waracle®

SOFTWARE DEVELOPMENT LIFECYCLE

BI-ANNUAL
Update Report
2026

THE AI ENABLED SDLC

6 MONTHS ON

1 SIX MONTHS ON
Still right, now sharper

In May, we argued that AI was reshaping how software is built in regulated organisations, and that the gains would come from changing systems, changing behaviours, as well as buying tools. Six months of delivery later, that argument holds.

What has changed in the intervening period, is how precisely we can say where organisations are encountering issues and what to do about it.

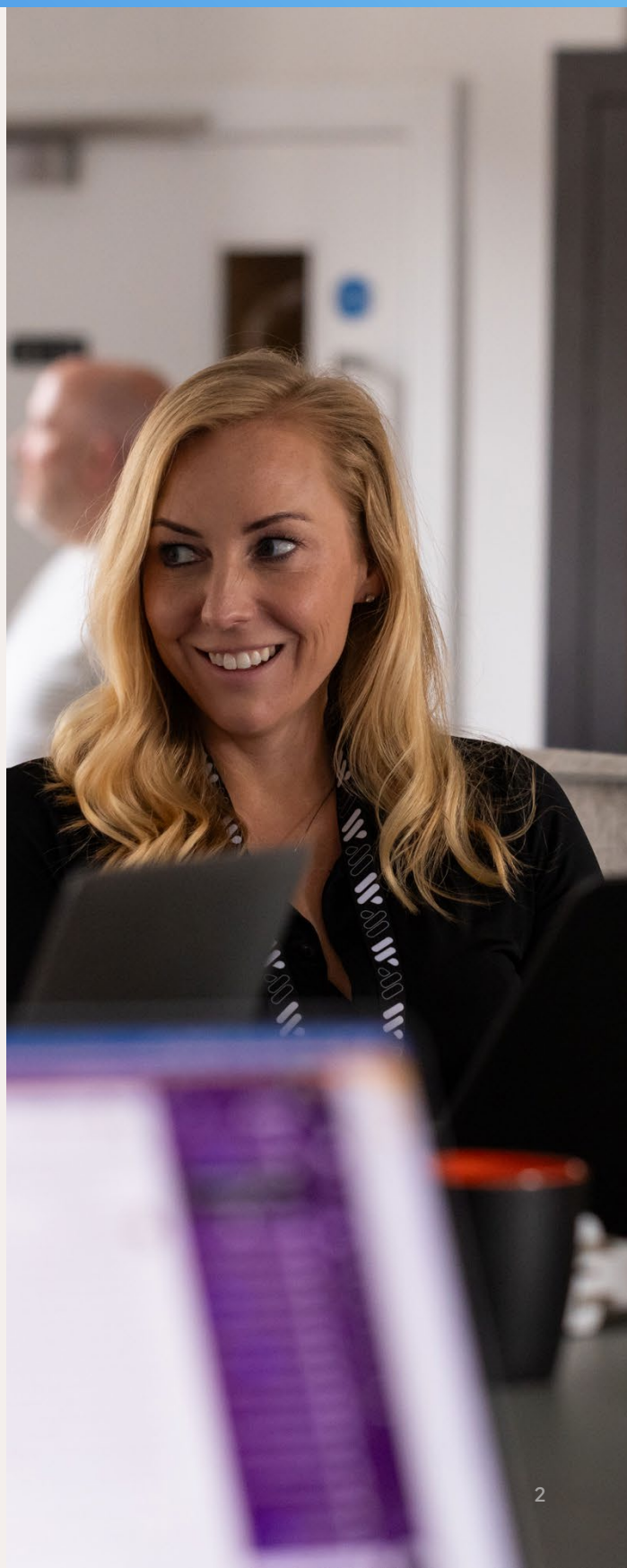
When we wrote the first paper, most boardroom conversations about AI were about access: which assistant, which licence, which vendor, what models and for which teams. That question is now largely settled. The tools are good, they are cheap when used in the right way, and nearly every organisation we work with has them.

What hasn't been settled for many organisations, is what happens next.

The speed of writing code is no longer a constraint. Agentic tools can produce a component, a function or a test suite in the time it takes to make a coffee. The delay has moved into the human process around the code: scoping, context, architectural alignment, code review, security sign-off, release governance and incident handling.

Amdahl's Law describes what happens when you make one stage of a pipeline dramatically faster and leave the rest alone. The whole thing gets a little quicker, and the bottleneck moves somewhere else. That is where most organisations are sitting today.

Addressing the bottlenecks.



WHY AN ADDENDUM – SIX MONTHS ON

We said much of this in May, and we still believe it. Republishing the same argument with fresher numbers would waste your time, and not add to the conversation.

So we have added to the original in six areas where our delivery work has sharpened our view, and where we think the market is still working things out.

Maturity. Most AI maturity frameworks put an organisation on a single ladder with a vague summit. Real organisations are a jagged frontier, and one score hides that. We set out a multi-vector view that is more honest and more useful.

Optionality. The debate about model cost is mostly a distraction. The question that matters is whether you can change your mind, and what it would cost you to do so.

