FinEthica
ADVISORY

FINETHICA WHITE PAPER SERIES/01

Monitoring AI Adoption in the Financial Sector

Building evidence for informed decisions



Ilham Tamimi
Elanur Türküz, PhD

November 2026

A practitioner's perspective



Ilham Tamimi

Founder and Managing Director
FinEthica Advisory FZCO

Ilham Tamimi is the Founder and Managing Director of FinEthica Advisory and a Senior Regulatory, Financial Crime Compliance and Governance Executive with thirty years at Citigroup, HSBC, First Abu Dhabi Bank and Zand Bank. She is an accredited Chief Compliance Officer and MLRO from UAE regulators. A Fellow Member of the ICA UK with Oxford University Saïd Business School credentials in Fintech, AI and Governance as a Certified AI Ethicist, she is a thought leader and sought after keynote speaker on digital assets, stablecoins, cross-border payments, financial crime prevention and AI governance in financial services.

Putting adoption evidence to work

Financial institutions increasingly need to explain where AI is used, what it is permitted to do and how its use is overseen. A count of adopters provides a useful starting point, but it cannot answer all the questions facing a board, a compliance officer or a financial authority.

This paper examines how monitoring can provide a more useful evidence base. Its concern extends to benefits and implementation barriers as well as risks. An institution needs to know when an application merits expansion, when a reported benefit remains untested and when a change in use calls for closer scrutiny.

Our interest is practical. In licensing, financial crime compliance and governance work, the same question takes different forms: a board asks how much AI the firm uses; an MLRO needs to establish whether a transaction-monitoring tool falls within its AI inventory; a licence applicant must explain a vendor stack it did not build. Each needs to connect the technology to the activity it supports, the actions it can take and the people responsible for it.

This paper gives those conversations a shared vocabulary. Table 3 and the appendix offer records an institution can begin keeping now, whether or not a regulator has requested them. Start with a defined activity, connect the information already held and identify what remains unknown. The aim is a record that helps the firm make decisions and supports external inquiries where their definitions align. The international evidence and UAE examples provide context for that practical work.

Ilham Tamimi

© 2026 FinEthica Advisory FZCO. All rights reserved.

A research perspective



Elanur Türküz, PhD

Academic Advisor, FinEthica Advisory
Department of Data Science and Analytics
İstanbul Nişantaşı University

Elanur Türküz, PhD, is an econometrician and Academic Advisor at FinEthica Advisory. Her work focuses on big data, machine learning and causal inference, including applications using large-scale international datasets. She connects measurement design and analytical methods with practical questions about AI adoption. Her contribution to this paper centres on interpreting monitoring evidence and distinguishing between describing patterns, predicting outcomes and explaining causes.

From measurement to useful evidence

As an econometrician, I approach AI adoption through a question that sits behind every headline statistic: what does the evidence allow us to conclude? An adoption rate is valuable because it describes the spread of a technology within a defined population. Its interpretation depends on what was counted, when it was observed and how consistently those choices were applied.

As AI becomes part of financial activity, the questions become more demanding. Knowing that an institution uses AI does not tell us how much of its work depends on it or whether a claimed improvement has been measured. More information can help, provided it is connected to a clear question and its limitations remain visible.

My interest in this collaboration is to bring measurement and analytical reasoning into the conversations that shape institutional decisions. Research contributes by making the basis of a claim explicit. A reported improvement, a predictive association and a causal effect answer different questions. Monitoring can reveal changes and guide further inquiry; it cannot, on its own, establish why an outcome occurred or which intervention will work.

This paper brings that research perspective together with experience in compliance and governance. The aim is evidence that readers can question, compare and use. For me, the practical test is whether a record helps someone distinguish what is known from what still needs investigation, and make a better-supported decision as a result.

Elanur Türküz, PhD

© 2026 FinEthica Advisory FZCO. All rights reserved.

Contents

A practitioner's perspective 2
A research perspective 3
How to use this paper 5
Abbreviations 6
Executive summary 7
01 The changes an adoption rate cannot show 8
1.1 A familiar technology with an expanding role 8
1.2 What recent evidence establishes 9
1.3 Monitoring as a continuing inquiry 9
02 Lessons from existing monitoring 10
2.1 Different exercises illuminate different parts of the sector 10
2.2 UAE implications for monitoring 12
2.3 Connecting use to its significance 12
03 Following AI into financial activity 13
3.1 Establishing whose adoption is visible 13
3.2 Understanding the role AI performs 14
3.3 Identifying the dependencies surrounding use 14
3.4 Examining governance in practice 15
3.5 Examining what use is producing 15
04 Building evidence that remains useful 16
4.1 Combining breadth with selective depth 17
4.2 Preserving meaning through change 18
4.3 Making evidence credible enough to use 18
4.4 Organising responsibility for the evidence 19
05 From monitoring findings to action 19
5.1 Giving findings a practical destination 20
06 Conclusion 21
Appendix Designing a monitoring exercise 21
Establish the purpose before selecting fields 22
Choose a manageable starting point 22
Connect the minimum records 22
Test questions against the intended decision 23
Publish the basis of the finding 23
Applying the approach within an institution 23
References 24
How FinEthica can help 26

© 2026 FinEthica Advisory FZCO. All rights reserved.

How to use this paper

This paper connects institutional records with sector monitoring for boards, compliance teams, financial businesses, authorities and sector bodies. Choose the route below for the decision you face, and adapt the recommendations to your institution, monitoring purpose and applicable framework.

Where to start, depending on your desk

Board members and business leaders	**Sections 1 and 5; then Section 3.2.** Ask which systems can act without prior approval, where staff can intervene and what evidence supports expansion. Bring a specific unresolved question to the next management or risk meeting.
Chief Compliance Officers and MLROs	**Sections 3 and 4; Table 3 and the appendix.** Include established compliance tools as well as newer assistants. Start with one live deployment and connect its task, permissions, owner, review evidence and outstanding actions. Check what "human review" means in practice.
Fintech and virtual asset leaders	**Section 3.3 and the institutional steps in the appendix.** Map the activity supported by each onboarding, analytics, model and cloud provider. Identify relevant upstream dependencies, what would stop during an outage and which fallback has been tested. This gives a licensing or governance discussion a concrete basis.
Regulators and sector bodies	**Section 2 and the appendix.** Use Table 1 to understand the boundaries of existing evidence. Define the decision your exercise should inform, then test a manageable baseline and one focused inquiry before expanding the collection.
Technology builders and providers	**Sections 3.2 and 3.5.** Help customers describe the actions the system permits and distinguish expected benefits from observed results. Keep evidence on service quality, review, correction and total operating cost that customers can interpret in their own setting.

A useful first step: choose one important deployment, complete the relevant parts of Table 3, mark unknowns and assign a dated follow-up. The record earns its place when it helps someone decide what to examine or do next.

This structured narrative review draws on selected public official reports, surveys, supervisory publications and research available by 1 November 2026. Sources were selected for relevance to financial-sector adoption, measurement transparency and monitoring design. It is not a systematic review or a new empirical survey. Definitions and populations follow each source. Public reports do not disclose all supervisory information; recommendations and fictional examples are identified as such.

© 2026 FinEthica Advisory FZCO. All rights reserved.

Abbreviations

AI	Artificial intelligence	**FINMA**	Swiss Financial Market Supervisory Authority
AML/CFT	Anti-money laundering and countering the financing of terrorism	**FSB**	Financial Stability Board
BIS	Bank for International Settlements	**FZCO**	Free Zone Company
CBUAE	Central Bank of the United Arab Emirates	**ICA**	International Compliance Association
CCO	Chief Compliance Officer	**IMF**	International Monetary Fund
DFSA	Dubai Financial Services Authority	**MLRO**	Money Laundering Reporting Officer
DIFC	Dubai International Financial Centre	**NIST**	National Institute of Standards and Technology
DOI	Digital object identifier	**OECD**	Organisation for Economic Co-operation and Development
EBA	European Banking Authority	**OSFI**	Office of the Superintendent of Financial Institutions
EEA	European Economic Area	**SDG**	Sustainable Development Goal
EIOPA	European Insurance and Occupational Pensions Authority	**SEO**	Senior Executive Officer
EU	European Union	**UAE**	United Arab Emirates
FATF	Financial Action Task Force	**UK**	United Kingdom
FCA	Financial Conduct Authority	**US**	United States
FCAC	Financial Consumer Agency of Canada	**VARA**	Virtual Assets Regulatory Authority

© 2026 FinEthica Advisory FZCO. All rights reserved.

