Apporine Research | November 2026

The Accountable *Hour*

AI in bank outages, and who decides at 2am



Apporine

Foreword

Every bank now asks the same question about AI. Not whether it works, but who answers for it when it acts.

The hardest hour in banking technology is the first hour of a major outage. Customers cannot pay. Phones light up. A small team works through the night to find the cause and fix it, and in the morning a named Senior Manager explains what happened.

AI is about to join that team. It can read thousands of alerts in seconds and gather the clues an engineer would spend an hour collecting. It can also change a live system faster than any person can stop it.

This paper asks a simple question: in a bank's worst hour, which decisions can AI safely take, and which must stay with a person?

We studied what UK regulators said this year, what independent tests show AI can and cannot do, and what really happens in the first hour of major outages. We build software for this moment, and we say so plainly. We have tried to write the paper we would want to read if our own name were on the line at 9am.

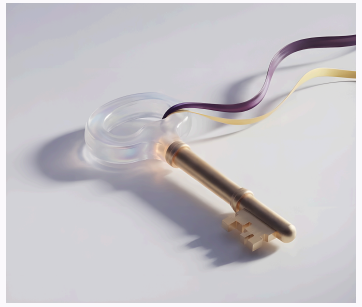


Aoon Syed

Author, Cofounder, Apporine



Contents



01
The 2am question
Who answers when AI acts
Page 6



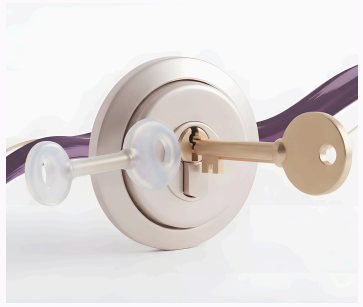
02
What AI can and can't do yet
Trust earned, not assumed
Page 8



03
The 2am Decision Count
What really happens in the first hour
Page 10



04
The Autonomy Ladder
Five steps, from looking to acting
Page 13



05
Reasonable steps
The evidence a Senior Manager needs
Page 15



06
Built around the human
How we designed Relyxus
Page 18

By the numbers
Page 4

In brief
Page 5

Ten board questions
Page 17

About the author
Page 20

Method and sources
Page 21

The Accountable Hour: *by the numbers*

75%

of UK financial firms already use AI

Bank of England and FCA, 2024

46%

say they have only a partial understanding of the AI they use

Bank of England and FCA, 2024

78%

of first hour actions in 24 major outages changed nothing in a live system

Apporine Decision Count, 2026

25%

of medium difficulty cases where the best AI model found the full root cause

ORCA-Bench, Traversal

158

IT failures at major UK banks and building societies, January 2023 to February 2025

Treasury Committee, 2025

24 hours

to file an initial incident report under UK rules from 18 March 2027

FCA and PRA

In brief

1. **Accountability does not move to the machine.** Under the UK Senior Managers regime, a named person still answers for what AI does. The FCA's October 2026 good practice paper says so plainly.

2. **AI is not yet ready to decide alone in an outage.** In a large independent test on real on call tasks, the best model found the full root cause in about one in four medium difficulty cases.

3. **Four in five first hour actions change nothing.** Across 165 actions in 24 public reports of major outages, only about one in five changed a live system. That is where a person must hold the key.

4. **The FCA's five human roles make a ready ladder.** Applied to incident response, our view is that live changes should stop at the approver step today.

5. **Reasonable steps can be evidenced.** We list nine pieces of evidence a Senior Manager should hold before AI is allowed to act.

A note on independence

Apporine builds Relyxus, AI incident response software for banks. To keep that out of the research, we:

- used only public sources
- wrote our selection rules before counting
- publish our full data sheet
- kept our product to one chapter, at the end

