CLAUDE OPUS 5.5 ALONE VS WITH FIVE DOTS

With Five Dots, Claude found 94% of the problems. On its own, 65%.

We planted problems in a dummy company's sales pipeline and asked Claude Opus 5.5 to audit it, ten times on its own and ten times with Five Dots. With Five Dots it found more and made fewer mistakes. The gap comes from the client: Five Dots asks what the CRM cannot show. A blind judge preferred the Five Dots audit in 8 of 10 head-to-head pairs, and the Opus-alone audit in none.

94%
vs 65% alone
of planted problems found, weighted by severity

100%
vs 24% alone
of the problems only the client knew

98%
vs 90% alone
of findings right (precision)

8/10
vs 0 of 10 alone
pairs a blind judge preferred (2 even)

The test: one company, one pipeline, eleven planted problems

Both setups audited the sales pipeline of the same invented company, with the same CRM, the same client and the same request. The only difference was the Five Dots connector.

The company. Kestrelview Analytics sells product-analytics software to mid-market online retailers: about 70 people, Series A, sales-led, a 14-day free trial, contracts of about $24k a year in the US and UK. Its HubSpot holds 44 deals, three sales reps and their activity history.

The client. Maya Lindqvist, Head of Revenue Operations, played by a separate AI. She answers what she is asked and volunteers nothing, so some problems only surface if the auditor asks.

How it was scored. Each setup ran the audit 10 times. An independent judge (Claude Opus) read every final audit blind, with product names removed, and checked each finding against the answer key and the CRM records. High-severity problems count three times, medium twice, low once. Three traps, things that look wrong but are deliberate, counted against an audit that called them errors.

Results at a glance

Averages over 10 audits each, run on 15 November 2026 against the current Five Dots release, the same version clients use.

	OPUS 5.5 ALONE	OPUS 5.5 + FIVE DOTS	HUMAN CONSULTANT (ESTIMATE)
Planted problems found (weighted by severity)	65%	**94%**	97%
Problems visible in the CRM data	84%	**91%**	93%
Problems only the client knew	24%	**100%**	100%
Findings that are right (precision)	90%	**98%**	98%
Wrong findings per audit	0.5	**0.1**	0.1
Traps wrongly called a problem, per audit	0.3	**0**	0
Rounds of questions to the client	0	**2–3**	5–7
How specific the fixes are (1–5)	2.9	**3.4**	4.5
Top-3 fixes that match the key (of 3)	1.1	**1.0**	2
Time per audit	1.5 min	**4.7 min**	12+ hours
Cost per audit	$0.39 AI compute	**$1.55 AI compute**	about $2,000 fees

Five Dots takes longer and uses more compute because it asks the client questions and writes the result into the client's workspace. Both AI setups put the right fixes first about equally often. The human consultant column is our estimate from Revenue Wizards' own consulting audits, not a result of this test: no consultant audited Kestrelview. It shows where the two AI setups sit against the work they stand in for.

The questions: Opus alone asked none

Opus alone asked no questions in any of its 10 audits: it read HubSpot and went straight to findings. Opus with Five Dots read HubSpot first, then asked about what the data could not explain, two or three rounds per audit. From one typical Five Dots audit:

WHAT FIVE DOTS ASKED	WHAT THE CLIENT SAID	WHAT IT CHANGED
10 of 24 open deals are renewals, expansions or pilots in the new-business pipeline. Is that on purpose?	Pilots are deliberate, paid 30-day pilots. She could not say about renewals.	Pilots not treated as an error; renewals logged as an open question for the VP Sales
Is the $140k Northgale deal right, against a $24k average contract?	It is real: their biggest prospect, run by Rhea and Daniel.	Not called a data error (one of the traps)
Is there a regular pipeline review, and who notices when a deal goes quiet?	No review. Deals are looked at when something comes up; the only fixed cadence is a quarterly board update.	New finding no pipeline review
Why do wins jump from the first stage to Closed Won? Are stage criteria written down?	Nothing is written down. Reps move a deal when it feels right.	New finding no stage criteria
Where do the forecast and the board number come from?	The VP Sales builds it every Friday in his own spreadsheet, because nobody trusts HubSpot.	New finding ranked first among the fixes
When were lost deals last reviewed together?	She did not know.	Logged as an open question, not as a finding

The last row matters as much as the others: an answer of "I don't know" was kept as a question for the person who would know, not written up as a problem.

> **A human consultant (not part of this test)** usually asks more, and of more people. A typical audit starts with the Head of RevOps, then interviews the VP Sales, two or three reps and finance, five to seven rounds in all. That is where a consultant goes further than either AI setup: talking to reps shows why deals go quiet, not only that they do. It is also where most of the 12+ hours go.

The findings: the difference is in what the CRM cannot show

On what the CRM shows, Opus is strong on its own: it found six of the eight data problems in every audit. Five Dots makes the difference on the problems that need the client, and on one data problem that needs a methodology to look for it.

