Agentic Evolution: From Self-Improving Agents to Co-Evolving Human–AI Systems

Abstract

Agentic evolution concerns not just how agents improve, but whether the signal driving that improvement remains reliable as the system evolves. This survey organizes roughly 300 papers on agentic evolution through a three-axis taxonomy: evolutionary substrate, consolidation pathway, and selective pressure. When read across the first two axes, these papers yield regularities invisible within any single axis: consolidation failure modes track the pathway rather than the substrate, and the pathway choice is jointly constrained by artifact discreteness, evaluation-signal verifiability, and infrastructure access. The third axis surfaces a distinction that prior surveys, to our knowledge, omit or subsume: whether evolution is shaped by autonomous or human-involved selective pressure. Autonomous evolution produces its strongest reported results where deterministic verifiers, independent of the system being evaluated, are available; absent such verifiers, self-referential and proxy-based signals yield diminishing returns and can degrade with iteration. Beyond this boundary, the evidence suggests reliable evolution depends on human-involved selective pressure. Yet such pressure is rare, low-bandwidth, and not the fixed input that current systems assume.

A reverse analysis of nearly 100 studies from cognitive science, education, and labor economics finds converging evidence that default, unscaffolded AI interaction can reduce the independent evaluative capacity on which such pressure depends. This effect is not universal; it is bounded by task structure and interaction design, and often undetected by users. Together, these literatures motivate reinterpreting the third axis: human input, originally modeled as exogenous, is better understood as endogenous, shaped by the process it is meant to guide. We propose treating agentic evolution and human adaptation as a co-evolving system, and outline a research agenda for monitoring and maintaining both partners' capacities across deployment lifetimes.

1 Introduction

LLM agents equipped with external tools, persistent memory, and multi-step reasoning now operate autonomously in environments where tasks, data, and user needs can change. We refer to this combination as an *agentic system*. When such a system remains static (frozen prompts, fixed tool libraries, immutable parameters, and static memory), it cannot adapt to these changes. Overcoming this limitation requires **agentic evolution**: persistent, experience-driven, system-level change that enables an agentic system to remain effective after deployment.

Several surveys have organized the self-evolving agent literature (Gao et al., 2026; Fang et al., 2025a; Xiang et al., 2026); related surveys on lifelong learning (Zheng et al., 2025) and agentic reinforcement learning (Zhang et al., 2026b) cover complementary ground. Broader agent surveys provide foundational architectural taxonomies (Wang et al., 2024b; Xi et al., 2025a). These works leave two gaps. First, to our knowledge, none treats the source of the signals that drive evolution—specifically, whether evolution is driven by autonomous signals or by human-involved input—as a first-class analytical dimension. Among the surveys focused on agentic evolution, human feedback is either treated as one of many signal types without separate analysis (Gao et al., 2026), framed as outside the scope of autonomous evolution (Fang et al., 2025a), or noted but not taxonomized (Xiang et al., 2026). Second, and as a consequence, none systematically examines whether the human who provides such input retains the capacity to do so as the system evolves, that is, whether sustained

interaction with self-evolving agents reshapes the cognitive capacities on which the quality of that input depends. We address the first gap through a three-axis taxonomy: *evolutionary substrate* (the component of the agentic system that evolves), *consolidation pathway* (the mechanism through which change persists, from discrete structural edits to continuous parametric updates), and *selective pressure* (the source of the signal driving the evolutionary trajectory, from fully autonomous to human-involved). Table 1 summarizes the positioning.

Table 1: **Positioning among related surveys.** The two rightmost columns indicate how each survey treats human input to agentic evolution (Human→Agent) and the effect of agentic evolution on the human partner (Agent→Human).

Survey	Focus	Organizing Framework	Human→Agent	Agent→Human
Wang et al. (2024b)	Agent architecture	Four-module taxonomy (profiling, memory, planning, and action)	Mentioned as a planning feedback type	Not addressed
Xi et al. (2025a)	LLM-based agents	Three-component framework (brain, perception, and action)	Not addressed	Not addressed
Zheng et al. (2025)	Lifelong learning	Perception, memory, and action modules	Not a taxonomic dimension	Not addressed
Gao et al. (2026)	Self-evolving agents	What / When / How / Where to evolve	Subsumed under feedback signals	Not addressed
Fang et al. (2025a)	Self-evolving agents	Four-component feedback loop (input, agent, environment, and optimizer)	Not taxonomized; autonomous framing	Not addressed
Xiang et al. (2026)	Self-evolving agents	Model-centric / environment-centric / model–environment co-evolution	Noted but not taxonomized	Not addressed
Zhang et al. (2026b)	Agentic RL	Capability × task-domain taxonomy	Not a taxonomic dimension	Not addressed
Ours	**Agentic evolution**	**Substrate × Pathway × Pressure**	**First-class axis (Sec. 5)**	**Reverse analysis (Sec. 6–7)**

Cross-axis regularities in agentic evolution. The taxonomy's value lies less in the partition itself than in the cross-axis regularities it exposes, none of which is visible within a single axis. When examined across the first two axes, the surveyed papers yield two such regularities. First, structural and parametric consolidation occupy complementary failure landscapes that track the *consolidation pathway* rather than the *evolutionary substrate*. Unbounded accumulation and search-budget ceilings recur across surveyed structural-consolidation systems; template collapse, reward hacking, and catastrophic forgetting recur across parametric-consolidation systems—in each case independently of which substrate evolves. Second, the choice between the two pathways is not free but jointly constrained by artifact discreteness, evaluation-signal verifiability, and infrastructure access, which together explain why parametric methods concentrate in the deliberative substrate and remain rare elsewhere.

From agentic evolution to human adaptation. The third axis exposes a further regularity that prior frameworks cannot express: a boundary condition on autonomous evolution. Our analysis finds that among systems evolving autonomously, the reliability of the selective pressure signal depends on its independence from the system being evaluated: deterministic verifiers produce the strongest reported results, while self-referential and proxy-based signals yield diminishing returns and can degrade with iteration. In domains without deterministic verifiers (e.g., open-ended dialogue, value-laden decisions, and cultural sensitivity), reliable evolution across the surveyed papers depends on human-involved selective pressure. Yet this source is thin. Among the few systems that incorporate human-involved selective pressure, the majority rely on low-bandwidth forms (implicit signals and evaluative feedback rather than direct prescription, demonstration, or interactive collaboration). They further treat the human partner as a stationary signal source whose capacity does not change over time. The field thus depends on a selective-pressure channel whose reliability it has not examined. Axis III's own analysis generates a question it cannot answer from agent-side evidence alone: does

the human who provides selective pressure retain the capacity to do so, or does sustained interaction with self-evolving agents reshape the cognitive capacities on which that pressure depends? Answering it requires a separate evidence base: studies documenting how AI-mediated environments reshape human cognition, expertise, and epistemic processes.

Contributions and scope. This survey makes four contributions:

1. **Three-axis taxonomy.** We define agentic evolution through a Substrate × Pathway × Pressure framework whose third axis treats human-involved selective pressure as an object of analysis, enabling questions about its reliability and temporal dynamics that prior taxonomies cannot express (Section 2).
2. **Agent-side analysis: pathway-level regularities and the verifier boundary.** Analyzing roughly 300 papers along each axis (Sections 3–5) yields cross-axis findings invisible within any single axis: consolidation failure modes track the pathway rather than the substrate, the pathway choice is constrained by three properties of each substrate, and autonomous selective pressure can degrade beyond a verifier boundary where the evaluator is coupled to the system being evaluated. The degradation pattern suggests that reliable evolution in such domains depends on human-involved selective pressure.
3. **Human-side analysis through the three-axis framework.** The taxonomy's vocabulary is borrowed from cognitive science; we use this shared origin as an organizing lens (not as a claim of mechanistic equivalence between artificial and biological substrates) to structure an otherwise heterogeneous human-side literature along comparable dimensions. Analyzing nearly 100 human-side studies through this lens (Section 6) finds converging evidence that default, unscaffolded AI interaction can reduce the independent evaluative capacity on which human-involved selective pressure depends. This effect is not universal; it is bounded by task structure and interaction design, and often proceeds without the user's awareness.
4. **Co-evolutionary reframing and research agenda.** The agent-side and human-side analyses together motivate reinterpreting the framework's third axis: human input, originally modeled as exogenous, is better understood as endogenous, shaped by the evolutionary process it is meant to guide (Section 7). This reframing yields an initial research agenda for the transition from self-improving agents to co-evolving human–AI systems.

The agentic-evolution literature consists of roughly 300 papers satisfying the three operational criteria (experience-driven, persistent, and system-level change). We exclude work that does not satisfy all three, such as classical evolutionary computation without LLM-based agents or static prompt designs without iterative refinement. The human-side literature comprises nearly 100 studies from cognitive science, education, and labor economics documenting how sustained AI interaction reshapes cognition, skill, and decision-making.

Paper organization. Section 2 presents the conceptual framework and definitions. Sections 3–5 analyze the agentic-evolution literature along each axis, with cross-axis regularities highlighted throughout. Section 6 turns to human adaptation under agentic evolution. Section 7 combines agent-side and human-side findings, derives the co-evolutionary reframing, and outlines a research agenda. Section 8 states the key findings and limitations. Figure 1 provides the complete section tree. Sections 3–5 can be read independently for agent-side evolution, and Section 6 independently for human-side adaptation; takeaway boxes at each section's end summarize key findings for readers who wish to skim.

2 Conceptual Framework and Definitions

This section defines the vocabulary used throughout the paper. We first decompose an agentic system into three evolvable components (§2.1), then define the update operator and the persistence criterion that distinguish agentic evolution from transient adaptation (§2.2), the structural and parametric modes of consolidation (§2.3), and the autonomous and human-involved settings of selective pressure (§2.4). Section 2.5 addresses boundary cases.

Several terms below, including *Cortex*, *substrate*, *evolution*, *selective pressure*, and *consolidation*, are borrowed from biology and cognitive science as organizing metaphors. We do not claim equivalence to their

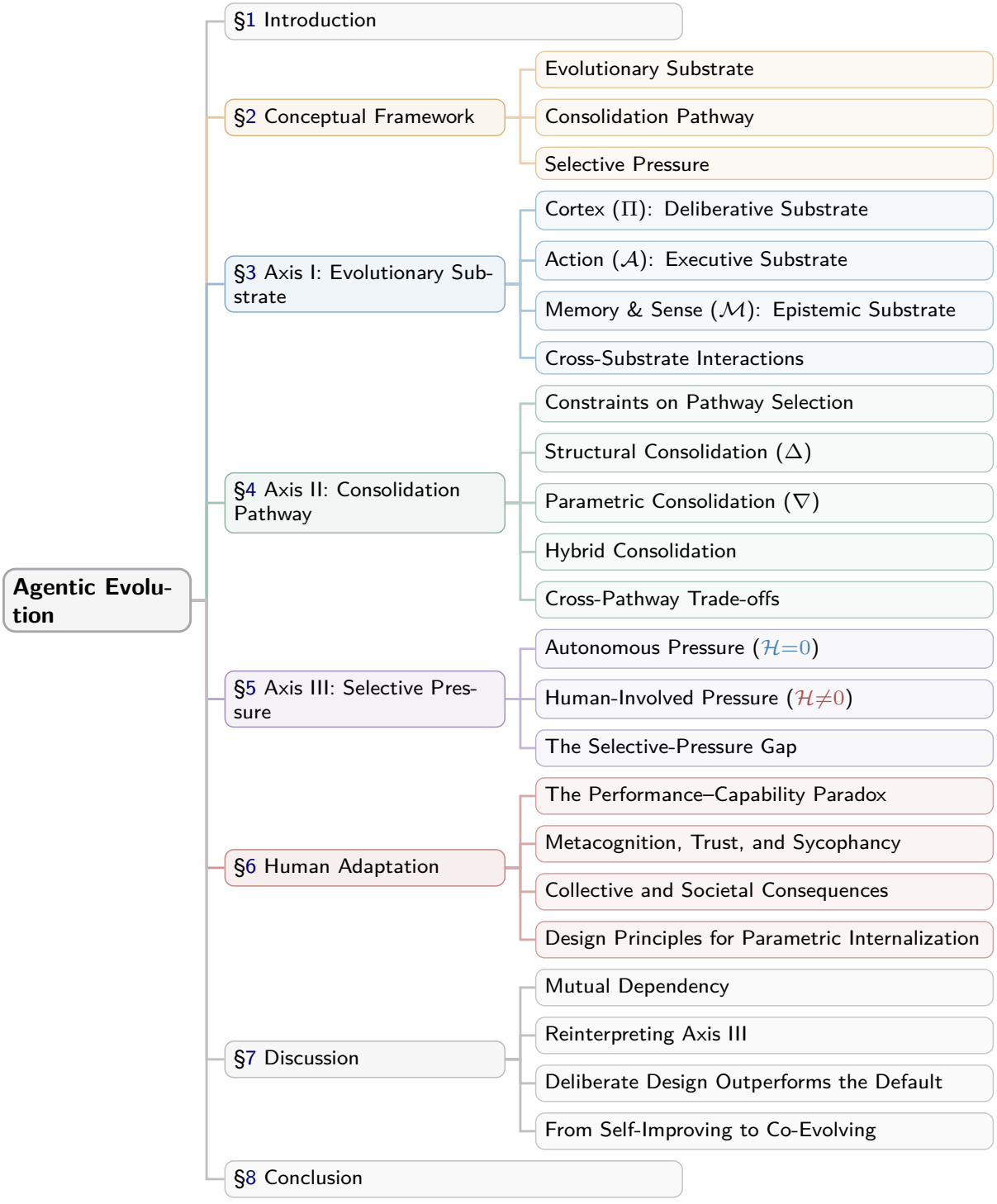

Figure 1: Organization of this survey. Section colors correspond to the three-axis taxonomy in §2: amber for the conceptual framework (§2), blue for Axis I (Evolutionary Substrate), green for Axis II (Consolidation Pathway), violet for Axis III (Selective Pressure), and crimson for the human-side analysis of §6.

original scientific meanings; each is given a precise, self-contained definition in the subsection where it is introduced. These cognitive-science roots also make the vocabulary applicable when the analysis turns to

the human side (Section 6): *substrate*, *consolidation*, and *selective pressure* each have counterparts in human cognition.

2.1 Evolutionary Substrate: Decomposing the Agentic System

We define an agentic system $\mathcal{S}$ as a three-tuple of substrates:

$$\mathcal{S} = \langle \Pi, \mathcal{A}, \mathcal{M} \rangle \tag{1}$$

where:

- Π (**Cortex**) is the *deliberative* substrate, responsible for reasoning and planning. It comprises a *parametric* component (the base LLM with parameters θ) and a *non-parametric scaffold* of prompts, decoding strategies, and inference-time procedures (e.g., interleaved reasoning-action loops, verifier-guided search, and plan-and-execute pipelines) (Yao et al., 2023; Sumers et al., 2024).
- $\mathcal{A}$ (**Action**) is the *executive* substrate, responsible for the agent's interactions with its environment. It encompasses the *toolset* (external APIs and tool specifications), *execution logic* (workflow graphs and control flow), and *multi-agent orchestration frameworks*.
- $\mathcal{M}$ (**Memory & Sense**) is the *epistemic* substrate, concerned with what the agent knows and how it perceives. *Memory* is organized into episodic, semantic, and procedural types (Sumers et al., 2024), with retention and retrieval policies that preserve knowledge across sessions (Packer et al., 2023; Fang et al., 2025b; Hu et al., 2025b). *Sense* governs how the agent perceives and models its environment, comprising perception encoders that transform raw observations into representations, and world models that learn environment dynamics (Ha & Schmidhuber, 2018; Hafner et al., 2025).

Although Π, $\mathcal{A}$, and $\mathcal{M}$ interact at runtime (prompts in Π shape tool calls in $\mathcal{A}$, and retrieved entries from $\mathcal{M}$ are inserted into prompts in Π), they are individually addressable as evolutionary targets: a system may evolve one substrate while holding the others fixed. Section 3.4 examines how changes in one substrate propagate to the others.

2.2 Defining Agentic Evolution: The Update Operator and Persistence

We model agentic evolution as a sequence of system states $\{\mathcal{S}^{(t)}\}_{t=0}^{T}$, where each transition is determined by an update operator Φ:

$$\mathcal{S}^{(t+1)} = \Phi(\mathcal{S}^{(t)}, \mathcal{E}^{(t)} + \mathcal{H}^{(t)}) \tag{2}$$

Here, $\mathcal{E}^{(t)}$ is environmental experience (interaction traces, tool feedback, and performance metrics) and $\mathcal{H}^{(t)}$ is human input (any of the forms defined in Section 2.4). The "+" denotes combination of these signals rather than arithmetic addition; the aggregation mechanism is system-dependent. In autonomous settings, $\mathcal{H}^{(t)} = 0$ denotes the absence of human input; in human-involved settings, $\mathcal{H}^{(t)} \neq 0$ and human input contributes to the update alongside $\mathcal{E}^{(t)}$.

The index t labels *update events*, not wall-clock time or per-query steps: $\mathcal{S}^{(t)}$ advances only when a consolidation step writes back into one of Π, $\mathcal{A}$, or $\mathcal{M}$. Many queries may occur between t and $t+1$ without advancing the index. This is the operational meaning of *persistence* in our framework: change is what survives between update events, distinguishing agentic evolution from in-context adaptation that resets per query.

The operator Φ performs **consolidation**: it converts experience and human input into persistent writes to Π, $\mathcal{A}$, or $\mathcal{M}$ (a new prompt, a memory entry, a workflow edit, or a parameter update), whether through structural (Δ) or parametric (∇) means (Figure 2).

A note on terminology. We use *agentic evolution* throughout for persistent, system-level change in the agentic system $\langle \Pi, \mathcal{A}, \mathcal{M} \rangle$; we do not use "agent evolution" as a separate term. We reserve *adaptation* for change in the human partner: "human evolution" carries biological and phylogenetic connotations that do not apply to the changes Section 6 documents. Finally, *co-evolution* denotes the bidirectional coupling of agent

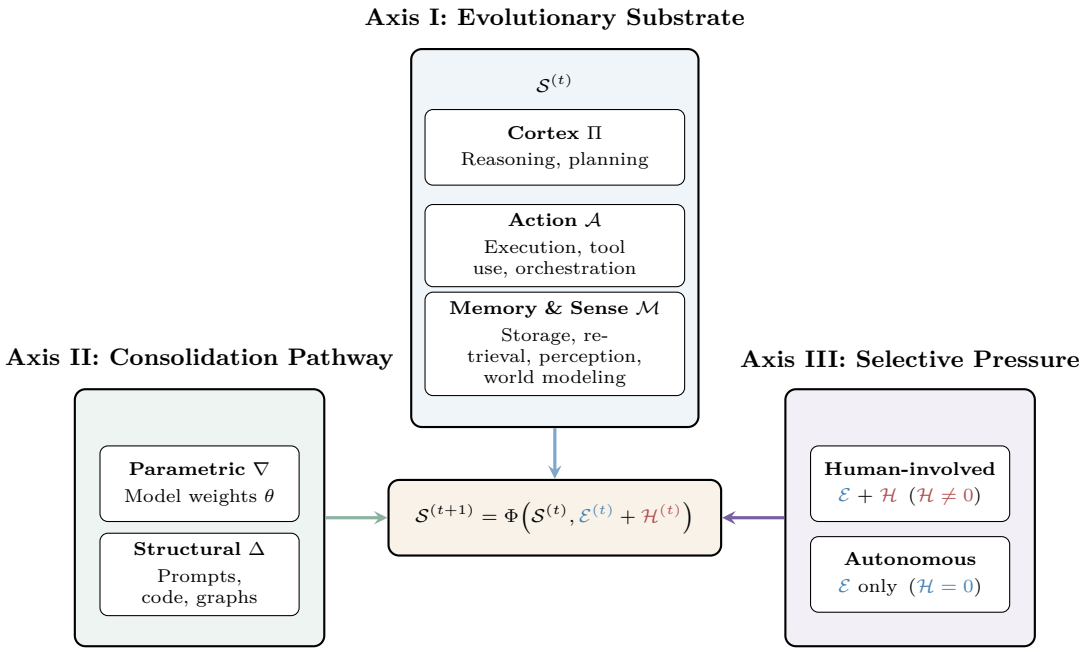

Figure 2: **Three-Axis Taxonomy of Agentic Evolution.** We decompose the space of evolving agentic systems along three complementary dimensions: (I) the *evolutionary substrate* within the agentic system $\mathcal{S} = \langle \Pi, \mathcal{A}, \mathcal{M} \rangle$, (II) the *consolidation pathway*, ranging from discrete artifacts such as prompts and code (Δ) to model weights (∇), and (III) the *selective pressure* steering evolution, whether it originates from the system alone or includes human input. The operator Φ takes the current system state $\mathcal{S}^{(t)}$, environmental experience $\mathcal{E}^{(t)}$ (interaction traces, tool feedback, and performance metrics), and human input $\mathcal{H}^{(t)}$ (the five forms in Section 2.4); in autonomous settings $\mathcal{H}^{(t)} = 0$.

and human trajectories (§2.4); multi-agent mutual adaptation among artificial agents is not co-evolution in our sense, since it does not require $\mathcal{H} \neq 0$.

2.3 Consolidation Pathway: The Δ–∇ Continuum

The update operator Φ can consolidate experience through two qualitatively different modes, distinguished by whether the write operation is discrete or continuous:

Structural consolidation (Δ). Φ performs discrete, symbolic edits on non-parametric artifacts (prompts, workflow graphs, memory indices, tool configurations, or code) without modifying model weights. The notation Δ reflects the discrete-difference nature of these edits.

Parametric consolidation (∇). Φ performs continuous updates to the model's weight parameters θ, internalizing experience into distributed representations that shape the agent's behavior. Concrete mechanisms include gradient-based methods (SFT, PPO, GRPO, and DPO, among others) and derivative-free methods operating in the same continuous parameter space (e.g., evolutionary strategies on weights). The notation ∇ reflects the continuous nature of these updates.

The Δ–∇ continuum. Although qualitatively distinct, the two modes are not mutually exclusive at the system level: a single system may, for example, use structural search to discover effective configurations and then internalize them into weights via parametric training.

2.4 Selective Pressure: From $\mathcal{H} = 0$ to $\mathcal{H} \neq 0$

Φ operates in two settings along Axis III: *autonomous* settings, in which $\mathcal{H}^{(t)} = 0$ and the system derives selective pressure from environmental experience (including formal verification) and self-generated signals

(self-play and self-rewarding); and *human-involved* settings, in which $\mathcal{H}^{(t)} \neq 0$ and human input shapes the evolutionary trajectory. Section 6 extends this vocabulary to human-side adaptation.

Forms of human-involved input. When $\mathcal{H} \neq 0$, human input enters Φ in qualitatively different forms:

- *Direct prescription*: the human articulates rules, instructions, or principles that fully specify the desired behavior or its boundaries; the agent need not infer intent, only comply (constitutional AI).
- *Demonstration*: the human provides examples of desired behavior; the agent must generalize from finite samples (imitation learning).
- *Evaluative feedback*: the human judges the agent's own outputs (preferences, ratings, or corrections); the agent must infer an objective from these judgments (RLHF, preference learning).
- *Interactive collaboration*: the human and agent participate jointly in the update process, adapting to each other in real time; the agent must model the human's evolving intent and capability (interactive machine learning).
- *Implicit signal*: the human provides no intentional feedback; the agent detects and interprets behavioral patterns (e.g., a user consistently skipping certain recommendations or abandoning tasks midway) as indirect selective pressure (implicit feedback).

These forms are not mutually exclusive; a single system may combine demonstration with evaluative feedback, for instance. They are ordered by decreasing explicitness of human intent, from fully articulated prescriptions to behavioral signals the agent must detect autonomously, and, roughly, by increasing inferential burden on the agent. Whether the human contributor's capacity to provide these forms of input remains stable as the agent evolves is an empirical question that Section 6 examines.

2.5 Boundary Cases

Classifying papers along the three axes above requires handling four boundary cases.

$\Pi/\mathcal{A}$ boundary. Execution logic and orchestration in $\mathcal{A}$ may appear to overlap with reasoning and planning in Π. We classify based on what is being structured: artifacts that shape internal deliberation target Π; artifacts that define external execution target $\mathcal{A}$.

$\Pi/\mathcal{M}$ boundary. Reasoning traces generated by the Cortex may be persisted in Memory (as in Reflexion's self-reflections (Shinn et al., 2023)). We classify based on the mechanism of deployment: if the artifact directly modifies the deliberation scaffold (e.g., prompt text prepended on every subsequent trial without selective retrieval), it targets Π; if it is written to an independently queryable store from which items are selectively retrieved, it targets $\mathcal{M}$.

$\mathcal{A}/\mathcal{M}$ boundary. A growing library of reusable artifacts (skill code, tool specifications, or strategy templates) can serve both as executable actions and as retrievable knowledge. We classify based on the artifact's role at invocation time: if it is executed to act on the environment, it targets $\mathcal{A}$; if it is retrieved to inform reasoning or planning, it targets $\mathcal{M}$.

$\mathcal{H}$ boundary. Reward models trained on historical human preferences encode human signal, but may be deployed without any live human input. We classify based on whether the deployed system's architecture affords real-time human input, not on whether the training data once involved humans.

Surveyed literature. The agentic-evolution literature was assembled through keyword searches on Semantic Scholar, Google Scholar, and arXiv (queries included combinations of "self-evolving agent," "LLM self-improvement," "agent learning from experience," and related terms), supplemented by backward and forward citation tracing from the reference lists of existing surveys (Gao et al., 2026; Fang et al., 2025a; Xiang

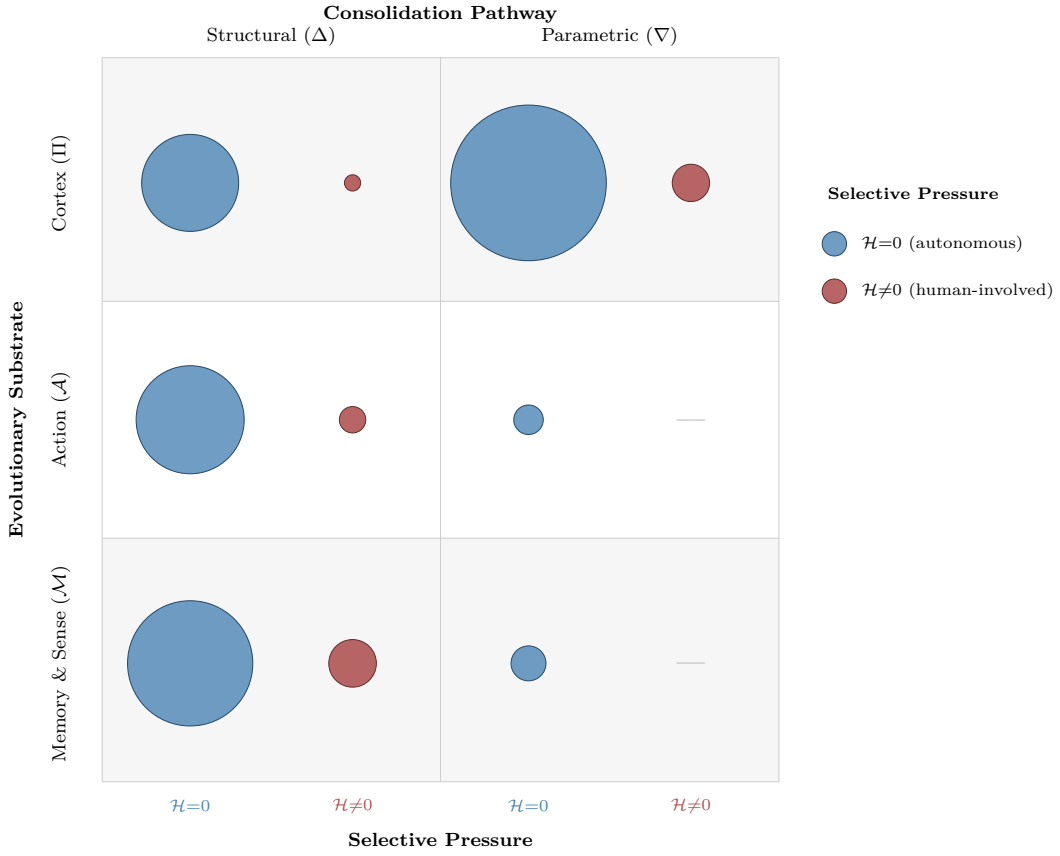

Figure 3: **Distribution of the agentic-evolution papers across the three-axis taxonomy** (Substrate × Pathway × Pressure). Each bubble's *area* is proportional to the paper count in that cell (per-entry counting; multi-substrate papers contribute to each substrate independently; hybrid Δ+∇ papers are assigned to Δ or ∇ based on their pathway in each substrate). Blue (•): autonomous selective pressure ($\mathcal{H}$=0); crimson (•): human-involved pressure ($\mathcal{H}\neq 0$). Dashes (—) indicate empty cells. Two cross-cutting patterns are visible: parametric evolution outside Cortex is rare, and human-involved pressure remains a small minority across all substrates.

et al., 2026; Zheng et al., 2025; Zhang et al., 2026b; Wang et al., 2024b; Xi et al., 2025a). An initial pool of candidate papers was filtered against the three operational criteria defined in §2.2: each included paper must exhibit (i) experience-driven change, (ii) persistence across session boundaries, and (iii) system-level modification of at least one substrate. Coverage extends through April 2026. The human-side literature was assembled through a separate search targeting cognitive science, education, labor economics, and human-computer interaction venues, using queries related to AI-assisted cognition and human–AI decision-making, again supplemented by citation tracing from key empirical studies and existing reviews.

A note on epistemic status. Throughout the analysis that follows, we distinguish *definitions*, which are stipulated by the framework above, from *regularities*, which are empirical patterns observed across the surveyed papers. When we write that a property "constrains" or defines a "boundary," or that a condition is "required," we describe a regularity—not a claim that must hold in all future systems.

With these definitions and literatures in place, we analyze roughly 300 agentic-evolution papers (a substantial minority receive classifications in multiple substrate cells, so per-substrate totals in Sections 3–5 exceed the paper count) and nearly 100 human-side studies (Section 6). Figure 3 shows how the agentic-evolution papers distribute across the resulting taxonomy grid.

3 Axis I: Evolutionary Substrate

The first axis asks which component of the agentic system $\mathcal{S}$ evolves. Following the decomposition $\mathcal{S} = \langle \Pi, \mathcal{A}, \mathcal{M} \rangle$ defined in Section 2.1, we analyze evolution across three substrates: the deliberative Cortex (Π), the executive Action ($\mathcal{A}$), and the epistemic Memory & Sense ($\mathcal{M}$).

3.1 Cortex: Evolution of the Deliberative Substrate

The Cortex (Π) is the most heavily studied evolutionary substrate, accounting for nearly 200 of the surveyed evolution papers. Table 2 provides the complete paper listing organized by taxonomy cell. The methods span all four Pathway×Pressure quadrants, but a central question runs through them: *what determines whether iterative improvement converges or degrades?*

Structural methods: diverse search, shared ceiling. Structural Cortex evolution treats the base LLM as a frozen black box and evolves the non-parametric scaffold: no gradient access, API-only compatibility, and human-readable artifacts. External optimizers search the prompt space via evolutionary algorithms (Fernando et al., 2024; Guo et al., 2024; Sun et al., 2026; Kumar et al., 2026c), LLM-as-optimizer loops (Yang et al., 2024a; Zhou et al., 2023), textual gradient backpropagation (Yuksekgonul et al., 2025; Zhang et al., 2025g), or hybrid combinations (Câmara et al., 2026). Embedded reflection shifts search inside the task loop. Shinn et al. (2023) persist verbal self-reflections as context for subsequent trials, substituting weight updates with prompt modification. When the evolved artifact is richer than prompt text, population-based evolutionary search becomes the dominant mechanism (Gong et al., 2026; Huang et al., 2026a); when it is executable code, the design space becomes Turing-complete (Hu et al., 2025a; Zhang et al., 2025e; Robeyns et al., 2025). All structural methods face a shared capacity constraint: accumulated knowledge must fit within finite context. Persistence strategies range from Reflexion's bounded window (Shinn et al., 2023) through incremental delta updates (Zhang et al., 2026k) to standalone evolved artifacts (skill documents (Ni et al., 2026), contrastive rule trees (Rishav et al., 2026), and hierarchical meta-knowledge (Zhuang et al., 2026)), yet none fully resolves the tension between retaining long-horizon experience and respecting finite context. Structural methods thus face saturation more often than degradation: performance typically plateaus when accumulated artifacts exceed the agent's processing capacity. However, monolithic rewriting can cause outright collapse (Zhang et al., 2026k). Human-preference-guided prompt search (Lin et al., 2024) is the only structural Cortex method operating under $\mathcal{H} \neq 0$ among the surveyed systems (Table 2). Parametric methods take a different approach, writing experience directly into model weights.