Executive summary

An institution can remain an AI adopter while almost everything that matters about its use of AI changes. An experiment may become a routine service; an assistant may gain access to customer information or authority to act. A deployment may create useful improvements, become essential to operations or prove unsuitable. An adoption rate cannot distinguish these developments by itself.

Existing monitoring initiatives provide a substantial foundation. Financial authorities have examined uses, deployment stages, benefits, governance and risks for years. Differences across exercises often reflect legitimate differences in mandate, population and technology scope. Their findings should be interpreted within those boundaries, and public reporting should not be treated as a complete account of supervisory knowledge.

This paper argues that monitoring should connect the institution to its deployments, the financial activities they influence and the evidence needed to understand their significance. Relevant information includes purpose, reach, decision authority, dependencies, oversight and experienced outcomes. Keeping these connections current helps distinguish experimentation from operational reliance and expected benefits from observed results.

Governance becomes more visible when monitoring goes beyond the existence of a policy. For consequential uses, evidence about reviews, findings, corrective-action owners and verified responses helps assess whether oversight remains appropriate. This is one dimension of a broader inquiry that also examines barriers to adoption, abandoned applications and opportunities worth supporting.

A practical programme combines a consistent baseline with selected deeper inquiries. It needs clear definitions, proportionate validation, a way to capture significant changes and named users responsible for considering the findings. An institution can start with one important activity, link the records it already holds and assign follow-up for missing evidence. Table 3 and the appendix support that first review. Maintaining the connections requires resources, even where existing records can be reused.

Monitoring can identify problems and inform decisions. Routine observations alone do not establish causal effects or determine which intervention will work best. Preserving the basis, timing and limitations of a finding allows further evaluation to investigate those questions without overstating what the monitoring evidence demonstrates.

© 2026 FinEthica Advisory FZCO. All rights reserved.

01 The changes an adoption rate cannot show

An AI adoption rate can remain unchanged while the financial sector's dependence on AI changes substantially. Once an institution qualifies as an adopter, further developments inside it may have no effect on that percentage. Yet those developments may be exactly what a supervisor, a policymaker or the institution's own management needs to understand.

Consider a fictional financial institution that introduces a generative AI assistant to help employees find internal guidance. Initially, the tool searches a limited collection of documents and produces draft answers for trained staff. Later, it is connected to customer records and begins preparing responses that staff approve and send. Management then considers allowing it to complete selected service requests. The institution would answer "yes" to an adoption question throughout this sequence, although the tool's authority, reach and operational importance would change. It might improve service, or the time needed for review and correction might consume the expected savings. This example illustrates a measurement problem; it does not describe an observed institution.

The problem extends beyond a single firm. If many institutions follow similar paths, an unchanged sector adoption rate may conceal a widening role for AI in customer interactions. If they rely on a common external service, the pattern may also create an operational dependence that is difficult to see from institution counts. Conversely, a rising rate could reflect numerous small experiments with little connection to consequential financial decisions. The percentage alone cannot distinguish these situations.

This does not make adoption rates uninformative. A well-defined rate establishes how widely a technology is being used within an observed population. Repeated on a consistent basis, it can show diffusion and help identify which types of institutions participate. It can support decisions about the reach of guidance, the need for training, or the population from which to select further inquiry. Its value depends on retaining the question it actually answers. A measure of the presence of AI becomes misleading when it is treated as a measure of operational importance, effective governance or realised value.

1.1 A familiar technology with an expanding role

Financial institutions were using AI well before the recent attention to generative systems. Fraud detection, credit assessment, suspicious-transaction analysis, insurance pricing and trading strategies already provided reasons to examine the technology's role in finance. The Financial Stability Board (FSB, 2017) discussed financial applications and concerns about shared data and third-party dependence. Established model, data and operational questions remain relevant as new applications emerge.

Financial crime compliance provides another established application. The Financial Action Task Force's (2021) report on new technologies for anti-money laundering and counter-terrorist financing discusses machine learning applications and the opportunities and challenges of their use. For a CCO or MLRO, an inventory review should therefore consider established transaction-monitoring and screening tools alongside newer assistants. Whether a particular tool counts as AI depends on the definition and technology used; a rules-based threshold alone does not establish AI use. Established deployments may also change in scope and significance.

Generative AI expands this history through its ability to produce and work with language, code and other content. Information retrieval, document analysis, customer support and software development bring AI into tasks whose outputs are often less narrowly specified than a score or classification. The Organisation for Economic Co-operation and Development's (OECD, 2023) analysis described such opportunities in finance. For monitoring, the practical implication is that a technology label reveals less than the task performed

© 2026 FinEthica Advisory FZCO. All rights reserved.

and the conditions of use. An internal drafting assistant and a system that responds directly to customers may use a similar underlying model while creating different demands on evidence and oversight.

Some risks associated with these applications are familiar: poor data, inappropriate disclosure, biased treatment and weak access controls did not originate with generative AI. What can change is the route through which they arise and the difficulty of recognising them. A plausible but unsupported answer may enter a workflow as if it were verified information. An interface may expose information or invoke tools beyond the intended task. The US National Institute of Standards and Technology (NIST) identifies confabulation, harmful bias and information-security concerns in its generative AI risk profile, a voluntary cross-sectoral resource (Autio et al., 2024).

The same reasoning applies to systems described as agentic. The monitoring question is what actions the system is permitted to take, under what constraints and with what opportunity for intervention. A label cannot reveal whether the system merely proposes an action, prepares it for approval, or executes it. Nor does a recorded human approval step establish that the reviewer has the time, information and authority to challenge the output. The OECD's (2026) work on AI supervision emphasises the contextual nature of human involvement and the challenge of automation bias.

A model can remain technically unchanged while its significance changes because an institution connects new data, widens its user population or grants additional permissions. The same reasoning applies to established AI: a fraud system that moves from ranking alerts to blocking transactions acquires a different role even if the institution has used AI for years. Monitoring should make such changes visible and prompt reassessment of the relevant controls. The appropriate response may be to adapt a deployment, retain its existing limits or discontinue it; continual modification is not an objective in itself.

1.2 What recent evidence establishes

Recent official findings show extensive engagement with AI, but engagement has several meanings. The 2024 Bank of England and Financial Conduct Authority (FCA) survey found that 75% of 118 respondents used AI (Bank of England & Financial Conduct Authority, 2024). The European Insurance and Occupational Pensions Authority (EIOPA, 2026) published findings in February 2026 from 347 insurers in 25 EU and EEA countries, surveyed from May to July 2025. While 65% reported generative AI use, 64% of the reported use cases were experiments or proofs of concept. The latter percentage counts use cases, not insurers.

Institutional participation and deployment maturity therefore answer different questions. The UK and insurance figures cannot rank adoption: they cover different technologies, populations and observation periods. An institution may use generative AI while most applications remain experimental.

The Dubai Financial Services Authority (DFSA, 2025) offers evidence from a clearly bounded UAE population. Its 2025 survey covered 661 Authorised Firms in the Dubai International Financial Centre (DIFC), with an 88% participation rate. The published results reported AI use by 52% of firms, compared with 33% in the previous round, alongside developing governance arrangements and gaps in accountability. These findings help describe both adoption and reported governance. They do not, by themselves, establish whether controls work effectively in every deployment. That requires evidence about the particular activity, its oversight and the results of relevant checks.

These findings support a practical conclusion: the question of whether AI is present is already accompanied by questions about how it is used and how far deployment has progressed. They do not support a single adoption rate for the financial sector worldwide. Nor do they establish that widespread use necessarily produces widespread value or harm. Those conclusions require evidence closer to the activity and outcome concerned.

