

THE 2026 AI CAPABILITY DELTA MAP

An empirical mapping of what actually shipped & shifted across frontier labs in 2026

CRUCIAL METHODOLOGY & SCOPE

An empirical delta of 2026 shipments — not a baseline capability encyclopedia.

This document represents an empirical delta map tracking **50 observable capabilities that publicly shipped, previewed, or were scheduled for retirement in 2026** across the four frontier foundation labs. It explicitly excludes developer-tool aggregators (like GitHub Copilot) to evaluate primary platform bets on their own architectural merits.

8

L1 DOMAINS

Core enterprise jobs

32

L2 CLUSTERS

Architectural families

50

L3 CAPABILITIES

Observed shipments

4

FRONTIER LABS

Primary platform targets

Key Findings from the 2026 Release Feeds:

- **Commoditized Foundation:** >60% feature overlap between Anthropic and OpenAI (31/50 items shared).
- **The State Trap:** Zero vendors offer export paths for persistent memory, locking customer history into walled gardens.
- **Battle Moves to the Harness:** >50% of observed shipment notes focus on execution loops, IDEs, and CLI tools.
- **Active Deprecations:** Q4 2026 is an aggressive retirement quarter, forcing decoupling and middleware portability.

THE 3 LEVELS OF CAPABILITY

How the modern AI stack is engineered, scored, and decoupled

LEVEL
L1

Macro Domains (8 Market-Level Jobs)

WHERE BUDGET FLOWS

The fundamental problem spaces every enterprise AI platform must deliver: Foundation Intelligence (FND), Agentic Autonomy (AGT), Software Engineering (SWE), Interaction Surfaces (INT), Memory & Context (MEM), Trust & Governance (TRU), Platform Distribution (PLT), and Domain Verticals (DOM).

Architecture Impact: Establishes the 8 enterprise boundaries where budget, compliance, and systems integration live.

LEVEL
L2

Architectural Clusters (32 Capability Families)

HOW SYSTEMS CONNECT

Functional groupings of technical capabilities that solve distinct operational challenges. Examples include Model Families (FND-FAM), Long-Horizon Workers (AGT-LONG), Native Coding CLIs (SWE-CLI), Context Compaction (MEM-CTX), and Zero-Retention Safety (TRU-CMP).

Architecture Impact: Defines where vendor interfaces are standardized (e.g. MCP) vs. proprietary walled gardens.

LEVEL
L3

Atomic Capabilities (50 Observable Deliverables)

WHAT ACTUALLY SHIPPED

Discrete, testable capabilities scored Y/N based on publicly shipped or announced 2026 releases. Examples: Selectable Reasoning Sliders, Computer/Browser Use, Claude Tag in Slack, Dreaming Memory, EFS Zero-Retention, and 500k-class context windows.

Architecture Impact: Separates real, working production code from aspirational keynote demos and marketing claims.

STRATEGIC INSIGHT

Why the 3-level structure matters: Evaluating vendors only at L1 leads to naive multi-model contracts. Evaluating at L3 proves that while models are interchangeable, **runtime harnesses (L2/L3) and memory layers (L2/L3) are where severe vendor lock-in occurs.**

2026 shipments across Foundation Intelligence (FND) and Agentic Autonomy (AGT)