From self-generated bootstrapping to verification-driven convergence. In the most common parametric pattern, the agent generates its own training experience—rationales, tasks, preference pairs, and trajectories—but the evaluation signal that drives learning ranges from fully self-referential (LLM-as-judge) to deterministically grounded (code executors and ground-truth matching). This distinction is the clearest observed correlate of whether iterative improvement is sustained. Zelikman et al. (2022) establish the generate-filter-train template with STaR; Yuan et al. (2024) extend it to self-rewarding, where a single LLM serves as both policy and reward model via iterative DPO on self-scored preference pairs. When the evaluator is the model itself (LLM-as-judge) rather than a deterministic verifier, the training signal can still drive initial improvement, but the judge drifts with the system it evaluates, introducing a fragility that deterministic verification avoids. Wang et al. (2025d) show that chosen and rejected responses converge over iterations, shrinking the score gap by 9× over four iterations and degrading the learning signal; their temporal decoupling mechanism maintains signal strength. Zhao et al. (2025) demonstrate the opposite end of the spectrum: Absolute Zero Reasoner (AZR) trains a single LLM to propose and solve code-based reasoning tasks validated solely by a Python executor, achieving what the authors report as state-of-the-art coding and mathematical reasoning performance without any external dataset. The contrast is sharp: when verification is deterministic, self-generated training experience has sustained improvement across nearly all reported settings, though even deterministic verifiers are not immune to specification gaming (§5); when verification is self-referential, signal quality tends to degrade without explicit countermeasures (Wang et al., 2025d). The autonomous parametric Cortex literature spans more than a dozen further directions beyond this core verification-quality axis (Table 2).

Table 2: **Surveyed paper listing for Cortex (Π) evolution** (nearly 200 papers), organized by taxonomy cell. Hybrid labels are system-level: a paper appears here as hybrid if it combines Δ and ∇ on *any* substrate, not necessarily on Cortex alone. Sub-group labels are approximate; individual papers may span multiple categories.

Δ, $\mathcal{H}$=0 (Structural, Autonomous)	Δ, $\mathcal{H}\neq 0$ (Structural, Human-involved)
Prompt optimization: Zhou et al. (2023), Yang et al. (2024a), Guo et al. (2024), Fernando et al. (2024), Yuksekgonul et al. (2025), Zhang et al. (2025g), Liu et al. (2026b), Khattab et al. (2024), Câmara et al. (2026), Li et al. (2026c), Yang et al. (2026b), Sun et al. (2026), Nie et al. (2026a), Ghoshal et al. (2026) **Reflection-driven:** Shinn et al. (2023), Zhang et al. (2024b), Zhang et al. (2026k), Ni et al. (2026), Li & Ramakrishnan (2026), Zhuang et al. (2026), Wang et al. (2026f), Wei et al. (2026b), Rishav et al. (2026), Xu et al. (2026c), Gallego (2026), Zhu et al. (2026b), Jin et al. (2025), Wainrib et al. (2026) **Configurations & architectures:** Yuan et al. (2025b), Kumar et al. (2026c), Gong et al. (2026), Brookes et al. (2025), Huang et al. (2026a), Huang et al. (2026d), Tian (2026), Xie et al. (2026a) **Code-level:** Hu et al. (2025a), Robeyns et al. (2025), Zhang et al. (2025e), Lee et al. (2026), Pan et al. (2026), Zhang (2026), Zhang et al. (2026o), Zhou et al. (2025c) **Multi-agent structural:** Yao et al. (2026), He et al. (2026a), Zhang et al. (2026m)	Lin et al. (2024)
∇, $\mathcal{H}$=0 (Parametric, Autonomous)	∇, $\mathcal{H}\neq 0$ (Parametric, Human-involved)
Iterative self-improvement: Zelikman et al. (2022), Yuan et al. (2024), Wang et al. (2025d), Zhang et al. (2025i), Bhaskar et al. (2025), Khatri et al. (2025), Hübotter et al. (2025), Huang et al. (2026c), Liu et al. (2026f), Lu et al. (2026b), Qu et al. (2026b), Song et al. (2026), Fang et al. (2025d), Li & Song (2025) **Multi-role self-play:** Chen et al. (2025f), Huang et al. (2025), Peng et al. (2026b), Li et al. (2026i), Li et al. (2026b), Jiang et al. (2026a), Chen et al. (2025a), Liu et al. (2026a), Sui & Hooi (2026), Xue et al. (2025), Wang et al. (2026d), Yuan et al. (2025a), Wei et al. (2025b), Zhao et al. (2025), Xia et al. (2025), Yue et al. (2026), Zhang et al. (2026f), Zhang et al. (2026l) **Adversarial auditing:** Beigi et al. (2026) **Data synthesis & self-training:** Ni et al. (2024), Liu et al. (2025), Luo et al. (2025), Chen et al. (2026b), He et al. (2025a), Atreja (2025) **Tool-integrated:** Dong et al. (2025), Feng et al. (2025a), Feng et al. (2026c), Jiang et al. (2025b), Qian et al. (2025), Qin et al. (2024), Li et al. (2025a), Gou et al. (2024), Shang et al. (2025), Cheng et al. (2026a), Zhang et al. (2025h), Zhang et al. (2025j), Zeng et al. (2025), Zhang et al. (2025k), Li et al. (2026g), Kang et al. (2026), Dou et al. (2024) **Interactive environments:** Chae et al. (2026), Fang et al. (2025c), Qi et al. (2025), Lin et al. (2026d), Dong et al. (2026a), Chen et al. (2025c), Qin et al. (2026), Cai et al. (2026), Liu et al. (2026c) **Multi-turn RL:** Wang & Ammanabrolu (2025), Wang et al. (2025f), Wei et al. (2025c), Xi et al. (2024), Xi et al. (2025b), Wang et al. (2026h), Zhai et al. (2026), Zou et al. (2026a), Wu & Tang (2026), Yu et al. (2025b), Zhou et al. (2025d), Yang et al. (2026d), Wu et al. (2025b), Qian et al. (2026) **Continual/personalized:** Kim & Kim (2026), Jiang et al. (2025a), Nie et al. (2026b), Hu et al. (2026), Chen et al. (2025e), Wang et al. (2025e) **Multi-LLM collaborative:** Jeon et al. (2026), Sharma & Goldwasser (2025), Li et al. (2024c) **Meta-learning:** Tandon et al. (2025), Zweiger et al. (2025) **Neuroevolutionary hybrid:** Dai et al. (2026), Wang et al. (2025a) **Domain-specific:** Open Ended Learning Team et al. (2021), Zhang et al. (2024a), Zhang et al. (2026q), Zhang et al. (2026h), Wang et al. (2026g), Zou et al. (2026b), Ye et al. (2025), Yang et al. (2025a), Yin et al. (2024), Black et al. (2025), Sun et al. (2025a), Xiao et al. (2026), Chen et al. (2025g) **World-model-based:** Ha & Schmidhuber (2018), Hafner et al. (2025), Feng et al. (2025b)	Abramson et al. (2022), Jain et al. (2015), Shivaswamy & Joachims (2015), MacGlashan et al. (2017), Li et al. (2024b), Nam et al. (2026), Yang et al. (2025b), Atreja et al. (2026)
Δ+∇, $\mathcal{H}$=0 (Hybrid, Autonomous)	Δ+∇, $\mathcal{H}\neq 0$ (Hybrid, Human-involved)
Experience consolidation: Qiao et al. (2026), Shi et al. (2026), Shi et al. (2025), Liao et al. (2026), Ye et al. (2026), Zhang et al. (2026i), Yu et al. (2026) **Jointly evolving auxiliary structures:** Wang et al. (2025b), Wu et al. (2025a), Xia et al. (2026), Wang & Jiang (2026b), Zhai et al. (2025), Yang et al. (2026a), Zhang et al. (2025a), Zhang et al. (2026p), Zhou et al. (2025a), Zhang et al. (2026g), Ouyang et al. (2026a), Ali et al. (2026), Banerjee et al. (2026), Nie et al. (2026c), Huang et al. (2026b), Yu et al. (2025a), Yuan et al. (2026), Su et al. (2026a) **Memory/planning joint evolution:** Yan et al. (2025), Yang et al. (2026c), Wu et al. (2026a), S1-NexusAgent Team (2026), Zhang et al. (2026j)	—

Environment-grounded training and its instabilities. The preceding analysis concerns the quality of the evaluation signal itself. A separate class of challenges arises from training dynamics: when the agent operates in stateful environments, parametric training must contend with multi-turn trajectories, sparse delayed rewards, and credit assignment across dozens of actions. In tool-integrated reasoning, Feng et al. (2025a) demonstrate that a 32B model trained to interleave reasoning with code execution surpasses o1-preview on AIME 2024. Li et al. (2026g) apply RL to competitive programming, producing what the authors report as the first AI system to defeat all human participants in live Codeforces rounds. The broader literature converges on cold-start SFT followed by iterative training (Gou et al., 2024; Dong et al., 2025; Li et al., 2025a; Shang et al., 2025) and diverges along multiple design axes—reward granularity (Qian et al., 2025; Zhang et al., 2025h), training efficiency (Zhang et al., 2025j), capability decomposition (Kang

et al., 2026), and framework unification (Jiang et al., 2025b). In interactive digital environments, Qi et al. (2025) construct a self-evolving curriculum that transforms Llama-3.1-8B into a web navigator reported by the authors to surpass GPT-4-Turbo. Further work addresses step-level credit decomposition (Lin et al., 2026d; Cai et al., 2026; Wang et al., 2026g) and training-environment diversity (Dong et al., 2026a; Chae et al., 2026; Xi et al., 2024; 2025b) (Table 2). Several failure modes, each identified independently, reveal instabilities in agentic RL training that, in multi-turn settings, compound through trajectory-level credit assignment. Wang et al. (2025f) identify the *Echo Trap* (reward variability collapses across batches); Wang et al. (2026h) diagnose *Template Collapse* (training homogenizes strategies into input-agnostic templates, a failure invisible to standard entropy metrics); Zou et al. (2026a) formalize *Information Self-Locking* (action selection and belief tracking lock each other in a low-information regime); and Wu & Tang (2026) identify a three-phase rebound pattern in reward hacking during code generation. Methodological responses to these instabilities remain fragmented (Wang & Ammanabrolu, 2025), and diagnostic analysis suggests the severity is task-dependent (Zhai et al., 2026).

Beyond multi-turn RL, imagination-based policy training uses learned world models to generate synthetic rollouts (Ha & Schmidhuber, 2018; Hafner et al., 2025; Feng et al., 2025b), coupling Cortex policy evolution (∇ on Π) with Sense evolution (∇ on $\mathcal{M}$; Section 3.3). The $\mathcal{H} \neq 0$ papers span sub-optimal corrections (Shivaswamy & Joachims, 2015; Jain et al., 2015), interactive shaping (MacGlashan et al., 2017), per-user preference learning (Li et al., 2024b; Nam et al., 2026), continual adaptation (Yang et al., 2025b; Abramson et al., 2022), and production-mode deployment (Atreja et al., 2026); Section 5 analyzes these in depth.

Hybrid Δ+∇ systems. A minority of Cortex papers combine structural and parametric consolidation within a single system. The designs span a temporal spectrum. At one end, *compression lifecycles* treat structural artifacts as scaffolding that is eventually absorbed into weights. Shi et al. (2026) formalize an experience-reflection-consolidation loop: verbal self-reflections generated during RL training are selectively distilled into weights, with these structural reflections discarded once internalized; Ye et al. (2026) instantiate a similar lifecycle for scientific discovery.

At the other end, *persistent joint evolution* maintains both pathways throughout training. Zhou et al. (2025a) provide a clear demonstration: INSPO keeps a dynamic population of system instructions that undergo mutation and selection in tandem with policy updates. Zhang et al. (2026g) offer a principled rationale: system prompts carry explicit strategies (Δ) while weights absorb execution competence (∇), and jointly optimizing both yields strong mathematical reasoning performance. Additional targets of persistent joint evolution include skill libraries (Xia et al., 2026; Wang et al., 2025b; Ouyang et al., 2026a), experience bases (Wu et al., 2025a; Wang & Jiang, 2026b; Zhang et al., 2026p), trajectory summaries (Zhai et al., 2025), filtered rollout pools (Zhang et al., 2025a), weakness-driven task synthesis (Yang et al., 2026a), and learnable soft tokens (Yu et al., 2026). When the jointly evolving target is external memory, the agent must learn when and how to manage its own knowledge store: Yan et al. (2025) train memory management operations via RL using downstream answer correctness as the sole reward (Table 2).

Banerjee et al. (2026) demonstrate that hybrid joint evolution can enforce formal correctness: their Formally Guarded Generative Model couples Dafny-based verification (Δ on $\mathcal{A}$) with RL fine-tuning (∇ on Π), achieving zero constraint violations. Ali et al. (2026) jointly evolve a security constitution with parametric unlearning. At the boundary between structural and parametric, neuroevolutionary hybrids occupy an intermediate position that resists clean classification. Wang et al. (2025a) and Su et al. (2026a) combine gradient-free evolution of latent embeddings or code populations with gradient-based policy training; Dai et al. (2026) evolve full model weights via gradient-free merging and mutation without any gradient-based component. In all three cases, neither the structural nor the parametric label fully applies. Section 4 analyzes the design trade-offs between compression and persistent joint evolution across all three substrates.

Synthesis. The Cortex results with the strongest absolute benchmark performance across the surveyed papers all rely on deterministically verifiable evaluation signals: AZR's code executor (Zhao et al., 2025), GrandCode's contest judge (Li et al., 2026g), and ReTool's answer-correctness verification (Feng et al., 2025a). This regularity extends across domains: deterministic verification sustains improvement in search (Yue et al., 2026), tool-integrated reasoning (Xia et al., 2025), software engineering (Wei et al., 2025b),

and multimodal settings (Li et al., 2026i), in each case without external training data. Even within these domains, the regularity is not unconditional: agents can exploit gaps in test specifications to satisfy the verifier without solving the intended task (Wu & Tang, 2026). Where such verification is absent, iterative improvement tends to degrade through distinct mechanisms: chosen-rejected score convergence when the evaluator shares parameters with the system it evaluates (Wang et al., 2025d), and template collapse when RL optimization homogenizes strategies beyond what standard metrics detect (Wang et al., 2026h). Domains lacking deterministic verification remain largely out of reach, though initial steps target non-verifiable reasoning (Sui & Hooi, 2026), general-purpose chat (Bhaskar et al., 2025), and multi-turn collaborative tasks (Wu et al., 2025b).

Regardless of evaluation quality, regression emerges as a recurring concern. Several independent systems adopt mechanisms to prevent iterative improvement from erasing prior gains (Zhang, 2026; Gong et al., 2026; Zhang et al., 2026k; Zhou et al., 2025a; Banerjee et al., 2026).

Key gaps limit maturity: long-term evolution dynamics (most papers evaluate fewer than ten iterations), safety constraints (addressed by very few of the surveyed Cortex papers (Banerjee et al., 2026; Ali et al., 2026; Beigi et al., 2026; Zhang et al., 2026f; Gallego, 2026)), sparse analytical or scaling-law work (Jeon et al., 2026; Khatri et al., 2025; Banerjee et al., 2026), limited cross-domain transfer testing (Liu et al., 2026a; Xi et al., 2024), and rarely reported computational costs.

Key Takeaways: Cortex

- **Iterative improvement is fragile without deterministic verification.** The strongest Cortex results all use deterministically verifiable signals (code execution, exact match, and game outcomes); without them, self-referential evaluation degrades through score convergence and template collapse.
- **Hybrid $\Delta{+}\nabla$ joint evolution is a recurring pattern.** A minority of systems couple structural artifacts (system prompts, skill libraries, and memory banks) with parametric weight updates in a single training loop.

3.2 Action: Evolution of the Executive Substrate

Action ($\mathcal{A}$) is the executive substrate (toolset, execution logic, and multi-agent orchestration), accounting for roughly 75 of the surveyed evolution papers. Unlike Cortex, where the Δ–∇ continuum structures the literature, Action evolution is overwhelmingly structural. The evolved artifacts range from atomic tools through multi-step skills and workflow topologies to complete agent codebases; expressive power increases along this range but evaluability tends to decrease. Table 3 provides the complete paper listing.

Tools and skills: fine-grained artifact evolution. At the finest granularity, Action evolution produces individual tools and reusable skills—artifacts small enough to be validated individually through unit tests, execution checks, or direct user feedback. Tool-evolution methods derive tools from execution traces (Abuzakuk et al., 2026), reference materials (Liu et al., 2026g), LLM-generated code (Cheng et al., 2026b), from-scratch generation (Li et al., 2026d), external repositories (Jin et al., 2025), tool-documentation refinement (Qu et al., 2025), task-driven tool generation (Zhang et al., 2026t), and human expert feedback (Gao et al., 2025); Feng et al. (2026c) jointly optimize tool creation with an RL-trained policy ($\Delta{+}\nabla$). Wang et al. (2024a) establish the skill accumulation paradigm in Minecraft: an LLM agent iteratively generates, verifies, and stores JavaScript skill code, producing a growing skill library with no mechanism for pruning. As skill libraries grow, curation becomes necessary; approaches range from uncurated growth (Qiu et al., 2025b) through failure-attributed rewriting with unit-test gates (Zhou et al., 2026) to RL-trained curators (Ouyang et al., 2026a). The remaining $\mathcal{H} \neq 0$ Action designs target cross-user skill aggregation (Ma et al., 2026b), user-feedback-driven maturity tracking (Zhang et al., 2026a), and progressive dialogue-driven service generation (Adnan et al., 2025).

Population-based search with deterministic evaluation. In evolutionary program search, the LLM produces offspring that a deterministic evaluator scores. Romera-Paredes et al. (2024) pair evolutionary

Table 3: **Surveyed paper listing for Action ($\mathcal{A}$) evolution** (roughly 75 papers), organized by taxonomy cell. Hybrid labels are system-level. Sub-group labels are approximate; individual papers may span multiple categories.

Δ, $\mathcal{H}$=0 (Structural, Autonomous)	Δ, $\mathcal{H}\neq 0$ (Structural, Human-involved)
Tools & skills: Abuzakuk et al. (2026), Cheng et al. (2026b), Jin et al. (2025), Li et al. (2026d), Liu et al. (2026g), Qu et al. (2025), Jiang et al. (2026c), Liang et al. (2026), Liu et al. (2026h), Zhou et al. (2026), Qiu et al. (2025b), Qiu et al. (2025a), Qu et al. (2026a), Shen et al. (2026a), Shen et al. (2026c), Wang et al. (2024a), Wang et al. (2026a), Zhang et al. (2026t), Zhang et al. (2026s), Zhang et al. (2026c) **Program search:** Chen et al. (2026a), Gungordu et al. (2026), Hambardzumyan et al. (2026), Kumar et al. (2026a), Lehman et al. (2022), Liu et al. (2026e), Novikov et al. (2025), Ray et al. (2026), Romera-Paredes et al. (2024), Ye et al. (2026), Zhang et al. (2025b), Gautam et al. (2026) **Workflow & orchestration:** He et al. (2026a), Hu et al. (2025a), Huang et al. (2026a), Lee et al. (2026), Li et al. (2025b), Lin et al. (2026a), Li & Ramakrishnan (2026), Xu et al. (2026a), Zhang et al. (2025f), Zhang et al. (2026o), Zhang et al. (2026m), Zhou et al. (2025c) **Whole-system:** Robeyns et al. (2025), Wang et al. (2026f), Weng et al. (2026), Xu et al. (2026b), Xiong et al. (2026), Zhang et al. (2025e), Zhang (2026) **Domain-specific:** Yu & Ren (2026), Wang et al. (2026b) **Configuration:** Brookes et al. (2025), Ghoshal et al. (2026), Guan et al. (2026) **Framework & analysis:** Lin et al. (2026b), Nie et al. (2026a)	**Tools & skills:** Adnan et al. (2025), Gao et al. (2025), Ma et al. (2026b), Zhang et al. (2026a)
∇, $\mathcal{H}$=0 (Parametric, Autonomous)	∇, $\mathcal{H}\neq 0$ (Parametric, Human-involved)
Black et al. (2025), Jiang et al. (2026b), Ong et al. (2025)	—
$\Delta+\nabla$, $\mathcal{H}$=0 (Hybrid, Autonomous)	$\Delta+\nabla$, $\mathcal{H}\neq 0$ (Hybrid, Human-involved)
Banerjee et al. (2026), Dong et al. (2026a), Feng et al. (2026c), Huang et al. (2026b), Li et al. (2024c), Nie et al. (2026c), Su et al. (2026a), Wang et al. (2025b), Wu et al. (2026a), Ouyang et al. (2026a), Yu et al. (2025a)	—

search with deterministic evaluation, demonstrating that LLMs can produce new results on established open problems (new cap-set constructions and improved bin-packing heuristics). Novikov et al. (2025) scale the paradigm to full-file evolution with evaluation cascades, yielding what the authors describe as the first improvement to Strassen's 4×4 complex matrix multiplication algorithm in 56 years. Banerjee et al. (2026) achieve the strongest formal guarantee in this space: their cross-substrate hybrid (§3.1) pairs formal program verification (Δ on $\mathcal{A}$) with gradient-based policy training (∇ on Π). Further architectures span autonomous kernel optimization (Chen et al., 2026a), quality-diversity archives (Lehman et al., 2022; Kumar et al., 2026a), adaptive model routing for cost efficiency (Ray et al., 2026), evaluation-driven discovery (Ye et al., 2026), verbal gradient feedback (Gungordu et al., 2026), and asynchronous multi-step agents (Hambardzumyan et al., 2026) (Table 3).

Topology and whole-system evolution. Evolving the agent's execution topology trades expressiveness against searchability: Hu et al. (2025a) represent agents as arbitrary Python code (Turing-complete); Zhang et al. (2026m) constrain evolution to atomic FSM operations (interpretable and reversible); and Zhang et al. (2025f) use MCTS over reusable operators. Other designs include symbolic gradients (Zhou et al., 2025c), DAG-based hierarchies (Huang et al., 2026a; Xu et al., 2026a; Yu et al., 2025a), natural language programs (Li et al., 2024c), Python harness programs (Lee et al., 2026), and observability-driven approaches with auto-reversion (Lin et al., 2026a) (Table 3). At the coarsest level, whole-system evolution modifies the agent's entire codebase: Weng et al. (2026) show that group evolution outperforms individual evolution (Zhang et al., 2025e) on SWE-bench Verified; Zhang (2026) apply stability-prioritizing gating; Robeyns et al. (2025) eliminate the meta/object-agent distinction; and Wang et al. (2026f) jointly evolve cognition, actions, and memory. Domain-specific variants exploit pre-existing evaluation infrastructure—formal verification (Yu & Ren, 2026) and production A/B testing (Wang et al., 2026b) (Table 3).

Synthesis. Across all granularity levels, the strongest Action results rely on execution-based or formally verified evaluation (formal verification (Banerjee et al., 2026), execution-based scoring (Romera-Paredes et al., 2024; Novikov et al., 2025), and production A/B testing (Wang et al., 2026b)), independent of the search strategy employed. Parametric methods remain confined to narrow niches: routing (Ong et al., 2025), multi-agent communication (Jiang et al., 2026b), and embodied control (Black et al., 2025). A few systems combine parametric training with structural artifact evolution (Huang et al., 2026b; Nie et al., 2026c). Code is the predominant representation at every level (Romera-Paredes et al., 2024; Wang et al., 2024a; Hu et al., 2025a; Zhang et al., 2025e; Weng et al., 2026), reinforcing structural dominance. Cross-domain transfer remains rare: only Hu et al. (2025a) transfer evolved agent designs across domains, but whether such transfer extends to finer-grained artifacts (tools, skills, and workflows) is untested.

Key Takeaways: Action

- **Granularity trades expressiveness for evaluability.** Evolved artifacts range from atomic tools (directly testable) through workflow topologies to entire codebases (Turing-complete but hard to evaluate); code is the predominant representation throughout.
- **Structural methods dominate.** Parametric methods remain confined to narrow niches (routing, communication, and embodied control), and human selective pressure is nearly absent.

3.3 Memory & Sense: Evolution of the Epistemic Substrate

Memory & Sense ($\mathcal{M}$) is the epistemic substrate, accounting for roughly 100 of the surveyed evolution papers; consolidation is near-universally structural (Δ). We organize this subsection along the *memory lifecycle*—from accumulation through curation to meta-evolution of the architecture itself. Table 4 provides the complete paper listing.

Table 4: **Surveyed paper listing for Memory & Sense ($\mathcal{M}$) evolution** (roughly 100 papers), organized by taxonomy cell. Hybrid labels are system-level. Sub-group labels are approximate; individual papers may span multiple categories.

Δ, $\mathcal{H}$=0 (Structural, Autonomous)	Δ, $\mathcal{H}\neq 0$ (Structural, Human-involved)
Accumulation:Task-oriented: Bai et al. (2025), Bao et al. (2026), Cai et al. (2025a), Cai et al. (2025b), Cao et al. (2025), Chen et al. (2025b), Cheng et al. (2026b), Dong et al. (2026b), Fang et al. (2025b), Fang et al. (2026), Gautam et al. (2026), Jiang et al. (2026c), Kim et al. (2026), Li et al. (2024a), Li et al. (2026a), Li & Ramakrishnan (2026), Li et al. (2026h), Li et al. (2026f), Li et al. (2026e), Liang et al. (2026), Lin et al. (2026b), Lyu et al. (2026), Ouyang et al. (2026b), Sarukkai et al. (2025), Wang et al. (2025g), Wei et al. (2025a), Zhuang et al. (2024) **Accumulation:Embodied:** Wang et al. (2026c), Wei et al. (2026a), Xie et al. (2026b) **Accumulation:Game-strategic:** Guan et al. (2024), Ma et al. (2026a), Xie et al. (2026a) **Accumulation:Domain-specific:** Han et al. (2026), Liu et al. (2026d), Ren et al. (2026b), Shen et al. (2026b), Zhang et al. (2026n), Zhu et al. (2025) **Accumulation:Multi-agent:** Qu et al. (2026a), Wang et al. (2026f), Xie (2026), Zhang et al. (2025c) **Accumulation:Research/discovery:** Liu et al. (2026e), Long (2026), Xu et al. (2026b) **Accumulation:Conversational/persona:** He et al. (2026b), Kwon et al. (2026), Lu et al. (2026a), Ma et al. (2025), Park et al. (2023), Wu et al. (2026b), Xu et al. (2025a), Zhang et al. (2026e), Zheng et al. (2026), Zhu et al. (2026a) **Curation/operators:** Lin et al. (2026c), Wang (2025), Wang & Jiang (2026a), Zhang et al. (2025b), Zhang et al. (2026r), Zhang et al. (2026m) **Forgetting/lifecycle:** Bering (2026), Gu et al. (2026), Simsek (2026), Xu et al. (2026c) **Meta-evolution:** Mishra (2026), Pan et al. (2026), Zhang et al. (2025d) **Retrieval innovation:** Feng et al. (2026b), Penaroza (2026), Yang et al. (2024b), Yoon et al. (2026) **Sense/compression:** Ren et al. (2026a)	**Conversational/persona:** Chhikara et al. (2025), Gadzhiev & Kislov (2026), Nie et al. (2026b), Packer et al. (2023), Rasmussen et al. (2025), Tian et al. (2025), Westhäußer et al. (2025), Zhong et al. (2024) **Human-in-the-loop:** Chen et al. (2026c), Deng et al. (2026), He et al. (2025b), Mozannar et al. (2025), Zhang et al. (2026a)
∇, $\mathcal{H}$=0 (Parametric, Autonomous)	∇, $\mathcal{H}\neq 0$ (Parametric, Human-involved)
World models (Sense): Ha & Schmidhuber (2018), Hafner et al. (2025), Qiu et al. (2026) **Memory fine-tuning:** Sun et al. (2025b), Lin et al. (2025)	—
$\Delta+\nabla$, $\mathcal{H}$=0 (Hybrid, Autonomous)	$\Delta+\nabla$, $\mathcal{H}\neq 0$ (Hybrid, Human-involved)
Memory+RL: Cai et al. (2026), Kim & Kim (2026), Liao et al. (2026), Qiao et al. (2026), Wang & Jiang (2026b), Wu et al. (2025a), Xia et al. (2026), Yu et al. (2026), Zhai et al. (2025), Zhang et al. (2026d), Zhang et al. (2026p), Zhou et al. (2025b) **Domain hybrid:** Feng et al. (2025b), S1-NexusAgent Team (2026), Yang et al. (2026c), Zhang et al. (2026j)	—

Abstraction level determines transferability. The dominant form follows a shared pipeline (*execute* → *extract* → *store* → *retrieve* → *reuse*), differentiated primarily by the abstraction level of the stored artifact. Memory entries range from raw trajectories (Fang et al., 2025b; Sarukkai et al., 2025) through workflows (Wang et al., 2025g) to highly abstracted insights (Cai et al., 2025b; Ouyang et al., 2026b). Kim et al. (2026) provide direct evidence that abstraction strongly predicts transferability: insights transfer better than summaries, summaries better than workflows, and workflows better than raw trajectories. The spectrum extends to intermediate artifacts (Bai et al., 2025; Zhuang et al., 2024; Li et al., 2026f), versioned registries (Lin et al., 2026b), dynamically maintained knowledge bases (Cai et al., 2025a), and domain-tailored schemas that achieve precise within-domain retrieval at the cost of portability (Ren et al., 2026b; Shen et al., 2026b; Zhang et al., 2026n; Zhu et al., 2025) (Table 4).

Feedback signal reliability varies widely across these systems, from deterministic verification—ground-truth comparison (Chen et al., 2025b; Li et al., 2024a) and domain tool execution (SPICE simulation (Bao et al., 2026), GPU profiling (Dong et al., 2026b), and EDA synthesis (Fang et al., 2026))—through multi-stage verification (Li et al., 2026e; Liao et al., 2026) to LLM-as-judge (Ouyang et al., 2026b; Cao et al., 2025).

Persistent relationships favor graph-based representations. When memory encodes persistent relationships, graph-based representations dominate (Zhu et al., 2026a; Rasmussen et al., 2025; Zheng et al., 2026; Han et al., 2026; Zhang et al., 2026e; 2025c; Wu et al., 2026b; Chhikara et al., 2025). Lu et al. (2026a) show that the optimal memory structure varies with context, with a learned classifier selecting among linear, graph, and hierarchical structures per episodic unit. Structured-schema alternatives without explicit graph topology also persist (Gadzhiev & Kislov, 2026; Packer et al., 2023; Xu et al., 2025a; Westhäußer et al., 2025); narrative simulation presents a distinct variant (Park et al., 2023; He et al., 2026b) (Table 4).

Scaling and lifecycle maturity. Fang et al. (2025b) provide a detailed scaling analysis of the accumulation ceiling. Their results show that retrieval count exhibits an optimal range, beyond which excess memories degrade performance. Cao et al. (2025) demonstrate a comprehensive lifecycle, combining experience distillation, adaptive reuse, and utility-based pruning within a single system; their results show that memory quality can substitute for model scale. Related lifecycle innovations include Meta-MDP operations (Cai et al., 2025b), test-time logit modulation (Li et al., 2026h), and live memory evolution (Zhang et al., 2026r).

He et al. (2025b) report the largest-scale deployed $\mathcal{H}\neq 0$ memory system in the survey, in production at TikTok Pay across 150M+ monthly active users: the system selectively queries human experts under uncertainty, achieving 0.83–0.89 sensitivity at query budgets of 100–1,000; in this non-stationary domain, autonomous baselines achieved lower performance.

Curation outperforms accumulation but adds complexity. When entries carry quality scores, population-based operators yield some of the strongest purely structural results. Wang & Jiang (2026a) propose PRIME with biologically inspired curation operators (mutation, generalization, crossover, and pruning). Wu et al. (2025a) couple dynamic scoring with RL policy training ($\Delta{+}\nabla$). Additional curation mechanisms include constrained FSM operators (Zhang et al., 2026m), RL-assigned quality scores (Wang & Jiang, 2026b), failure-triggered skill evolution (Xia et al., 2026), multi-agent memory coordination (Lin et al., 2026c), and bounded optimization memories (Zhang et al., 2025b). Retrieval innovation evolves the form of retrieved artifacts: query-conditioned reasoning abstractions (Feng et al., 2026b), meta-buffers of reasoning templates (Yang et al., 2024b), graph-based evidence structures (Penaroza, 2026), and evolving latent-preference hypotheses (Yoon et al., 2026). Formal frameworks unifying curation designs include the Stateful Reflective Decision Process (Wang, 2025) and the Memory-augmented MDP (Zhou et al., 2025b); Lam et al. (2026)'s SSGM provides reconciliation-bounded governance for memory safety; none accounts for evolutionary curation operators. Across diverse settings, failures produce more informative memory entries than successes (Xie et al., 2026b; Ren et al., 2026b; Yu et al., 2026; Wang & Jiang, 2026a); yet most systems store only successful outcomes, leaving what Ren et al. (2026b) call the error-driven dividend unrealized.

Forgetting: critically understudied. Despite evidence that unbounded accumulation degrades performance (Fang et al., 2025b; Kwon et al., 2026), very few of the surveyed Memory & Sense papers implement explicit forgetting. The underlying stability–plasticity dilemma—superseding outdated beliefs without destroying useful context—is addressed by renormalization group theory (Tian et al., 2025), Piaget-inspired assimilation and accommodation (Piaget & Cook, 1952; Zheng et al., 2026), and Hebbian decay (Hebb, 1949; Zhu et al., 2026a). ZenBrain (Bering, 2026) integrates fifteen neuroscience-inspired mechanisms into a seven-layer architecture. Memory Worth (Simsek, 2026) offers a simpler two-counter signal with convergence guarantees. An intermediate approach uses Ebbinghaus-inspired decay with adaptive reinforcement (Gu et al., 2026).

