FROM PILOTS TO COPILOTS

Scaling Enterprise Generative AI with Data, Controls and Measurable Value

31 AUGUST 2023 • ENTERPRISE AI SERIES



Paul Forrest

PAPER FOUR · ENTERPRISE AI, 2023 TO 2024

About this paper

This is the fourth of nine papers on the enterprise adoption of generative artificial intelligence.

The first three papers dealt with learning, individual capability and workforce design. This one takes the question those three defer, which is how an organisation decides that a pilot has earned the right to become a product.

The evidence in this paper is thinner than the argument, and I have tagged the joins so that a reader can see where a claim rests on a published source and where it rests on my own practice.

HOW TO CITE

> Forrest, P. (2023). *From Pilots to Copilots: Scaling Enterprise Generative AI with Data, Controls and Measurable Value.* Enterprise AI, 2023 to 2024, Paper Four. Version 1.0, 31 August 2023.

HOW TO READ THE EVIDENCE TAGS

Every substantive claim carries a tag stating what kind of evidence stands behind it. A claim resting on two kinds carries both.

PRIMARY An institutional or official document, cited from the issuing body.

VENDOR A vendor announcement or a figure the vendor reports about its own product.

SURVEY Survey or polling data, cited with its sample and fieldwork dates.

AUTHOR My own framework or judgement, proposed and not yet proven.

A NOTE ON DATES

Research for this paper closed on 29 August 2023. One source needs care and the text gives it. The international labour evidence cited here was collected in early 2022, before the systems this paper is about reached the public, and it is used for what it can support.

Contents

FRONT MATTER

	Abstract and summary of the argument	4

THE PAPER

1	What changed over the summer	5
2	Portfolio before platform	7
3	Five gates and what each one is for	8
4	A reference stack, and the layer everybody skips	10
5	Deployment that treats people as participants	12
6	A scorecard with seven dimensions	13
7	Scale, hold or stop	15

END MATTER

	Method, evidence and limits	16
	Sources considered and set aside	17
	About the author	18
	References	19
	Further sources consulted	20

ABSTRACT AND SUMMARY OF THE ARGUMENT

Nine months of experiments, and the question of which ones deserve a budget

Three days ago an enterprise product with contractual data terms became available, which changes the procurement conversation and leaves the evidence where it was.[1] VENDOR Most large organisations now hold a collection of pilots that were never designed to be compared with one another, and a growing expectation from their boards that something should be scaled.

This paper sets out how to convert that collection into a governed portfolio. It proposes five stage gates, a reference architecture for an assisted workflow, a balanced value scorecard and an explicit scale, hold or stop decision with written thresholds. The design principle throughout is that scale is a portfolio decision supported by evidence, and a compelling demonstration is not evidence.

A worker-centred section sits in the middle, where it can affect the design. The available international evidence on artificial intelligence in workplaces reports both improved job quality and material concern, from a survey conducted before these systems existed in their current form.[2] SURVEY Used carefully, it supports an argument about consultation and task redesign. Used as evidence about copilots, it supports nothing at all.

WHAT THIS PAPER ARGUES

Scale is a decision about a portfolio, taken against baselines and thresholds set in advance. A pilot that cannot say what it would take to stop has not produced the evidence that scaling requires.

- Portfolio comes before platform. Demand intake, a use-case taxonomy and duplication control cost nothing and prevent the common failure, which is four teams building the same retrieval assistant in parallel.
- Five gates, each with an exit criterion, take a pilot from problem proof through data readiness, controlled pilot and benefit proof to operational readiness. The fourth is where most portfolios stall and the fifth is where most organisations discover they have no support model. AUTHOR
- Enterprise terms are now purchasable, and they are not the same as assurance. The vendor states that business data and conversations are not used for training, and describes encryption at rest and in transit, single sign-on, domain verification, an administrative console and a compliance attestation.[1] VENDOR Each of those is a vendor claim, and each belongs in a contract.
- Seven dimensions need measuring rather than one. Time, quality, throughput, experience, adoption, risk and total operating cost belong on the scorecard together, because a scorecard reporting only the first will recommend scaling something that degrades the second.
- The stop threshold should be written before the pilot runs. Among surveyed organisations reporting any adoption of artificial intelligence, only 21 per cent said they had policies governing employee use of generative tools, which suggests the governance layer is thinner than the deployment layer almost everywhere.[3] SURVEY

SECTION 1

What changed over the summer

Two developments, of unequal weight. One changes what can be bought and the other describes a workplace that predates the thing being bought.

