



POINT OF VIEW • JULY 2026

From Review to Reality

What the Mills Review and the AI Adoption Plan mean for financial services, and what firms should do now

Alan Paterson • Founder and Chief Innovation Officer

Executive summary

July 2026 will be remembered as the month UK financial services regulation stopped circling AI and committed to a direction.

Within a matter of weeks, two important publications set out how the UK intends to govern and accelerate the use of the technology now reshaping the sector. The Mills Review, commissioned by the FCA Board and led by Sheldon Mills, examines how AI could transform retail financial services by 2030 and sets out seven priority recommendations.¹ HM Treasury's Financial Services AI Adoption Plan provides ten recommendations aimed at accelerating safe adoption across the sector.²

Both publications should be welcomed. Together, they provide the clearest strategic direction the industry has yet received: retain the UK's principles-based and outcomes-focused regulatory model, clarify how existing obligations apply to AI, prepare for agentic systems, strengthen skills and supervisory capability, and develop a credible market for independent assurance. The decision not to create a standalone AI rulebook is significant. It recognises that a rapidly changing technology is better governed through adaptable outcomes and evidence than through a parallel regime likely to become outdated quickly.

Read together, the two publications also answer a question firms have been asking for several years. The direction is not to wait for a new set of AI-specific rules, but to demonstrate how existing obligations (including the Consumer Duty, SM&CR, model risk management and operational resilience) are being met as AI takes on more of the work.

The regulatory question has therefore shifted from
“What rule applies?” to **“What can you evidence?”**

In this respect AI has not so much changed the FCA's philosophy as brought it into sharper focus. For two decades regulation has moved steadily from prescribing behaviour towards requiring firms to demonstrate good outcomes, through Treating Customers Fairly, the Senior Managers Regime and

the Consumer Duty. Culture and intent were previously accepted as sufficient proxies. Autonomous systems have neither. The shift from *"What rule applies?"* to *"What can you evidence?"* is therefore the natural next step in a long regulatory journey.

That is an important and constructive shift. It moves the industry beyond abstract principles and towards the practical foundations required for trusted adoption: clear accountability, robust evaluation, independent challenge, production controls, monitoring and evidence that can withstand Board and regulatory scrutiny.

It also reflects what we heard from senior leaders at our AI Tipping Point briefing in June to examine the arrival of agentic AI in financial services, held before either publication was released.³ The consensus in the room was that technical capability has moved ahead of the governance and assurance needed to deploy it confidently. The constraint is no longer whether the technology can work, but whether firms can demonstrate that it is safe, controlled and accountable at production scale. This is consistent with recommendations we made in our submissions to the FCA's AI Input Zone⁴ and in our response to IOSCO's AI Supervisory Toolkit,^{5 6} several of which are now reflected in the direction both publications take.

"The governance challenge is no longer defining principles. It is operationalising them."

PLENITUDE FEEDBACK TO IOSCO, AI SUPERVISORY TOOLKIT, 2026

This paper sets out what the two publications say, the five strategic themes that emerge when they are read together, and our assessment of what firms should do next.

Taken together, the Government, the FCA and the authors of both reviews have established a clear direction and a strong foundation for AI adoption in UK financial services. An outcomes-based framework, a review of the regulatory perimeter, a proposed third-party assurance scheme, preparation for agentic finance and a serious focus on skills and supervisory capability are the right foundations.

The next challenge is execution.

The operational layer is not yet fully defined: how evaluation evidence will be graded; what production readiness looks like beyond a successful pilot; how systems should be re-certified when models, data or orchestration components change; and how senior managers should sign off autonomous decisions under SM&CR. These are not criticisms of the direction taken. They are the natural next questions created by it, and they are the questions currently holding deployment back.

The final section sets out six practical actions institutions can take now. None requires new regulation. All are no-regret moves under any likely form of future guidance, and all will help firms build the evidence, accountability and assurance the new direction of travel will require.

The institutions that act now will be best placed not only to satisfy supervisors, but to turn trusted AI deployment into a source of competitive advantage.

Why this matters now

The two publications arrive at a moment when the gap between what AI can do and what firms can confidently deploy has become the defining constraint on adoption.