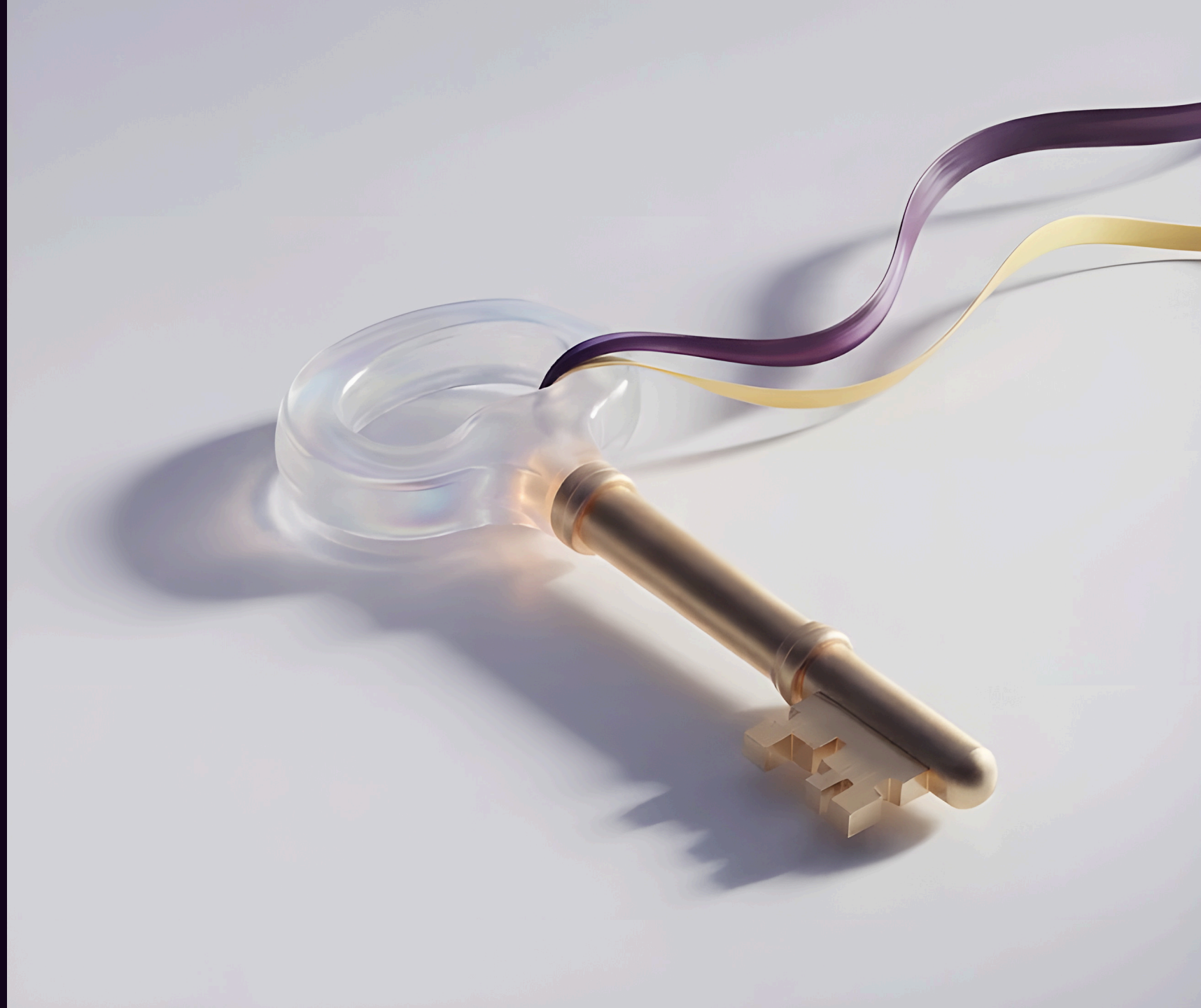


01

The 2am question

Who answers when AI acts

In a UK bank, accountability has a name. When AI starts acting inside live systems, that name does not change. The real question is not whether AI works, but whether a Senior Manager can show they took reasonable steps.



AI can act at 2am. A named person still answers at 9am.

Under the Senior Managers and Certification Regime, each important area of a UK bank has a named Senior Manager who is personally accountable for it.

In January 2026, the Treasury Committee's report on AI in financial services heard that a Senior Manager can be held responsible for failing to take reasonable steps with AI, and asked the FCA for practical guidance by the end of the year.

The FCA's answer came on 19 October 2026. Senior managers stay accountable. They do not need to be technical experts, but they must understand the risks, challenge them and keep controls that fit.

Outages make this sharper. In the first hour, decisions are made fast, by tired people, with incomplete information. If AI joins that team, three questions follow. What may it do alone? What needs a person's approval? And what evidence will show, afterwards, that the bank stayed in control?

From 18 March 2027, UK firms must also report major incidents faster: an initial report generally within 24 hours, and a final report within 30 working days. The FCA expects significant AI failures and near misses to be reported in the same way.

What are "reasonable steps"?

A Senior Manager can be held responsible if something goes wrong in their area and they did not take the steps a reasonable person in their role would have taken to prevent it.

With AI, those steps must be shown with evidence, not assumed.

The question is not whether AI works. It is whether you can prove you stayed in control when it did.



