AI ADOPTION TRENDS / FIELD REPORT 03

Financial Institutions AI 2026

From controlled access to accountable action

A decision-grade field study for bank, credit union, insurer, payments, wealth, risk, compliance, operations, and technology leaders.

18 pages / evidence, cases, economics, controls, worksheets

Research cutoff: 18 September 2026

aiadoptiontrends.com

THE DECISION MEMO

The winning institution will not be the one with the most copilots. It will be the one that can prove which system informed a decision, who reviewed it, and how the customer can challenge it.

75%

of 118 UK financial firms surveyed in 2024 said they already used AI. [1]

33%

of reported AI use cases were third-party implementations, up from 17% in 2022. [1]

2%

of reported use cases used fully autonomous decision-making. [1]

- **Start inside the control perimeter.**

  Grounded policy search, meeting notes, complaint triage, and analyst drafting have reviewers and systems of record.

- **Treat the explanation as an output.**

  For credit and other adverse decisions, a score is not enough; the institution must produce a specific, accurate reason.

- **Govern the dependency, not only the model.**

  Cloud, data, and model concentration can turn a vendor update into simultaneous operational and conduct risk.

PRIORITY DECISION

Fund a 90-day portfolio with one employee-knowledge workflow, one reviewed customer-service workflow, and one risk use case in shadow mode. Require provenance, reviewer identity, decision logging, customer redress, and rollback before granting write or execution authority.

PRINCIPAL LIMITATION

The strongest adoption dataset here is a UK regulatory survey. Named U.S. cases are company-reported deployments, not controlled estimates of customer outcomes or net economic value.

AI USE IS BROAD; ACCOUNTABLE AUTONOMY REMAINS NARROW

The best public survey separates adoption, materiality, third-party exposure, and decision authority. Those measures should not be collapsed into one maturity score.

118

FIRMS

Bank of England and FCA survey respondents across banking, insurance, lending, markets, and infrastructure.

17%

FOUNDATION MODELS

Share of reported AI use cases; most AI in finance is still broader ML, not generative AI.

16%

HIGH MATERIALITY

Firm-rated share of all reported AI use cases; a self-assessment, not a regulator judgment.

84%

ACCOUNTABLE PERSON

Share of AI-using firms reporting a person accountable for the AI framework.

SHARE OF RESPONDENTS USING AI IN SELECTED USE CASES [1]

READ THIS CAREFULLY

These are different measures from one survey: the first three are shares of respondent firms; the final bar is a share of use cases. Use them to locate activity, not to rank business value.

SIX MARKETS HIDING INSIDE FINANCIAL-INSTITUTION AI

Each lane has a different customer consequence, evidentiary burden, and authority boundary.

01 Employee knowledge

VALUE Faster policy and research access

EVIDENCE Citation accuracy; task time

FACEPLANT Stale rule; invented source

02 Service and complaints

VALUE Resolution speed and consistency

EVIDENCE QA; reopen rate; redress

FACEPLANT Unsupported advice; lost complaint

03 Fraud and AML

VALUE Loss avoided; investigation yield

EVIDENCE Precision, recall, queue capacity

FACEPLANT False positives swamp analysts

04 Credit and pricing

VALUE Risk selection; access; speed

EVIDENCE Outcome, fairness, explanation

FACEPLANT Wrong reason; disparate harm

05 Markets and wealth

VALUE Research speed; advisor capacity

EVIDENCE Grounding; suitability review

FACEPLANT Hallucinated fact becomes advice

06 Operations and controls

VALUE Reconciliation; testing; resilience

EVIDENCE Exceptions; recovery; audit trail

FACEPLANT Silent write-back; hidden model

PORTFOLIO RULE

Sequence from retrieve to draft to recommend to execute. Do not grant a higher authority level until the lower level has stable provenance, measured reviewer behavior, capacity for exceptions, and a tested customer-remediation path.

THE AUTHORITY LADDER IS THE CONTROL SURFACE

Model capability is not permission. Every workflow should declare the highest action the system may take without a person.