75%

of firms responding to the Bank of England / FCA 2024 survey reported using AI⁷

1 in 5

UK adults are already open to AI making financial decisions for them (Mills Review consumer research)¹

26%

trust general-purpose tools such as ChatGPT, Claude or Gemini for financial advice, outside the FCA-regulated advice framework¹

9%

of UK adults are reached by regulated financial advice: the gap unregulated AI tools are filling¹

Adoption itself is no longer the story. Around 75% of firms responding to the Bank of England and FCA's 2024 survey reported using AI,⁷ and a 2026 global survey cited by the Mills Review found 81% of firms adopting AI at some level, with 40% at more advanced stages of scaling or transformation.⁸ Agentic AI, barely present in surveys two years ago, is now being piloted or deployed by more than half of industry respondents. Sheldon Mills notes in his foreword that more than 20 frontier models were released during the course of his Review alone.¹ The technology moved faster than the review commissioned to study it.

What has changed is the nature of the systems being deployed. The Mills Review frames this as an autonomy spectrum: five stages in which the human role shifts from operator, using AI as a tool, through collaborator, consultant and approver, to observer, monitoring systems that act continuously within agreed limits. At our June briefing the same trajectory was described in operational terms: from one human supported by one AI assistant, to one human overseeing ten AI-driven processes, to one human overseeing a thousand or more. Every step along that path changes what oversight means and what evidence a firm needs to hold.

Consumers are moving along the same path, with or without the industry. The Review's consumer research found that one in five UK adults is already open to AI making financial decisions for them, and around 26% trust general-purpose tools for financial advice, mostly unaware that FCA advice protections, suitability obligations and formal routes to redress will not apply.¹

Regulators, meanwhile, have been gathering evidence at pace. The FCA's AI Input Zone closed in June; IOSCO published its final Supervisory Toolkit for AI Use in Capital Markets in May;⁶ the FCA's good and poor practice publication is expected later this year. Plenitude has engaged at each stage: a joint response with our partner ExoBrain to the FCA's AI Live Testing proposals in 2025, in which we first argued that trust in AI must be engineered in rather than retrofitted;⁹ our submission to the AI Input Zone in June; and formal feedback to IOSCO on the Toolkit. Each drew on our advisory work, our own governed use of AI in regulated delivery, and our AI assurance practice. This paper draws on that evidence base, on the two July publications, and on the discussion among senior leaders from banking and investment management at our AI Tipping Point briefing on 9 June.

The window for shaping how this framework gets operationalised is open now. The good and poor practice publication arrives later this year, the perimeter review is due to begin within months, and the assurance scheme's standards will be set by the assessments already in use. Windows like these close when the documents publish

The five strategic themes

The Mills Review runs to seven priority recommendations; the Adoption Plan to ten. Read separately, they can look like two long lists. Read together, five strategic themes emerge, and they are the themes on which firms should organise their response.

One acknowledgement before the themes. Our reading concentrates on assurance, controls and deployment, because that is where our work sits and where firms must act first. It is not the whole of the policy problem. The Mills Review is also a study of market power and consumer agency: who controls the AI-mediated customer interface, whether value migrates away from financial institutions, how personalised manipulation and opaque pricing might emerge, and whether unequal access to high-quality tools widens exclusion. Its recommendation of a trusted public-interest AI financial capability service belongs to that agenda. Assurance is necessary, but market structure, distribution and consumer agency may prove equally consequential, and boards should read the Review with both lenses.

THEME 1

Regulatory clarity without a new rulebook

The clearest signal from both publications is what they do not do. Neither proposes standalone AI-specific regulation, although the Adoption Plan expressly leaves room for targeted clarification or adaptation where AI raises novel questions. The Mills Review finds the overall framework "remains sound", noting that engagement respondents did not seek changes to the regime but wanted clarity on how to interpret it as AI use grows. The Adoption Plan opens its recommendations by asking regulators to confirm that the UK will maintain its principles-based, outcomes-focused approach, and reports that firms strongly support this over new AI-specific regimes.

The problem both documents identify is not the framework but its practical application. Firms consistently report that guidance is fragmented across regulators, that sandbox access is inconsistent, and that high-level principles have not been translated into operational expectations for AI and generative AI use cases, particularly on the Consumer Duty, model risk management, explainability and accountability. The Adoption Plan's answer is cross-regulator coordination and a possible Financial Services AI Adoption Support Hub; the Mills Review's is regular clarification through supervision, the Scale-up Unit and the AI Lab.

Extending established risk frameworks to AI is the right approach. The gap is specificity: the single most valuable contribution the regulator could now make is to set out how existing frameworks extend to AI, where they reach, where they need extending, and how to treat risk inherited from third-party model providers. Every firm is currently solving that mapping from first principles, and the duplication of effort across the market is considerable.