1.3 Monitoring as a continuing inquiry

Monitoring AI adoption is the organised, continuing process of collecting, checking, interpreting and updating evidence about AI use and its development. This paper focuses on sector-level inquiry across institutions and financial activities. Technical monitoring of an individual model and an institution's internal oversight can supply evidence to that inquiry, but have different immediate purposes. Monitoring can identify problems and guide action; routine observations alone do not establish causal attribution.

© 2026 FinEthica Advisory FZCO. All rights reserved.

This distinction is consistent with the broader treatment of monitoring and evaluation in public governance (OECD, 2019).

A survey is one instrument within this process. The process also includes deciding whom to observe, identifying what has changed, resolving unclear responses, and determining whether a finding calls for action or further investigation. A publication may be a valuable output, but producing it does not by itself establish that the evidence is current enough, complete enough or sufficiently connected to a decision.

The fictional assistant illustrates why continuity matters. A snapshot can describe the initial deployment accurately and still become inadequate after the application gains access to customer records. Repeating the original question without recognising the changed role would create a time series with little insight into the important transition. Monitoring has to preserve enough consistency to detect change while remaining responsive to changes in the object being observed.

This is not an entirely new demand on financial authorities. Existing exercises have already gone well beyond asking whether AI is used. Before proposing improvements, it is necessary to establish what those exercises contribute and where their evidence leaves a decision unresolved. Their experience provides a stronger foundation than an assumption that monitoring must start again.

02 Lessons from existing monitoring

The history of AI adoption monitoring does not follow a simple progression from counting users to understanding risks. Earlier exercises already examined deployment, benefits, governance and obstacles. What has changed is the combination of technologies, applications and dependencies that monitoring is asked to explain. Reading the history in this way recognises the accumulated work and makes it possible to identify improvements with greater precision.

The Bank of England and FCA's sequence of surveys provides a useful example. The 2019 exercise examined machine learning deployment alongside governance, benefits and risks. Its sample deliberately included larger, faster-growing and technologically advanced firms, and the report cautioned that the results did not represent the whole sector (Bank of England & FCA, 2019). By 2022, the survey distinguished several stages of application development and deployment. Its headline finding combined institutions using machine learning with those developing applications (Bank of England & FCA, 2022). The 2024 exercise used a broader AI scope and reported foundation-model use, autonomy, materiality and third-party implementation (Bank of England & FCA, 2024).

The lesson is a development in the questions and the detail available, rather than the late discovery that adoption has several dimensions. The sequence also exposes a tension that every recurring programme must manage. Updating a survey can make it more relevant to current technology while making its results harder to compare with earlier rounds. The label "adoption" does not make a sequence consistent when one round covers machine learning, another covers broader AI, and categories or responding populations differ. A plausible trend should not be constructed simply by placing the published headline percentages in chronological order.

For a monitoring designer, the implication is constructive. Preserve the observations needed to understand continuity but document the points at which a change in scope improves relevance at the cost of comparability. Where feasible, a transition round can record both an earlier category and a revised one. Where that is not feasible, a clearly explained break in the series is more informative than an uninterrupted line with an unstable meaning.

2.1 Different exercises illuminate different parts of the sector

Variation across authorities is another source of information. A prudential authority may need to understand whether an institution can maintain a critical service. A conduct authority may need evidence

© 2026 FinEthica Advisory FZCO. All rights reserved.

about the treatment of customers. An insurance supervisor will encounter a different distribution of activities from a securities regulator. A sector-development body may be trying to understand why promising applications are not reaching production. These responsibilities can overlap, but they do not create identical information requirements.

Table 1 brings together selected official exercises to show this diversity. It compares the scope and contribution of the evidence rather than ranking jurisdictions by adoption. The selection illustrates monitoring approaches; it is not an exhaustive inventory or an assessment of the full supervisory work of the institutions concerned.

Table 1 Selected official evidence on AI adoption

Exercise	Population and evidence base	What the published work helps illuminate
Bank of England & FCA (2019, 2022, 2024)	Successive UK financial-services surveys with changing samples and technology scope	Deployment and governance over time, with more recent detail on autonomy and external implementation
OSFI & FCAC (2024)	Voluntary questionnaire shared with federally regulated financial institutions in December 2023, supplemented by research	Uses, preparedness and risks viewed through prudential and consumer interests
FINMA (2025)	Around 400 licensed institutions, surveyed from late November 2024 to mid-January 2025	Adoption and development alongside governance and risk management across financial subsectors
DFSA (2025)	June 2025 survey of DFSA Authorised Firms in the DIFC, with 661 respondents and an 88% response rate	AI use, generative applications and governance within a clearly bounded financial centre population
EIOPA (2024, 2026)	Digitalisation monitoring followed by a dedicated generative AI market survey	Insurance-specific applications and the distinction between experimentation and production
EBA (2025)	AI evidence drawn from the Risk Assessment Questionnaire	Integration of technology questions into an established banking risk-assessment channel
FSB (2025)	Survey of authorities, with 28 responses covering 19 jurisdictions and one international body	How authorities monitor adoption and related vulnerabilities, rather than a direct global estimate of firm adoption

Source: The publications cited in each row. Dates identify publication years. A public report does not disclose every item collected or every analysis conducted internally.

Other abbreviations in Table 1: OSFI, Office of the Superintendent of Financial Institutions; FCAC, Financial Consumer Agency of Canada; FINMA, Swiss Financial Market Supervisory Authority; EBA, European Banking Authority.

The Canadian example is particularly instructive for the relationship between responsibility and evidence. The involvement of both OSFI and FCAC brings prudential and financial-consumer perspectives into the same risk report (OSFI & FCAC, 2024). This does not remove the difference between their interests. It shows why one information base may need to support several interpretations. The same customer-service application can matter because it reduces an institution's operating costs, changes the quality of information given to customers, and creates dependence on an external provider. Each question follows a different consequence of the same adoption decision.

EIOPA's move from a broader digitalisation exercise to a dedicated generative AI survey illustrates a different form of adaptation. A new collection can investigate a development in greater depth without replacing the wider picture of technological change (EIOPA, 2024, 2026). This suggests that continuity need not depend on making every successive survey longer. A recurring baseline and a focused inquiry can serve complementary purposes.

The DFSA example shows the value of describing the reporting boundary and participation clearly. Its findings concern DFSA Authorised Firms in the DIFC and should not be extended to all financial activity in the UAE. Different respondent counts across rounds also mean that a published year-on-year comparison is not automatically a measure of change within the same firms (DFSA, 2025). These qualifications identify where the evidence is useful and where additional information is needed.

© 2026 FinEthica Advisory FZCO. All rights reserved.

2.2 UAE implications for monitoring

For institutions operating in the UAE, the monitoring argument has a practical connection to existing governance and supervisory arrangements. The relevant instruments differ in scope and legal status. They should be read according to the institution, activity and jurisdiction concerned, rather than combined into a single set of duties applying to every UAE firm.

The Central Bank of the UAE's Guidance Note on consumer protection and the responsible use of AI and machine learning, issued in February 2026, sets out principles for licensed financial institutions. Its governance section calls for an inventory of AI models, systems or technologies, with metadata including model name, purpose and risk rating. It also addresses board and senior-management accountability and reporting on performance and risk (Central Bank of the United Arab Emirates, 2026). Such records offer a starting point for understanding institutional use, although a sector inquiry may need additional information about deployment scale, dependencies and experienced outcomes.

In the DIFC, the DFSA issued a Dear SEO letter on its regulatory expectations for AI risk management on 4 June 2026 (DFSA, 2026). This communication and the DFSA's adoption surveys serve related but distinct purposes: supervisory expectations concern how firms manage their activities, while surveys provide evidence about the population and developments within it. A monitoring programme should identify where existing institutional records can answer its questions, rather than assuming that a new questionnaire is always necessary.

DIFC Data Protection Regulation 10, introduced in 2023, addresses personal data processing through autonomous and semi-autonomous systems (Dubai International Financial Centre, 2023). Its data-protection scope should remain distinct from the wider task of monitoring AI adoption. Information maintained for a specific regulatory purpose may contribute to that task, but an inventory, a personal-data record and a sector return are not interchangeable.