On 28 August 2023 OpenAI announced an enterprise offering, describing enterprise-grade security and privacy alongside expanded capability.[1] VENDOR The announcement states that the company does not train on business data or conversations and that models do not learn from usage, and it describes encryption in transit and at rest, single sign-on, domain verification, an administrative console with an analytics dashboard, and a compliance attestation.[1] VENDOR

Each of those is a vendor statement about a vendor product, and that is how it should be cited and how it should be treated commercially. The useful move is to take the list into a contract negotiation and ask which of the claims the supplier will warrant, what happens on breach, and what notice is given when the underlying model changes. A published announcement is a starting position.

> WHAT THE ANNOUNCEMENT DOES NOT SAY
>
> It does not mention data residency or regional hosting, configurable retention, audit logs or an audit interface, or integration with data loss prevention and legal discovery tooling. It also discloses no pricing and no customer numbers.
>
> None of that means the capabilities are absent. It means this source does not show them, and an organisation that lists them in a board paper on the strength of the announcement is citing something that was never said.

The labour market evidence, and its actual scope

The Organisation for Economic Cooperation and Development published its Employment Outlook on 11 July 2023 with a chapter on artificial intelligence and the labour market.[2] SURVEY The underlying employer and worker surveys covered 5,334 workers and 2,053 firms, in manufacturing and finance only, across seven countries, and the fieldwork ran from mid-January to mid-February 2022.[4] SURVEY

That fieldwork predates the public availability of conversational generative systems by some ten months. The chapter is careful about this and its readers frequently are not. It is evidence about workplaces that had adopted earlier forms of artificial intelligence, and it is not evidence about copilots, assistants or anything released since November 2022. Cited for what it covers it is valuable. Cited as early evidence on generative deployment it is simply the wrong document.

Within its scope the findings are worth carrying. Around 80 per cent of workers using artificial intelligence said it had improved their performance at work, against 8 per cent who said it had worsened it, and across four working-condition measures users were more than four times as likely to report improvement as deterioration.[4] SURVEY On the other side, 20 per cent of finance workers and 15 per cent in manufacturing knew someone at their company who had lost a job because of it, and 66 per cent of employers in finance and 72 per cent in manufacturing said it had automated tasks workers used to perform.[4] SURVEY

TWO WORRY FIGURES THAT ARE NOT THE SAME

The Outlook's editorial reports that three in five workers are worried about losing their job to artificial intelligence within ten years. The underlying working paper reports that 19 per cent in finance and 14 per cent in manufacturing are very or extremely worried, with around half not worried at all. Both are accurate and they measure different things. Placing the first beside the 80 per cent improvement figure produces a contrast that the data does not contain.

SECTION 2

Portfolio before platform

The first scaling decision is which requests are the same request. The tool comes later.

By the autumn a large organisation typically holds somewhere between 15 and 60 proposals, arriving through no common route, described in incompatible language, and sponsored at wildly different levels. Some fraction of them are the same idea. Establishing which is the cheapest useful work available.

1. A single intake route, meaning one form, eight fields, a named reviewer and a response within two weeks. The response may be no, and a route that only ever says yes generates no information.
2. A use-case taxonomy that classifies by what the system does, whichever department asked. Retrieval over internal documents, drafting from a brief, classification, code assistance, and summarisation of supplied material will cover most of the intake.
3. Duplication control, which means grouping the proposals by taxonomy class before reading them individually. Four teams building retrieval assistants over four document sets is one problem, and treating it as four wastes the evaluation effort that matters.
4. Sponsorship at the right level, so that each surviving proposal has an operational owner who would be accountable for the workflow whether or not any technology were involved.

The fourth decides the outcome. A proposal sponsored by a technology function with no operational owner can pass the first four gates and will fail the fifth, because at operational readiness somebody has to accept the thing into a service they run. Establishing that at intake costs one conversation. Discovering it at gate five costs the whole pilot. AUTHOR

SECTION 3

Five gates and what each one is for

A stage gate is a decision point with an exit criterion. Without the criterion it is a meeting.