OUR VIEW

The commitment to the existing framework is the right call, but clarity deferred is adoption deferred. The good and poor practice publication expected later this year is now the single most important regulatory document in the pipeline, and firms should engage with it as such.

THEME 2

The moving perimeter

The most consequential recommendation in either document concerns who is inside the regulatory perimeter at all. Both publications land on the same fact pattern: consumers are taking financial guidance from general-purpose AI tools that sit entirely outside regulation, with no suitability standards, no accountability and no route to redress. The Mills Review recommends the FCA launch a review into the scale, nature and impact of general-purpose LLMs outside the perimeter, suggesting it begin within three to six months. The Adoption Plan makes the same review its Recommendation 2, marked high priority, paired with a Recommendation 3 that industry adopt a consistent, voluntary consumer disclosure to distinguish regulated from unregulated AI-driven services.

The asymmetry both documents describe is stark. Regulated firms carry strict obligations and liability while unregulated AI providers scale rapidly without equivalent safeguards, and consumers frequently cannot tell the difference. Left unresolved, this shifts trust and engagement away from

regulated institutions precisely as AI-driven advice becomes capable of closing the advice gap at population scale.

For firms, the perimeter review cuts both ways. It is a defensive measure against unregulated competition, but it is also the gateway to a substantial commercial opportunity: if proportionate guardrails allow AI-enabled guidance within the perimeter, the institutions best placed to provide it are those that can already evidence safe deployment.

OUR VIEW

The case for an early perimeter review is strong: consumer use, regulatory asymmetry and the potential for harm are already visible, and the Mills Review proposes beginning the work within three to six months. Firms should prepare on two fronts: disclosure practices for their own AI-driven services, and a position on what proportionate guardrails for AI-enabled guidance should look like, in anticipation of the policy and consultation process.

THEME 3

Resilience, dependency and the assurance market

THE ASSURANCE STACK

Outcome assurance	fairness · accuracy · drift · consumer impact
Implementation assurance	prompts · data · orchestration · human review
Provider assurance	security · transparency · resilience · capability
FOUNDATION MODELS	

The proposed scheme certifies the provider layer. Firms retain the two layers above it.

Both publications treat dependency on a small number of AI and cloud providers as a systemic issue rather than a procurement detail. The Mills Review warns that shared reliance on similar models, datasets and infrastructure could produce correlated behaviour, herding and common points of failure across the system. The Adoption Plan's response spans three recommendations: accelerate

the Critical Third Parties regime and assess key AI and cloud providers under it; establish voluntary, unattributed AI incident and near-miss sharing across the sector; and launch a voluntary, industry-led AI third-party assurance scheme, with standards developed "as soon as possible" using the assurance assessments firms already have in use as the benchmark. The CTP regime itself is no longer prospective: the Government has designated AWS, Google Cloud, Microsoft and Oracle as the first Critical Third Parties, with oversight commencing on 13 July 2026, and assessment of key AI providers is the logical next step.¹⁰

The assurance scheme is the recommendation we would draw boards' attention to first. The design described in the Adoption Plan, in which qualified third-party assessors evaluate AI providers against agreed standards and regulators potentially accept certification as evidence of baseline due diligence, would reduce the duplicated model-risk assessments currently run separately by every firm, and would begin to standardise what credible assurance evidence looks like.

"Internal checks are essential but not sufficient. You cannot mark your own homework."

PLENITUDE SUBMISSION TO THE FCA AI INPUT ZONE, JUNE 2026

It also reflects the position we put forward in our regulatory submissions: assurance that is credible to customers, regulators and counterparties requires independence, external challenge, reproducible tests and evidence others can verify, and supervisors will increasingly need to distinguish between internal documentation, self-attestation and genuinely independent assurance. The Adoption Plan's scheme is the institutional form of that argument.

One distinction will matter as the scheme takes shape, and the Adoption Plan itself points to it: certification of a provider is not certification of your use of that provider. We think of assurance as operating at three levels. Provider assurance covers the security, transparency, resilience and foundational capability of the model or platform itself; this is what the proposed scheme would standardise. Implementation assurance covers everything the firm builds around it: prompts, retrieval and data, orchestration, access controls, human review and integration. Outcome assurance covers the system in operation: fairness, accuracy, consumer impact, drift and continuing compliance. The scheme addresses the first level. Firms retain responsibility for the second and third, a point the Adoption Plan makes explicitly, and a certified model can still sit inside an unsafe use case.

