



The Delivery Intelligence Gap

Why organisations are accumulating delivery intelligence debt and how to close the gap with AI

By Callum Campbell and Sahar Khan

June 2026



Table of contents

| | |
|---|----|
| Foreword | 2 |
| Executive summary | 3 |
| 1. Delivery intelligence debt | 4 |
| 2. The three stages of AI maturity..... | 7 |
| Three stages of AI maturity in enterprise delivery..... | 8 |
| 2a. Practitioner lens – which stage are you actually in?..... | 9 |
| 3. Stage 2 is the missing layer | 11 |
| 3a. Practitioner lens – what the transition feels like inside the team..... | 11 |
| 4. Metadata is the scaling layer | 13 |
| 5. Why software engineering is pulling ahead..... | 15 |
| 6. What leading financial institutions reveal..... | 17 |
| 7. Limits, counter-arguments and regression risks | 19 |
| 7a. Practitioner lens – when Stage 2 collapses back to Stage 1 | 19 |
| 8. A 24-month path from personal AI to delivery intelligence | 21 |
| A 24-month path from personal AI to delivery intelligence..... | 21 |
| Conclusion: the leadership choice | 23 |
| Appendix..... | 24 |
| AI definitions..... | 24 |
| Endnotes | 24 |
| SmartCo | 26 |
| About the authors | 27 |

Foreword

AI is creating real momentum in programme delivery and many organisations are already seeing the benefits. Across delivery teams, it is reducing manual work and enabling teams to, communicate more clearly.

At the same time, as adoption grows, a more structural challenge is beginning to emerge. Alongside these gains, there is an increasing need to ensure that the work being accelerated today also strengthens how the organisation learns, improves and makes decisions over time. In many cases, delivery activity is moving ahead, but the underlying evidence is not yet being captured in a way that is comparable or reusable. Over time, this can create a disconnect between what organisations are delivering and what they can fully explain and improve.

In this paper, we introduce the concept of **Delivery Intelligence Debt**, the gap between rising AI activity and an organisation's ability to turn that activity into usable, governed insight. Like any form of debt, it does not appear immediately. It builds gradually, often alongside positive progress, and becomes visible when organisations look to answer broader questions around performance, risk, and outcomes across their portfolios.

This matters because programme delivery has always been as much about insight as execution. In complex, regulated environments, leaders rely on structured, trusted signals to make decisions, intervene early and manage risk with confidence. Without that foundation, there is a risk that AI enhances activity without fully strengthening organisational intelligence.

Callum Campbell, Head of Clients and Growth at SmartCo

Sahar Khan, Transformation and AI expert in financial services

Executive summary

Most large organisations are mistaking AI adoption for AI transformation in programme delivery. Copilots, note-takers and ad hoc analysis are producing real time savings, but the outputs sit in email threads, decks and meeting transcripts. The enterprise gets faster typing without building reusable delivery intelligence. We call this gap Delivery Intelligence Debt: the compounding cost of accelerating AI use without the data structure, metadata and governance that turn it into institutional learning.

Programme delivery sits on a three-stage maturity curve. Stage 1 is AI-assisted delivery – useful but largely personal. Stage 2 is structured delivery automation, where AI outputs flow into governed workflows with shared taxonomies, metadata and ownership. Stage 3 is enterprise delivery intelligence, where accumulated history forecasts schedule slippage, dependency risk and benefits drift. Without Stage 2, Stage 3 is unreachable.

Four observations frame the argument. AI use in UK project management has nearly doubled in two years to 70%. Gartner finds that 63% of organisations lack the data-management practices AI needs. Deloitte reports that 68% have moved 30% or fewer of their generative-AI experiments fully into production. And research on the "jagged frontier" of AI capability shows that AI used outside well-understood task boundaries can degrade performance by 19 percentage points.

This paper is for **Heads of PMO, programme directors, transformation leaders and the CIOs they report to**. The recommended sequence is practical: scope one delivery domain in the first 90 days, build the data and metadata substrate in months 3–9, and only then attempt predictive intelligence in months 9–24.

The argument in brief

AI becomes transformational in programme delivery only when it turns delivery activity into structured, attributable and reusable evidence. Anything less creates faster documents, not better delivery intelligence.

70%

of UK project professionals say their organisation now uses AI – but only 27% say it is fully embedded into workflows

63%

of organisations lack the data-management practices AI projects require, according to Gartner

1. Delivery intelligence debt

Delivery Intelligence Debt is the gap between an organisation's AI activity and its ability to learn from that activity. It accumulates whenever AI is used to go faster on a local task without leaving behind structured, attributable, reusable evidence: tagged RAID items, decision records with provenance, dependency classifications, forecast deltas, benefits assumptions. The work happens; the institutional memory does not.

The debt is invisible at first because Stage 1 productivity gains feel like progress. Status reports get drafted faster. Meeting summaries appear automatically. None of this shows up as a problem on a single programme. It shows up two years later, when leadership asks why – despite all the AI investment – the enterprise still

cannot reliably forecast slippage, compare like for like across portfolios, or detect leading indicators of trouble.