The gates below are deliberately conventional. Nothing about generative systems requires a new governance vocabulary, and organisations that invent one generally end up with two processes and no clarity about which applies. What is specific to this technology sits in the evidence each gate demands.

THE PORTFOLIO FUNNEL AND ITS FIVE GATES

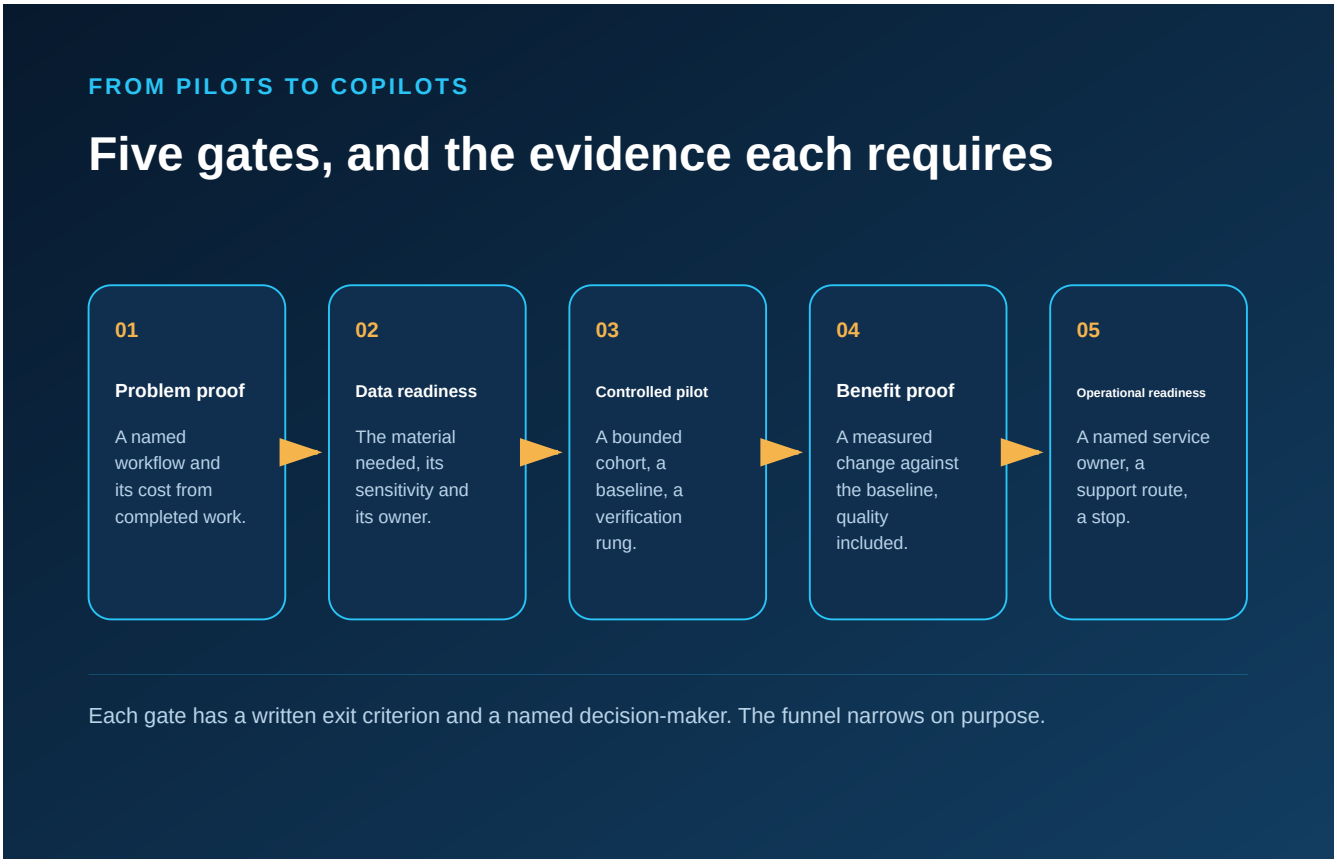


FIGURE 1 Each gate has a written exit criterion and a named decision-maker. The funnel narrows on purpose, and a portfolio in which nothing is ever stopped is not being governed.