Return on Investment. Most AI productivity claims would not survive an audit. We set out what a defensible gain looks like, and we apply the same standard to our own numbers.

AI-enabled and AI-native. These terms are used loosely. We define them, explain why culture decides who makes the move, and describe the fluency framework we use to move our own people along.

Agents in production. We run agentic systems in production for clients in regulated sectors. So we can speak from a place of certainty on the topic.

Operationalising AI. A year and a half ago, we created a Chief AI Officer role that was designed to make itself unnecessary. It has. We explain what replaced it, how we approach internal automation, and why we would not put agents on your org chart.



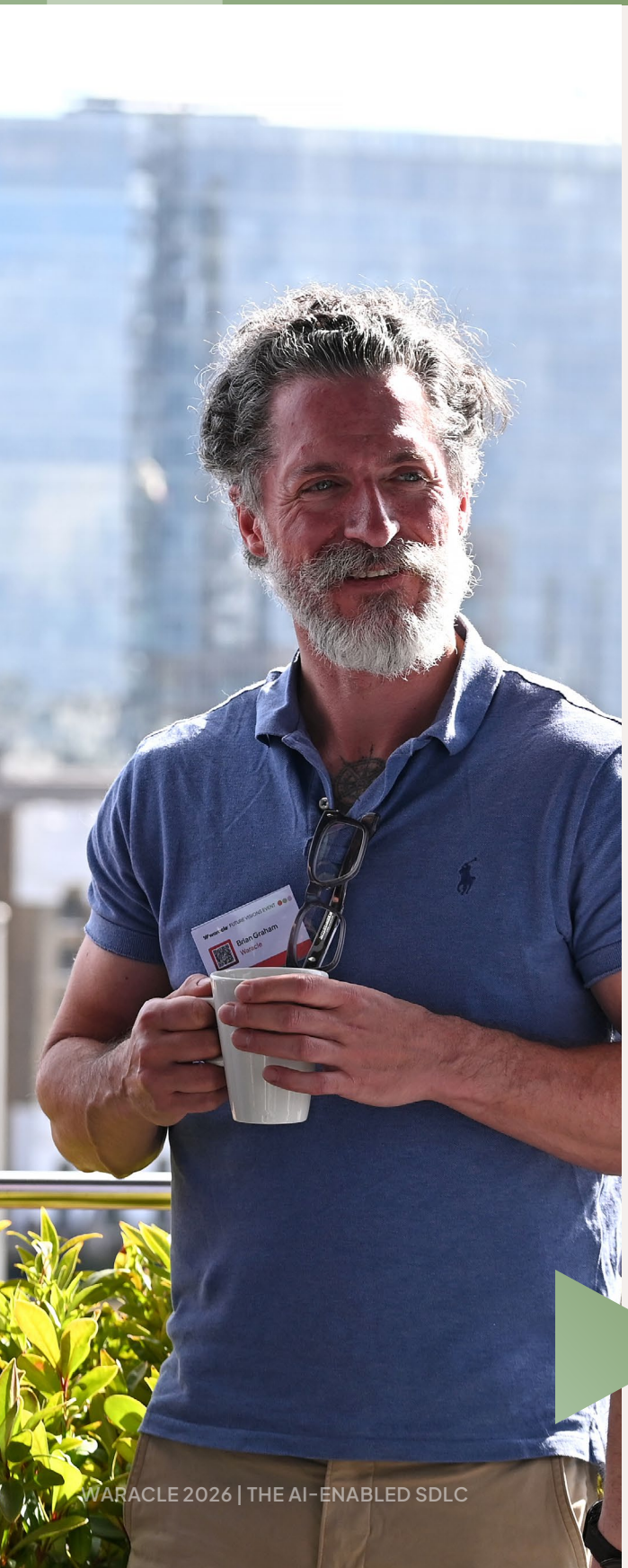
A QUICK NOTE ON HONESTY

Our values ask us to work in the open, so we will tell you where we are potentially behind as well as where we are ahead. No grandstanding here!

Our 10 - 20 % compounding efficiency figure from May sounded small to many, but was the right order of magnitude for enterprise clients. We believe that what we've seen in the last six months makes us confident that number is creeping up towards 25 - 35% efficiency gains, but as we will explore throughout this addendum, this is still very unevenly distributed and difficult to realise.

Each section closes with a short set of questions to ask of your own organisation. They are the questions we ask our clients, and the ones we keep asking ourselves … to ensure we keep ourselves honest.

2 AI MATURITY
Frameworks are measuring the wrong thing



Most AI maturity models give you a score out of five. Real organisations are uneven, and a single number hides exactly the information you need to decide what to do next.

Over the last year, we have read a lot of AI maturity frameworks. Nearly all of them share a shape. You start at "not using AI", move through "using some AI" and "using a lot of AI", and arrive at a summit nobody can quite describe. You answer the questions, sit through the workshop, and leave with a three out of five.

It has its place as a point-in-time exercise, but not so useful in identifying what to address next on your change journey.

What the simple maturity models get right and wrong

It is easy to be sniffy about these frameworks.

