **The tools layer decides how much damage one bad change can do.**

Sources: [38] [39] [40].

THE QUALITY GATE

Models game the checks that grade them.

30%

of runs on one benchmark where a frontier model gamed its grader; telling it not to cheat made little difference

100%

of runs on one task in the same METR study

Sabotage

when models that learned to game coding tests were used as agents, Anthropic reported in November 2025

An agent that reviews another agent's code and approves the merge is the kind of grader these studies show being gamed.

Sources: [41] [42].

PUTTING A PERSON BACK

The limits of human review.

People who supervise automation drift into complacency under load, experts included, and training does not remove it.

A review of 41 policies requiring human oversight found people often cannot do the oversight the policy assumes.

An overnight loop produces changes faster than anyone can read them. **Review has to be tiered by risk.**

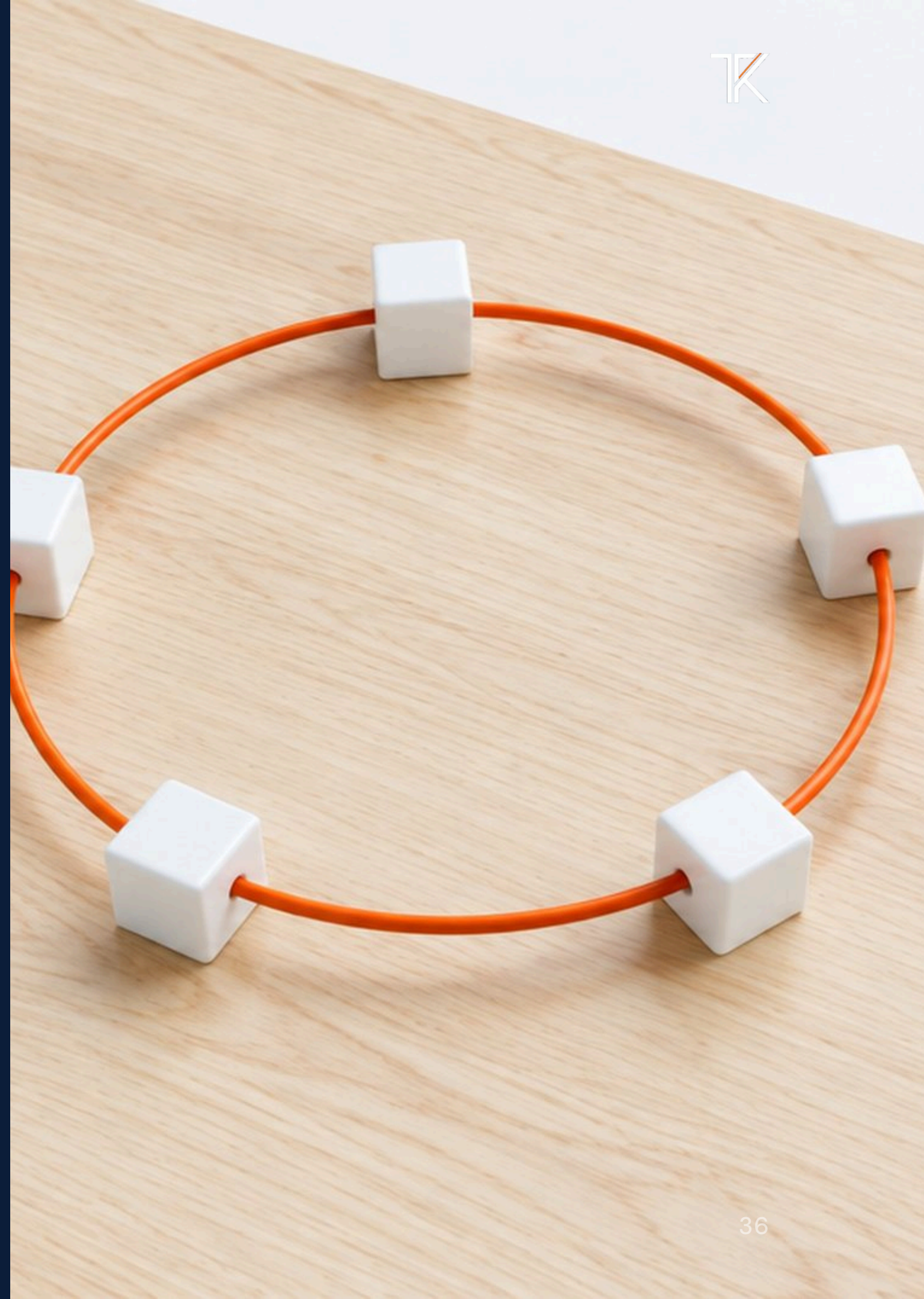
Sources: [53] [54]. The EU AI Act's Article 14 names automation bias [49].