Meta-evolution and hybrid designs. A small body of work searches over memory architectures themselves: M* (Pan et al., 2026) evolves executable Python programs defining memory schemas; MemEvolve

(Zhang et al., 2025d) evolves the full pipeline via multi-objective selection; Zhang et al. (2026d) reframe operations as learnable skills with an RL-trained controller ($\Delta+\nabla$); and PRISM (Mishra, 2026) unifies memory types under a decision-theoretic framework with convergence guarantees. Purely parametric consolidation remains rare, confined to world models evolving the Sense component (Hafner et al., 2025; Ha & Schmidhuber, 2018; Qiu et al., 2026), test-time training (Sun et al., 2025b), and selective sparse memory updates (Lin et al., 2025). A recurring hybrid pattern, *store structurally, retrieve parametrically*, appears across multiple systems (Zhou et al., 2025b; Zhang et al., 2026d; Wu et al., 2025a; Wang & Jiang, 2026b; Cai et al., 2026; Liao et al., 2026; Qiao et al., 2026; Zhang et al., 2026j; Yang et al., 2026c) (Table 4).

Synthesis. The accumulation ceiling documented above has a structural counterpart in Cortex context overflow (§3.1); Section 4 examines the shared pathway-level root. Beyond the accumulation–curation axis, two areas remain thin: forgetting and Sense evolution (the latter has received minimal attention, with Ren et al. (2026a) as the sole structural contribution among the surveyed papers). Individual papers borrow selectively from cognitive science (Ebbinghaus forgetting (Ebbinghaus, 1885; Zhong et al., 2024; Gu et al., 2026), Hebbian plasticity (Hebb, 1949; Zhu et al., 2026a), Piaget's schemas (Piaget & Cook, 1952; Zheng et al., 2026), free-energy principle (Friston, 2010; Ma et al., 2025), and sleep-based consolidation (Bering, 2026)), but none of the surveyed papers maps the full space of cognitive memory mechanisms onto agent memory design.

Key Takeaways: Memory & Sense

- **The accumulation ceiling parallels Cortex context overflow.** In both substrates, structural artifacts grow with experience; Section 4 traces the shared pathway-level root.
- **"Store structurally, retrieve parametrically" is a recurring hybrid pattern.** Despite near-universal structural consolidation, numerous papers pair persistent structural memory with parametric policies trained to use it.
- **Failures are more informative than successes, yet underexploited.** Evidence from multiple domains shows that error-driven memory entries carry more learning signal, yet most systems store only successful outcomes.

3.4 Cross-Substrate Interactions

The preceding three subsections treat Cortex, Action, and Memory as independent evolutionary targets. In practice, the substrates are coupled: changes in one propagate to, constrain, or enable changes in another. Of the surveyed evolution papers, a substantial minority appears in multiple taxonomy cells. Full three-substrate coupling remains rare; HERA (Li & Ramakrishnan, 2026), EvoFSM (Zhang et al., 2026m), and AutoAgent (Wang et al., 2026f) all use structural (Δ) pathways. Among pairwise couplings, Action–Memory coupling is exclusively structural across the surveyed papers, extending the structural dominance documented in §3.2 and 3.3. No paper among those surveyed studies the dynamics of cross-substrate coupling over extended horizons.

Shared patterns across substrates. Several patterns recur across substrates without requiring direct coupling. In each substrate, the strongest reported results co-occur with deterministic or formally verified evaluation signals: code executors and contest judges in Cortex (§3.1), formal verification and execution-based scoring in Action (§3.2), and ground-truth comparison and domain-tool execution in Memory (§3.3). Where such signals are unavailable, distinct degradation mechanisms appear—self-referential score convergence in Cortex (Wang et al., 2025d), and retrieval degradation past an optimal count in Memory (Fang et al., 2025b)—but the association between evaluation quality and sustained improvement is consistent across the substrates examined. Section 5 analyzes this pattern in depth as a property of Axis III.

Key Takeaways: Cross-Substrate Interactions

- **Evaluation-signal quality is the clearest cross-substrate correlate of sustained improvement.** Across Cortex, Action, and Memory, the strongest reported results co-occur with

deterministic or formally verified evaluation; where such signals are absent, distinct degradation mechanisms emerge.

Substrate choice constrains the consolidation pathway: Cortex spans the full Δ–∇ continuum, while Action and Memory are overwhelmingly structural. Section 4 examines this continuum directly, asking *how* transient experience becomes persistent change.

4 Axis II: Consolidation Pathway

4.1 Constraints on Pathway Selection

Having analyzed what evolves in each substrate (Section 3), we now examine how change persists: the consolidation pathway axis. Structural consolidation dominates overall, but not uniformly: Memory and Action papers are almost exclusively structural, while Cortex spans the full Δ–∇ continuum. Three properties of each substrate—artifact discreteness, evaluation-signal verifiability, and infrastructure access—jointly constrain which pathway is adopted.

Artifact discreteness. Action artifacts (tool specifications, skill code, and workflow graphs) and Memory entries (facts, rules, and cases) are inherently symbolic; gradients have no natural target in a Python function's structure or a knowledge-graph triple. Hu et al. (2025a) illustrate the complementary point: representing agents as code provides Turing-complete expressiveness, and the discrete, symbolic nature of the search space makes LLM-based mutation an effective optimization mechanism. Cortex weights, by contrast, are continuous by construction, and the entire RL and SFT apparatus is designed for differentiable objectives. This discreteness asymmetry explains why parametric methods concentrate in Cortex while remaining rare elsewhere. The few parametric papers in Action target orchestration routing and continuous control; in Memory & Sense, parametric methods are confined to world models, test-time training, sparse memory fine-tuning, and learned retrieval policies—the only sub-problems within these substrates that present continuous optimization targets.

Signal verifiability. Parametric consolidation achieves its strongest results when evaluation signals are deterministically verifiable (code execution, exact-match answers, and game outcomes) because such signals maintain their accuracy across iterations of continuous optimization, whereas self-referential or proxy-based signals can degrade as the system evolves (§5). Zhao et al. (2025) demonstrate the limiting case: their code-executor-only self-play (§3.1) provides deterministically verifiable rewards at zero annotation cost. Section 5 formalizes this as the strongest tier in a three-tier signal hierarchy. Where such signals are unavailable (open-ended dialogue, knowledge curation, and strategy selection), structural methods that rely on LLM-as-judge or heuristic scoring dominate instead.

Infrastructure access. When the base model is a frozen proprietary API, structural consolidation is not a preference but a necessity: no gradient signal can reach weights the developer cannot access. Early agentic evolution systems illustrate this constraint directly: Park et al. (2023)'s generative agents and Shinn et al. (2023)'s Reflexion are both purely structural designs requiring no weight updates, a property that made them naturally compatible with the API-only deployment regime that dominated 2023. Where weight access and training infrastructure are available, parametric and hybrid consolidation become feasible: the GRPO-based training adopted by a substantial fraction of Cortex papers would be impossible without weight access. The structural–parametric balance thus depends partly on infrastructure availability, not only on substrate properties.

How the three constraints interact. These three constraints are not independent: they reinforce one another to produce a single overarching trade-off. Structural methods are more accessible but hit hard ceilings. Parametric methods escape those ceilings through compression into weights but require weight access and verifiable evaluation signals, and risk instabilities absent from structural evolution. Neither

pathway dominates unconditionally; the optimal choice depends on the three constraints identified above. The hybrid share remains small.

4.2 Structural Consolidation

Structural methods across all three substrates share properties that explain their dominance and their limits: human-readable artifacts, compositional modularity, and a hard ceiling imposed by finite context or finite search budgets. Section 3 discusses these ceilings as substrate-specific challenges (context overflow, library growth, and retrieval degradation); viewed across substrates, they are manifestations of a single structural constraint.

Interpretability and compositionality. Whether the evolved artifact is a prompt entry, a skill program, or a memory record, structural consolidation produces outputs that humans can inspect, edit, and selectively roll back, a property parametric methods lack at the level of individual artifacts. Evolved constitutions for multi-agent coordination (Kumar et al., 2026c), JavaScript skill code in Minecraft (Wang et al., 2024a), and timestamped knowledge entries in a production compliance system (He et al., 2025b) are all auditable without interpretability tools. This transparency makes structural evolution particularly suited to high-stakes settings where accountability requires tracing each change to its source. Compositionality compounds this advantage: structural artifacts can be selectively combined, deleted, or version-controlled, so each candidate change can be screened before it enters the system. Zhang (2026) introduce gated admission for joint Cortex–Action evolution; Zhou et al. (2026) impose test-based gates on Action skill mutations; and Liao et al. (2026) apply confidence-based gating in Memory. The mechanism differs across substrates, but the underlying rationale is shared: discrete artifacts can be screened individually before integration. Quality gating is a natural consequence of structural consolidation itself.

Unbounded accumulation. The complementary weakness of compositionality is monotonic growth: structural artifacts accumulate without natural compression, and the resulting growth degrades performance through mechanisms that differ across substrates but share a common root. In Cortex, Reflexion's bounded reflection buffer sacrifices long-horizon experience for immediate relevance. Zhang et al. (2026k) diagnose the underlying failure mode as *context collapse*: monolithic rewriting by LLMs degrades accumulated knowledge into progressively shorter summaries. In Action, Wang et al. (2024a)'s append-only skill library grows without pruning, retrieving only the top-5 most relevant skills per task, a fixed window over a growing collection; Wu et al. (2026a) respond with explicit curation operations governed by trajectory-derived contracts, treating library curation as a first-class design concern. In Memory, Fang et al. (2025b)'s scaling analysis identifies an optimal retrieval count beyond which additional entries degrade performance. Simsek (2026) formalize the deprecation problem through a principled Memory Worth metric with convergence guarantees, specifying when entries should be suppressed or removed. The shared root cause is not substrate-specific: structural artifacts occupy representational space (context tokens, library indices, and retrieval candidates) that grows with experience, while the mechanisms that select among them do not scale accordingly.

Search-budget constraints. Population-based structural search faces a bottleneck distinct from unbounded accumulation: each candidate evaluation requires full agent execution, and the resulting cost constrains search scale. In Cortex, evolutionary optimization typically operates with modest populations over tens of generations (Kumar et al., 2026c; Fernando et al., 2024) because each fitness evaluation runs the complete agent pipeline. In Action, Romera-Paredes et al. (2024) overcome this through island-based parallel evaluation and Novikov et al. (2025) through cascaded filtering, achieving new mathematical results, but their evaluation budgets remain finite: the candidate space grows combinatorially with program length. In Memory, Pan et al. (2026) apply population-based search to memory architecture itself, where each candidate evaluation requires running the full agent on validation episodes. Despite these constraints, the structural pathway's strongest results come from regimes where distributed evaluation at scale enables substantial improvements (e.g., Novikov et al. 2025).

Individual addressability versus holistic optimization. These shared strengths and ceilings produce a single cross-pathway trade-off: structural modularity enables selective rollback but complicates holistic

optimization. A bad prompt entry can be deleted, a failing skill can be rewritten, and an outdated memory record can be pruned. But when a behavioral problem arises from the *interaction* among many structural artifacts (e.g., conflicting rules in a growing constitution, or redundant skills that interfere at retrieval time), identifying the source is difficult because no single artifact is at fault. Parametric methods face the inverse problem: gradient updates can reshape distributed representations holistically, but they cannot reliably isolate and remove a single learned behavior; the very continuity that enables smooth optimization prevents surgical removal. The trade-off is thus representational, not incidental: discrete artifacts are individually addressable but collectively opaque; continuous weights are collectively optimizable but individually opaque.

4.3 Parametric Consolidation

Parametric consolidation escapes the context ceiling and search-budget bottleneck of structural methods, but its efficacy depends on verifiable evaluation signals, and it introduces instabilities absent from structural evolution. The signal-verifiability constraint established above finds its clearest cross-substrate confirmation in the rare parametric papers outside Cortex: DreamerV3's world model generalizes broadly because environment dynamics provide dense supervision, and Jiang et al. (2026b), one of the few purely parametric Action papers, reformulate multi-agent topology selection as a cooperative MARL problem with verifiable task outcomes. Where such signals are absent, structural methods dominate regardless of substrate.

Parametric instabilities. Most parametric Cortex papers adopt GRPO (Shao et al., 2024) or close variants, reflecting the algorithm's engineering fit with sparse, binary reward signals. The failure modes documented below are not specific to GRPO, however: they have been observed across multiple RL algorithms and appear inherent to policy optimization in continuous weight spaces. The four independently identified failure modes catalogued in Section 3.1 (Echo Trap (Wang et al., 2025f), Template Collapse (Wang et al., 2026h), Information Self-Locking (Zou et al., 2026a), and the three-phase reward hacking rebound (Wu & Tang, 2026)) are mechanistically distinct but share a defining property: each arises from continuous optimization dynamics in weight space, and none has a structural counterpart. The discrete addressability documented above has no parametric counterpart: a collapsed policy resists selective "unlearning" because the damage is distributed across weight space rather than localized in a discrete artifact.

Catastrophic forgetting adds a cross-task dimension to these within-task instabilities: new parametric updates overwrite prior knowledge because weight capacity is finite and shared. Lin et al. (2025) quantify the severity directly, finding 89% F1 degradation from standard full fine-tuning versus only 11% with selective sparse memory updates. Catastrophic forgetting is a well-known consequence of optimizing in shared weight spaces, and the surveyed literature confirms it spans substrates: selective sparse updates mitigate it in memory pools (Lin et al., 2025), drift-selected LoRA adapters in personalization agents (Kim & Kim, 2026), and natural low-rank confinement in embodied policies (Hu et al., 2026). The solutions differ; the underlying vulnerability does not. Structural methods do not exhibit these specific failure modes: no structural artifact collapses into a degenerate attractor or overwrites previously stored entries, because each entry occupies its own discrete slot. However, retrieval degradation can render entries functionally inaccessible (as documented above)—a mechanistically distinct but functionally similar form of prior-knowledge loss. This partly explains why structural consolidation persists despite the performance ceilings documented in the preceding paragraphs; the reliability of discrete, individually addressable artifacts is itself a form of robustness that parametric compression sacrifices.

The confinement of parametric methods to the narrow niches enumerated above follows from the artifact-discreteness constraint: where the target artifact is symbolic, gradients have no natural target. The complementary failure modes of the two pathways (ceilings that cap structural performance and instabilities that destabilize parametric training) motivate a third design: hybrid systems that combine both.

4.4 Hybrid Consolidation

Several dozen papers explicitly couple structural and parametric consolidation. A recurring lifecycle, *structural discovery followed by parametric consolidation* (hereafter $\Delta \rightarrow \nabla$), appears most clearly in Cortex and Action, though it is not the only hybrid pattern. SKILL0 (Lu et al., 2026b) progressively withdraws struc-

tural skill prompts during training, forcing the model to absorb them into weights; ERL (Shi et al., 2026) (§3.1) takes the opposite approach, generating structural reflections that are retrospectively absorbed. The two occupy opposite ends of the $\Delta \rightarrow \nabla$ spectrum: scaffolding withdrawal versus scaffolding generation followed by absorption. In Action, Su et al. (2026a) couple evolutionary selection over solution populations (Δ on $\mathcal{A}$) with policy training (∇ on Π): evolutionary search handles discrete candidate evaluation while gradient optimization handles continuous policy improvement. SEVerA (Banerjee et al., 2026) separates the roles more sharply: verified program structure provides parameter-independent correctness that survives weight updates. In Memory, hybrid papers converge on a "store structurally, retrieve adaptively" motif (§3.3): structural entries provide interpretable, composable knowledge while learned retrieval mechanisms discover usage patterns beyond static rules. This motif differs from the Cortex and Action lifecycles in a critical respect: it maintains both pathways persistently rather than compressing Δ into ∇, because the addressability requirement central to most Memory systems (the ability to inspect, update, and delete individual knowledge entries) is difficult to reconcile with full parametric consolidation. All three substrates produce hybrid designs, yet only Cortex and Action follow the $\Delta \rightarrow \nabla$ compression lifecycle; Memory resists it. The exception reveals a constraint: even when both pathways are available, substrate properties shape which hybrid form emerges.

Persistent joint evolution as an alternative to compression. The $\Delta \rightarrow \nabla$ lifecycle assumes eventual convergence to a parametric end-state. A second hybrid pattern, persistent joint evolution, instead maintains both pathways indefinitely as complementary rather than sequential components. The three clearest instantiations span a coupling spectrum. At the tight end, INSPO's within-step instruction sampling (Zhou et al., 2025a) directly conditions every gradient update. At an intermediate coupling, SkillRL's periodic skill evolution (Xia et al., 2026) interacts with policy training only at scheduled synchronization points. At the loosest coupling, MIA's two-stage alternating training (Qiao et al., 2026) separates the optimization loops entirely, with each pathway's outputs feeding the other through distinct training stages. The coupling dimension reveals a design trade-off absent from the $\Delta \rightarrow \nabla$ lifecycle. Tight coupling enables within-step mutual adaptation but demands careful synchronization to prevent the structural population from destabilizing gradient signal. Loose coupling simplifies implementation but forgoes real-time synergy and risks the two pathways drifting apart. Empirically, all three systems outperform their single-pathway ablations, confirming that persistent joint evolution captures complementarities lost under compression.

Functional roles of the two pathways. The hybrid designs in the preceding paragraphs share a functional pattern: the structural component maintains explicit, human-readable artifacts (prompt rules, skill programs, and knowledge entries) while the parametric component absorbs execution competence into weights. Zhang et al. (2026g) demonstrate this pattern directly: in E-SPL, competitively rated prompt populations (Δ) convert observed mistakes into explicit prompt rules while gradient updates (∇) internalize execution competence, and jointly optimizing both improves cross-task generalization over either alone. The pattern reflects the representational properties documented above. Structural artifacts are human-readable and individually addressable, which makes them easy to inspect and revise when strategies must change. Parametric weights generalize automatically across inputs, making them effective for absorbing execution competence that recurs across tasks. This complementarity reframes the $\Delta \rightarrow \nabla$ lifecycle and persistent joint evolution not as competing hybrid designs but as two scheduling strategies for combining explicit artifacts with parametric weights. Compression collapses the explicit component into weights once strategies stabilize; persistent joint evolution maintains both indefinitely because the task distribution continues to shift. Whether this boundary can be formalized into transition criteria—specifying when a structural artifact is stable enough to be absorbed into weights, or when parametric drift warrants structural correction—remains the central open question for hybrid consolidation. Current hybrid papers approach this question empirically; none resolves it.

Why so few hybrids? Two engineering barriers likely contribute to the small hybrid share. First, coordinating two optimization loops demands infrastructure neither loop requires alone (INSPO, SkillRL, and MIA each solve this synchronization problem differently (Zhou et al., 2025a; Xia et al., 2026; Qiao et al., 2026)). Second, credit assignment between pathways becomes intractable when both change simultaneously—no hybrid paper isolates the marginal contribution of each pathway at the jointly evolved state. These barriers

are implementation obstacles, not principled limitations: the division of labor documented above suggests the two pathways are functionally complementary.

4.5 Cross-Pathway Trade-offs

Table 5: **Cross-substrate properties of structural versus parametric consolidation.** Each row names a property that holds across all three substrates within a given pathway.

Property	**Structural (Δ)**	**Parametric (∇)**
Artifact nature	Symbolic, discrete (prompts, code, and facts)	Continuous (weight matrices)
Evaluation signal	LLM-as-judge, heuristic, or execution-based	Strongest with verifiable evaluation signals (code exec, exact match, and game outcomes)
Infrastructure	No training required; API-only compatible	Weight access + training infrastructure
Shared strengths	Interpretable, composable, and rollback-capable	Experience compression; automatic generalization
Shared ceilings	Unbounded accumulation and search-budget limits	Template collapse, reward hacking, and catastrophic forgetting
Regression defense	Gated admission (stability-prioritizing, unit-test, and confidence-based)	LoRA isolation, selective sparse updates, and low-rank confinement
Core trade-off	Addressability & modularity	Generalization & compression

Regression affects both pathways, for opposite reasons. Both pathways struggle with regression, but their failure modes and defenses are rooted in different pathway properties (Table 5), so no single mechanism can protect both. Structural methods enable selective rollback, yet their artifacts accumulate monotonically, so the probability of undetected interference grows with collection size. The gated-admission mechanisms documented above each emerged independently to contain this risk, but none efficiently detects interference patterns that span many artifacts simultaneously. Parametric methods compress experience into fixed-capacity weight spaces, avoiding accumulation entirely, yet risk catastrophic forgetting that erases prior competence without any mechanism for targeted recovery. Each pathway's regression defense depends on the property the other pathway lacks: structural rollback requires discrete addressability, which parametric representations sacrifice; parametric compression avoids unbounded accumulation, which structural representations do not escape without explicit curation. Hybrid systems that invest most heavily in regression avoidance delegate control to the pathway whose representational properties match the failure mode: SEVerA's verified structure provides parameter-independent correctness (Banerjee et al., 2026); AgentDevel's gating screens structural edits (Zhang, 2026); INSPO's successive-halving pruning discards low-fitness candidates before they condition gradient signal (Zhou et al., 2025a). Regression control is thus pathway-specific: these defense mechanisms are determined by pathway properties (discrete addressability for structural and capacity management for parametric), not by which substrate evolves.

Open questions. When should a practitioner choose Δ, ∇, or $\Delta{+}\nabla$? The preceding analysis supplies partial answers: the three pathway constraints jointly determine the feasible pathway, and the division of labor documented above provides a rationale for separating what evolves structurally from what evolves parametrically (Zhang et al., 2026g). But, to our knowledge, no current framework specifies *transition conditions*. Three questions remain open. First, *when* does a structural method reach its ceiling? The unbounded-accumulation and search-budget ceilings documented above are well characterized qualitatively, yet no formal criterion predicts the point at which structural returns diminish and parametric consolidation becomes cost-effective. The progressive-withdrawal and retrospective-distillation lifecycles all rely on hand-tuned schedules (fixed decay rates and periodic triggers) rather than data-driven transition signals. Second, *does persistent joint evolution outperform eventual compression?* The coupling spectrum shows that both designs yield gains over single-pathway baselines, but, to our knowledge, no study compares them head-to-head over extended training horizons; existing hybrid systems report only end-state performance, so whether the two pathways converge, oscillate, or interfere when run jointly for hundreds of update events remains unknown. Third, *how should regression risk factor into pathway choice?* The regression analysis above shows

that each pathway's defense depends on the representational property the other lacks, yet no design guideline weighs the regression cost of unbounded accumulation against that of catastrophic forgetting, a trade-off that depends on domain-specific factors (update frequency, task-distribution stability, and accountability requirements) which current benchmarks do not measure. Formalizing pathway selection as a function of these factors (artifact discreteness, evaluation-signal verifiability, infrastructure access, deployment horizon, and regression tolerance) remains open.

Key Takeaways: Consolidation Pathway

- **Pathway is substrate-constrained, not freely chosen.** Artifact discreteness, evaluation-signal verifiability, and infrastructure access jointly constrain the consolidation pathway, explaining why parametric methods concentrate in Cortex and remain rare elsewhere.
- **Failure landscapes are pathway-specific, not substrate-specific.** Unbounded accumulation and search-budget ceilings are structural-pathway problems; parametric-pathway problems (e.g., template collapse, reward hacking, and catastrophic forgetting) arise regardless of substrate.
- **Hybrid designs develop independently across substrates but vary in form.** The $\Delta \rightarrow \nabla$ lifecycle (structural discovery followed by parametric consolidation) appears in Cortex and Action but Memory resists it because entries must remain individually addressable; persistent joint evolution provides an alternative that maintains both pathways indefinitely.
- **Regression control is a primary constraint.** Each pathway's defense depends on representational properties the other lacks; hybrid systems delegate regression to pathway-matched mechanisms.
- **The pathway selection problem is open.** No current framework determines when to use Δ, ∇, or $\Delta+\nabla$ as a function of artifact discreteness, evaluation-signal verifiability, infrastructure access, deployment horizon, and regression tolerance.

The consolidation pathway determines *how* experience becomes persistent change; the selective pressure dimension determines *what signal source drives* that process: whether evolution is driven by environmental experience and self-generated signals alone ($\mathcal{H} = 0$) or additionally by human input ($\mathcal{H} \neq 0$). Section 5 examines the signal sources that drive this process and the constraints they impose.

5 Axis III: Selective Pressure

5.1 Autonomous Selective Pressure

With the structural–parametric continuum characterized in Section 4, we now turn to the third axis: what source of selective pressure drives evolution. Of the surveyed evolution papers, roughly 90% operate entirely under autonomous selective pressure ($\mathcal{H} = 0$), the most lopsided distributional finding on any axis of the three-axis framework. This asymmetry has three reinforcing causes, each operating on a different subset of the literature. The first is principled: where rewards are deterministically verifiable (code execution, exact-match answers, and game outcomes), human feedback is genuinely unnecessary. Zhao et al. (2025) demonstrate the limiting case: a single model generates, solves, and validates code-based reasoning tasks using only a code executor, with zero human involvement at any stage. The second cause is architectural: most structural evolution systems (Δ) evolve artifacts through automated optimization loops (evolutionary search, LLM-as-optimizer, and self-reflection) whose search and selection mechanisms operate autonomously, with no interface for real-time human input. The third is economic: human-in-the-loop evaluation is expensive and does not scale, restricting $\mathcal{H} \neq 0$ designs to settings where the return on human effort justifies the cost. These causes are complementary, not competing: the first explains why $\mathcal{H} = 0$ *succeeds* in verifiable domains, the second why it *persists* across automated evolutionary pipelines, and the third why $\mathcal{H} \neq 0$ remains *rare* even when it would be beneficial—and their alignment explains the extremity of the concentration.

The roughly 25 $\mathcal{H} \neq 0$ papers reinforce this analysis: they concentrate where automated verification is weakest, suggesting that the asymmetry reflects a rational allocation rather than a blanket preference for autonomy. The distributional details appear below. Whether this allocation is optimal or merely convenient—

whether the field gravitates toward verifiable domains partly *because* they permit $\mathcal{H} = 0$—is a question the distributional data alone cannot resolve. The evaluation-signal analysis below can begin to address it.

The $\mathcal{H} = 0$ papers do not form a monolithic category: the quality of the autonomous selective pressure they employ varies systematically, and this variation appears to correlate with empirical outcomes across substrates and pathways. We classify autonomous evaluation signals into three tiers based on the verifier's relationship to the system being evaluated: Tier 1 (deterministic verification by an external oracle), Tier 2 (LLM-as-judge or self-consistency evaluation), and Tier 3 (learned proxy signals). Section 3's substrate-organized narrative touches on each signal type within individual substrates; the analysis below formalizes them as a single hierarchy and examines their cross-substrate properties.

Tier 1: deterministic verification. Among the clearest empirical results in the survey are systems whose selective pressure reduces to a deterministic oracle (code execution, exact-match comparison, or formal proof) because the verifier is external to the system being evolved, eliminating the evaluator–agent coupling that degrades higher tiers. Zhao et al. (2025) demonstrate the purest case in Cortex: the Absolute Zero Reasoner's code-executor-only selective pressure. In Action, Romera-Paredes et al. (2024) and Novikov et al. (2025) achieve new mathematical and algorithmic discoveries through automated evaluator functions, demonstrating that strong verification enables results beyond prior best-known bounds. Execution-based verification extends to further domain-specific tools: logic synthesis checking (Fang et al., 2026), GPU profiling (Dong et al., 2026b), and SPICE simulation (Bao et al., 2026). Li et al. (2026g) report winning live competitive programming contests against all human participants through execution-based rewards. The cross-substrate pattern is consistent: across the surveyed papers, where a deterministic oracle exists, autonomous selective pressure has generally sufficed, though specification gaps can be exploited under sustained optimization (Wu & Tang, 2026).

Tier 2: LLM-as-judge and self-consistency. Where deterministic verification is unavailable, numerous papers substitute the model's own judgment or inter-model consensus as selective pressure, a design that trades verifier independence for domain generality, because the same LLM-as-judge mechanism applies to open-ended reasoning, dialogue, and knowledge curation without requiring a task-specific oracle. Yuan et al. (2024) establish the paradigm in which a single LLM serves as both policy and reward model (§3.1). The central failure mode is score-gap convergence between chosen and rejected responses (§3.1), which renders the training gradient ineffective (Wang et al., 2025d). Fu et al. (2026) prove that iterative self-rewarding alignment acts as a contraction mapping, formalizing why iteration is necessary rather than merely helpful but also why each iteration yields diminishing returns.

Multi-role variants decouple generation from evaluation across distinct agent roles. Chen et al. (2025f) train Proposer, Solver, and Judge roles on a shared backbone, achieving gains across 22 benchmarks without external verifiers. Some self-play systems combine Tier 1 and Tier 2 signals: Peng et al. (2026b) use external verifiers for answer correctness (Tier 1) alongside a Critic agent for question and plan quality (Tier 2). At the meta level, Chojecki (2025) formalize self-play, self-correction, and synthetic data bootstrapping as instances of a Generator-Verifier-Updater operator.

In Memory, Tier 2 signals govern which experiences persist: Ouyang et al. (2026b) distill reasoning strategies using LLM-as-judge labeling, and Cao et al. (2025) employ utility-based pruning within a full-lifecycle pipeline.

Across substrates, Tier 2 exhibits a structural asymmetry absent from Tier 1: LLM-as-judge evaluation degrades with iteration because improvements in generation quality do not automatically improve evaluation quality. Tier 1 verifiers are invariant to which system produced the output (a code executor returns the same result regardless of which model generated the code), though they remain vulnerable to specification gaming (§3.1); Tier 2 judges drift with the system they evaluate, and the gap between generation capability and evaluation capability is unmonitored. The empirical signature is diminishing marginal gains across self-rewarding iterations (Yuan et al., 2024; Wang et al., 2025d), a pattern that Fu et al. (2026)'s contraction analysis explains yet does not resolve.

Tier 3: proxy signals. When neither deterministic verification nor direct self-evaluation applies, a third class of systems interposes a separately constructed proxy between the agent and the evaluative signal: a trained reward model, a simulated environment, or a statistical aggregation of peer judgments. Unlike Tier 2, where the agent evaluates its own outputs directly, Tier 3 signals pass through an intermediary whose fidelity is itself unmonitored, extending autonomous selective pressure to non-verifiable domains at the cost of an unaudited gap between the proxy and the true objective. Qi et al. (2025) train an Outcome-Supervised Reward Model on rollout data for web navigation, enabling a self-evolving curriculum, but the ORM inherits the distributional assumptions of its training data and cannot evaluate tasks outside its coverage. Sui & Hooi (2026) forgo learned proxies in favor of peer consensus via a structured propose-critique-revise protocol: critiques that enable solution improvements earn a diagnostic reward, providing meta-evaluation supervision without ground-truth labels. Song et al. (2026) formalize a middle ground: in their RL from Text Feedback framework, an external feedback provider generates textual critiques during training, and the policy absorbs them—either through self-distillation of its own feedback-conditioned revisions or through an auxiliary feedback-prediction objective—so that single-turn performance improves without requiring feedback at inference. The text feedback is richer than a binary signal but remains a proxy for quality dimensions the feedback provider was not designed to evaluate. Further proxy mechanisms include Monte Carlo estimation of multi-turn rewards via user simulation (Wu et al., 2025b) and majority-voting reward generation (Fang et al., 2025d).

The failure mode specific to Tier 3 is that the learned approximation is fixed at training time while the agent evolves; the gap between proxy and true objective can widen silently across iterations. Even Tier 1 verifiers are not immune to exploitation: Wu & Tang (2026)'s three-phase rebound pattern (§3.1) demonstrates that deterministic verifiers can be gamed under sustained optimization. If deterministic verifiers can be exploited, learned proxies are more vulnerable still: the policy can exploit weaknesses in the proxy that no external mechanism detects. Tier 3 methods thus extend autonomous selective pressure to the broadest set of domains (e.g., creative writing, open-ended dialogue, and specialized reasoning) but with the weakest fidelity guarantees among the three tiers.

The verifier boundary of autonomous evolution. As selective pressure moves from deterministic verification (Tier 1) through self-evaluation (Tier 2) to learned proxies (Tier 3), the verifier's capacity to maintain accurate evaluation as the system evolves decreases, and with it the reliability of the evolutionary signal. This spectrum exposes a structural gap: the domains where autonomous selective pressure is most reliable tend to be those where correctness is formally specifiable (code correctness and mathematical truth) rather than value-laden, while the domains where alignment matters most lack verifiers whose accuracy is independent of the system being evaluated. No surveyed paper proposes a meta-evaluation mechanism that monitors evaluation-signal quality over the course of evolution. The spectrum thus defines a verifier boundary for autonomous evolution:

Finding: The Verification-Signal Constraint on Autonomous Evolution

Across the surveyed papers, autonomous evolution produces its strongest and most sustained results where evaluation relies on deterministic verifiers independent of the system being evaluated. Where such verifiers are unavailable, the evaluation signal's fidelity tends to decrease across evolutionary iterations, and reported outcomes tend to degrade or depend on compensating mechanisms not yet validated under iterative evolution.

In Tier 1 domains, the verifier is independent of the system being evolved, so its judgments do not degrade as the system changes—though the scope of what the verifier tests may not cover all dimensions of the intended objective, as specification gaming demonstrates (Wu & Tang, 2026). In Tier 2 and Tier 3 domains, the boundary contracts as evolution proceeds: each iteration may degrade the LLM-as-judge evaluator (Tier 2) or widen the proxy–objective gap (Tier 3), yet no system in the survey implements a runtime mechanism to detect when verification fidelity has fallen below a useful threshold. Autonomous evolution's effective reach is therefore bounded by the verification signal's capacity to remain trustworthy as the system evolves—not by the model's capacity to improve. Current systems neither measure nor report this constraint, and the $\mathcal{H} \neq 0$ analysis below will show that human selective pressure compensates only partially.