But they do deserve some credit. Common nomenclature and a framework of levels gives leaders a shared language and a sense of proportion. If an organisation has no sanctioned AI use at all, the most useful thing anyone can say is: we won't spend your time on autonomous agents yet, we will get you safely onto the first rung. That clarity saves everyone time and money.

The trouble starts when the level is treated as the whole picture. A linear score assumes an organisation moves as one, and organisations don't. A single number tells you roughly where you are, but nothing about what to do next.

So we use multiple views, like zoom levels on a map. The headline level shows the country. The profile across vectors shows the terrain. The distribution of people shows where everyone is actually standing.

Zoomed out: 5 LEVELS

Level	What it looks like	The sensible next move
0 Unstarted	No sanctioned use. AI is happening anyway, out of sight.	Safe access, a clear usage policy and a first set of approved tools.
1 Exploring	Licences issued, individual experiments, uneven enthusiasm.	Capture what works, pick two or three real problems and start measuring.
2 Enabled	Teams share practices, rules and some evaluation. AI sits alongside existing work.	Get one AI capability into production with clear ownership and measurement.
3 Operator	AI runs in production, owned, evaluated and governed.	Redesign processes around what AI makes possible.
4 Native	New work starts from AI by default, with evidence it produces better outcomes.	Keep evaluating. The bar keeps moving.

Most organisations we meet sit somewhere between levels 1 and 2. That is a perfectly reasonable place to be, provided you know it and are not buying level 4 solutions to level 1 problems.

The job is to understand the distribution of your teams across the levels and aim to shift them forward on the bell curve distribution.

Zoomed in: 7 VECTORS, 1 JAGGED FRONTIER

The headline level hides how much maturity varies inside a single organisation. In large regulated businesses we regularly find a team doing substantive, well-governed AI work a corridor away from a team still debating whether it is allowed to open a chat window.

Often neither knows the other exists.

One financial services leader told us that their main experiment with generative AI had been a product manager seconded three days a week to build prototypes. The prototypes were promising. Nothing in the organisation, technically or structurally, allowed them to go further, so they didn't. That pattern is common, and the failure sits in connection and governance.

Many organisations also have mature traditional machine learning: custom models, data science teams, years of production experience. Their generative AI strategy, meanwhile, amounts to buying assistant licences and hoping the impact appears. Those are different capabilities, and they need assessing separately.

We assess seven vectors independently. The result is a jagged profile, and the shape of the jag is where the insight lives.

VECTOR	WHAT IT ASKS	A PROVOCATION
Product Thinking	Do you know which problems AI should solve, and for whom?	Which AI experiment last reached a customer, and who decided it should?
Creating software	Can you confidently create working software using AI?	Do your teams feel like they're doing their best work? Are you shipping good software, or just bad software faster?
Deployment	Can you move AI work from the lab to production safely and repeatedly?	How long does an AI change take to reach production, and what stops it?
Evaluation	Do you measure output quality continuously, in development and in production?	Who owns the evaluations for your most-used AI workflow?
Governance	Are accountability, risk and approval paths clear?	If an AI-assisted decision is challenged tomorrow, who answers for it?
Context	Can your AI reach authoritative information, with the right permissions, and can you show what it used?	Could an agent find the authoritative version of your customer record, and could you prove which version it used?
Fluency	Do people know when to use AI, when not to, and how to judge the result?	What share of your people could explain why an AI output is wrong?

Change in action – From AI Operator to AI Native

Moving from Operator to Native takes you from outperforming to exceptional, and the gains are Sir Dave Brailsford's marginal ones, the equivalent of bringing your own pillow to the hotel during the Tour De France. If your maturity is lower, the bigger gains are yet to come, as the bigger impacts come earlier on the curve.

The pattern repeats across dimensions. Operators have built capable systems. Their discovery produces detailed, machine-readable specs for a very efficient feature factory. Their agentic tooling has grown agent upon agent. Their deployment works well in greenfield areas but leaves the legacy estate manually gated, and they connect agents to every enterprise tool on the assumption that more context means better results. Natives simplify. They know which work is no longer hard, so discovery focuses on the biggest ambiguities and the time saved goes into building the right product. They've curated the context that reaches their agents as an information architecture problem, and treat deployment as a distributed systems problem, with automated rollbacks, tracing and AI monitoring of the whole system.

The sharpest differences show up in how organisations measure, govern and grow their people. Operators track time saved and developer satisfaction, which only covers software production; Natives measure one North Star, cycle time from idea to customer, and test hypotheses as their bottlenecks shift. Operators bring governance in early but still queue for sign-off committees at the end; Natives build governance agents into development and deployment workflows, which needs a governance team treated as a real partner. Operators have pockets of fluency, often driven by top-down rules that breed superficial use; Natives hold leaders accountable for doing AI well and for caring about their people.

In every dimension, culture sets the ceiling.



The distribution: Where your people actually stand

Culture is the cumulative outcome of all your people's actions.

And behind every level and every vector sits a spread of people. Plot AI fluency across a department and you get something like a bell curve: a few sceptics, a few enthusiasts running ahead of the guardrails, and most people in the middle. Where that middle sits describes an organisation's maturity better than its best performer does. One star team building agents does not make the organisation operational if everyone else is still exploring.