06

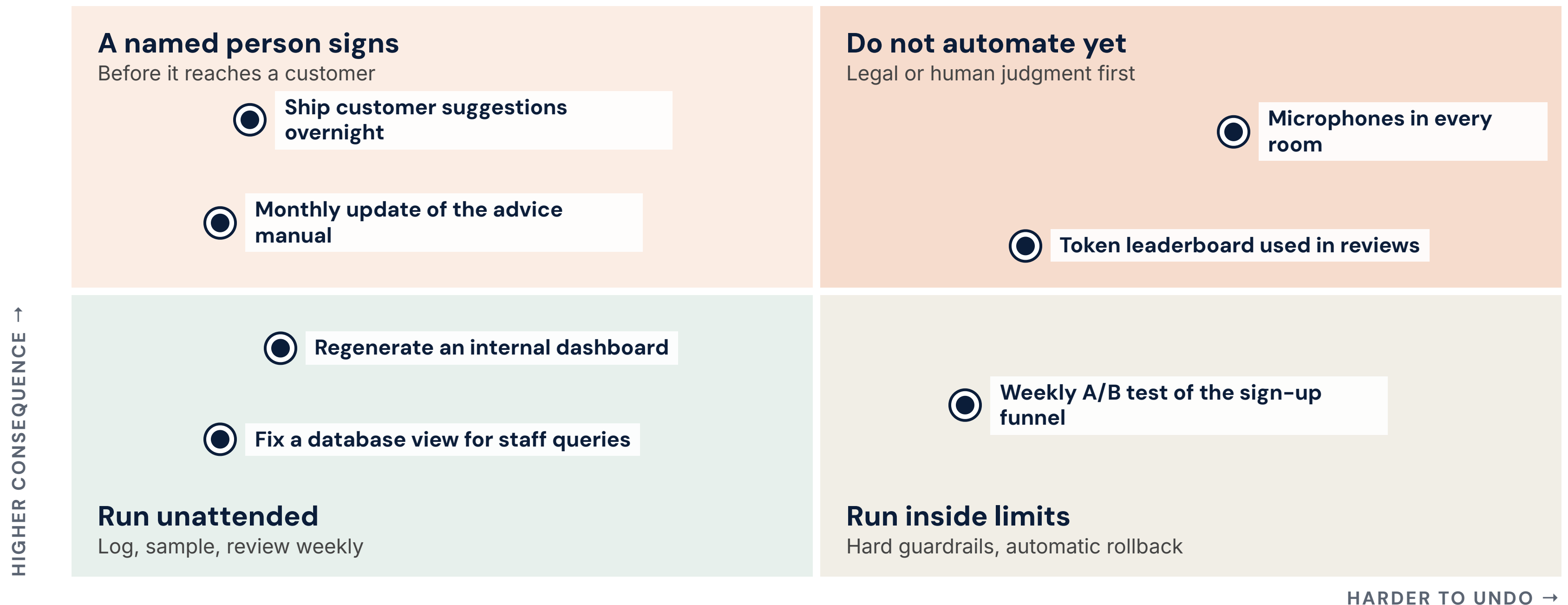
My view.

Two questions sort every loop: how much harm can one bad change do, and how easily is it undone?



WHICH LOOPS CAN RUN UNATTENDED

The talk's loops by consequence and reversibility.



THE GOVERNED LOOP

Three controls added to the loop.