These examples support reusing relevant records where their definitions and permitted uses align. They do not establish that every field proposed in this paper is already required, or that linking existing records involves little additional work. The appendix is a design aid for monitoring, to be adapted to the applicable framework and the decision it is intended to support.

2.3 Connecting use to its significance

Existing work provides evidence of adoption and the conditions surrounding it. A continuing challenge is to connect observations closely enough to interpret their significance. Knowing that a firm uses AI, relies on a third party and has a governance policy does not establish whether that policy is implemented in the deployment on which a critical service depends. The relevant inquiry concerns which controls apply, what their operation reveals and how unresolved issues are handled. Public reports may not disclose the deployment-level evidence available internally to supervisors, so gaps in published information should not be treated as proof that an authority lacks that information.

The FSB's (2025) monitoring report makes this challenge explicit at the level of financial vulnerabilities. It discusses the development of monitoring, the distinction between direct measures and proxies, and the importance of criticality and substitutability in assessing dependencies. Its survey concerns authorities' practices, so it should not be read as a direct census of global financial-firm use. This is evidence that the monitoring agenda extends beyond documenting the spread of technology to examining how that spread may matter for the financial system.

A similar issue arises when evidence is assembled from different sources. The International Monetary Fund's (IMF, 2024) work on AI in capital markets draws on information including patents, job vacancies, professional profiles and stakeholder outreach. These sources can reveal activity or capabilities that a firm survey alone may miss. They do not all observe the same thing: hiring for AI expertise may signal capacity building without establishing that a consequential application is in production. Combining sources is useful when their relationship to the question is understood. Simply accumulating them does not make the conclusion more secure.

The strongest lesson from these examples is that monitoring quality cannot be judged by the number of questions, the sophistication of a dashboard or the prominence of a headline. It depends on whether the evidence preserves the relationships needed for

© 2026 FinEthica Advisory FZCO. All rights reserved.

interpretation. An authority with a narrow, well-defined purpose may obtain more useful evidence from a focused exercise than from an extensive questionnaire with ambiguous categories. Equally, a broad baseline may be indispensable when the location of relevant activity is not yet known.

This leads to a more precise design problem. The aim is not to give every stakeholder the same information or to collect every possible characteristic of AI. It is to determine which connections make adoption intelligible for the decisions at hand. The next step is to follow those connections from the institution that adopts AI into the financial activity that AI influences.

03 Following AI into financial activity

A useful account of adoption must connect several levels of observation. Institutions make investment and governance decisions. Applications perform particular tasks. Models and external services supply capabilities that may be shared across applications. Financial activities are where benefits, dependencies and adverse outcomes are experienced. Treating these levels as interchangeable obscures the relationships that monitoring is meant to reveal.

An institution may have one model contract, several applications and many users. These counts each serve a purpose, but none establishes how much customer activity depends on the technology. Monitoring must specify what is counted and connect that unit to the decision it informs.

Consider a fictional virtual asset service provider. It uses one model service in four applications: checking onboarding documents, drafting support replies, summarising blockchain analytics alerts and proposing treasury trades for staff approval. A separate analytics vendor embeds a risk-scoring engine in its product. The firm appears once in an adoption count, but its exposure runs through several activities and includes a system it did not build. For a licensing or internal governance review, the useful questions are which applications can act, what each one touches and which activities would stop if either provider became unavailable. A supplier list cannot answer those questions. A connected deployment record can distinguish support for a reviewer from authority to reject an application or execute a trade, while identifying the recovery arrangements that need checking.

3.1 Establishing whose adoption is visible

The starting point remains the institution. Its size, business model and activities help explain the setting in which AI is being introduced. Similar deployments may have different implications in a large retail bank, a specialist investment firm and a small insurer. An aggregate result can therefore be useful for describing the observed population while concealing differences that matter for a particular decision.

This is especially important for understanding constraints. If a sector body wants to help institutions adopt useful technology, a survey of prominent adopters will provide an incomplete account. Institutions that have not adopted may face difficulties with data, skills, integration, cost or procurement. Others may have considered an application and found that it did not fit their needs. Treating all non-use as a capability deficit would turn a measurement category into an unsupported policy conclusion. Monitoring should give respondents room to distinguish inability, uncertainty and a considered decision not to proceed.

The same principle applies within adopting institutions. An abandoned pilot may provide more information about implementation barriers than a count of active experiments. A deployment that has been scaled back may reveal a problem with costs or service quality. A programme that records only new and continuing applications can make adoption appear to move in one direction. Including discontinued uses helps explain how the sector is learning and where an intervention might be useful.

© 2026 FinEthica Advisory FZCO. All rights reserved.

Coverage also depends on the reporting boundary. A group response can combine activity across subsidiaries, while legal-entity returns may duplicate information about a shared service. Embedded AI features may be known to operational teams but be absent from a central inventory. Staff use of public tools may be governed separately from approved production systems. These are reasons to specify the perimeter and test how respondents interpret it. They do not justify assuming that every omission is unreported misconduct or that a survey can observe every form of use.

Once the population and boundary are clear, the institution profile provides a map for deeper inquiry. It can identify where activity is concentrated, whose experience is missing and which differences deserve explanation. The profile becomes more useful when it is connected to the role of particular deployments.

3.2 Understanding the role AI performs

For the assistant, the move from searching internal guidance to preparing customer responses changes the relevant activity even if the underlying model is the same. The monitoring record must therefore describe purpose and actual use, as well as technology. A category such as "customer service" may be a sensible starting point, but it does not reveal whether AI retrieves information for an employee, drafts a message, recommends a product or executes a service request.

Information can influence a decision before a system has formal authority to make it. Staff may rely heavily on a recommendation while retaining the final approval step. Conversely, a system may execute a narrow action within tightly defined limits. Monitoring should examine where AI enters the workflow, which actions it can take and where people can change or stop the result. A recorded human-in-the-loop arrangement is meaningful only in relation to the task and evidence of how review operates.

Scale adds another dimension. A small pilot and a routine service may perform the same task but reach very different numbers of people. The useful denominator is the relevant activity: for example, eligible customer contacts handled with assistance, rather than all transactions undertaken by the institution. A low share of a broad denominator can conceal heavy dependence within a particular service. Equally, high use in a routine administrative task should not automatically be equated with high influence over consequential financial decisions.

In the 2024 UK survey, respondents rated 62% of use cases low materiality, 22% medium and 16% high (Bank of England & FCA, 2024). Adoption status therefore does not establish significance. Monitoring designers should specify materiality criteria that match their purpose, including relevant financial, service and customer consequences.

The resulting evidence serves several readers without requiring them to reach the same conclusion. Management may see a case for expanding a useful service. A conduct authority may seek closer evidence about the accuracy of customer communications. A prudential supervisor may focus on continuity if the service becomes essential. The differences in response arise from responsibilities and consequences, rather than from a disagreement over whether the institution uses AI.

3.3 Identifying the dependencies surrounding use

Following the activity also reveals why institution-level information alone can be insufficient for understanding shared exposures. The assistant may be delivered by a software supplier, rely on a foundation model from another provider and operate through a cloud service. Several institutions can buy differently branded applications while depending on a common upstream component. Gambacorta and Shreeti's (2025) analysis of the AI supply chain describes these interdependent layers of infrastructure, data, models and applications.

The monitoring implication is to trace dependencies far enough to answer the relevant question. If the question concerns continuity of a customer service, knowing the immediate supplier may not reveal the component whose interruption would stop it. But mapping the whole technical supply chain for every small experiment would impose costs without an equivalent benefit. The depth of inquiry should follow the activity's importance and the uncertainty that needs to be resolved.

A shared provider is an exposure signal, not a complete assessment of vulnerability. Its significance depends on which activities rely on it, the alternatives available and the time needed to restore service. A firm with a credible, tested fallback may face a different continuity

© 2026 FinEthica Advisory FZCO. All rights reserved.

problem from one with the same provider but no workable alternative. At sector level, the question also concerns whether many institutions would need to switch or recover at the same time.

The FSB's (2024) assessment identifies third-party concentration, market correlations, cyber risk and model and data weaknesses as potential financial-stability channels. It does not establish that the presence of AI has already produced systemic damage. Monitoring can help investigate such channels by connecting exposure to critical activities and recovery arrangements. It should retain the distinction between identifying a plausible mechanism and demonstrating that the mechanism is operating.