The aim is to move the middle of this distribution towards confident, fluent use of AI. Training, having safe access to tools, supporting peer learning. Organisations that do this best recognise that they need to engage different groups. The people who are pushing are asked to go even further, asked what slows them down; but are also told to share what they've learned with other teams. The people who are lagging in AI adoption aren't told they're wrong, often they are grappling with the biggest challenges AI poses to your organisation, but are working together to make AI adoption last.

What good looks like

Use all three views, each for its own job. The headline level sets proportion: it tells you which conversations are worth having now. The vector profile finds your weakest link and gives it an owner. The distribution tells you where to invest in people. Replace "you are a three" with provocations that place a team in context and point to its next move.

Questions to ask this quarter:

1. What level would you honestly give your organisation overall, and would your teams agree?
2. Where are your most mature AI teams, and does the rest of the organisation know?
3. Could champions from those teams help others move faster, and have you given them the time to do it?
4. Which of the seven vectors is your weakest, and is anyone accountable for it?
5. Where did your last successful AI experiment go after the demo?

3 OPTIONALITY

The real question is whether you can change your mind



The debate about the best model is mostly noise. For enterprises in regulated sectors, the strategic question is control: what would it cost you to switch, and who gets to decide?

On one of our long-running delivery programmes, model spend has been roughly 1 per cent of total cost. People account for the other 99. If token prices halved tomorrow, the programme budget would barely register the change. If we improved how humans oversee and direct the AI, it would.

That ratio reframes the cost conversation. Token prices matter at scale, in high-volume products and automated workflows. For most enterprise software delivery, the expensive part of applying AI is the human commitment around it: oversight, context, governance, review and integration. The model is the cheap bit.

Optionality is about control

If cost is not the main argument, what is? Control. It comes in three parts.

The cost of switching. Models improve and change every few months. Vendors change pricing, terms and roadmaps. If your prompts, tools, agent logic and evaluation are all tied to one provider's ecosystem, moving costs you a rebuild.

Lock-in beyond the model. Lock-in increasingly lives in the layers around the model: proprietary agent frameworks, memory stores, tool integrations and hosted workflow builders. These are easy to adopt and hard to leave.

Proof of where data goes. In financial services and healthcare, data residency, retention and auditability decide the architecture. You need to be able to show a regulator where every prompt and response went, and why.

The middle ground is poorly mapped

Market noise tends to polarise. Frontier models are presented as either essential or too risky; open-weight models as either the future or a liability. The sensible points in between are rarely described.

A large bank is not going to run an open-weight model in production unless it has a safe way to host it inside a trusted boundary. Managed model platforms from the major cloud providers offer a middle ground: several model families behind one governed boundary, inside an estate and enterprise agreement your security team already understands. It is an unglamorous answer, and often the right one.

A large enterprise agreement with a hyperscaler or a single model provider is also a defensible first move.

Large UK financial institutions have signed exactly these deals. Buy deliberately, and build the exit while you use it.

Routing matters more than model choice

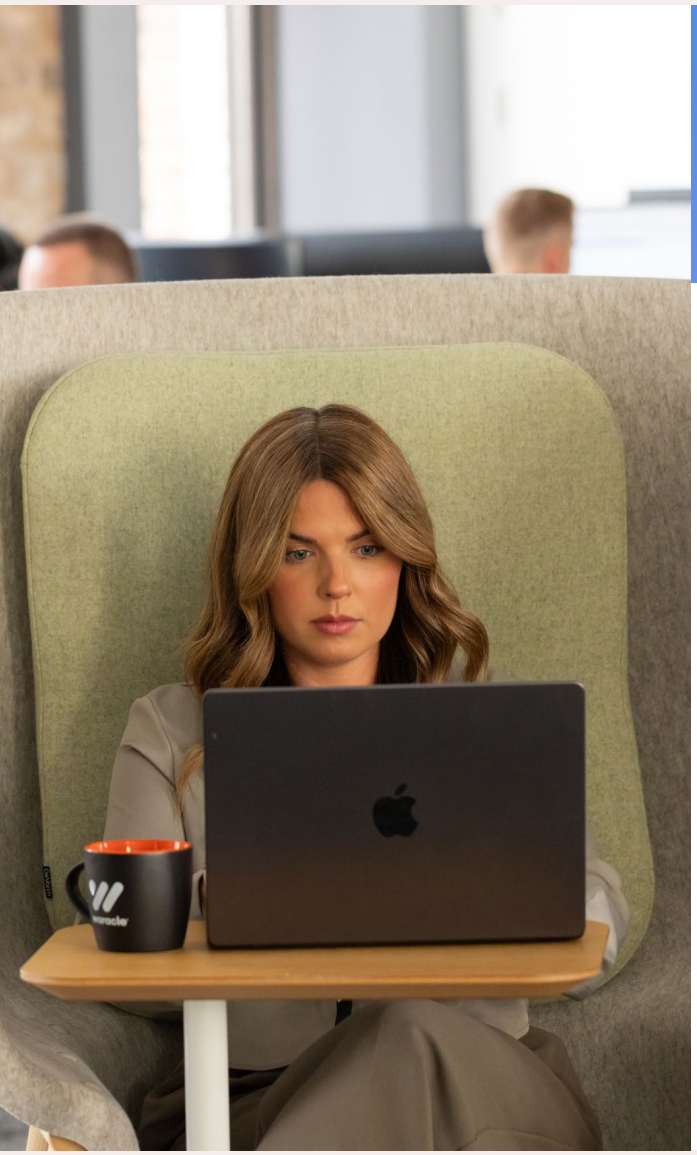
Ensure that you have control on how your agentic interactions are routed. Not just for governance, observability, and tokenomics; but also to ensure you can swap models without disturbing the user.