Three things make the debt expensive. First, free-text outputs cannot be retrofitted with meaning at scale; semantic layers help but cannot recover decision context that was never captured. Second, fragmented adoption can actively degrade outcomes when AI is used outside its competence boundary, while making the output look more polished. Third, the longer a firm operates in Stage 1, the harder the cultural shift to Stage 2 becomes, because individual productivity has been habituated as the success metric.

DORA's 2025 research helps explain the mechanism. It frames AI adoption as a J-curve – an initial productivity dip, or "tuition cost", before longer-term returns materialise – and treats AI as an amplifier that magnifies the strengths of high-performing organisations and the weaknesses of struggling ones. The dip in the maturity curve is not a generic adoption tax. It is the result of AI being layered onto an operating model that is not yet ready to compound its outputs. The same lens reframes shadow AI: not rogue users, but a change-saturation symptom of leadership failing to sequence AI adoption coherently.

Where the J-curve dip actually falls is the most important leadership question in the maturity journey. Without a structured approach, Stage 2 is when productivity declines – Stage 1 individual gains stop compounding, informal workarounds proliferate, and the absence of a governed data layer becomes a drag on the work. With a structured approach, Stage 2 is instead a deliberate phased build: productivity holds steady on Stage 1 gains while the data, taxonomy and governance substrate is laid in parallel, then steps up sharply when Stage 3 capability comes online. The gap between the two paths – the value the unstructured organisation never captures – is Delivery Intelligence Debt.

Value gained over the course of AI adoption in Enterprise Delivery

A phased journey that compounds value – from individual productivity to enterprise-wide delivery intelligence.

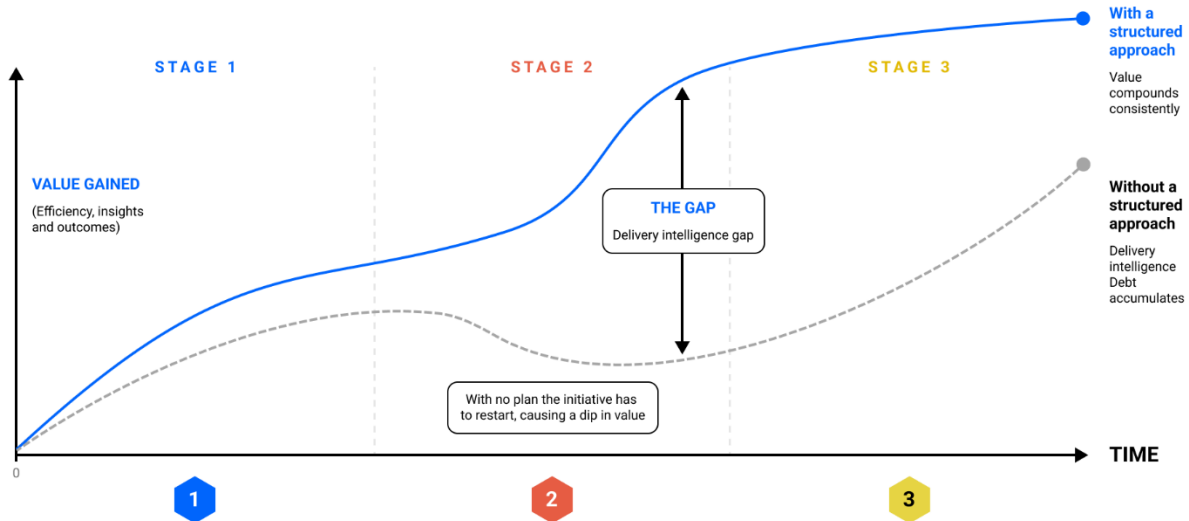


Figure 1. Value gained over the course of AI adoption in enterprise delivery. The gap between the structured and unstructured paths is Delivery Intelligence Debt accumulating over time.

| | | |
|--|---|---|
| <p>STAGE 1 · PERSONAL PRODUCTIVITY</p> <p>Individuals use AI tools to support day-to-day delivery tasks. Gains are real but largely personal – not yet captured as reusable institutional evidence.</p> <p><i>"Faster typing, no learning."</i></p> | <p>STAGE 2 · DELIVERY AUTOMATION</p> <p>AI outputs flow into governed workflows with shared taxonomy, metadata and ownership. The bridge most organisations underbuild – and the source of the dip.</p> <p><i>"The missing layer."</i></p> | <p>STAGE 3 · DELIVERY INTELLIGENCE</p> <p>Structured history drives predictive visibility across the portfolio. Without the Stage 2 foundation, benefits cannot be realised or sustained at scale.</p> <p><i>"Learning, not just doing."</i></p> |
|--|---|---|

2. The three stages of AI maturity

The maturity journey is best understood as three stages. They are not technology categories; they describe how work, data and governance change as AI moves from the individual desktop into the delivery operating model.