This distinction matters beyond resilience. Competition and sector-development questions may concern the affordability of access, switching costs or the ability of smaller institutions to obtain suitable services. Evidence about a common dependency can therefore inform several inquiries, with different additional information required for each. Coordination is valuable because the underlying facts overlap, even when the decisions remain separate.

3.4 Examining governance in practice

Governance is one of the conditions that monitoring needs to understand. An institution may have an AI policy, a committee and a review process while important questions remain about a particular deployment. For consequential uses, the inquiry should connect the responsible owner and applicable controls to evidence of how those controls operate. This supports a judgement about oversight without treating the presence of a framework as proof of its effectiveness.

For the assistant, evidence might include the date and scope of the last review, the problems identified, any operating restrictions and the status of agreed corrections. An unresolved issue should have a responsible owner and a route for escalation. Closure should identify what was checked and by whom, rather than simply record that an action was marked complete. The detail collected should reflect the significance of the deployment and the purpose of the inquiry.

Findings require interpretation. Numerous control findings may indicate extensive testing, a deteriorating deployment or both. Few findings may reflect effective safeguards, limited exposure or weak detection. A review count or a list of completed actions therefore needs context about coverage, severity and the basis for concluding that a problem has been addressed. Monitoring should preserve that context instead of rewarding a low number of reported issues.

This approach connects governance to changes in use. If the assistant acquires authority to complete a service request, the monitoring record should show whether the change was assessed, who authorised it and whether the oversight arrangements were reconsidered. The appropriate response may be to maintain existing limits, introduce further controls or stop the use. Monitoring supplies evidence for that decision; a full governance assessment may require additional examination.

3.5 Examining what use is producing

Purpose, scale, dependence and oversight describe the conditions in which AI operates. They do not establish its effects. If the assistant is intended to improve customer service, monitoring should distinguish an expectation of shorter response times from evidence that customers receive accurate, timely and useful answers. A statement that staff save time also needs to account for review, correction and work displaced elsewhere in the process.

For management, this distinction affects whether expansion is worthwhile. For a sector body, it affects whether a celebrated application offers a transferable lesson. For authorities, it may affect the balance between enabling innovation and investigating adverse outcomes. Monitoring that records perceived benefits can still be valuable, especially early in deployment, provided those reports are identified as perceptions. They become a stronger basis for decisions when linked to an observable measure and an appropriate comparison.

A before-and-after change is not automatically an AI effect. Staff training, a redesigned workflow or a different customer mix may have changed at the same time. This paper does not set out the analytical methods needed to establish causation. Its narrower point is that the monitoring record must preserve information about timing, conditions and the basis of a claim, so that descriptive findings are not presented as conclusions they cannot support. Evaluation can then investigate questions that monitoring has made visible.

© 2026 FinEthica Advisory FZCO. All rights reserved.

Adverse outcomes require the same care. A rise in reported errors may reflect more use, better detection or a real deterioration. A low complaint count may reflect good performance, but it may also reflect limited awareness or an ineffective route to redress. Interpreting a count requires some understanding of exposure and detection. Where appropriate and lawful, the distribution of outcomes across affected groups may also matter. A disparity is a reason for inquiry, not by itself a complete explanation or a legal finding.

The New York State Department of Financial Services' Apple Card investigation illustrates why a reported concern and an established finding must be distinguished. Following allegations about unequal credit outcomes, the department examined underwriting data for nearly 400,000 New York applicants. Its 2021 report did not find evidence of unlawful discrimination, but identified deficiencies in customer service and transparency (New York State Department of Financial Services, 2021). The relevance here is methodological: complaints justified inquiry, while the investigation supported a more specific conclusion about the problem requiring attention. The case is not presented as evidence that AI caused the alleged disparity.

Table 2 shows how these distinctions change the evidence needed for the same illustrative application. The table is a set of analytical examples, not an account of any institution or a prescribed reporting obligation.

Table 2 Different decisions about the same customer service assistant

Decision	Evidence that would help	What that evidence would not establish alone
Expand the service	Completed tasks, service quality and total cost including review and rework	That AI caused all observed improvement or that results will transfer to another setting
Examine customer treatment	Verified errors, affected contacts, complaints, correction and recourse	That low complaint numbers prove good outcomes or that every error has the same consequence
Assess continuity	Activities dependent on the service, upstream dependencies and tested fallback performance	That multiple supplier names provide independent alternatives
Support wider adoption	Implementation obstacles, discontinued pilots, skills and integration requirements	That every non-adopter should adopt or that the same support suits every institution
Assess oversight	Review coverage, significant findings, action owners and evidence of resolution	That a policy or a completed action alone demonstrates effective governance

Together, these connections provide what this paper calls AI adoption visibility: an interpretable account of who uses AI, how it enters financial activity and why that use matters. The term describes the understanding that monitoring supports, rather than a new indicator or a claim that the entire sector can be observed. It includes recognising where evidence is incomplete or insufficient for a decision.

That visibility will deteriorate if the underlying records become stale or cannot be checked. The assistant's role can change after it is assessed, and a reported safeguard can cease to match its permissions. A useful snapshot therefore needs an institutional process that keeps the evidence connected, credible and current. This is where the design of monitoring becomes decisive.

04 Building evidence that remains useful

The preceding analysis creates a demanding brief for monitoring. It asks for enough breadth to understand the distribution of adoption and enough depth to interpret consequential uses. It also asks for continuity without freezing the questions around yesterday's technology. These aims compete for limited reporting

© 2026 FinEthica Advisory FZCO. All rights reserved.

and analytical capacity. A practical design must make the trade-offs explicit.

The first choice is the question the programme is expected to inform. An authority looking for emerging dependence on external AI services will need a different inquiry from a sector body examining barriers to deployment. They may share institution and application records but should not assume that one set of additional questions serves both purposes. A useful commissioning brief identifies the intended decision, the population involved and the uncertainty that prevents a well-supported response. This provides a basis for deciding which information earns its place in the collection.

The brief should also recognise the starting position. If little is known about the distribution of use, a broad mapping exercise may be warranted. If the distribution is already reasonably understood but a consequential dependency remains unclear, another general adoption survey may add little. A focused inquiry can be the more informative next step. Monitoring becomes cumulative when each collection addresses what the existing evidence leaves unresolved.

4.1 Combining breadth with selective depth

A consistent baseline can establish where AI is used and provide the records needed to connect later inquiries. It should distinguish institutional boundaries, actual deployment from plans, the activity supported and the stage of use. The appropriate detail will depend on the population and available records. The aim is a foundation that respondents can maintain and users can interpret, rather than a miniature version of every possible specialist investigation.

Existing inventories, procurement records and oversight processes can supply part of this baseline. The UAE examples in Section 2.2 illustrate that opportunity. An institution can start by matching an inventory entry to its operational owner and supplier record, then linking its activity, permissions and review evidence under a stable deployment reference. This makes it easier to see what is already known and what needs follow-up. Reuse still requires work: definitions may differ, embedded applications may be missing and teams may hold incompatible records. The cost of establishing and maintaining those links should be included in the design.

Deeper inquiries can then follow particular purposes. A resilience inquiry may examine the services dependent on an external component and the feasibility of fallback. An inquiry into customer outcomes may examine selected interactions and how mistakes are corrected. An adoption-support initiative may investigate why pilots fail to progress. The baseline helps identify relevant institutions and deployments; the focused work produces the evidence needed to interpret them.

Selection must be visible in the findings. Examining a group of high-exposure firms can be appropriate for investigating a vulnerability, but its results should not be reported as the average experience of all institutions. Equally, selecting successful adopters can help understand implementation practices without demonstrating that other institutions would achieve the same results. The purpose of selection and the claim made from it have to remain aligned. A small exploratory element is also valuable: open questions about unexpected uses, discontinued applications or newly observed dependencies can reveal developments outside the programme's current priorities. This gives monitoring a way to revise what it considers important, rather than investigating only what it already recognises.

