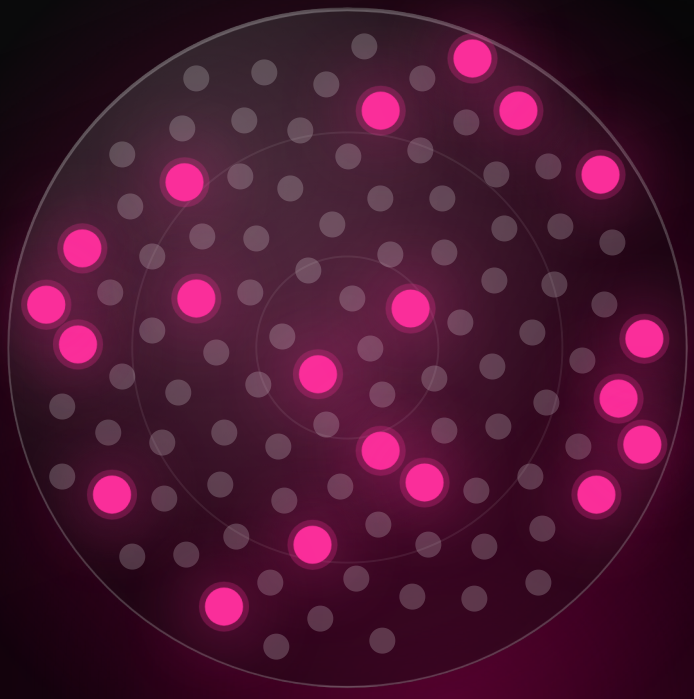
Aveni Detect platform data study

The Evidence Gap

What 674,000 customer conversations reveal



Each circle is 1% of calls. Pink: at least one complaint, dissatisfaction or vulnerability signal.

673,671
real customer calls analysed

28
UK financial services firms

12 months
1 Nov 2025 – 31 Oct 2026

At a glance

Key findings at a glance

We analysed 674,000 real customer calls across 28 UK financial services firms over 12 months. Here is what the data shows.

1 in 5
calls carried at least one complaint, dissatisfaction or vulnerability signal

1 in 5
calls had a vulnerability signal, at the typical firm

53%
contained a vulnerable customer call where the firm's own QA said vulnerability wasn't addressed

84%
of calls that had shown expressions of dissatisfaction contain no complaint flag

12%
of signals were held by the 'riskiest' 10% of advisers

0.8%
calls reviewed by a person

1.3 hours vs 9 days
Median time to an AI review, vs a completed human review

The conclusion

Risk is routine, it's spread across every team and every hour, and it sits in the calls that cost most to review by hand. Sampling cannot see it, and firms that want to prove good outcomes need to see every conversation.

Why it matters

Proving good customer outcomes is critical

Consumer Duty asks firms to prove good outcomes, not just to have good intentions. That proof lives in customer conversations, and most of those conversations are never reviewed.



Circles drawn to scale by area.