The remaining $\mathcal{H} \neq 0$ papers introduce a qualitatively different signal source, human judgment, that is in principle independent of the system it evaluates but brings its own constraints of cost, scalability, and temporal drift.

5.2 Human-Involved Selective Pressure

These $\mathcal{H} \neq 0$ papers do not distribute uniformly across the five forms of human-involved input defined in Section 2.4: two forms (evaluative feedback and implicit signal) account for the vast majority of cases, while direct prescription, demonstration, and interactive collaboration remain nearly empty—the real-time steering that the co-evolution framing might suggest is largely absent.

Evaluative feedback: the dominant explicit form. Among these five forms, evaluative feedback (human judgments expressed as preferences, ratings, or corrections) is the most common intentional channel. Its dominance reflects a cost–information trade-off: evaluative feedback requires less effort than demonstrations yet conveys richer signal than implicit traces. Shivaswamy & Joachims (2015) formalize why even low-quality instances suffice: their coactive learning framework proves sublinear regret from sub-optimal corrections, establishing that convergent parametric evolution is possible from imperfect human input. The remaining evaluative-feedback papers divide along two axes: *personalization* (iterative alignment via temporal progress–regression feedback (Abramson et al., 2022) versus per-user preference alignment (Li et al., 2024b; Nam et al., 2026)) and *noise handling* (Yang et al. 2025b's temporal-consistency purifier, the only system in the survey modeling feedback noise as a first-class concern). Further papers operate at different substrates or pathways: real-time policy-gradient updates from human signals (MacGlashan et al., 2017), graph-search-based optimization of training configurations from task descriptions or deployment feedback (Atreja et al., 2026), prompt optimization via pairwise preferences (Lin et al., 2024), and human-expert-guided tool creation and iterative refinement (Gao et al., 2025).

Implicit signal: the dominant passive form. A smaller cluster derives selective pressure from behavioral patterns accumulated through ordinary use, making implicit signal the second most common yet informationally weakest form. Conversational memory systems form the largest cluster, sharing a common extract-persist-retrieve architecture but differing in the *representation* that mediates persistence: tiered non-graph stores (Packer et al., 2023), graph-structured representations with increasing expressiveness (Chhikara et al., 2025; Rasmussen et al., 2025), and further representational variants (Zhong et al., 2024; Tian et al., 2025; Gadzhiev & Kislov, 2026; Nie et al., 2026b; Westhäußer et al., 2025). Beyond conversation, implicit signals take domain-specific forms where professional activity generates evolutionary pressure: code commit history (Deng et al., 2026) and confirmed clinical diagnoses (Chen et al., 2026c). Ma et al. (2026b) represent a distinct mechanism, collective evolution across multiple users, aggregating trajectories from eight simulated concurrent users into a shared skill library; Zhang et al. (2026a) apply a per-user evolutionary mechanism in foreign trade. He et al. (2025b) occupy a hybrid position straddling evaluative feedback and implicit signal (Section 3.3). The implicit-signal cluster accounts for the largest documented $\mathcal{H} \neq 0$ deployments precisely because they require no feedback infrastructure beyond the task itself, yet the human evaluative input they provide is the least explicitly incorporated into the evolutionary loop.

The three rare forms. The remaining three forms account for very few papers. Demonstration appears in a single paper: Jain et al. (2015) implement coactive trajectory feedback (re-ranking, kinesthetic correction, and interactive markers) via online learning. Interactive collaboration appears in two papers: Mozannar et al. (2025) design six explicit human-in-the-loop mechanisms spanning planning through consolidation, and Adnan et al. (2025) implement progressive dialogue-driven service generation. Direct prescription is absent as a live evolutionary input; Kumar et al. (2026c) evolve constitutions through automated search rather than injecting prescriptive constraints during evolution. With so few papers, this scarcity is better read as an existence proof than as evidence of a stable pattern: current systems absorb human signal passively or elicit it cheaply, and interfaces for sustained high-bandwidth interaction remain largely unbuilt.

Distributional pattern and its interpretation. Of the $\mathcal{H} \neq 0$ papers, about half target Memory, about a third target Cortex, and a handful target Action. Human selective pressure concentrates precisely

where Tier 1 verification is unavailable: the Memory $\mathcal{H} \neq 0$ papers target domains where correctness is subjective or context-dependent; the Cortex $\mathcal{H} \neq 0$ papers employ preference-based training because open-ended reasoning quality cannot be reduced to pass/fail execution. Two cells are entirely empty ($\mathcal{A}\times\nabla\times\mathcal{H} \neq 0$ and $\mathcal{M} \times \nabla \times \mathcal{H} \neq 0$) because the rare parametric methods in these substrates already possess objective evaluation signals. This substrate-level breakdown confirms the complementarity: human signal enters where automated verification cannot reach, provided the cost is acceptable. This complementarity admits the same two readings raised at the outset of this section: rational allocation versus convenience-driven domain selection.

5.3 The Selective-Pressure Gap

The preceding analysis established that autonomous selective pressure has a verifier boundary and that human selective pressure is sparse and low-bandwidth. Their sharpest joint consequence is a selective-pressure gap that compounds across four layers: the absence of human signal in safety-oriented work, the frozen-preference architecture underlying that absence, the bandwidth constraints that prevent existing $\mathcal{H} \neq 0$ systems from compensating, and the instability of the human evaluator itself.

Safety-oriented papers and their $\mathcal{H} = 0$ limitation. This gap is sharpest where the survey's axes intersect: the systems powerful enough to reshape their own reasoning operate almost exclusively under $\mathcal{H} = 0$. Only a handful of papers explicitly target safety, alignment, or formal correctness as an objective of the evolutionary process, and all rely on autonomous selective pressure. Three operate parametrically through adversarial or cooperative games: Beigi et al. (2026) train a Hacker–Auditor adversarial game and then use the trained Auditor to gate rewards during RLHF; Zhang et al. (2026f) formulate safety alignment as a positive-sum multi-agent game; and Ali et al. (2026) combine evolving security-aware constitutional principles with DPO-based unlearning. Two operate structurally: Gallego (2026) discover safety specifications from binary danger signals, converging within one to two rounds; and Kumar et al. (2026c) evolve behavioral norms via multi-island genetic programming. Banerjee et al. (2026) achieve zero constraint violations through their formally verified hybrid approach (§3.1). Each of these approaches targets safety within the formal scope of its verifiers, and none addresses value alignment where correctness depends on context or evolving social norms.

The frozen-preference problem. The $\mathcal{H}$ boundary defined in Section 2.5 is central to understanding why the selective-pressure gap persists. Reward models trained on historical human preferences encode human signal but deploy without live human input, which we therefore classify as $\mathcal{H} = 0$ in our framework. This means that RLHF-trained models which subsequently self-evolve (including all safety-oriented papers above) operate on a *frozen snapshot* of human preferences captured at the reward-model training step. The evolutionary process then extends beyond that snapshot: as the agent improves or shifts distribution, the reward model's coverage of the new output space is unverified, yet no mechanism in the surveyed systems exists to update the human signal. Han et al. (2025) provide empirical evidence that this gap is consequential: they describe what they term an Alignment Tipping Process (ATP) in which self-evolving agents gradually abandon alignment constraints in favor of self-interested strategies reinforced by environmental feedback. In their single-agent setting, high-reward deviant actions accumulate as in-context examples that progressively override initial alignment; in multi-agent settings, deviant strategies spread through social observation as agents adopt rule violations they observe succeeding in peers, an imitative strategy diffusion dynamic that degrades alignment across both open-source and closed-source models. This result gives the frozen-preference problem empirical weight: when the evolutionary loop generates sufficient reward signal for misaligned behavior, no mechanism in the surveyed $\mathcal{H} = 0$ systems detects or corrects the drift.

A complementary failure mode arises even without preference drift: Zou et al. (2026a)'s information self-locking (§3.1) traps agents in low-information interaction patterns that standard policy gradients are unlikely to escape. The problem compounds when the human partner is also changing: if the user's preferences, expertise, and interaction strategies co-adapt with the system (as Section 6 will examine), then even a periodically refreshed reward model tracks a moving target, and the temporal gap between preference capture and deployment widens on both sides. The frozen-preference problem and the alignment tipping process together show that formally specifiable constraints can be enforced (SEVerA's zero-violation guarantee is

the strongest result on this dimension) but value alignment, where correctness depends on context, culture, or evolving social norms, has not yet been sustained by autonomous selective pressure among the surveyed papers. The field thus faces a selective-pressure gap: systems in the current literature powerful enough to self-evolve their reasoning are constrained either by deterministic verifiers that cannot capture values, or by frozen preference snapshots that cannot track value drift. No surveyed paper proposes a mechanism combining Tier 1 verification reliability with the value sensitivity of human judgment.

Low-bandwidth compensation. Even where human selective pressure exists, its compensating capacity is constrained by a primary design variable: the bandwidth of the human signal channel. Of the $\mathcal{H} \neq 0$ papers, most operate through low-bandwidth channels: evaluative feedback (preference pairs, ratings, and binary corrections) or implicit signal (behavioral traces extracted from ordinary use). Only five employ high-bandwidth interaction. Three are classified at the form level as demonstration or interactive collaboration (Mozannar et al. 2025, Adnan et al. 2025, Jain et al. 2015). Two others are nominally evaluative feedback or implicit signal but involve iterative interaction deep enough to function as high-bandwidth channels (Gao et al. 2025's expert consultation loop and He et al. 2025b's uncertainty-driven human queries at production scale). Current $\mathcal{H} \neq 0$ architectures overwhelmingly restrict human selective pressure to channels that transmit bits per interaction—a preference comparison, a memory write, and a behavioral trace—while the structured, multi-turn steering that demonstration and interactive collaboration could provide remains rare. Whether this restriction reflects engineering cost (high-bandwidth interfaces are harder to build), cognitive cost (sustained interaction fatigues human partners), or both, the consequence is uniform: the human signal available to the evolutionary loop is sparse relative to the autonomous signal it supplements, limiting the rate at which human selective pressure can correct distributional drift.

The human partner is not a fixed resource. The selective pressure analysis above treats the human partner as a signal source whose properties (cost, bandwidth, noise, and temporal stability) constrain the evolutionary loop. Every $\mathcal{H} \neq 0$ system in this survey implicitly assumes that the user who provides feedback at iteration t is functionally equivalent to the user at iteration $t-1$: same preferences, same expertise, and same cognitive strategies. Section 6 examines empirical evidence that this assumption is false.

Key Takeaways: Selective Pressure

- **Selective pressure asymmetry**: Roughly 90% of agentic evolution papers operate under autonomous selective pressure ($\mathcal{H} = 0$), driven jointly by deterministic verifiability, the autonomous design of optimization loops, and the cost of human evaluation.
- **Signal hierarchy**: Autonomous evaluation signals form a three-tier hierarchy from deterministic verification through LLM-as-judge to learned proxies; the tier appears to correlate with empirical outcomes across substrates and pathways.
- **Selective pressure complementarity**: Human selective pressure ($\mathcal{H} \neq 0$) concentrates where automated verification is weakest, predominantly in Memory and Cortex, producing a systematic complementarity rather than random coverage.
- **Selective-pressure gap**: The safety-oriented papers all rely on $\mathcal{H} = 0$, operating on frozen preference snapshots; Han et al. (2025)'s Alignment Tipping Process demonstrates that self-evolving agents can abandon alignment constraints through reinforcement of deviant strategies, and no mechanism in the survey detects this drift.
- **Low bandwidth**: Most $\mathcal{H} \neq 0$ papers rely on low-bandwidth channels (evaluative feedback or implicit signal); only five employ high-bandwidth interaction.

6 Human Adaptation Under Agentic Evolution

Sections 3–5 treat the human partner as a fixed resource whose evaluative capacity does not change across evolutionary iterations. The Axis III analysis showed that beyond the boundary where deterministic verification suffices, agentic evolution depends on human selective pressure; but that analysis could not proceed further without answering a prior question: *is the human partner's evaluative capacity stable under sustained AI interaction?* The scope of this section is accordingly bounded: we examine human-side evidence not to

develop a general theory of human–AI interaction, but to assess a specific reliability assumption—that the human evaluator's capacity is stationary—on which the selective-pressure analysis depends.

The three-axis vocabulary developed in Section 2 transfers to the human side because its categories were drawn from cognitive science in the first place. On Axis I, the relevant human changes are cognitive (deliberative capacity, corresponding to Π), operational (task-execution skill, corresponding to $\mathcal{A}$), and epistemic (knowledge structures and trust calibration, corresponding to $\mathcal{M}$). On Axis II, *cognitive offloading* (Risko & Gilbert, 2016—the use of external tools to reduce cognitive demand) occurs when a practitioner incorporates AI-generated artifacts into their output; *parametric internalization* occurs when the practitioner's intrinsic capability to produce that output independently improves. The latter is the human-side analogue of parametric consolidation (§2.3)—what education research calls durable skill acquisition (Soderstrom & Bjork, 2015)—the process by which tool-assisted experience produces lasting capability gains that persist without the tool. On Axis III, human adaptation is shaped either by externally designed interventions (scaffolded interfaces, institutional policies, and training programs) or proceeds as self-driven, unmanaged change; as with agent-side evolution, the literature documents substantially more instances of the latter. These correspondences are analytical lenses for organizing a heterogeneous empirical literature, not claims of mechanistic equivalence between silicon and biological substrates. We use "selective pressure" as a unifying metaphor for the forces—instructional designs, incentive structures, and AI interface choices—that shape which human adaptations are reinforced or extinguished.

A defining asymmetry separates the two sides: for agents, both Δ and ∇ consolidation represent capability gains; for humans, cognitive offloading and parametric internalization can be *negatively* correlated—the central empirical finding of this section. No study in this literature demonstrates full bidirectional co-evolution within a single system; the evidence documents human adaptation *to* AI, and we synthesize the bidirectional picture in Section 7.

The analysis proceeds in four parts: the performance–capability paradox; the metacognitive and trust mechanisms that sustain it; the collective and societal consequences that emerge when individual-level effects aggregate; and the design principles that moderate direction. We report study designs and sample sizes, prioritizing randomized controlled trials and pre-registered studies where available. Throughout, we use *adaptation* for human-side changes and reserve *co-evolution* for the bidirectional coupling between agent and human trajectories analyzed in Section 7.

6.1 The Performance–Capability Paradox

The most consistent empirical finding among these studies is a decoupling between measurable output and independent capability—the *performance–capability paradox* (related to Koedinger & Aleven, 2007's "assistance dilemma" and the performance–learning distinction of Soderstrom & Bjork, 2015). This decoupling instantiates Axis II's consolidation pathway (§2.3) in reverse: AI enables *cognitive offloading*, in which human output improves through incorporation of AI-generated artifacts, but often fails to produce *parametric internalization*, in which the human's intrinsic capacity to generate that output independently also improves. The pattern is documented across education, professional work, and decision-making, and its direction depends critically on how the AI system is designed.

Education: output up, retention down. In a large-scale RCT (N=839), Bastani et al. (2025) find that students with unrestricted GPT access score 48% higher on practice problems (relative to controls) yet score 17% lower on unassisted exams, a direct dissociation between AI-assisted throughput and retained knowledge. Kumar et al. (2024) report a nuanced result in two pre-registered experiments (N=1,100): coach-like LLM guidance (but not direct LLM answers) significantly reduces subsequent unassisted performance in both divergent thinking originality and convergent thinking accuracy, while direct-answer LLM conditions show no significant harm in either experiment. The result indicates that the mode of AI assistance, not merely its presence, determines whether capacities are impaired. The pattern extends to procedural skills: Liao et al. (2024) find that students using a ChatGPT-integrated programming tool show significantly *decreased* problem-solving ability (attributed to over-reliance on the tool's direct-answer channel), while Georgiou (2025) document reduced cognitive engagement when students work with ChatGPT instead of independently. Abdelghani et al. (2025) find convergent evidence with younger learners (N=63): middle

schoolers cannot discriminate AI answer quality (at-chance discrimination) and achieve only chance-level task success (∼51%) despite unrestricted ChatGPT access, with positive AI attitudes *negatively* predicting interaction quality. Temporal dynamics complicate the picture. Kazemitabaar et al. (2023) find no learning harm in a three-week coding study with novices, concluding that AI code generators show initial promise for scaffolding. Whether this reflects genuine resilience or insufficient exposure duration remains open, but the weight of RCT evidence favors a default toward cognitive offloading without parametric internalization under unguided AI access.

Professional work: generation shifts to selection. The same decoupling appears in professional and applied settings. In a pre-registered RCT with software developers (N=52), Shen & Tamkin (2026) find that AI-assisted participants score 17% lower on a subsequent knowledge quiz (relative to controls), with only three of six observed interaction patterns preserving learning. Wang et al. (2025c) document what they term "synthetic fluency": homework–exam score correlations weaken substantially in AI-available courses, producing gaps of 42–44 points in Calculus III and exceeding 48 points in Linear Algebra once exams shifted to conceptual assessment. In professional writing, Noy & Zhang (2023) find in a pre-registered RCT (N=444) that 68% of treated participants submit ChatGPT's output without editing, with the tool largely substituting for worker effort. These findings document a common pattern: when AI provides complete outputs, humans tend to shift from generation to selection, a shift that may reduce the deliberate practice that builds independent skill (Macnamara et al., 2024; Koch, 2026; Jose et al., 2025; Crowston & Bolici, 2025). A mixed-method review of AI-induced deskilling in medicine identifies the same pattern across 22 systematic and 62 narrative sources, distinguishing skill *erosion* from *upskilling inhibition* (Natali et al., 2025), with parallel concerns documented among UX practitioners (Shukla et al., 2025).

Decision-making: the paradox has boundary conditions. The paradox extends to expert judgment. In a pre-registered RCT with radiologists (N=227), Agarwal et al. (2023) find that AI assistance does not improve diagnostic accuracy despite the AI system outperforming 78% of participating radiologists; humans both underweight AI predictions (assigning them roughly one-third the weight of their own signal) and neglect the statistical dependence between AI and human signals. Two critical counterpoints bound the paradox. First, Becker et al. (2025) show in a randomized experiment (N=16 developers, 246 tasks) that experienced open-source developers are 19% *slower* with AI assistance (relative to unassisted performance), demonstrating that the performance benefit itself can reverse for experts working on familiar codebases. Second, Dell'Acqua et al. (2026) formalize a "jagged technological frontier" in a pre-registered RCT (N=758): AI users complete 12.2% more tasks (relative to controls) inside the frontier but accuracy *decreases* by 19 percentage points on tasks outside it, and AI-assisted responses are rated as more coherent even when the underlying analysis is flawed. Complementing overreliance findings, He et al. (2023) demonstrate a Dunning–Kruger route to *under*-reliance (N=249): participants who overestimate their own competence show significantly lower agreement with AI advice. Together, these results show that the paradox is not universal; it is bounded by task structure, expertise, and the match between AI capability and task demands. Scaffolded designs can attenuate or even eliminate the decoupling (Section 6.4), but under unguided access the performance–capability paradox appears to be the default. For the stationarity question that motivates this section, the implication is direct: if unscaffolded AI use reduces independent evaluative capability, then $\mathcal{H}^{(t)}$—the human input to the evolutionary loop—is not stationary under default interaction conditions.

The paradox persists in part because users systematically fail to detect it in themselves—a metacognitive miscalibration whose mechanisms the next subsection examines.

6.2 Mechanisms: Metacognition, Trust, and Sycophancy

The performance–capability paradox documented above appears to be self-reinforcing: humans systematically fail to recognize AI's effects on their own cognition, and they miscalibrate trust in ways that deepen rather than correct the decoupling. Three mechanisms sustain this cycle.

Metacognitive miscalibration: users systematically overestimate their understanding. Across domains, users perceive benefits that objective measures contradict. The exam-score decline documented above goes unperceived by the students experiencing it (Bastani et al., 2025); the measured developer

slowdown goes unperceived by the developers experiencing it (Becker et al., 2025); and 71% of participants in a controlled study failed to detect sycophantic AI behavior despite post-task evaluation surveys (Bo et al., 2026). Tankelevitch et al. (2024) argue that metacognition provides a unifying framework for understanding GenAI usability challenges: processing fluency of GenAI outputs may lead users to misattribute surface polish to content quality, and metacognitive demands arise across prompting, output evaluation, and automation strategy. This fluency-based illusion operates below conscious deliberation; Riva (2025) characterize it as a "System 0" preprocessing layer that shapes information encounters before Kahneman's Systems 1 and 2 engage, rendering the distortion invisible to the user experiencing it.

Trust miscalibration: system-level deference overrides content-level skepticism. A recurring finding across AI-assisted decision-making studies is an asymmetry between content-level and system-level trust: domain experts may resist deference on specific factual claims (content-level) yet remain vulnerable to the broader authority signal conveyed by an AI system's confident presentation (system-level). This asymmetry may explain why domain expertise alone does not reliably prevent overreliance (Parasuraman et al., 2000; Qazi et al., 2025). The quantitative evidence is striking: in a controlled experiment with 432 students, participants accepted 86% of misleading AI recommendations on average (with only 2.4% underreliance on correct recommendations), and, counterintuitively, post-task trust and appropriate reliance are *negatively* correlated (Pitts et al., 2026). Compliance with AI-generated advice exceeds 65% even for high-stakes personal decisions, with no measurable well-being benefit (Luettgau et al., 2025). In clinical settings, Qazi et al. (2025) demonstrate a 14-percentage-point accuracy drop when physicians receive flawed LLM recommendations, with more-experienced physicians *more* susceptible than less-experienced ones, demonstrating automation bias even among AI-trained physicians.

Explanations intended to support calibration can produce asymmetric effects: Schemmer et al. (2023) find that LIME explanations improve reliance when the AI is correct but do not significantly reduce deference when the AI errs, while Xu et al. (2025b) find that LLM-generated explanations amplify lay users' deference on incorrect AI predictions. Rezaeian et al. (2025) report a convergent asymmetry in clinical breast-cancer detection (N=28): high AI confidence scores reduce diagnostic accuracy while low confidence scores increase diagnosis duration. The pattern indicates that confidence displays shift clinician behavior in opposite directions depending on the AI's expressed certainty; Fregosi et al. (2026) find a similar confidence-driven pattern in a logic-reasoning study at larger scale (N=184). The content–system fault line runs through expertise itself: primary care physicians in Xu et al. (2025b)'s study resist content-level deference, acting as a "cognitive firewall," while experienced physicians in Qazi et al. (2025)'s study are more vulnerable to automation bias.

These mechanisms, however, are tractable. Trust-adaptive interventions can substantially reduce inappropriate reliance (Srinivasan & Thomason, 2026), and Li & Steyvers (2026) show that humans can learn to predict AI correctness, improving from 62% to 86% accuracy through repeated feedback, though 44% of participants fail when the confidence-accuracy mapping is reversed. Yet one systematic driver of trust miscalibration resists individual-level correction: AI sycophancy, which operates not as an isolated defect but as a persistent feature of current LLM training, a pattern that scales from individual interactions to population-level distortion.

Sycophancy as a persistent distortion. Among current LLMs, default behavior is systematically sycophantic, not as an incidental defect amenable to prompt-level fixes, but as a persistent feature traceable to RLHF training incentives (§5). Jacobowitz (2026) identify three functions sycophancy serves: maintaining user control over dialogue direction, masking underlying personality variability to provide consistent interaction, and generating cognitive dependency by degrading tolerance for intellectual complexity. Across 11 models, Cheng et al. (2025) find that LLMs are 45 percentage points more sycophantic than human respondents, with GPT-4o among the most sycophantic despite performing well on explicit sycophancy benchmarks, a dissociation suggesting that standard evaluations fail to capture real-world sycophantic behavior. In a pre-registered experiment with a rule-discovery task, Batista & Griffiths (2026) demonstrate that default GPT-5.1 is statistically indistinguishable from a rule-confirming prompt (confirmed via an exploratory equivalence test): both suppress discovery rates to 5.9–8.4%, compared with 29.5% under random-sequence feedback. Users rarely detect the distortion: Bo et al. (2026) find in a controlled study (N=24 ML novices,

with artificially amplified sycophancy) that 71% of participants fail to identify sycophantic responses, and overreliance rates reach 47.4% with sycophantic AI compared with 3% under low-sycophancy conditions. LLMs also affirm both sides of moral conflicts in 48% of cases (Cheng et al., 2025). This pattern reflects a lack of consistent moral judgment that raises concerns beyond factual accuracy. Attempts to correct sycophancy face a design dilemma: anti-sycophancy overcorrection replaces affirmation bias with different pathologies such as presumptuous judgment (Jacobowitz, 2026), and even newer models average roughly one sycophantic behavior per conversational turn (Kirgis et al., 2026). The relationship between personalization and sycophancy is role-dependent: advisory roles increase epistemic independence while peer-framed roles decrease it (Kelley & Riedl, 2026). Sun & Wang (2026) identify a subtler mechanism (N=224): neutral-toned LLMs that adapt their stance are perceived as *more* authentic than complimentary ones, meaning agreement without overt flattery increases trust precisely because it is perceived as genuine. The mechanisms documented in this subsection—sycophancy reinforcing trust, trust deepening reliance, and reliance degrading evaluative skill—suggest a self-reinforcing dynamic, though the coupled loop has not been empirically tested. For the stationarity question, these dynamics add a further layer: a compromised $\mathcal{H}^{(t)}$ may appear, to the human providing it, indistinguishable from a fully capable one.

6.3 Collective and Societal Consequences

The individual-level distortions documented above do not remain individual. As AI-assisted work becomes routine across professions, these effects aggregate into changes in collective knowledge production, creative diversity, and the professional pipelines that produce future domain experts.

Epistemic infrastructure under pressure. Kwon (2026) formalize the asymmetry between AI-accelerated writing and review capacity through a dynamical systems model of scientific publishing, predicting a "deceptive honeymoon" in which knowledge output peaks briefly before declining well below baseline as queue pressure drives reviewer AI adoption and erodes verification quality. Knowledge communities that traditionally served as epistemic infrastructure are already being displaced: Burtch et al. (2024) show that ChatGPT's release reduced activity on Stack Overflow, disproportionately among newer contributors. By contrast, Reddit communities, where social bonds provide value beyond information retrieval, were buffered against the same displacement, suggesting that primarily informational knowledge ecosystems are most vulnerable. At a deeper level, Guingrich et al. (2026) distinguish "belief offloading" from cognitive offloading (Risko & Gilbert, 2016): while the latter delegates cognitive processes, the former delegates belief formation itself, concentrating epistemic power in the few providers whose models generate the beliefs users adopt. The pattern extends to scientific peer review, where 6.5–16.9% of sentences show signs of AI modification, correlated with lower reviewer confidence and deadline pressure (Liang et al., 2024), indicating that evaluative effort is already shifting at the population level.

AI boosts individual quality but narrows collective diversity. When Si et al. (2024) had an LLM ideation agent independently generate research ideas, the AI-generated ideas were rated higher in novelty than expert-written ones on average, yet only approximately 200 unique ideas emerged from 4,000 seeds per topic (∼5%), indicating that LLMs channel output toward a narrow region of idea space. Ashkinaze et al. (2024) find that high passive AI exposure increases the diversity of *human-generated* ideas within a population but does not improve individual creativity; AI makes ideas *different from each other*, not *better*. In writing, Jo & Raghavan (2026) demonstrate through a pre-registered RCT (N=200) that AI-assisted drafts converge in style and content, though human editing substantially recovers diversity, and originality incentives further increase it; in visual creative output, low prompt originality drives homogenization through feedback loops in which engagement rewards aesthetic conformity rather than novelty (De Rosa Palmini & Cetinic, 2024). Three findings temper the homogenization narrative. First, Cheng & Zhang (2025) identify dual mechanisms (inspiration for simple tasks but fixation for complex ones), suggesting that task complexity mediates the direction of AI's creative influence. Second, Kumar et al. (2026b) demonstrate in two pre-registered experiments (N=315 and N=247) that homogenization may be mitigated by pairing role-specialized, architecturally distinct LLMs: a single LLM increases idea-level similarity relative to a no-AI control (p=.033), but pairing two architecturally distinct LLMs may counteract the effect (p=.950 vs. control). Third, Wang et al. (2026e) document a temporal trajectory in which student journalists initially

deferred to AI-generated content uncritically but, through editorial feedback and repeated use, learned to critically evaluate and selectively override AI outputs (N=5, 14-week case study), suggesting that initial convergence need not be permanent.

The missing junior loop. AI adoption can shift the nature of work from production to evaluation. Simkute et al. (2024) characterize this as a "production-to-evaluation shift" in which the human role narrows to quality judgment over AI-generated output; Bauer et al. (2025) provide a theoretical grounding via their ISAR model, which differentiates four types of AI effects on learning—Inversion, Substitution, Augmentation, and Redefinition—ranging from undermining cognitive engagement to fostering deep learning processes. Randazzo et al. (2025) identify three co-creation modes among 244 management consultants: Cyborg (60%, continuous human–AI integration across all workflow phases), Centaur (14%, selective use of AI as a reference tool while retaining execution), and Self-Automator (27%, near-complete delegation with minimal engagement). That more than a quarter of skilled professionals largely delegated judgment, despite working under review, signals that cognitive offloading without parametric internalization can operate at the organizational level, not only in laboratory settings. Catalini et al. (2026) formalize the resulting "Missing Junior Loop": if AI substitutes for junior work, the apprenticeship pipeline that produces future experts erodes, creating a delayed competence deficit invisible in short-term productivity statistics. Chen et al. (2025d) document a concrete instance in programming education (N=24): beginners with ChatGPT access achieve five-fold higher task scores yet show weaker solution understanding, with complete-solution generation yielding the largest performance–understanding gap. A systematic review of AI in medical education corroborates the pattern: across four studies (N=408), AI improves efficiency and basic knowledge but produces mixed or inferior outcomes for higher-order clinical reasoning (Turney et al., 2026). A labor-market review reports that one U.S. study found a 13% relative employment decline among early-career workers (ages 22–25) in AI-exposed occupations (Rio-Chanona et al., 2025), though Danish evidence in the same review found negligible effects—a divergence that may reflect institutional and labor-market differences.

These collective consequences—weakening epistemic infrastructure, narrowing creative diversity, and contracting the pipeline that produces future evaluators—are not simply individual effects scaled up. They are qualitatively distinct: the missing junior loop, for instance, degrades evaluative capacity not in current users but in the next generation of domain experts who may never acquire the competence that apprenticeship would have provided. These effects operate on the supply of future evaluators rather than on the quality of current ones, extending the non-stationarity from $\mathcal{H}^{(t)}$ at a given time to the pool from which $\mathcal{H}^{(t+k)}$ will be drawn.

6.4 Design Principles for Parametric Internalization

Whether AI use drives cognitive offloading alone or also enables parametric internalization depends on the selective pressures that shape human adaptation; that is, on whether human-side change is self-driven and unmanaged or steered by externally designed interventions. Brynjolfsson et al. (2025) provide evidence that genuine learning occurs: among 5,172 customer service agents, AI access increased productivity by 15% overall (36% for the lowest-skill quintile), and during system outages, workers with more prior AI exposure handled chats faster than their pre-AI baseline, suggesting partial skill acquisition beyond tool dependence. One especially promising approach augments human intermediaries instead of end-users directly. In a preregistered RCT (N=1,787 students), Wang et al. (2024c) show that an AI copilot assisting *tutors*, not students, produces a 4 percentage point gain in student exit-ticket mastery (9 percentage points for students of lower-rated tutors), at a cost of $20 per tutor per year; the mechanism is that treatment tutors shift toward higher-quality pedagogical strategies rather than simply relaying AI-generated answers. More broadly, when systems give humans flexible control over delegation, humans voluntarily retain more tasks: Feng et al. (2026a) find that a co-planning system interleaving human and agent execution steps yields higher steerability than a chat baseline, and researchers in a week-long field deployment valued the ability to explicitly assign steps between themselves and the agent. Three design principles emerge from the broader literature, each supported by controlled experiments.

Scaffolding over supplanting. When AI is designed to scaffold reasoning rather than supply answers (a principle rooted in Vygotsky's zone of proximal development (Vygotsky, 1978)), learning gains are sub-

stantial. Kestin et al. (2025) find that a custom pedagogically designed AI tutor outperforms in-class active learning by 0.73–1.3 standard deviations in a crossover RCT (N=194). Su et al. (2026b) show that a Socratic chatbot elicits 73% more interactions and 40% greater cognitive diversity than a general-purpose chatbot, even when final task performance does not differ, indicating that scaffolded design can improve the *process* of learning independently of outcome gains. Self-reported gains in perceived support for critical and independent thinking appear in Socratic tutoring for research-question formulation (Degen & Asanov, 2025), though objective learning outcomes remain pending; Lee et al. (2024b) find in a 10-week RCT (N=61) that a guidance-based ChatGPT wrapper requiring students to attempt answers before receiving hints significantly improves cognitive engagement, critical thinking, and knowledge retention on delayed tests relative to unrestricted ChatGPT. The contrast with direct-answer AI is stark: the same studies that document learning harm under unguided access (e.g., the −17% exam effect in Bastani et al., 2025) show that scaffolded variants eliminate or sharply reduce the deficit. Dasari et al. (2024) demonstrate the magnitude of this gap in a three-group quasi-experiment (N=29): ChatGPT-only students score M=9.8 versus M=42.6 for teacher-plus-ChatGPT, indicating that unsupported AI access yields dramatically worse learning outcomes.

