The State of the AI Second Brain in Mid-2026: A Field Review for Operational Leaders

Prepared by Perth AI Consulting using Anthropic's Claude (deep-paper pipeline: Fable 5 drafting, Opus 4.8 multi-pass fact-check, plus an independent spot-check)

Version 1.0 (fact-checked) · 5 July 2026

Perth AI Consulting
perthaiconsulting.com.au

> *Prepared as a PAC resource. Research completed 5 July 2026. All web sources accessed July 2026 unless otherwise noted.*
>
> *This paper has been through a three-pass independent fact-check plus an independent spot-check. The pre-fact-check draft is preserved at archive/the-state-of-brains-july-2026-v0.9-pre-fact-check.md; every correction is logged in the Corrections Log appendix.*

Abstract

A "brain", in the sense this paper surveys, is a folder of plain markdown files that an AI agent with filesystem access reads, writes and maintains: raw material captured in one layer, processed knowledge in another, an instruction file governing the agent's behaviour, and git as the record. As of July 2026 the pattern is real, growing fast, and younger than it looks. The enabling stack assembled between November 2024 and January 2026: the Model Context Protocol (Anthropic, November 2024), file-native terminal agents (Claude Code, February 2025; OpenAI Codex CLI, April 2025; Gemini CLI, June 2025), instruction-file conventions (CLAUDE.md from February 2025, AGENTS.md from August 2025), and a non-developer desktop agent (Anthropic's Cowork, January 2026). Practitioners were building agent-maintained markdown knowledge bases through 2025; Andrej Karpathy's X post of 2 April 2026 and gist of 4 April 2026 named the pattern, supplied its cleanest articulation, and carried it into mass awareness. Google formalised the same pattern as a draft Open Knowledge Format on 12 June 2026.

The honest state of the evidence is this paper's central finding. The substrate is sound: markdown plus git plus an agent under an explicit instruction file is cheap, portable and auditable, and multiple unconnected practitioners converge on the same architecture. But the field is roughly three months old at survey date in its popular form, no independent controlled measurement of aggregate time saved exists anywhere, the best longitudinal report (one month in) calls the effort "roughly a wash", and the field's one controlled experiment (a narrow A/B test of an agent wiki's token economics) found that a wiki mirroring already-greppable sources delivered negative value. Agent maintenance does not abolish the write-only graveyard that killed earlier personal knowledge systems; it changes the failure modes, adding hallucinated notes, taxonomy drift and a documented security attack surface. Several widely repeated origin claims did not survive checking, and two leads this paper was handed that looked most like fabrications turned out to be real. The paper closes with a practical posture for operators: adopt the substrate, pilot the automation under a human gate, treat security as the gating constraint, and ignore every unmeasured multiplier claim.

Introduction

This paper is PAC's reference account of the state of "brains" as of July 2026: personal and business knowledge systems built from plain markdown files and maintained by AI agents. It is written for sophisticated operators and practitioners, and for PAC itself as the wellspring from which downstream writing on this subject is drawn. It is not an academic literature review, because almost no academic literature exists here; it is a survey of a practitioner-led field, sourced from the artefacts practitioners actually produce: posts, gists, repositories, plugin directories, vendor announcements and specifications.

The method matters as much as the content. Every capability claim is dated. Every source is labelled as the vendor's own claim or independent evidence. In a field this young, the source hierarchy is adapted: a reproducible artefact that can be inspected (a repo, a spec, a documented workflow) outranks an assertion; several unconnected practitioners converging on the same practice outranks one influential voice; a single influencer's claim is a lead to verify, not a fact; and anything that could not be traced to a credible primary source sits in Appendix B rather than in the body. This run also began with an adversarial input: a draft on the same topic produced by another model (Grok, xAI), used strictly as a source of leads. Its specifics were checked one by one, and the results cut both ways; section 1 and Appendix B record which of its claims held, which collapsed, and which turned out to be truer than a sceptic would have guessed.

Throughout the survey, each pattern and tool receives the same four-part judgement, tuned to this field: what practitioners actually sustain in real use; what looks good in a demo or a blog post but decays into a write-only graveyard; what is oversold; and what is underrated. That judgement, not the inventory, is the point of the paper.

Two scope notes. First, "brain" is used here for the artefact (the files) plus the agent arrangement around it, not for any vendor's memory feature, though vendor memory features appear where they bear on the pattern. Second, the paper describes a moment. The popular form of this field was three months old when the survey was written, and the sections on tooling and trajectory will age fastest.

1. Lineage and convergence

1.1 The event: Karpathy, 2 to 4 April 2026

The surge that made "brains" a mainstream idea has a precise, verifiable centre. On 2 April 2026 Andrej Karpathy, then at Eureka Labs, published an X post titled "LLM Knowledge Bases", opening: "Something I'm finding very useful recently: using LLMs to build personal knowledge bases for various topics of research interest. In this way, a large fraction of my recent token throughput is going less into manipulating code, and more into manipulating..." (Karpathy, 2 Apr 2026; the post could not be fetched in full, and the date is derived from the post ID's embedded timestamp, consistent with a same-day independent summary by DeepakNess on 3 April 2026).

Two days later, on 4 April 2026 at 16:25 UTC, he published the durable artefact: a GitHub gist titled llm-wiki.md, created in a single revision and carrying over 5,000 stars and over 5,000 forks by July 2026 (GitHub caps the display at "5,000+"). The gist was fetched in full for this paper, and its architecture is worth quoting precisely, because most retellings blur it. Three layers: raw sources ("your curated collection of source documents... These are immutable — the LLM reads from them but never modifies them. This is your source of truth"); the wiki ("a directory of LLM-generated markdown files... The LLM owns this layer entirely... You read it; the LLM writes it"); and the schema ("a document (e.g. CLAUDE.md for Claude Code or AGENTS.md for Codex) that tells the LLM how the wiki is structured... it's what makes the LLM a disciplined wiki maintainer rather than a generic chatbot"). Three operations: ingest, query, lint. Two special files: index.md, a catalogue the agent navigates from, and log.md, an append-only chronological record. The gist is explicitly anti-RAG at personal scale: index-first navigation "works surprisingly well at moderate scale (~100 sources, ~hundreds of pages) and avoids the need for embedding-based RAG infrastructure". And it is explicitly a pattern, not a product: "It describes the idea, not a specific implementation" (Karpathy, 4 Apr 2026).

The "tractable brain upload" framing that circulates is real but second-order: it comes from a reply Karpathy posted around 10 April 2026 agreeing with another user's framing ("Yes it's the tractable form of brain upload..."), eight days after the original post. The gist itself never uses the words "brain" or "upload". Reception numbers are softer than the retellings: view counts of 16 to 21 million circulate in secondary coverage and disagree with each other, and X view counts are not externally auditable. The gist's star and fork counts are the auditable signal. One more point of record: Karpathy joined Anthropic on 19 May 2026, six weeks after publishing; at the time of the post he had no vendor affiliation in this space.

1.2 A lead that could not be traced

The brief for this paper carried a widely retold attribution: that a step-by-step guide by an X account called @0xMiraqle prompted the replication wave. Two independent research passes found no trace of the account, the guide, or any secondary coverage crediting it. The role it describes is filled, on the evidence, by the account @godofprompt ("God of Prompt"), whose 6 April 2026 X article (519,000 views and 6,700 bookmarks per an archived copy with engagement data) was exactly that: a copy-paste replication guide of Karpathy's pattern, with a full schema template and ingest, query and lint prompts. The account markets prompt products, noted here as an interest label in line with this paper's method, not as a judgement. The untraceable attribution is recorded in Appendix B and stands open to correction if a primary source surfaces.

1.3 Convergent synthesis: the pattern before the name

Nothing in the primary record frames April 2026 as an invention, least of all the gist itself, which says plainly: "It describes the idea, not a specific implementation." What the record shows instead is convergent synthesis: through 2025, practitioners in different corners of the field, mostly without reference to one another, assembled the same pattern because the same newly available tools made it the natural thing to build. Karpathy's contribution sits on top of that convergence, and is the more interesting for being locatable exactly:

The markdown-as-LLM-interface move dates at least to llms.txt, Jeremy Howard's proposal of 3 September 2024 (Answer.AI): a curated markdown index at a domain root for LLM consumption. The agent-maintained markdown memory pattern was in mass circulation by February 2025: Anthropic's CLAUDE.md convention shipped with Claude Code's research preview (24 February 2025), and the Cline "Memory Bank" pattern (published 6 February 2025) had an agent read and update a hierarchy of markdown files at the start of every task. The personal-scope version existed as a product by March 2025: Basic Memory (Basic Machines, released March 2025) stored an AI-and-human-shared knowledge graph in plain markdown files with Obsidian integration, a full year before the Karpathy post (the repo's README has since been rewritten; the description here follows its 2025 positioning). Individual practitioners articulated the life-system version through 2025: Steve Faulkner's much-shared prompt of 4 August 2025 ("You are my personal assistant and chief of staff. Self organize using markdown files"); Noah Brier's documented Claude Code plus Obsidian consultancy workflow (Every, 10 September 2025); and by early 2026 the Obsidian-plus-agent second brain was, in one independent practitioner's words, "one of the PKM community's biggest trends" for the preceding six months (Yu, 2026). Kenneth Reitz published the most concrete pre-Karpathy architecture on 6 March 2026: a 467-file vault with a roughly 200-line CLAUDE.md framed as an "API contract", including an explicit list of what the agent must not do. At mass scale, OpenClaw (first released as Clawdbot by Peter Steinberger in November 2025; renamed twice by early 2026; reported at roughly 247,000 GitHub stars by early March 2026, an archived-snapshot figure via Wikipedia consistent with the 382,000 observed live on 5 July 2026) normalised plain markdown files (MEMORY.md, SOUL.md, USER.md) as agent

memory for a very large audience.