Dependency was also one of the sharpest discussions at our June briefing. Participants questioned whether current frontier-model pricing reflects sustainable long-run economics; one observation shared in the room was that monthly token spend had doubled month on month. The concern raised went beyond budgets: if a small number of providers become deeply embedded in critical workflows,

their failure could begin to exhibit the characteristics of systemic dependency. That is precisely the territory the CTP regime is designed for, and the Adoption Plan is right to ask for its extension to AI providers.

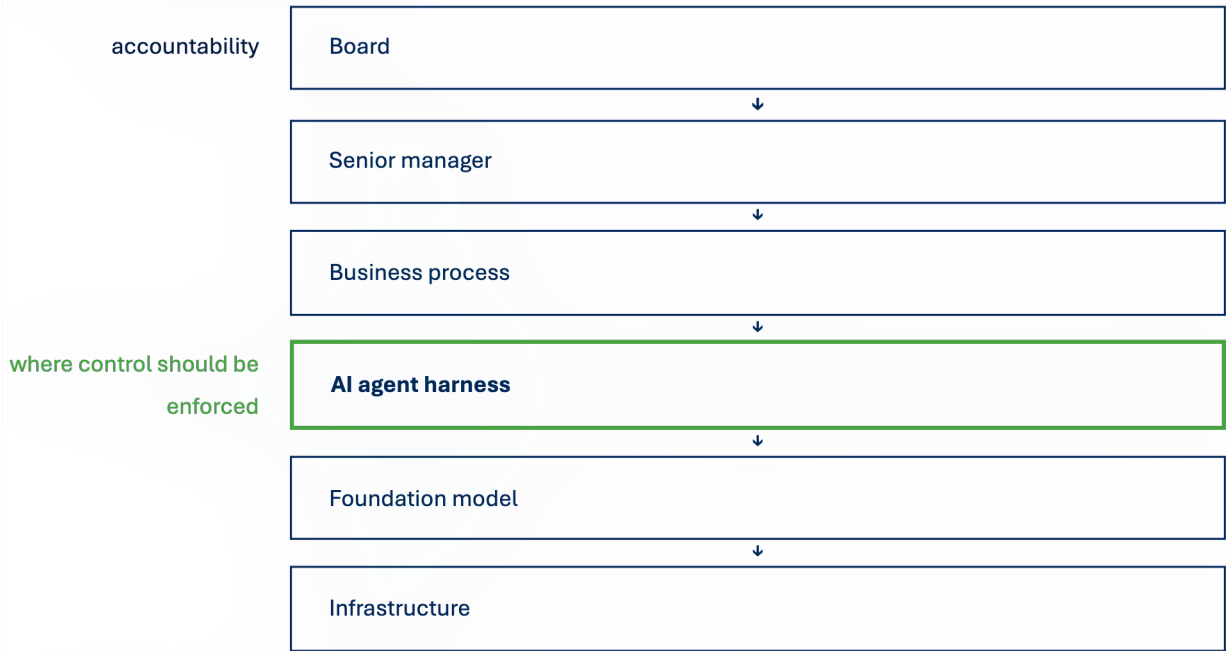
OUR VIEW

The assurance scheme will take time to stand up, and the plan itself notes that existing firm-level assurance assessments will be the benchmark for its standards. Firms that build credible evaluation and assurance evidence now will find themselves ahead of the scheme rather than retrofitting to it.

THEME 4

Agentic finance and the trust question

THE GOVERNANCE STACK



Both publications take agentic AI seriously as a near-term reality rather than a horizon topic. The Mills Review recommends the FCA lead development of a trusted framework for AI agent participation in financial services, clarifying how agents can be authorised, identified and held accountable, with clear expectations for consent mandates, identity, control and liability. The Adoption Plan concentrates on agentic payments as the practical proving ground, recommending a trust

framework built on three pillars: legal and liability constructs for autonomous transactions, Know Your Agent protocols for identity and verification, and interoperable standards for machine-to-machine authentication.

The questions these frameworks must answer were being asked, with some urgency, at our June briefing. Who verifies an agent's identity, provenance and behaviour when agents make decisions and execute transactions? What happens when an agent operating inside a firm's infrastructure is manipulated, or its guardrails are bypassed? The guardrail question is not hypothetical: independent testing presented at the briefing showed that prompt-level guardrails can be bypassed under adversarial conditions, reinforcing the need for system-level enforcement rather than reliance on model instructions alone.

"Telling an agent not to do something is not the same as preventing it."