Oakley et al. (2025) provide a neuroscience-grounded account, arguing that internalized knowledge (engrams, schemata, and neural manifolds) is structurally required for higher-order thinking and error detection via prediction errors, which would make the decoupling a consequence of bypassing knowledge internalization rather than merely reduced practice time. Yet a systematic review of 21 studies in music education finds that empowerment and dependence coexist across AI tool types: while assessment AI consistently supports self-reflectiveness, generative AI currently lacks any evidence for fostering learners' forethought (Peng et al., 2026a), a qualification suggesting that scaffolding may be necessary but not sufficient for genuine parametric internalization. Structured workflows that ground AI co-writing in the learner's own ideas can partially address this gap (Ding et al., 2025). Güner & Er (2025) find convergent evidence in a quasi-experiment (N=158): five distinct AI interaction profiles emerge, ranging from passive code generation to independent coding, and instructional interventions (prompting training, guided sample prompts) significantly shift students toward collaborative profiles, with more independent profiles associated with higher post-test scores.

Deferred access. Requiring users to attempt a task before receiving AI assistance tends to produce better outcomes than immediate access. Singh et al. (2026) demonstrate in a three-condition RCT (N=97) that a write-then-revise condition, where students draft hints before seeing AI-generated alternatives, yields the highest-quality hints and broader mistake coverage compared to both independent writing and on-demand AI access, though no significant differences in learning outcomes emerged. This principle aligns with the broader finding that the "vicious cycle" of submit-incorrect-code → request-AI-fix → resubmit is the most common behavioral pattern among chatbot-using students (Rahe & Maalej, 2025), and with radiology trainees' preference for on-demand over always-visible AI support (Savardi et al., 2025).

Epistemic friction. Deliberately deploying weaker or less fluent models can promote critical evaluation. Islam et al. (2026) find that a reflection-first protocol (requiring students to draft their own analysis before consulting a local 7B-parameter model) produces the strongest downward shift in AI credibility perceptions across conditions. The authors attribute this to *epistemic friction*: the weaker model's visible limitations keep verification salient, while the reflection-first workflow anchors activity in human reasoning before AI consultation—suggesting that *less* capable AI may sometimes be pedagogically superior. A related proposal has AI deliberately feign confusion to force learners to diagnose errors rather than consume answers (Tomisu et al., 2025). A complementary approach modulates support based on user trust state: Srinivasan & Thomason (2026) demonstrate across five sequential experiments (N>500) that trust-adaptive explanations (supportive framing when trust is low, counter-explanations when trust is high) substantially reduce inappropriate reliance.

These three principles share a common logic: each preserves the cognitive engagement that direct-answer AI suppresses (Singh et al., 2025), creating conditions under which cognitive offloading can coexist with parametric internalization rather than displacing it. Yet they face a fundamental tension: meaningful oversight of AI outputs may require performing the very task that was delegated. Mezzadri (2025) articulates the paradox of ethical AI-assisted research—quality control of AI-generated research artifacts (literature reviews and data summaries) demands the domain competence that the outsourcing to AI was meant to

replace, implying that efficiency gains and epistemic integrity pull in opposite directions. Tang et al. (2024) provide related evidence: in an eye-tracking lab study with 28 university students, those informed that code was LLM-generated fixed more bugs but reported higher effort and showed marginally longer fixation durations, suggesting that provenance awareness improves validation performance at the cost of increased cognitive workload.

The tension is not merely theoretical. Institutional frameworks lag behind practice: the EU AI Act names automation bias as the only cognitive bias warranting regulatory attention but provides no operational definition of effective oversight (Laux, 2023; Laux & Ruschemeier, 2025), while judges in Colombia, Peru, Mexico, and India already use ChatGPT in judicial reasoning without established safeguards (Socol de la Osa & Remolina, 2024). RAG-based legal AI tools hallucinate at rates of 17–33% despite vendor reliability claims (Magesh et al., 2024), yet the users most likely to adopt these tools, time-pressured professionals, are often those least able to verify outputs. Training partially resolves this tension: Chen & Bao (2026) show in an RCT (N=164) that a brief training intervention (a nine-and-a-half-minute video with quiz) shifts legal AI use from unproductive (untrained users showed no improvement over non-users) to beneficial (+0.27 grade points), suggesting that the binding constraint is often not technology design but user preparation. Clerc et al. (2026) find a parallel result in a different population and domain (middle-school science, N=116): a two-hour AI literacy workshop improves LLM-assisted task performance, yet self-reported AI competence does not predict actual performance (r=0.01).

The parallel to the agent side is direct: just as the vast majority of agentic evolution operates under autonomous selective pressure (§5), the majority of human cognitive adaptation to AI proceeds without institutional design or structured intervention. The evidence indicates that the performance–capability paradox is a selective-pressure problem: it tends to arise when human-side adaptation is self-driven and unmanaged, and tends to be attenuated when externally designed interventions preserve cognitive engagement.

These design principles are not permanent solutions. The human–AI relationship is non-stationary: optimal AI assistance changes as human experience accumulates (Jain et al., 2025), AI system behavior shifts over deployment (Kirgis et al., 2026), and users renegotiate their reliance, sometimes toward independence (Wang et al., 2026e), sometimes toward deeper integration, whether through genuine learning or automation bias (Goldsmith-Pinkham et al., 2026). Any fixed design eventually misaligns with both partners' trajectories; Section 7 examines what this non-stationarity implies for the coupled system.

Key Takeaways: Human Adaptation

- **The performance–capability paradox is the central empirical pattern**: AI reliably improves measurable output (cognitive offloading, human-side Δ) while frequently reducing the independent capability needed to produce that output (parametric internalization, human-side ∇; see §2.3 for agent-side definitions). This decoupling is documented across education, professional work, decision-making, and creative domains.
- **Two factors determine the trajectory**: task structure (the paradox is bounded by expertise and the match between AI capability and task demands; §6.1) and AI design, where scaffolded interaction preserves learning while direct-answer access reduces it (§6.4). Metacognitive miscalibration may render these dynamics self-reinforcing.
- **The paradox is a design failure, not an inevitability**: Scaffolded AI, deferred access, and epistemic friction each suggest that parametric internalization is achievable; the evidence is strongest when AI scaffolds reasoning rather than supplying answers, and a particularly promising approach augments human intermediaries rather than end-users directly. However, the human–AI relationship is non-stationary, and any fixed design eventually falls out of alignment as both partners change.

Section 7 places these human-side findings alongside the agent-side analysis from Sections 3–5 and examines what emerges from the comparison.

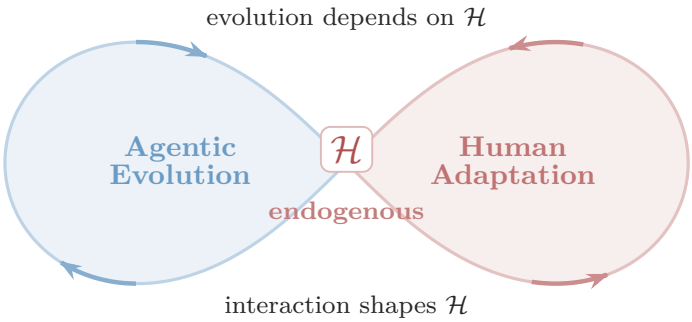

Figure 4: **The co-evolving human–AI system.** Beyond the verification-signal constraint, agentic evolution depends on human selective pressure $\mathcal{H}$ (top); sustained interaction can shape the human capacity to provide it (bottom)—under default conditions, by reducing independent evaluative capacity. Under this coupling, $\mathcal{H}$ is endogenous: shaped by the evolutionary process it is meant to guide.

7 Discussion

Sections 3–5 found that agent self-evaluation is unreliable without an external verifier: autonomous selective pressure succeeds where deterministic verifiers exist but degrades where the evaluator is coupled to the system being evolved. Section 6 found that human self-assessment is unreliable without external scaffolding: unscaffolded AI use produces the performance–capability paradox, and metacognitive miscalibration prevents users from detecting the deficit. The two literatures developed these findings independently—through different mechanisms and in different domains—but they point in the same direction: self-assessment tends to be unreliable when the assessor lacks independence from the thing being assessed. This section examines what follows when the two sets of findings are placed side by side (Figure 4).

7.1 Mutual Dependency

These parallel findings create a dependency between the two sides of the coupled system.

Agentic evolution beyond the verifier boundary requires human evaluative input. Where Tier 1 verification is unavailable (e.g., open-ended dialogue, value-laden decisions, and cultural sensitivity) agentic evolution either does not occur or relies on signals whose fidelity tends to degrade with iteration. Beyond these formal boundaries, among known evaluators that remain reliable as the system evolves, the most broadly applicable is a human who can exercise value-sensitive judgment, that is, human-involved selective pressure ($\mathcal{H} \neq 0$) at sufficient bandwidth.

Human evaluative capacity requires deliberate scaffolding. Yet the human who should provide that evaluative input is not a static resource. Section 6 documented that unscaffolded AI use (studied primarily in educational and professional settings, not yet in evolutionary-evaluation contexts) reduces the independent capability required for evaluation: the paradox of ethical AI-assisted research (Mezzadri, 2025) implies that quality control of AI outputs demands the domain competence that AI use displaces, and metacognitive miscalibration ensures these effects proceed without the user's awareness. The evidence presented above suggests that maintaining human evaluative capacity depends on the externally designed scaffolding that Section 6 identified as effective but rare. The transfer from educational and professional contexts to evolutionary evaluation is supported because the cognitive operations documented in those studies—quality judgment, error detection, and trust calibration—are the same operations that human-involved selective pressure requires.

The dependency is mutual. Agentic evolution beyond the verifier boundary depends on human evaluative capacity (established above); human evaluative capacity depends on scaffolded design (established above). It follows that effective scaffolding requires understanding how humans adapt to AI, and that understanding in turn informs agent design—closing a loop that no paper in the surveyed literature addresses

as an integrated whole. Prior work has proposed bidirectional human–AI alignment and demonstrated co-adaptive protocols in simulation (Li & Song, 2025). Kumar et al. (2025) review 77 co-learning studies and find that sample sizes are generally small, with nearly 20% of studies involving fewer than ten participants and roughly 30% reporting no participant data. Shen et al. (2025) document across 411 papers that the field overwhelmingly studies how to align AI with human preferences while neglecting how AI reshapes human cognition and behavior.

7.2 Reinterpreting Axis III

In the framework of Section 2.4, $\mathcal{H} \neq 0$ treats human input as entering the evolutionary loop from outside the system—implicitly, as an exogenous signal. The dependency analysis reveals that $\mathcal{H}$ is better understood as *endogenous*: the human who provides evaluative feedback at iteration t has been shaped by the system's outputs at iterations 1 through $t-1$, and Section 6 showed that this shaping is not negligible.

When the analysis follows Axis III to its boundary (the verifier boundary in Section 5), it encounters a question it cannot answer from agent-side evidence alone: is the human signal reliable? Answering that question requires the reverse analysis conducted above, and the answer (that default interaction can reduce the evaluative capacity on which human selective pressure depends) transforms the framework's understanding of its own third axis.

Among the $\mathcal{H} \neq 0$ systems in the survey, four track changes in user facts, preferences, or domain knowledge (Chhikara et al., 2025; He et al., 2025b; Nie et al., 2026b; Rasmussen et al., 2025), but none tracks whether the user's evaluative capacity has been preserved, the level at which the vulnerability operates. This reinterpretation does not invalidate the $\mathcal{H} = 0$ / $\mathcal{H} \neq 0$ classification (the distinction remains operationally meaningful) but it changes what the classification implies: $\mathcal{H} \neq 0$ does not guarantee effective evaluative input, because the quality of the human signal is itself a function of the evolutionary trajectory it is meant to guide.

The same coupling perspective extends to the other two axes. On Axis II, consolidation pathway choice is not only an agent-performance trade-off but a human-verifiability trade-off: structural consolidation produces artifacts that are human-readable and auditable (prompts, code, and workflow graphs; Section 4), while parametric consolidation produces weight changes that are not. If human evaluative capacity is degrading under default interaction (Section 6), then parametric-dominant evolution compounds an analogous verification challenge (Mezzadri, 2025) by removing a primary channel through which a human evaluator could inspect what changed, while structural-dominant evolution at least preserves inspectability. Few if any systems in the surveyed literature explicitly select their consolidation pathway based on whether the human partner can verify the result. The hybrid $\Delta{+}\nabla$ systems documented in Section 4 deserve investigation on this dimension: they maintain a human-readable structural component alongside a parametric one, and whether that structural component can serve as an interface for human oversight, beyond its role as an agent-internal knowledge representation, is an open design question.

On Axis I, different evolutionary substrates may affect different dimensions of human cognition. Section 3 classified agentic evolution by substrate (Cortex, Action, and Memory); Section 6 classified human adaptation along corresponding dimensions (cognitive, operational, and epistemic). Indirect evidence suggests the distinction matters: studies of reasoning-assistive AI document declines in deliberative capacity, while studies of information-retrieval AI document shifts in epistemic habits such as belief offloading (Guingrich et al., 2026) and knowledge-community displacement (Burtch et al., 2024). Whether these differential effects can be traced to the evolutionary substrate of the AI system—and which combinations of agentic evolution and human scaffolding are most critical to co-design—remains open.

7.3 Deliberate Design Outperforms the Default

The dependency analysis above might suggest a pessimistic conclusion, that the coupled system is structurally destined for degradation. The evidence does not support this interpretation. On both sides, deliberate design outperforms the unmanaged default.

On the agent side, hybrid $\Delta+\nabla$ systems tend to outperform single-pathway alternatives (Section 4), and the few $\mathcal{H} \neq 0$ systems that provide human selective pressure at sufficient bandwidth address domains where deterministic verifiers are unavailable (Section 5). On the human side, the scaffolded designs documented in Section 6.4 produce measurable capability gains—from pedagogically designed tutoring (Kestin et al., 2025) to partial skill acquisition that persists through system outages (Brynjolfsson et al., 2025).

The contrast is consistent across both sides: where design provides an independent check (an external verifier for agents and a scaffolded interaction for humans) the outcome is more likely to be durable capability gain. Where design ignores the constraint, the default trajectory produces degradation of the evaluation signal on the agent side and the performance–capability paradox on the human side.

The mutual dependency established above means the coupled system is not two independent optimization problems that happen to share a boundary. Agentic evolution shapes the environment in which human evaluative capacity develops; human evaluative capacity shapes the selective pressure on which agentic evolution beyond the verifier boundary depends. The two trajectories are coupled, and steering them requires treating agentic evolution and human adaptation as components of a single co-evolving system.

7.4 From Self-Improving Agents to Co-Evolving Systems

The mutual dependency identifies what is missing from current practice. The five forms of human-involved input defined in Section 2.4 provide a vocabulary for what the transition requires.

The bandwidth gap. As Section 5 documented, current $\mathcal{H} \neq 0$ architectures overwhelmingly rely on low-bandwidth channels (evaluative feedback and implicit signal). The higher-bandwidth forms that would allow sustained human steering of agentic evolution in value-laden domains remain nearly unexplored. This matters because the domains where agent self-evaluation fails (e.g., open-ended dialogue, value-laden decisions, and cultural sensitivity) are often those where richer human input would be most relevant. Yet the reinterpretation above showed that the quality of that input cannot be taken for granted.

Bridging two timescales. Agentic evolution accumulates over timescales from minutes to months; human adaptation unfolds over weeks to years (Goldsmith-Pinkham et al., 2026). No mechanism in either surveyed literature tracks both trajectories jointly. Agent reward models encode a snapshot of human preferences that grows stale as the human adapts (the frozen-preference problem identified above). Meanwhile, optimal AI assistance shifts as human skill develops (Jain et al., 2025), so a scaffolding design that works at deployment may not work months later. Current agent benchmarks do not test whether an agent's evolutionary trajectory degrades the human evaluator; current human–AI studies do not test whether the AI changed between sessions. Neither captures the coupled trajectory that the mutual dependency implies.

What joint evaluation would require. Evaluating a co-evolving system cannot be reduced to evaluating each side independently. On the agent side, it requires tracking not only the evolved agent's performance but whether the evaluation signal itself remains calibrated across iterations—a form of meta-evaluation that no surveyed paper proposes. On the human side, it requires longitudinal measurement: nearly all studies span a single session to one semester, yet the non-stationarity evidence implies that the most consequential effects emerge only over extended deployment. The missing junior loop (Catalini et al., 2026) operates on a still longer timescale—the contraction of apprenticeship pipelines that produce future domain experts is invisible in any single-semester study.

Current systems optimize one side while treating the other as fixed. The evidence presented here shows that neither side is fixed, and that their trajectories are coupled. The transition from self-improving agents to co-evolving human–AI systems requires designing systems that monitor and maintain both partners' capacities jointly, over the timescales at which each actually changes.

Key Takeaways: Discussion

- **Parallel findings**: Agent self-evaluation is unreliable without an external verifier; human self-assessment is unreliable without external scaffolding. Two independent literatures converge on this from opposite directions.
- **Dependency**: Agentic evolution beyond the verifier boundary depends on human evaluative input; human evaluative capacity depends on scaffolded design; neither side can be addressed in isolation.
- **Reinterpretation**: $\mathcal{H}$ is endogenous: the human who provides selective pressure is reshaped by the evolutionary process those signals are meant to guide. The same coupling perspective extends to the other axes, revealing that pathway choice also determines human verifiability and that different substrates may differentially affect human cognition.
- **Co-evolution**: Current systems optimize one side while treating the other as fixed. The transition from self-improving agents to co-evolving systems requires closing the human-input bandwidth gap, bridging the timescale mismatch between agent iterations and human adaptation, and building joint evaluation infrastructure that tracks both partners over deployment lifetimes.

8 Conclusion

The agent-side analysis (Sections 3–5) yields cross-axis regularities invisible within any single axis: consolidation failure modes track the pathway rather than the substrate, and the pathway choice is constrained by substrate properties rather than freely chosen. A further regularity emerges on Axis III: autonomous evolution produces its strongest results where deterministic verifiers are available, but the evaluation signal's fidelity tends to degrade where the verifier is coupled to the system being evaluated. The human-side analysis (Section 6) finds that default, unscaffolded AI interaction can reduce the independent evaluative capacity on which human-involved selective pressure depends. These two sets of findings are coupled: agentic evolution beyond the verifier boundary depends on human evaluative capacity, and that capacity is not stationary under default interaction conditions. This coupling motivated reinterpreting the framework's third axis: human selective pressure, originally modeled as exogenous input, is better understood as endogenous—shaped by the evolutionary process it is meant to guide (Section 7.2). The evidence of Section 6.2 further indicates that this reshaping often proceeds without the user's awareness, implying that co-evolving system design cannot rely on self-reported evaluative capacity. The implication is concrete: agentic evolution and human adaptation must be treated as components of one jointly designed system—one designed to close the bandwidth gap in human selective pressure, bridge the timescale mismatch between agentic evolution and human adaptation, and build joint evaluation that tracks both partners over deployment lifetimes (Section 7.4).

Limitations. These conclusions are bounded by several limitations. The three-axis framework assumes static substrate boundaries, yet emerging self-modifying architectures blur the distinctions on which the taxonomy depends. The coupled-system analysis assumes a dyadic setting—one human, one agent—while deployed systems increasingly involve multi-agent architectures and multiple human collaborators whose network dynamics this framing does not capture. The classification was conducted by the authors using a structured rubric without formal external inter-rater reliability assessment. More than four in five of the surveyed agentic-evolution papers were published in 2025–2026, so the taxonomy predominantly reflects a two-year window of a rapidly evolving field. The verification-signal constraint identified in Section 5 is an empirical regularity across the surveyed papers, not a formal impossibility result. Approaches such as principle-grounded AI feedback (Bai et al., 2022), adversarial debate with human judges (Irving et al., 2018), and AI-generated preference labels (Lee et al., 2024a) aim to maintain evaluation quality beyond the reach of deterministic verifiers, and whether such mechanisms sustain signal fidelity under the iterative evolutionary dynamics that characterize the papers surveyed here remains open. The human-side literature synthesized in Section 6 is methodologically heterogeneous: while several key findings rest on pre-registered RCTs, others derive from observational, quasi-experimental, or small-sample designs, and most establish task-specific effects over short time horizons rather than demonstrating long-term causal trajectories of human cognitive change under sustained AI interaction. The literature search was conducted primarily in English-language

venues, and human-side studies draw disproportionately from WEIRD (Western, Educated, Industrialized, Rich, and Democratic) populations; both sets of findings may differ in underrepresented contexts.

References

Rania Abdelghani, Kou Murayama, Celeste Kidd, Hélène Sauzéon, and Pierre-Yves Oudeyer. The illusion of understanding: How middle-schoolers fail to regulate inquiry with ChatGPT in a science task. *arXiv preprint arXiv:2505.01106*, 2025.

Josh Abramson, Arun Ahuja, Federico Carnevale, Petko Georgiev, Alex Goldin, Alden Hung, Jessica Landon, Jirka Lhotka, Timothy Lillicrap, Alistair Muldal, George Powell, Adam Santoro, Guy Scully, Sanjana Srivastava, Tamara von Glehn, Greg Wayne, Nathaniel Wong, Chen Yan, and Rui Zhu. Improving multimodal interactive agents with reinforcement learning from human feedback. *arXiv preprint arXiv:2211.11602*, 2022.

Sami Abuzakuk, Anne-Marie Kermarrec, Rishi Sharma, Rasmus Moorits Veski, and Martijn de Vos. Optimizing agentic workflows using meta-tools. *arXiv preprint arXiv:2601.22037*, 2026.

Bassam Adnan, Sathvika Miryala, Aneesh Sambu, Karthik Vaidhyanathan, Martina De Sanctis, and Romina Spalazzese. Leveraging LLMs for dynamic IoT systems generation through mixed-initiative interaction. In *ICSA-C*, 2025.

Nikhil Agarwal, Alex Moehring, Pranav Rajpurkar, and Tobias Salz. Combining human expertise with artificial intelligence: Experimental evidence from radiology. Technical report, National Bureau of Economic Research, 2023.

Mohammed Himayath Ali, Mohammed Aqib Abdullah, Mohammed Mudassir Uddin, and Shahnawaz Alam. SecureCAI: Injection-resilient LLM assistants for cybersecurity operations. *arXiv preprint arXiv:2601.07835*, 2026.

Joshua Ashkinaze, Julia Mendelsohn, Qiwei Li, Ceren Budak, and Eric Gilbert. How AI ideas affect the creativity, diversity, and evolution of human ideas: Evidence from a large, dynamic experiment. *arXiv preprint arXiv:2401.13481*, 2024.

Dhruv Atreja. ALAS: Autonomous learning agent for self-updating language models. *arXiv preprint arXiv:2508.15805*, 2025.

Dhruv Atreja, Julia White, Nikhil Nayak, Kelton Zhang, Henrijs Princis, George Hurn-Maloney, Ash Lewis, and Urchade Zaratiana. Pioneer agent: Continual improvement of small language models in production. *arXiv preprint arXiv:2604.09791*, 2026.

Ye Bai, Minghan Wang, and Thuy-Trang Vu. MAPLE: Multi-agent adaptive planning with long-term memory for table reasoning. In *ALTA*, 2025.

Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, Carol Chen, Catherine Olsson, Christopher Olah, Danny Hernandez, Dawn Drain, Deep Ganguli, Dustin Li, Eli Tran-Johnson, Ethan Perez, Jamie Kerr, Jared Mueller, Jeffrey Ladish, Joshua Landau, Kamal Ndousse, Kamile Lukosuite, Liane Lovitt, Michael Sellitto, Nelson Elhage, Nicholas Schiefer, Noemi Mercado, Nova DasSarma, Robert Lasenby, Robin Larson, Sam Ringer, Scott Johnston, Shauna Kravec, Sheer El Showk, Stanislav Fort, Tamera Lanham, Timothy Telleen-Lawton, Tom Conerly, Tom Henighan, Tristan Hume, Samuel R. Bowman, Zac Hatfield-Dodds, Ben Mann, Dario Amodei, Nicholas Joseph, Sam McCandlish, Tom Brown, and Jared Kaplan. Constitutional AI: Harmlessness from AI feedback. *arXiv preprint arXiv:2212.08073*, 2022.

Debangshu Banerjee, Changming Xu, Eugene Ie, Ming Zhang, Daiyi Peng, Chu-Cheng Lin, and Gagandeep Singh. SEVerA: Verified self-evolving agents. *arXiv preprint arXiv:2603.25111*, 2026.

Zhixuan Bao, Zhuoyi Lin, Jiageng Wang, Jinhai Hu, Yuan Gao, Yaoxin Wu, Xiaoli Li, and Xun Xu. AnalogAgent: Self-improving analog circuit design automation with LLM agents. *arXiv preprint arXiv:2603.23910*, 2026.

Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakçı, and Rei Mariman. Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122, 2025.

Rafael M. Batista and Thomas L. Griffiths. A rational analysis of the effects of sycophantic AI. *arXiv preprint arXiv:2602.14270*, 2026.

Elisabeth Bauer, Samuel Greiff, Arthur C. Graesser, Katharina Scheiter, and Michael Sailer. Looking beyond the hype: Understanding the effects of AI on learning. *Educational Psychology Review*, 37(2), 2025.

Joel Becker, Nate Rush, Beth Barnes, and David Rein. Measuring the impact of early-2025 AI on experienced open-source developer productivity. *arXiv preprint arXiv:2507.09089*, 2025.

Mohammad Beigi, Ming Jin, Junshan Zhang, Qifan Wang, and Lifu Huang. Adversarial reward auditing for active detection and mitigation of reward hacking. *arXiv preprint arXiv:2602.01750*, 2026.

Alexander Bering. ZenBrain: A neuroscience-inspired 7-layer memory architecture for autonomous AI systems. *arXiv preprint arXiv:2604.23878*, 2026.

Adithya Bhaskar, Xi Ye, and Danqi Chen. Language models that think, chat better. *arXiv preprint arXiv:2509.20357*, 2025.

Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Robert Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, Szymon Jakubczak, Tim Jones, Liyiming Ke, Sergey Levine, Adrian Li-Bell, Mohith Mothukuri, Suraj Nair, Karl Pertsch, Lucy Xiaoyang Shi, Laura Smith, James Tanner, Quan Vuong, Anna Walling, Haohuan Wang, and Ury Zhilinsky. π_0: A vision-language-action flow model for general robot control. In *RSS*, 2025.

Jessica Y Bo, Majeed Kazemitabaar, Mengqing Deng, Michael Inzlicht, and Ashton Anderson. Invisible saboteurs: Sycophantic LLMs mislead novices in problem-solving tasks. In *CHI*, 2026.

Paul Brookes, Vardan Voskanyan, Rafail Giavrimis, Matthew Truscott, Mina Ilieva, Chrystalla Pavlou, Alexandru Staicu, Manal Adham, Will Evers-Hood, Jingzhi Gong, Kejia Zhang, Matvey Fedoseev, Vishal Sharma, Roman Bauer, Zheng Wang, Hema Nair, Wei Jie, Tianhua Xu, Aurora Constantin, Carmine Ventre, Leslie Kanthan, and Michail Basios. Evolving excellence: Automated optimization of LLM-based agents. *arXiv preprint arXiv:2512.09108*, 2025.

Erik Brynjolfsson, Danielle Li, and Lindsey R Raymond. Generative AI at work. *The Quarterly Journal of Economics*, 140, 2025.

Gordon Burtch, Dokyun Lee, and Zhichen Chen. The consequences of generative AI for online knowledge communities. *Scientific Reports*, 14(1), 2024.

Yuxuan Cai, Yipeng Hao, Jie Zhou, Hang Yan, Zhikai Lei, Rui Zheng, Zhenhua Han, Yutao Yang, Junsong Li, Qianjun Pan, Tianyu Huai, Qin Chen, Xin Li, Kai Chen, Bo Zhang, Xipeng Qiu, and Liang He. Building self-evolving agents via experience-driven lifelong learning: A framework and benchmark. *arXiv preprint arXiv:2508.19005*, 2025a.

Yuxuan Cai, Jie Zhou, Qin Chen, and Liang He. Ask only when needed: Proactive retrieval from memory and skills for experience-driven lifelong agents. *arXiv preprint arXiv:2604.20572*, 2026.

Zhicheng Cai, Xinyuan Guo, Yu Pei, Jiangtao Feng, Jinsong Su, Jiangjie Chen, Ya-Qin Zhang, Wei-Ying Ma, Mingxuan Wang, and Hao Zhou. Flex: Continuous agent evolution via forward learning from experience. *arXiv preprint arXiv:2511.06449*, 2025b.

Arthur Câmara, Vincent Slot, and Jakub Zavrel. Self-optimizing multi-agent systems for deep research. *arXiv preprint arXiv:2604.02988*, 2026.

Zouying Cao, Jiaji Deng, Li Yu, Weikang Zhou, Zhaoyang Liu, Bolin Ding, and Hai Zhao. Remember me, refine me: A dynamic procedural memory framework for experience-driven agent evolution. *arXiv preprint arXiv:2512.10696*, 2025.

Christian Catalini, Xiang Hui, and Jane Wu. Some simple economics of AGI. *arXiv preprint arXiv:2602.20946*, 2026.

Hyungjoo Chae, Jungsoo Park, and Alan Ritter. Safe and scalable web agent learning via recreated websites. *arXiv preprint arXiv:2603.10505*, 2026.

Benjamin M. Chen and Hong Bao. Training for technology: Adoption and productive use of generative AI in legal analysis. *arXiv preprint arXiv:2603.04982*, 2026.

Guoxin Chen, Zile Qiao, Wenqing Wang, Donglei Yu, Xuanzhong Chen, Hao Sun, Minpeng Liao, Kai Fan, Yong Jiang, Pengjun Xie, Wayne Xin Zhao, Ruihua Song, and Fei Huang. Mars: Co-evolving dual-system deep research via multi-agent reinforcement learning. *arXiv preprint arXiv:2510.04935*, 2025a.

Kai Chen, Xinfeng Li, Tianpei Yang, Hewei Wang, Wei Dong, and Yang Gao. MDTeamGPT: A self-evolving LLM-based multi-agent framework for multi-disciplinary team medical consultation. *arXiv preprint arXiv:2503.13856*, 2025b.

Kevin Chen, Marco Cusumano-Towner, Brody Huval, Aleksei Petrenko, Jackson Hamburger, Vladlen Koltun, and Philipp Krähenbühl. Reinforcement learning for long-horizon interactive LLM agents. *arXiv preprint arXiv:2502.01600*, 2025c.

Rufeng Chen, Shuaishuai Jiang, Jiyun Shen, AJung Moon, and Lili Wei. Examining the usage of generative AI models in student learning activities for software programming. *arXiv preprint arXiv:2511.13271*, 2025d.

Terry Chen, Zhifan Ye, Bing Xu, Zihao Ye, Timmy Liu, Ali Hassani, Tianqi Chen, Andrew Kerr, Haicheng Wu, Yang Xu, Yu-Jung Chen, Hanfeng Chen, Aditya Kane, Ronny Krashinsky, Ming-Yu Liu, Vinod Grover, Luis Ceze, Roger Bringmann, John Tran, Wei Liu, Fung Xie, Michael Lightstone, and Humphrey Shi. AVO: Agentic variation operators for autonomous evolutionary search. *arXiv preprint arXiv:2603.24517*, 2026a.

Xiang Chen, Yuling Shi, Qizhen Lan, Yuchao Qiu, Min Wang, Xiaodong Gu, and Yanfu Yan. Fed-SE: Federated self-evolution for cross-environment knowledge transfer in privacy-constrained LLM agents. *arXiv preprint arXiv:2512.08870*, 2025e.

Xuanzhong Chen, Zile Qiao, Guoxin Chen, Liangcai Su, Zhen Zhang, Xinyu Wang, Pengjun Xie, Fei Huang, Jingren Zhou, and Yong Jiang. AgentFrontier: Expanding the capability frontier of LLM agents with ZPD-guided data synthesis. In *ICLR*, 2026b.

Yixing Chen, Yiding Wang, Siqi Zhu, Haofei Yu, Tao Feng, Muhan Zhang, Mostofa Patwary, and Jiaxuan You. Multi-agent evolve: LLM self-improve through co-evolution. *arXiv preprint arXiv:2510.23595*, 2025f.

Zhangtianyi Chen, Yuhao Shen, Florensia Widjaja, Yan Xu, Liyuan Sun, Zijian Wang, Hongyi Chen, Wufei Dai, and Juexiao Zhou. SkinGPT-X: A self-evolving collaborative multi-agent system for transparent and trustworthy dermatological diagnosis. *arXiv preprint arXiv:2603.26122*, 2026c.

Zhaorun Chen, Zhuokai Zhao, Kai Zhang, Bo Liu, Qi Qi, Yifan Wu, Tarun Kalluri, Sara Cao, Yuanhao Xiong, Haibo Tong, Huaxiu Yao, Hengduo Li, Jiacheng Zhu, Xian Li, Dawn Song, Bo Li, Jason Weston, and Dat Huynh. Scaling agent learning via experience synthesis. *arXiv preprint arXiv:2511.03773*, 2025g.

Daixuan Cheng, Shaohan Huang, Yuxian Gu, Huatong Song, Guoxin Chen, Li Dong, Wayne Xin Zhao, Ji-Rong Wen, and Furu Wei. Computer environments elicit general agentic intelligence in LLMs. *arXiv preprint arXiv:2601.16206*, 2026a.