Stage 1 – AI-assisted delivery. Personal productivity. Meeting notes, summary emails, draft status reports, document search, action capture. Microsoft's early-user research found users 29% faster across search, writing and summarisation. A Forrester study reported time savings of up to 34% on content creation. Microsoft's case study with Commercial Bank of Dubai reports an IT PMO generating 3,200 AI meeting summaries in a single month. The defining property: outputs are personal – they help the user but do not accumulate into a shared, comparable record. APM's 2026 research makes the gap visible: 70% of project professionals say their organisation now uses AI, but only 27% say it is fully embedded into workflows.

Stage 2 – Structured delivery automation. AI output stops being a personal aid and becomes part of the controlled delivery system. The bankable value is not faster drafting but the automation of repeatable processes: extracting RAID items into canonical logs, generating actions with owners and dates, creating decision records with provenance, classifying dependencies, routing exceptions through the right controls. The hard work is metadata discipline – standard fields, consistent definitions, review steps, versioning. Morgan Stanley's AI Assistant reached 98% adoption across Financial Advisor teams, and its Debrief capability sits explicitly between colleagues and systems of record. That is not a copilot; it is a workflow layer.

Stage 3 – Enterprise delivery intelligence. Most organisations talk about Stage 3; few can credibly reach it yet. Here, the organisation uses historical delivery data to predict slippage, identify risk cascades, detect benefits drift and recommend interventions before issues materialise. Stage 3 is impossible without Stage 2. Microsoft's semantic-layer guidance is explicit that AI systems consume both data

and metadata, and that organisations need gold layers, AI-friendly naming, documentation and an enterprise ontology.

Three stages of AI maturity in enterprise delivery

Focus, capabilities, data posture and key evidence at each stage.

| STAGE 01 AI-assisted delivery | STAGE 02 Structured automation | STAGE 03 Delivery intelligence |
|--|---|--|
| FOCUS Personal productivity Outputs help the user but do not accumulate. | FOCUS Workflow integration AI output enters the controlled delivery system. | FOCUS Predictive insight Anticipates slippage, risk cascades, benefits drift. |
| CAPABILITIES <ul style="list-style-type: none"> • Meeting summaries • Draft status reports • Document search • Action capture | CAPABILITIES <ul style="list-style-type: none"> • Tagged RAID logs • Decision records w/ provenance • Routed exceptions • Governed metadata fields | CAPABILITIES <ul style="list-style-type: none"> • Schedule slippage models • Dependency risk cascades • Benefits drift detection • Recommendation engines |
| DATA & METADATA Free-text outputs. No standard taxonomy or ownership. | DATA & METADATA Standard fields, definitions, named ownership, human review gate. | DATA & METADATA Semantic layer over historical delivery data. Lineage and model monitoring documented. |
| EVIDENCE Microsoft: 29% faster on search/writing tasks. APM: 70% adoption, only 27% fully embedded. | EVIDENCE Morgan Stanley AI Assistant: 98% adoption. Debrief sits between calls and systems of record. | EVIDENCE Emerging – strongest in code delivery and customer ops. Delivery-specific evidence still developing in 2026. |
| TIME: IMMEDIATE | TIME: 3 – 9 MONTHS | TIME: 12+ MONTHS |

Figure 2. The three stages of AI maturity in enterprise delivery – focus, capabilities, data and metadata posture, and key evidence at each stage.

What this looks like in practice. Stage 3 is not a generic dashboard. It is a portfolio view that draws from the structured taxonomy built in Stage 2: risk scores derived

from tagged RAID items, milestone status from governed records, forecast variance from lineage-tracked spend data, and AI-generated recommendations whose inputs can be inspected. Without Stage 2, none of this view can be built – the underlying data is too fragmented to support it.