CODE	L3 CAPABILITY & 2026 SHIPMENT SCOPE	ANTHROPIC	OPENAI	GOOGLE	XAI
FND — FOUNDATION INTELLIGENCE (8 SHIPPED CAPABILITIES IN 2026)					
FND-FAM-01	Frontier flagship generation New top-tier general model (Fable/Mythos 5.x, GPT-6 Astra, Grok 4.6, Gemini 3.8).	●	●	●	●
FND-FAM-02	Mid / default workhorse Default chat or coding model for most seats (Sonnet 5, GPT-5.6 Sol, Gemini 3.8 Flash).	●	●	●	○
FND-FAM-03	Generation retirement Scheduled removal of prior generation from products.	○	●	○	○
FND-RSN-01	Selectable reasoning effort User/API control of think-time (low-xhigh, Instant vs Thinking, max/ultra).	●	●	○	●
FND-RSN-02	Live reasoning summaries Streaming or status-row traces of model thinking.	○	●	○	○
FND-TIER-01	Fast / cheap dual SKU Speed or cost twin of a flagship (Fast mode, Luna, Flash).	●	●	●	●
FND-TIER-02	Token price reset Published in/out price change or intro discount.	●	●	○	●
FND-SPEC-01	Domain-tuned model Law, cyber, life-science, or live-audio SKU.	●	●	○	○
AGT — AGENTIC AUTONOMY (7 SHIPPED CAPABILITIES IN 2026)					
AGT-LONG-01	Multi-hour agent sessions Sessions that survive compaction, disconnect, or cloud handoff.	●	●	●	●
AGT-LONG-02	Cloud-hosted agent runtime Vendor runs the loop off-device (Agents API, cloud sessions, CCA).	●	●	○	○
AGT-TOOL-01	Computer / browser use Click, type, navigate, fill forms, site tools / WebMCP.	●	●	●	○
AGT-TOOL-02	MCP and external tools Model Context Protocol servers, connectors, tool result hooks.	●	●	●	●
AGT-MULTI-01	Subagents and parallel threads Coordinator + workers, nested agents, project threads.	●	●	●	○
AGT-MULTI-02	Task objects and @mentions First-class tasks agents can create, hide, archive, message.	○	●	○	○
AGT-WORK-01	Persistent named bots Always-on teammates with identity and memory (Grok Bot, Scout, Tag).	●	○	○	●

Commodity Floor: Reasoning controls, MCP tool use, and fast/cheap twin models have reached complete parity across the labs. Differentiation has moved to hosted agent runtimes (Agents API) and persistent named teammates (Claude Tag, Grok Bot).

DEVELOPER HARNESS & SURFACES

2026 shipments across Software Engineering (SWE) and Interaction Surfaces (INT)

CODE	L3 CAPABILITY & 2026 SHIPMENT SCOPE	ANTHROPIC	OPENAI	GOOGLE	XAI
SWE — SOFTWARE ENGINEERING (7 SHIPPED CAPABILITIES IN 2026)					
SWE-IDE-01	In-editor agent host VS Code / Antigravity / Copilot app agent window.	●	●	●	●
SWE-IDE-02	Multi-harness IDE Same editor runs Copilot, Claude, Codex, or Gemini side by side.	○	●	●	○
SWE-CLI-01	Native coding CLI / TUI Claude Code, Codex TUI, Copilot CLI, Grok Build CLI.	●	●	●	●
SWE-REV-01	Agentic code review PR comments, auto-resolve, generated commit messages.	●	○	○	○
SWE-REV-02	PR creation from agent Agent session opens the pull request.	○	○	○	○
SWE-GIT-01	Git / worktree native Sidebar git, worktrees, path templates.	●	○	●	○
SWE-GIT-02	Hardened sandbox runtime OS sandbox, Rust rewrite, Windows/macOS permission prompts.	●	●	○	○
INT — INTERACTION SURFACES (6 SHIPPED CAPABILITIES IN 2026)					
INT-CHAT-01	Unified chat + work app Cowork merged with chat; ChatGPT Work; Gemini notebooks.	●	●	●	●
INT-CHAT-02	Slack / team channels Taggable agent in shared channels.	●	●	○	○
INT-VOICE-01	Realtime voice conversation Full-duplex or experimental /voice.	○	●	●	○
INT-VOICE-02	Speech-to-text product Dedicated STT model (Grok Voice Transcribe 2.0).	○	○	○	●
INT-BRW-01	First-party browser product Claude in Chrome, Cowork browser, ChatGPT Atlas, Antigravity Remote Control.	●	●	●	○
INT-DOC-01	Office artifact generation Docs, slides, sheets, Word add-in, generative UI.	●	●	●	●

PATTERN READ

Battle for the Execution Loop: All 4 labs have shipped native CLIs and editor agents. Google concentrates effort on the Antigravity IDE stack, OpenAI captures desktop knowledge workflows, and Anthropic focuses on terminal agent execution and sandbox safety.