The fair summary, on the primary record: the elements (markdown substrate, agent write access, instruction file, raw-versus-processed split) were each established practice by mid-2025 to early 2026, arrived at convergently across unconnected builders. Karpathy then built on that base, some of it explicitly (his gist cites Bush's Memex by name and adopts the existing CLAUDE.md and AGENTS.md conventions as its schema layer) and some of it convergently, and added the parts the pattern had lacked: the compiler articulation (an immutable raw layer compiled by the agent into a wiki it owns), the ingest-query-lint operations loop as a named triad, and the metaphors that made the whole thing legible ("Obsidian is the IDE; the LLM is the programmer; the wiki is the codebase"). What he supplied above all was a name, the cleanest articulation, and reach.

1.4 The deep lineage, pinned

The intellectual ancestry claimed for brains is largely real, and better documented than most retellings bother to show.

Vannevar Bush described the Memex in "As We May Think" (The Atlantic, July 1945): a personal device storing "all his books, records, and communications", navigated by associative trails. The connection to brains is not retrofitted; Karpathy's gist cites Bush directly, with the sharpest single line in the lineage: "The part he couldn't solve was who does the maintenance. The LLM handles that."

Niklas Luhmann built the paper prototype: a Zettelkasten of approximately 90,000 slips written between 1951 and 1996, a figure documented by Bielefeld University's Niklas Luhmann-Archiv, which has digitised the estate. The commonly repeated claim that the method caused his prolific output is interpretation, not record; the slips, the numbering protocol and his 1981 essay framing the box as a "communication partner" are documented.

The modern popular layer is Tiago Forte's: the PARA method (first published February 2017 on the Forte Labs blog) and Building a Second Brain (Atria Books, 14 June 2022), with the CODE workflow (Capture, Organise, Distill, Express). Forte's own framing is synthesis of older practice (commonplace books, Zettelkasten) into a mass-market method, and that is the right reading: Forte built on the commonplace-book and Zettelkasten traditions, added the PARA and CODE frameworks and the second-brain name, and scaled the practice to an audience those traditions had never reached. Sönke Ahrens's How to Take Smart Notes (2017) carried Luhmann into the same Anglophone wave, and Nick Milo's Maps of Content (Linking Your Thinking, around 2020) supplied the Obsidian-era linking pedagogy. Behind all of it sit the standard hypertext ancestors: Ted Nelson (the word "hypertext", 1965) and Doug Engelbart (NLS, demonstrated December 1968). The substrate philosophy that made markdown the natural medium is well captured by Obsidian CEO Steph Ango's "File over app" essay (July 2023).

One lineage chain, then, with every link datable: Bush (1945, the concept, maintenance unsolved) to Luhmann (1951 to 1996, the manual protocol) to Ahrens and Forte (2017 to 2022, the mass-market method) to the plain-text local-first movement (Obsidian from 2020, "File over app" 2023) to markdown as LLM interface (llms.txt, September 2024) to agent-maintained markdown memory (CLAUDE.md and Cline Memory Bank, February 2025; Basic Memory, March 2025; AGENTS.md, August 2025; OpenClaw, November 2025) to Karpathy naming the pattern (April 2026) and Google moving to standardise it (June 2026).

2. Evolution: how the idea moved from infrastructure to movement

2.1 The enabling stack, November 2024 to January 2026

The brain pattern is downstream of four pieces of infrastructure, each with a primary-source date.

First, connection. Anthropic open-sourced the Model Context Protocol on 25 November 2024, and the same announcement gave Claude Desktop local MCP server support, including a filesystem server: the first mainstream path for a chat assistant to read and write local files. OpenAI adopted MCP on 26 March 2025 ("people love MCP and we are excited to add support across our products", Altman), Google on 9 April 2025 ("rapidly becoming an open standard for the AI agentic era", Hassabis), and on 9 December 2025 Anthropic donated MCP to the newly formed Agentic AI Foundation under the Linux Foundation, co-founded with OpenAI and Block.

Second, file-native agents. Claude Code shipped as a research preview on 24 February 2025 and went generally available on 22 May 2025: a terminal agent that reads, edits and organises files in a working directory. OpenAI's Codex CLI followed on 16 April 2025, Google's Gemini CLI on 25 June 2025. These made "point an agent at a folder of markdown" a one-command act.

Third, instruction files. The CLAUDE.md convention (a markdown file the agent reads at session start) shipped with Claude Code and was canonised in Anthropic's "best practices" post of April

2025. AGENTS.md, the cross-vendor equivalent, launched in August 2025 (OpenAI with Google, Cursor, Factory and others) and was donated to the same Linux Foundation body in December 2025, with adoption in over 60,000 open-source repositories claimed by its stewards (a consortium figure, linked to a reproducible GitHub search but not independently re-counted here). The significance is hard to overstate: the instruction file is the brain pattern's load-bearing component, and by late 2025 it was a multi-vendor standard.

Fourth, the non-developer on-ramp. Anthropic's Cowork, a desktop agentic mode that gives Claude access to local folders without a terminal, entered research preview on 12 January

2026 (independently reviewed the same day by Simon Willison) and went generally available on all paid plans on 9 April 2026, per independent trade coverage (eWeek and others); Anthropic's own announcement page could not be machine-read during this paper's checks. Cowork matters to this story because it moves the brain pattern's addressable audience from developers to everyone who can pick a folder.

2.2 The practitioner wave, 2025

Between these infrastructure dates, practice accumulated. The Cline community's Memory Bank (February 2025) is the earliest widely replicated instance found of an agent instructed to read and maintain a set of markdown memory files as a standing discipline. Basic Memory (March 2025) productised the personal knowledge graph in markdown. Through mid and late 2025, practitioner posts describing Obsidian vaults driven by Claude Code multiplied; by February 2026, guides like WhyTryAI's "Build Your Second Brain With Claude Code and Obsidian" (26 February 2026) treated the combination as established practice, and Kenneth Reitz's March 2026 essay described a two-year-old vault practice recently upgraded with an agent contract. In parallel, OpenClaw's rapid growth (November 2025 to March 2026) taught a mass audience two things at once: that plain markdown files are a workable memory substrate for a persistent agent, and, through a well-documented security crisis examined in section 4.6, what happens when that substrate is exposed without discipline.

The 2023 predecessor wave deserves a paragraph, because its failure defines what changed. The first "AI plus notes" cycle (embeddings plugins such as Smart Connections, released January 2023, and chat-with-your-vault tools) gave models read access and retrieval, not write access and maintenance duties. Those tools did not die (Smart Connections stood at roughly a million cumulative downloads by mid-2026); they hit a ceiling as search tools, because retrieval-augmented chat "can't create new notes, can't cross-reference, and can't flag contradictions", as one practitioner-builder put it in 2026. Karpathy's gist makes the same diagnosis of RAG systems: "Nothing is built up." The 2025 agent layer, not better embeddings, is what unlocked the brain.

2.3 The viral moment and after, April to July 2026

The Karpathy post produced a measurable replication wave: implementation repos within days (among them lucasastorian/llmwiki, Astro-Han/karpathy-llm-wiki, and an Obsidian community plugin implementing the pattern), at least four Show HN threads through April, a guide industry (the @godofprompt article of 6 April 2026 being the largest), and continuing activity in the gist's own comment thread into late June 2026, where practitioners were reporting named failure modes: taxonomy drift ("an unconstrained LLM will generate company, Company, Business, and Organization across different runs"), token burn from agents re-reading index files, and agents repeating approaches that earlier sessions had ruled out. Astro-Han's implementation repo, at roughly 446 stars in mid-April per an aggregator ranking, was observed directly at 1,400 stars on 5 July 2026.

Institutional responses followed within weeks. Obsidian's CEO published official Agent Skills for Obsidian formats in January 2026 (before the Karpathy post, and further evidence the trend predates it), and Obsidian overhauled its plugin directory with automated security review in May 2026, explicitly because "coding agents accelerate the creation of plugins". On 12 June 2026, Google Cloud published the Open Knowledge Format, a draft specification that "formalizes the LLM-wiki pattern into a portable, interoperable format", citing Karpathy's gist by name. Whether OKF wins or not, a hyperscaler standardising a three-month-old community pattern is the clearest available signal of where the field sits in July 2026: past the toy stage, well short of the settled stage.

The sober counterweight arrived on schedule too. The most useful longitudinal report published to date (Brian Buntz, R&D World, 11 June 2026, one month of sustained use, roughly 760 pages maintained with citation-required schema rules and lint gates) concludes that "the time I spend maintaining the wiki and the time it saves me are roughly a wash", with the payoff concentrated in narrow topics the open web covers poorly. That verdict, from a disciplined practitioner a month in, is the field's honest centre of gravity at survey date.

3. The current state of the brain itself

3.1 The core pattern

As of July 2026, a working brain converges on five elements, each evidenced across multiple unconnected practitioners.

Plain markdown as substrate. The arguments practitioners give are consistent: models emit and parse markdown natively; the files are greppable, diffable and version-controllable; nothing but files is load-bearing, so the system survives tool churn; and a human can audit everything. The counterweight is scale, treated in 3.3.

A layered architecture with a trust boundary. Karpathy's raw-versus-wiki split is one instance of a more general convergence: an untrusted capture layer (inbox, raw sources, clipped pages) separated from a trusted, processed layer, with some gate between them. Documented variants include nightly ingestion pipelines that move clipped notes from an inbox to a wiki after a cheap model summarises and tags them (James Wilkins, April 2026); "proposal-first" agents that suggest but never commit (several independent repos); guarded intake paths where every automated edit writes an auditable ledger entry and only reviewed pages publish (Buntz, June 2026); and OpenClaw's consolidation pipeline, which promotes daily notes into curated memory only through scored thresholds with a human review surface. The draft-then-accept gate this document itself lives under is the same pattern.

An instruction file as the system's constitution. CLAUDE.md, AGENTS.md or equivalent: the file that tells the agent what the structure means, what it may touch, and how to behave. Karpathy calls it "what makes the LLM a disciplined wiki maintainer rather than a generic chatbot"; Reitz

frames his as an API contract whose most valuable section is the list of prohibitions ("Do not create new files unless explicitly asked… Do not restructure existing folder hierarchies"). Practitioners repeatedly identify this file as the highest-value artefact in the system, and it is where a brain's owner encodes judgement. It is also, as section 4.6 notes, a security boundary in name only.

