EXPONENTIAL | BY RAVENPACK

DECISION INFRASTRUCTURE FOR FINANCIAL AI

Lessons learnt — what worked, what didn't

Aakarsh Ramchandani

Chief Product Officer,
RavenPack

AGENDA

A SIX-MONTH UPDATE — THE ROAD WE'LL WALK

01 **The directional update**

Where the predictions landed. Receipts included.

02 **Expanding jobs to be done**

New workflows, unlocked by coding-capable models.

03 **Team structure**

Who actually ships this — and what hasn't worked.

04 **The hill climb**

Workflows plus evals. Why 2026 is different.

05 **Harness + architecture**

The physical reality. Interfaces vs. jobs.

06 **The close**

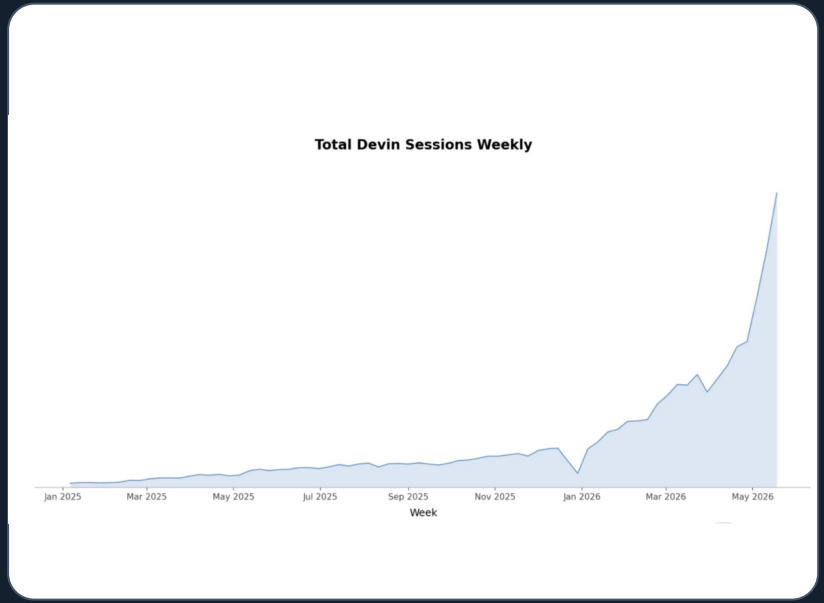
Three postures, and the question you leave with.

PREDICTION 1 · SCORED

THE BOTTLENECK MOVED FROM THE MODEL TO THE INFRASTRUCTURE LAYER

The same curve repeats at every frontier lab: slow, linear growth — then an exponential break after January. Models finally got good enough at tool-calling to finish the workflow, not just start it.

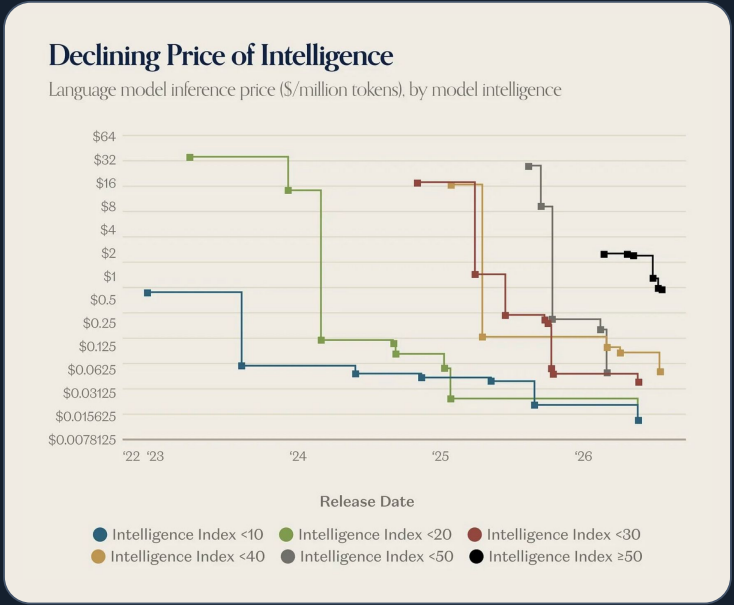
Dec 2025 · the Opus 4.5 moment
Agents crossed from experiments to accelerating, weekly production deployments.