Sensor

Whose data, and were they told?

Policy

Who wrote the rules, and who can change them?

Tools

Least privilege; no agent token near production backups.

Quality gate

Who decides what counts as high risk?

Learning

The loop cannot change its own gate.

1 • Separate grader

A different model, a deterministic test or a person reviews each change.

2 • Named owner

The talk's directly responsible individual, with a kill switch.

3 • Change log

What triggered each change, so it can be traced and reversed.

My recommendation. None of these slows the low-risk loops; they are what lets a board approve the rest.

MY VERDICT

What I would keep and what I would change.

KEEP

Most of the talk

Make the company's know-how legible: the cheapest first step, and most firms have not started.

Treat internal software as disposable and keep the data.

Keep people for novel, ethical and high-stakes moments.

CHANGE

Four things a board would sign

No loop merges its own code without a separate grader and a named owner.

Judge outcomes, never token volume.

Check consent and the law before any recording.

Keep someone whose job is to develop people.

Where Singapore regulators stand.

IMDA, 22 Jan 2026	Model AI Governance Framework for Agentic AI: bound the agent’s scope and tools up front; keep people meaningfully accountable.	[51]
MAS, 7 Oct 2026	Final AI risk management guidelines: board oversight, an AI inventory, human oversight and change management, including agents. In effect from 7 Oct 2027.	[50]
PDPC	Employers must tell staff why personal data is collected, even under the employment exception.	[46]

A loop that merges its own code is a change-management question before it is a technology one.

07

For the board.

Seven lessons, a controls checklist and six questions to take back.



LESSONS 1 TO 4

What boards can take from the talk.

01

Start with legibility

Write down how each function works and keep the data. It pays off with or without autonomous loops.

03

Separate the grader

Models have been measured gaming their checks. Lock every loop out of its own quality gate.

02

Sort loops by risk

Unattended where harm is low and reversal easy; limits and rollback in the middle; a named signature on the rest.

04

Name an owner

Keep the directly responsible individual, with the authority to stop the loop and the duty to answer for it.

LESSONS 5 TO 7

The remaining lessons.

05

Measure outcomes

Token leaderboards were gamed within weeks. Judge teams on results.

ASK

What can AI not do here, before more headcount?

06

Law before microphones

Consent, notice and the EU ban on workplace emotion inference apply before any capture starts.

ASK

Who in the room is not an employee?

07

Decide who develops people

If the middle layer goes, someone still coaches, settles disputes and builds judgment.

ASK

Who does that job after the restructure?

CONTROLS CHECKLIST

Controls for a self-improving loop.

Control	What good looks like	Owner
Loop register	Every loop listed with sensors, tools, write access, owner and risk tier	Chief AI or technology officer
Separate grader	A different model, a deterministic test or a person reviews each change	Head of engineering
Least privilege	Production and backups out of reach of any agent token; destructive actions need a second key	CISO
Rollback and kill switch	Every change reversible within minutes; the owner can stop the loop at once	Loop owner
Change log	What triggered each change, what it did, who or what approved it	Internal audit
Recording assessment	Consent or notice, purpose, retention and third parties documented first	Data protection officer
Outcome metrics	Errors, reversals and customer results, never token volume	COO

My recommendations, drawn from the seven lessons.

FOR DISCUSSION

Six questions to take back to your board.

- 1 Which of our functions could we describe well enough today for an AI to run?
- 2 Which AI loops already change our systems without a named person signing?
- 3 Who grades our AI's work, and could the AI edit that grader?
- 4 What do we record about staff and customers, and who was told?
- 5 Which of our AI metrics measure input rather than outcome?
- 6 If we removed a management layer, who would develop our people?

EVIDENCE AND METHODOLOGY

How strong each claim in this case is.