My framework. AUTHOR

FIVE GATES, WITH THE EVIDENCE EACH REQUIRES

GATE	WHAT MUST BE DEMONSTRATED	WHAT COMMONLY GOES WRONG
Problem proof	A named workflow, its current cost from completed work, and a person accountable for that workflow who wants it changed.	The sponsor is a technology function and no operational owner has asked for anything. The pilot then has no home at gate five.
Data readiness	The material the system needs, its sensitivity classification, the retention terms that apply, and who may see it.	A prototype is built against a copy of a dataset that nobody had authority to copy, which is discovered at the security review.
Controlled pilot	A bounded cohort, a baseline, a verification method and a written stop condition, run for long enough to see a second week.	The pilot becomes a demonstration. Enthusiastic users, no baseline, and a result that cannot be compared with anything.
Benefit proof	A measured change against the baseline on at least three of the seven scorecard dimensions, including quality.	Time saved is reported and quality is asserted. This is the gate at which most portfolios quietly stop applying their own rules.
Operational readiness	A named service owner, a support route, a monitoring plan, a change process for model updates, and a retirement path.	Nobody owns it in production. The pilot team disperses and the first model update changes behaviour with no one watching.

The gate where portfolios actually fail

Benefit proof is where the discipline breaks, and it breaks in a recognisable way. The pilot reports a time saving, which is real. Quality is described as maintained, which is an assertion, because measuring it requires a blinded comparison that nobody scheduled. The gate is passed on one dimension out of seven and the deployment inherits an unexamined assumption about the other six.

The repair is cheap and unpopular. Require the blinded quality comparison as a condition of entering the pilot rather than as a deliverable of it, and budget the reviewer time at the start. Where a sponsor will not fund the comparison, that is information about how much they believe the benefit claim. AUTHOR

SPECIFICATION

A blinded quality comparison that a small team can actually run

- Take 30 completed outputs, 15 assisted and 15 not, from the same task class and the same period.
- Strip identifying formatting so a reviewer cannot tell which is which, which usually means normalising layout and removing any tool artefacts.
- Have two experienced reviewers rank them independently against the standard that applied before the pilot.
- Report the disagreements between reviewers as well as the result. Where they disagree more than they agree, the task has no stable quality standard and the benefit claim rests on nothing.

SECTION 4

A reference stack, and the layer everybody skips

The stack has six layers. The model is one of them, and whether the thing is safe is decided elsewhere.

Architecture conversations in this field tend to be conversations about models, which is the layer most likely to be replaced and the layer over which an enterprise has least influence. The layers that determine whether an assisted workflow is defensible sit below and around it.