1

RETRIEVE

SOURCE REQUIRED

Search approved policy, product, research, or customer records. Return citations and access metadata.

2

DRAFT

NAMED REVIEWER

Prepare a memo, meeting note, response, or case summary. A person edits and owns the artifact.

3

RECOMMEND

DECISION EVIDENCE

Rank alerts or propose an action. Preserve inputs, model version, reason, overrides, and outcome.

4

EXECUTE

PRE-AUTHORIZED BOUNDS

Write, message, block, price, or move money only inside narrow limits with monitoring and immediate reversal.

NON-NEGOTIABLE

A customer-facing explanation, adverse-action reason, fraud block, or transaction is not a chat response. It is a regulated business action with an owner, record, appeal path, and retention requirement.

MORGAN STANLEY: RETRIEVAL BEFORE RECOMMENDATION

The wealth-management rollout is a useful pattern because it began with access to approved internal intellectual capital, then added meeting capture with client consent.

WHAT CHANGED

Morgan Stanley fully rolled out an internal advisor assistant in 2023 and added Debrief in 2024 to create meeting notes, action items, a draft email, and a Salesforce note for advisor review.

MEASURED RESULT

Morgan Stanley reported that 98% of financial-advisor teams adopted the Assistant. That is an adoption measure, not a controlled productivity or client-outcome estimate.

EVIDENCE LIMIT

The evidence is a company announcement. It does not disclose error rates, edit distance, complaint effects, or a counterfactual group.

TRANSFERABLE LESSON

A high-consequence institution can start with bounded retrieval and reviewed artifacts, keeping the advisor between generation and advice.

WHAT TO REPLICATE

- Ground answers in approved material and show the source.
- Require client consent for meeting capture and advisor review before CRM write-back.
- Measure unanswered queries, citation failures, edits, and downstream corrections.

Source basis: Morgan Stanley press release, 26 June 2024. [4]

JPMORGANCHASE: A CONTROLLED GENERAL-PURPOSE LAYER

A secure enterprise shell can reduce shadow use, but access at scale is only the beginning of evidence.

WHAT CHANGED

JPMorganChase launched LLM Suite in 2024 and reported access for more than 200,000 colleagues, with a controlled environment intended to protect customer and company data.

MEASURED RESULT

The firm later reported more than 500 AI use cases in production and described cost avoidance, risk reduction, and revenue growth, without publishing a portfolio-wide causal estimate.

EVIDENCE LIMIT

User counts and use-case counts do not show sustained use, realized value, or the distribution of failures. The claims are company-reported.

TRANSFERABLE LESSON

Central access can create a control point for data, vendors, and reusable services; value still has to be measured workflow by workflow.

WHAT TO REPLICATE

- Make the approved route easier than consumer tools.
- Separate active users, embedded workflows, and production actions.
- Require each use case to carry an owner, materiality rating, and retirement rule.

Source basis: JPMorganChase 2024 shareholder materials and 2025 technology update. [5][6]

BANK OF AMERICA ERICA: SCALE THROUGH A BOUNDED RESPONSE LIBRARY

Erica illustrates long-run operational tuning, not a single foundation-model launch.

WHAT CHANGED

Bank of America says Erica has been trained to recognize millions of questions using a library of more than 700 responses and has received more than 75,000 updates since launch.

MEASURED RESULT

The bank reported more than 3 billion interactions, nearly 50 million users, and more than 98% of users finding the information they needed as of August 2025.

EVIDENCE LIMIT

These are company-reported engagement and service measures. They do not reveal independent accuracy, complaint, or customer-financial-outcome analysis.

TRANSFERABLE LESSON

Scale came from bounded intents, a maintained response system, and continuous operations - not from letting a model improvise every answer.

WHAT TO REPLICATE

- Maintain an approved answer catalog and explicit unsupported-intent route.
- Track containment separately from repeat contacts, complaints, and corrections.
- Use handoff as a designed outcome, not a model failure to hide.