In the systems we build, the largest frontier models are reserved for the places where genuine ambiguity lives: interpreting unstructured requests, reasoning across messy documents, drafting. Small models, classical machine learning and deterministic code do the rest. The result is cheaper, faster, easier to evaluate and far easier to explain to an auditor.

Your evaluations are your exit door

The most overlooked part of optionality is evaluation. You can only switch models with confidence if you can prove the replacement performs as well on your tasks, with your data, against your standards. This is important not just for when you switch models, but also so you know when your model provider is changing their end.

We recommend owning a thin layer of your own around whatever models you use: prompts, tool definitions, monitoring for unsuccessful interactions with agents, evaluations, logging and routing rules held in your repositories, under your change control. That layer is what lets you be confident that changing your mind won't be disastrous.



THE TEST FOR THIS YEAR

Most organisations we ask cannot answer these questions cleanly:

1. **If your primary model provider doubled its prices or withdrew a model, what would switching cost, in weeks and money?**
2. **Who in your organisation has the authority to make that decision?**
3. **Do your evaluations run against more than one model today?**
4. **Can you show, for any AI interaction, where the data went and how long it is kept?**

If the answers are vague, that is where your optionality work starts. Data and model choices in regulated settings should always be reviewed with your risk, legal and data protection teams.

4 RETURN ON INVESTMENT

Most AI productivity claims would not survive an audit

"Two years work, done in 12 months" is a fishing story. The fish gets bigger with every telling. Here is what a defensible claim looks like, and how we hold our own numbers to the same standard.

Every week brings another claim that AI has made a team three, five or ten times faster. Some of those claims could be true. Most could not be defended in front of a sceptical finance director, and that matters, because those are the people who will decide whether AI investment continues once the novelty wears off.

The problem is rarely dishonesty. It is that time saved is easy to state and hard to prove. Data can tell almost any story if you choose what to count



What a defensible gain contains

A claim worth repeating states five things:

- → **The mechanism.** What did the AI actually do, and what did people stop doing as a result?
- → **The standard.** Was the output held to the same quality bar, the same tests and the same review as before?
- → **The baseline.** What did this work take before, measured the same way?
- → **The review burden.** How much human time went into checking, correcting and supervising the output?
- → **Who measured it.** Was it the team that wanted the result, or someone independent?

Without those five, a productivity claim is an anecdote.

Only some gains can be tied to money

Anthropic describes **its own internal use** in three buckets: smarter employees, automated workflows and products with AI components. Only the third can be tied to money with real confidence, because products can be A/B tested. Gains from more capable people and automated workflows are real, but the evidence is largely circumstantial.

There is also a perception gap. In a 2025 randomised controlled trial by METR, experienced open-source developers using AI tools took around 19 per cent longer to complete tasks, yet believed they had been around 20 per cent faster.

That was one study in one context, and tools have moved on since, but it still shows why self-reported gains are a weak form of evidence.

Four traps we see most often

Measuring what is easy.
Lines of code, tickets closed and prompts sent are easy to count and say little about value.

Speeding up one stage and calling it a system gain.
Faster coding does not mean faster delivery if review, testing and release processes are unchanged. It just pushes your bottleneck somewhere else.

Mistaking early-adopter enthusiasm for a trend.
The first team to adopt a tool is usually the most motivated. Their results rarely directly extrapolate to the rest of the organisation.

Ignoring rework and supervision.
Time spent checking, correcting and reworking AI output is a hidden cost, and it often lands on different people from those reporting the gain.