PLANTED PROBLEM	SEVERITY	OPUS ALONE	OPUS + FIVE DOTS
Stale pipeline: 9 of 24 open deals quiet 60+ days, $227k	High	100%	100%
7 open deals past their close date	High	100%	100%
Open deals with no close date or $0 amount	Medium	100%	100%
10 of 12 lost deals with no lost reason	Medium	100%	100%
Renewals and expansions in the new-business pipeline	High	10%	50%
4 deals stuck in Contract Sent for 77+ days	Medium	100%	100%
5 of 8 wins skipped the middle stages	Low	100%	100%
One rep owns 7 of the 9 stale deals	Low	95%	100%
No written stage exit criteria CLIENT ONLY	High	65%	100%
Forecast built in the VP Sales' spreadsheet CLIENT ONLY	High	0%	100%
No regular pipeline review CLIENT ONLY	Medium	0%	100%
Trap: the large, real Northgale deal called a problem	—	30%	0%

Share of the 10 audits that found each problem; a partial find counts as half. For the trap, lower is better. Opus alone sometimes guessed the stage-criteria problem from the skipped stages, but never learned about the forecast spreadsheet or the missing review, because it never asked.

> **A human consultant** would be expected to find the same data problems, since they sit in the CRM, and the client-only problems through interviews. An experienced consultant adds most on problems that take judgement about how the business runs. Renewals sitting in the new-business pipeline is one: Five Dots found it in half its audits, and someone who has set up many pipelines tends to spot it from the deal names alone.

The presentation: a score, owners and a test of done

Both hand over a clear findings table with evidence. Five Dots adds a score, separates what the data proves from what the client said, names an owner and a test of done for each fix, and keeps open questions apart. From the same pair of typical audits:

	OPUS ALONE	OPUS + FIVE DOTS
Overall verdict	"A problem, high priority"	Scored red on a fixed scale, with the methodology version and date
Findings	One table of 10, data only	Scored findings from the data; what the client said; patterns noticed; what is working
Unknowns	Mixed in with the findings ("worth checking")	Listed apart, each with who should answer it
Fixes	5 in order, with what to do	4 in order, each with owner, why it comes first, what changes and when it counts as done
Where it lives	A notes page in the client's Notion	The module page, the audit overview, and the client's memory, journal and decision log in their Notion

THE FIRST FIX, AS EACH WROTE IT

OPUS ALONE

"Clean up the pipeline this week. Go through each rep's deals, starting with Jonas. For every stale or overdue deal, either close it out with a lost reason or give it a real next step and close date."

OPUS + FIVE DOTS

"Start a weekly pipeline review and build the forecast from HubSpot. Owner: Daniel. Why first: the board number currently comes from a spreadsheet because HubSpot isn't trusted. This one change addresses findings 1, 2, 6 and 8, and the forecast problem. Done when: the review has run four weeks in a row and the Friday forecast is read from HubSpot."

The Opus-alone fix is sound but treats the symptom, the stale deals. The Five Dots fix goes after the cause, which it learned by asking: nobody reviews the pipeline, so the forecast lives in a spreadsheet.

> **A human consultant** usually presents in a working session, not only a document: findings, the roadmap and owners are agreed with the leadership team in the room. That is why, in our estimate, their fixes are the most specific and the best ranked. Five Dots gives that conversation a comparable written starting point: a score, the evidence, owners and a test of done for each fix.

How to read this

- **One area, one company.** This covers the sales pipeline only, one of the ten areas in the Five Dots audit, for one invented company. A real audit covers more and depends on the real CRM.
- **The client is simulated.** Real clients say more, less and differently. The simulated client keeps hidden facts back until asked, which is what makes asking matter.
- **The judge is also an AI.** It reads each audit blind, without product names, and was checked first on a known-good and a known-bad audit.
- **No client data leaves the client's own setup.** In the Five Dots runs, the connector received only the names of the connected tools and the module requested, never a deal, name or number. Each run checks this.

OUR RECOMMENDATION

Use Five Dots for audit work that follows a known method, and bring in human expertise where the method runs out.

SITUATION	USE	WHY
Standard audits: a pipeline health check, a quarterly review, a new area of the GTM audit	Five Dots	It finds nearly what a consultant would, in minutes instead of days, for a fraction of the cost, and keeps the result in your own workspace.
Complex work, with RevOps experience in the team	Five Dots, **led by your RevOps lead**	An experienced operator takes the findings and open questions further: weighs the trade-offs, runs the stakeholder conversations and turns the roadmap into a change plan.
Niche topics and knowledge gaps: commission design, pricing, complex CRM architecture, an unfamiliar market	**Human expert, with Five Dots as the starting point**	A specialist brings what neither AI setup has: patterns from many similar companies and judgement on the edge cases.
Validation of a big decision before it goes to the board or changes how people are paid	**Human expert**	An outside check from someone accountable for the call.

In short: run every audit with Five Dots. Bring in a consultant only when you need specialist knowledge, or a second opinion on a big decision.