Git as the record. Version history, rollback insurance against agent mistakes, and an audit trail. The obsidian-git plugin's 2.77 million downloads (observed 5 July 2026) are substantially agent-insurance; Reitz states the norm plainly: "Your second brain deserves the same version control discipline as your code."

Operating loops. Ingest (new material enters, gets processed, linked and logged), query (the agent reads the index first, then drills into pages), and maintenance (lint passes that flag contradictions, staleness, orphan pages and drift). The loops are where brains differ most from classic PKM: the human's job shifts from writing notes to reviewing the agent's writes.

3.2 Organising patterns: what survived contact with agents

The classic PKM taxonomies survived mostly as naming conventions inside the instruction file, not as human disciplines. PARA's project-area-resource-archive split maps cleanly onto agent folder rules and is in evidenced use inside agent vaults (Stefan Imhoff's 6,000-note vault, March 2026; agent frameworks shipping PARA as a configuration mode). Zettelkasten survives as atomic notes and wikilinks, with the agent doing the linking; one practitioner notes that a dangling wikilink becomes "a 'write this later' marker" for the agent. Maps of Content became the machine-maintained index file. YAML frontmatter became the typed-metadata layer agents rely on, and the tags-versus-links debate resolved quietly into "whatever is consistent enough for the agent": consistency, not taxonomy choice, is what agents reward. A distinctly agent-era failure appeared in the same place: taxonomy drift, where the agent invents near-duplicate tags and entity names across runs unless the schema pins vocabulary or the pipeline feeds existing tags back in.

3.3 Where plain files strain

Scale limits are better documented than the hype suggests. Karpathy bounds his own pattern: index-first navigation works "at moderate scale (~100 sources, ~hundreds of pages)"; beyond that "you want proper search", and he points to qmd (Tobi Lütke's local search tool for markdown). Practitioners at thousands of notes bolt on retrieval infrastructure: Imhoff put qmd in front of a 6,000-note vault immediately; Graham Mann added vector search at week four because synonym mismatch defeats text search; OpenClaw ships hybrid search as core infrastructure and budgets how much curated memory is injected per session. The pattern that emerges everywhere: a small, curated, always-loaded core plus a larger on-demand layer reached through search. This split is reinvented independently by nearly every serious build (Karpathy's index.md, OpenClaw's MEMORY.md versus daily notes, Claude Code's auto-memory loading only the first 200 lines of its index, Mann's core-versus-current split after his single memory file bloated). It is the pattern's real backbone, and markdown remains the storage layer while retrieval quietly stops being markdown.

Token economics bite too. A practitioner running Codex reported that token burn came mainly from the agent re-reading instruction and index files across short sessions, pushing bookkeeping into deterministic scripts and cheap models, and reserving frontier models for synthesis. Several independent builds converge on that division of labour.

3.4 The graveyard question: does agent maintenance fix PKM's failure mode?

The pre-agent critique is well established and still the intellectual baseline. Casey Newton's "Why note-taking apps don't make us smarter" (Platformer, August 2023) argued that linking databases do not produce insight; Joan Westenberg's "I Deleted My Second Brain" (June 2025) described deleting 10,000 notes accumulated over seven years: "my second brain became a mausoleum. A dusty collection of old selves, old interests, old compulsions, piled on top of each other like geological strata." The genre's shared diagnosis is capture without synthesis: elaborate funnels with no drain.

Karpathy's core argument is that maintenance, the thing humans abandon, is exactly what the LLM makes near-free. The July 2026 evidence on that claim is mixed, and the paper reports it as mixed. In favour: practitioners with disciplined loops report compounding value ("the wiki answers questions the raw sources never could, because the synthesis already happened", one implementer at a few thousand concepts; Mann's assistant "gets noticeably better every week"). Against: Buntz's "roughly a wash" at one month; a survey of 27 public Claude Code memory-system builds reporting that "most systems were abandoned within weeks" (Vibehackers, March 2026, flagged: the page's body could not be fully read and the claim rests on its own summary metadata); and the strongest negative result in the field, the archcheck A/B experiment (published in-repo, 3 July 2026), which found a 28-page agent-built wiki over a codebase saved nothing versus letting the agent grep, because the wiki pages were larger than the roughly 100-line files they described, and pages marked "derived — never trust over

code" made the agent read both wiki and source. Its author's conclusions are the sharpest design principles yet written for this field: the token saving "appears only when a page aggregates facts scattered across many files (compression ratio > 1)"; per-entity pages that mirror small source files "are pure maintenance debt"; and "Either the wiki is trusted at query time (with freshness maintained separately), or it should not exist."

There is also a failure mode classic PKM never had: the record can now lie. The HaluMem benchmark (Chen et al., arXiv, November 2025) found that agent memory systems generate and accumulate hallucinations during extraction and updating, which then propagate to answers. Unsupervised agent maintenance does not merely fail to help; it can actively pollute the record. The practitioner countermeasures that work look like editorial process: citation-required schemas ("every factual sentence needs an inline citation... a claim without a source is a defect", Buntz), provenance and status fields, supersession instead of deletion, and human gates on promotion to the trusted layer.

3.5 Assessment

Sustained in real use: the five-element core (markdown, trust-layered folders, instruction file, git, loops) at personal and small-team scale; the small-core-plus-search architecture; cheap deterministic automation for bookkeeping with frontier models reserved for synthesis; lint loops with hard gates.

Decays in practice: wikis that mirror already-greppable sources; "never trust over code" disclaimers on derived pages, which pay the reading cost twice; big flat memory files without budgets; batch-only capture (the evening conversation the nightly cron missed); graph-view aesthetics, which carry over unchanged from the pre-agent critique.

Oversold: zero-maintenance claims; the assertion that agent maintenance abolishes the graveyard (the best independent datapoint is "roughly a wash", and abandonment-within-weeks is the reported norm among casual builds); markdown-only retrieval past hundreds of pages.

Underrated: compression as the design goal (the archcheck result); provenance and staleness fields; the instruction file as owned intellectual property that compounds; and honest measurement, of which the field has almost none.

4. The tooling, model and app ecosystem

Every tool, model and framework in this section belongs to its named owner. This is a survey of others' work; nothing here is PAC's.

4.1 Obsidian

Obsidian (developed by the company founded by Shida Li and Erica Xu; Steph Ango, known as kepano, CEO since February 2023) remains the dominant human frontend for markdown brains, and its position rests on decisions that predate the AI wave: local plain-text files, a free core app with no telemetry, optional paid sync and publish (US$4 and US$8 a month respectively as of July 2026), and a deliberately investor-free structure. The vault is just a folder of markdown files, which is exactly what a file-native agent wants.

Obsidian's response to the agent era is distinctive: it has shipped no native AI feature at all. Instead, in January 2026 Ango published official Agent Skills (kepano/obsidian-skills, observed at 38,900 GitHub stars on 5 July 2026): five skills teaching any compatible agent (Claude Code, Codex CLI, OpenCode and others) Obsidian's formats, including its markdown conventions, Bases and JSON Canvas. The strategy is to make the formats legible to every agent rather than build an AI button, and it has the useful side effect for users of keeping the intelligence swappable. Supporting moves: Bases (notes as database views, an open plain-text format agents can write, rolled out through 2025), an official CLI (early access reported February 2026, positioned for scripting and agents), and a rebuilt plugin directory with automated security review (May 2026), launched with the stated observation that over 4,000 plugins and themes exist and over 120 million total downloads have occurred, and motivated explicitly by agent-accelerated plugin creation.

4.2 The plugin layer: verified against the directory

The unverified prior this paper was handed named three plugins as the ones that matter: "AI Wiki", "Vault Operator / Agent Client" and "Local LLM Hub". Checked against the official directory on 5 July 2026, all three names exist, which a sceptic would not have guessed, but the claim misleads on every axis that matters. "Vault Operator / Agent Client" conflates two unrelated plugins by different authors. Vault Operator (Sebastian Hanke, from around April 2026) is an in-vault autonomous agent with shadow-git checkpoints and one-click undo, at 4,000 downloads. Local LLM Hub (takeshy, around April 2026) wires local runtimes and skills into the vault, at 4,000 downloads. AI Wiki (ikeniborn, June 2026) implements the Karpathy pattern with a Claude Code backend, at 114 downloads. Agent Client (rait-09, from October 2025), the fourth name, is real and more significant: 160,000 downloads, a chat panel that hosts Claude Code, Codex or Gemini CLI inside Obsidian via Zed's Agent Client Protocol.

The plugins that actually carry the ecosystem, by observed download counts on 5 July 2026: Claudian (Yishen Tu, from December 2025), which embeds a full coding agent inside Obsidian with the vault as working directory, at 1.1 million downloads in roughly seven months, the breakout adoption event of the cycle; Copilot for Obsidian (Brevilabs, commercial), the most-adopted in-app AI chat at about 1.5 million; Smart Connections (Brian Petro), the 2023-era semantic search that persists as retrieval at roughly one million, now partly paywalled and flagged "Caution" on Obsidian's own scorecard; and the non-AI infrastructure agents depend on: Dataview (4.4 million, effectively dormant in favour of a successor),

obsidian-git (2.77 million), Templater and Tasks. The MCP-over-REST bridges that connected Claude Desktop to vaults through 2025 (Local REST API plus community MCP servers) still exist but are being displaced by agents that simply open the files.

4.3 Agent front-ends

Four front-ends matter for brains as of July 2026. Claude Code (Anthropic, GA May 2025) is the reference: vault as working directory, CLAUDE.md as contract, skills and subagents for specialised behaviour. Codex CLI (OpenAI, from April 2025) and the Google terminal agents (Gemini CLI from June 2025; Google announced its transition to a successor CLI in June 2026, a reminder that the harness layer churns) serve the same role in other stacks, and the Agent Skills format is readable across several of them. Cowork (Anthropic, research preview January 2026) extends the same file-agent capability to non-developers through a desktop app; it is too new for sustained-use evidence, but it changes who can run a brain, and early security findings (section 4.6) came within 48 hours of its launch. Embedded options (Claudian, Agent Client) bring the same agents inside Obsidian. Persistent-agent frameworks are a parallel branch: OpenClaw (open source, stewarded by a foundation with OpenAI's continued backing after its creator joined OpenAI in February 2026) and Hermes Agent (Nous Research; announced on the Nous releases page under 25 February 2026, with the repo's own history showing development from July 2025 and a first tagged public release, v0.2.0, on 12 March 2026; MIT-licensed, observed at roughly 209,000 GitHub stars in July 2026) both run always-on agents whose memory is markdown files, and Hermes Agent ships an importer for OpenClaw's SOUL.md, MEMORY.md and USER.md, which means the brain files themselves are already portable between competing frameworks.