A REFERENCE STACK FOR AN ASSISTED WORKFLOW

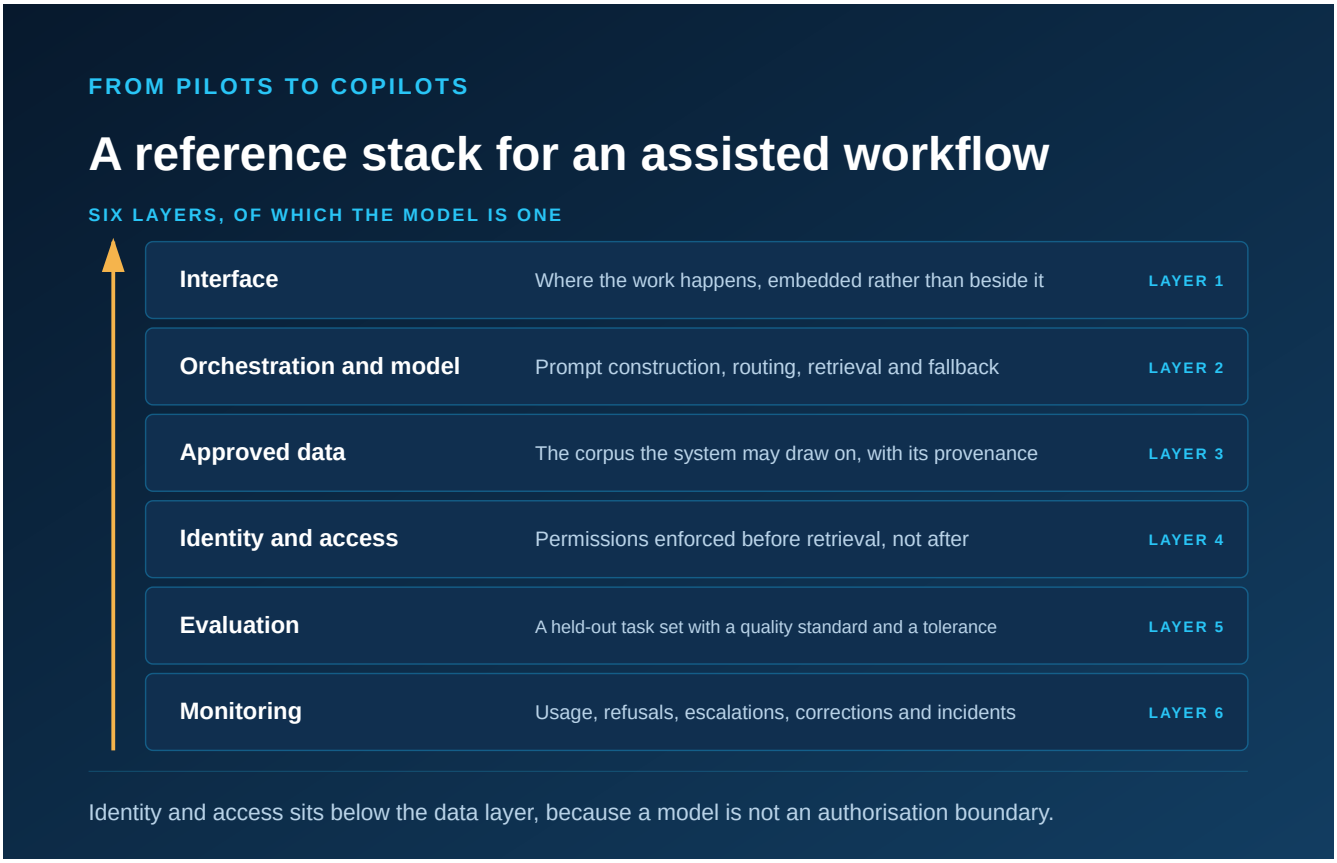


FIGURE 2 Six layers, of which the model is one. Identity and access sits below the data layer because a model is not an authorisation boundary and cannot be made into one.

My framework. AUTHOR

1. The interface is where the work happens. Assistance embedded in an existing tool is used, and assistance in a separate destination is visited twice and abandoned.
2. Orchestration and model covers prompt construction, routing, and the model itself. Treat the model as replaceable from the first design, because it will be replaced.
3. Approved data is the corpus the system may draw on, with its classification and its retention terms. Scope it narrowly at first, since a narrow corpus produces answers that can be checked.

4. Identity and access means permissions enforced before retrieval, so that the system can only reach what the individual user could already reach.
5. Evaluation is a held-out set of representative tasks with known good answers, run before release and again after any model change.
6. Monitoring covers usage, refusals, escalations, user corrections and incidents, with somebody whose job it is to look at them.

Why identity sits below data

The most common serious design error this year is to connect a capable model to a document store and filter the results afterwards. It fails because the filtering happens after the model has already seen the material, and because a sufficiently determined prompt can surface content the filter was meant to suppress. Permissions must be applied at retrieval, so that the material never enters the context. A model is not an authorisation layer and no amount of instruction makes it one. AUTHOR

The evaluation layer is new. Conventional software is tested against expected outputs and passes or fails. A probabilistic system requires a representative task set, a quality standard, and a tolerance, and most organisations have none of the three. The published risk frameworks name the requirement without specifying the method. That is a reasonable division of labour, and it leaves the method to be built locally.[5] PRIMARY

SECTION 5

Deployment that treats people as participants

Consultation, task redesign and a route to challenge an output. The evidence here is thin and the argument does not depend on it.

The international evidence discussed in section one reports that among employers who had adopted artificial intelligence, 43 per cent in finance and 45 per cent in manufacturing consulted workers or their representatives.[4] SURVEY Set against the two thirds of the same employers reporting that it had automated tasks workers previously performed, the asymmetry is the finding.