Sources: Treasury Committee, January 2026. FCA, 19 October 2026. FCA and PRA incident reporting rules.

02

What AI can and can't do yet

Trust earned, not assumed

AI is excellent at reading, sorting and summarising at speed. Independent tests show it is not yet reliable at finding the true cause of complex failures on its own. Banks need to know exactly where that line sits.



The best AI found the full cause in one case in four.

ORCA-Bench, built by Traversal with researchers at Columbia University and Cornell Tech, is one of the largest public tests of AI on real on call work: 1,079 tasks drawn from realistic production incidents. The best model found the full root cause in 25.3% of medium difficulty cases, and in 10% of the hardest. Microsoft's OpenRCA benchmark, presented at ICLR 2025, points the same way.

Inside firms, understanding is thin. In the Bank of England and FCA's 2024 survey, 46% of UK financial firms said they had only a partial understanding of the AI they use, and a third of all AI use cases came from third parties.

Let AI do the looking. Keep the deciding with people until the evidence says otherwise.

Exhibit 1

Share of cases where the best AI model found the full root cause

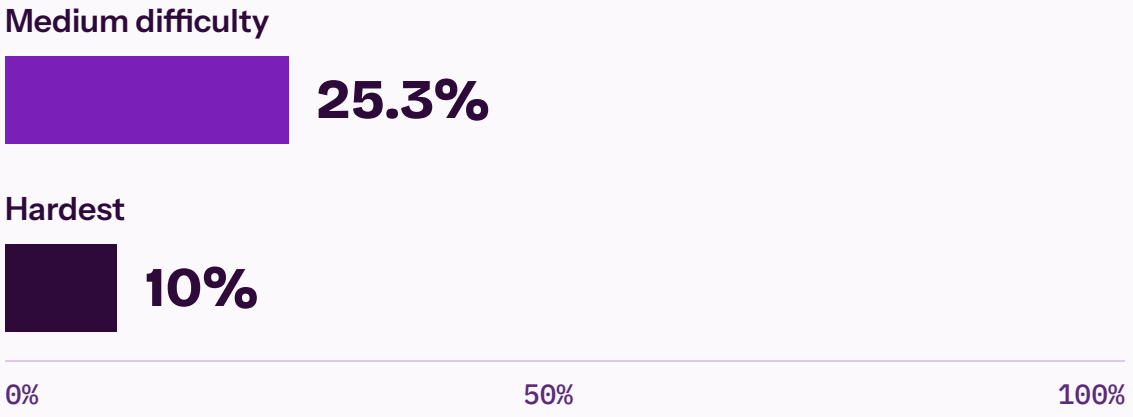


Chart takeaway: *AI is ready to gather the evidence. It is not yet ready to decide alone.*

Source: Traversal, ORCA-Bench, with Columbia University and Cornell Tech.

03

The 2am Decision Count

What really happens in the first hour

We sort every action taken in the first hour of real, publicly reported major outages: what was looking, and what was changing. As far as we know, it is the first count of its kind.



How we count

1. **Choose the reports.** 24 public post incident reports, 2017 to 2025, from banks, payment firms and the cloud, network and software firms they rely on. Rules written before counting.
2. **List every action** taken in the first hour after the outage was detected.
3. **Sort each action** into one of three boxes.
4. **Publish every line.** The sort was prepared with AI assistance using this rulebook. Every action and how it was sorted is public, so anyone can check our work.
5. **Publish the data sheet,** including the 7 reports we left out and why.

Safe for AI: looking and checking

Reading alerts, logs, dashboards and recent changes. Building the timeline. Nothing in the live system changes.

Grey zone

Actions that change nothing on their own but could mislead if wrong, such as paging teams or drafting customer updates.

Needs a human: changing

Anything that changes a live system: rollbacks, restarts, failovers and configuration changes.

What this cannot tell you. Public reports are written after the event and tidied up. They show what was reported, not every keystroke.



Four in five first hour actions change nothing

Exhibit 3

165 actions from the first hour of 24 publicly reported major outages, sorted by what each action did



78%

of first hour actions changed nothing in a live system. This is where AI can help most, and most safely.

22%

changed a live system: rollbacks, restarts, configuration changes. This is where the second key belongs to a person.

42%

were about people and messages: paging, declaring and updating customers. The biggest box, and the most overlooked.

Sorted with AI assistance using the published rulebook. All 165 actions, and how each was sorted, are in the data sheet at apporine.com/research.