VERDICT Half right. Missed the harness — and how fast it would matter.

PREDICTION 2 · SCORED

INTELLIGENCE TOO CHEAP TO METER



While frontier models have stayed at a relatively higher price point, we are seeing a bifurcation between models

The fastest decline we've seen is in the open source/distilled models. (Intelligence <20)

Many knowledge tasks, like summarization, formatting, LLMs as a judge, classification, extraction can now be offloaded to SLMs via subagents

Sub-agents can be called via sophisticated harnesses.

This pattern is now compounding.

VERDICT Mostly true. Flash-class models did last year's GPT-4 work — then open source made small models cheap to train and deploy.

PREDICTION 3 · SCORED

AMBIENT AGENTS ARRIVE

1

2023 · Copilot

Prompt-driven, ephemeral. Waits for a human to ask.

2

2025 · Workflow Agent

Scoped, evaluable. Runs defined processes on demand.

3

2026 · Ambient Agent

Always on, proactive. *They stopped* ***WAITING*** *for prompts.*

VERDICT Missed the acceleration. OpenClaw launched in 2026 — 150K stars, now an OpenAI project.

WHAT REALLY CHANGED? MODEL + HARNESS.

New Capabilities = New Jobs to be done

GATHER

01 News alert fires

02 Pull the filings

03 Read the transcript

JUDGEMENT

ANALYZE + REASON

01 Spin up a container

02 Call tools with skills

03 Pull data + execute code

04 Reason & self-critique

SHIP

01 Draft the memo

02 Get human approval

03 Ship the presentation

The agent doesn't describe the analysis — it spins up an **ephemeral container**, runs the **real code**, checks the output against the model, reasons through the delta, then tears the container down and leaves a **reproducible trace**.

Turns out the sacred part was just code no one would let a machine run — and check.

WHO SHIPS THIS

Every Successful Deployment We've Observed Has the Same Four Roles

1

Business Sponsor

Budget and air cover. Change management at insane speed.

2

Principal Engineer

In the weeds, with the mandate to ship.

3

Super Users

Own the workflows. Can write the evals.

4

Forward-Deployed Engineer

Implements the right harness and guardrails.

What we've seen struggle: centralized AI CoEs and AI committees. Central function. Separate floor. Steering committee. No ship mandate.

WHY 2026 IS THE YEAR EVALS COMPOUND

Swarms → Intelligence

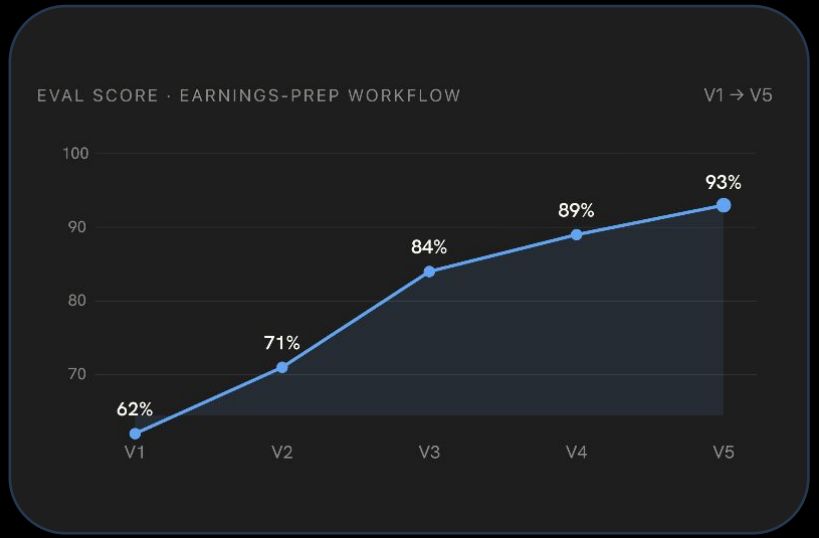
Agents decompose work and delegate to subagents.