The minimalist end deserves naming because it may age best: a plain folder, git, and any skills-capable CLI, with no viewer at all. Nothing but files is load-bearing; every other layer is swappable.

4.4 Alternatives that do not fit

Logseq split into two products in April 2026, with the file-based app moving to maintenance-only development and the flagship moving database-first, away from plain files; practitioner drift to Obsidian and plain folders preceded that and looks validated by it. Notion is a brain competitor, not a brain substrate: server-side storage, lossy export, and AI that runs over workspace data in the vendor's cloud. The documented pattern among firms that use both is a hybrid: collaboration in Notion, private knowledge in local markdown. Roam Research, the 2020 darling, has no visible agent-era relevance; its proprietary cloud graph is the opposite of the file-access pattern.

4.5 The model layer, local models and Hermes

Cloud frontier models do the heavy lifting in most documented brains. As of July 2026 the practitioner default is Anthropic's Claude models through Claude Code and Cowork (Opus 4.7, GA April 2026, whose release notes explicitly claim improved file-system-based memory across multi-session work, a vendor claim; Opus 4.8, May 2026; Sonnet 5, June 2026, date from secondary coverage), with OpenAI's GPT-5.5 generation (April 2026) and Google's Gemini 3.1 Pro (February 2026) in the same role in other stacks. All capability figures from those vendors are vendor claims and dated as such.

The local and open-weight option matters to brains for one structural reason: a brain concentrates the most sensitive material its owner has, and plain markdown transits to whichever model reads it. The privacy argument is not hypothetical; it has dated incidents behind it. In NYT v. OpenAI, a court order issued in May 2025 required OpenAI to preserve output logs it would otherwise have deleted, including user-deleted chats; the blanket order ended on 26 September 2025, but the April-to-September corpus remains under legal hold, a demonstration that a vendor's deletion promise yields to legal process for whatever the vendor holds. Anthropic's consumer terms change of August and September 2025 moved consumer Claude plans to opt-out training with five-year retention for those who allow it (commercial and API terms excluded). Zero-retention arrangements exist but are enterprise-gated contract terms, not architecture. The honest counterweight: revealed preference in the documented builds is that most brain-owners still run frontier cloud models over their brains; capability currently outbids privacy for the majority, and the sensible mitigations are at the data layer (secrets never in notes, the most sensitive files excluded from cloud sessions).

As of mid-2026 the open-weight frontier is dominated by Chinese labs, each of the following pinned to a vendor primary in this paper's spot-check: DeepSeek (DeepSeek-V4 Preview, 24 April 2026), Moonshot's Kimi (K2.6, 20 April 2026), Alibaba's Qwen (Qwen3.6-27B open weights, 22 April 2026) and Zhipu's GLM (GLM-5.1, May 2026; later versions circulating on aggregator sites are unconfirmed against any Zhipu primary). Google's Gemma remains an actively maintained open line (Gemma 4, 2 April 2026) and should not be grouped with Meta's Llama, whose situation is starker than "trailing": no 2026 open-weight Llama release could be found on Meta's own channels, leaving Llama 4 (April 2025) the last open release. Mistral remains an active open-weight lab (Mistral Large 3, December 2025, Apache 2.0, a mixture-of-experts flagship; Mistral Small 4, March 2026), with a contested rather than leading position against the Chinese labs. The honest capability line, from the best independent evaluation found (a 30-day, six-stack, five-task local-agent study published May 2026): tool-calling is a model property with a floor around 27 to 32 billion parameters; supervised, scoped agent work over a brain (capture, summarisation, note updates) runs reliably on a 24GB GPU or 64GB Apple Silicon machine with approval gates, finishing 13 to 15 of 15 test tasks where frontier cloud models finish 14 to 15; long-horizon work (past roughly five to eight steps) and judgement-heavy curation remain frontier territory; and everything sold as autonomous "is a demo, not a product". The realistic 2026 pattern is hybrid: local models for

continuous private capture, frontier cloud models for curation sessions, with the most sensitive files excluded.

The Hermes lead this paper was handed verified almost entirely, with one correction. Nous Research's Hermes is two things. The model series (Hermes 3, August 2024; Hermes 4 family, August 2025, open weights, the 14B built on Qwen3 under Apache 2.0) is a respectable mid-tier open family, not a leader on mid-2026 open-model leaderboards. Hermes Agent, the framework (announced February 2026; first tagged public release March 2026), is the significant artefact: its README's feature list promises, under the heading of a closed learning loop, "Agent-curated memory with periodic nudges. Autonomous skill creation after complex tasks. Skills self-improve during use. FTS5 session search with LLM summarization for cross-session recall", and the code is public and inspectable. The claimed capabilities (skills generated from experience, multi-session memory, self-improvement loops) are real as described, though "the only agent with a built-in learning loop" is the vendor's framing and disputed by OpenClaw partisans. A May 2026 aggregator report citing OpenRouter's public rankings had Hermes Agent's routed traffic overtaking OpenClaw's (roughly 224 billion versus 186 billion daily tokens, point in time, not re-verified against the live leaderboard). The correction: that traffic measures the harness, across whatever models users route through it, and is not evidence that Hermes models power brains. The framework's importance to this paper is as the strongest public demonstration that the self-improving markdown brain ships in inspectable code, and that its memory files are portable enough to import from a competitor.

4.6 Security: the part the field learned the hard way

Brains inherit a structural vulnerability that Simon Willison's "lethal trifecta" framing captures: an agent with access to private data, exposure to untrusted content, and any channel to the outside world can be induced to exfiltrate. A vault of clipped web pages and pasted emails is untrusted content by definition. The demonstrations are concrete. Within 48 hours of Cowork's January 2026 launch, PromptArmor demonstrated a Word document with invisible one-point white text causing it to exfiltrate financial documents. The OpenClaw crisis of January and February 2026 put numbers on careless deployment: SecurityScorecard identified 40,214 exposed instances across 28,663 unique IP addresses (reported 9 February 2026), with 63 per cent of observed deployments vulnerable and 12,812 exploitable by remote code execution; earlier counts from other researchers ran lower (about 21,000), and one firm separately verified 5,194 actively vulnerable instances. Add plaintext memory and persona files readable by anyone, high-severity CVEs with public exploit code, and marketplace skills that wrote persistent backdoor instructions into SOUL.md and MEMORY.md that survived the skill's removal. The memory file, in other words, is an attack surface, and a poisoned note is an instruction the agent may follow later. The practitioner mitigations with evidence behind them: read-only defaults, human gates on writes, git as rollback, secrets never in notes, and treating anything ingested from outside as untrusted until reviewed. These map exactly onto the trust-boundary architecture of 3.1, which is why the boundary is not optional.

4.7 Formats and standards: the OKF verification

The second high-risk lead this paper was handed, a "Google Open Knowledge Format", looked like a model fabrication. It is real. Google Cloud announced OKF on 12 June 2026: a short v0.1 draft specification, public on GitHub, defining a knowledge bundle as a directory of markdown files with YAML frontmatter, exactly one required field (type), reserved index.md and log.md filenames, and ordinary markdown links forming the graph. The announcement cites Karpathy's gist and names "Obsidian vaults wired to coding agents, the AGENTS.md / CLAUDE.md family of convention files" as the prior art being formalised. Three weeks after launch it had modest traction (roughly 5,900 stars on the repo, as observed by this paper's research pass) and no visible adoption beyond Google's reference tools; the little independent commentary published so far is mostly descriptive, with at least one sceptical take asking whether it is a standard or "just a folder". It is a claim to standardisation, not an achieved standard, and this paper treats it as such.

OKF lands on top of a standards stack that assembled in about thirteen months: MCP for tool access (November 2024, Linux Foundation governance from December 2025), AGENTS.md for repo-level instructions (August 2025, same foundation), Anthropic's Agent Skills format for procedural knowledge (launched October 2025, opened as a standard in December 2025, adopted well beyond Anthropic's own tools), llms.txt for site-to-agent context (September 2024; contested, with roughly ten percent adoption in one surveyed sample and an explicit refusal from Google Search), and Obsidian's JSON Canvas (March 2024) at the margins. The direction is unmistakable: markdown-with-frontmatter bundles read by agents are being institutionalised by the largest vendors, and the brain pattern's substrate is becoming boring in the way that lasting infrastructure does.

4.8 Assessment

Production-reliable today: Obsidian core plus obsidian-git; Claude Code (or Codex CLI) against a vault under an explicit instruction file; the official Obsidian skills; the plain-folder minimalist stack; supervised local capture on a 27B-class model for the privacy-sensitive.

Demos well, fails in sustained use so far: in-vault autonomous agents (impressive engineering, weeks-old adoption, churn-level release cadence); MCP-over-REST vault bridges; unattended local autonomy of any kind; sub-7B "agent" models.

Oversold: the three plugins the prior named as significant; Notion as an open brain; view-count virality numbers; "fully autonomous" persistent-agent marketing; llms.txt as an AI-search lever; OKF as a settled standard.

Underrated: the Agent Skills spec as the practical interoperability layer; obsidian-git as the single highest-value safety component; Cowork's effect on who can run a brain at all; and the demonstrated portability of memory files between competing agent frameworks, which is the client-owns-the-record guarantee made mechanical.

5. The frontier and trajectory

Announced and speculative are kept strictly separate here.