This arrangement supports proportionality on both sides of the collection. Respondents are not asked to document every minor experiment as if it were a critical service. Monitoring teams can devote more attention to evidence that could change a consequential judgement. Proportionality should not, however, become a reason to exclude small institutions automatically. Their experience may be central to a question about access, capability or common service providers. Importance is determined by the question, not solely by organisational size.

Shared definitions can reduce unnecessary differences without eliminating legitimate ones. Institutions, deployments and providers need sufficiently clear identifiers and boundaries for information to be linked. Terms such as production, planned use and discontinuation need practical explanations. Specialist questions can remain tailored to the responsibility concerned. The resulting arrangement is a common information foundation with targeted inquiries, rather

© 2026 FinEthica Advisory FZCO. All rights reserved.

than a universal questionnaire that tries to anticipate every decision.

4.2 Preserving meaning through change

The experience of successive surveys, discussed earlier, shows why continuity cannot be assumed from repeated wording. A programme needs to retain definitions, reporting boundaries and observation dates, and explain changes that affect comparison. Apparent growth may partly reflect a newly recognised embedded feature, a revised group boundary or the discovery of uses that were previously missing. Where classifications are updated, bridge questions or a documented break in the series can preserve the meaning of the findings.

Changes inside deployments deserve equal attention. Returning to the assistant, the monitoring record should recognise when it moves into a different activity, reaches a much larger population, gains authority to act or becomes essential to service delivery. A vendor update can matter, but so can a change in local configuration, data access or workflow. Focusing only on whether the model has been replaced would leave important changes unobserved.

The programme should define which changes warrant reassessment or an interim update. Expanded permissions, access to a new category of data or a change in a critical provider are possible triggers. Their significance depends on the deployment and the applicable framework; they should not be declared legally material simply because they appear in a monitoring template. Record the change, its date and the review decision when it occurs. Those details are easier to establish at the time than to reconstruct for an annual return.

Periodic collection is useful for a stable picture of the population. Material changes may require an update between rounds, especially where they alter an earlier assessment. A feasible programme can use existing channels for significant changes and incidents where those channels fit the purpose. The frequency and triggers should reflect the pace of change and the consequences of outdated evidence. This is a design recommendation, not a statement that every jurisdiction already imposes such a reporting duty.

Timeliness also includes the interval between observation and use. Data collected promptly can still be unhelpful if validation and reporting take too long for the decision concerned. A monitoring team should be able to tell users when the underlying activity was observed, when the evidence was last checked and what may have changed since. The distinction is particularly important when a newly published report contains observations from an earlier deployment environment.

4.3 Making evidence credible enough to use

A richer collection does not automatically produce stronger knowledge. Detailed responses can contain inconsistent categories, uncertain estimates and claims that have not been tested. More fields may also encourage hurried completion or false precision. Evidence quality depends on how information is produced and checked, as well as on what the questionnaire asks.

The first opportunity to improve quality is before collection. Testing questions with a small, varied group of respondents can reveal terms that are interpreted differently and requests that cannot be answered from available records. A respondent may know that an application is widely used without having a reliable measure of the share of activities it influences. Allowing an explicit estimate or an “unknown” response is more informative than forcing an exact number that has no defensible basis.

Unknowns should then influence interpretation. If upstream providers are unidentified for a substantial part of the relevant population, the resulting map cannot be read as a complete account of dependence. If institutions cannot distinguish pilots from production uses, a maturity result needs qualification. Missing information is itself a finding about the limits of visibility. It should not be silently converted into non-use, absence of exposure or evidence that a safeguard exists.

Checking should be proportionate to the claim. Basic validation can identify conflicting dates, duplicate records or a production deployment with no named activity. Selected follow-up can examine whether a consequential self-report corresponds to operational evidence. For the assistant, an assertion that staff review responses may warrant examining a sample of reviewed outputs and the handling of identified errors. The appropriate check depends on the decision and

© 2026 FinEthica Advisory FZCO. All rights reserved.

available resources; a sector survey need not become an exhaustive audit of every deployment.

Other sources can strengthen the inquiry if they address the same uncertainty. Inventories may clarify the deployment boundary. Incident records may show problems not captured by a questionnaire. Interviews may explain why an application was discontinued. Public information can reveal a development worth asking about. Sources that disagree should prompt an examination of dates, definitions and coverage. Their disagreement should not be removed by averaging quantities that represent different phenomena.

The evidence should also retain its status. A reported benefit, an independently checked outcome and a hypothesis about future impact have different strengths. A policy document demonstrates that an expectation has been recorded; it does not demonstrate that the expectation is implemented in every relevant workflow. Keeping these distinctions visible allows users to decide whether the evidence supports action or whether further inquiry is needed.

4.4 Organising responsibility for the evidence

The practical constraints on monitoring are organisational as well as technical. Information may be distributed across business teams, procurement, risk functions and service providers. Without a clear internal owner, an institution may return an answer assembled from incompatible records. Without a clear owner at the collecting body, an unresolved response may pass into a report without follow-up. Responsibility is necessary at both ends of the process.

Coordination can help when several bodies need related information. Reusing a clearly defined deployment record may reduce duplicate requests and make findings easier to connect. It does not require all bodies to share every confidential detail or surrender their distinct judgements. Decisions about access, retention and publication should follow the purpose of the information and applicable constraints. A public report can communicate patterns and limitations while more detailed material supports authorised inquiry.

Confidentiality matters to the quality of participation. Institutions need to understand how sensitive information about systems, commercial arrangements or incidents will be handled. Collecting identifiable customer information is often unnecessary for a sector monitoring purpose that can be served through appropriately aggregated evidence. The objective is to obtain information sufficient for the question, with a clear justification for sensitive detail, rather than to assemble an unrestricted repository.

Monitoring also depends on analytical capacity. A collecting body needs people able to resolve ambiguous responses, understand the activity described and judge the limits of a conclusion. Expanding the questionnaire without expanding this capacity can leave the organisation with more material than it can interpret. The resource decision should therefore cover the work after collection, including engagement with users of the findings.

These arrangements turn a collection into a continuing evidence base. Yet even well-maintained evidence can remain unused if the programme ends when a report is issued. The final design question is what the evidence changes in practice: who receives a finding, what response it can support and how that response informs the next observation.

05 From monitoring findings to action

The purpose of stronger evidence is to improve judgement. A monitoring result may support an immediate operational response, justify further inquiry, reveal a need for capability support or show that an earlier concern has diminished. It may also establish that the available information is insufficient to choose among those responses. The programme is useful when that distinction becomes clearer.

Consider how a collecting authority could use the information illustrated by the fictional assistant. Its

© 2026 FinEthica Advisory FZCO. All rights reserved.

baseline might show little change in the share of institutions using AI, while linked deployment records reveal that applications are moving from internal tasks into customer interactions. Records of permissions and activity would identify where the transition is most consequential. The authority would then have a reason to select particular uses for closer inquiry, even though the headline adoption rate gives no such signal.

A focused inquiry could distinguish several situations within that group. Some institutions may have examined service quality and total operating costs, with evidence that review arrangements remain effective. Others may report benefits without measuring correction and rework, leaving the value claim uncertain. A further group may have expanded permissions without reassessing the application. These differences would support different next steps: learning from documented implementation experience, seeking better outcome evidence or examining whether safeguards still fit the activity. They would not justify treating all customer-facing AI uses as one category of concern.

The same information could reveal a practical need for support. If smaller institutions repeatedly struggle to evaluate customer-facing applications, a sector body might investigate whether shared training or clearer assessment guidance would address the constraint. That decision would require engagement with the affected institutions, including those that chose not to deploy. The evidence would give the engagement a specific purpose, rather than presuming that low adoption always calls for encouragement.

Across the population, linked provider information could identify material services that depend on a common component. A financial authority could then examine recovery arrangements in the activities most affected. The fact of shared dependence gives the inquiry a reason, while evidence about criticality, fallback and switching helps determine its significance. A supplier count alone would not support the same judgement. Nor would a shared component justify an automatic conclusion that the sector is unsafe.

For the institutions themselves, participation can also improve the basis for management decisions. A clearer account of outcomes, costs and changing permissions can support a decision to expand, redesign or discontinue an application. The sector exercise and internal management process benefit from related evidence, while retaining their different responsibilities.