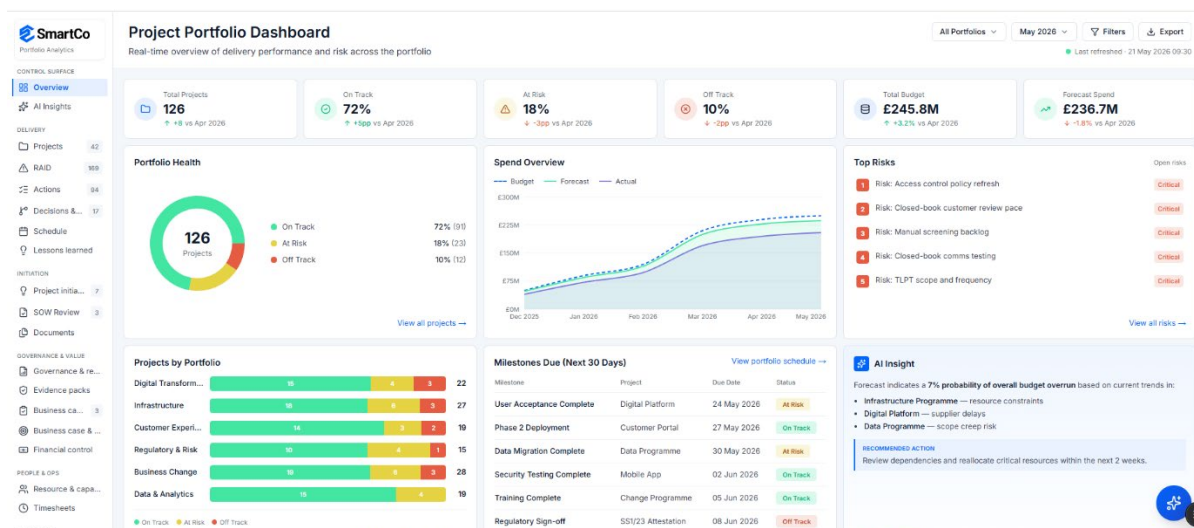


Figure 2b. Illustrative Stage 3 portfolio view – predictive risk, governed lineage, AI-generated recommendations. Built on the Stage 2 data substrate.

2a. Practitioner lens – which stage are you actually in?

By Sahar Khan, Transformation and AI expert in financial services

The three-stage model is most useful as a mirror, not a map. Most delivery leaders will instinctively place themselves in Stage 2. In practice, the majority are still in Stage 1 with Stage 2 ambitions. The questions below are drawn from PMO assessment work across UK public sector and financial services programmes. Each has a yes/no answer that does not depend on interpretation.

| Stage 1 indicators | Stage 2 indicators | Stage 3 indicators |
|--|---|---|
| <ul style="list-style-type: none"> • AI tools used by some team members. • Meeting summaries or draft status reports in regular use. • Outputs reviewed by the individual who generated them. • No standard storage, tagging or ownership model. • Core terms (risk, milestone, dependency) not consistently defined. | <ul style="list-style-type: none"> • AI outputs flow into at least one system of record. • RAID, actions or decisions populated with named ownership. • Agreed taxonomy for at least one artefact type. • Human review before AI output becomes a governance artefact. • The accountable person can be named if the output is wrong. | <ul style="list-style-type: none"> • Six to twelve months of structured delivery history exists. • A model or analytical layer uses that history. • A predictive signal has influenced a decision before the problem materialised. • Lineage and governance are documented. • Model performance is monitored, not merely demonstrated. |

Figure 3. A practical self-assessment. The point is not to label the whole organisation, but to identify the highest stage that is genuinely true for one delivery domain today.

Three Stage 1 yeses plus two Stage 1 noes (on storage and taxonomy) and you are in Stage 1. Five Stage 2 yeses and you are genuinely in Stage 2 for that domain. Most organisations cannot honestly tick the Stage 3 boxes in 2026, and that is fine. The point is to know it – and to stop investing as though Stage 3 is one tool purchase away.

3. Stage 2 is the missing layer

The most common failure in enterprise AI adoption is to jump from Stage 1 excitement to Stage 3 aspiration without doing Stage 2 design. The language of AI strategy moves quickly to agents, predictive analytics and portfolio intelligence. The delivery operating model is less glamorous: definitions, fields, owners, review steps, exceptions, audit trails – and the unglamorous work of getting one term to mean one thing across a portfolio.

That work matters because programme delivery is full of ambiguous language. A milestone can mean a contractual date, an internal planning point or a sponsor-facing narrative marker. A dependency can mean a technical sequence, a commercial constraint or a resource contention. A risk can be a true uncertainty, a disguised issue or a politically safer way to describe a decision that has not been made.

AI does not remove that ambiguity. It amplifies whatever structure is already present. In a strong delivery environment, AI helps extract, classify, summarise and route work. In a weak environment, AI produces cleaner narrative around inconsistent definitions: the output looks more polished, but the underlying signal stays weak.

Stage 2 therefore has two jobs. The first is operational: reduce manual effort in repeatable delivery processes. The second is strategic: create a governed evidence base for future intelligence. The second job is the one most organisations underinvest in.

3a. Practitioner lens – what the transition feels like inside the team

By Sahar Khan, Transformation and AI expert in financial services

The maturity journey is correctly described as an operating-model problem, but that framing omits the experience of the people living through it. The transition from

Stage 1 to Stage 2 is not primarily a data architecture challenge for most delivery teams. It is a trust and identity challenge.

The experienced PM problem. Senior PMs often resist Stage 2 not because they distrust AI, but because structured tagging feels like a demotion of their judgement. A PM who has spent fifteen years reading a room and sensing a dependency risk before it appears in any log rightly asks what a taxonomy does for them. The answer is that Stage 2 captures the output of their judgement in a reusable form – it does not replace the judgement itself. That distinction has to be made explicitly. Implementations that skip it will see their most experienced people produce the least structured outputs, which is the worst possible outcome.

The job security question. Delivery teams ask about this, directly or indirectly, in almost every AI rollout. Stage 2 automation does reduce time on administrative tasks – status drafting, RAID population, action capture – and that will eventually affect resourcing. What it does not do, when implemented well, is remove the need for delivery judgement, stakeholder navigation, risk escalation and the capacity to hold a programme together under pressure. The PMO role shifts from producing governance artefacts to owning the quality of the data those artefacts are built from. That is a more skilled role, not a lesser one – but the shift needs to be named, and teams need to see a credible path to it.