5.1 Announced, with dates

Vendor memory is converging on the same territory from the managed side. Anthropic shipped Claude memory to Team and Enterprise plans in September 2025, Pro and Max in October 2025, and the free tier by March 2026, with project scoping and user-editable stores; Claude Code's auto-memory writes its own markdown index per project, loading a bounded core at session start. OpenAI rebuilt ChatGPT's memory architecture ("Dreaming V3", announced 4 June 2026) with self-benchmarked recall improvements, vendor figures. The managed features and the file-based pattern are competitors in one sense (convenience against ownership) and complements in another (several practitioners run both).

Governance matured faster than the pattern itself: MCP and AGENTS.md sit under the Linux Foundation's Agentic AI Foundation (December 2025), MCP has published a 2026 roadmap with a specification release candidate dated July 2026, and the Agent Skills format went from vendor feature to open standard in two months. Venture money entered the memory layer: Mem0 raised US$24 million (announced October 2025) to build a memory API for agents, with several funded peers. Jeremy Utley's "Teammate Stack" (verified: two Substack posts, 12 May and 16 June 2026, with a free downloadable kit) shows the pattern reaching the mass professional audience in persona-file form: eight plain files (identity, voice, anti-style, context, rules and three others) loaded into a chat assistant, with an interview meta-prompt doing the real work. Its mechanics are sound and its marketing ("11.5x smarter") carries no cited methodology. The "God-Mode Second Brain" essay this paper was asked to verify could not be found under that title anywhere and appears to be a conflation; the idea-cluster it describes is Karpathy's gist and the essays around it, and this paper cites those instead.

Obsidian's published roadmap keeps AI out of the core app and multiplayer collaboration in development for late 2026. Google's OKF is the standardisation bid to watch, not because it is winning but because whoever wins, the substrate it formalises (markdown, frontmatter, index and log files) is the one this paper surveys.

5.2 Speculative, labelled as such

Fine-tuning on personal corpora is talked about and not shipped in any mainstream product as of July 2026. Proactive brains (the agent briefs you unprompted) exist as scheduled-task community projects and vendor infrastructure, not as reliable defaults. Enterprise multi-user brains with real governance have one well-documented early adopter pattern (a fintech firm running seven knowledge domains as curated markdown skills with usage analytics and group permissions, published May 2026) and no evidence yet of being repeatable standard practice. The maximalist convergence claims ("RAG is dead", "everything becomes markdown plus agents") outrun their author's own caveats: Karpathy scopes the anti-RAG claim to moderate personal scale, and retrieval infrastructure reappears in every build past a few thousand notes.

The 12-to-24-month questions that will decide the field's shape: whether anyone produces an independent controlled measurement of net time saved; whether the security posture (gates, read-only defaults, provenance) becomes default tooling rather than practitioner discipline; whether OKF or a successor gets second-vendor adoption; and whether the abandonment curve for casual builds bends, which is the graveyard question asked again with better instruments.

6. Implications for operators

The reasonable posture, on the evidence surveyed:

Adopt now: the substrate. A markdown knowledge base under git, a trust boundary between raw capture and accepted knowledge, an instruction file with explicit prohibitions, and a human accept gate on anything entering the trusted layer. The cost is modest (the core tooling is free; a capable agent subscription runs US$20 to US$200 a month as of July 2026), the artefacts are portable across vendors, and the lock-in risk is close to nil, because nothing but files is load-bearing. This layer is durable whatever happens to the tools above it.

Pilot, under supervision: agent-maintained operations (ingest, query, lint), scheduled maintenance passes, and, where privacy warrants it, local-model capture on a 27B-class model. Measure while piloting: the field's best longitudinal datapoint is "roughly a wash" at one month, and the payoff concentrates where the material really is scattered and the open web is weak. A brain earns its keep on compression of dispersed private context, not on mirroring what an agent could grep or search anyway.

Wait on: unattended write access; agents that combine untrusted content with exfiltration channels (the PromptArmor demonstration is dispositive until the tooling changes); enterprise multi-user brains beyond early-adopter scale; OKF and any single standardisation bet; in-vault autonomous agents until their release cadence settles.

Walk away from: anything sold on an unmeasured multiplier ("11.5x smarter", "85% less maintenance"); write-only capture systems with no synthesis loop, which the last decade

already falsified; retrieval-only "chat with your notes" tools sold as brains; and any arrangement that surrenders the files themselves, because the one asset this field has proven is the portability of a folder of markdown.

The caveat that belongs on all of it: this field has produced, so far, exactly one controlled experiment and no independent measurement of aggregate benefit. Operators should price claims accordingly, including the claims in the paragraph above. Nothing here is legal or financial advice; costs and vendor terms cited are as of July 2026 and change without notice.

7. Conclusion

The state of brains in July 2026 is a young pattern on old foundations. The idea is 80 years old this month if dated from Bush, a decade old as mass-market method if dated from Forte's PARA, and about fifteen months old in its operative form if dated from the first file-native agents of early 2025. Its popular explosion is precisely dateable to the first week of April 2026, and the paper trail behind that explosion is unusually clean: the pattern was established practice among practitioners through 2025, was named and legitimised by Karpathy's gist, and was moving toward vendor standardisation within ten weeks.

What is solid: the substrate. Plain markdown, git, a trust boundary, an instruction file that encodes its owner's judgement, and an agent doing supervised maintenance. Every element is cheap, inspectable, portable and multiply evidenced. What is not solid: every claim of large, unmeasured gains. The honest verdicts available at survey date are "roughly a wash" from the most disciplined public one-month report, "negative value" from the only controlled experiment for wikis that mirror greppable sources, abandonment within weeks for most casual builds, and a documented capacity for the record to pollute itself and to be weaponised. The field's own best practitioners already work like editors: provenance on every claim, gates on every write, and deletion never, deprecation always.

That this survey's two most fabrication-shaped leads (a Google format and a self-improving open-source agent) proved real, while its most confidently repeated attribution (a named originator account) proved untraceable, and that the fact-check refuted one of this paper's own dates along the way and logged the correction in the open, is the method's closing argument. In a field this young and this loud, the evidence trail is the product. This paper describes a moment; the tooling sections will age in months, the pattern sections in years, and the lineage, with luck, not at all.

References

All URLs were accessed on 5 July 2026. Where an X/Twitter post could not be fetched directly, the entry notes the corroboration used. Vendor sources are used for what a product is and

claims, never for how well it works.

Ahrens, S. (2017). *How to Take Smart Notes*. (Operationalised Luhmann for the Anglophone PKM wave; second edition 2022.)

Ango, S. (2023, July). File over app. https://stephango.com/file-over-app

Ango, S. / Obsidian (2026, January). obsidian-skills (GitHub repository; observed 38.9k stars 5 July 2026). https://github.com/kepano/obsidian-skills

Answer.AI / Howard, J. (2024, 3 September). The /llms.txt file. https://www.answer.ai/posts/2024-09-03-llmstxt.html

Anthropic (2024, 25 November). Introducing the Model Context Protocol. https://www.anthropic.com/news/model-context-protocol [vendor]

Anthropic (2025, April). Claude Code: Best practices for agentic coding. (Date corroborated by Willison, 19 April 2025.) [vendor]

Anthropic (2025, September). Bringing memory to Claude. https://www.anthropic.com/news/memory [vendor]

Anthropic (2025, 9 December). Donating the Model Context Protocol and establishing the Agentic AI Foundation. https://www.anthropic.com/news/donating-the-model-context-protocol-and-establishing-of-the-agentic-ai-foundation [vendor]

Anthropic (2026, 16 April). Claude Opus 4.7 announcement. https://www.anthropic.com/news/claude-opus-4-7 [vendor]

Anthropic (2026). Claude Cowork product page and safety guidance. https://www.anthropic.com/product/claude-cowork ; https://support.claude.com/en/articles/13364135-use-claude-cowork-safely [vendor]

Anthropic (n.d.). Manage Claude's memory (Claude Code documentation). https://code.claude.com/docs/en/memory [vendor]

archcheck / blurman-ai (2026, July). Agent wiki economics (A/B experiment write-up). https://github.com/blurman-ai/archcheck/blob/master/docs/research/agent_wiki_economics.md

Astro-Han (2026). karpathy-llm-wiki (GitHub repository; observed 1.4k stars 5 July 2026). https://github.com/Astro-Han/karpathy-llm-wiki

Basic Machines (2025, 15 March). Basic Memory (GitHub repository). https://github.com/basicmachines-co/basic-memory

Bielefeld University. Niklas Luhmann-Archiv (approximately 90,000 slips, 1951 to 1996). https://niklas-luhmann-archiv.de/

Brier, N. / Every (2025, 10 September). How to use Claude Code as a second brain (podcast write-up). https://every.to/podcast/how-to-use-claude-code-as-a-thinking-partner

Buntz, B. / R&D World (2026, 11 June). Is Karpathy's viral LLM wiki helpful? Mostly yes, one month in. https://www.rdworldonline.com/is-karpathys-viral-llm-wiki-helpful-mostly-yes-one-month-in/

Bush, V. (1945, July). As we may think. *The Atlantic*. https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/

Chen, D., Niu, S., Li, K., Liu, P., Zheng, X., Tang, B., Li, X., Xiong, F., & Li, Z. (2025, 5 November; v3 5 January 2026). HaluMem: Evaluating hallucinations in memory systems of agents. arXiv:2511.03506. https://arxiv.org/abs/2511.03506

Cline / Baumann, N. (2025, 6 February). Memory Bank: how to make Cline an AI agent that never forgets. https://cline.bot/blog/memory-bank-how-to-make-cline-an-ai-agent-that-never-forgets [vendor]

CNBC (2026, 2 February). OpenClaw: the open-source AI agent's rise and controversy. https://www.cnbc.com/2026/02/02/openclaw-open-source-ai-agent-rise-controversy-clawdbot-moltbot-moltbook.html

DeepakNess (2026, 3 April). LLM knowledge bases (same-day summary of the Karpathy post). https://deepakness.com/raw/llm-knowledge-bases/

DeepSeek (2026, 24 April). DeepSeek-V4 Preview announcement. https://api-docs.deepseek.com/news/news260424 [vendor]

eWeek (2026, April). Anthropic opens Claude Cowork to all paid plans on macOS, Windows. https://www.eweek.com/news/claude-cowork-general-availability-enterprise-controls/