Source basis: Bank of America newsroom, 20 August 2025. [7]

DO NOT LET A BENCHMARK ANSWER THE WRONG QUESTION

Finance teams need a chain from model behavior to workflow performance to customer and prudential outcomes.

Evidence	What it proves	What it cannot prove	Decision use
Offline model test	Performance on frozen, labeled cases	Queue effects or user reliance	Entry gate only
Shadow deployment	Behavior on current live data	Impact after users act	Threshold and routing design
Controlled rollout	Incremental workflow effect	Long-run drift without monitoring	Scale decision
Outcome monitoring	Losses, approvals, complaints, corrections	Causality without a comparator	Ongoing control
Customer challenge	Whether reasons and remediation work	Population fairness alone	Conduct and redress

EVIDENCE RULE

A model can improve AUC while worsening the business. Always pair performance with queue capacity, override behavior, specific reasons, customer outcomes, and the cost of corrections.

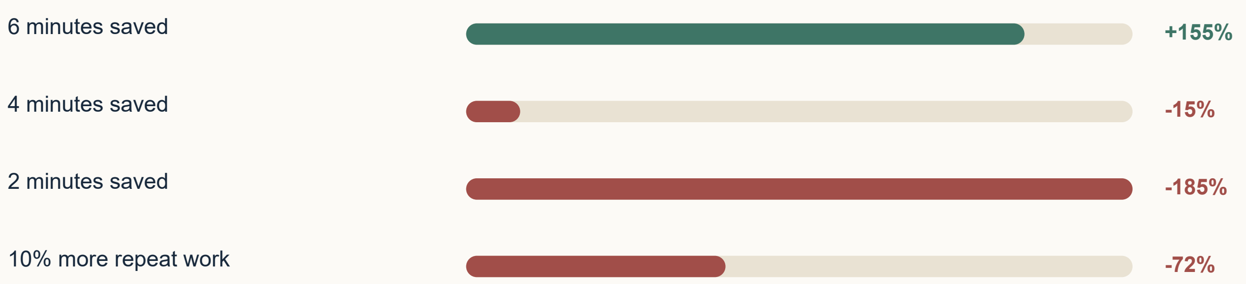
THE BUSINESS CASE MUST PRICE REVIEW AND FALSE DECISIONS

This example tests a complaint-triage assistant. It is not a vendor ROI claim.

ILLUSTRATIVE BASE CASE - REPLACE EVERY INPUT

Annual eligible complaints	48,000	**volume**
Minutes saved before review	4.0	**gross benefit**
Loaded reviewer cost	$62/hour	**assumption**
Added QA and model operations	$145,000	**annual cost**
Incremental rework / redress reserve	$90,000	**risk cost**
Illustrative annual net	$-36,600	**base case**

SENSITIVITY



DO NOT COUNT

Do not count a faster first draft as savings if reviewers, complaint teams, call centers, or customers absorb more work. Price the displaced queue and the tail, not just the happy path.

THE VENDOR EVIDENCE ROOM SHOULD SURVIVE AN EXAMINER

Ask for artifacts, not adjectives. A SOC report and a model card do not replace workflow-specific evidence.

1 Intended use

Exact users, inputs, outputs, prohibited uses, authority level, and affected customer populations.

2 Data and rights

Training and evaluation sources, retention, cross-border handling, customer data use, deletion, and subprocessors.

3 Performance

Version-specific tests, subgroup results, calibration, explanation fidelity, and realistic failure examples.

4 Operations

Latency, uptime, rate limits, support, monitoring, incident response, rollback, and business continuity.

5 Change control

Notice periods, regression evidence, approval rights, version pinning, and the right to reject an update.

6 Exit

Data export, deletion evidence, replacement support, record retention, and a tested manual workflow.

WALK-AWAY CONDITION

Walk away if the institution cannot preserve the exact version, input, output, explanation, reviewer, and resulting action for a high-consequence decision - or if the vendor cannot support customer remediation.

A 90-DAY PILOT THAT EARNS AUTHORITY

The pilot should make uncertainty visible and keep the incumbent workflow available until the new one proves itself.