1 in 5

Calls carrying at least one complaint, dissatisfaction or vulnerability signal

0.8%

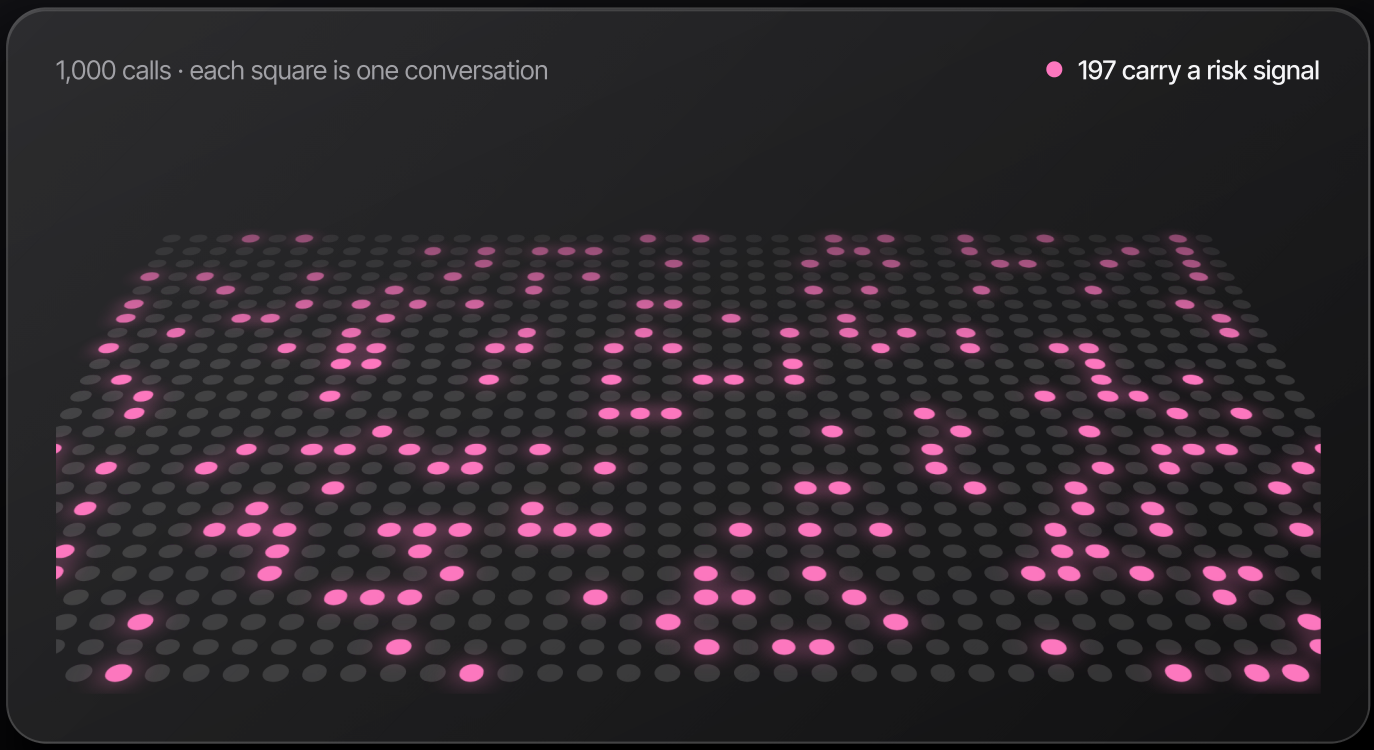
Calls reviewed by a person

The usual answer is to sample: review a small share of calls and assume it represents the rest. This report tests that assumption against what actually happens across hundreds of thousands of real interactions drawn from anonymised Aveni Detect platform data, rather than a survey of opinions.

Finding 01

Risk is routine

About one in five calls (197 per 1,000) carried at least one AI-flagged sign of a complaint, dissatisfaction or customer vulnerability, showing that they're not edge cases. In a team handling 1,000 calls a week, that is roughly 200 conversations a week that a conduct or complaints reviewer would want to see.



1 in 4 signals

A quarter of all signals sit on calls carrying more than one. Where a customer is unhappy, they are often also vulnerable, or on the way to a complaint.

Key takeaway

Risk is a feature of everyday customer contact. An outcome testing approach built around finding the occasional problem call will never be sufficient.

Finding 02

Vulnerability is common, and half the time it isn't addressed

221 per 1,000 calls

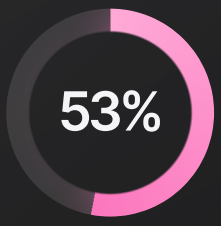
At the typical firm, around one in five calls showed signs of customer vulnerability.

53% of flagged calls

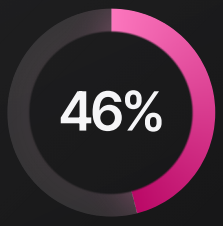
The firm's own QA answered No to: if the customer is vulnerable, does the adviser address it?

