

State of the AI Agent Space

six paradigms · one converging shape · where you fit.
may 2026 · grounded at ai dev sf, apr 28-29, 2026

file	state-of-space
ver	v1.0
date	2026-05-16
scope	builders

Walking the floor at AI Dev SF, the insight that hit me: everyone is rebuilding their agent harness from scratch.

THE HOOK

Karpathy. Tomby. Kashef. Bedrock. CodeRabbit. Six paradigms – independently arriving at the same five primitives.

THE SIX PARADIGMS · FRAGMENTING

P1 Personal IOS

Intelligent OS. Single operator, agent fleet, persistent memory. The system that runs your business.

WHO · Kashef · Lucas · Dan · Tomby

ME · home base. 33 agents · 124 skills · 28 hooks · 116 memory files. wake-check converts dormant capacity into action.

P2 Karpathy Minimalism

If I can't build it, I don't understand it. Frameworks are debt. 200 lines beats 20K.

WHO · Karpathy · microgpt · autoresearch (75K*) · nanochat · llm-wiki

ME · design philosophy. the LLM Wiki at business-ops/llm-wiki/ runs the append-only frontmatter pattern.

P3 Frameworks · Platform

Heavyweight orchestration, vendor lock-in. The enterprise default – declining among builders.

WHO · AWS Bedrock · Databricks · Snowflake · Temporal · LangChain · CrewAI

ME · avoided. the framework race optimizes for enterprise, not builders. memory is retrieval-LLM, not embeddings.

P4 Vertical SaaS Agents

One job, done well. Domain context = the moat. Cursor for X – recurring revenue.

WHO · Cursor · LandingAI · CodeRabbit · Chroma · Procore add-ons

ME · Struvo. voice → structured field reports. Per-report + Pro subscription tiers. In beta with a lead architecture firm.

P5 AAA · Productized Services

Five-client guarantees. Standardized chatbot offers. Volume play – build the agency, not the agent.

WHO · Kashef AAA · Saraev · Ottley · the AAA YouTube subculture

ME · adjacent. AFP is the cold-outreach surface – but the offer is upmarket: bespoke vertical builds.

P6 Hardware-Attached

Embodied agents. Voice ID, persistent memory, spatial intake. Front-of-house in the real world.

WHO · Pollen Robotics (Reachy Mini) · World Labs · Karpathy on nanochat-as-hardware

ME · Niero. tonight's mascot. plus the 5-7yr robotics endgame for Struvo – robot dispatch.

WHERE LUCAS FITS · THE INTERSECTION

P1 + P4 + P6

P1 · the meta-factory. IOS produces everything else. **P4** · Struvo. the product IOS is shipping today. **P6** · Niero. the front of house – what the world sees first.

Recursive: IOS gets stronger as Struvo ships. Niero is what people see first. Struvo is what closes the deal. IOS is what stays.

THE COMMON SHAPE · CONVERGING

Every paradigm builds on the same five primitives.

- 01

Memory > everything
What the agent remembers across sessions, machines, projects. Memory engineering is the emerging discipline. Not a vector index by default – frontmatter + LLM-as-retrieval is winning at the builder tier.
- 02

Verifiable execution
You can't automate what you can't verify. Karpathy's val_bpb. Struvo's entity-extraction F1. CodeRabbit's review-pass rate. Eval is the unsexy work that unlocks every loop above it.
- 03

Tool / skill library
Reusable, addressable capabilities. 124 skills here. MCP servers in Claude Code. Bedrock action groups. Cursor's tool integrations. Surfaces vary; the pattern is the same.
- 04

Routing / decision
Something decides what runs when. Wake-check hooks. Capability matrices. Thenvoi's interaction control plane. This layer is what separates an OS from a pile of scripts.
- 05

Front vs back of house
Customer-facing surface (chat, agent, mascot) is different from the engine. Niero out front, OS in the back. Cursor's chat UI vs the underlying agent runtime. Confuse them and you build the wrong thing.

YOUR MOVE · PICK THE READ THAT MATCHES YOU

IF YOU'RE A BUILDER

Pick a paradigm. Build the substrate.

- 1 Pick the paradigm closest to your shape – most people are P1 or P2.
- 2 Inventory: agents, skills, hooks, memory. Counts as visual emphasis.
- 3 Diagram it. Pair the diagram with a fact-base markdown. Ship the draft.

IF YOU RUN A BUSINESS

Build one bespoke instance. Use it. Decide.

- 1 One Personal Intelligent OS (IOS) – scaled to your business. Not a SaaS – your stack.
- 2 Use it 60-90 days. Steal what works. Throw out what doesn't.
- 3 Then choose: keep for yourself, or productize for your industry.

IF YOU'RE INVESTING

The shape is converging.

- 1 The space is fragmenting on UI; the substrate underneath is converging.
- 2 Bet on memory-engineering, eval primitives, routing layers – the picks-and-shovels.
- 3 P1+P4+P6 founders are rare and undervalued – recursive moat.

THE CLOSER

The space is fragmenting. The shape underneath is converging.

Pick your paradigm. Build the substrate. Ship the surface.
Compound for ten years.

HONEST SCOPE

Not a market map. Not a forecast. Six paradigms is the right number – fifteen is busywork.

GROUNDING IN

AI Dev SF (DeepLearning.AI), Apr 28-29 2026.
Karpathy github · Bedrock + AgentCore (Brooker, Chambers) · Databricks (Mathur) · Snowflake AI · Temporal (Herrera) · Cursor · LandingAI · CodeRabbit · Chroma · Pollen Robotics · World Labs · plus direct work with Tomby, Dan, Kashef.

THREE DOORS · PICK ONE



EXPLORE THE SYSTEM
lucface.com
SCAN WITH PHONE



BOOK A 25 MIN CALL
lunacal.ai/lets-go
SCAN WITH PHONE



NEXT SD AI STUDIO
sandiegoaistudio.com
SCAN WITH PHONE