PLENITUDE SUBMISSION TO THE FCA AI INPUT ZONE, JUNE 2026

The engineering answer is that control must be enforced at system and orchestration level, in the harness, and cannot rely primarily on instructions embedded in prompts. As models commoditise, the harness is where governance, constraint and auditability are enforced, and it is where commercial value is accruing. We put forward to the FCA that supervisory attention should extend to the orchestration layer as a primary point of control, including architectural patterns such as deterministic orchestration that keep routing and decision logic predictable and auditable even though the underlying models are not. The same holds at supervisory level: runtime governance for agentic systems, covering tool gating, continuous authorisation, trajectory-level monitoring and orchestration drift, is the largest gap in current supervisory toolkits.

OUR VIEW

The trust frameworks proposed in both publications are the right destination, but they describe the network layer of a stack whose firm-level foundations most institutions have not yet built. A model-agnostic harness strategy, owned at senior level and treated as a strategic asset, is the prerequisite for participating in agentic finance on defensible terms.

People, skills and the supervisory model

The final theme is the one most often treated as soft, and both publications refuse to treat it that way. The Adoption Plan devotes three recommendations to skills: industry participation in the Financial Services Skills Compact, a sector-wide AI skills plan covering boards, specialists and frontline teams, and visa reform to attract global AI talent. The Mills Review, for its part, recommends the FCA build an AI-enabled agentic supervisory model, on the logic that system-wide risks created by AI cannot be seen through firm-by-firm supervision, and that the regulator will need many of the same capabilities as the firms it supervises.

The workforce discussion at our June briefing went further than the skills-gap framing. The leaders in the room described AI adoption as an organisational change programme that touches how work is allocated and reviewed, how performance is measured, and how talent develops. Two concerns stood out. The first is generational: today's compliance professionals have the contextual knowledge to supervise AI outputs, but a generation that grows up normalising AI-generated work may not retain that independent capability. The second is continuity: what is the minimum viable business if the AI systems go down, and can anyone still do the work by hand? One firm represented at the briefing already operates a formal roll-back to human processing as an operational resilience mechanism.

We made the corresponding recommendation to the FCA: that adoption and change control be treated as a supervisory theme in its own right, because poorly managed adoption (insufficient steering, shared tooling and literacy) is one of the most common routes to ungoverned shadow AI use, and it is rarely addressed in guidance that focuses on controls alone.

OUR VIEW

Firms that treat AI as a technology procurement will find their governance describing a deployment that their workforce has already worked around. The skills recommendations in the Adoption Plan are a floor, not a ceiling; the differentiating investment is in AI literacy deep enough that accountable individuals can ask hard questions of the systems they sign off.

What the Mills Review and the AI Adoption Plan got right

It is easy for a paper like this to concentrate on gaps. It is worth being clear first about how much the two publications get right, because the things they get right were not inevitable.

The restraint is the achievement. The pressure to legislate something visible after three years of AI headlines was real, and both publications resisted it. The decision to extend the existing framework rather than build a parallel AI regime matches the evidence of what works. In our own practice we have managed AI risk the same way, extending established disciplines rather than inventing new ones and drawing on the NIST AI Risk Management Framework, the model-risk logic of SR 11-7 and the FCA's own risk-management expectations. Firms told both reviews the same thing. A new rulebook would have obsolesced on contact with the technology; an outcomes-based framework can keep pace with it.

Assurance is recognised as market infrastructure. The Adoption Plan's third-party assurance scheme is one of the clearest UK government statements to date that independent AI assurance should become shared market infrastructure rather than a service some firms buy. The design logic of standardised assessment, certification that regulators may accept as baseline due diligence, and firms retaining responsibility for implementation risk mirrors how audit built institutional trust in financial reporting. We think it is the recommendation with the greatest long-term consequence for how AI scales safely in the sector.

The accountability analysis is precise. The Review does not simply flag senior manager accountability as an open question; it maps it. It finds the Senior Managers Regime robust while humans operate as Operator, Collaborator or Consultant, and locates the emerging pressure point where delegation reaches Approver and Observer, where opacity and model drift could open a gap between AI-mediated outcomes and any identifiable responsible party exercising meaningful human control.¹ Just as significant for firms is what the Review says about the mitigation: assurance tools, spanning pre-deployment and ongoing checks, could better enable senior managers to take "reasonable steps" to prevent regulatory breaches, including risks inherited from third-party models operating upstream. That is the clearest regulatory statement yet that assurance evidence is how accountability will be discharged as AI autonomy grows.