The practical case for consultation does not rest on the survey. It rests on something simpler. The people doing the work hold the information the pilot needs. They know which parts of the task are hard, which outputs get quietly rewritten downstream, and where the existing process already fails. A pilot designed without them will measure the wrong task competently.

SPECIFICATION

Four things to establish before an assisted workflow reaches a team

- What the task looks like afterwards, described by the people who do it now.
- Which verification rung applies, and whether the time for it has been added to the schedule or silently removed from it.
- A route to challenge an output, and a route to decline to use the system on a particular piece of work, both without it counting against the person.
- What happens to the time saved, answered in advance. An unanswered version of this question is how a productivity gain becomes a vacancy nobody decided on.

Accessibility, briefly

Assisted drafting removes a real barrier for some people and introduces one for others, and I am not aware of any published work that measures either effect in a workplace setting. The observation is a practitioner's note, and it belongs in the pilot's experience measure. AUTHOR

SECTION 6

A scorecard with seven dimensions

One number will always be available and will always be time. The other six are what stop it being misleading.

The pressure to reduce the assessment to a single figure is constant and comes from people who have to compare this investment with others. The answer is to give them a scorecard with a small number of dimensions rather than a composite index, because a composite index hides exactly the trade-off the decision turns on.

THE SEVEN-DIMENSION VALUE SCORECARD



FIGURE 3 Time and throughput are what sponsors ask for. Quality, risk and total operating cost are what determine whether the first two survive contact with production.

My framework. AUTHOR

THE SEVEN DIMENSIONS, AND HOW EACH IS ACTUALLY CAPTURED

DIMENSION	WHAT IS MEASURED	WHERE THE MEASUREMENT USUALLY FAILS
Time	Elapsed time on a defined task class, before and during, same people.	The baseline is remembered rather than recorded, which inflates the gain by roughly the size of the gain.
Quality	Blinded preference between assisted and unassisted output, judged by a reviewer who does not know which is which.	Replaced by user satisfaction, which measures how the tool felt rather than what it produced.
Throughput	Units of work completed per period at the team level, not the individual level.	Individual gains are summed and reported as team gains, ignoring the queue that did not move.
Experience	Whether the work became better or worse to do, asked of the people doing it.	Collected once, during the novelty period, and never repeated.
Adoption	Repeat use on real work, not logins and not first use.	Licence activation is reported as adoption. It measures procurement.
Risk	Incidents, near misses, escalations, and unverified outputs reaching a decision.	No near-miss category exists, so the count is zero until the first incident.
Total operating cost	Licence, compute, integration, evaluation, review time and support, per delivered unit.	Review time is excluded, which is the single largest cost in a well-governed deployment.

The dimension that is nearly always wrong

Total operating cost is routinely reported as the licence fee. In a well-governed deployment that fee is a minority of the real figure. The costs that get omitted are the review time required by the verification rung, the evaluation effort at each model change, the integration and identity work, and the support route. A deployment whose business case excludes review time has assumed away the control that made it defensible, and it will report a return that cannot be reproduced.

Where a summer survey offers a useful external reference, it is this. Of organisations surveyed in April 2023, around a third reported regular use of generative tools in at least one business function.[3] SURVEY Among the narrower group reporting adoption of artificial intelligence of any kind, 21 per cent said their organisation had established policies governing employee use of generative tools.[3] SURVEY The two figures rest on different denominators and cannot be subtracted into a single gap. Both describe a self-selected panel, and what they describe between them is a population in which regular use is already common and written policy is not.

SECTION 7

Scale, hold or stop

There are three outcomes, and the thresholds for each are written in advance, along with a retirement path designed at the same time as the launch.

The decision at the end of a pilot should be capable of going three ways, and in most organisations it is capable of going one. That is a consequence of never having written down what a stop would look like, which leaves stopping looking like an admission.

1. Scale where there is a measured improvement on at least three dimensions including quality, no unresolved incidents, a named service owner who has accepted it, and a support route that exists. Scaling means a larger cohort with the same measurement, and it does not mean an enterprise licence.
2. Hold where the evidence is ambiguous, and it usually is. Extend with a tightened method and a specific question, and set a date at which holding again is not available.
3. Stop where a threshold was breached, the quality comparison failed, or no operational owner will take it. Write down what was learned and put it somewhere the next team will find it.