Consider a fictional Chief Compliance Officer (CCO) at a mid-sized UAE bank answering a supervisory questionnaire. A deployment record distinguishes a transaction-monitoring model that ranks alerts but cannot block transactions from an onboarding tool that can reject applications below a confidence threshold. It also identifies a complaints assistant whose replies require staff approval, with logs showing a median review time of eleven seconds. The same record helps the CCO prepare the response and bring specific questions to the next risk committee. Eleven seconds does not establish inadequate oversight: the task, information available and quality of completed reviews need examination. The onboarding tool's rejection authority prompts a check of its approval, oversight and customer recourse. The value lies in the questions the record makes possible and the follow-up assigned to them.

This progression from observation to interpretation should be visible in reporting. Readers need to know what was found, how firmly it is established, why it matters for the stated purpose and what remains uncertain. A report can recommend follow-up without presenting a provisional concern as a demonstrated failure. It can describe a promising result without implying that every institution should reproduce it. Making these distinctions explicit improves the usefulness of the evidence to readers with different responsibilities.

5.1 Giving findings a practical destination

Before collection begins, the commissioning body should identify who will review the findings and what decisions they may inform. For a general sector exercise, the destination may be a programme of engagement or a choice of topics for deeper examination. For a focused inquiry, it may be a decision about remediation, support or further evidence. The decision process need not be automatic, but it should exist. Responsibility for collecting evidence and authority to respond may sit with different teams. Their handover should make the finding, its limitations and the unresolved question clear.

© 2026 FinEthica Advisory FZCO. All rights reserved.

Feedback from users and respondents should influence the next cycle. If a reported indicator repeatedly triggers inquiries that find only a definitional inconsistency, the indicator needs revision. If a targeted investigation identifies a previously unrecognised form of dependence, the baseline may need to capture enough information to locate it elsewhere. If a question has no identifiable use and imposes substantial effort, its continued inclusion should be reconsidered. Monitoring must be capable of learning about its own performance.

Useful signs of that performance include the time taken to resolve important information gaps, the currency of evidence for consequential uses and examples of decisions that became better informed. Reporting burden and recurrent missing information also matter. None of these measures proves that the programme caused better financial outcomes. They help assess whether it is doing the work needed to support those outcomes.

06 Conclusion

An adoption rate establishes how widely AI is present within an observed population. Decisions about innovation, governance and supervision also require evidence about the financial activities it influences, the authority it holds, the dependencies it creates and the outcomes associated with its use. Monitoring should preserve these connections as deployments change.

The international exercises reviewed here provide a substantial foundation. Their different mandates, populations and definitions explain much of the variation in published evidence. The opportunity is to build on that work through clearly defined records, proportionate follow-up and timely updates. Public reporting should make the basis and limitations of findings visible without claiming to reveal the full supervisory information available to authorities.

The contribution of a governance perspective is practical. A useful record distinguishes a policy from evidence of its implementation, a completed review from an unresolved finding, and a planned correction from a verified response. A firm can begin before a new questionnaire arrives: select one important live deployment, connect its existing records and take the first unresolved question to the responsible owner. The same evidence can support internal oversight and sector inquiry where purposes and definitions align. Its reuse requires coordination, resources and attention to confidentiality.

Monitoring can identify emerging problems, inform action and preserve evidence for further analysis. It does not, on its own, establish that AI caused an observed improvement or that a particular intervention will produce the desired result. The second paper in this series examines the distinction between prediction, model explanation and causal explanation, and what those distinctions mean for decisions. For this paper, the immediate task is to define a useful monitoring purpose, establish maintainable records and assign responsibility for interpreting and responding to the findings.

Appendix Designing a monitoring exercise

This appendix can support a sector monitoring exercise or an institution's first review of its own AI use. It is an adaptable design aid, not a validated universal framework or a statement of regulatory duties. Table 3 sets out the connected records; the prompts help test their usefulness. An institutional team can begin with the final subsection, "Applying the approach within an institution", and use Table 3 to structure its first record.

© 2026 FinEthica Advisory FZCO. All rights reserved.

Establish the purpose before selecting fields

Record the decision or responsibility the exercise will inform, the population concerned and the uncertainty to be reduced. Identify the evidence already available and explain what a new collection is expected to add. Name the intended users of the findings and the person responsible for resolving unclear or incomplete responses.

A practical brief should also state the observation period, reporting unit, permitted uses of confidential information and intended publication approach. Define whether the exercise is a broad baseline or a selected inquiry. Explain the basis for selection and how it limits the conclusions that can be drawn.

Choose a manageable starting point

The first implementation cycle should be deliberately manageable. Use a clearly bounded population, test the core definitions and establish how records will be checked. Select one focused inquiry that addresses a consequential uncertainty. This gives the programme an opportunity to learn whether respondents can provide the information and whether users find it useful. Expanding after that experience is more defensible than requiring a comprehensive dataset before any evidence can be used.

A programme with an established survey can begin differently. It can review which findings have actually informed decisions, which questions repeatedly produce ambiguity and which relevant developments arrive too late to be captured. Some improvements may require a new field. Others may require clearer definitions, better follow-up or a change in reporting frequency. A larger collection should be an answer to an identified need, rather than the default sign of progress.

Connect the minimum records

Table 3 A starting point for connected monitoring records

Record	Information to define	Why the connection matters
Institution	Identifier, group or legal-entity boundary, subsector, size measure and reporting contact	Establishes the population and supports interpretation across institutions
Deployment	Stable reference, activity, technology description, stage, start or retirement date and institution link	Distinguishes the changing role of an application from adoption status
Activity and authority	Relevant activity volume, affected population, actions permitted and review points	Shows reach and influence within a defined workflow
Dependence	Relevant providers or components, activities affected and fallback evidence	Connects supplier relationships to operational consequences
Oversight and response	Responsible owner, review date and scope, significant open issues, action owner and resolution status	Links governance arrangements to review and follow-up for the deployment
Evidence and change	Observation date, source, measured or estimated status, validation status, unknowns and significant changes	Makes the currency, basis and limitations of an assessment visible

Note: Select detailed outcome, assurance and dependency fields for the inquiry. Do not interpret missing information as zero exposure. One deployment can involve several models or providers, and one model can support several deployments. Oversight records should distinguish a completed review from verified remediation.

© 2026 FinEthica Advisory FZCO. All rights reserved.

Test questions against the intended decision

The following prompts illustrate how a question can seek interpretable evidence rather than an unqualified assertion. They are not a complete questionnaire.

- Which activities used AI during the reference period, and which reported uses were experiments, production deployments or discontinued applications? Record planned use separately.
- For the deployments selected for closer inquiry, what task did AI perform, what actions could it take and what required human review? Describe actual arrangements rather than a generic policy.
- What share or volume of the relevant activity involved the deployment? State the denominator, period and whether the answer is measured or estimated.
- What has changed since the last assessment in data access, users, permissions, provider arrangements or the activity supported? Identify changes whose consequences have not yet been assessed.
- Which expected benefits have been examined in practice, using what measures and comparisons? Include material review, correction and operating costs where relevant.
- What adverse outcomes or near misses were identified, how were they detected and what remedy followed? Explain the exposure and detection limits of the reported counts.
- For selected deployments, when and how was oversight last reviewed, what significant findings remain open and who owns the response? Where an issue is recorded as resolved, identify the evidence supporting closure.
- Which obstacles prevented a planned or experimental use from progressing? Distinguish a capability constraint from a decision that the application was unsuitable.

Publish the basis of the finding

Accompany important results with the population boundary, observation period, unit counted, definition and denominator. State participation and relevant missingness. Distinguish reported information from checked evidence and explain selection or changes that affect comparison. Identify what the result can support and which question needs further inquiry.

For example, “the share of responding institutions with any production use” and “the share of customer contacts supported by a selected deployment” are different measures. Either can be useful. Their labels and supporting information must preserve that difference.

Applying the approach within an institution

Start with an important live deployment in one activity, such as transaction monitoring or onboarding. Bring together its business owner, compliance or risk reviewer and supplier manager. Use Table 3 and the steps below to connect existing information, identify unknowns and assign follow-up actions. Test whether the record helps a real decision before expanding coverage. This initial review does not replace applicable requirements or a full assessment.