The uneven adoption pattern. Junior analysts adopt fastest, because tools are immediately useful for first drafts. Mid-level PMs adopt selectively, integrating what fits their workflow. Senior PMs and programme directors adopt last and least consistently, often relying on their teams to surface AI-assisted outputs without generating them. This is not a training problem – it is a relevance and trust problem. If the taxonomy asks senior PMs to do something they find meaningless, the design needs to change, not the people.

What good change management looks like. Three things work in practice. Design with sceptics – not to be managed, but to make the design better; the experienced PM who resists tagging will identify the genuinely ambiguous fields faster than any

consultant. Measure quality and completeness of structured outputs, not just adoption rates; early completeness problems are design problems, not culture problems. And protect the review step: under delivery pressure, human review is the first thing to be dropped, and the whole Stage 2 value proposition depends on it holding.

The takeaway

If the best delivery people in the organisation produce the least structured data, Stage 2 has been designed badly. The fix is in the design, not in the people.

4. Metadata is the scaling layer

The hardest lesson in enterprise AI is that models are not the binding constraint for long. Data design is. Gartner is more direct: most organisations lack the data-management practices AI requires, and unsupported projects will be abandoned. In programme delivery, that means the design work begins before the first automation is built. If the organisation has not agreed how to identify a milestone, classify a risk, record a dependency, define a control failure or capture a decision, then the predictive layer has no stable foundation to stand on.

Metadata is what turns narrative into comparable signals. It tells an AI system what a field means, where it came from, when it changed, who owns it and how much to trust it. The UK government's AI-ready data guidance, the Open Data Institute's framework and NIST's generative-AI profile converge on the same conclusion: data must be self-describing, governed and contextually complete. Translated into programme delivery, a serious Stage 2 design should at least standardise the following fields.

| Category | Minimum viable delivery fields |
|-----------------|---|
| Identity | Portfolio ID · programme ID · workstream ID · stage gate · version |
| Classification | Milestone type · dependency type and criticality · RAID category · issue cause and resolution code · decision type · benefit category |
| Ownership | Action owner · data steward · accountable executive |
| Temporal | Created date · due date · timestamp · update frequency · reporting period |
| Quality & trust | Forecast confidence · source system · provenance · synthetic vs manual flag · review status |
| Governance | Policy tags · regulatory tags · access role · retention status · approval status |

Figure 4. Minimum viable metadata for AI-enabled programme delivery.

These fields are not bureaucracy. They are the minimum viable semantics that allow automation to be auditable and analytics to become predictive. Retrofitting meaning into a year of free-text updates is expensive and unreliable. If the goal is predictive analytics in year two, structured tagged delivery data must start in year one. For one PMO domain, a useful foundation can usually be built in one or two quarters with tight scope. Enterprise-wide governance is slower – McKinsey recommends domain-first sequencing and warns against disconnected pilots; vendor rollouts commonly run 12–18 months.

What this means for a delivery leader

Can you point to a single agreed definition of "milestone", "risk" and "dependency" across your portfolio – and a single owner for each? If not, that is where Stage 2 begins. You are not commissioning a data programme; you are commissioning the discipline of having one set of words mean one set of things.

5. Why software engineering is pulling ahead

Software engineering moves faster than programme delivery because it already has an operating system for AI to land in. Code repositories, pull-request reviews, build pipelines, security scans, code ownership and audit logs exist as standard machinery. AI plugs into a system that already knows who changed what, when, and with whose approval. Programme delivery often lacks that substrate – most of what happens still lives in documents and email.

GitHub's well-architected guidance for Copilot is built around licensing, access control, model restrictions, human review gates, audit streaming and spend controls. It explicitly warns against enabling AI agents without audit logging. DORA's AI Capabilities Model makes the substrate explicit: a clear AI stance, healthy data ecosystems, AI-accessible internal data, version control, small-batch work, user-centric focus, and quality internal platforms. Each has an obvious delivery-side analogue, and the contrast is not flattering.

| Operating-system element | Software engineering | Programme delivery |
|--------------------------|---|---|
| Change record | Code commits and pull requests | Status decks, emails, free-text logs |
| Review model | Reviewer approval before merge | Steering pack review after the narrative is formed |
| Ownership | Code owners, repository permissions | Often implicit across PM, PMO and sponsor |
| Audit trail | Pipeline logs, branch history, security scans | Weak lineage from source data to reported position |
| Reusable structure | Schemas, tests, release gates | Narrative artefacts that are hard to compare |
| AI consequence | AI lands inside a controlled system | AI speeds up documents without creating reusable intelligence |

Figure 5. Software engineering already has a substrate for AI; programme delivery often has a thinner one.

Financial-services engineering shows the pattern. J.P. Morgan integrates AI into code assistance, peer code review and internal LLM workflows, with reported efficiency improvements of 10–20%. JPMorgan’s LLM Suite went from zero to 200,000 users in eight months and is being extended into internal data access and workflow agents. Programme delivery, by contrast, still sits much closer to the horizontal-copilot model: powerful tools, thin control system. The point is not to copy engineering —

delivery work is not code – but to recognise that without an equivalent operating system, AI in delivery will keep producing personal productivity without compounding into intelligence.