Agentic finance is being prepared for before it arrives. Regulators have historically engaged with technology after deployment reveals harm. Both publications instead do the foundational work early: the autonomy spectrum gives supervision a shared language for delegation; the agent trust framework and Know Your Agent protocols address identity and liability before autonomous transactions scale; the agentic payments focus gives the abstract questions a concrete, near-term proving ground. The Mills Review's observation that risk shifts from harm within a single firm towards system-wide harms as autonomy grows is exactly the reframing supervision needs.

The threat picture is honest. The Mills Review's fourth system shift states plainly that AI will make fraud and cyber attacks faster, cheaper, more scalable and more persuasive, and that defensive capability must evolve at least as quickly. The Adoption Plan's broader considerations on cross-sector dependency mapping and scenario testing acknowledge that AI risk transmits across sector boundaries. Neither document pretends the technology is safe by default. Attendees at our briefing had already reached the same conclusion; at least one had placed frontier AI models on their corporate risk register, and the discussion treated that as prudent rather than premature.

The innovation infrastructure is already delivering. The recommendation to scale up the FCA's AI Lab extends something that demonstrably works. We saw this first-hand in July as the delivery lead for the synthetic AML dataset behind the FCA's Synthetic Data AML Solution Sprint, alongside the FCA, the Alan Turing Institute and Napier AI: ten teams, from specialist vendors to major regulated firms, used a shared synthetic dataset to build and benchmark detection approaches without exposing a single item of real customer data. Synthetic data of this kind also addresses one of AML's oldest weaknesses, the absence of labelled ground truth against which detection can actually be measured. That is the model the Mills Review proposes scaling: a safe, common platform where firms, vendors and the regulator generate the evidence supervision needs. The recommendation deserves support because it builds on a proven approach, not an aspiration¹¹

The supervisor is included in the transformation. The agentic supervisory model recommendation applies to the FCA the same standard the FCA applies to firms: if the work changes, the capability must change with it. A regulator that cannot see cross-firm patterns cannot supervise a system in which risk is correlated across firms. Asking the FCA to build AI-enabled supervision, and to assess AI models and systems through a scaled-up AI Lab, is a commitment to supervising the system as it will exist rather than as it used to.

The operational challenge

The strategic direction is right. The operational layer is not there yet, and the gaps that remain are precisely the ones holding deployment at proof-of-concept stage today. We highlight five, each of which we raised in our regulatory submissions, and each of which remains open after both publications.

"The toolkit helps supervisors ask the right questions, but is thinner on how to grade the answers."

PLENITUDE FEEDBACK TO IOSCO, AI SUPERVISORY TOOLKIT, 2026

THE PUBLICATIONS PROVIDE

Outcomes-based framework confirmed^{1 2}

CTP extension towards key AI providers^{2 10}

Agent trust framework and Know Your Agent^{1 2}

Perimeter review of general-purpose LLMs^{1 2}

Third-party assurance scheme and incident sharing²

Skills compact and sector-wide AI skills plan²



THE OPERATIONAL LAYER STILL TO COME

Benchmarks for grading governance and evaluation evidence

A revalidation standard for externally controlled model changes

Incentives that give voluntary schemes early critical mass

Production-readiness criteria beyond a working proof of concept

SM&CR sign-off treatment for autonomous decisions

Nobody has said what good looks like. The supervisory toolkits and principles now in circulation are effective at generating questions: about governance, accountability, third-party risk, monitoring. What they do not yet do is help anyone grade the answers. In our feedback to IOSCO we put it directly, noting the absence of any maturity model, adequacy benchmarks, or worked examples of good versus deficient control frameworks. The same gap runs through both July publications. Until expectations

are anchored in benchmarks and examples of acceptable, weak and insufficient evidence, supervisory consistency will depend on individual judgement, and firms will keep guessing at the standard they are being held to. The FCA's good and poor practice publication is the natural vehicle to close this gap; whether it does so is the single most important open question of the next six months.

The pilot-to-production gap is underestimated. Modern tools make a convincing proof of concept remarkably easy to build, and that convincingness is misleading. Small defects that are invisible in a demonstration are amplified sharply by real data at production volume. We know this first-hand: an independent assurance review of an AI pipeline built by our own Innovation Hub found release-blocking defects that were undetectable in normal use, in a tool whose every output looked plausible. That review was not an afterthought; it was the pre-release gate our own framework requires, with the evaluation designed before the pipeline was built. The system was rebuilt and passed the same tests before release; the point is that only systematic, component-level evaluation surfaced the problems at all. The leaders at our June briefing described the same pattern from the other side: proof-of-concept fatigue, with learning that does not transfer between pilots and no defined pathway to production. Explicit production-readiness criteria are needed: what a pilot has and has not proven, and what evidence should be required before deployment. Neither July publication takes them up. The Adoption Plan is a plan for adoption; it is largely silent on the discipline of the step from pilot to production, which is where adoption actually fails.