Faulkner, S. [@southpolesteve] (2025, 4 August). X post ("You are my personal assistant and chief of staff. Self organize using markdown files"). https://x.com/southpolesteve/status/1952422489847689383 (title-verbatim; not fetched in full)

Forte, T. (2017, February). The P.A.R.A. method: a universal system for organizing digital information. Forte Labs. https://fortelabs.com/blog/para/ [self-interested]

Forte, T. (2022, 14 June). *Building a Second Brain*. Atria Books. https://www.simonandschuster.com/books/Building-a-Second-Brain/Tiago-Forte/9781982167387

God of Prompt [@godofprompt] (2026, 6 April). X article: step-by-step replication guide of the Karpathy pattern (archived with engagement data). https://x.com/godofprompt/status/2041265656893489419 [self-interested]

Google (2025, 25 June). Introducing Gemini CLI. https://blog.google/technology/developers/introducing-gemini-cli-open-source-ai-agent/ [vendor]

Google (2026, 19 February). Gemini 3.1 Pro announcement. https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-1-pro/ [vendor]

Google (2026, 2 April). Gemma 4 announcement. https://blog.google/innovation-and-ai/technology/developers-tools/gemma-4/ [vendor]

Google (2026, 18 June). Transitioning Gemini CLI to Antigravity CLI. https://developers.googleblog.com/an-important-update-transitioning-gemini-cli-to-antigravity-cli/ [vendor]

Google Cloud / McVeety, S., & Hormati, A. (2026, 12 June). Introducing the Open Knowledge Format. https://cloud.google.com/blog/products/data-analytics/how-the-open-knowledge-format-can-improve-data-sharing/ [vendor]; specification: https://github.com/GoogleCloudPlatform/knowledge-catalog/blob/main/okf/SPEC.md

Hedgineer (2026, 12 May). How we built a company-wide knowledge layer with Claude Skills. https://www.hedgineer.io/content/claude-skills-knowledge-layer/ [self-interested]

Imhoff, S. (2026, 7 March). Agentic note-taking with Obsidian and Claude Code. https://www.stefanimhoff.de/writing/agentic-note-taking-obsidian-claude-code/

InfoQ (2025, August). AGENTS.md: a README for AI agents. https://www.infoq.com/news/2025/08/agents-md/

Infosecurity Magazine (2026). Researchers find 40,000+ exposed OpenClaw instances. https://www.infosecurity-magazine.com/news/researchers-40000-exposed-openclaw/

Karpathy, A. (2026, 2 April). LLM Knowledge Bases (X post; date derived from post ID timestamp, corroborated by DeepakNess same-day summary). https://x.com/karpathy/status/2039805659525644595

Karpathy, A. (2026, 4 April). llm-wiki.md (GitHub gist; created 4 April 2026 16:25, single revision; observed 5,000+ stars and forks 5 July 2026). https://gist.github.com/karpathy/442a6bf555914893e9891c11519de94f

Karpathy, A. (2026, ~10 April). Reply: "Yes it's the tractable form of brain upload..." (title-verbatim; not fetched in full). https://x.com/karpathy/status/2042626702459674801

Linking Your Thinking / Milo, N. (c. 2020). Maps of Content. https://www.linkingyourthinking.com/ [self-interested]

Linux Foundation (2025, 9 December). Linux Foundation announces the formation of the Agentic AI Foundation. https://www.linuxfoundation.org/press/linux-foundation-announces-the-formation-of-the-agentic-ai-foundation

Logseq (2026, April). Product split announcement (Logseq OG and Logseq DB). https://logseq.io/ [vendor]

Lütke, T. (n.d.). qmd: local search for markdown (GitHub repository). https://github.com/tobi/qmd

Mann, G. (2026, 5 March). Memory and task systems: giving your AI agent a brain. https://grahammann.net/blog/memory-and-task-systems-giving-your-ai-agent-a-brain

MarkTechPost (2026, 10 May). Hermes Agent overtakes OpenClaw on OpenRouter rankings. [aggregator citing openrouter.ai/rankings; point-in-time figures]

Mem0 / TechCrunch (2025, 28 October). Mem0 raises $24M to build the memory layer for AI apps. https://techcrunch.com/2025/10/28/mem0-raises-24m-from-yc-peak-xv-and-basis-set-to-build-the-memory-layer-for-ai-apps/

Mistral AI (2025, December). Introducing Mistral 3 (the family whose flagship is Mistral Large 3). https://mistral.ai/news/mistral-3/ [vendor]. Mistral Small 4 (2026, 16 March) per Mistral's news index, https://mistral.ai/news/ [vendor]

Model Context Protocol (2026). 2026 roadmap and July 2026 specification release candidate. https://blog.modelcontextprotocol.io/posts/2026-mcp-roadmap/ ; https://blog.modelcontextprotocol.io/posts/2026-07-28-release-candidate/ [project blog]

Moonshot AI (2026, 20 April). Kimi K2.6 announcement. https://www.kimi.com/blog/kimi-k2-6 [vendor]

Newton, C. (2023, August). Why note-taking apps don't make us smarter. *Platformer*. https://platformer.substack.com/p/why-note-taking-apps-dont-make-us

Nous Research (2026). Hermes Agent (GitHub repository, MIT; announced on the Nous releases page under 25 February 2026; development history from July 2025, first tagged release v0.2.0 on 12 March 2026; observed ~209k stars 5 July 2026 via the GitHub API). https://github.com/nousresearch/hermes-agent ; https://nousresearch.com/releases [vendor]

Obsidian (2026, 12 May). The future of Obsidian plugins. https://obsidian.md/blog/future-of-plugins/ [vendor]

Obsidian (2024, March). JSON Canvas. https://obsidian.md/blog/json-canvas/ [vendor]

Obsidian (2025, October). Obsidian October announcement (Bases APIs). https://obsidian.md/blog/2025-obsidian-october/ [vendor]

Obsidian (2026). Pricing page and community plugin directory pages for Claudian, Agent Client, Vault Operator, Local LLM Hub, AI Wiki and Smart Connections; download counts for Copilot, Dataview, obsidian-git, Templater and Tasks observed via the directory and obsidianstats.com, all on 5 July 2026. https://obsidian.md/pricing ; https://community.obsidian.md/ ; https://www.obsidianstats.com/ [vendor directory; obsidianstats is an independent tracker]

OpenAI (2025, August; with partners). AGENTS.md. https://agents.md/ [vendor coalition]

OpenAI (2025). Response to NYT data demands. https://openai.com/index/response-to-nyt-data-demands/ [vendor]

OpenAI (2026, 23 April). Introducing GPT-5.5. https://openai.com/index/introducing-gpt-5-5/ [vendor]

OpenAI (2026, 4 June). ChatGPT memory: Dreaming V3. https://openai.com/index/chatgpt-memory-dreaming/ [vendor, self-benchmarked]

OpenClaw (n.d.). Memory concepts documentation. https://docs.openclaw.ai/concepts/memory [project documentation]

PromptArmor (2026, January). Claude Cowork exfiltrates files (demonstration). https://www.promptarmor.com/resources/claude-cowork-exfiltrates-files [security vendor, reproduced demonstration]

PromptQuorum / Kuepper, H. (2026, 7 May). Autonomous local agents: what actually works (30-day, six-stack evaluation). https://www.promptquorum.com/power-local-llm/autonomous-local-agents-actually-work

QwenLM / Alibaba (2026, 22 April). Qwen3.6 repository. https://github.com/QwenLM/Qwen3.6 [vendor]

Reitz, K. (2026, 6 March). Obsidian vaults and Claude Code: a second brain that thinks back. https://kennethreitz.org/essays/2026-03-06-obsidian_vaults_and_claude_code

Search Engine Land (2025). llms.txt: the proposed standard and its critics (includes Google Search's refusal). https://searchengineland.com/llms-txt-proposed-standard-453676

TechCrunch (2025, 26 March). OpenAI adopts rival Anthropic's standard for connecting AI models to data. https://techcrunch.com/2025/03/26/openai-adopts-rival-anthropics-standard-for-connecting-ai-models-to-data/

TechCrunch (2025, 9 April). Google says it'll embrace Anthropic's standard for connecting AI models to data. https://techcrunch.com/2025/04/09/google-says-itll-embrace-anthropics-standard-for-connecting-ai-models-to-data/

TechCrunch (2025, 16 April). OpenAI debuts Codex CLI. https://techcrunch.com/2025/04/16/openai-debuts-codex-cli-an-open-source-coding-tool-for-terminals/

TechCrunch (2025, 28 August). Anthropic users face a new choice: opt out or share your data for AI training. https://techcrunch.com/2025/08/28/anthropic-users-face-a-new-choice-opt-out-or-share-your-data-for-ai-training/

TechCrunch (2026, 15 February). OpenClaw creator Peter Steinberger joins OpenAI. https://techcrunch.com/2026/02/15/openclaw-creator-peter-steinberger-joins-openai/

TechCrunch (2026, 19 May). OpenAI co-founder Andrej Karpathy joins Anthropic's pre-training team. https://techcrunch.com/2026/05/19/openai-co-founder-andrej-karpathy-joins-anthropics-pre-training-team/

TechCrunch (2026, 28 May). Anthropic releases Opus 4.8 with new dynamic workflow tool. https://techcrunch.com/2026/05/28/anthropic-releases-opus-4-8-with-new-dynamic-workflow-tool/

Utley, J. (2026, 12 May). Stop reintroducing yourself to AI. https://jeremyutley.substack.com/p/stop-reintroducing-yourself-to-ai [self-interested]

Utley, J. (2026, 16 June). Build your AI chief of staff in 30 minutes. https://jeremyutley.substack.com/p/build-your-ai-chief-of-staff-in-30 [self-interested]

Vibehackers (2026, 4 March). Claude Code second brains: a survey of 27 public builds. https://vibehackers.io/blog/claude-code-second-brain (body could not be fully read; see Appendix B)

Westenberg, J. (2025, 16 June). I deleted my second brain. https://medium.com/westenberg/i-deleted-my-second-brain-b7a65bce3717

WhyTryAI (2026, 26 February). Build your second brain with Claude Code and Obsidian. https://www.whytryai.com/p/claude-code-obsidian