What this means for a delivery leader

Ask one question of your AI rollout: who would I call if it produced a wrong answer that ended up in a steering pack? If the answer is unclear – no named owner, no log of inputs used, no review trail – the rollout is still in the copilot stage. Engineering teams have an answer to this for every line of code AI suggests.

6. What leading financial institutions reveal

The strongest public examples in financial services do not look like generic AI deployment. They look like operating-model redesign. The direction of travel is not simply more prompts – it is secure platforms, workflow integration and controls that connect AI output into existing systems.

Morgan Stanley rolled out AI Assistant in September 2023 and launched Debrief in 2024. Assistant reached 98% adoption across Financial Advisor teams; Debrief is positioned not as a note-taker but as a layer sitting between colleagues and execution, CRM, reporting and risk tools. The important feature is the movement from conversation to reviewed workflow record.

Lloyds Banking Group is more explicit about the operating model: a dedicated AI Centre of Excellence, a new AI operating model, a cross-functional GenAI Control Tower, structured upskilling and a Data & AI Summer School with over 250 sessions.

By October 2025, Lloyds reported nearly 30,000 Copilot licences rolled out with 93% active use. This is centrally governed scaling, not random bottom-up adoption.

JPMorgan Chase provides a third pattern. LLM Suite reached 200,000 users in eight months and is being extended into internal-data integration and workflow agents. Combined with domain-specific developer tools, this is exactly the secure-platform-plus-workflow combination programme delivery needs to move beyond Stage 1.

The regulatory lens reinforces the pattern. The 2024 Bank of England / FCA survey found 75% of UK financial firms using AI, but flagged concentration risk: one third of use cases are third-party implementations, the top three providers account for 44% of model providers cited, and 46% of firms admit only partial understanding of the AI they use. Any delivery-intelligence strategy must be designed with governance, provenance and portability in mind from the outset.

98%

adoption of Morgan Stanley's AI Assistant across Financial Advisor teams — embedded into workflow, not bolted on

46%

of UK financial firms admit only partial understanding of the AI they use (Bank of England / FCA, 2024)

The pattern to notice

The strongest examples are not 'AI everywhere' stories. They are 'AI inside a governed workflow' stories. That is the distinction programme delivery should copy.

7. Limits, counter-arguments and regression risks

A credible position should be able to state its limits. Stage 3 delivery intelligence is not yet a mature, widely proven capability across the market. Public evidence is strongest in knowledge work, software engineering and customer operations. Delivery-specific evidence is still emerging. That does not weaken the argument; it clarifies it: most organisations should be more honest about how early they are.

Is Stage 3 actually achievable? Honestly, both achievable and aspirational. Most cited examples sit between Stage 1 and Stage 2. The argument is not that Stage 3 is imminent – it is that without Stage 2 foundations, Stage 3 is structurally unreachable, and most current investment patterns are not building those foundations.

Does structured taxonomy slow delivery down? It can, if introduced badly. Heavy upfront ontology programmes that try to standardise everything before producing anything tend to fail. The recommended sequence – one domain, tight scope, parallel data workstream, iterate – is designed to avoid that. The question is not whether to add structure; it is whether structure is added by design or extracted painfully after the fact.

Is AI being oversold? Should firms wait? Waiting is not a neutral choice. Shadow-AI evidence is that employees bring their own tools anyway; deferring the operating-model question just guarantees adoption without governance. Some current use cases are genuinely productive, some are theatre, and the discipline of distinguishing them is itself a Stage 2 capability.

7a. Practitioner lens – when Stage 2 collapses back to Stage 1

By Sahar Khan, Transformation and AI expert in financial services

The maturity model is presented as a forward progression. In practice, programmes regress. Stage 2 gains are more fragile than they look, and the conditions that collapse them are predictable. Naming them is the basis for designing against them.

Sponsor or leadership change. The most common cause of regression. A new SRO arrives with their own reporting preferences and impatience with governance that feels bureaucratic before its value is visible. Within two or three reporting cycles, the taxonomy bends, tagging becomes inconsistent, and the data layer degrades. The mitigation: make the Stage 2 design legible to incoming leaders from day one – not as rules, but as a system already producing value they can point to. If the first thing a new sponsor sees is compliance overhead, they will rationalise it away. If the first thing they see is a portfolio risk report that caught something, they will protect it.

Tooling mandates from the centre. Organisations frequently impose enterprise tool changes – a new PPM platform, a consolidated Copilot licence, a mandatory data-warehouse migration – without accounting for the Stage 2 infrastructure that sat on the previous tool. Taxonomies, metadata fields and workflow integrations do not migrate automatically. Six months of governed RAID work can vanish in a quarter. The mitigation: treat the Stage 2 substrate as a programme asset, documented separately from any specific tool, with an explicit migration plan that is part of any tooling change request.