Re-certification has no answer. Model versions change under firms continuously, and moving between versions materially changes a system's outputs. Model-risk, change-control and software assurance standards all touch this problem, but no settled, sector-wide standard yet defines the extent of revalidation required when an underlying foundation model, data source or orchestration component changes, who re-approves, or what evidence re-approval requires. This question was raised repeatedly at our briefing as one of the two most directly delaying production deployment, and neither publication addresses it. Every firm is currently inventing its own answer. Our own approach is a reproducible evaluation harness that re-runs the full test suite on every model or data change, but a market of firms solving this independently is exactly the duplication the Adoption Plan elsewhere seeks to eliminate.

SM&CR accountability for autonomous decisions remains unresolved. The Mills Review maps the problem carefully, finding the regime robust at lower autonomy levels and identifying the pressure point at Approver and Observer, and it points to assurance tools as the route to "reasonable steps". But a route named is not a route specified. Senior leaders at our briefing were specific: the absence of clarity on who is personally accountable when an autonomous system makes a wrong decision, and on what basis they would sign off, is directly preventing production deployment of systems that have already proven technical capability. Our recommendation to the FCA stands: explicit treatment of

senior manager accountability for AI-supported and autonomous decisions, including what good sign-off evidence looks like. Accountability that cannot be operationalised is accountability that stalls adoption.

Voluntary schemes need a reason to join. The Adoption Plan's incident-sharing repository and assurance scheme are both voluntary, and both face the same collective-action problem: the firms whose participation matters most have the least short-term incentive to move first. The cyber information-sharing precedent the plan cites took years and regulatory encouragement to reach critical mass. The plan gestures at eventual standardisation and regulatory recognition, but sequencing matters; recognition that arrives after the scheme has stalled will not revive it. A clear early signal from regulators about how certified assurance will be treated in supervision would change the adoption calculus overnight.

A final observation on pace rather than content. The Mills Review proposes its perimeter review begin within three to six months; the technology it concerns ships new frontier models monthly. More than 20 were released during the Review itself, a point its own foreword concedes. None of the gaps above requires waiting for the next publication cycle to address, and the firms in the strongest position a year from now will be the ones that did not wait.

Practical actions for financial institutions

The broad regulatory direction is now considerably clearer, the operational detail is not, and the evidence expectations implied by both publications can be built now. Six actions follow.



Map your AI estate against the autonomy spectrum

The Mills Review's spectrum, from human as operator to human as observer, is about to become supervisory vocabulary. Inventory every AI use case, including the ungoverned ones, and place each on the spectrum. The exercise surfaces the two things a supervisor will ask about first: where humans are quietly becoming approvers and observers without the governance changing to match, and where shadow AI use sits entirely outside the map. At the briefing, a recurring theme was that firms discover more AI in their estate than their governance describes.



Assign named accountability that AI does not move

For every material AI use case, a named senior owner, close enough to the use case to challenge it, with documented sign-off criteria covering the risks that matter for that use case. Sign-off should be a genuine control event with evidence behind it, not a one-line approval. Nothing in either July publication softens this position; the SM&CR questions that remain open make demonstrable ownership more valuable, not less.



Design evaluation before you build, and make it reproducible

The evaluation framework (what will be measured, against which thresholds, producing what evidence) should exist before the first prompt is written, and should be re-runnable on every model or data change. This single discipline addresses three of the gaps above: production readiness, re-certification and evidence quality. Ask of every AI initiative the question we ask in assurance reviews: if the regulator wanted the evidence pack tomorrow, how long would it take to assemble? If the answer is more than a day, the assurance architecture was bolted on.

"The evaluation framework is designed before the first prompt is written."

PLENITUDE SUBMISSION TO THE FCA AI INPUT ZONE, JUNE 2026



Get ahead of the third-party assurance scheme

The Adoption Plan says the scheme's standards should be developed using the assurance assessments firms already have in use as the benchmark. That is an invitation. Commission independent assurance of your material AI systems now, against a defined framework with reproducible tests, and you will hold exactly the evidence the scheme will formalise, while shaping the benchmark rather than receiving it. Interrogate vendor claims with the same discipline: guardrail assurances deserve independent verification, not acceptance.



