Accuracy you can trust — and the evidence behind it

How accurate is Aveni, and how we prove it

Accuracy matters at every stage of a client conversation: from the words spoken in the room, to the meeting summary, to the compliance checks that follow. Here's how Aveni keeps it right across Assist and Detect, and how we measure it.

Assist

Capturing the conversation accurately

Assist records and transcribes your client meetings, then writes an AI meeting summary. A summary is only ever as good as the transcript beneath it, so accuracy starts with getting the words right.

Accurate transcription

Accent-robust speech models handle strong UK regional and international accents without losing quality.

Financial terminology. Product names, providers, regulatory terms and adviser shorthand are recognised, not misheard.

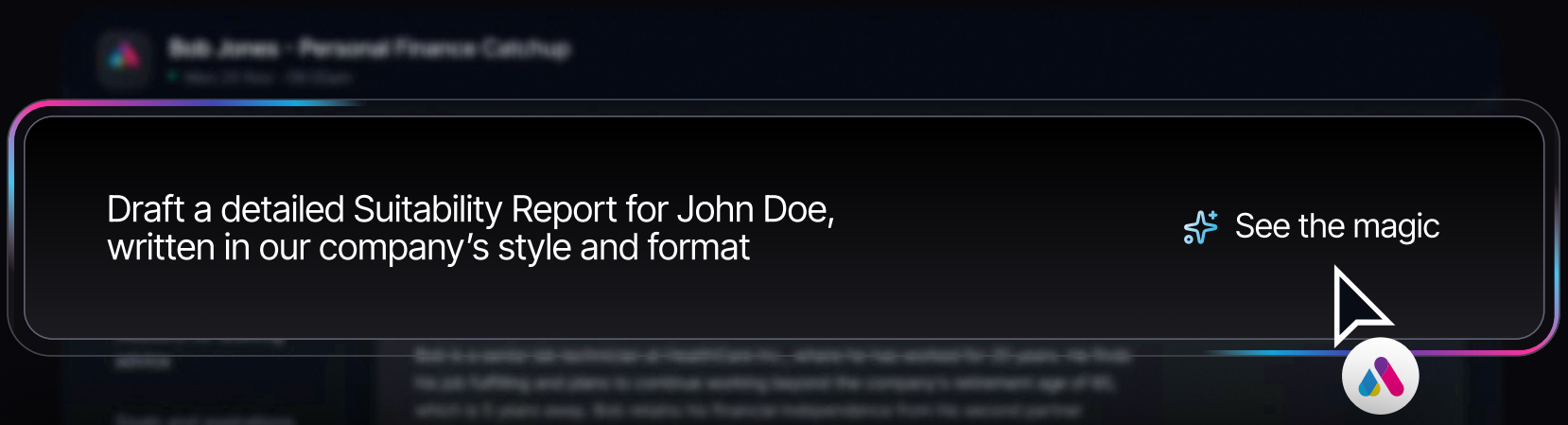
Speaker separation keeps adviser and client turns apart, so the record clearly shows who said what.

Faithful meeting summaries

Grounded in the transcript. Summaries only draw on what was actually said, so the content stays traceable back to the meeting.

Structured for advice. Key points, client circumstances, risks and agreed actions are captured consistently in every meeting.

You stay in control. The adviser reviews and confirms each summary before it's saved, so a person always signs off the final record.



Detect

Automated compliance you can evidence

Detect uses AI to run compliance checks automatically across your customer interactions and their documents. For every check, Detect gives you a clear outcome, a plain-English explanation of why, and the supporting evidence from the transcript or document.

"Were the client's objectives established?" Yes No Not applicable

The AI behind Detect

Leading AI models, including OpenAI models hosted securely in EU data centres, combined with Aveni's own prompts, controls and guardrails built for UK financial services.

FinLLM, Aveni's own model, is fine-tuned in-house for compliance checking and powers Detect.

A dedicated vulnerability model for the subtle, context-dependent job of spotting signs of customer vulnerability.

How we measure it

We measure every check three ways: **precision** (how often Detect is right when it flags something), **recall** (how many genuine issues it catches) and **F1** (the two brought together in a single score).

Why not "accuracy"? Most interactions are fine, so a system that always said "no issue" could look 95% accurate while catching none of the real problems. Precision, recall and F1 tell you what's really happening.

Checked by real experts. Compliance specialists using the platform review Detect's outcomes, explanations and evidence, and give feedback on each one. That expert feedback is what we use to score precision, recall and F1 for every check, so we can track accuracy over time and keep improving it.

How Detect performs today

~0.90
Average F1 score
across automated compliance checks

0.85–1.00
Typical range
precision, recall & F1 per check

~0.80
Vulnerability detection F1
evaluated separately

Performance varies a little between clients and checks, reflecting differences in processes, products and data. Vulnerability detection scores lower by nature, not by quality: it relies on subtle language and context that even human experts read differently, so around 0.80 is a strong result for this kind of task.

Figures are indicative and vary with your configuration and data. Current figures for your set-up are available on request from your Aveni account team.