MEMORY, GOVERNANCE & VERTICALS

2026 shipments across Memory (MEM), Trust (TRU), Platform (PLT) & Domains (DOM)

CODE	L3 CAPABILITY & 2026 SHIPMENT SCOPE	ANTHROPIC	OPENAI	GOOGLE	XAI
MEM — MEMORY & CONTEXT (7 SHIPPED CAPABILITIES IN 2026)					
MEM-USR-01	Cross-surface memory Memory shared across chat, cowork, build; editable topics.	●	●	○	●
MEM-USR-02	Dreaming / consolidated memory Offline synthesis of long-term user memory.	○	●	○	○
MEM-CTX-01	Long context window Published 500k-class windows.	○	○	○	●
MEM-CTX-02	Compaction and resume Token-budget compaction, session fork/resume.	●	●	○	○
MEM-SKL-01	Skills / plugins marketplace Installable skills, plugin CLI, marketplaces.	●	●	○	○
MEM-SKL-02	Repo instruction files AGENTS.md / CLAUDE.md project instructions.	●	○	●	○
MEM-PRJ-01	Project workspace objects Projects with library, shared memory, parallel threads.	●	○	●	○
TRU — TRUST, SAFETY & GOVERNANCE (7 SHIPPED CAPABILITIES IN 2026)					
TRU-SAF-01	Capability classifiers Cyber / bio dual-use routing and fallbacks.	●	○	○	○
TRU-SAF-02	Alignment disclosure Misalignment reports, jailbreak framework, system cards.	●	●	○	○
TRU-IDM-01	Enterprise identity SSO, SCIM, MDM, GitHub Enterprise auth.	●	●	●	●
TRU-CMP-01	Regulatory compliance pack Watermark / C2PA, EU AI Act, FERPA, youth safety, HIPAA path.	●	●	○	○
TRU-CMP-02	Zero-retention / customer-cloud safety EFS-style privacy plus misuse detection.	●	○	○	○
TRU-OPS-01	Usage and ROI analytics Admin consoles, impact dashboards, feature engagement.	●	●	●	○
TRU-OPS-02	Budget and credit controls Seat limits, budget-increase requests, token pricing meters.	○	●	○	●
PLT — PLATFORM DISTRIBUTION (4 SHIPPED CAPABILITIES IN 2026)					
PLT-CLD-01	AWS / Azure / GCP listing Bedrock, Foundry, Vertex / Agent Platform.	●	●	●	●
PLT-PRT-01	Third-party coding host Shipped inside Copilot, Cursor, VS Code agent host.	●	●	○	●
PLT-API-01	Public agent / skills API Skills API, Files API, Agents API, AHP.	●	●	●	●
PLT-PKG-01	Plan and SKU expansion New seats (Teachers, SMB, Bot on more plans).	●	●	●	●
DOM — DOMAIN APPLICATIONS (4 SHIPPED CAPABILITIES IN 2026)					
DOM-PRO-01	Legal / sales / finance pack Astra for Law, Salesforce skills, Copilot Excel finance.	●	●	○	○
DOM-SCI-01	Science and cyber programs Claude Science, LSVP, Mythos cyber, hardware standard.	●	○	○	○
DOM-EDU-01	Education / teen products Claude for Teachers, ChatGPT for Teens / EDU.	●	●	●	○
DOM-MED-01	Image, ads, or media API Grok Imagine, ChatGPT Ads, Gemini Live audio.	○	●	●	●

PATTERN READ

The Sovereignty Chasm: Memory and Governance have the lowest cross-lab parity. Zero export paths exist for vendor memory. Anthropic leads on zero-retention (EFS) and enterprise steering; OpenAI leads on consumer memory & legal/word tooling.