Designing the retirement path first

Every assisted workflow put into production this year will be retired or rebuilt within a few years, because the underlying products are changing faster than enterprise software normally does. Designing the exit at launch is therefore ordinary prudence. What data would have to be extracted, what the workflow reverts to, who would notice if it stopped working, and how long the organisation could operate without it.

> **WHAT A GOOD PORTFOLIO LOOKS LIKE AT THE END OF THE YEAR**
>
> Somewhere between two and four workflows in production, each with a named owner and a monitored evaluation set. A larger number held or stopped, with written reasons. A total operating cost figure that includes review time. And an evaluation capability that can be pointed at the next proposal without being rebuilt.
>
> What it does not look like is an enterprise-wide licence with high activation and low repeat use, which is the most likely outcome of the current market pressure and the hardest to reverse once the budget has been committed.

One open question sits under all of this. Nobody knows whether the governance overhead described here falls as the technology matures or becomes a permanent cost of operating it. I would guess the former for well-bounded tasks and the latter for anything touching a customer, and a guess is all that it is.

END MATTER

Method, evidence and limits

How the sources in this paper were chosen and dated, and what the reader should distrust.

How the sources were chosen

Primary institutional publications, peer-reviewed studies, one vendor announcement cited as such, and one industry survey cited with its sampling method stated. Consultancy projections of market size do not appear, because none I could find rests on observed enterprise deployment at scale.

A note on the labour market chapter

The Employment Outlook chapter used here draws on surveys whose fieldwork ran in the first quarter of 2022, covering manufacturing and finance in seven countries, with worker responses collected through opt-in online panels. The organisation itself notes that the literature it reviews largely predates the current generation of generative systems. This paper uses it for what it covers and says so in the text.

On vendor claims

The enterprise product announcement of 28 August is the only vendor source carrying weight here, and every claim taken from it is attributed to the vendor in the sentence that uses it. Where the announcement is silent, section one says so explicitly, because an enterprise reading a capability into a silence is the most common procurement error in this market.

Three things I could not find

No independent measurement of enterprise deployment at scale exists. No published evaluation method for assisted professional work that an organisation could adopt without designing it could be located. And no evidence addresses whether governance overhead falls with maturity, which is the open question section seven ends on.

Instruments rather than findings

The five gates, the reference stack, the seven-dimension scorecard and the blinded comparison procedure are my own. The claim that identity must be enforced at retrieval rather than after generation is a design judgement supported by argument, and no published study tests it.

A declaration of interest

I advise on technology adoption and could plausibly supply the portfolio and evaluation work described here. The paper argues against enterprise-wide licensing, the larger transaction, and readers can decide what that is worth.

END MATTER

Sources considered and set aside

Five sources a reader might expect to see cited here. I read them and set them aside for the reasons given.

SOURCE	PUBLISHED	WHY IT WAS SET ASIDE
KPMG, Generative AI survey	3 August 2023	Executives reporting intentions on a self-selected panel. Intentions are cheap, and the survey does not follow up.
Salesforce, Generative AI snapshot research series	June 2023	Vendor-commissioned polling of workers in a market the vendor sells into. Read for direction and not for magnitude.
Gartner, Hype Cycle for Artificial Intelligence, 2023	19 July 2023	Places generative artificial intelligence at the peak of inflated expectations, which is an opinion I share and cannot cite as evidence.
OpenAI, ChatGPT plugins announcement	23 March 2023	A capability that changes what an assistant can reach, opened to paying subscribers in May and described only in blog posts. Section four treats integration as something to design rather than something to buy.
Deloitte, State of AI in the Enterprise, fifth edition	18 October 2022	The last large enterprise survey before the November release. It measures a different technology and I have not used it as a baseline.

END MATTER

About the author

Paul Forrest advises boards and executive teams on the adoption of new technology in operating businesses, with a practice grounded in delivery rather than in advisory work alone.