- Identify AI uses within the chosen boundary, including relevant features embedded in vendor products. Distinguish experiments, production deployments, planned uses and discontinued applications. Record known gaps in coverage.
- For each selected production deployment, record its task, permitted actions and human review arrangements. Examine the adequacy of review using evidence appropriate to the task; an approval label or elapsed time alone is insufficient.
- Record the activity supported and its relevant volume or share, stating the denominator, period and whether the figure is measured or estimated.
- Identify relevant providers and upstream components where known. For important activities, record recovery arrangements, when they were tested and any unresolved limitations.
- Record changes in data access, users, permissions, providers or workflow, together with the decision on whether reassessment is needed.
- Distinguish measured benefits from expectations. Record significant adverse outcomes, relevant control findings, corrective-action owners and the evidence used to close issues.
- Give each deployment a stable reference, date the evidence and assign an owner. Give important unknowns and unresolved findings a follow-up

© 2026 FinEthica Advisory FZCO. All rights reserved.

owner and review date. Record management's response. Reuse the evidence for external inquiries where definitions and reporting boundaries align.

References

Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., … Roberts, K. (2024). *Artificial intelligence risk management framework: Generative artificial intelligence profile* (NIST AI 600-1). Gaithersburg, MD: National Institute of Standards and Technology. doi:10.6028/NIST.AI.600-1

Bank of England, & Financial Conduct Authority. (2019). *Machine learning in UK financial services*. Retrieved from https://www.bankofengland.co.uk/report/2019/machine-learning-in-uk-financial-services

Bank of England, & Financial Conduct Authority. (2022). *Machine learning in UK financial services*. Retrieved from https://www.bankofengland.co.uk/report/2022/machine-learning-in-uk-financial-services

Bank of England, & Financial Conduct Authority. (2024). *Artificial intelligence in UK financial services – 2024*. Retrieved from https://www.bankofengland.co.uk/report/2024/artificial-intelligence-in-uk-financial-services-2024

Central Bank of the United Arab Emirates. (2026). *Guidance note on the consumer protection and responsible adoption and use of artificial intelligence and machine learning by licensed financial institutions in the U.A.E.* Retrieved from https://rulebook.centralbank.ae/en/entiresection/6960

Dubai Financial Services Authority. (2025). *DFSA artificial intelligence survey 2025: An overview of regulatory insights and trends in financial services in the DIFC*. Retrieved from https://www.dfsa.ae/download_file/4087/0

Dubai Financial Services Authority. (2026, June 4). *DFSA regulatory expectations on artificial intelligence risk management in the DIFC* [Dear SEO letter]. Retrieved from https://www.dfsa.ae/your-resources/publications-reports/seo-letters-1

Dubai International Financial Centre. (2023). *Regulation 10: Processing personal data through autonomous and semi-autonomous systems*. Retrieved from https://www.difc.com/business/registrars-and-commissioners/commissioner-of-data-protection/regulation-10

European Banking Authority. (2025). *Rising application of AI in EU banking and payments sector* [Fact sheet]. Retrieved from https://www.eba.europa.eu/sites/default/files/2025-09/146b3558-d026-47bf-a872-f05e93ed30d2/Rising%20application%20of%20AI%20in%20EU%20banking%20and%20payments%20sector.pdf

European Insurance and Occupational Pensions Authority. (2024). *Report on the digitalisation of the European insurance sector*. Retrieved from https://www.eiopa.europa.eu/publications/eiopas-report-digitalisation-european-insurance-sector_en

European Insurance and Occupational Pensions Authority. (2026). *Generative AI market survey: Outlook, use cases and risk management*. Retrieved from https://www.eiopa.europa.eu/publications/generative-ai-market-survey-outlook-use-cases-and-risk-management_en

Financial Action Task Force. (2021). *Opportunities and challenges of new technologies for AML/CFT*. Retrieved from https://www.fatf-gafi.org/en/publications/Digitaltransformation/Opportunities-challenges-new-technologies-for-aml-cft.html

Financial Stability Board. (2017). *Artificial intelligence and machine learning in financial services: Market developments and financial stability implications*. Retrieved from https://www.fsb.org/2017/11/artificial-intelligence-and-machine-learning-in-financial-service/

Financial Stability Board. (2024). *The financial stability implications of artificial intelligence*. Retrieved from https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/

Financial Stability Board. (2025). *Monitoring adoption of artificial intelligence and related vulnerabilities in the financial sector*. Retrieved from https://www.fsb.org/2025/10/monitoring-adoption-of-artificial-intelligence-and-related-vulnerabilities-in-the-financial-sector/

Gambacorta, L., & Shreeti, V. (2025). *The AI supply chain* (BIS Papers No. 154). Basel, Switzerland: Bank for International Settlements. Retrieved from https://www.bis.org/publications/paper-154-ai-supply-chain

International Monetary Fund. (2024). *Global financial stability report: Steadying the course—Uncertainty, artificial intelligence, and financial stability*. Retrieved

© 2026 FinEthica Advisory FZCO. All rights reserved.

from https://www.imf.org/en/Publications/GFSR/Issues/2024/10/22/global-financial-stability-report-october-2024

New York State Department of Financial Services. (2021). *Report on Apple Card investigation*. Retrieved from https://www.dfs.ny.gov/system/files/documents/2021/03/rpt_202103_apple_card_investigation.pdf

Office of the Superintendent of Financial Institutions, & Financial Consumer Agency of Canada. (2024). *OSFI–FCAC risk report: AI uses and risks at federally regulated financial institutions*. Retrieved from https://www.osfi-bsif.gc.ca/en/about-osfi/reports-publications/osfi-fcac-risk-report-ai-uses-risks-federally-regulated-financial-institutions

Organisation for Economic Co-operation and Development. (2019). *Governance as an SDG accelerator: Country experiences and tools*. Paris, France: OECD Publishing. doi:10.1787/0666b085-en

Organisation for Economic Co-operation and Development. (2023). *Generative artificial intelligence in finance* (OECD Artificial Intelligence Papers No. 9). Paris, France: OECD Publishing. doi:10.1787/ac7149cc-en

Organisation for Economic Co-operation and Development. (2026). *Supervision of artificial intelligence in finance: Challenges, policies and practices* (OECD Artificial Intelligence Papers No. 54). Paris, France: OECD Publishing. doi:10.1787/92743dc1-en

Swiss Financial Market Supervisory Authority. (2025, April 24). *FINMA survey: Artificial intelligence gaining traction at Swiss financial institutions* [Press release]. Retrieved from https://www.finma.ch/en/news/2025/04/20250424-mm-umfrage-ki/

© 2026 FinEthica Advisory FZCO. All rights reserved.

FinEthica
ADVISORY

How FinEthica can help

FinEthica Advisory works with banks, fintechs, virtual asset firms and industry bodies in the UAE on licensing, financial crime compliance and governance. We can help institutions, industry associations and financial authorities develop and use evidence about AI adoption by:

- reviewing existing information on AI use and identifying gaps relevant to a defined decision;
- designing or improving inventories and deployment records, with attention to applicable regulatory requirements and supervisory guidance;
- developing surveys and focused inquiries with clear definitions, proportionate reporting and a practical plan for analysis;
- briefing boards and senior management on interpreting monitoring findings, testing governance claims and prioritising follow-up;
- developing empirical studies and applying machine learning or causal analysis, where appropriate, to investigate AI adoption patterns and evaluate outcomes.

FinEthica also welcomes research and analytical collaborations with financial institutions, industry bodies and academic partners. Its academic advisory expertise supports the design and interpretation of evidence, from framing a research question to interpreting results and their limits.

For advisory support or a research collaboration, contact FinEthica to discuss your objectives and the evidence you need.

CONTACT

Ilham Tamimi

Founder and Managing Director

ilham.tamimi@aefinethica.com

+971 50 421 5056

aefinethica.com

FinEthica Advisory FZCO · Dubai Silicon Oasis · Dubai, United Arab Emirates

© 2026 FinEthica Advisory FZCO. All rights reserved. | aefinethica.com