The harder finding is what happens next. When firms ask their own QA question, across 12,678 flagged calls at 11 firms, the typical firm's answer was No in 53% of cases. Counting every question firms ask about identifying, acknowledging or acting on vulnerability, the typical firm recorded at least one failure on 46% of flagged calls.

Vulnerability handling at the typical firm, measured against the firm's own QA questions



Adviser did not address vulnerability



At least one failure on identifying, acknowledging or acting

Base: 12,678 flagged calls 11 firms Median of each firm's own rate

Key finding

This is measured against firms' own standards, not ours. Vulnerable customers are present in a fifth of conversations, and around half the time the firm's own checklist says they were not properly supported. Without full coverage, most of those calls are never seen.

Finding 03

Most dissatisfaction never becomes a complaint flag

84% of calls that had shown expressions of dissatisfaction contain no complaint flag.



Dissatisfied, no complaint flag Dissatisfied, complaint flagged

6 in 100

That is roughly six in every hundred calls overall: a dissatisfied customer, and no complaint recorded.

Not every unhappy customer is making a complaint. Under the FCA's DISP rules, a complaint must allege financial loss, material distress or material inconvenience. But these are exactly the conversations a complaints reviewer would check first, and the kind that can resurface later at the Financial Ombudsman Service.

> "The gap between what customers say and what firms record is where complaints-handling risk hides."

Key finding

A firm can only decide whether a dissatisfied call was a complaint if someone sees it.

Finding 04

You can't sample your way to the risk

The common defence of sampling is that a smart QA team can aim it: at weaker advisers, or at busy periods. The data says there is little to aim at. Risk isn't concentrated by person or by time. The only thing that concentrates it is call length, and the long calls are both nearly half the volume and the most expensive to review by hand.

Not by person | Not by time | **Only by length**

Not by person

Ranked within each firm, the top 10% of advisers hold just 12% of signals. The top 20% hold 26%. There is no small group of problem advisers whose calls would catch most of the risk.



Bars drawn on a 0–100% scale. Advisers ranked by signals within each firm.

Not by time

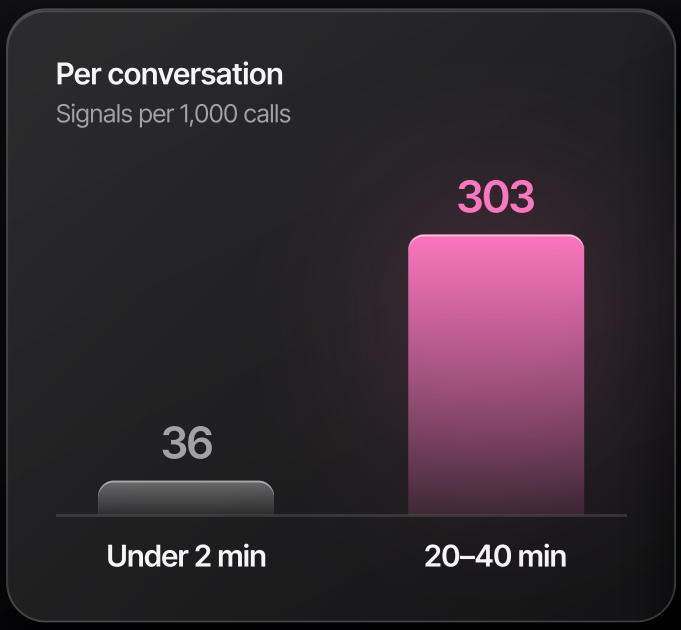
Signal rates barely move across the working week or the working day. The highest-rate hours hold almost exactly their share of calls.