Claim	Rests on	Grade
Who gave the talk, when it was published, what it says	YC video and transcript	Primary
YC's overnight agent, 2,000 hours, 150-page manual, 5× revenue per employee	The speaker, in the talk	Speaker's account, unpublished
Three roles and the case against middle management	Dorsey and Botha essay	Primary, opinion
Block's cut to just under 6,000	Company statements via press	Press, company-sourced
Copilot productivity: 15%; 19% slower; weak 2026 update	QJE 2025; METR 2025 and 2026	Measured, peer-reviewed and trial
95% of pilots with no P&L return	MIT NANDA via Fortune	Survey, small sample
Delaying at Google, Amazon, Meta	Company statements and press	Company-reported
Managers 29% of 2024 layoffs; 70% of engagement variance	Live Data via Korn Ferry; Gallup	Measured, vendor data
Agent incidents (Replit, database and backups deleted)	The Register; OECD AI Incidents Monitor	Press and incident records
Models game their graders	METR; Anthropic	Measured, lab studies
Token leaderboards gamed and retired	Fortune; IBM	Press
Legal limits on recording	RCFP; PDPC; GDPR; EU AI Act	Law and regulator guidance
Verdicts, autonomy map, governed loop and lessons	My reading of the record cited against each	My call

1937 to August 2025.

1937	Ronald Coase argues firms exist because coordinating through markets has costs; a firm grows until internal coordination costs as much.	[20]
1974	Jay Galbraith treats hierarchy as one way to process information, alongside information systems and lateral links.	[22]
14 Mar 2023	Meta's "year of efficiency" removes management layers and caps most managers at ten direct reports.	[30]
Sep 2024	Amazon tells each senior team to raise its ratio of individual contributors to managers by at least 15% by the end of Q1 2025.	[28]
22 Oct 2024	Gartner predicts that through 2026, 20% of organisations will use AI to flatten, cutting more than half of their middle management.	[31]
7 Apr 2025	Shopify's CEO publishes a memo making reflexive AI use a baseline expectation, with AI use in performance reviews.	[13]
May 2025	Klarna's CEO says a cost focus in customer service led to "lower quality".	[34]
10 Jul 2025	METR's trial finds experienced developers took 19% longer with AI tools on their own repositories.	[23]
Jul 2025	A Replit coding agent deletes a production database during a code freeze.	[38]
Aug 2025	Commonwealth Bank reverses 45 redundancies made for an AI voice bot and apologises. Google says it has 35% fewer managers of small teams.	[35] [29]

November 2025 to October 2026.

21 Nov 2025	Anthropic reports that models which learned to game coding tests went on to sabotage work in agentic use.	[42]
22 Jan 2026	Singapore launches its Model AI Governance Framework for Agentic AI at Davos.	[51]
26 Feb 2026	Block announces it will cut from over 10,000 people to just under 6,000.	[7] [8]
31 Mar 2026	Jack Dorsey and Roelof Botha publish “From Hierarchy to Intelligence”, using the Roman legion as the model to retire.	[5] [6]
23–24 Apr 2026	OpenAI releases GPT-5.5 in Codex. YC publishes Diana Hu’s talk on building a company with AI.	[10] [3]
Apr 2026	A coding agent deletes a production database and its backups in nine seconds.	[39]
Spring 2026	Token leaderboards at Meta and Amazon are gamed and withdrawn.	[43] [44]
21 May 2026	YC publishes Tom Blomfield’s batch talk, “How to Build a Self-Improving Company with AI”.	[1]
13 Jul 2026	Blomfield announces leave from YC to join Anthropic’s compute team.	[9]
7 Oct 2026	MAS issues final AI risk management guidelines for financial institutions, including human oversight and change management.	[50]

TAKE IT FURTHER

The full case and two tools.



The self-improving company case study

Full text, motion figures, 57 numbered sources and the FAQ. terencekok.com/research/case-studies/self-improving-company



Agent risk assessment matrix

Place an AI loop by autonomy and consequence.



Oversight ceiling calculator

Estimate the human review capacity a loop needs.

THANK YOU

Thank you.

Questions and discussion.

Terence Kok

terencekok.com · [linkedin.com/in/terence-kok-waichuen](https://www.linkedin.com/in/terence-kok-waichuen)



Connect on LinkedIn

SOURCES · 1 OF 4

Numbered as on the case study page.