Delivery pressure. When a programme hits a major release, regulatory deadline or go-live, Stage 2 discipline is the first casualty. Teams revert to whatever produces output fastest – unstructured AI use, manual workarounds, governance artefacts produced outside the system of record. This is understandable, but it is also the moment when high-quality delivery data would be most valuable for the postmortem. The mitigation: design the minimum viable compliance path – the three or four fields that matter most – to require less effort than the workaround, not more. If the structured path is slower under pressure, teams will abandon it under pressure.

Three conditions for Stage 2 resilience

1. A named data owner senior enough to defend the design when it is challenged.
2. At least one leadership-valued output that depends on the structured data — so that regression becomes visibly costly.
3. A design documented as an organisational asset, not held in one person's head or one team's shared drive.

Stage 2 is fragile until it becomes infrastructure. The transition from fragile to infrastructure is the hardest and least-discussed part of the maturity journey.

8. A 24-month path from personal AI to delivery intelligence

The next 12 to 24 months should not be framed as a choice between buying copilots and building agents. The more important choice is whether the organisation stays in Stage 1 or deliberately builds the bridge to Stages 2 and 3. The path below is the sequence a serious delivery leader can use to avoid accumulating Delivery Intelligence Debt while still capturing near-term AI value.

A 24-month path from personal AI to delivery intelligence

Three phases, each with a single sequencing principle.

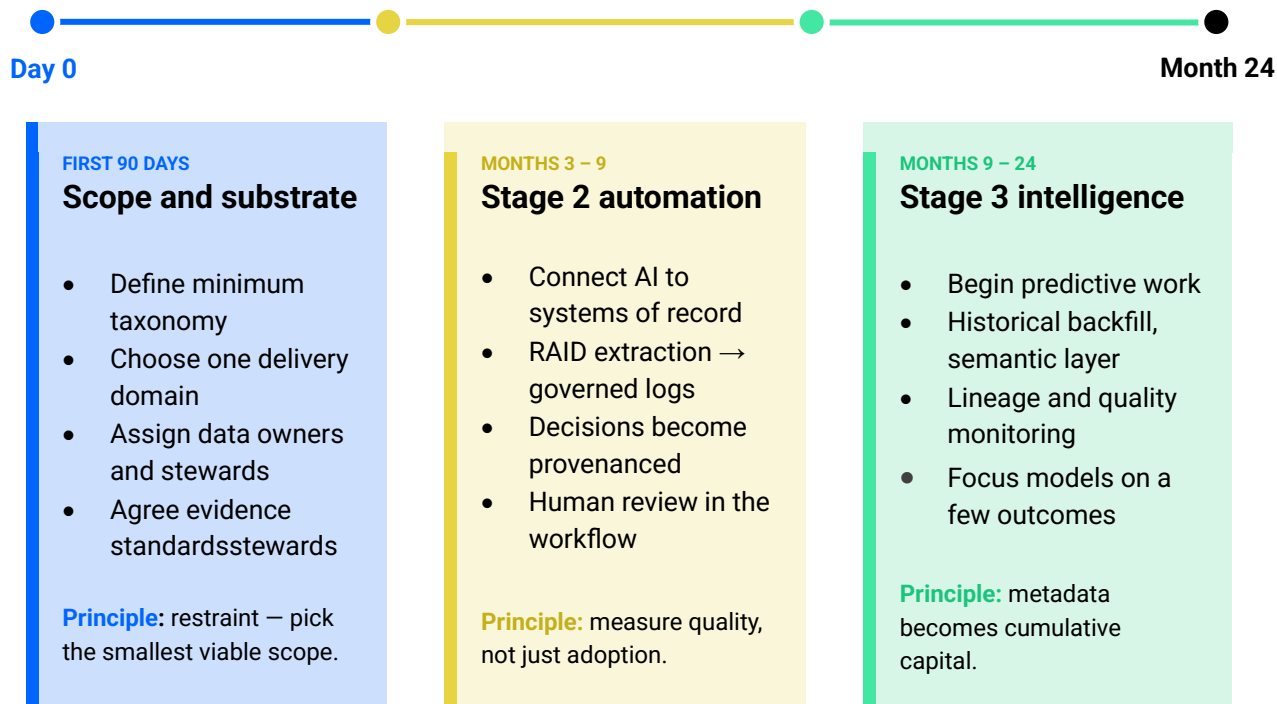


Figure 6. A practical 24-month sequence from personal AI to delivery intelligence.

Two principles run across the whole journey. The first is governance. Regulatory and assurance design is not a final-stage concern: NIST’s AI Risk Management Framework, the EU AI Act and the Bank of England / FCA evidence base all point the same way. Every meaningful automation should have a visible owner, logged provenance, an approved review path and a documented failure mode.