His background is in business process reengineering, which he has practised since the 1990s, and in applied natural language processing, which he has worked on since 2014. The stage-gate structure in this paper is unfashionable and deliberate. Reengineering programmes failed in the 1990s for reasons that had nothing to do with the technology and everything to do with nobody owning the process afterwards, and the fifth gate exists because of that.

He writes on the condition that a claim carries its evidence, and this series is an attempt to hold that line while a subject moves quickly.

END MATTER

References

Five works cited in the text, each with its original date of publication. None is later than 29 August 2023.

1. **OpenAI.** *Introducing ChatGPT Enterprise.* Product announcement. Cited for the statements on training, encryption, single sign-on, domain verification, the administrative console and the compliance attestation, each attributed to the vendor.
   PUBLISHED 28 AUGUST 2023

2. **Organisation for Economic Cooperation and Development.** *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market.*
   PUBLISHED 11 JULY 2023

3. **McKinsey and Company.** *The state of AI in 2023: Generative AI's breakout year.* Survey of 1,684 respondents, fielded 11 to 21 April 2023, self-selected panel rather than a probability sample.
   PUBLISHED 1 AUGUST 2023

4. **Lane, M., Williams, M., and Broecke, S.** *The impact of AI on the workplace: Main findings from the OECD AI surveys of employers and workers.* OECD Social, Employment and Migration Working Paper 288. Survey of 5,334 workers and 2,053 firms in manufacturing and finance across seven countries, fieldwork mid-January to mid-February 2022.
   PUBLISHED 27 MARCH 2023

5. **National Institute of Standards and Technology.** *Artificial Intelligence Risk Management Framework 1.0.* NIST AI 100-1, with the accompanying Playbook.
   PUBLISHED 26 JANUARY 2023

END MATTER

Further sources consulted

Nine sources I read for this paper and did not cite. They are listed so that the reading behind it can be seen in full, and none is later than 29 August 2023.

- **Peng, S., Kalliamvakou, E., Cihon, P., and Demirer, M.** *The Impact of AI on Developer Productivity: Evidence from GitHub Copilot.* arXiv:2302.06590.
  PUBLISHED 13 FEBRUARY 2023
- **Noy, S., and Zhang, W.** *Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence.* Science, volume 381.
  PUBLISHED 13 JULY 2023
- **Brynjolfsson, E., Li, D., and Raymond, L.** *Generative AI at Work.* National Bureau of Economic Research working paper 31161.
  PUBLISHED APRIL 2023
- **World Economic Forum.** *The Future of Jobs Report 2023.*
  PUBLISHED 30 APRIL 2023
- **Microsoft.** *Will AI Fix Work?* Work Trend Index annual report.
  PUBLISHED 9 MAY 2023
- **McKinsey and Company.** *The economic potential of generative AI: the next productivity frontier.* Cited as a modelled projection rather than a measurement.
  PUBLISHED 14 JUNE 2023
- **Microsoft.** *Introducing Bing Chat Enterprise, Microsoft 365 Copilot pricing, and Microsoft Sales Copilot.* Cited for the published per-user price and for the early access status of the product at that date.
  PUBLISHED 18 JULY 2023
- **Eloundou, T., Manning, S., Mishkin, P., and Rock, D.** *GPTs are GPTs.* arXiv:2303.10130.
  PUBLISHED 17 MARCH 2023
- **Stanford Institute for Human-Centered Artificial Intelligence.** *Artificial Intelligence Index Report 2023.*
  PUBLISHED 3 APRIL 2023

WHERE THIS GOES NEXT

The next paper takes governance on its own terms

This paper treated controls as a condition of scaling. The fifth will treat them as the subject, and it is due at the end of November.

The United Kingdom has called a summit on frontier risk for the autumn, the White House has extracted voluntary commitments from the largest developers, and the European proposal is in trilogue. The next paper will be written to whatever has actually been issued by then.

WHAT TO HAVE IN PLACE BEFORE THE NEXT PAPER IS USEFUL

- A single intake route with a taxonomy, and a portfolio view showing what is duplicated.
- At least one pilot that has passed benefit proof with a blinded quality comparison rather than an assertion.
- A total operating cost figure that includes review time, even if the figure is uncomfortable.

Paul Forrest · 31 August 2023 · Version 1.0 · Research closed 29 August 2023