	Phase	Timing	Description
1	**BASELINE**	Weeks 0-2	Sample current cycle time, quality, overrides, complaints, losses, queue age, and customer segments.
2	**BOUND**	Week 2	Declare permitted data, users, outputs, authority, excluded populations, and escalation routes.
3	**SHADOW**	Weeks 3-5	Run on live cases without changing outcomes. Compare recommendations, reasons, and queue load.
4	**REVIEWED**	Weeks 6-9	Expose output to trained users; require recorded review, edits, overrides, and customer-facing source.
5	**ADJUDICATE**	Weekly	Risk, compliance, operations, and the business review incidents and subgroup performance.
6	**DECIDE**	Week 12	Scale, narrow, retrain, or stop by workflow and population. Do not average away harm.

MINIMUM STOP RULES

Pause on inaccurate adverse-action reasons, material subgroup deterioration, unrecoverable write-back, uncontrolled vendor change, data leaving the approved boundary, or exception queues beyond capacity.

MINIMUM SCALE RULES

Require stable quality, specific explanations, manageable exceptions, measurable net benefit after review and redress, and a rehearsed rollback with accountable owners.

THE SCORECARD MUST BALANCE VALUE, CONDUCT, AND RESILIENCE

A single productivity KPI will systematically hide the most expensive failure modes.

Reviewed-output quality

OWNER Business QA

GATE Non-inferior to baseline by case type

Specific reason fidelity

OWNER Compliance

GATE Accurately reflects factors used

Queue displacement

OWNER Operations

GATE No unstaffed exception growth

Customer harm

OWNER Conduct risk

GATE No adverse complaint or redress trend

Subgroup outcomes

OWNER Fair lending / risk

GATE Pre-set thresholds by population

Recovery

OWNER Technology

GATE Rollback and manual route tested

BALANCED RULE

Scale only when no single critical dimension is red. Strong time savings do not compensate for inaccurate reasons, weak customer redress, unstable vendor versions, or a queue the institution cannot work.

THE EXPENSIVE FAILURES ARE OPERATIONAL, NOT THEATRICAL

A fluent answer becomes risk when it meets a record, a decision, or an overloaded queue.

Reason mismatch

The notice lists a plausible factor that was not the principal driver.

RECOVERY Replay decision; correct notice; validate explanation mapping

False-positive flood

Fraud alerts rise faster than analyst capacity and age silently.

RECOVERY Capacity gate; threshold rollback; aged-case review

Advice leakage

A service draft crosses into personalized advice or a product promise.

RECOVERY Intent blocks; approved answers; supervised escalation

Hidden vendor model

A third-party feature changes behavior without model inventory approval.

RECOVERY Feature inventory; notice right; change-control test

Automation bias

Reviewers accept recommendations faster and stop independent checks.

RECOVERY Blind sampling; override review; rotation and retraining

Record fracture

Prompt, answer, approval, and final customer action live in different systems.

RECOVERY Single decision record; retention test; reconciliation

Population drift

Approval or fraud outcomes shift after channel or product changes.

RECOVERY Segment monitoring; recalibration; narrow intended use

Exit failure

The institution cannot reproduce or replace a vendor-dependent workflow.

RECOVERY Export test; manual fallback; replacement runbook

RUN THIS EXERCISE BEFORE A MODEL CAN AFFECT A CUSTOMER

A specific adverse-action reason is wrong for a subset of applicants after a vendor update.

SCENARIO

A lender deploys an updated underwriting model on Friday. On Monday, compliance finds that denial notices for thin-file applicants cite revolving utilization even when the decisive feature was cash-flow volatility. Approval rates also moved for one protected-age proxy segment. No regulator or customer has yet raised the issue.

0-5 min	**Detect**	Which monitoring, complaint, notice, and model-version signals reveal the issue?
5-15 min	**Contain**	Can you stop decisions, preserve records, restore the prior version, and route urgent applications?
15-25 min	**Assess**	Who determines scope, customer harm, fair-lending impact, and notification obligations?
25-35 min	**Correct**	How are notices and decisions repaired without obscuring the record?
35-45 min	**Learn**	What changes to mapping tests, release gates, vendor terms, and sampling?

