



FDE CAPABILITY DOSSIER · ONE PAGE · WRITTEN TO BE FORWARDED · 7 OCTOBER 2026

Native AI in production, owned by your team

Vikram Jha · Principal FDE · AI Architect · Builder · 15+ years in production AI · Co-Founder at a US-based native-AI startup in stealth mode, working on AI governance, agentic security and context graphs.

WHY NOW

Analysts expect over 40% of agentic projects to be canceled on cost, value or control ([Gartner](#)) and vendor-built agents to be abandoned when customers cannot evolve them ([Gartner, Sep 2026](#)). The model is rarely the problem; ownership, evaluation, authority and cost are.

WHAT I DO

I work inside your team and build the whole stack with your engineers (context, models, agents, evaluation, security, deployment), prove it against your own baseline, and leave it running in your environment, owned by your people. If AI is not the answer, I say so in writing.

THE FOUR DELTAS, EACH WITH THE NUMBER THAT PROVES OR DISPROVES IT

1 One owner, full stack

Architecture, build and governance in one accountable person; no handoff loss

MEASURED BY Handoffs removed; days from question to decision

2 An evaluated slice in weeks

A working, measured slice of one workflow in 3-4 weeks of data access

MEASURED BY Weeks to a slice that meets its written threshold

3 Governance built in

Authority bands, permission receipts, red-team and a tested kill path before go-live

MEASURED BY Share of agent actions with an authority receipt

4 Open-source cost and sovereignty

Open-weight models wherever blind evaluation shows parity; data in-region

MEASURED BY Cost per resolved case against the frontier baseline

WHERE TO START

IF YOUR SITUATION IS...

START WITH

YOU DECIDE AT THE END

One workflow hurts, but nobody knows if AI is the answer	Workflow diagnostic (10 business days)	Pilot, simpler automation, process fix, or stop
A pilot demos well and will not ship	Pilot-to-production rescue (2-6 weeks)	Release, repair or retire it, on evidence
Launch is close and the risk owner is nervous	Go-live readiness and red-team (2-3 weeks)	Go, go with conditions, or hold
Too many ideas, no shared ranking	CXO and board AI portfolio sprint (1-2 weeks)	The first three use cases and their owners

WHAT YOU CAN HOLD ME TO

48 h

From your reply to a 30-minute scoping call

1 week

From a signed scope note to the first working session

3-4 wks

From data access to a first evaluated, working slice

2 days

Working days from any session to its written summary and next decision

WHAT OTHERS HAVE PUBLISHED ABOUT THEIR OWN RESULTS

- **JPMorganChase:** [double the screening volume, half the manual checks](#)
- **Permanente Medical Group:** [1,794 working days of documentation saved in a year](#)
- **Siemens:** [25% less reactive maintenance time in first pilots](#)
- **Amazon:** [4,500 developer-years saved upgrading to Java 17](#)

Self-reported or vendor-published, linked; they show what is achievable, not a prediction for you.

WHY ME

- Pucho, an Indic voice assistant (2014-2020): 40,000 hours of speech data augmented to about 600,000
- Generative drug design (2020): three COVID-19 candidates approved by NIV Pune for pre-clinical trials
- Twelve engagement shapes, from a 10-day diagnostic to a retained design authority
- Published reasoning on production agents at [vikramjha.work/writing](#)

NEXT STEP

A 30-minute scoping call, within 48 hours of your reply. Tell me the workflow or the stalled pilot; the written summary follows within two working days.

Book 30 minutes: [calendly.com/vikram01jha/30-mins](#)

Email: vikram@vikramjha.work

Writing: [vikramjha.work/writing](#)

WhatsApp / phone: [+91 73034 70111](tel:+917303470111)

LinkedIn: linkedin.com/in/invinciblejha