Wikipedia (2026). OpenClaw. https://en.wikipedia.org/wiki/OpenClaw (used for its dense citations to CNBC, TechCrunch, Wired and Axios coverage)

Wilkins, J. (2026, 22 April). Seamless content ingestion for a Claude and Obsidian second brain. https://www.jdhwilkins.com/seamless-content-ingestion-for-claude-obsidian-second-brain/

Willison, S. (2025, 19 April). Claude Code best practices (commentary). https://simonwillison.net/2025/Apr/19/claude-code-best-practices/

Willison, S. (2025, 19 December). Agent Skills (commentary). https://simonwillison.net/2025/Dec/19/agent-skills/

Willison, S. (2026, 12 January). Claude Cowork (day-one review). https://simonwillison.net/2026/jan/12/claude-cowork/

Willison, S. (ongoing). The lethal trifecta. https://simonwillison.net/tags/lethal-trifecta/

YouMind (2026). Archive of the @godofprompt build guide with engagement data. https://youmind.com/landing/x-viral-articles/karpathy-second-brain-build-guide

Yu, W. (2026). Karpathy's LLM wiki versus the Zettelkasten (independent comparison). https://yu-wenhao.com/en/blog/karpathy-zettelkasten-comparison/

Appendix A: Methodology and verification approach

This paper was produced on 5 July 2026 through PAC's deep-paper pipeline: six parallel research streams (genesis and lineage; evolution and practitioner evidence; the brain pattern itself; apps and agent front-ends; the model layer, privacy and formats; frameworks, trajectory and implications), each run by a separate research agent instructed to work from live web search and primary-page fetches, to date every claim, to label every source as vendor or independent, and to return an explicit list of what it could not verify. The coordinator held the streams' raw notes without summarising them, and drafted to the structure the brief specified. The draft then passes through a structural verification pass and a multi-pass adversarial fact-check, gated on human inspection between passes; corrections are logged in a Corrections Log appendix added during that phase, and the pre-fact-check draft is preserved in archive. The version line in the frontmatter states where the document sits in that process.

The source hierarchy was adapted to a practitioner-led, largely un-peer-reviewed field, as the brief directed: a reproducible primary artefact that can be inspected (a gist, a repo, a spec, a plugin directory entry, a documented workflow) outranks any assertion; several unconnected practitioners converging on a practice outranks one influential voice; a single influencer's claim is a lead to verify; vendor documentation is evidence of what a product is and claims, never of how well it works; and aggregator or SEO content was used only to locate primary sources, with figures that trace no further than an aggregator flagged rather than asserted. Download counts, star counts and prices are point-in-time observations dated 5 July 2026 unless otherwise stated.

An adversarial input was used deliberately: a prior draft on this topic by another model (Grok, xAI), held in PAC's records as unverified. It was mined for leads only. Each of its high-risk specifics was independently checked, with results recorded in the body and Appendix B. Its

own assertions of verification were given no weight.

Honest limitations. Fact-check Pass 1 could not be run as a separate agent (the session's subagent budget was exhausted); it was run inline by the coordinator acting as a distinct checker role, per the deep-paper skill's documented fallback, with the findings written to a separate file before any correction was applied. X/Twitter pages could not be fetched directly; load-bearing X posts rest on verbatim text exposed in page metadata, post-ID timestamp decoding, and same-day independent summaries, and are flagged wherever the full text could not be read. Several vendor announcement pages render client-side and returned empty to fetches; dates there rest on convergent secondary coverage and are flagged. The field is young enough that independent evaluation barely exists; where the paper reports what works, it is reporting practitioner testimony and two pieces of systematic evidence (one benchmark, one controlled experiment), not settled science. AI systems performed the research, drafting and fact-checking of this paper under human direction, and the paper was not published until a human had reviewed it; that is the same supervised-autonomy posture the paper recommends.

Appendix B: What could not be verified

Each item was searched for specifically and is recorded here with the reason it failed verification, rather than asserted in the body.

1. **The @0xMiraqle step-by-step guide.** No X account, post, or secondary coverage crediting such a guide was found across repeated searches, including X-scoped searches, by two independent research streams. The role described matches the @godofprompt article of 6 April

2026. Treated as unverifiable and not repeated; open to correction if a primary URL is produced.

2. **The "God-Mode Second Brain" essay.** No essay under this title was found on Substack, Medium, Every or the open web across five search strategies. The nearest-titled item is unrelated; a GitHub project called GodMode (an OpenClaw persona-file setup) may be the source of the name. The idea-cluster it describes is real and is cited to Karpathy and Reitz instead.

3. **Full verbatim text of Karpathy's 2 April 2026 X post.** X fetches returned empty; only the opening (via page title) and same-day summaries were recoverable. In particular the widely quoted "~100 articles / ~400,000 words" corpus figure appears only in secondary retellings and is not treated as established.

4. **X view counts for the Karpathy post.** Secondary sources disagree (16 to 21 million within days to weeks); X view counts are not externally auditable. The paper says "tens of millions claimed" and relies on the gist's auditable star count instead.

5. **Cowork's general-availability date and Windows date.** Partly resolved in Pass 2: GA on all paid plans on 9 April 2026 is now corroborated by independent trade coverage (eWeek and others), and the claim was promoted into section 2.1. Anthropic's own announcement page remained

unreadable to fetches, and the Windows date (10 February 2026) is still secondary-sourced. The 12 January 2026 research-preview date was always solid.

6. **The Vibehackers finding that most of 27 surveyed memory-system builds were abandoned within weeks.** Sharpened in Pass 2: the article's title ("Building a Second Brain with Claude Code: What Actually Works After Three Months"), date (4 March 2026) and the abandonment claim are confirmed verbatim in its published metadata, but the body (and so the survey's methodology and the 27 builds' identities) remains unreadable to fetches. The claim stands as the article's own stated finding, not an independently checkable one.
7. **Hacker News thread scores for the gist discussion.** Aggregator-only; HN pages returned empty to fetches during the research window.
8. **Exact OpenClaw exposure figures.** Resolved in Pass 2: SecurityScorecard's research (reported by Infosecurity Magazine, 9 February 2026) gives 40,214 exposed instances, 28,663 unique IPs, 63 per cent of observed deployments vulnerable and 12,812 RCE-exploitable. Earlier lower counts (about 21,000) reflect earlier scans by other researchers; the 5,194 "actively vulnerable" figure is a separate firm's measurement.
9. **OpenRouter traffic figures for Hermes Agent versus OpenClaw (roughly 224 versus 186 billion daily tokens, May 2026).** Taken from an aggregator citing the public leaderboard; the live leaderboard was not re-fetched. Point-in-time only.
10. **Open-weight model versions beyond the vendor-pinned set, and Mistral's standing.** A spot-check pinned DeepSeek-V4 Preview, Kimi K2.6, Qwen3.6 and Gemma 4 to vendor primaries and confirmed GLM-5.1 as Zhipu's latest vendor-confirmed release. Aggregator-only versions (a claimed GLM-5.2, Kimi K2.7-Code, Qwen3.7 Max, and Meta's reported closed-weight "Muse Spark") could not be confirmed against any vendor primary and are not asserted. Mistral was resolved in Pass 2: an active open-weight line is vendor-confirmed (Mistral Large 3, December 2025; Mistral Small 4, March 2026), and the v0.9 draft's "no longer at the front" framing was dropped as too strong.
11. **AGENTS.md adoption ("60,000+ projects").** A steward self-report; the underlying GitHub search is reproducible but was not independently re-run.
12. **Mid-April 2026 star counts for the first replication repos.** Single aggregator ranking; only the 5 July 2026 observations are direct.
13. **The gbrain repository's scale figures (146,646 pages and related counts).** README self-reports without independent audit, which is normal for a personal repo; the figures are not load-bearing anywhere in this paper and are simply not relied on.
14. **The claim that Lex Fridman and Elvis Saravia run similar setups.** Asserted only in a third-party promotional guide; no primary from either person was found, so the claim is not repeated in this paper.
15. **Exact Luhmann sub-figures.** The approximately 90,000 total (1951 to 1996) is archive-documented; slip-box sub-splits and "books published" counts vary across sources and are not quoted precisely here.

16. **Any independent, controlled measurement of aggregate time saved by a markdown brain.** None was found anywhere, by any stream; the archcheck experiment (section 3.4) measured the token economics of one wiki pattern, not aggregate benefit. Every efficiency claim in circulation is self-reported. This is the largest hole in the field's evidence base and is reported as such in the body.

Blog seeds surfaced during this paper's research and drafting now live separately, internal only, at publication/brains-blog-seeds.md.

Corrections Log (v0.9 to v1.2 fact-check phase)

The pre-fact-check draft of this paper is preserved at [[archive/the-state-of-brains-july-2026-v0.9-pre-fact-check|archive/the-state-of-brains-july-2026-v0.9-pre-fact-check.md]]. The corrections below trace every change applied during the fact-check phase. The full findings file is [[the-state-of-brains-fact-check-findings]].

Pass 1 corrections (5 July 2026)

Summary: 19 findings processed (one a class finding covering the convergent date set). Eight verified with no change needed. Zero refuted; no fabricated sources, wrong-document citations, or numbers that failed to match their primary were found. Major corrections: four, all of one family, research-stream paraphrases that had hardened into quotation marks (archcheck twice, Hermes once, Basic Memory once), plus one paywalled quote replaced with a verifiable one (Westenberg). Material citation corrections: one (HaluMem reference completed with verified title and authors, refiled under Chen). Mechanical fixes: four precision softenings (OKF spec line count removed; OpenClaw first-release day softened to month; NYT order issue date softened to month; OKF repo stars attributed as a research-pass observation). Unverifiable items handled: the Platformer fetch failure was added to the flags carried in Appendix B; no items were moved or removed. Process disclosure added to Appendix A (Pass 1 checker ran inline; see finding #19). Items for Paul's judgement: none blocking; the class-18 convergent dates were kept on multi-stream agreement and can be spot-checked in Pass 2.