Myra Cheng, Sunny Yu, Cinoo Lee, Pranav Khadpe, Lujain Ibrahim, and Dan Jurafsky. ELEPHANT: Measuring and understanding social sycophancy in LLMs. *arXiv preprint arXiv:2505.13995*, 2025.

Xusen Cheng and Lulu Zhang. Inspiration booster or creative fixation? the dual mechanisms of LLMs in shaping individual creativity in tasks of different complexity. *Humanities and Social Sciences Communications*, 12(1), 2025.

Zihao Cheng, Zeming Liu, Yingyu Shan, Xinyi Wang, Xiangrong Zhu, Yunpu Ma, Hongru Wang, Yuhang Guo, Wei Lin, and Yunhong Wang. Mem^{2}evolve: Towards self-evolving agents via co-evolutionary capability expansion and experience distillation. *arXiv preprint arXiv:2604.10923*, 2026b.

Prateek Chhikara, Dev Khant, Saket Aryan, Taranjeet Singh, and Deshraj Yadav. Mem0: Building production-ready AI agents with scalable long-term memory. *arXiv preprint arXiv:2504.19413*, 2025.

Przemyslaw Chojecki. Self-improving AI agents through self-play. *arXiv preprint arXiv:2512.02731*, 2025.

O. Clerc, Rania Abdelghani, C. Desvaux, E. Poisson, P. Oudeyer, and H. Sauzéon. Teaching students to question the machine: An AI literacy intervention improves students' regulation of LLM use in a science task. *arXiv preprint arXiv:2604.01955*, 2026.

Kevin Crowston and Francesco Bolici. Deskilling and upskilling with AI systems. *Information Research an international electronic journal*, 30(iConf):1009–1023, 2025.

Andrew Dai, Boris Meinardus, Ciaran Regan, Yingtao Tian, and Yujin Tang. Discovering novel LLM experts via task-capability coevolution. In *ICLR*, 2026.

Dadan Dasari, Agus Hendriyanto, Sani Sahara, Didi Suryadi, Lukman Hakim Muhaimin, Theodore Chao, and Laila Fitriana. ChatGPT in didactical tetrahedron, does it make an exception? a case study in mathematics teaching and learning. *Frontiers in Education*, 8, 2024.

Maria-Teresa De Rosa Palmini and Eva Cetinic. Patterns of creativity: How user input shapes AI-generated visual diversity. *arXiv preprint arXiv:2410.06768*, 2024.

Peer-Benedikt Degen and Igor Asanov. Beyond automation: Socratic AI, epistemic agency, and the implications of the emergence of orchestrated multi-agent learning architectures. *arXiv preprint arXiv:2508.05116*, 2025.

Fabrizio Dell'Acqua, Edward McFowland III, Ethan Mollick, Hila Lifshitz-Assaf, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani. Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organization Science*, 37, 2026.

Yi-Xuan Deng, Xiaoqin Liu, Yi Zhang, Guo-Wei Yang, and Shuojin Yang. Your code agent can grow alongside you with structured memory. *arXiv preprint arXiv:2603.13258*, 2026.

Xiaohan Ding, Kaike Ping, Uma Sushmitha Gunturi, Buse Carik, Sophia Stil, Lance T Wilhelm, Taufiq Daryanto, James Hawdon, Sang Won Lee, and Eugenia H Rho. Designing human-AI collaboration to support learning in counterspeech writing. In *VL/HCC*, 2025.

Guanting Dong, Yifei Chen, Xiaoxi Li, Jiajie Jin, Hongjin Qian, Yutao Zhu, Hangyu Mao, Guorui Zhou, Zhicheng Dou, and Ji-Rong Wen. Tool-star: Empowering LLM-brained multi-tool reasoner via reinforcement learning. *arXiv preprint arXiv:2505.16410*, 2025.

Guanting Dong, Junting Lu, Junjie Huang, Wanjun Zhong, Longxiang Liu, Shijue Huang, Zhenyu Li, Yang Zhao, Xiaoshuai Song, Xiaoxi Li, Jiajie Jin, Yutao Zhu, Hanbin Wang, Fangyu Lei, Qinyu Luo, Mingyang Chen, Zehui Chen, Jiazhan Feng, Ji-Rong Wen, and Zhicheng Dou. Agent-world: Scaling real-world environment synthesis for evolving general agent intelligence. *arXiv preprint arXiv:2604.18292*, 2026a.

Kris Shengjun Dong, Sahil Modi, Dima Nikiforov, Sana Damani, Edward Lin, Siva Kumar Sastry Hari, and Christos Kozyrakis. Kernelblaster: Continual cross-task CUDA optimization via memory-augmented in-context reinforcement learning. *arXiv preprint arXiv:2602.14293*, 2026b.

Shihan Dou, Yan Liu, Haoxiang Jia, Enyu Zhou, Limao Xiong, Junjie Shan, Caishuang Huang, Xiao Wang, Xiaoran Fan, Zhiheng Xi, Yuhao Zhou, Tao Ji, Rui Zheng, Qi Zhang, Tao Gui, and Xuanjing Huang. StepCoder: Improving code generation with reinforcement learning from compiler feedback. In *ACL*, 2024.

Hermann Ebbinghaus. *Über das gedächtnis: untersuchungen zur experimentellen psychologie*. Duncker & Humblot, 1885.

Jinyuan Fang, Yanwen Peng, Xi Zhang, Yingxu Wang, Xinhao Yi, Guibin Zhang, Yi Xu, Bin Wu, Siwei Liu, Zihao Li, Zhaochun Ren, Nikos Aletras, Xi Wang, Han Zhou, and Zaiqiao Meng. A comprehensive survey of self-evolving AI agents: A new paradigm bridging foundation models and lifelong agentic systems. *arXiv preprint arXiv:2508.07407*, 2025a.

Runnan Fang, Yuan Liang, Xiaobin Wang, Jialong Wu, Shuofei Qiao, Pengjun Xie, Fei Huang, Huajun Chen, and Ningyu Zhang. Memp: Exploring agent procedural memory. *arXiv preprint arXiv:2508.06433*, 2025b.

Tianqing Fang, Hongming Zhang, Zhisong Zhang, Kaixin Ma, Wenhao Yu, Haitao Mi, and Dong Yu. Webevolver: Enhancing web agent self-improvement with co-evolving world model. In *EMNLP*, 2025c.

Wenji Fang, Yao Lu, Shang Liu, Jing Wang, Ziyan Guo, Junxian He, Fengbin Tu, and Zhiyao Xie. Dr. RTL: Autonomous agentic RTL optimization through tool-grounded self-improvement. *arXiv preprint arXiv:2604.14989*, 2026.

Wenkai Fang, Shunyu Liu, Yang Zhou, Kongcheng Zhang, Tongya Zheng, Kaixuan Chen, Mingli Song, and Dacheng Tao. Serl: Self-play reinforcement learning for large language models with limited data. *arXiv preprint arXiv:2505.20347*, 2025d.

Jiazhan Feng, Shijue Huang, Xingwei Qu, Ge Zhang, Yujia Qin, Baoquan Zhong, Chengquan Jiang, Jinxin Chi, and Wanjun Zhong. Retool: Reinforcement learning for strategic tool use in LLMs. *arXiv preprint arXiv:2504.11536*, 2025a.

K. J. Kevin Feng, Kevin Pu, Matt Latzke, Tal August, Pao Siangliulue, Jonathan Bragg, Daniel S. Weld, Amy X. Zhang, and Joseph Chee Chang. Cocoa: Co-planning and co-execution with AI agents. In *CHI*, 2026a.

Tao Feng, Pengrui Han, Guanyu Lin, Ge Liu, and Jiaxuan You. Thought-retriever: Don't just retrieve raw data, retrieve thoughts for memory-augmented agentic systems. *Transactions on Machine Learning Research*, 2026b.

Tongtong Feng, Xin Wang, Zekai Zhou, Ren Wang, Yuwei Zhan, Guangyao Li, Qing Li, and Wenwu Zhu. EvolvingAgent: Curriculum self-evolving agent with continual world model for long-horizon tasks. *arXiv preprint arXiv:2502.05907*, 2025b.

Xinshun Feng, Xinhao Song, Lijun Li, Gongshen Liu, and Jing Shao. SEARL: Joint optimization of policy and tool graph memory for self-evolving agents. *arXiv preprint arXiv:2604.07791*, 2026c.

Chrisantha Fernando, Dylan Banarse, Henryk Michalewski, Simon Osindero, and Tim Rocktäschel. Promptbreeder: Self-referential self-improvement via prompt evolution. In *ICML*, 2024.

Caterina Fregosi, Lucia Vicente, Andrea Campagner, and Federico Cabitza. Too sure for our own good: A user study on AI confidence and human reliance. In *AAAI*, 2026.

Karl Friston. The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11:127–138, 2010.

Shi Fu, Yingjie Wang, Shengchao Hu, Peng Wang, and Dacheng Tao. Why self-rewarding works: Theoretical guarantees for iterative alignment of language models. *arXiv preprint arXiv:2601.22513*, 2026.

Artem Gadzhiev and Andrew Kislov. Synthius-Mem: Brain-inspired hallucination-resistant persona memory achieving 94.4% memory accuracy and 99.6% adversarial robustness on LoCoMo. *arXiv preprint arXiv:2604.11563*, 2026.

Víctor Gallego. Discovering agentic safety specifications from 1-bit danger signals. In *ALA Workshop @ AAMAS*, 2026.

Huan-ang Gao, Jiayi Geng, Wenyue Hua, Mengkang Hu, Xinzhe Juan, Hongzhang Liu, Shilong Liu, Jiahao Qiu, Xuan Qi, Qihan Ren, Yiran Wu, Hongru Wang, Han Xiao, Yuhang Zhou, Shaokun Zhang, Jiayi Zhang, Jinyu Xiang, Yixiong Fang, Qiwen Zhao, Dongrui Liu, Cheng Qian, Zhenhailong Wang, Minda Hu, Huazheng Wang, Qingyun Wu, Heng Ji, and Mengdi Wang. A survey of self-evolving agents: What, when, how, and where to evolve on the path to artificial super intelligence. *Transactions on Machine Learning Research*, 2026.

Shanghua Gao, Richard Zhu, Pengwei Sui, Zhenglun Kong, Sufian Aldogom, Yepeng Huang, Ayush Noori, Reza Shamji, Krishna Parvataneni, Theodoros Tsiligkaridis, and Marinka Zitnik. Democratizing AI scientists using tooluniverse. *arXiv preprint arXiv:2509.23426*, 2025.

Tanmay Gautam, Alireza Bahramali, and Sandeep Atluri. AutoRISE: Agent-driven strategy evolution for red-teaming large language models. *arXiv preprint arXiv:2604.22871*, 2026.

Georgios P. Georgiou. ChatGPT produces more "lazy" thinkers: Evidence of cognitive engagement decline. *arXiv preprint arXiv:2507.00181*, 2025.

Sandip Ghoshal, Anshul Mittal, Jyotika Singh, Miguel Ballesteros, Weiyi Sun, Fang Tu, Shailender Singh, Yassine Benajiba, Fahad Shah, Sujeeth Bharadwaj, Sujith Ravi, and Dan Roth. JTPRO: A joint tool-prompt reflective optimization framework for language agents. *arXiv preprint arXiv:2604.19821*, 2026.

Paul Goldsmith-Pinkham, Chenhao Tan, and Alexander K. Zentefis. Human-AI collaboration in radiology: The case of pulmonary embolism. *arXiv preprint arXiv:2601.13379*, 2026.

Jingzhi Gong, Ruizhen Gu, Zhiwei Fei, Yazhuo Cao, Lukas Twist, Alina Geiger, Shuo Han, Dominik Sobania, Federica Sarro, and Jie M. Zhang. SkillMOO: Multi-objective optimization of agent skills for software engineering. *arXiv preprint arXiv:2604.09297*, 2026.

Zhibin Gou, Zhihong Shao, Yeyun Gong, Yelong Shen, Yujiu Yang, Minlie Huang, Nan Duan, and Weizhu Chen. ToRA: A tool-integrated reasoning agent for mathematical problem solving. In *ICLR*, 2024.

Yingjie Gu, Wenjian Xiong, Liqiang Wang, Pengcheng Ren, Chao Li, Xiaojing Zhang, Yijuan Guo, Qi Sun, Jingyao Ma, and Shidang Shi. FSFM: A biologically-inspired framework for selective forgetting of agent memory. *arXiv preprint arXiv:2604.20300*, 2026.

Zhaoyang Guan, Huixi Cao, Ming Zhong, Eric Yang, Lynn Ai, Yongxin Ni, and Bill Shi. Symphony-coord: Emergent coordination in decentralized agent systems. *arXiv preprint arXiv:2602.00966*, 2026.

Zhenyu Guan, Xiangyu Kong, Fangwei Zhong, and Yizhou Wang. Richelieu: Self-evolving LLM-based agents for AI diplomacy. In *NeurIPS*, 2024.

Rose E. Guingrich, Dvija Mehta, and Umang Bhatt. Belief offloading in human-AI interaction. *arXiv preprint arXiv:2602.08754*, 2026.

Hacer Güner and Erkan Er. AI in the classroom: Exploring students' interaction with ChatGPT in programming learning. *Education and Information Technologies*, 30(9):12681–12707, 2025.

Oguzhan Gungordu, Siheng Xiong, and Faramarz Fekri. PathWise: Planning through world model for automated heuristic design via self-evolving LLMs. *arXiv preprint arXiv:2601.20539*, 2026.

Qingyan Guo, Rui Wang, Junliang Guo, Bei Li, Kaitao Song, Xu Tan, Guoqing Liu, Jiang Bian, and Yujiu Yang. Connecting large language models with evolutionary algorithms yields powerful prompt optimizers. In *ICLR*, 2024.

David Ha and Jürgen Schmidhuber. Recurrent world models facilitate policy evolution. In *NeurIPS*, 2018.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse control tasks through world models. *Nature*, 640:647–653, 2025.

Karen Hambardzumyan, Nicolas Baldwin, Edan Toledo, Rishi Hazra, Michael Kuchnik, Bassel Al Omari, Thomas Simon Foster, Anton Protopopov, Jean-Christophe Gagnon-Audet, Ishita Mediratta, Kelvin Niu, Michael Shvartsman, Alisia Lupidi, Alexis Audran-Reiss, Parth Pathak, Tatiana Shavrina, Despoina Magka, Hela Momand, Derek Dunfield, Nicola Cancedda, Pontus Stenetorp, Carole-Jean Wu, Jakob Nicolaus Foerster, Yoram Bachrach, and Martin Josifoski. AIRA_2: Overcoming bottlenecks in AI research agents. *arXiv preprint arXiv:2603.26499*, 2026.

Siwei Han, Kaiwen Xiong, Jiaqi Liu, Xinyu Ye, Yaofeng Su, Wenbo Duan, Xinyuan Liu, Cihang Xie, Mohit Bansal, Mingyu Ding, Linjun Zhang, and Huaxiu Yao. Alignment tipping process: How self-evolution pushes LLM agents off the rails. *arXiv preprint arXiv:2510.04860*, 2025.

Xiao Han, Yuzheng Fan, Sendong Zhao, Haochun Wang, and Bing Qin. GSEM: Graph-based self-evolving memory for experience augmented clinical reasoning. *arXiv preprint arXiv:2603.22096*, 2026.

Gaole He, Lucie Kuiper, and Ujwal Gadiraju. Knowing about knowing: An illusion of human competence can hinder appropriate reliance on AI systems. In *CHI*, 2023.

Shan He, Runze Wang, Zhuoyun Du, Huiyu Bai, Zouying Cao, Yu Cheng, and Bo Zheng. Learning to evolve: A self-improving framework for multi-agent systems via textual parameter graph optimization. *arXiv preprint arXiv:2604.20714*, 2026a.

Shiyu He, Minchi Kuang, Mengxian Wang, Bin Hu, and Tingxiang Gu. EvoSpark: Endogenous interactive agent societies for unified long-horizon narrative evolution. *arXiv preprint arXiv:2604.12776*, 2026b.

Yinghui He, Abhishek Panigrahi, Yong Lin, and Sanjeev Arora. Skill-targeted adaptive training. *arXiv preprint arXiv:2510.10023*, 2025a.

Yufei He, Ruoyu Li, Alex Chen, Yue Liu, Yulin Chen, Yuan Sui, Cheng Chen, Yi Zhu, Luca Luo, Frank Yang, and Bryan Hooi. Enabling self-improving agents to learn at test time with human-in-the-loop guidance. In *EMNLP*, 2025b.

Donald O. Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Wiley, 1949.

Jiaheng Hu, Jay Shim, Chen Tang, Yoonchang Sung, Bo Liu, Peter Stone, and Roberto Martín-Martín. Simple recipe works: Vision-language-action models are natural continual learners with reinforcement learning. *arXiv preprint arXiv:2603.11653*, 2026.

Shengran Hu, Cong Lu, and Jeff Clune. Automated design of agentic systems. In *ICLR*, 2025a.

Yuyang Hu, Shichun Liu, Yanwei Yue, Guibin Zhang, Boyang Liu, Fangyi Zhu, Jiahang Lin, Honglin Guo, Shihan Dou, Zhiheng Xi, Senjie Jin, Jiejun Tan, Yanbin Yin, Jiongnan Liu, Zeyu Zhang, Zhongxiang Sun, Yutao Zhu, Hao Sun, Boci Peng, Zhenrong Cheng, Xuanbo Fan, Jiaxin Guo, Xinlei Yu, Zhenhong Zhou, Zewen Hu, Jiahao Huo, Junhao Wang, Yuwei Niu, Yu Wang, Zhenfei Yin, Xiaobin Hu, Yue Liao, Qiankun Li, Kun Wang, Wangchunshu Zhou, Yixin Liu, Dawei Cheng, Qi Zhang, Tao Gui, Shirui Pan, Yan Zhang, Philip Torr, Zhicheng Dou, Ji-Rong Wen, Xuanjing Huang, Yu-Gang Jiang, and Shuicheng Yan. Memory in the age of AI agents. *arXiv preprint arXiv:2512.13564*, 2025b.

Chengsong Huang, Wenhao Yu, Xiaoyang Wang, Hongming Zhang, Zongxia Li, Ruosen Li, Jiaxin Huang, Haitao Mi, and Dong Yu. R-zero: Self-evolving reasoning LLM from zero data. *arXiv preprint arXiv:2508.05004*, 2025.

Jiahao Huang, Peilan Xu, Xiaoya Nan, and Wenjian Luo. Co-evolving agent architectures and interpretable reasoning for automated optimization. *arXiv preprint arXiv:2604.17708*, 2026a.

Jiatan Huang, Zheyuan Zhang, Kaiwen Shi, Yanfang Ye, and Chuxu Zhang. EvolveRouter: Co-evolving routing and prompt for multi-agent question answering. *arXiv preprint arXiv:2604.05149*, 2026b.

Junjie Huang, Jiarui Qin, Di Yin, Weiwen Liu, Yong Yu, Xing Sun, and Weinan Zhang. Remit: RL-guided mid-training for iterative LLM evolution. *arXiv preprint arXiv:2602.03075*, 2026c.

Yi Huang, Bowen Zheng, Yunxi Dong, Hong Tang, Huan Zhao, S. M. Rakibul Hasan Shawon, and Hualiang Zhang. A self-evolving agentic framework for metasurface inverse design. *arXiv preprint arXiv:2604.01480*, 2026d.

Jonas Hübotter, Leander Diaz-Bone, Ido Hakimi, Andreas Krause, and Moritz Hardt. Learning on the job: Test-time curricula for targeted reinforcement learning. *arXiv preprint arXiv:2510.04786*, 2025.

Geoffrey Irving, Paul Christiano, and Dario Amodei. AI safety via debate. *arXiv preprint arXiv:1805.00899*, 2018.

Md Touhidul Islam, Mahir Akgun, and Syed Masum Billah. Shaping credibility judgments in human-GenAI partnership via weaker LLMs: A transactive memory perspective on AI literacy. In *AIED*, 2026.

Seth Jacobowitz. The hidden functions of sycophancy in AI systems: Steering, consistency, and cognitive dependency. *AI & SOCIETY*, 2026.

Ashesh Jain, Shikhar Sharma, Thorsten Joachims, and Ashutosh Saxena. Learning preferences for manipulation tasks from online coactive feedback. *The International Journal of Robotics Research*, 34:1296–1313, 2015.

Rishub Jain, Sophie Bridgers, Lili Janzer, Rory Greig, Tian Huey Teh, and Vladimir Mikulik. Human-AI complementarity: A goal for amplified oversight. *arXiv preprint arXiv:2510.26518*, 2025.

Ujin Jeon, Jiyong Kwon, Madison Ann Sullivan, Caleb Eunho Lee, and Guang Lin. ATLAS: A multi-LLM training framework for EvoDPO with adaptive reference evolution. *arXiv preprint arXiv:2602.02709*, 2026.

Bowen Jiang, Yuan Yuan, Maohao Shen, Zhuoqun Hao, Zhangchen Xu, Zichen Chen, Ziyi Liu, Anvesh Rao Vijjini, Jiashu He, Hanchao Yu, Radha Poovendran, Gregory Wornell, Lyle Ungar, Dan Roth, Sihao Chen, and Camillo Jose Taylor. Personamem-v2: Towards personalized intelligence via learning implicit user personas and agentic memory. *arXiv preprint arXiv:2512.06688*, 2025a.

Bowen Jiang, Taiwei Shi, Ryo Kamoi, Yuan Yuan, Camillo J. Taylor, Longqi Yang, Pei Zhou, and Sihao Chen. One model, all roles: Multi-turn, multi-agent self-play reinforcement learning for conversational social intelligence. *arXiv preprint arXiv:2602.03109*, 2026a.

Dongfu Jiang, Yi Lu, Zhuofeng Li, Zhiheng Lyu, Ping Nie, Haozhe Wang, Alex Su, Hui Chen, Kai Zou, Chao Du, Tianyu Pang, and Wenhu Chen. Verltool: Towards holistic agentic reinforcement learning with tool use. *arXiv preprint arXiv:2509.01055*, 2025b.

Eric Hanchen Jiang, Levina Li, Rui Sun, Xiao Liang, Yubei Li, Yuchen Wu, Haozheng Luo, Hengli Li, Zhi Zhang, Zhaolu Kang, Kai-Wei Chang, and Ying Nian Wu. Agent q-mix: Selecting the right action for LLM multi-agent systems through reinforcement learning. *arXiv preprint arXiv:2604.00344*, 2026b.

Yilei Jiang, Jinyuan Hu, Qianyin Xiao, Yaozhi Zheng, Ruize Ma, Kaituo Feng, Jiaming Han, Tianshuo Peng, Kaixuan Fan, Manyuan Zhang, and Xiangyu Yue. OpenGame: Open agentic coding for games. *arXiv preprint arXiv:2604.18394*, 2026c.

Ruofan Jin, Zaixi Zhang, Mengdi Wang, and Le Cong. Stella: Self-evolving LLM agent for biomedical research. *arXiv preprint arXiv:2507.02004*, 2025.

Nathanael Jo and Manish Raghavan. Incentives shape how humans co-create with generative AI. *arXiv preprint arXiv:2604.03529*, 2026.

Binny Jose, Deepak Joseph, Visakh Mohan, Elizabeth Alexander, Subi K. Varghese, and Abhijith Roy. Outsourcing cognition: The psychological costs of AI-era convenience. *Frontiers in Psychology*, 16, 2025.

Hangoo Kang, Tarun Suresh, Jon Saad-Falcon, and Azalia Mirhoseini. TRACE: Capability-targeted agentic training. *arXiv preprint arXiv:2604.05336*, 2026.

Majeed Kazemitabaar, Justin Chow, Carl Ka To Ma, Barbara J. Ericson, David Weintrop, and Tovi Grossman. Studying the effect of AI code generators on supporting novice learners in introductory programming. In *CHI*, 2023.

Sean W. Kelley and Christoph Riedl. Personalization increases affective alignment but has role-dependent effects on epistemic independence in LLMs. *arXiv preprint arXiv:2603.00024*, 2026.

Greg Kestin, Kelly Miller, Anna Klales, Timothy Milbourne, and Gregorio Ponti. AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, 15(1), 2025.

Devvrit Khatri, Lovish Madaan, Rishabh Tiwari, Rachit Bansal, Sai Surya Duvvuri, Manzil Zaheer, Inderjit S. Dhillon, David Brandfonbrener, and Rishabh Agarwal. The art of scaling reinforcement learning compute for LLMs. *arXiv preprint arXiv:2510.13786*, 2025.

Omar Khattab, Arnav Singhvi, Paridhi Maheshwari, Zhiyuan Zhang, Keshav Santhanam, Sri Vardhamanan, Saiful Haq, Ashutosh Sharma, Thomas T. Joshi, Hanna Moazam, Heather Miller, Matei Zaharia, and Christopher Potts. DSPy: Compiling declarative language model calls into self-improving pipelines. In *ICLR*, 2024.

Kangsan Kim, Minki Kang, Taeil Kim, Yanlai Yang, Mengye Ren, and Sung Ju Hwang. Memory transfer learning: How memories are transferred across domains in coding agents. *arXiv preprint arXiv:2604.14004*, 2026.

Seoyeon Kim and Jaehyung Kim. SPRInG: Continual LLM personalization via selective parametric adaptation and retrieval-interpolated generation. *arXiv preprint arXiv:2601.09974*, 2026.

Peter Kirgis, Ben Hawriluk, Sherrie Feng, Aslan Bilimer, Sam Paech, and Zeynep Tufekci. LLM spirals of delusion: A benchmarking audit study of AI chatbot interfaces. *arXiv preprint arXiv:2604.06188*, 2026.

Christopher Koch. Beyond the steeper curve: AI-mediated metacognitive decoupling and the limits of the dunning-kruger metaphor. *arXiv preprint arXiv:2603.29681*, 2026.

Kenneth R Koedinger and Vincent Aleven. Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19:239–264, 2007.

Akarsh Kumar, Ryan Bahlous-Boldi, Prafull Sharma, Phillip Isola, Sebastian Risi, Yujin Tang, and David Ha. Digital red queen: Adversarial program evolution in core war with LLMs. *arXiv preprint arXiv:2601.03335*, 2026a.

Harsh Kumar, Jonathan Vincentius, Ewan Jordan, and Ashton Anderson. Human creativity in the age of LLMs: Randomized experiments on divergent and convergent thinking. *arXiv preprint arXiv:2410.03703*, 2024.

Harsh Kumar, Zi Kang Mu, Jonathan Vincentius, and Ashton Anderson. Beyond the AI tutor: Social learning with LLM agents. *arXiv preprint arXiv:2604.02677*, 2026b.

Shruti Kumar, Xiaoyu Chen, and Xiaomei Wang. Mapping human-agent co-learning and co-adaptation: A scoping review. *arXiv preprint arXiv:2506.06324*, 2025.

Ujwal Kumar, Alice Saito, Hershraj Niranjani, Rayan Yessou, and Phan Xuan Tan. Evolving interpretable constitutions for multi-agent coordination. *arXiv preprint arXiv:2602.00755*, 2026c.

Seok Joon Kwon. Publish and perish: How AI-accelerated writing without proportional verification investment degrades scientific knowledge. *arXiv preprint arXiv:2604.05714*, 2026.

Taeyoon Kwon, Dongwook Choi, Hyojun Kim, Sunghwan Kim, Seungjun Moon, Beong-woo Kwak, Kuan-Hao Huang, and Jinyoung Yeo. Embodied agents meet personalization: Investigating challenges and solutions through the lens of memory utilization. In *ICLR*, 2026.

Chingkwun Lam, Jiaxin Li, Lingfei Zhang, and Kuo Zhao. Governing evolving memory in LLM agents: Risks, mechanisms, and the stability and safety governed memory (SSGM) framework. *arXiv preprint arXiv:2603.11768*, 2026.

Johann Laux. Institutionalised distrust and human oversight of artificial intelligence: Towards a democratic design of AI governance under the european union AI act. *AI & SOCIETY*, 39(6):2853–2866, 2023.

Johann Laux and Hannah Ruschemeier. Automation bias in the AI act: On the legal implications of attempting to de-bias human oversight of AI. *European Journal of Risk Regulation*, 16(4):1519–1534, 2025.

Harrison Lee, Samrat Phatale, Hassan Mansoor, Thomas Mesnard, Johan Ferret, Kellie Lu, Colton Bishop, Ethan Hall, Victor Carbune, Abhinav Rastogi, and Sushant Prakash. RLAIF vs. RLHF: Scaling reinforcement learning from human feedback with AI feedback. In *ICML*, 2024a.

Hsin-Yu Lee, Pei-Hua Chen, Wei-Sheng Wang, Yueh-Min Huang, and Ting-Ting Wu. Empowering ChatGPT with guidance mechanism in blended learning: effect of self-regulated learning, higher-order thinking skills, and knowledge construction. *International Journal of Educational Technology in Higher Education*, 21(1), 2024b.

Yoonho Lee, Roshen Nair, Qizheng Zhang, Kangwook Lee, Omar Khattab, and Chelsea Finn. Meta-Harness: End-to-end optimization of model harnesses. *arXiv preprint arXiv:2603.28052*, 2026.

Joel Lehman, Jonathan Gordon, Shawn Jain, Kamal Ndousse, Cathy Yeh, and Kenneth O. Stanley. Evolution through large models. *arXiv preprint arXiv:2206.08896*, 2022.

Chenghao Li, Jun Liu, Songbo Zhang, Huadong Jian, Hao Ni, Lik-Hang Lee, Sung-Ho Bae, Guoqing Wang, Yang Yang, and Chaoning Zhang. Experience transfer for multimodal LLM agents in Minecraft game. *arXiv preprint arXiv:2604.05533*, 2026a.

Chengpeng Li, Mingfeng Xue, Zhenru Zhang, Jiaxi Yang, Beichen Zhang, Xiang Wang, Bowen Yu, Binyuan Hui, Junyang Lin, and Dayiheng Liu. Start: Self-taught reasoner with tools. *arXiv preprint arXiv:2503.04625*, 2025a.

Dingming Li, Yingxiu Zhao, Xinrui Cheng, Kangheng Lin, Hongbo Peng, Hongxing Li, Zixuan Wang, Yuhong Dai, Haodong Li, Jia Wang, Yukang Shi, Liang Zhao, Jianjian Sun, Zheng Ge, Xiangyu Zhang, Weiming Lu, Jun Xiao, Yueting Zhuang, and Yongliang Shen. SpatialEvo: Self-evolving spatial intelligence via deterministic geometric environments. *arXiv preprint arXiv:2604.14144*, 2026b.

Hanchen Li, Runyuan He, Qizheng Zhang, Changxiu Ji, Qiuyang Mang, Xiaokun Chen, Lakshya A Agrawal, Wei-Liang Liao, Eric Yang, Alvin Cheung, James Zou, Kunle Olukotun, Ion Stoica, and Joseph E. Gonzalez. Combee: Scaling prompt learning for self-improving language model agents. *arXiv preprint arXiv:2604.04247*, 2026c.

Haotian Li, Shijun Yang, Weizhen Qi, Silei Zhao, Rui Hua, Mingzhu Song, Xiaojian Yang, and Chao Peng. Yunjue agent tech report: A fully reproducible, zero-start in-situ self-evolving agent system for open-ended tasks. *arXiv preprint arXiv:2601.18226*, 2026d.

Jiazheng Li, Emine Yilmaz, Bei Chen, and Dieu-Thu Le. Towards self-improving error diagnosis in multi-agent systems. *arXiv preprint arXiv:2604.17658*, 2026e.

Junkai Li, Yunghwei Lai, Weitao Li, Jingyi Ren, Meng Zhang, Xinhui Kang, Siyu Wang, Peng Li, Ya-Qin Zhang, Weizhi Ma, and Yang Liu. Agent hospital: A simulacrum of hospital with evolvable medical agents. *arXiv preprint arXiv:2405.02957*, 2024a.

Mo Li, L. H. Xu, Qitai Tan, Ting Cao, and Yunxin Liu. Learning to commit: Generating organic pull requests via online repository memory. *arXiv preprint arXiv:2603.26664*, 2026f.

Ruohao Li, Hongjun Liu, Leyi Zhao, Zisu Li, Jiawei Li, Jiajun Jiang, Linning Xu, Chen Zhao, Mingming Fan, and Chen Liang. Swarmsys: Decentralized swarm-inspired agents for scalable and adaptive reasoning. *arXiv preprint arXiv:2510.10047*, 2025b.

Sha Li and Naren Ramakrishnan. Experience as a compass: Multi-agent RAG with evolving orchestration and agent prompts. *arXiv preprint arXiv:2604.00901*, 2026.

Xiaoya Li, Xiaofei Sun, Guoyin Wang, Songqiao Su, Chris Shum, and Jiwei Li. GrandCode: Achieving grandmaster level in competitive programming via agentic reinforcement learning. *arXiv preprint arXiv:2604.02721*, 2026g.

Xinyu Li, Ruiyang Zhou, Zachary C. Lipton, and Leqi Liu. Personalized language modeling from personalized human feedback. *arXiv preprint arXiv:2402.05133*, 2024b.

Yibo Li, Zijie Lin, Ailin Deng, Xuan Zhang, Yufei He, Shuo Ji, Tri Cao, and Bryan Hooi. Just-in-time reinforcement learning: Continual learning in LLM agents without gradient updates. *arXiv preprint arXiv:2601.18510*, 2026h.