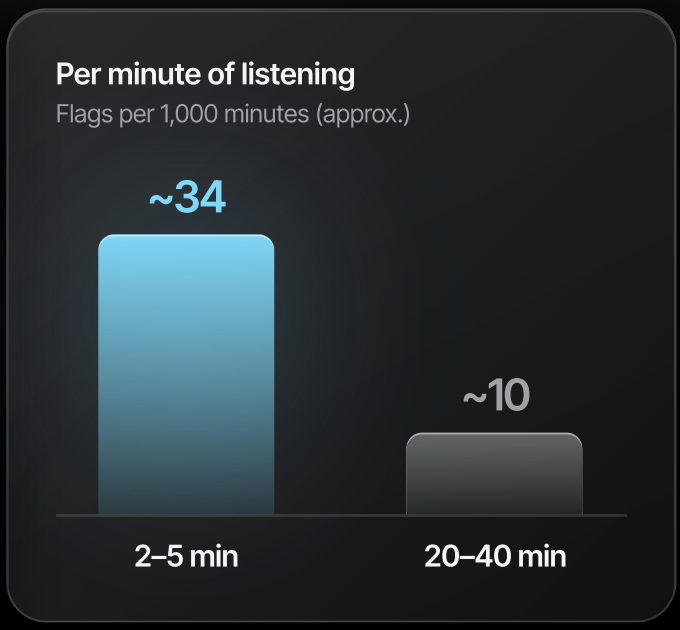
Continued on the next page: the one thing that does concentrate risk. →

Finding 04 · continued

Only by length, and that's the trap

Signals climb steadily with call length, from 36 per 1,000 calls under two minutes to 303 per 1,000 for calls of 20 to 40 minutes. But per minute of listening, shorter calls turn up more flags. A cost-conscious outcome testing team is naturally pulled towards shorter calls, while the risk per conversation sits in the long ones.





The per-minute view rests on average call lengths, so treat it as directional. Each chart has its own scale.

10×+

Vulnerability signals rise more than tenfold from the shortest calls to calls of 20 to 40 minutes.

~45%

of call volume is calls of 10 minutes or more: the costliest to review by hand.

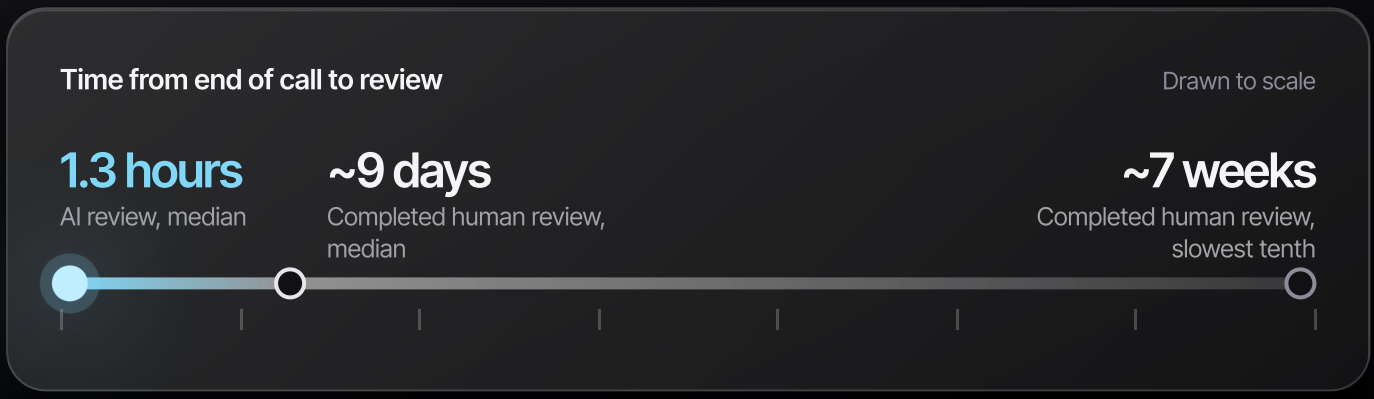
Key finding

Better targeting can't close the gap. The risk is everywhere, and the part that is concentrated is the part that is hardest to review by hand.

