DARK ARTS

WHITE PAPER
NOVEMBER 2026

The Unwatched Machine

Operational AI in Financial Services

Executive Summary

Artificial intelligence has left the pilot stage in financial services. The Bank of England and Financial Conduct Authority survey published in October 2026 found that 86% of the firms surveyed now use AI, 80% have deployed AI agents, and 16% already use agents to automate customer communications.

The controls have not kept pace. Almost every firm has an AI policy, yet only around half monitor how their AI behaves once it is live. More than half of the most consequential AI systems are supplied by third parties, and most firms rely on vendor assurances rather than their own testing to understand them. In April 2026, the Australian Prudential Regulation Authority reported that few of the entities it examined could detect drift or bias in their models.

This paper argues that firms are trusting AI with real decisions about customers faster than they can explain or evidence those decisions. The risk is not that AI occasionally gets something wrong. Every system does. The risk is that a firm cannot tell when it does. Regulators in New Zealand and Australia have now said plainly that boards must know where AI is used and who is accountable for it.

> *The question is no longer whether a firm uses AI. It is whether the firm can show that its AI is working correctly today.*

1. The Operational Gap: AI Has Left the Pilot Stage

AI is no longer confined to innovation teams and proofs of concept. It now sits inside ordinary business processes: credit assessment, claims handling, customer service, fraud screening and complaint triage. The Bank of England and FCA's October 2026 survey of UK financial services shows how far that shift has gone.

86%	80%	16%	53%
Use AI	**Run AI agents**	**Automate customer messages**	**Key systems bought in**
Up from 75% in 2024	Only 2% with broad autonomy	34% once plans go live	High-materiality AI from third parties

> **In plain English.** An AI agent is software that can work through a task and take steps on its own, such as drafting and sending a customer email, rather than simply answering a question.

Governance on paper has kept up. Governance in operation has not. The same survey found that 94% of firms have firm-wide AI policies, principles or standards, but ongoing monitoring, leakage testing and periodic review were each reported by only around half.

UK, October 2026

94% of firms surveyed by the Bank of England and FCA have firm-wide AI policies. Only 39% monitor the impact of their AI on customers.

Australia, April 2026

APRA's letter to industry reported that many entities were still treating AI as "just another technology", with weak post-deployment monitoring and few able to continuously validate their models for drift or bias.

New Zealand regulators have now drawn the same line. On 15 October 2026, the Reserve Bank of New Zealand and the Financial Markets Authority outlined supervisory expectations on the risks associated with frontier AI, naming evolving governance and oversight challenges and growing third-party and concentration risk among their three areas of concern.

> *"Boards and senior management need to know where AI is being used in their business."*
> Louise Unger, FMA

> These expectations build on existing prudential, conduct and operational resilience obligations. They are not a new AI law. They do, however, make clear what supervisors will ask.

2. When AI Changes or Gets Things Wrong

Systems That Change Without a Release

Traditional change control assumes a system behaves the same way until someone changes it. Hosted AI services weaken that assumption. Their behaviour can shift when the provider updates the model, its safeguards or the surrounding service, with no release by the firm that relies on it.

A peer-reviewed study published in PLOS One in February 2026 put the same 240 prompts to three major AI services every week for ten weeks. One stayed stable, one improved and one deteriorated sharply partway through. Nothing changed on the user's side.

Fixing drift brings its own risks. The FCA's October 2026 review describes a firm whose guardrails successfully reduced drift, but were designed so narrowly that the tool gave overly generic results and steered some consumers towards unsuitable options. The technical metric improved. The customer outcome did not.

Few firms are watching closely enough to notice either problem. The Bank of England and FCA survey found:

52%	48%	39%
Monitor for drift	**Always re-test**	**Monitor customer impact**
Model drift, bias or data drift	After vendor model updates (small sample)	Of their AI systems

APRA has observed the same weakness, noting reliance on point-in-time and sample-based assurance that is poorly suited to systems whose outputs are probabilistic. Aggregate measures can look stable while outcomes for particular customers or document types change.

> *An AI system that can change without your knowledge cannot be assured by testing it once.*

You Own What Your AI Says

When AI gets something wrong in front of a customer, the firm carries the consequences. New Zealand consumers are already meeting confident, wrong AI-generated financial information, and tribunals overseas have made clear who is responsible for it.

The Ombudsman's Warning	Moffatt v Air Canada	Klarna's U-turn
In April 2026, New Zealand's Insurance and Financial Services Ombudsman warned that AI search summaries were telling people up to 80 or 90% of insurance complaints succeed if they persist, and that AI tools were citing previous cases that did not exist.	An airline chatbot told a bereaved customer he could claim a bereavement fare after travelling, contradicting the airline's own policy. The airline argued the chatbot was responsible for its own statements. A Canadian tribunal disagreed. The award was small. The principle was not.	Klarna, the Swedish buy-now-pay-later and payments company, moved its customer service to AI to cut costs. In May 2025 its chief executive acknowledged the result was lower quality, and the company began recruiting human agents again.

3. The Accountability Problem: Decisions Nobody Can Reconstruct

Dependence Without Understanding

Most firms now depend on AI they did not build. The Bank of England and FCA survey found that firms understand their own AI far better than the AI they buy in.

77%	24%	81%
Deep knowledge	**Deep knowledge**	**Rely on the vendor**
Of key systems built in-house	Of key systems bought in	Assurances and contracts

> A contract can allocate liability. It cannot explain a decision to a customer, a dispute resolution scheme or a regulator.

Governance After the Fact

In October 2026, the UK Financial Conduct Authority published examples from its supervisory work. One firm deployed AI tools before it had established clear governance. When risks emerged, each team assumed oversight sat with someone else, and there was no clear route to escalate or challenge. Retrofitting governance proved difficult and left gaps in accountability.