04

The Autonomy Ladder

Five steps, from looking to acting

In July 2026, the FCA's Mills Review described five roles a person can play beside AI. We apply them, step by step, to the worst hour in banking technology.



Who holds the key at each step

Step	FCA role	What AI does in an outage	Who holds the key
1	**Operator**	Runs searches and summaries when an engineer asks	The engineer
2	**Collaborator**	Gathers alerts, logs and recent changes, and builds the timeline as the team works	The engineer
3	**Consultant**	Suggests the likely cause and a fix, with its evidence	The engineer decides
4	**Approver**	Prepares the change. Nothing runs until a named person approves it	**Two keys: AI prepares, a person turns**
5	**Observer**	Acts alone within agreed limits. People watch and can stop it	AI, with people watching

Our view: for live changes during a major outage, UK banks should stop at step 4 today. Step 5 should be earned, one action at a time, with evidence.



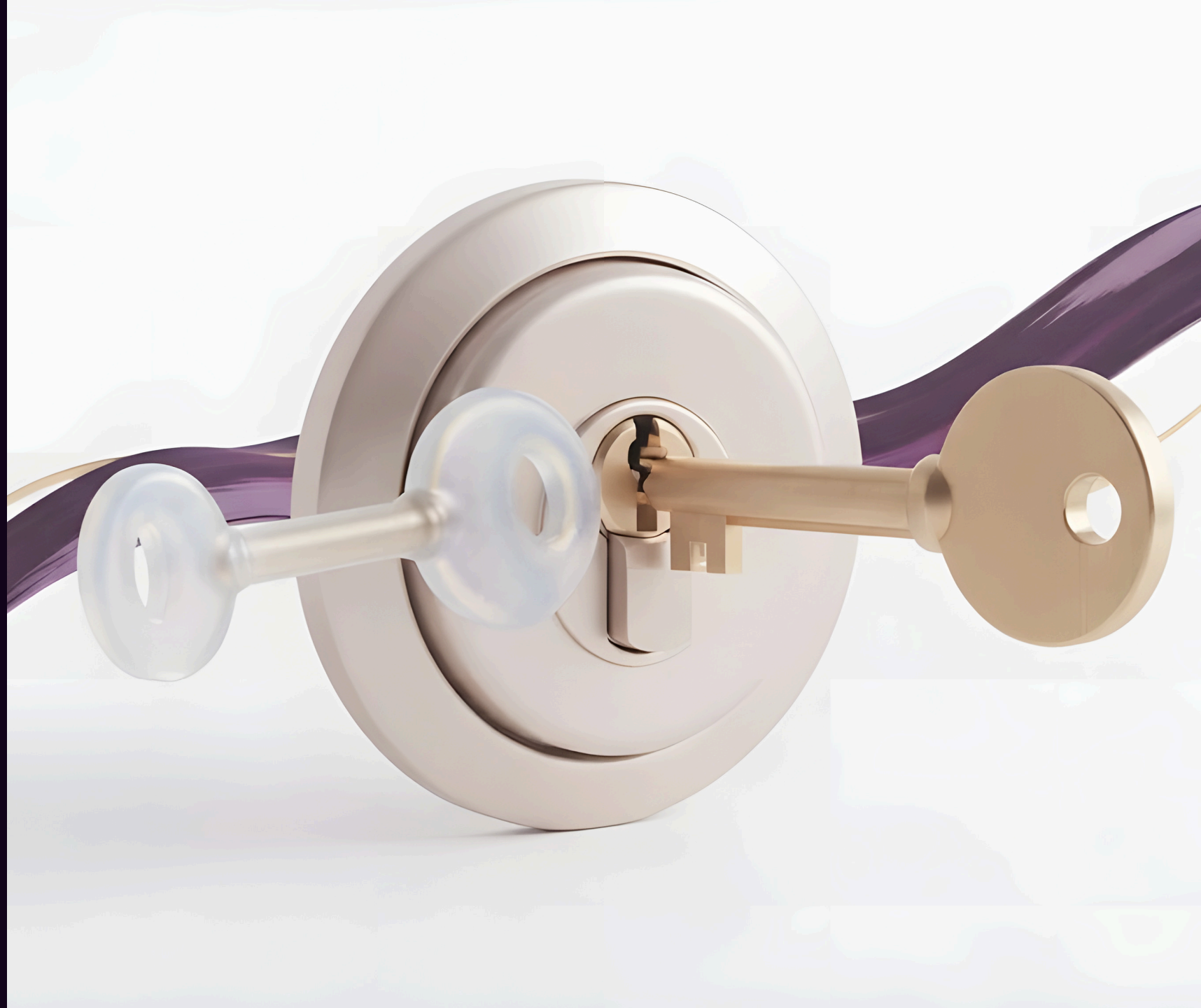
Roles from the FCA Mills Review, July 2026. Applied to incident response by Apporine.