The second is restraint. The most important decision in the first 90 days is what not to do. Do not standardise the whole enterprise. Do not build a predictive model. Do not create a taxonomy so complex that teams avoid it. Build the smallest structure that makes one high-value process auditable and reusable. In months 3–9, measure quality as much as adoption – adoption tells you people use the tool; completeness, consistency and correction rates tell you whether the output is becoming trustworthy delivery evidence. In months 9–24, the most important governance question is not whether the model is impressive, but whether leaders understand what evidence it uses, where its boundaries are and who is accountable when its signal influences a decision.

A non-sales call to action

The right first move is not to buy another tool. It is to pick one delivery domain and test whether AI-assisted work can become structured, governed evidence. If it can, the organisation has started to build delivery intelligence. If it cannot, the organisation is still in Stage 1, regardless of how many AI licences it owns.

Conclusion: the leadership choice

AI will become normal in programme delivery. That is no longer the strategic question. The question is whether it becomes a layer of personal productivity on top of fragmented delivery practice, or a route to a more intelligent delivery operating model.

The organisations that benefit most will not be the ones with the most ambitious agent roadmap. They will be the ones that make delivery work more observable, comparable and reusable. They will define the terms that matter, create metadata as work happens, connect AI outputs into systems of record, preserve human review, and treat delivery data as an asset rather than exhaust.

Delivery Intelligence Debt is avoidable, but only if leaders recognise it early. Stage 1 is useful. Stage 3 is attractive. Stage 2 is the bridge. The leadership choice for the next two years is whether to build that bridge deliberately, or to discover later that the organisation accelerated thousands of delivery tasks without learning enough from any of them.

That is the difference between an organisation that uses AI and one that learns with it.

Appendix

AI definitions

Artificial Intelligence (AI). Technology that enables computer systems to perform tasks typically requiring human intelligence – such as recognising patterns, analysing data, solving problems, and making decisions.

Predictive AI. The use of artificial intelligence to analyse historical and current data and make predictions about future events or trends.

Generative AI. The use of artificial intelligence to generate new, realistic content by learning and replicating patterns.

Agentic AI. A model of artificial intelligence execution with autonomous agents that can observe, reason, plan, use tools, and act, coordinating across workflows, tools, and systems with minimal human input.

Endnotes

1. Association for Project Management, "AI use in project management nearly doubles in just two years", 9 September 2025.
2. Gartner, "Lack of AI-ready data puts AI projects at risk", February 2025.
3. Deloitte AI Institute, "State of Generative AI in the Enterprise", 2024.
4. Dell'Acqua et al., "Navigating the Jagged Technological Frontier", 2023; MIT Sloan summary.
5. Microsoft WorkLab, "What can Copilot's earliest users teach us about generative AI at work?", November 2023.
6. Forrester Consulting, "The Total Economic Impact of Microsoft 365 Copilot", commissioned by Microsoft.

7. Microsoft customer story, "Commercial Bank of Dubai boosts AI literacy with Microsoft 365 Copilot", 2025.
8. UK Government, "Guidelines and best practices for making government datasets ready for AI".
9. Open Data Institute, "A framework for AI-ready data", 2025.
10. NIST, "AI Risk Management Framework: Generative AI Profile", NIST AI 600-1, July 2024.
11. Microsoft Learn, "Prepare the Semantic Layer for AI in Microsoft Fabric".
12. GitHub Docs, "Managing Copilot for your enterprise".
13. DORA / Google Cloud, "2025 DORA AI Capabilities Model".
14. Morgan Stanley, "Launch of AI @ Morgan Stanley Debrief", 26 June 2024.
15. Lloyds Banking Group, "Lloyds scales adoption of Copilot", 21 October 2025.
16. JPMorgan Chase, "LLM Suite named 2025 Innovation of the Year by American Banker", June 2025.
17. J.P. Morgan Payments, "How J.P. Morgan developers leverage AI", April 2025.
18. Bank of England and Financial Conduct Authority, "Artificial Intelligence in UK financial services 2024", November 2024

SmartCo

SmartCo is a digital transformation partner, enabling global organisations to scale AI and data, re-risk M&A, optimise cloud and ERP estates and deliver complex technology programmes with confidence. Combining advisory, hands-on delivery and continuous optimisation, SmartCo brings senior expertise and rapid mobilisation to ensure your technology programme delivers measurable outcomes and long-term value.

www.smart-co.co.uk

For further contact

If you would like to discuss this white paper, please contact the authors.

Contact the authors



Callum Campbell,
SmartCo, Head of Clients and Growth

callum.campbell@smart-co.co.uk [linkedin.com/in/callum-campbell-consulting](https://www.linkedin.com/in/callum-campbell-consulting)



Sahar Khan,
Transformation and AI expert in financial services

sahar_khan@live.co.uk <https://www.linkedin.com/in/sahar-k-88a0091>

About the authors

Callum Campbell leads on clients and growth at SmartCo, working closely with organisations to shape and deliver transformation outcomes. Alongside this, he plays a key role in driving SmartCo's AI agenda, specialising in how AI can be applied practically to improve delivery and performance.

Sahar Khan has been a Transformation and AI expert for several global financial groups. Her practitioner perspective in this paper draws on big scale change assessment and delivery work across global public-sector and financial-services programmes.