WHERE THE 4 LABS PUT THEIR CAPITAL

Distinct architectural bets across the 2026 shipments

Anthropic

39 / 50 (78%)

The Enterprise Control Plane & Safety Harness

Harness: **46%** Enterprise/Trust: **28%** Model: **13%**

- **Core Bet:** Claude Code TUI, computer use, and deterministic enterprise steering.
- **Unique 2026 Releases:** Claude Tag in Slack, EFS zero-retention privacy, Life-Sciences & Hardware Standards (LSVP).
- **Architectural Role:** The safest anchor for regulated workflows and deep developer execution.

OpenAI

39 / 50 (78%)

The Ubiquitous Workflow & Knowledge Layer

Harness: **44%** Enterprise/Trust: **26%** Model: **18%**

- **Core Bet:** End-to-end coverage across consumer, office, and developer loops.
- **Unique 2026 Releases:** Astra for Law, Word artifacts, Dreaming Memory, Live-I voice, ChatGPT Ads, and Agents API as a billed loop.
- **Architectural Role:** The highest-distribution ecosystem with aggressive vertical capture.

Google

24 / 50 (48%)

The Developer IDE & Workspace Gravity

Harness: **54%** Enterprise/Trust: **21%** Model: **12%**

- **Core Bet:** First-party developer tooling and enterprise workspace data integration.
- **Unique 2026 Releases:** Antigravity IDE (2.4-2.11) with native terminal/git control, Generative UI, and Gemini 3.8 Live audio.
- **Architectural Role:** Winning developer mindshare through integrated environments rather than pure model APIs.

xAI

21 / 50 (42%)

Raw Compute Density, Media & Scale

Harness: **38%** Enterprise/Trust: **29%** Model: **19%**

- **Core Bet:** Massive context capacity, speed, and real-time multimodal media generation.
- **Unique 2026 Releases:** Grok 4.6 multi-cloud blast, 500k context window, Imagine image API, and priced Voice Transcribe 2.0 STT.
- **Architectural Role:** The high-throughput alternative for media processing and uncensored data feeds.

STRATEGIC TAKEAWAY

Divergent Postures: While Anthropic builds governance guardrails and Google builds IDEs, OpenAI fights to own the workplace desktop and xAI fights for raw compute volume.

THE 2026 SHIFT: MODELS TO RUNTIMES

Three structural shifts and three architectural imperatives for tech leaders

01 The Foundation Has Converged (>60% Shared Floor)

Anthropic and OpenAI share 31 of 50 capabilities. All 4 labs now sell high-reasoning models, native CLIs, MCP tools, and price ladders. **Raw intelligence is a commodity.** The real competition is who captures the runtime where agents execute.

02 The Three Shipment Waves of 2026

Wave 1 (Q1–Q2): Models became agents (Sonnet 5, Codex tooling). **Wave 2 (Jul–Aug):** Distribution & runtime (Antigravity IDE, computer use GA, multi-cloud). **Wave 3 (Sep–Oct):** Vertical workplace lock-in (Law, Salesforce) alongside scheduled endpoint retirements.

03 Q4 Deprecations: The Portability Test

October brings major model retirements: GPT-5.5 (14 Oct), Grok 4.5, and preview cutoffs. **Specialized models do not switch overnight.** If your middleware cannot swap an endpoint in an afternoon without rewriting logic, you have engineered a single point of failure.

The Core Architectural Imperative:

Vendor memory is a convenience trap. The system of record must remain yours.

License horizontal foundation models from 1–2 anchor labs for operational resilience. But build, version, and own your business rules, evaluation harnesses, plain-text playbooks, and state internally. Everything else should be swappable.