05

Reasonable steps

The evidence a Senior Manager needs

Reasonable steps are shown with evidence. Built on the FCA's October 2026 good practice paper, this is what we believe a Senior Manager should hold before AI is allowed to act in an outage.



Nine pieces of evidence before AI acts

1 A named owner from day one

Accountability set at design and mapped to the right Senior Manager.

2 One inventory of AI agents

Owner, data access and controls for each, in one place.

3 Clear escalation triggers

Thresholds that decide when AI must stop and call a person.

4 Real human review

Reviewers with the authority, information and time to overrule.

5 A complete audit trail

Every AI action, its evidence and the decision behind it, recorded.

6 Tests before go live

Stress tested before deployment, and again after model changes.

7 A playbook with a stop button

Set roles and thresholds for closer monitoring or temporary shutdown.

8 Supplier control

Notice of model changes, a fallback plan and a tested exit route.

9 AI in incident reporting

Significant AI failures and near misses reported like any incident.

Built on FCA, AI in financial services: good practice and areas for improvement, 19 October 2026.



Ten questions for your next board meeting

Answer each with a plain yes or no. Every no is a gap to close before AI is allowed to change a live system.

1. Can you name the Senior Manager accountable for each AI tool used in incident response?
2. Do you keep one inventory of every AI agent that can reach production?
3. Does every AI change to a live system need a named person's approval?
4. Are the triggers that make AI stop written down and tested?
5. Could you show a regulator every step an AI tool took in your last major incident?
6. Do your reviewers have the authority and the time to overrule AI?
7. Was each AI tool stress tested before go live, and after its last model change?
8. Can you switch any AI tool off in one step, and does the team know how?
9. Would your supplier tell you before changing the model underneath?
10. Are AI failures and near misses part of your incident reporting?

Reading your answers

10 yes: ready for step 4 of the ladder.
7 to 9: close. Fix the gaps first.
Under 7: keep AI to looking and checking for now.



06

Built around the human

How we designed Relyxus

We build Relyxus, AI incident response for banks. This is the one chapter about us: the principles, not the parts.



Five principles we will not break

1	**Look first**	It starts by gathering alerts, logs and recent changes, so engineers begin the hour ahead.
2	**Propose, never presume**	It suggests the likely cause and a fix, and shows its evidence.
3	**Two keys for every change**	Nothing in a live system changes until a named person approves it.
4	**Everything on the record**	Every step, every piece of evidence and every approval is kept for audit.
5	**Stop means stop**	One command halts it, and the bank sets the limits it works within.

See it work on a replay of a real public outage: relyxus.com/replay



About the author

Aoon Syed

Author

Cofounder, Apporine



Data and method

The rulebook and the full data sheet are public at apporine.com/research.



About Apporine Research

Apporine Research publishes practical work on how AI should be governed inside the systems that banks and enterprises cannot afford to lose. Every paper uses public sources and an open method.

Contact: apporine.com/contact

Method and sources

This paper uses only public material: regulatory publications, parliamentary reports, independent benchmarks and public post incident reports. Chapters 4 and 5 are Apporine's own views, built on the sources listed. Nothing in this paper is legal or regulatory advice.



1. Bank of England and FCA. *Artificial intelligence in UK financial services*, November 2024. bankofengland.co.uk
2. FCA. *AI in financial services: good practice and areas for improvement*, 19 October 2026. fca.org.uk
3. FCA. *The Mills Review: the future of AI in retail financial services*, July 2026.
4. House of Commons Treasury Committee. *Artificial intelligence in financial services*, January 2026.
5. House of Commons Treasury Committee. Evidence on IT failures in the banking sector, January 2023 to February 2025.
6. FCA and PRA. Operational incident and third party reporting rules, in force 18 March 2027.
7. Traversal, with Columbia University and Cornell Tech. *ORCA-Bench*. traversal.com
8. Microsoft Research. *OpenRCA*, ICLR 2025.
9. Addleshaw Goddard. *Senior managers liable under UK regulatory regime for decisions made by AI*, February 2026.

Apporine

Who holds the key at 2am?

Read this paper online, and download the full data sheet, at **apporine.com/research**



Wembley, London, United Kingdom

© 2026 Apporine. For information only.
Not legal, regulatory or financial advice.