Finding 05

Hours, not weeks

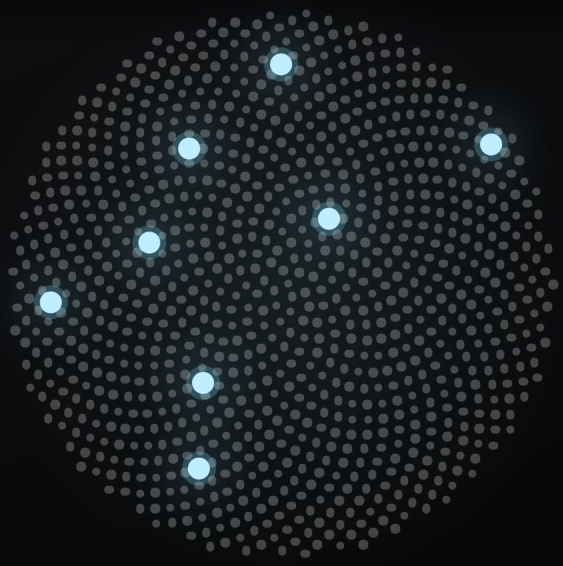
The median time from the end of a call to an AI review was **1.3 hours.** The median time to a completed human review in Detect was **about 9 days,** and the slowest tenth took around seven weeks.



0.8%

of calls were looked at by human reviewers in Detect.

That is not a criticism of outcome testing teams, but more the arithmetic of reviewing conversations one at a time.



1,000 calls. The 8 in blue are the ones a person reviewed.

Key finding

A vulnerable customer flagged within hours can still be helped. One found weeks later, if at all, becomes a remediation case. Speed turns outcome testing from a record of what went wrong into a chance to put it right.

Recommendations

What firms should do next

The evidence gap is not a failure of effort but of the method. Five steps close it.

01

Move from sampling to full coverage

Review every conversation automatically, so evidence of outcomes covers the whole book, not a sliver of it. Keep people for judgement, not for finding the calls.

02

Point human reviewers at flagged calls, not random ones

With AI triage, a reviewer's hour goes on the one call in five that carries a signal, and on the long, complex conversations sampling tends to skip.

03

Make vulnerability a same-day process

Flag vulnerability within hours and route it to someone who can follow up while the customer can still be helped. Track how often your own handling questions are failed, and why.

04

Close the dissatisfaction-to-complaint loop

Review every dissatisfied call against DISP, and record the decision either way. That record is your defence if a case reaches the Ombudsman.

05

Report outcomes to the board from the full population

Replace sample-based assurance with rates drawn from every interaction: signals per 1,000 calls, vulnerability handling, time to action. That is evidence a regulator can rely on.

Aveni Detect

See what full coverage would reveal in your own conversations

Aveni Detect was built to do exactly this: review every customer interaction against your own standards, surface the conversations that matter within hours, and give conduct, compliance and operations teams the evidence to prove good outcomes.

Book a Detect demo aveni.ai

Methodology

About the data

This report draws on 673,671 customer calls analysed by Aveni Detect's AI between 1 November 2025 and 31 October 2026. They come from 28 UK financial services firms, mainly in wealth management and consumer credit, with some insurance and banking.

Signals, not verdicts

Every figure counts AI-flagged signals of complaint, dissatisfaction or vulnerability. They are not confirmed issues, and reviewers overturn a share of them.

The typical firm

Where firm size would skew a figure, we report the typical firm: the median of each firm's own rate.

Firms' own standards

Vulnerability handling is measured against the questions in each firm's own QA forms, at the 11 to 14 firms whose forms ask them.

Human review

Review figures cover review inside Detect only. Firms may also run QA elsewhere.

Privacy

All data is anonymised and aggregated. No figure covers fewer than three firms or 100 calls, and no single firm makes up more than half of any figure. No firm, adviser or conversation can be identified.