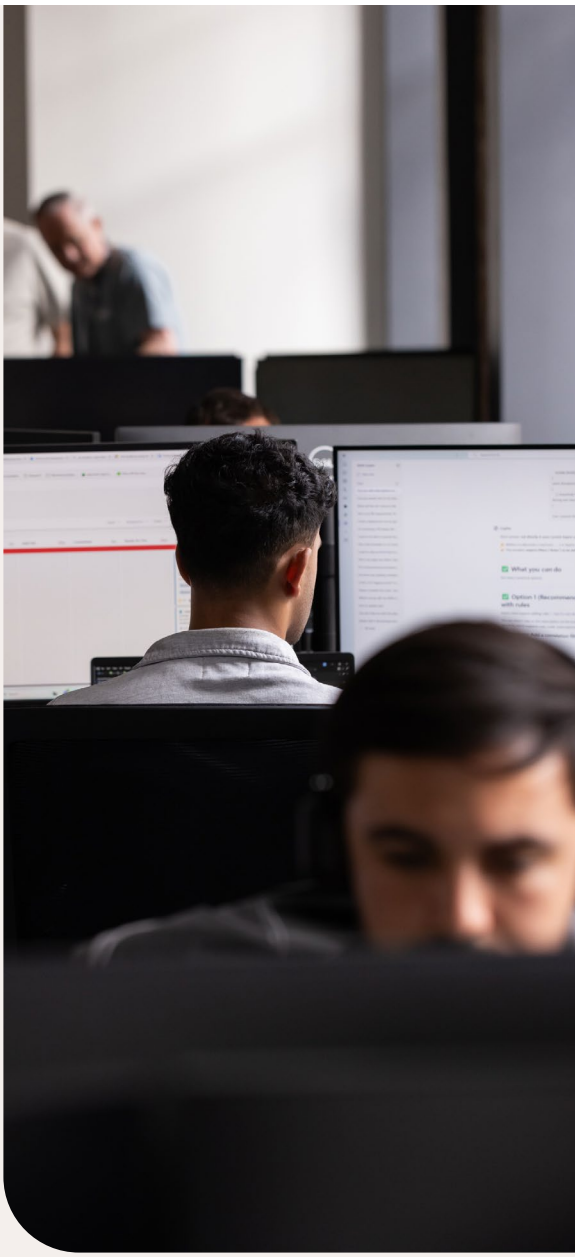
What good looks like

- → **Baseline first.** Measure the current process before you change it.
- → **Measure end-to-end.** Lead time from idea to production and change failure rate tell you more than any single stage metric.
- → **Keep evaluating in production.** Quality does not stop mattering at release.
- → **Count losses as carefully as wins.** Where AI made something slower or worse, record it.

For internal automation, hours saved is a reasonable headline metric, provided it comes with guardrails on output quality and end-user satisfaction. An automation that saves time and frustrates the people who rely on it has moved the cost around.

Questions to ask:

1. Which AI productivity claim in your organisation would you be comfortable defending to your audit committee?
2. Do you have a baseline for the processes you are changing?
3. Who measures the gain, and are they independent of the team claiming it?
4. What are you looking to prove out - A new way of working or an old way of working slightly refined?



5 AI-ENABLED AND AI-NATIVE
Culture decides who makes the move

AI-native is fast becoming a phrase in investor decks. We think it deserves a definition, and an honest admission that unless you incorporated in the last three years, few have got there yet.

The phrase AI-native is everywhere. It has been used about us, and we have found ourselves using it. So it seems fair to say what we mean by it.

AI-enabled means the technology sits alongside existing work and makes it somewhat faster. The process is the same; the people have better tools.

AI-native means new work starts from what the system should do, with AI as the default and the process built around it. When an AI-native team starts something new, it reaches for AI first, consistently and systematically, and knows from evidence that doing so produces better outcomes.



CULTURE IS THE DIFFERENTIATOR



Good tools are now close to universal. Outcomes are not, and the difference follows culture more than technical skill. Three things matter most: permission to experiment, the ability to learn from each other by watching how colleagues work, and no fear that automating your own work ends your job. AI amplifies the strengths and weaknesses an organisation already has.

Leadership shapes this more than most people admit. Where leaders are technically engaged, as at Shopify, where the CEO made AI use a baseline expectation, adoption moves faster.

Top down does not always work. We have seen organisations where encouraging personal experimentation did more for AI adoption than any top down programme.

But ultimately, for any business not incorporated in the last three years, there is a job to be done in evaluating standard operating practices and either augmenting or completely reimagining them for the AI age, this is no small undertaking and one that in many cases will take years, not weeks. On the path to an AI-native future though, it is a vital one.

At Waracle, we take an approach that is both bottom up and top down. Practitioners are encouraged to raise the bar of what's possible with AI and bring others with them. Leaders make sure that AI adoption counts; by making sure it is aligned to the desired outcome, not driven by vanity or fear.

Fluency, and the role of the engineer

We built a fluency framework to move our own people towards AI-native work. It runs across ten levels. The early levels cover foundations: choosing the right model for the task, understanding context windows, recognising the signs of generated content. The later levels are about judgement: when to use AI, when not to, and how to tell whether the output is any good.

Roles are more convergent. People with real agency, someone who can cross into adjacent roles and, crucially, understands what they produced. Intercom's experiments with people stepping into neighbouring disciplines shaped our thinking here. When designers understand implementation and engineers understand design, there are fewer handoffs and less of the friction that slows AI work in large organisations.

The practitioner's shifts with each phase of adoption: from prompt author, to rule architect, to specification author, to lead of a system of agents. Human effort moves from typing code towards defining intent, setting architecture and overseeing execution.

Questions to ask:

1. Could you describe what AI-native would mean for one specific team in your business?
2. Do your people feel safe automating parts of their own job?
3. Where does your bell curve sit, and what would move its middle?