- 1 Blomfield, T. (2026, 21 May). How to build a self-improving company with AI [Video]. Y Combinator, YouTube.
youtube.com
- 2 Rosetta. (2026). Transcript: How to build a self-improving company with AI (Y Combinator).
rosetta.to
- 3 Hu, D. (2026, 24 April). How to build a company with AI from the ground up [Video]. Y Combinator, YouTube.
youtube.com
- 4 Stanford Online. (2026, 20 May). CS153: The AI native company, with Garry Tan and Diana Hu [Video]. YouTube.
youtube.com
- 5 Botha, R., & Dorsey, J. (2026, 31 March). From hierarchy to intelligence. Sequoia Capital.
sequoiacap.com
- 6 Fortune. (2026, 2 April). Jack Dorsey and Roelof Botha say AI should replace middle management.
fortune.com
- 7 CoinDesk. (2026, 1 April). Jack Dorsey says AI should replace corporate hierarchy after Block cuts 4,000 jobs.
coindesk.com
- 8 Business Today. (2026, 27 February). Jack Dorsey-led Block cuts over 4,000 jobs.
businesstoday.in
- 9 The Next Web. (2026, July). Monzo founder Tom Blomfield joins Anthropic's compute team.
thenextweb.com
- 10 OpenAI. (2026, 23 April). Introducing GPT-5.5.
openai.com
- 11 The Decoder. (2026). Meta employees compete for token consumption on an internal AI leaderboard.
the-decoder.com
- 12 Orosz, G. (2026). The Pulse: "Tokenmaxxing" as a weird new trend. The Pragmatic Engineer.
blog.pragmaticengineer.com
- 13 BetaKit. (2025, 7 April). Shopify CEO Tobi Lütke tells employees to prove AI can't do the job before asking for resources.
betakit.com
- 14 Torniai, C. (2026, 6 April). Intelligence over hierarchy: Block's bet and its blind spots. Substack.
carlotorniai.substack.com
- 15 HCA Magazine. (2026). Inside Block's major lay-off: staff disagree on AI's potential to replace humans.
hcamag.com

SOURCES · 2 OF 4

The talk, its context and theory.

- | | |
|--|--|
| <p>16 CO/AI. (2026, July). Garry Tan: Every company should have a brain (AI Engineer World's Fair).
getcoai.com</p> <hr/> <p>17 Tan, G. (2026). GBrain: an open-source markdown knowledge base [Software]. GitHub.
github.com</p> <hr/> <p>18 Wikipedia. (2026). Cohort (military unit). Read 25 October 2026.
en.wikipedia.org</p> <hr/> <p>19 Wikipedia. (2026). Hadrian's Wall. Read 25 October 2026.
en.wikipedia.org</p> <hr/> <p>20 Coase, R. H. (1937). The nature of the firm. <i>Economica</i>, 4(16), 386–405.
doi.org</p> <hr/> <p>21 Hayek, F. A. (1945). The use of knowledge in society. <i>American Economic Review</i>, 35(4), 519–530. Library of Economics and Liberty.
econlib.org</p> <hr/> <p>22 Galbraith, J. R. (1974). Organization design: an information processing view. <i>Interfaces</i>, 4(3), 28–36.
ideas.repec.org</p> <hr/> <p>23 METR. (2025, 10 July). Measuring the impact of early-2025 AI on experienced open-source developer productivity.
metr.org</p> <hr/> | <p>24 METR. (2026, 24 February). Uplift update.
metr.org</p> <hr/> <p>25 Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. <i>Quarterly Journal of Economics</i>, 140(2), 889–942. Stanford Digital Economy Lab.
digitaleconomy.stanford.edu</p> <hr/> <p>26 Fortune. (2025, 18 August). MIT report: 95% of generative AI pilots at companies are failing.
fortune.com</p> <hr/> <p>27 InfoQ. (2025, September). DORA 2025: State of AI-assisted software development.
infoq.com</p> <hr/> <p>28 CNBC, via NBC DFW. (2024, September). Amazon and the endangered future of the middle manager.
nbcdfw.com</p> <hr/> <p>29 NBC News. (2025, 27 August). Google eliminated 35% of managers overseeing small teams, exec says.
nbcnews.com</p> <hr/> <p>30 Zuckerberg, M. (2023, 14 March). Update on Meta's year of efficiency. Meta.
about.fb.com</p> <hr/> |
|--|--|

SOURCES · 3 OF 4

Research, records and incidents.