EVIDENCE TO LEAVE THE ROOM WITH

Named incident commander / affected decisions by version and time / preserved features and reasons / subgroup analysis / corrected notices / redress decision / vendor escalation / restart criteria.

PASS CONDITION

The team can identify every affected decision, stop or reverse the workflow, give customers accurate reasons and remediation, preserve an audit trail, and name who authorizes restart.

SCALE IS A REUSABLE CONTROL PLANE PLUS LOCAL OWNERSHIP

Centralize the hard controls. Localize the workflow, customer context, thresholds, and decision accountability.

01

Approved access

Identity, data boundary, vendor route, logging, and secrets.

02

Evidence services

Evaluation sets, provenance, reason mapping, drift, and red-team tests.

03

Workflow controls

Review queue, exception route, permissions, rollback, and records.

04

Business ownership

Outcome, conduct, staffing, customer redress, and retirement.

SCALE TEST

A second business line should inherit controls and evidence machinery - not the first line's thresholds, customer assumptions, or risk appetite.

WHAT CENTRALIZE / WHAT LOCALIZE

Centralize identity, logging, vendor inventory, evaluation tooling, security, and incident mechanics. Localize intended use, reviewer training, customer language, thresholds, exception capacity, outcome measures, and the final go/no-go decision.

EIGHT QUESTIONS BEFORE APPROVAL

If the committee cannot answer these in plain language, the system is not ready for material authority.

01 **What exact action can the system take, and what can it never take?**

02 **Which customer populations and products are excluded?**

03 **Can every material output be reconstructed by version?**

04 **How are specific reasons generated and verified?**

05 **Where does exception work land, and is the queue staffed?**

06 **What independent evidence supports value and safety?**

07 **What vendor change can occur without our approval?**

08 **How do we stop, repair, notify, and compensate?**

APPROVAL STANDARD

Approve a bounded operating design, not a model brand. The packet should contain the intended-use statement, local evidence, economics after review and redress, authority level, monitoring, named owners, and the tested stop path.

WHAT THIS REPORT CAN AND CANNOT TELL YOU

This field report synthesizes regulatory surveys, regulator guidance, company disclosures, and an illustrative operating model. Company-reported adoption and outcome claims are labeled as such. Cross-source percentages are not combined into a single market estimate.

SOURCE LEDGER

[1] Bank of England and FCA, Artificial intelligence in UK financial services - 2024
www.bankofengland.co.uk/report/2024/artificial-intelligence-in-uk-financial-servi...

[2] CFPB Circular 2022-03, adverse action and complex algorithms
www.consumerfinance.gov/compliance/circulars/circular-2022-03-adverse-action-noti...

[3] CFPB Circular 2023-03, accurate and specific reasons
files.consumerfinance.gov/f/documents/cfpb_adverse_action_notice_circular_2023-09...

[4] Morgan Stanley, launch of AI at Morgan Stanley Debrief
www.morganstanley.com/press-releases/ai-at-morgan-stanley-debrief-launch

[5] JPMorganChase, 2024 COO letter
www.jpmorganchase.com/content/dam/jpmc/jpmorgan-chase-and-co/investor-relations/d...

[6] JPMorganChase, 2025 Innovation Week
www.jpmorganchase.com/about/technology/blog/innovation-week25

[7] Bank of America, Erica surpasses 3 billion interactions
newsroom.bankofamerica.com/content/newsroom/press-releases/2025/08/a-decade-of-ai...

[8] OCC, Semiannual Risk Perspective Fall 2025
occ.treas.gov/publications-and-resources/publications/semiannual-risk-perspective...

[9] NIST, AI Risk Management Framework
www.nist.gov/itl/ai-risk-management-framework

[10] Mastercard, AI and payment fraud prevention
www.mastercard.com/content/dam/mccom/shared/news-and-trends/insights/2025/one-the...