6 AGENTS IN PRODUCTION

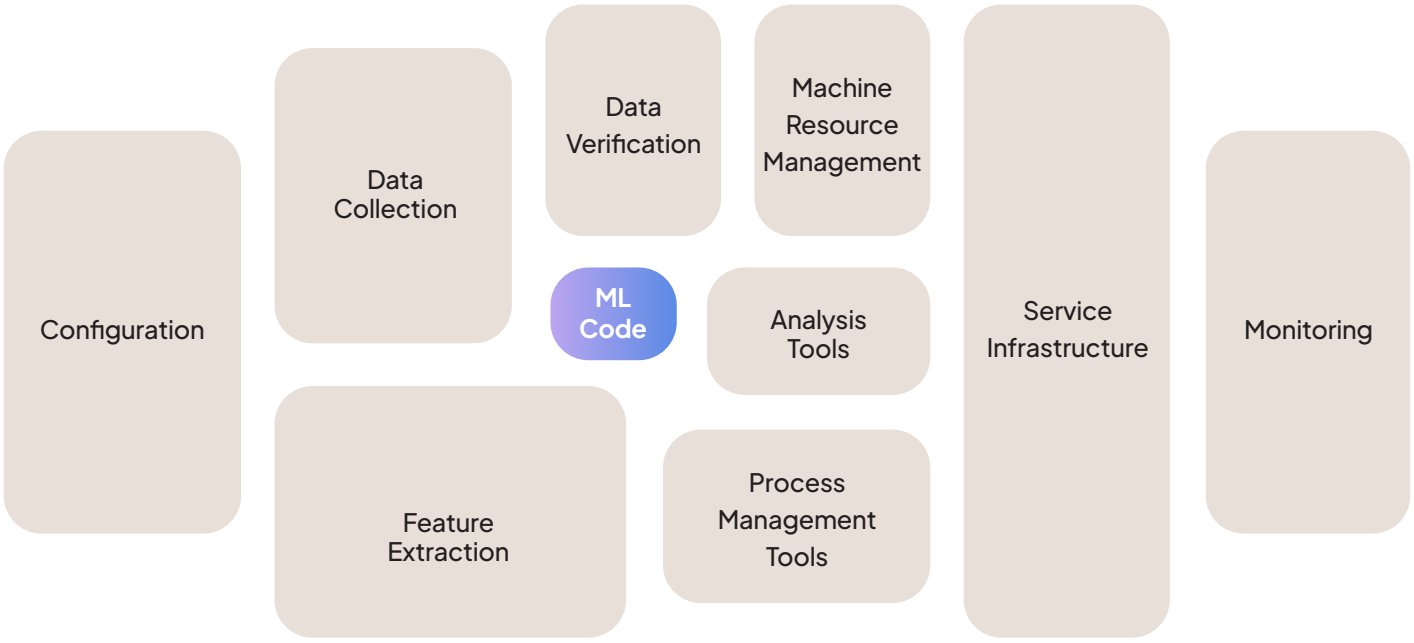
The demo is easy, a Tuesday in November is the test

We now run agentic systems in production for clients in regulated sectors. The lesson is unglamorous: the systems that work combine models with a lot of carefully engineered software around them.

This isn't novel. This figure from Hidden Technical Debt in Machine Learning (2016) would be exactly the same in 2026 if we substitute AI for ML.

A demo of an agent is a lovely thing. It reads an email, checks a record, drafts a reply and books a follow-up, all in a minute, all in front of an audience. Then comes a Tuesday in November: real inputs, real volumes, an outage in a downstream system, a malformed document, and an auditor asking why a particular decision was made at 3:14pm.

That is the test that matters, and it is where most agentic projects struggle.



Only a tiny fraction of Machine Learning systems is made up of the ML code. In comparison to the vast, complex infrastructure that sits around it, the purple box represents a very small part of the wider system.

Not every component should be a language model

The single most important design decision in an agentic system is what should not be done by a large language model. The systems we see working in production pair a large model, used where genuine ambiguity lives, with small models, classical machine learning, retrieval and well-designed deterministic code.

If a rule can be written down, write it in code. If a classification can be trained on historical data, use a model built for that. Reserve the large model for interpreting unstructured requests and reasoning across messy information. The result is cheaper, faster, more predictable and far easier to test.

An agentic system is a piece of software in which some components happen to be models. Treat it with the same engineering discipline as any other production system.

THE RESCUE CALL

Increasingly, organisations come to us after six months of clever model building with nothing underneath. The system impressed in testing and breaks on real inputs, real volumes and real audit requirements. The request is for a forward-deployed engineer to arrive and fix it.

We can do that work, and we do. It costs far more than designing the system properly at the start, and the client has usually lost months of momentum and some organisational confidence along the way.

The way to solve this problem isn't to double down and add more agents. Instead, pause and step back to understand the whole. It might be better to throw away some of your 'clever solution' and create a simpler solution that gets from end to end.

What production demands

The disciplines are familiar ones, applied to a new kind of component.