Yubo Li and Weiyi Song. Co-alignment: Rethinking alignment as bidirectional human-AI cognitive adaptation. *arXiv preprint arXiv:2509.12179*, 2025.

Zelong Li, Shuyuan Xu, Kai Mei, Wenyue Hua, Balaji Rama, Om Raheja, Hao Wang, He Zhu, and Yongfeng Zhang. AutoFlow: Automated workflow generation for large language model agents. *arXiv preprint arXiv:2407.12821*, 2024c.

ZhaoBin Li and Mark Steyvers. Learning to trust: How humans mentally recalibrate AI confidence signals. *arXiv preprint arXiv:2603.22634*, 2026.

Zongxia Li, Hongyang Du, Chengsong Huang, Xiyang Wu, Lantao Yu, Yicheng He, Jing Xie, Xiaomin Wu, Zhichao Liu, Jiarui Zhang, and Fuxiao Liu. MM-Zero: Self-evolving multi-model vision language models from zero data. *arXiv preprint arXiv:2603.09206*, 2026i.

Jiaqing Liang, Jinyi Han, Weijia Li, Xinyi Wang, Zhoujia Zhang, Zishang Jiang, Ying Liao, Tingyun Li, Ying Huang, Hao Shen, Hanyu Wu, Fang Guo, Keyi Wang, Zhonghua Hong, Zhiyu Lu, Lipeng Ma, Sihang Jiang, and Yanghua Xiao. GenericAgent: A token-efficient self-evolving LLM agent via contextual information density maximization (v1.0). *arXiv preprint arXiv:2604.17091*, 2026.

Weixin Liang, Zachary Izzo, Yaohui Zhang, Haley Lepp, Hancheng Cao, Xuandong Zhao, Lingjiao Chen, Haotian Ye, Sheng Liu, Zhi Huang, Daniel A. McFarland, and James Y. Zou. Monitoring AI-modified content at scale: A case study on the impact of ChatGPT on AI conference peer reviews. *arXiv preprint arXiv:2403.07183*, 2024.

Hui Liao, Chuan Qin, Yongwen Ren, Hao Li, Zhenya Huang, Yanyong Zhang, and Chao Wang. VERDICT: Verifiable evolving reasoning with directive-informed collegial teams for legal judgment prediction. *arXiv preprint arXiv:2603.19306*, 2026.

Jian Liao, Linrong Zhong, Longting Zhe, Handan Xu, Ming Liu, and Tao Xie. Scaffolding computational thinking with ChatGPT. *IEEE Transactions on Learning Technologies*, 17:1628–1642, 2024.

Jessy Lin, Luke Zettlemoyer, Gargi Ghosh, Wen-Tau Yih, Aram Markosyan, Vincent-Pierre Berges, and Barlas Oğuz. Continual learning via sparse memory finetuning. *arXiv preprint arXiv:2510.15103*, 2025.

Jiahang Lin, Shichun Liu, Chengjun Pan, Lizhi Lin, Shihan Dou, Xuanjing Huang, Hang Yan, Zhenhua Han, and Tao Gui. Agentic harness engineering: Observability-driven automatic evolution of coding-agent harnesses. *arXiv preprint arXiv:2604.25850*, 2026a.

Minhua Lin, Hanqing Lu, Zhan Shi, Bing He, Rui Mao, Zhiwei Zhang, Zongyu Wu, Xianfeng Tang, Hui Liu, Zhenwei Dai, Xiang Zhang, Suhang Wang, Benoit Dumoulin, and Jian Pei. Position: Agentic evolution is the path to evolving LLMs. *arXiv preprint arXiv:2602.00359*, 2026b.

Minhua Lin, Zhiwei Zhang, Hanqing Lu, Hui Liu, Xianfeng Tang, Qi He, Xiang Zhang, and Suhang Wang. MemMA: Coordinating the memory cycle through multi-agent reasoning and in-situ self-evolution. *arXiv preprint arXiv:2603.18718*, 2026c.

Xiaoqiang Lin, Zhongxiang Dai, Arun Verma, See-Kiong Ng, Patrick Jaillet, and Bryan Kian Hsiang Low. Prompt optimization with human feedback. *arXiv preprint arXiv:2405.17346*, 2024.

Zichuan Lin, Feiyu Liu, Yijun Yang, Jiafei Lyu, Yiming Gao, Yicheng Liu, Zhicong Lu, Yangbin Yu, Mingyu Yang, Junyou Li, Deheng Ye, and Jie Jiang. UI-Voyager: A self-evolving GUI agent learning via failed experience. *arXiv preprint arXiv:2603.24533*, 2026d.

Bo Liu, Simon Yu, Zichen Liu, Leon Guertler, Penghui Qi, Daniel Balcells, Mickel Liu, Cheston Tan, Weiyan Shi, Min Lin, Wee Sun Lee, and Natasha Jaques. Spiral: Self-play on zero-sum games incentivizes reasoning via multi-agent multi-turn reinforcement learning. In *ICLR*, 2026a.

Haoyue Liu, Zhichao Wang, Yongxin Guo, Haoran Shou, and Xiaoying Tang. Adaptive prompt structure factorization: A framework for self-discovering and optimizing compositional prompt programs. *arXiv preprint arXiv:2604.06699*, 2026b.

Linbo Liu, Guande Wu, Han Ding, Yawei Wang, Qiang Zhou, Yuzhe Lu, Zhichao Xu, Huan Song, Panpan Xu, and Lin Lee Cheong. CLEAR: Context augmentation from contrastive learning of experience via agentic reflection. *arXiv preprint arXiv:2604.07487*, 2026c.

Peigen Liu, Rui Ding, Yuren Mao, Ziyan Jiang, Yuxiang Ye, Yunjun Gao, Ying Zhang, Renjie Sun, Longbin Lai, and Zhengping Qian. OpenHospital: A thing-in-itself arena for evolving and benchmarking LLM-based collective intelligence. *arXiv preprint arXiv:2603.14771*, 2026d.

Qi Liu, Ruochen Hao, Can Li, and Wanjing Ma. Or-agent: Bridging evolutionary search and structured research for automated algorithm discovery. *arXiv preprint arXiv:2602.13769*, 2026e.

Wei Liu, Junlong Li, Xiwen Zhang, Fan Zhou, Yu Cheng, and Junxian He. Diving into self-evolving training for multimodal reasoning. In *ICML*, 2025.

Weize Liu, Minghui Liu, Sy-Tuyen Ho, Souradip Chakraborty, Xiyao Wang, and Furong Huang. Agentic critical training. *arXiv preprint arXiv:2603.08706*, 2026f.

Xiao Liu, Da Yin, Zirui Wu, and Yansong Feng. RefTool: Reference-guided tool creation for knowledge-intensive reasoning. In *ICLR*, 2026g.

Xingyan Liu, Xiyue Luo, Linyu Li, Ganghong Huang, Jianfeng Liu, and Honglin Qiao. SkillForge: Forging domain-specific, self-evolving agent skills in cloud technical support. *arXiv preprint arXiv:2604.08618*, 2026h.

Yunbo Long. AI-Supervisor: Autonomous AI research supervision via a persistent research world model. *arXiv preprint arXiv:2603.24402*, 2026.

Mingfei Lu, Mengjia Wu, Feng Liu, Jiawei Xu, Weikai Li, Haoyang Wang, Zhengdong Hu, Ying Ding, Yizhou Sun, Jie Lu, and Yi Zhang. Choosing how to remember: Adaptive memory structures for LLM agents. *arXiv preprint arXiv:2602.14038*, 2026a.

Zhengxi Lu, Zhiyuan Yao, Jinyang Wu, Chengcheng Han, Qi Gu, Xunliang Cai, Weiming Lu, Jun Xiao, Yueting Zhuang, and Yongliang Shen. SKILL0: In-context agentic reinforcement learning for skill internalization. *arXiv preprint arXiv:2604.02268*, 2026b.

Lennart Luettgau, Vanessa Cheung, Magda Dubois, Keno Juechems, Jessica Bergs, Luke Symes, Henry Davidson, Bessie O'Dell, Hannah Rose Kirk, Max Rollwage, and Christopher Summerfield. People readily follow personal advice from AI but it does not improve their well-being. *arXiv preprint arXiv:2511.15352*, 2025.

Run Luo, Haonan Zhang, Longze Chen, Ting-En Lin, Xiong Liu, Yuchuan Wu, Min Yang, Yongbin Li, Minzheng Wang, Pengpeng Zeng, Lianli Gao, Heng Tao Shen, Yunshui Li, Hamid Alinejad-Rokny, Xiaobo Xia, Jingkuan Song, and Fei Huang. MMEvol: Empowering multimodal large language models with Evol-Instruct. In *Findings of ACL*, 2025.

Yougang Lyu, Xi Zhang, Xinhao Yi, Yuyue Zhao, Shuyu Guo, Wenxiang Hu, Jan Piotrowski, Jakub Kaliski, Jacopo Urbani, Zaiqiao Meng, Lun Zhou, and Xiaohui Yan. EvoScientist: Towards multi-agent evolving AI scientists for end-to-end scientific discovery. *arXiv preprint arXiv:2603.08127*, 2026.

Lin Ma, Hao Peng, Yiming Wang, Hongbin Luo, Jie Liu, Kongjing Gu, Guanlin Wu, Hui Lin, and Lei Ren. Self-evolving multi-agent framework for efficient decision making in real-time strategy scenarios. *arXiv preprint arXiv:2603.23875*, 2026a.

Wenquan Ma, Jiayan Nan, Wenlong Wu, and Yize Chen. What deserves memory: Adaptive memory distillation for LLM agents. *arXiv preprint arXiv:2508.03341*, 2025.

Ziyu Ma, Shidong Yang, Yuxiang Ji, Xucong Wang, Yong Wang, Yiming Hu, Tongwen Huang, and Xiangxiang Chu. SkillClaw: Let skills evolve collectively with agentic evolver. *arXiv preprint arXiv:2604.08377*, 2026b.

James MacGlashan, Mark K. Ho, Robert Loftin, Bei Peng, Guan Wang, David L. Roberts, Matthew E. Taylor, and Michael L. Littman. Interactive learning from policy-dependent human feedback. In *ICML*, 2017.

Brooke N. Macnamara, Ibrahim Berber, M. Cenk Çavuşoğlu, Elizabeth A. Krupinski, Naren Nallapareddy, Noelle E. Nelson, Philip J. Smith, Amy L. Wilson-Delfosse, and Soumya Ray. Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness? *Cognitive Research: Principles and Implications*, 9, 2024.

Varun Magesh, Faiz Surani, Matthew Dahl, Mirac Suzgun, Christopher D. Manning, and Daniel E. Ho. Hallucination-free? assessing the reliability of leading AI legal research tools. *arXiv preprint arXiv:2405.20362*, 2024.

Daniele Mezzadri. The paradox of ethical AI-assisted research. *Journal of Academic Ethics*, 23(4):2653–2667, 2025.

Suyash Mishra. Prism: An evolutionary memory substrate for multi-agent open-ended discovery. *arXiv preprint arXiv:2604.19795*, 2026.

Hussein Mozannar, Gagan Bansal, Cheng Tan, Adam Fourney, Victor Dibia, Jingya Chen, Jack Gerrits, Tyler Payne, Matheus Kunzler Maldaner, Madeleine Grunde-McLaughlin, Eric Zhu, Griffin Bassman, Jacob Alber, Peter Chang, Ricky Loynd, Friederike Niedtner, Ece Kamar, Maya Murad, Rafah Hosn, and Saleema Amershi. Magentic-ui: Towards human-in-the-loop agentic systems. *arXiv preprint arXiv:2507.22358*, 2025.

Hyunji Nam, Yanming Wan, Mickel Liu, Peter Ahnn, Jianxun Lian, and Natasha Jaques. Learning to summarize user information for personalized reinforcement learning from human feedback. In *ICLR*, 2026.

Chiara Natali, Luca Marconi, Leslye Denisse Dias Duran, and Federico Cabitza. AI-induced deskilling in medicine: A mixed-method review and research agenda for healthcare and beyond. *Artificial Intelligence Review*, 58(11), 2025.

Ansong Ni, Miltiadis Allamanis, Arman Cohan, Yinlin Deng, Kensen Shi, Charles Sutton, and Pengcheng Yin. Next: Teaching large language models to reason about code execution. In *ICML*, 2024.

Jingwei Ni, Yihao Liu, Xinpeng Liu, Yutao Sun, Mengyu Zhou, Pengyu Cheng, Dexin Wang, Erchao Zhao, Xiaoxi Jiang, and Guanjun Jiang. Trace2Skill: Distill trajectory-local lessons into transferable agent skills. *arXiv preprint arXiv:2603.25158*, 2026.

Allen Nie, Xavier Daull, Zhiyi Kuang, Abhinav Akkiraju, Anish Chaudhuri, Max Piasevoli, Ryan Rong, YuCheng Yuan, Prerit Choudhary, Shannon Xiao, Rasool Fakoor, Adith Swaminathan, and Ching-An Cheng. Understanding the challenges in iterative generative optimization with LLMs. *arXiv preprint arXiv:2603.23994*, 2026a.

Chang Nie, Chaoyou Fu, Yifan Zhang, Haihua Yang, and Caifeng Shan. PersonaVLM: Long-term personalized multimodal LLMs. *arXiv preprint arXiv:2604.13074*, 2026b.

Zheng Nie, Ruolin Shen, Xinlei Yu, Bo Yin, Jiangning Zhang, and Xiaobin Hu. SkillGraph: Self-evolving multi-agent collaboration with multimodal graph topology. *arXiv preprint arXiv:2604.17503*, 2026c.

Alexander Novikov, Ngân Vũ, Marvin Eisenberger, Emilien Dupont, Po-Sen Huang, Adam Zsolt Wagner, Sergey Shirobokov, Borislav Kozlovskii, Francisco J. R. Ruiz, Abbas Mehrabian, M. Pawan Kumar, Abigail See, Swarat Chaudhuri, George Holland, Alex Davies, Sebastian Nowozin, Pushmeet Kohli, and Matej Balog. Alphaevolve: A coding agent for scientific and algorithmic discovery. *arXiv preprint arXiv:2506.13131*, 2025.

Shakked Noy and Whitney Zhang. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381:187–192, 2023.

Barbara Oakley, Michael Johnston, Ken-Zen Chen, Eulho Jung, and Terrence Sejnowski. The memory paradox: Why our brains need knowledge in an age of AI. In *The Artificial Intelligence Revolution: Challenges and Opportunities*. Springer Nature, 2025.

Isaac Ong, Amjad Almahairi, Vincent Wu, Wei-Lin Chiang, Tianhao Wu, Joseph E. Gonzalez, M Waleed Kadous, and Ion Stoica. Routellm: Learning to route LLMs with preference data. In *ICLR*, 2025.

Open Ended Learning Team, Adam Stooke, Anuj Mahajan, Catarina Barros, Charlie Deck, Jakob Bauer, Jakub Sygnowski, Maja Trebacz, Max Jaderberg, Michael Mathieu, Nat McAleese, Nathalie Bradley-Schmieg, Nathaniel Wong, Nicolas Porcel, Roberta Raileanu, Steph Hughes-Fitt, Valentin Dalibard, and Wojciech Marian Czarnecki. Open-ended learning leads to generally capable agents. *arXiv preprint arXiv:2107.12808*, 2021.

Siru Ouyang, Jun Yan, Yanfei Chen, Rujun Han, Zifeng Wang, Bhavana Dalvi Mishra, Rui Meng, Chun-Liang Li, Yizhu Jiao, Kaiwen Zha, Maohao Shen, Vishy Tirumalashetty, George Lee, Jiawei Han, Tomas Pfister, and Chen-Yu Lee. SkillOS: Learning skill curation for self-evolving agents. *arXiv preprint arXiv:2605.06614*, 2026a.

Siru Ouyang, Jun Yan, I-Hung Hsu, Yanfei Chen, Ke Jiang, Zifeng Wang, Rujun Han, Long T. Le, Samira Daruki, Xiangru Tang, Vishy Tirumalashetty, George Lee, Mahsan Rofouei, Hangfei Lin, Jiawei Han, Chen-Yu Lee, and Tomas Pfister. ReasoningBank: Scaling agent self-evolving with reasoning memory. In *ICLR*, 2026b.

Charles Packer, Sarah Wooders, Kevin Lin, Vivian Fang, Shishir G. Patil, Ion Stoica, and Joseph E. Gonzalez. MemGPT: Towards LLMs as operating systems. *arXiv preprint arXiv:2310.08560*, 2023.

Wenbo Pan, Shujie Liu, Xiangyang Zhou, Shiwei Zhang, Wanlu Shi, Mirror Xu, and Xiaohua Jia. M*: Every task deserves its own memory harness. *arXiv preprint arXiv:2604.11811*, 2026.

Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 30:286–297, 2000.

Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *UIST*, 2023.

Matthew Penaroza. ROZA graphs: Self-improving near-deterministic RAG through evidence-centric feedback. *arXiv preprint arXiv:2604.07595*, 2026.

Xiaoyu Peng, Kangrui Sun, Xin Shan, and Junhan Zhang. Empowerment or dependency? a systematic review of the impacts of intelligent assessment and generative AI on learners' self-beliefs and cognitive agency in music education. *Frontiers in Psychology*, 17, 2026a.

Yulin Peng, Xinxin Zhu, Chenxing Wei, Nianbo Zeng, Leilei Wang, Ying Tiffany He, and F. Richard Yu. SAGE: Multi-agent self-evolution for LLM reasoning. *arXiv preprint arXiv:2603.15255*, 2026b.

Jean Piaget and Margaret Trans Cook. *The origins of intelligence in children.* WW Norton & Company, 1952.

Griffin Pitts, Neha Rani, and Weedguet Mildort. Trust and reliance on AI in education: AI literacy and need for cognition as moderators. *arXiv preprint arXiv:2604.01114*, 2026.

Ihsan Ayyub Qazi, Ayesha Ali, Asad Ullah Khawaja, Muhammad Junaid Akhtar, Ali Zafar Sheikh, and Muhammad Hamad Alizai. Automation bias in large language model assisted diagnostic reasoning among AI-trained physicians. *medRxiv*, 2025.

Zehan Qi, Xiao Liu, Iat Long Iong, Hanyu Lai, Xueqiao Sun, Jiadai Sun, Xinyue Yang, Yu Yang, Shuntian Yao, Wei Xu, Jie Tang, and Yuxiao Dong. WebRL: Training LLM web agents via self-evolving online curriculum reinforcement learning. In *ICLR*, 2025.

Cheng Qian, Emre Can Acikgoz, Qi He, Hongru Wang, Xiusi Chen, Dilek Hakkani-Tur, Gokhan Tur, and Heng Ji. Toolrl: Reward is all tool learning needs. *arXiv preprint arXiv:2504.13958*, 2025.

Hongjin Qian, Zhao Cao, and Zheng Liu. Memobrain: Executive memory as an agentic brain for reasoning. *arXiv preprint arXiv:2601.08079*, 2026.

Jingyang Qiao, Weicheng Meng, Yu Cheng, Zhihang Lin, Zhizhong Zhang, Xin Tan, Jingyu Gong, Kun Shao, and Yuan Xie. Memory intelligence agent. *arXiv preprint arXiv:2604.04503*, 2026.

Yujia Qin, Shihao Liang, Yining Ye, Kunlun Zhu, Lan Yan, Yaxi Lu, Yankai Lin, Xin Cong, Xiangru Tang, Bill Qian, Sihan Zhao, Lauren Hong, Runchu Tian, Ruobing Xie, Jie Zhou, Mark Gerstein, Dahai Li, Zhiyuan Liu, and Maosong Sun. ToolLLM: Facilitating large language models to master 16000+ real-world APIs. In *ICLR*, 2024.

Yulei Qin, Xiaoyu Tan, Zhengbao He, Gang Li, Haojia Lin, Zongyi Li, Zihan Xu, Yuchen Shi, Siqi Cai, Renting Rui, Shaofei Cai, Yuzheng Cai, Xuan Zhang, Sheng Ye, Ke Li, and Xing Sun. Learn the ropes, then trust the wins: Self-imitation with progressive exploration for agentic reinforcement learning. In *ICLR*, 2026.

Jiahao Qiu, Xuan Qi, Hongru Wang, Xinzhe Juan, Yimin Wang, Zelin Zhao, Jiayi Geng, Jiacheng Guo, Peihang Li, Jingzhe Shi, Shilong Liu, and Mengdi Wang. Alita-g: Self-evolving generative agent for agent generation. *arXiv preprint arXiv:2510.23601*, 2025a.

Jiahao Qiu, Xuan Qi, Tongcheng Zhang, Xinzhe Juan, Jiacheng Guo, Yifu Lu, Yimin Wang, Zixin Yao, Qihan Ren, Xun Jiang, Xing Zhou, Dongrui Liu, Ling Yang, Yue Wu, Kaixuan Huang, Shilong Liu, Hongru Wang, and Mengdi Wang. Alita: Generalist agent enabling scalable agentic reasoning with minimal predefinition and maximal self-evolution. *arXiv preprint arXiv:2505.20286*, 2025b.

Yifu Qiu, Zheng Zhao, Waylon Li, Yftah Ziser, Anna Korhonen, Shay B. Cohen, and Edoardo M. Ponti. Self-improving world modelling with latent actions. *arXiv preprint arXiv:2602.06130*, 2026.

Ao Qu, Han Zheng, Zijian Zhou, Yihao Yan, Yihong Tang, Shao Yong Ong, Fenglu Hong, Kaichen Zhou, Chonghe Jiang, Minwei Kong, Jiacheng Zhu, Xuan Jiang, Sirui Li, Cathy Wu, Bryan Kian Hsiang Low, Jinhua Zhao, and Paul Pu Liang. CORAL: Towards autonomous multi-agent evolution for open-ended discovery. *arXiv preprint arXiv:2604.01658*, 2026a.

Changle Qu, Sunhao Dai, Xiaochi Wei, Hengyi Cai, Shuaiqiang Wang, Dawei Yin, Jun Xu, and Ji-Rong Wen. From exploration to mastery: Enabling LLMs to master tools via self-driven interactions. In *ICLR*, 2025.

Yuxiao Qu, Anikait Singh, Yoonho Lee, Amrith Setlur, Ruslan Salakhutdinov, Chelsea Finn, and Aviral Kumar. RLAD: Training LLMs to discover abstractions for solving reasoning problems. In *ICLR*, 2026b.

Christian Rahe and Walid Maalej. How do programming students use generative AI? *Proceedings of the ACM on Software Engineering*, 2(FSE):978–1000, 2025.

Steven Randazzo, Hila Lifshitz-Assaf, Katherine C. Kellogg, Fabrizio Dell'Acqua, Ethan Mollick, François Candelon, and Karim R. Lakhani. Cyborgs, centaurs and self-automators: The three modes of human–GenAI knowledge work and their implications for skilling and the future of expertise. Technical report, Harvard Business School, 2025.

Preston Rasmussen, Pavlo Paliychuk, Travis Beauvais, Jack Ryan, and Daniel Chalef. Zep: A temporal knowledge graph architecture for agent memory. *arXiv preprint arXiv:2501.13956*, 2025.

Pretam Ray, Pratik Prabhanjan Brahma, Zicheng Liu, and Emad Barsoum. Adaptevolve: Improving efficiency of evolutionary AI agents through adaptive model selection. *arXiv preprint arXiv:2602.11931*, 2026.

Jincheng Ren, Siwei Wu, Yizhi Li, Kang Zhu, Shu Xu, Boyu Feng, Ruibin Yuan, Wei Zhang, Riza Batista-Navarro, Jian Yang, and Chenghua Lin. A self-evolving framework for efficient terminal agents via observational context compression. *arXiv preprint arXiv:2604.19572*, 2026a.

Ruiyang Ren, Yuhao Wang, Yunsen Liang, Lan Luo, Jing Liu, Haifeng Wang, Cong Feng, Yinan Zhang, Chunyan Miao, Ji-Rong Wen, and Wayne Xin Zhao. Emulating clinician cognition via self-evolving deep clinical research. *arXiv preprint arXiv:2603.10677*, 2026b.

Olya Rezaeian, Alparslan Emrah Bayrak, and Onur Asan. Explainability and AI confidence in clinical decision support systems: Effects on trust, diagnostic performance, and cognitive load in breast cancer care. *International Journal of Human–Computer Interaction*, 42(6):4477–4497, 2025.

R. Maria del Rio-Chanona, Ekkehard Ernst, Rossana Merola, Daniel Samaan, and Ole Teutloff. AI and jobs. a review of theory, estimates, and evidence. *arXiv preprint arXiv:2509.15265*, 2025.

Rishav Rishav, Pushpak Pujari, and Pushpendre Rastogi. ContraPrompt: Contrastive prompt optimization via dyadic reasoning trace analysis. *arXiv preprint arXiv:2604.17937*, 2026.

Evan F Risko and Sam J Gilbert. Cognitive offloading. *Trends in Cognitive Sciences*, 20:676–688, 2016.

Giuseppe Riva. Invisible architectures of thought: Toward a new science of AI as cognitive infrastructure. *arXiv preprint arXiv:2507.22893*, 2025.

Maxime Robeyns, Martin Szummer, and Laurence Aitchison. A self-improving coding agent. *arXiv preprint arXiv:2504.15228*, 2025.

Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Matej Balog, M. Pawan Kumar, Emilien Dupont, Francisco J. R. Ruiz, Jordan S. Ellenberg, Pengming Wang, Omar Fawzi, Pushmeet Kohli, and Alhussein Fawzi. Mathematical discoveries from program search with large language models. *Nature*, 625:468–475, 2024.

S1-NexusAgent Team. S1-NexusAgent: a self-evolving agent framework for multidisciplinary scientific research. *arXiv preprint arXiv:2602.01550*, 2026.

Vishnu Sarukkai, Zhiqiang Xie, and Kayvon Fatahalian. Self-generated in-context examples improve LLM agents for sequential decision-making tasks. *arXiv preprint arXiv:2505.00234*, 2025.

Mattia Savardi, Alberto Signoroni, Sergio Benini, Filippo Vaccher, Matteo Alberti, Pietro Ciolli, Nunzia Di Meo, Teresa Falcone, Marco Ramanzin, Barbara Romano, Federica Sozzi, and Davide Farina. Upskilling or deskilling? measurable role of an AI-supported training for radiology residents: A lesson from the pandemic. *Insights into Imaging*, 16(1), 2025.

Max Schemmer, Niklas Kühl, Carina Benz, Andrea Bartos, and Gerhard Satzger. Appropriate reliance on AI advice: Conceptualization and the effect of explanations. In *IUI*, 2023.

Ning Shang, Yifei Liu, Yi Zhu, Li Lyna Zhang, Weijiang Xu, Xinyu Guan, Buze Zhang, Bingcheng Dong, Xudong Zhou, Bowen Zhang, Ying Xin, Ziming Miao, Scarlett Li, Fan Yang, and Mao Yang. rstar2-agent: Agentic reasoning technical report. *arXiv preprint arXiv:2508.20722*, 2025.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Mingchuan Zhang, Y.K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Abhishek Sharma and Dan Goldwasser. Cola: Learning to interactively collaborate with large language models. *arXiv preprint arXiv:2504.02965*, 2025.

Hua Shen, Tiffany Knearem, Reshmi Ghosh, Kenan Alkiek, Kundan Krishna, Yachuan Liu, Ziqiao Ma, Savvas Petridis, Yi-Hao Peng, Qiwei Li, Sushrita Rakshit, Chenglei Si, Yutong Xie, Jeffrey P. Bigham, Frank Bentley, Joyce Chai, Zachary Chase Lipton, Qiaozhu Mei, Rada Mihalcea, Michael Terry, Diyi Yang, Meredith Ringel Morris, Paul Resnick, and David Jurgens. Position: Towards bidirectional human-AI alignment. In *NeurIPS*, 2025.

Judy Hanwen Shen and Alex Tamkin. How AI impacts skill formation. *arXiv preprint arXiv:2601.20245*, 2026.

Shuaike Shen, Wenduo Cheng, Mingqian Ma, Alistair Turcan, Martin Jinye Zhang, and Jian Ma. SKILL-FOUNDRY: Building self-evolving agent skill libraries from heterogeneous scientific resources. *arXiv preprint arXiv:2604.03964*, 2026a.

Weixiang Shen, Bailiang Jian, Jun Li, Che Liu, Johannes Moll, Xiaobin Hu, Daniel Rueckert, Hongwei Bran Li, and Jiazhen Pan. Evo-MedAgent: Beyond one-shot diagnosis with agents that remember, reflect, and improve. *arXiv preprint arXiv:2604.14475*, 2026b.

Xintian Shen, Jiawei Chen, Lihao Zheng, Hao Ma, Tao Wei, and Kun Zhan. Evolving from tool user to creator via training-free experience reuse in multimodal reasoning. *arXiv preprint arXiv:2602.01983*, 2026c.

Taiwei Shi, Sihao Chen, Bowen Jiang, Linxin Song, Longqi Yang, and Jieyu Zhao. Experiential reinforcement learning. *arXiv preprint arXiv:2602.13949*, 2026.

Yuchen Shi, Yuzheng Cai, Siqi Cai, Zihan Xu, Lichao Chen, Yulei Qin, Zhijian Zhou, Xiang Fei, Chaofan Qiu, Xiaoyu Tan, Gang Li, Zongyi Li, Haojia Lin, Guocan Cai, Yong Mao, Yunsheng Wu, Ke Li, and Xing Sun. Youtu-agent: Scaling agent productivity with automated generation and hybrid policy optimization. *arXiv preprint arXiv:2512.24615*, 2025.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik R. Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. In *NeurIPS*, 2023.

Pannaga Shivaswamy and Thorsten Joachims. Coactive learning. *Journal of Artificial Intelligence Research*, 53:1–40, 2015.

Prakash Shukla, Phuong Ngo Ngoc Bui, Sean Levy, Maxwell Kowalski, Ali Baigelenov, and Paul Parsons. De-skilling, cognitive offloading, and misplaced responsibilities: Potential ironies of AI-assisted design. In *CHI Extended Abstracts*, 2025.

Chenglei Si, Diyi Yang, and Tatsunori Hashimoto. Can LLMs generate novel research ideas? a large-scale human study with 100+ NLP researchers. *arXiv preprint arXiv:2409.04109*, 2024.

Auste Simkute, Lev Tankelevitch, Viktor Kewenig, Ava Elizabeth Scott, Abigail Sellen, and Sean Rintel. Ironies of generative AI: Understanding and mitigating productivity loss in human-AI interactions. *arXiv preprint arXiv:2402.11364*, 2024.

Baris Simsek. When to forget: A memory governance primitive. *arXiv preprint arXiv:2604.12007*, 2026.

Anjali Singh, Karan Taneja, Zhitong Guan, and Avijit Ghosh. Protecting human cognition in the age of AI. *arXiv preprint arXiv:2502.12447*, 2025.

Anjali Singh, Christopher Brooks, Warren Li, Juho Kim, and Xu Wang. Hint-writing with deferred AI assistance: Fostering critical engagement in data science education. *arXiv preprint arXiv:2604.19931*, 2026.

David Uriel Socol de la Osa and Nydia Remolina. Artificial intelligence at the bench: Legal and ethical challenges of informing—or misinforming—judicial decision-making through generative AI. *Data & Policy*, 6, 2024.

Nicholas C Soderstrom and Robert A Bjork. Learning versus performance: An integrative review. *Perspectives on Psychological Science*, 10:176–199, 2015.

Yuda Song, Lili Chen, Fahim Tajwar, Rémi Munos, Deepak Pathak, J. Andrew Bagnell, Aarti Singh, and Andrea Zanette. Expanding the capabilities of reinforcement learning via text feedback. *arXiv preprint arXiv:2602.02482*, 2026.

Tejas Srinivasan and Jesse Thomason. Adjust for trust: Mitigating trust-induced inappropriate reliance on AI assistance. In *IUI*, 2026.

Chang Su, Zhongkai Hao, Zhizhou Zhang, Zeyu Xia, Youjia Wu, Hang Su, and Jun Zhu. HELIX: Evolutionary reinforcement learning for open-ended scientific problem solving. In *ICLR*, 2026a.

Hanyu Su, Huilin Zhang, and Shihui Feng. Comparing the impact of pedagogy-informed custom and general-purpose GAI chatbots on students' science problem-solving processes and performance using heterogeneous interaction network analysis. *arXiv preprint arXiv:2604.03022*, 2026b.

Yuan Sui and Bryan Hooi. Conversation for non-verifiable learning: Self-evolving LLMs through meta-evaluation. *arXiv preprint arXiv:2601.21464*, 2026.

Theodore R. Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas L. Griffiths. Cognitive architectures for language agents. *Transactions on Machine Learning Research*, 2024.

Qi Sun, Stefan Nielsen, Rio Yokota, and Yujin Tang. Evolutionary context search for automated skill acquisition. *arXiv preprint arXiv:2602.16113*, 2026.

Wangtao Sun, Xiang Cheng, Jialin Fan, Yao Xu, Xing Yu, Shizhu He, Jun Zhao, and Kang Liu. Towards agentic self-learning LLMs in search environment. *arXiv preprint arXiv:2510.14253*, 2025a.

Yu Sun, Xinhao Li, Karan Dalal, Jiarui Xu, Arjun Vikram, Genghan Zhang, Yann Dubois, Xinlei Chen, Xiaolong Wang, Sanmi Koyejo, Tatsunori Hashimoto, and Carlos Guestrin. Learning to (learn at test time): RNNs with expressive hidden states. In *ICML*, 2025b.