Treat adoption as an organisational change programme

Fund AI literacy for boards and accountable senior managers, not just delivery teams; define the minimum viable business if AI systems are unavailable, including roll-back to human processing for critical workflows; and manage the widening competency gap directly. The firms that answered these questions well at our briefing had one thing in common: they treated the human transformation as the programme, with the technology inside it.



Stress-test your provider dependency

The first CTP designations have been made, and AI providers are the logical next candidates. Map your dependency on foundation-model, cloud and orchestration providers; establish how quickly you could switch providers if required; and have the board stress-test AI running costs at normalised, unsubsidised pricing over a three-to-five-year horizon. A model-agnostic harness strategy, owned at senior level, is the practical mitigation for most of what this exercise will find.

OUR VIEW

None of these actions depends on the regulation landing a particular way. All six are no-regret moves under the direction now emerging, and four of them (evaluation design, independent assurance, dependency mapping and accountability) generate exactly the evidence the two publications suggest supervisors will ask for.

The debate has moved on. The question is no longer whether AI belongs in financial services, it already does.

The firms that will pull ahead and gain a competitive advantage are those that can evidence safe, controlled and accountable AI at production scale.

If you are a Board member, CRO, COO or senior risk leader working through these questions, we'd welcome the opportunity to discuss how your firm is approaching them and share our practical experience.

References

- ¹ The Mills Review: AI and the future of retail financial services. Financial Conduct Authority, published 6 July 2026.
- ² Financial Services AI Adoption Plan. HM Treasury independent report, published 14 July 2026.
- ³ The AI Tipping Point: what leaders need to know now. Post-event briefing paper, Plenitude and ExoBrain, June 2026.
- ⁴ Plenitude Consulting, response to the FCA AI Input Zone, submitted June 2026.
- ⁵ Plenitude Consulting, feedback to IOSCO on the AI Supervisory Toolkit, 2026.
- ⁶ IOSCO, AI Supervisory Toolkit for AI Use in Capital Markets, final report, May 2026.
- ⁷ Artificial intelligence in UK financial services. Bank of England and Financial Conduct Authority, November 2024.
- ⁸ The 2026 Global AI in Financial Services Report. Cambridge Centre for Alternative Finance, 2026; cited in the Mills Review.
- ⁹ Plenitude and ExoBrain, joint response to the FCA's proposal for AI Live Testing, 2025.
- ¹⁰ Critical Third Parties regime: first designations (AWS, Google Cloud, Microsoft and Oracle), oversight commencing 13 July 2026.
- ¹¹ FCA Synthetic Data AML Solution Sprint, showcase held at the FCA, July 2026. Plenitude was the delivery lead for the typologies and synthetic dataset, with the FCA, the Alan Turing Institute and Napier AI.

Plenitude

Built to combat the threat, Plenitude is a specialist economic crime firm driven by experience and accelerated by AI.

We work across every stage of the customer lifecycle: advising, transforming and assuring, empowering clients to combat economic crime, fulfilling regulatory obligations and delivering confidence in their ability to protect themselves, improve the customer experience and reduce the harm economic crime does to society. Appointed to the FCA Skilled Persons panel for Financial Crime, Market Abuse and Conduct, we help clients detect, prevent and remediate these threats with confidence.

We have a 14-year delivery track record, leading some of the most challenging programmes in the industry with some of the world's largest financial institutions across global banks, insurers and asset managers.

THE AI LAYER

We engineer trust from the outset, not as a by-product. Assurance holds the standard, the delivery model runs on every engagement, and our RegTech & AI solutions drive more effective risk management and operational efficiency.

01

AI Assurance Framework

Adapted from the NIST AI Risk Management Framework, operationalised through our Risk Taxonomy and G3MA Assurance Framework (Govern, Map, Measure, Manage, Assure), with model validation and workflow red-teaming: a defensible basis for Board sign-off.

02

AI-enabled delivery model

On every engagement. Engineered, not improvised: persistent context grounded in Plenitude methodology, engineered guardrails, and a named person accountable for every output. Days compressed to hours.

03

RegTech & AI solutions

An 8-year track record building and deploying RegTech products and AI solutions used by some of the world's leading financial institutions: Supernova, RegSight and Compass.



plenitudeconsulting.com

LinkedIn: Plenitude Consulting

ai@plenitudeconsulting.com