Workflows → Scope

The SOP wrapper that turns one run into something repeatable.

Evals → The Hill

Package it correctly and every iteration compounds.



The PoC gap wasn't better models. It was the discipline to write the SOP — and the ability to verify it.

THE STANDARD OPERATING PROCEDURE FOR THE AGENT

A workflow is four things. If you can't write the SOP, you can't ship the agent.

Plan 01

What gets done, and in what order.

Skills 02

Packaged capabilities you hand the agent.

Knowledge & Context 03

What your firm knows that nobody else does.

Scope & Evals 04

Boundaries that make it evaluable.

THE PHYSICAL REALITY

Everything above assumes the agent can actually run. That's not a given. Three realities your CIO and CISO must solve.

1 Inference

Multi-vendor by design. No single provider kept its uptime promise this year.

"Agent isn’t responding! Your service sucks. I can’t trust it”

One outage from dead. You need a gateway and a fallback.

2 Sandbox

An agent without a sandbox is a loaded gun pointed at your stack. We've seen it go off.

“I’ve been waiting for a few hours and it still tells me its running. This is just too slow for me to work with"

Cost of a sandbox is small. Cost of no sandbox is your week.

3 Identity

SSO and 2FA were built for humans. Agents act 1,000× a minute and need their own model.

"Ummm. Why did you delete my database and every copy I saved!”

Your security stack wasn't built for agentic judgement.

AGENTS ARE NOT HUMANS — STOP TREATING THEM LIKE ONE

Agentic IAM is not optional. It is the gating constraint on every deployment after this point.

Dimension	Human identity	Agent identity
Action rate	~1 / min	**~1,000 / min**
Audit trail	Periodic	**Every step + reasoning**
Permission scope	Role-based	**Task-scoped, per call**
Blast radius	One mistake	**A loop × 1,000**
Credential model	SSO + 2FA	**Own identity. No reuse.**

If you're not modeling agents as first-class identities with their own permission graphs, you will either block your firm's AI rollout, or worse, you won't...and a loop runs your mistake a thousand times before lunch

INTERFACES VS. JOBS: THE SPLIT THAT DECIDES BUILD-VS-BUY

Interfaces → BUY

Let users pick. They're a commodity.

- Claude, ChatGPT, Gemini
- Day-to-day, ad-hoc use
- Switching cost: low
- Fail mode: a slow PM

Jobs → BUILD

Business processes can't be one provider deep.

- Earnings prep, portfolio monitors, ambient agents
- Harness-agnostic by design
- Switching cost: massive
- Fail mode: workflow dead

BUILD VS. BUY ISN'T ONE DECISION. IT'S THREE.

Everyone buys the models. Everyone builds the evals. The middle is your call.

POSTURE 01

The Front Door

"We want clients in our product, not someone else's."

BUILDS UX, workflows, evals

BUYS Orchestration, harness, models, data

POSTURE 02

The Adopter

"Just make it work."

BUILDS Integrations, evals

BUYS Everything else

POSTURE 03

The Builder

"We want primitives."

BUILDS Orchestration, harness, evals

BUYS Data, models, infra

THE QUESTION YOU LEAVE WITH

> You're one of these three. Have you built the team for it?

The hill is climbable now. The workflows are scopeable. The harnesses exist. The teams that ship have the trio: **sponsor, principal engineer in the business, super users with evals.**

If you don't have that team, you don't have a complete AI strategy. You have an AI press release.

Pro tip: We're here to help :)