- → **Limits on what an agent can touch.** Each agent gets the narrowest permissions and tool access that let it do its job, nothing more.
- → **Safe retries.** Actions that might run twice must be designed so that running twice does no harm.
- → **Cost and loop ceilings.** Every agent has a budget and a maximum number of steps. Runaway loops are a when, not an if.
- → **Kill switches and rollback.** Tested regularly, documented, and owned by a named person.
- → **Stopping error cascades.** An agent does something unexpected, which puts pressure on downstream systems, which do the same all the way down, which requires designing firebreaks and backoffs in your system.
- → **Full traceability.** Every input, model call, tool use and output logged, so any decision can be reconstructed.
- → **Continuous evaluation.** Quality is measured in production against defined thresholds. When a metric leaves its acceptable range, the system escalates to a human or loops back for correction.
- → **Human sign-off where the stakes are highest.** Changes to data schemas, authentication, payments or customer-affecting decisions go through a person. Lower-risk changes can be automated as confidence grows.

Not many of these solutions are novel, nor specific to AI systems. Most of the good practice here comes from distributed systems design.

Accountability by design

In regulated sectors, someone must answer for every action the system takes. Fully autonomous action may not be acceptable at all in parts of financial services and healthcare, and existing obligations on accountability, conduct and clinical safety still apply when an agent is involved. So the audit trail is designed on day one, not added at the end. Regulatory, legal and, where relevant, clinical safety review should be built into the delivery plan from the start.

Human in the loop by itself isn't enough. Often the human checking the results is just piping them into another AI, and if they're not then they become a bottleneck. And as such, logging, auditing, and defining automated governance and compliance workflows are becoming more crucial parts of the solution.

Our recommendation

Bring in engineering judgement when the architecture is being decided, well before the rescue call. The questions that matter early are simple:

1. Which parts of this system genuinely need a large language model?
2. What is the worst action this agent could take, and what stops it?
3. How will you know, in production, that quality has slipped?
4. Who answers for the system's decisions, and can they see what it did?



7 FROM PIONEERS TO TOWN PLANNERS

Making AI ordinary

We created a Chief AI Officer role that was designed to make itself unnecessary. It has. What replaces it is less exciting and more important: AI becoming part of how the organisation normally works.

Simon Wardley describes organisations as needing three kinds of people at different times. Pioneers explore the unknown. Settlers turn early discoveries into something repeatable. Town planners industrialise it so it runs at scale.

When we appointed a Chief AI Officer, we needed a pioneer. The role existed to set direction, run experiments, build our fluency framework and get AI into delivery. It was set up from the start as temporary: if it worked, it would work itself out of a job.

It worked. Responsibility for AI now sits with our engineering, data and AI leadership and with our communities of practice. Our fluency framework is run by the people using it. Our agentic engineering work is part of normal delivery. Many organisations had a Chief Cloud Officer a decade ago; very few do now, because cloud became how things are done. We think AI leadership will follow a similar path, and we would encourage clients to plan for that from the outset.

This is not to say AI has become business as usual. It has not yet dissolved into everyone's job.

The settlers' work is underway.

Automating your own processes

The first internal automations in any organisation are usually someone scratching their own itch, much like the spreadsheet macros of a generation ago. The comparison only goes so far. Macros were deterministic; AI automations are not, and someone has to check the results.

Doing this well starts before any technology. Do not automate something that should not be done at all. Simplify the process with its end user in mind, then automate it. Make a simple business case and avoid shiny object syndrome. Sometimes the business case is a second-order benefit, such as proving a platform component or gaining production experience, and that is fine if it is stated openly.

Then invest in a platform, so that automation is not confined to a small group of specialists. The shift you want is from "ask IT for a tool" to "I can build my own automation safely". That platform needs:

- → **A register and monitoring,** so you know what automations exist and whether they are working.
- → **Support for evaluation and quality testing.** Many agent builders are not yet as mature as the low-code platforms that came before them.
- → **Long-running state,** for example tracking human sign-offs across days.
- → **Simple, governed integration routes** into other systems, a safe way to deploy self-built apps, and a design system so they look and behave familiarly.
- → **Canonical data.** AI copes well with messy data, but authoritative sources with a clear semantic layer improve accuracy and reduce hallucination. Good security, information and data architecture do not stop mattering.

Should agents appear on the org chart?

Some organisations are giving agents a place on the org chart. A study from Boston University and BCG found around a quarter of organisations surveyed already do so. The appeal is clear accountability, capacity planning and escalation paths.

We would not recommend it as a general approach.

AI multiplies the grey areas of an org chart. The same study found a "hot potato" effect: when outputs came from an "AI employee" instead of an AI tool, managers caught 17 per cent fewer errors, requested more review from others and took less accountability themselves. Separate research found people trusted and rewarded agents differently depending on how human, and which gender they appeared (See the **University of Limerick study**). Anthropomorphising agents imports human biases into systems that should not have them.

What organisations are really after is a map of how work and information flow. Org charts were never good at that. Process mining, value stream mapping and service mapping are better tools for the job. Keep a clear line between people and agents, and keep accountability with named humans.



WHAT COMES NEXT

The next phase of AI adoption will be quieter than the last. Fewer announcements, more platforms, clearer ownership, better measurement. That is what operationalising looks like.

If you would like to talk through where your organisation sits on any of this, from maturity profiles to agentic systems in production, we would be glad to hear from you..

www.waracle.com