In another, human review of AI recommendations, meant to be the safeguard, became a tick-box exercise. Reviewers increasingly accepted recommendations without challenge, even when there were signs of error, and unclear escalation criteria let some low-confidence outputs through.

> Having a human in the loop is not the same as having a control.

Obligations That Already Apply

New Zealand firms do not need to wait for AI-specific legislation to be accountable. Where AI produces or relies on information about identifiable customers, existing privacy law applies.

Binding: Privacy Act IPP 8	**Expected: Commissioner's AI Guidance**
An agency must take reasonable steps to ensure personal information is accurate, up to date, complete, relevant and not misleading before using or disclosing it.	The Privacy Commissioner expects organisations to test that AI tools are accurate and fair for their purpose, and to complete privacy impact assessments before use. These are expectations, not statutory rules, but they signal how reasonable steps will be judged.

Records That Cut Both Ways

AI also creates records that firms did not set out to keep. Bell Gully noted in October 2026 that AI prompts, outputs and chat histories are "documents" for discovery purposes under the High Court Rules, and that conversations with AI tools do not attract legal privilege simply because they concern legal questions. Overseas courts have already admitted executives' chatbot conversations as evidence of intent.

The result is an uncomfortable combination. A firm may lack the evidence it needs to explain an automated decision, while holding unstructured AI records that can be used against it.

4. From Experiment to Controlled System

None of this requires a new discipline. Financial services firms already operate critical systems through regression testing, release management, change control, audit trails and incident response. Operating AI safely means applying those disciplines to a component that can change on its own. Five questions test whether a firm has done so.

1. Do we know where AI is being used?

A current inventory of every AI system: what it does, which model or supplier sits behind it, who owns it, what data it uses and how much it decides without a person. Include AI embedded in vendor products, which is where much of the growth is occurring.

2. Can we prove it works as intended?

A fixed set of representative, anonymised cases with known correct outcomes, including difficult edge cases and comparisons across customer groups. Aggregate accuracy is not enough; the evidence has to hold for individual decisions.

3. Will we know when its behaviour changes?

Model versions tracked, the test set re-run on a schedule and after every vendor update, and alerts when important outcomes shift, such as approval rates, referral rates or complaint themes.

4. Can we reconstruct and explain a decision?

Enough evidence for each consequential decision (inputs, model version, instructions, output and any human approval) to investigate a complaint, answer a regulator or correct a record. This is deliberate record-keeping, not an accidental archive of chat logs.

5. Can we intervene when something goes wrong?

A named owner, a clear escalation path, the ability to switch a system off or roll it back, a route for customers to reach a person, and a process to put things right for customers already affected.

> **For smaller firms.** None of this demands a model risk department. It demands an owner, a list, a test pack and the habit of using them.

5. Closing: Trust Is Not a Control

> *The risk is not simply that AI gets something wrong. It is that an organisation may not know when it does, may not be able to explain why, and may have no reliable way to prevent the same mistake happening again.*

AI is not unsuited to financial services. The productivity and service gains are real, and firms that ignore them will fall behind. But a system that influences who receives credit, how a claim is handled or what a customer is told deserves the same operating discipline as a payments engine or a loan management system.

Regulators on both sides of the Tasman have stopped asking whether firms use AI. They are asking whether firms understand it, own it and can control it. As the Reserve Bank's Angus McGregor put it, entities are responsible for having "the systems and controls in place" to manage the risks.

Firms that can answer the five questions in this paper will be able to adopt AI faster, because they will know when it is working. Firms that cannot are relying on vendor assurances and good fortune.

> *Trust is not a control. Evidence is.*

About Dark Arts

Dark Arts is a boutique software development firm that specialises in financial systems rescue and modernisation. We work with lenders, insurers, debt litigators, credit bureaus, fund managers and other financial services organisations to modernise their technology, ensuring it can support the compliance, assurance and governance requirements of an AI-enabled world.

If you are unsure whether you could demonstrate that the AI in your business is working correctly today, or if you recognise the gap between your current controls and where they need to be, we should talk.

Schedule a confidential discussion about your operational AI readiness

info@darkarts.it | 0800 DARK ARTS | darkarts.it

Sources

1. Bank of England and Financial Conduct Authority, Artificial Intelligence in UK Financial Services 2026, 26 October 2026
2. Financial Conduct Authority, AI in Financial Services: Good Practice and Areas for Improvement, 19 October 2026
3. Australian Prudential Regulation Authority, Letter to Industry on Artificial Intelligence, 30 April 2026
4. Reserve Bank of New Zealand and Financial Markets Authority, MR No. 2026-47, "RBNZ and FMA Outline Supervisory Expectations on Cyber and AI Risk", 15 October 2026
5. Wiese, "Human-Anchored Longitudinal Comparison of Generative AI with a Bias-Calibrated LLM-as-Judge", PLOS One, 2 February 2026
6. RNZ / NZ Herald, Insurance and Financial Services Ombudsman warning on AI-generated complaint advice, April 2026
7. Moffatt v Air Canada, British Columbia Civil Resolution Tribunal, 14 February 2024
8. Fortune, Klarna chief executive on AI customer service quality, 9 May 2025
9. Privacy Act 2020 (New Zealand), Information Privacy Principle 8
10. Office of the Privacy Commissioner, Artificial Intelligence and the Information Privacy Principles (current guidance)
11. Bell Gully, "AI and Legal Privilege: Key Risks for Businesses", October 2026

© 2026 Dark Arts Limited. All rights reserved. This document is provided for informational purposes only and does not constitute legal or regulatory advice.