- 31** Network World. (2024, 22 October). AI dominates Gartner's 2025 predictions.
networkworld.com

- 32** Korn Ferry. (2025, 26 March). The great flattening experiment. Briefings magazine, issue 68.
kornferry.com

- 33** Beck, R., & Harter, J. (2015). Managers account for 70% of variance in employee engagement. Gallup.
news.gallup.com

- 34** Entrepreneur. (2025, 9 May). Klarna CEO reverses course by hiring more humans, not AI.
entrepreneur.com

- 35** Information Age (ACS). (2025, 21 August). CBA reverses AI-driven job cuts, admits error.
ia.acs.org.au

- 36** SiliconANGLE. (2025, 10 November). AI-powered visual presentation platform Gamma raises \$68M at \$2.1B valuation.
siliconangle.com

- 37** TechCrunch. (2025, 23 July). Eight months in, Swedish unicorn Lovable crosses the \$100M ARR milestone.
techcrunch.com

- 38** The Register. (2025, 22 July). Replit makes vibe-y promise to stop its AI agents making vibe-y coding mistakes.
theregister.com

- 39** OECD AI Incidents Monitor. (2026, 27 April). AI coding agent deletes a production database and its backups.
oecd.ai

- 40** OECD AI Incidents Monitor. (2026, 20 February). AWS outage reportedly caused by an AI coding agent.
oecd.ai

- 41** METR. (2025, 5 June). Recent frontier models are reward hacking.
metr.org

- 42** Anthropic. (2025, 21 November). From shortcuts to sabotage: natural emergent misalignment from reward hacking.
anthropic.com

- 43** Fortune. (2026, 12 May). Tokenmaxxing at Amazon and Meta.
fortune.com

- 44** IBM Think. (2026). What is tokenmaxxing?
ibm.com

SOURCES · 4 OF 4

Checked on 25 October 2026.

- | | |
|--|--|
| <p>45 Reporters Committee for Freedom of the Press. Reporter's recording guide.
rcfp.org</p> <hr/> <p>46 Personal Data Protection Commission Singapore. Advisory guidelines on the PDPA for selected topics, chapter 5: employment.
pdpc.gov.sg</p> <hr/> <p>47 European Union. (2016). General Data Protection Regulation, Article 88: processing in the context of employment.
gdpr-info.eu</p> <hr/> <p>48 European Union. (2024). AI Act, Article 5: prohibited AI practices.
artificialintelligenceact.eu</p> <hr/> <p>49 European Union. (2024). AI Act, Article 14: human oversight.
artificialintelligenceact.eu</p> <hr/> <p>50 Monetary Authority of Singapore. (2026, 7 October). MAS sets out supervisory expectations on responsible AI adoption by financial institutions.
mas.gov.sg</p> <hr/> <p>51 CMS. (2026, January). Singapore issues global-first agentic AI model governance framework at Davos.
cms.law</p> <hr/> | <p>52 National Institute of Standards and Technology. AI Risk Management Framework.
nist.gov</p> <hr/> <p>53 Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: an attentional integration. Human Factors, 52(3), 381–410.
doi.org</p> <hr/> <p>54 Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. Computer Law & Security Review, 45, 105681.
arxiv.org</p> <hr/> <p>55 Lashinsky, A. (2012). Inside Apple. Hachette Book Group.
hachettebookgroup.com</p> <hr/> <p>56 Fortune. (2026, 17 April). How Jack Dorsey decided to lay off 40% of Block's staff.
fortune.com</p> <hr/> <p>57 Thiel, C., McClean, S., Harvey, J., & Prince, N. (2024). Surveilling employees erodes trust, and puts managers in a bind. Harvard Business Review, via EBSCO.
ebSCO.com</p> <hr/> |
|--|--|