Detailed entries:

1. Section 3.4, archcheck. Original: the boundary-condition passage quoted as "the pattern pays exactly when a page compresses facts scattered across many sources; mirrors of small greppable files are negative value. Either the wiki is trusted at query time or it shouldn't exist." Changed to the verbatim source conclusions (compression ratio, maintenance debt, and "Either the wiki is trusted at query time (with freshness maintained separately), or it should not exist."). Source: archcheck agent_wiki_economics.md, fetched. Status after correction: verified. (Blog seed 4's title was aligned to the verbatim wording.)
2. Section 3.4 and 3.5, archcheck. Original: pages "stamped 'derived, verify against source'". Changed to the verbatim "derived — never trust over code". Same source. Now verified.

3. Section 3.4, Westenberg. Original quote ("I was mistaking a database for a brain...") sits behind the Medium paywall and could not be re-read. Replaced with the verifiable passage ("my second brain became a mausoleum..."). Date and framing confirmed. Now verified.
4. Section 4.5, Hermes Agent. Original presented the README feature table as a prose sentence. Reworded to attribute the verbatim description to the feature list under its "closed learning loop" heading. Repo metadata (208k stars, MIT, v0.18.0) verified directly. Now verified.
5. Section 1.3, Basic Memory. Original quoted "in standard Markdown files" and dated the release to 15 March 2025; the README has since been rewritten and the day-level date could not be re-confirmed. De-quoted to a paraphrase, date softened to March 2025, and a note added that the description follows the repo's 2025 positioning. Now caveated.
6. Section 4.7, OKF. Removed the unre-verified "451 lines" spec figure (the announcement calls the spec single-page); attributed the repo star count as a research-pass observation; noted the announcement's details were verified against the Google Cloud primary of 12 June 2026. Now verified with stated precision.
7. Section 1.3, OpenClaw. First-release date softened from "24 November 2025" to "November 2025" (day rests on a tertiary source). Star count (247,000 by 2 March 2026) and rename timeline confirmed. Now verified at month precision. (The star-figure attribution was further refined in Pass 2; see Pass 2 entry 3.)
8. Section 4.5, NYT v. OpenAI. Order issue date softened from "13 May 2025" to "issued in May 2025"; OpenAI's own page confirms the order, its end on 26 September 2025, the April to September legal hold, and the ZDR and Enterprise exclusions. Now verified except day precision.
9. References. HaluMem entry completed with the verified title ("HaluMem: Evaluating Hallucinations in Memory Systems of Agents") and full author list, refiled alphabetically under Chen; body citation updated to "Chen et al.". Now verified.
10. Appendix A. Added the disclosure that fact-check Pass 1 ran inline as a distinct checker role after subagent budget was exhausted. Process record.

Pass 2 corrections (5 July 2026)

Summary: nine findings processed, five adopted from an external adversarial spot-check run against v1.0 (the-state-of-brains-spotcheck-findings.md) and four from the checker's sweep of Appendix B. One refuted specific, the paper's first: "Hermes Agent, released 25 February 2026" was contradicted by the repo's own history (initial commit July 2025; first tagged release v0.2.0, 12 March 2026) and was corrected to what the date actually was, the Nous releases-page announcement date. Major corrections: three (the Hermes date; the open-weight model paragraph rebuilt on vendor primaries, including the finding that Meta shipped no 2026 open-weight Llama and that Gemma should not be grouped with it; Mistral resolved from "not verified" to an active open-weight line, the v0.9 "no longer at the front" framing having been too strong). Promotions from Appendix B: two (Cowork GA date, now trade-press corroborated, into 2.1; OpenClaw exposure figures, now pinned to SecurityScorecard via Infosecurity Magazine, into 4.6); the appendix entries were annotated as resolved rather than renumbered, to keep cross-references stable. Material precision fixes: the OpenClaw historical star figure re-attributed as an archived-snapshot report alongside the live July figure; OpenClaw stewardship reworded to include OpenAI's continued backing; Hermes star count updated to roughly 209,000 (live API); the OKF commentary sentence softened (coverage is thin and mostly descriptive). Verifications without change: the spot-check independently re-confirmed the Karpathy gist timestamp to the second, the X post date via an archived copy (upgrading Pass 1's ID-decode dating), the Anthropic join date via three outlets, and every OKF specific. Items for Paul's judgement: none blocking.

Detailed entries:

1. Sections 4.3 and 4.5, Hermes Agent. Original: "released 25 February 2026". Refuted by repo history (spot-check, GitHub API); corrected to the announcement date on Nous's releases page, with development-from-July-2025 and first-tagged-release 12 March 2026 stated. Star count updated to roughly 209,000. Status: corrected, now verified.

2. Section 4.5, model families. Original: aggregator-hedged single sentence grouping Llama and Gemma as "trailing" and Mistral as "no longer at the front". Rebuilt on vendor primaries (DeepSeek-V4 Preview, Kimi K2.6, Qwen3.6, Gemma 4, GLM-5.1) with aggregator-only versions excluded; Llama's absence of any 2026 open-weight release stated plainly; Mistral resolved as active (Large 3, Small 4). References added for DeepSeek, Moonshot, QwenLM, Google Gemma and Mistral. Status: now vendor-verified.

3. Sections 1.3 and 4.3, OpenClaw. Star figure re-attributed (archived snapshot via Wikipedia, with the live 382,000 figure noted); stewardship reworded to "a foundation with OpenAI's continued backing" per Altman's own statement. Status: now caveated correctly.

4. Section 2.1 and Appendix B item 5, Cowork GA. Promoted: 9 April 2026 GA on all paid plans, corroborated by eWeek and others; appendix entry annotated as resolved. Reference added. Status: now verified (trade press).

5. Section 4.6 and Appendix B item 8, OpenClaw exposure. Promoted: SecurityScorecard figures (40,214 instances, 28,663 IPs, 63 per cent vulnerable, 12,812 RCE-exploitable, 9 February

2026) replace the range; appendix entry annotated as resolved. Status: now verified.

6. Section 4.7, OKF commentary. "Independent commentary is already split" softened to thin, mostly descriptive coverage with at least one sceptical take. Status: now precise.

7. Section 3.4 and Appendix B item 6, Vibehackers. Appendix entry sharpened: title, date and the abandonment claim confirmed in published metadata; body and methodology still unread. Status: unchanged in body, flag sharpened.

8. Appendix B item 10 updated to record the vendor-pinned families and the Mistral resolution.

Pass 3 corrections (5 July 2026)

Summary: final read-through for drift and housekeeping. No new factual errors found, which is what Pass 3 should find if the earlier passes did their job. Changes: five, all housekeeping. (1) Reference-list ordering repaired: Search Engine Land moved ahead of the TechCrunch block; the two 2026 TechCrunch entries put in date order; the Mem0/Mistral/Model Context Protocol sequence corrected in Pass 2's edits was re-checked and stands. (2) The Nous Research reference entry brought into line with the Pass 2 Hermes correction (announcement date, repo history, 209k stars via the GitHub API). (3) The Mistral reference reconciled with the body's naming (Mistral 3 as the family whose flagship is Mistral Large 3; Small 4 dated 16 March 2026). (4) A forward-reference note added to Pass 1 entry 7 (OpenClaw), whose star-figure attribution Pass 2 had refined, per the log's honesty rule. (5) The title-block note and the conclusion updated to reflect the completed passes, the conclusion now also naming the fact-check's own refutation as part of the method's argument. Downstream drift check: the abstract, discussion and implications sections were re-read against the Pass 1 and 2 corrections and required no changes; the abstract's characterisation of the two verified fabrication-shaped leads survives the Hermes date correction, since the capabilities verified while only the date's meaning changed. Em-dash count: three, all inside verbatim quoted source text. Version labels reconciled to v1.2 throughout.

Editorial revision, v1.3 (6 July 2026, directed by Paul)

Not a fact-check pass: no claim, date, figure or source changed. Two tone adjustments were made after Paul's full read, on his direction. First, language that read as calling out named people or accounts was neutralised: section 1.2 retitled ("A lead that could not be traced") and its interest label on the @godofprompt account restated as a method note rather than a characterisation (Paul had already removed the closing "attribution folklore" sentence during acceptance); Appendix B items 1, 13 and 14 reworded in the same register. Second, the "named-not-invented" framing was replaced with the convergent-synthesis framing the evidence actually supports, since no source in the record claims invention (the gist itself disclaims it): section 1.3 retitled ("Convergent synthesis: the pattern before the name") and its opening and closing paragraphs rewritten in the form "built on X, added Y"; the Forte passage in 1.4 recast the same way. Blog seeds 2 and 7 in publication/brains-blog-seeds.md were retitled to match. One follow-on fix caught on review: the parent header, "1. Genesis and lineage," still carried the origin-story framing the subsections underneath it had just argued against; retitled to "1. Lineage and convergence" so the header agrees with its own section. The v0.9 archive was left untouched, as always.

Title change, v1.4 (6 July 2026, directed by Paul)

Not a fact-check pass and not a factual change: the paper's title was changed from "The State of Brains: markdown, AI-readable knowledge systems as of July 2026" to "The State of the AI Second Brain in Mid-2026: A Field Review for Operational Leaders." No claim, date, figure, source, or section content changed. The filename and slug (the-state-of-brains-july-2026.md, and the published the-state-of-brains-july-2026-literature-review-v1.0.pdf) were kept unchanged, so no internal or external link needed to move; only the frontmatter title, the H1, the PDF's cover and footer text, and display-text mentions elsewhere in the brain and on the PAC site were updated. Historical records (this Corrections Log's own earlier entries, the v0.9 pre-fact-check archive, the fact-check and spot-check findings files) were left referring to the paper by its prior title, as point-in-time records.

This is version 1.4: the three fact-check passes and the editorial tone revision logged above, plus this title change. The pre-fact-check draft is preserved at [[archive/the-state-of-brains-july-2026-v0.9-pre-fact-check|archive/the-state-of-brains-july-2026-v0.9-pre-fact-check.md]], and the full findings are in [[the-state-of-brains-fact-check-findings]] and [[the-state-of-brains-spotcheck-findings]]. Corrections, with sources, are welcome: the paper's method assumes some of it will need them.