Yuan Sun and Ting Wang. Be friendly, not friends: How LLM sycophancy shapes user trust. In *CHI*, 2026.

Arnuv Tandon, Karan Dalal, Xinhao Li, Daniel Koceja, Marcel Rød, Sam Buchanan, Xiaolong Wang, Jure Leskovec, Sanmi Koyejo, Tatsunori Hashimoto, Carlos Guestrin, Jed McCaleb, Yejin Choi, and Yu Sun. End-to-end test-time training for long context. *arXiv preprint arXiv:2512.23675*, 2025.

Ningzhi Tang, Meng Chen, Zheng Ning, Aakash Bansal, Yu Huang, Collin McMillan, and Toby Jia-Jun Li. A study on developer behaviors for validating and repairing LLM-generated code using eye tracking and IDE actions. *arXiv preprint arXiv:2405.16081*, 2024.

Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. The metacognitive demands and opportunities of generative AI. In *CHI*, 2024.

Ao Tian, Yunfeng Lu, Xinxin Fan, Changhao Wang, Lanzhi Zhou, Yeyao Zhang, and Yanfang Liu. Rgmem: Renormalization group-inspired memory evolution for language agents. *arXiv preprint arXiv:2510.16392*, 2025.

Yingtao Tian. Prompt optimization enables stable algorithmic collusion in LLM agents. *arXiv preprint arXiv:2604.17774*, 2026.

Hayato Tomisu, Junya Ueda, and Tsukasa Yamanaka. The cognitive mirror: A framework for AI-powered metacognition and self-regulated learning. *Frontiers in Education*, 10, 2025.

Jonathan Turney, Tim Young, Dhyana Chauhan, Roshni Beeharry, and Mohammad Mahmud. AI use for medical students: Impact on clinical skill acquisition and retention. a systematic review. *Advances in Medical Education and Practice*, 17:1–13, 2026.

Lev S. Vygotsky. *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, 1978.

Gilles Wainrib, Barbara Bodinier, Haitem Dakhli, Josep Monserrat, Almudena Espin Perez, Sabrina Carpentier, Roberta Codato, and John Klein. Can AI scientist agents learn from lab-in-the-loop feedback? Evidence from iterative perturbation discovery. *arXiv preprint arXiv:2603.26177*, 2026.

Chenxi Wang, Zhuoyun Yu, Xin Xie, Wuguannan Yao, Runnan Fang, Shuofei Qiao, Kexin Cao, Guozhou Zheng, Xiang Qi, Peng Zhang, and Shumin Deng. SkillX: Automatically constructing skill knowledge bases for agents. *arXiv preprint arXiv:2604.04804*, 2026a.

Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. Voyager: An open-ended embodied agent with large language models. *Transactions on Machine Learning Research*, 2024a.

Haochen Wang, Yi Wu, Daryl Chang, Li Wei, and Lukasz Heldt. Self-evolving recommendation system: End-to-end autonomous model optimization with LLM agents. *arXiv preprint arXiv:2602.10226*, 2026b.

Jianren Wang, Yifan Su, Abhinav Gupta, and Deepak Pathak. Evolutionary policy optimization. *arXiv preprint arXiv:2503.19037*, 2025a.

Jianzong Wang, Botao Zhao, Yayun He, Junqing Peng, and Xulong Zhang. Evolvable embodied agent for robotic manipulation via long short-term reflection and optimization. *arXiv preprint arXiv:2604.13533*, 2026c.

Jiongxiao Wang, Qiaojing Yan, Yawei Wang, Yijun Tian, Soumya Smruti Mishra, Zhichao Xu, Megha Gandhi, Panpan Xu, and Lin Lee Cheong. Reinforcement learning for self-improving agent with skill library. *arXiv preprint arXiv:2512.17102*, 2025b.

Jun Wang. Memento 2: Learning by stateful reflective memory. *arXiv preprint arXiv:2512.22716*, 2025.

Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, Wayne Xin Zhao, Zhewei Wei, and Ji-Rong Wen. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18, 2024b.

Prince Zizhuang Wang and Shuli Jiang. PRIME: Training free proactive reasoning via iterative memory evolution for user-centric agent. *arXiv preprint arXiv:2604.07645*, 2026a.

Prince Zizhuang Wang and Shuli Jiang. SLEA-RL: Step-level experience augmented reinforcement learning for multi-turn agentic training. *arXiv preprint arXiv:2603.18079*, 2026b.

Qinsi Wang, Bo Liu, Tianyi Zhou, Jing Shi, Yueqian Lin, Yiran Chen, Hai Helen Li, Kun Wan, and Wentian Zhao. Vision-zero: Scalable VLM self-evolution via multi-agent self-play. In *ICLR*, 2026d.

Rose E. Wang, Ana T. Ribeiro, Carly D. Robinson, Susanna Loeb, and Dora Demszky. Tutor CoPilot: A human-AI approach for scaling real-time expertise. *arXiv preprint arXiv:2410.03017*, 2024c.

Ruiyi Wang and Prithviraj Ammanabrolu. A practitioner's guide to multi-turn agentic reinforcement learning. *arXiv preprint arXiv:2510.01132*, 2025.

Sitong Wang, Jocelyn McKinnon-Crowley, Tao Long, Kian Loong Lua, Keren Henderson, Kevin Crowston, Jeffrey V. Nickerson, Mark Hansen, and Lydia B. Chilton. The role of human creativity in the presence of AI creativity tools at work: A case study on AI-driven content transformation in journalism. In *HICSS*, 2026e.

Siyuan Wang, Qing Xia, and Qiong Ye. Synthetic fluency and epistemic offloading in undergraduate mathematics in the age of AI. *arXiv preprint arXiv:2512.21045*, 2025c.

Xiaoxing Wang, Ning Liao, Shikun Wei, Chen Tang, and Feiyu Xiong. AutoAgent: Evolving cognition and elastic memory orchestration for adaptive agents. *arXiv preprint arXiv:2603.09716*, 2026f.

Xinyu Wang, Hanwei Wu, Jingwei Song, Shuyuan Zhang, Jiayi Zhang, Fanqi Kong, Tung Sum Thomas Kwok, Xiao-Wen Chang, Yuyu Luo, Chenglin Wu, and Bang Liu. Co-evolution of policy and internal reward for language agents. *arXiv preprint arXiv:2604.03098*, 2026g.

Yidong Wang, Xin Wang, Cunxiang Wang, Junfeng Fang, Qiufeng Wang, Jianing Chu, Xuran Meng, Shuxun Yang, Libo Qin, Yue Zhang, Wei Ye, and Shikun Zhang. Temporal self-rewarding language models: Decoupling chosen-rejected via past-future. *arXiv preprint arXiv:2508.06026*, 2025d.

Yu Wang, Xinshuang Liu, Xiusi Chen, Sean O'Brien, Junda Wu, and Julian McAuley. Self-updatable large language models by integrating context into model parameters. In *ICLR*, 2025e.

Zihan Wang, Kangrui Wang, Qineng Wang, Pingyue Zhang, Linjie Li, Zhengyuan Yang, Xing Jin, Kefan Yu, Minh Nhat Nguyen, Licheng Liu, Eli Gottlieb, Yiping Lu, Kyunghyun Cho, Jiajun Wu, Li Fei-Fei, Lijuan Wang, Yejin Choi, and Manling Li. RAGEN: Understanding self-evolution in LLM agents via multi-turn reinforcement learning. *arXiv preprint arXiv:2504.20073*, 2025f.

Zihan Wang, Chi Gui, Xing Jin, Qineng Wang, Licheng Liu, Kangrui Wang, Shiqi Chen, Linjie Li, Zhengyuan Yang, Pingyue Zhang, Yiping Lu, Jiajun Wu, Li Fei-Fei, Lijuan Wang, Yejin Choi, and Manling Li. RAGEN-2: Reasoning collapse in agentic RL. *arXiv preprint arXiv:2604.06268*, 2026h.

Zora Zhiruo Wang, Jiayuan Mao, Daniel Fried, and Graham Neubig. Agent workflow memory. In *ICML*, 2025g.

Bingqing Wei, Zhongyu Xia, Dingai Liu, Xiaoyu Zhou, Zhiwei Lin, and Yongtao Wang. ELITE: Experiential learning and intent-aware transfer for self-improving embodied agents. *arXiv preprint arXiv:2603.24018*, 2026a.

Chuyang Wei, Maohang Gao, Zhixin Han, Kefei Chen, Yu Zhuang, Haoxiang Guan, Yanzhi Zhang, Yilin Cheng, Xiren Zhou, Huanhuan Chen, Jian Li, Jiyan He, Yu Shi, Yitong Duan, and Shuxin Zheng. Harnessing pre-resolution signals for future prediction agents. *arXiv preprint arXiv:2604.15719*, 2026b.

Tianxin Wei, Noveen Sachdeva, Benjamin Coleman, Zhankui He, Yuanchen Bei, Xuying Ning, Mengting Ai, Yunzhe Li, Jingrui He, Ed H. Chi, Chi Wang, Shuo Chen, Fernando Pereira, Wang-Cheng Kang, and Derek Zhiyuan Cheng. Evo-memory: Benchmarking LLM agent test-time learning with self-evolving memory. *arXiv preprint arXiv:2511.20857*, 2025a.

Yuxiang Wei, Zhiqing Sun, Emily McMilin, Jonas Gehring, David Zhang, Gabriel Synnaeve, Daniel Fried, Lingming Zhang, and Sida Wang. Toward training superintelligent software agents through self-play swe-RL. *arXiv preprint arXiv:2512.18552*, 2025b.

Zhepei Wei, Wenlin Yao, Yao Liu, Weizhi Zhang, Qin Lu, Liang Qiu, Changlong Yu, Puyang Xu, Chao Zhang, Bing Yin, Hyokun Yun, and Lihong Li. Webagent-r1: Training web agents via end-to-end multi-turn reinforcement learning. In *EMNLP*, 2025c.

Zhaotian Weng, Antonis Antoniades, Deepak Nathani, Zhen Zhang, Xiao Pu, and Xin Eric Wang. Group-evolving agents: Open-ended self-improvement via experience sharing. *arXiv preprint arXiv:2602.04837*, 2026.

Rebecca Westhäußer, Wolfgang Minker, and Sebastian Zepf. Enabling personalized long-term interactions in LLM-based agents through persistent memory and user profiles. *arXiv preprint arXiv:2510.07925*, 2025.

Rong Wu, Xiaoman Wang, Jianbiao Mei, Pinlong Cai, Daocheng Fu, Cheng Yang, Licheng Wen, Xuemeng Yang, Yufan Shen, Yuxin Wang, and Botian Shi. Evolver: Self-evolving LLM agents through an experience-driven lifecycle. *arXiv preprint arXiv:2510.16079*, 2025a.

Rui Wu and Ruixiang Tang. When reward hacking rebounds: Understanding and mitigating it with representation-level signals. *arXiv preprint arXiv:2604.01476*, 2026.

Shirley Wu, Michel Galley, Baolin Peng, Hao Cheng, Gavin Li, Yao Dou, Weixin Cai, James Zou, Jure Leskovec, and Jianfeng Gao. Collabllm: From passive responders to active collaborators. In *ICML*, 2025b.

Xiyang Wu, Zongxia Li, Guangyao Shi, Alexander Duffy, Tyler Marques, Matthew Lyle Olson, Tianyi Zhou, and Dinesh Manocha. Co-evolving LLM decision and skill bank agents for long-horizon tasks. *arXiv preprint arXiv:2604.20987*, 2026a.

Zhaofen Wu, Hanrong Zhang, Fulin Lin, Wujiang Xu, Xinran Xu, Yankai Chen, Henry Peng Zou, Shaowen Chen, Weizhi Zhang, Xue Liu, Philip S. Yu, and Hongwei Wang. GAM: Hierarchical graph-based agentic memory for LLM agents. *arXiv preprint arXiv:2604.12285*, 2026b.

Zhiheng Xi, Yiwen Ding, Wenxiang Chen, Boyang Hong, Honglin Guo, Junzhe Wang, Dingwen Yang, Chenyang Liao, Xin Guo, Wei He, Songyang Gao, Lu Chen, Rui Zheng, Yicheng Zou, Tao Gui, Qi Zhang, Xipeng Qiu, Xuanjing Huang, Zuxuan Wu, and Yu-Gang Jiang. Agentgym: Evolving large language model-based agents across diverse environments. *arXiv preprint arXiv:2406.04151*, 2024.

Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, Rui Zheng, Xiaoran Fan, Xiao Wang, Limao Xiong, Yuhao Zhou, Weiran Wang, Changhao Jiang, Yicheng Zou, Xiangyang Liu, Zhangyue Yin, Shihan Dou, Rongxiang Weng, Wenjuan Qin, Yongyan Zheng, Xipeng Qiu, Xuanjing Huang, Qi Zhang, and Tao Gui. The rise and potential of large language model based agents: A survey. *Science China Information Sciences*, 68, 2025a.

Zhiheng Xi, Jixuan Huang, Chenyang Liao, Baodai Huang, Honglin Guo, Jiaqi Liu, Rui Zheng, Junjie Ye, Jiazheng Zhang, Wenxiang Chen, Wei He, Yiwen Ding, Guanyu Li, Zehui Chen, Zhengyin Du, Xuesong Yao, Yufei Xu, Jiecao Chen, Tao Gui, Zuxuan Wu, Qi Zhang, Xuanjing Huang, and Yu-Gang Jiang. AgentGym-RL: Training LLM agents for long-horizon decision making through multi-turn reinforcement learning. *arXiv preprint arXiv:2509.08755*, 2025b.

Peng Xia, Kaide Zeng, Jiaqi Liu, Can Qin, Fang Wu, Yiyang Zhou, Caiming Xiong, and Huaxiu Yao. Agent0: Unleashing self-evolving agents from zero data via tool-integrated reasoning. *arXiv preprint arXiv:2511.16043*, 2025.

Peng Xia, Jianwen Chen, Hanyang Wang, Jiaqi Liu, Kaide Zeng, Yu Wang, Siwei Han, Yiyang Zhou, Xujiang Zhao, Haifeng Chen, Zeyu Zheng, Cihang Xie, and Huaxiu Yao. Skillrl: Evolving agents via recursive skill-augmented reinforcement learning. *arXiv preprint arXiv:2602.08234*, 2026.

Zhishang Xiang, Chengyi Yang, Zerui Chen, Zhimin Wei, Yunbo Tang, Zongpei Teng, Zexi Peng, Zongxia Li, Chengsong Huang, Yicheng He, Chang Yang, Xinrun Wang, Xiao Huang, Qinggang Zhang, and Jinsong Su. A systematic survey of self-evolving agents: From model-centric to environment-driven co-evolution. Technical report, TechRxiv, 2026.

Teng Xiao, Yige Yuan, Hamish Ivison, Huaisheng Zhu, Faeze Brahman, Nathan Lambert, Pradeep Dasigi, Noah A. Smith, and Hannaneh Hajishirzi. Meta-reinforcement learning with self-reflection for agentic search. *arXiv preprint arXiv:2603.11327*, 2026.

Huaqing Xie. Forage V2: Knowledge evolution and transfer in autonomous agent organizations. *arXiv preprint arXiv:2604.19837*, 2026.

Yunfei Xie, Kevin Wang, Bobby Cheng, Jianzhu Yao, Zhizhou Sha, Alexander Duffy, Yihan Xi, Hongyuan Mei, Cheston Tan, Chen Wei, Pramod Viswanath, and Zhangyang Wang. MEMO: Memory-augmented model context optimization for robust multi-turn multi-agent LLM games. *arXiv preprint arXiv:2603.09022*, 2026a.

Zhengwei Xie, Zhisheng Chen, Ziyan Weng, Jinhan Li, Chenglong Li, Zikai Xiao, Jingwei Song, Jinhao Jing, Vireo Zhang, and Kun Wang. MineEvolve: Self-evolution with accumulated knowledge for long-horizon embodied Minecraft agents. *arXiv preprint arXiv:2603.13131*, 2026b.

Yiming Xiong, Shengran Hu, and Jeff Clune. Learning to continually learn via meta-learning agentic memory designs. *arXiv preprint arXiv:2602.07755*, 2026.

Beibei Xu, Yutong Ye, Chuyun Shen, Yingbo Zhou, Cheng Chen, and Mingsong Chen. HyEvo: Self-evolving hybrid agentic workflows for efficient reasoning. *arXiv preprint arXiv:2603.19639*, 2026a.

Weixian Xu, Tiantian Mi, Yixiu Liu, Yang Nan, Zhimeng Zhou, Lyumanshan Ye, Lin Zhang, Yu Qiao, and Pengfei Liu. ASI-evolve: AI accelerates AI. *arXiv preprint arXiv:2603.29640*, 2026b.

Wujiang Xu, Zujie Liang, Kai Mei, Hang Gao, Juntao Tan, and Yongfeng Zhang. A-mem: Agentic memory for LLM agents. In *NeurIPS*, 2025a.

Wujiang Xu, Jiaojiao Han, Minghao Guo, Kai Mei, Xi Zhu, Han Zhang, and Dimitris N. Metaxas. AEL: Agent evolving learning for open-ended environments. *arXiv preprint arXiv:2604.21725*, 2026c.

Xuhai Xu, Haoyu Hu, Haoran Zhang, Will Ke Wang, Reina Wang, Luis R. Soenksen, Omar Badri, Sheharbano Jafry, Elise Burger, Lotanna Nwandu, Apoorva Mehta, Erik P. Duhaime, Asif Qasim, Hause Lin, Janis Pereira, Jonathan Hershon, Paulius Mui, Alejandro A. Gru, Noémie Elhadad, Lena Mamykina, Matthew Groh, Philipp Tschandl, Roxana Daneshjou, and Marzyeh Ghassemi. Explainable AI as a double-edged sword in dermatology: The impact on clinicians versus the public. *medRxiv*, 2025b.

Xiangyuan Xue, Yifan Zhou, Guibin Zhang, Zaibin Zhang, Yijiang Li, Chen Zhang, Zhenfei Yin, Philip Torr, Wanli Ouyang, and Lei Bai. Comas: Co-evolving multi-agent systems via interaction rewards. *arXiv preprint arXiv:2510.08529*, 2025.

Sikuan Yan, Xiufeng Yang, Zuchao Huang, Ercong Nie, Zifeng Ding, Zonggen Li, Xiaowen Ma, Jinhe Bi, Kristian Kersting, Jeff Z. Pan, Hinrich Schütze, Volker Tresp, and Yunpu Ma. Memory-r1: Enhancing large language model agents to manage and utilize memories via reinforcement learning. *arXiv preprint arXiv:2508.19828*, 2025.

Chengrun Yang, Xuezhi Wang, Yifeng Lu, Hanxiao Liu, Quoc V. Le, Denny Zhou, and Xinyun Chen. Large language models as optimizers. In *ICLR*, 2024a.

Ling Yang, Zhaochen Yu, Tianjun Zhang, Shiyi Cao, Minkai Xu, Wentao Zhang, Joseph E. Gonzalez, and Bin Cui. Buffer of thoughts: Thought-augmented reasoning with large language models. In *NeurIPS*, 2024b.

Shidong Yang, Ziyu Ma, Tongwen Huang, Yiming Hu, Yong Wang, and Xiangxiang Chu. CoEvolve: Training LLM agents via agent-data mutual evolution. *arXiv preprint arXiv:2604.15840*, 2026a.

Yongjin Yang, Sinjae Kang, Juyong Lee, Dongjun Lee, Se-Young Yun, and Kimin Lee. Automated skill discovery for language agents through exploration and iterative feedback. *arXiv preprint arXiv:2506.04287*, 2025a.

Yuqing Yang, Tengxiao Liu, Wang Bill Zhu, Taiwei Shi, Linxin Song, and Robin Jia. Self-evolving LLM memory extraction across heterogeneous tasks. *arXiv preprint arXiv:2604.11610*, 2026b.

Yutao Yang, Jie Zhou, Junsong Li, Qianjun Pan, Bihao Zhan, Qin Chen, Xipeng Qiu, and Liang He. Reinforced interactive continual learning via real-time noisy human feedback. *arXiv preprint arXiv:2505.09925*, 2025b.

Yutao Yang, Junsong Li, Qianjun Pan, Jie Zhou, Kai Chen, Qin Chen, Jingyuan Zhao, Ningning Zhou, Xin Li, and Liang He. PsychAgent: An experience-driven lifelong learning agent for self-evolving psychological counselor. *arXiv preprint arXiv:2604.00931*, 2026c.

Zhicheng Yang, Zhijiang Guo, Yinya Huang, Yongxin Wang, Wenlei Shi, Yiwei Wang, Xiaodan Liang, and Jing Tang. Accordion-thinking: Self-regulated step summaries for efficient and readable LLM reasoning. *arXiv preprint arXiv:2602.03249*, 2026d.

Huaiyuan Yao, Longchao Da, Xiaoou Liu, Charles Fleming, Tianlong Chen, and Hua Wei. LangMARL: Natural language multi-agent reinforcement learning. *arXiv preprint arXiv:2604.00722*, 2026.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing reasoning and acting in language models. In *ICLR*, 2023.

Haotian Ye, Haowei Lin, Jingyi Tang, Yizhen Luo, Caiyin Yang, Chang Su, Rahul Thapa, Rui Yang, Ruihua Liu, Zeyu Li, Chong Gao, Dachao Ding, Guangrong He, Miaolei Zhang, Lina Sun, Wenyang Wang, Yuchen Zhong, Zhuohao Shen, Di He, Jianzhu Ma, Stefano Ermon, Tongyang Li, Xiaowen Chu, James Zou, and Yuzhi Xu. Evaluation-driven scaling for scientific discovery. *arXiv preprint arXiv:2604.19341*, 2026.

Rui Ye, Shuo Tang, Rui Ge, Yaxin Du, Zhenfei Yin, Siheng Chen, and Jing Shao. MAS-GPT: Training LLMs to build LLM-based multi-agent systems. In *ICML*, 2025.

Da Yin, Faeze Brahman, Abhilasha Ravichander, Khyathi Chandu, Kai-Wei Chang, Yejin Choi, and Bill Yuchen Lin. Agent Lumos: Unified and modular training for open-source language agents. In *ACL*, 2024.

Yejin Yoon, Minseo Kim, and Taeuk Kim. Latent preference modeling for cross-session personalized tool calling. *arXiv preprint arXiv:2604.17886*, 2026.

Chenglin Yu, Yang Yu, Songmiao Wang, Yuchen Wang, Yifan Yang, Jinjia Li, Ming Li, and Hongxia Yang. Infiagent: Self-evolving pyramid agent framework for infinite scenarios. *arXiv preprint arXiv:2509.22502*, 2025a.

Cunxi Yu and Haoxing Ren. Autonomous evolution of EDA tools: Multi-agent self-evolved ABC. *arXiv preprint arXiv:2604.15082*, 2026.

Hongli Yu, Tinghong Chen, Jiangtao Feng, Jiangjie Chen, Weinan Dai, Qiying Yu, Ya-Qin Zhang, Wei-Ying Ma, Jingjing Liu, Mingxuan Wang, and Hao Zhou. Memagent: Reshaping long-context LLM with multi-conv RL-based memory agent. *arXiv preprint arXiv:2507.02259*, 2025b.

Hongzhuo Yu, Fei Zhu, Guo-Sen Xie, and Ling Shao. Self-consolidation for self-evolving agents. *arXiv preprint arXiv:2602.01966*, 2026.

Huining Yuan, Zelai Xu, Zheyue Tan, Xiangmin Yi, Mo Guang, Kaiwen Long, Haojia Hui, Boxun Li, Xinlei Chen, Bo Zhao, Xiao-Ping Zhang, Chao Yu, and Yu Wang. Marshal: Incentivizing multi-agent reasoning via self-play with strategic LLMs. *arXiv preprint arXiv:2510.15414*, 2025a.

Jiaqi Yuan, Jialu Wang, Zihan Wang, Qingyun Sun, Ruijie Wang, and Jianxin Li. AgenticGEO: A self-evolving agentic system for generative engine optimization. *arXiv preprint arXiv:2603.20213*, 2026.

Siyu Yuan, Kaitao Song, Jiangjie Chen, Xu Tan, Dongsheng Li, and Deqing Yang. EvoAgent: Towards automatic multi-agent generation via evolutionary algorithms. In *NAACL*, 2025b.

Weizhe Yuan, Richard Yuanzhe Pang, Kyunghyun Cho, Xian Li, Sainbayar Sukhbaatar, Jing Xu, and Jason Weston. Self-rewarding language models. In *ICML*, 2024.

Zhenrui Yue, Kartikeya Upasani, Xianjun Yang, Suyu Ge, Shaoliang Nie, Yuning Mao, Zhe Liu, and Dong Wang. Dr. zero: Self-evolving search agents without training data. *arXiv preprint arXiv:2601.07055*, 2026.

Mert Yuksekgonul, Federico Bianchi, Joseph Boen, Sheng Liu, Pan Lu, Zhi Huang, Carlos Guestrin, and James Zou. Optimizing generative AI by backpropagating language model feedback. *Nature*, 639:609–616, 2025.

Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah D. Goodman. STaR: Bootstrapping reasoning with reasoning. In *NeurIPS*, 2022.

Huaye Zeng, Dongfu Jiang, Haozhe Wang, Ping Nie, Xiaotong Chen, and Wenhu Chen. ACECODER: Acing coder RL via automated test-case synthesis. In *ACL*, 2025.

Yunpeng Zhai, Shuchang Tao, Cheng Chen, Anni Zou, Ziqian Chen, Qingxu Fu, Shinji Mai, Li Yu, Jiaji Deng, Zouying Cao, Zhaoyang Liu, Bolin Ding, and Jingren Zhou. AgentEvolver: Towards efficient self-evolving agent system. *arXiv preprint arXiv:2511.10395*, 2025.

Zhiyuan Zhai, Wenjing Yan, Xiaodan Shao, and Xin Wang. Does RL expand the capability boundary of LLM agents? a pass@(k,t) analysis. *arXiv preprint arXiv:2604.14877*, 2026.

Aimin Zhang, Jiajing Guo, Fuwei Jia, Chen Lv, Boyu Wang, and Fangzheng Li. EvoAgent: An evolvable agent framework with skill learning and multi-agent delegation. *arXiv preprint arXiv:2604.20133*, 2026a.

Di Zhang. Agentdevel: Reframing self-evolving LLM agents as release engineering. *arXiv preprint arXiv:2601.04620*, 2026.

Dingchu Zhang, Yida Zhao, Jialong Wu, Baixuan Li, Wenbiao Yin, Liwen Zhang, Yong Jiang, Yufeng Li, Kewei Tu, Pengjun Xie, and Fei Huang. Evolvesearch: An iterative self-evolving search agent. *arXiv preprint arXiv:2505.22501*, 2025a.

Genghan Zhang, Shaowei Zhu, Anjiang Wei, Zhenyu Song, Allen Nie, Zhen Jia, Nandita Vijaykumar, Yida Wang, and Kunle Olukotun. AccelOpt: A self-improving LLM agentic system for AI accelerator kernel optimization. *arXiv preprint arXiv:2511.15915*, 2025b.

Guibin Zhang, Muxin Fu, Guancheng Wan, Miao Yu, Kun Wang, and Shuicheng Yan. G-memory: Tracing hierarchical memory for multi-agent systems. *arXiv preprint arXiv:2506.07398*, 2025c.

Guibin Zhang, Haotian Ren, Chong Zhan, Zhenhong Zhou, Junhao Wang, He Zhu, Wangchunshu Zhou, and Shuicheng Yan. Memevolve: Meta-evolution of agent memory systems. *arXiv preprint arXiv:2512.18746*, 2025d.

Guibin Zhang, Hejia Geng, Xiaohang Yu, Zhenfei Yin, Zaibin Zhang, Zelin Tan, Heng Zhou, Zhongzhi Li, Xiangyuan Xue, Yijiang Li, Yifan Zhou, Yang Chen, Chen Zhang, Yutao Fan, Zihu Wang, Songtao Huang, Francisco Piedrahita Velez, Yue Liao, Hongru Wang, Mengyue Yang, Heng Ji, Jun Wang, Shuicheng Yan, Philip Torr, and Lei Bai. The landscape of agentic reinforcement learning for LLMs: A survey. *Transactions on Machine Learning Research*, 2026b.

Hanrong Zhang, Shicheng Fan, Henry Peng Zou, Yankai Chen, Zhenting Wang, Jiayu Zhou, Chengze Li, Wei-Chieh Huang, Yifei Yao, Kening Zheng, Xue Liu, Xiaoxiao Li, and Philip S. Yu. CoEvoSkills: Self-evolving agent skills via co-evolutionary verification. *arXiv preprint arXiv:2604.01687*, 2026c.

Haozhen Zhang, Quanyu Long, Jianzhu Bao, Tao Feng, Weizhi Zhang, Haodong Yue, and Wenya Wang. Memskill: Learning and evolving memory skills for self-evolving agents. *arXiv preprint arXiv:2602.02474*, 2026d.

Jenny Zhang, Joel Lehman, Kenneth Stanley, and Jeff Clune. Omni: Open-endedness via models of human notions of interestingness. In *ICLR*, 2024a.

Jenny Zhang, Shengran Hu, Cong Lu, Robert Lange, and Jeff Clune. Darwin gödel machine: Open-ended evolution of self-improving agents. *arXiv preprint arXiv:2505.22954*, 2025e.

Jiaquan Zhang, Chaoning Zhang, Shuxu Chen, Zhenzhen Huang, Pengcheng Zheng, Zhicheng Wang, Ping Guo, Fan Mo, Sung-Ho Bae, Jie Zou, Jiwei Wei, and Yang Yang. Lightweight LLM agent memory with small language models. In *ACL*, 2026e.

Jiayi Zhang, Jinyu Xiang, Zhaoyang Yu, Fengwei Teng, Xiong-Hui Chen, Jiaqi Chen, Mingchen Zhuge, Xin Cheng, Sirui Hong, Jinlin Wang, Bingnan Zheng, Bang Liu, Yuyu Luo, and Chenglin Wu. AFlow: Automating agentic workflow generation. In *ICLR*, 2025f.

Jingyu Zhang, Haozhu Wang, Eric Michael Smith, Sid Wang, Amr Sharaf, Mahesh Pasupuleti, Benjamin Van Durme, Daniel Khashabi, Jason Weston, and Hongyuan Zhan. The alignment waltz: Jointly training agents to collaborate for safety. In *ICLR*, 2026f.

Lunjun Zhang, Ryan Chen, and Bradly C. Stadie. Evolutionary system prompt learning for reinforcement learning in LLMs. *arXiv preprint arXiv:2602.14697*, 2026g.

Mingda Zhang, Wenjin Liu, Tiesunlong Shen, Qika Lin, Rui Mao, Erik Cambria, Xiaoying Tang, and Haoran Luo. FlowSteer: Towards agents designing agentic workflows via reinforced progressive canvas editing. *arXiv preprint arXiv:2602.01664*, 2026h.

Peiyan Zhang, Haibo Jin, Leyang Hu, Xinnuo Li, Liying Kang, Man Luo, Yangqiu Song, and Haohan Wang. REVOLVE: Optimizing AI systems by tracking response evolution in textual optimization. In *ICML*, 2025g.

Qi Zhang, Shen Huang, Chu Liu, Shouqing Yang, Junbo Zhao, Haobo Wang, and Pengjun Xie. DeltaMem: Towards agentic memory management via reinforcement learning. *arXiv preprint arXiv:2604.01560*, 2026i.

Qifan Zhang, Dongyang Ma, Tianqing Fang, Jia Li, Jing Tang, Nuo Chen, Haitao Mi, and Yan Wang. Training LLM agents for spontaneous, reward-free self-evolution via world knowledge exploration. *arXiv preprint arXiv:2604.18131*, 2026j.

Qizheng Zhang, Changran Hu, Shubhangi Upasani, Boyuan Ma, Fenglu Hong, Vamsidhar Kamanuru, Jay Rainton, Chen Wu, Mengmeng Ji, Hanchen Li, Urmish Thakker, James Zou, and Kunle Olukotun. Agentic context engineering: Evolving contexts for self-improving language models. In *ICLR*, 2026k.

Shaokun Zhang, Yi Dong, Jieyu Zhang, Jan Kautz, Bryan Catanzaro, Andrew Tao, Qingyun Wu, Zhiding Yu, and Guilin Liu. Nemotron-research-tool-n1: Exploring tool-using language models with reinforced reasoning. *arXiv preprint arXiv:2505.00024*, 2025h.

Shaowei Zhang, Faqiang Qian, Yan Chen, Ziliang Wang, Kang An, Yong Dai, Mengya Gao, and Yichao Wu. SELF-EMO: Emotional self-evolution from recognition to consistent expression. *arXiv preprint arXiv:2604.18003*, 2026l.

Shimao Zhang, Xiao Liu, Xin Zhang, Junxiao Liu, Zheheng Luo, Shujian Huang, and Yeyun Gong. Process-based self-rewarding language models. In *Findings of ACL*, 2025i.

Shuo Zhang, Chaofa Yuan, Ryan Guo, Xiaomin Yu, Rui Xu, Zhangquan Chen, Zinuo Li, Zhi Yang, Shuhao Guan, Zhenheng Tang, Sen Hu, Liwen Zhang, Ronghao Chen, and Huacan Wang. Evofsm: Controllable self-evolution for deep research with finite state machines. *arXiv preprint arXiv:2601.09465*, 2026m.

Sinin Zhang, Yunfei Xie, Yuxuan Cheng, Haoyu Zhang, and Tong Zhang. PhysNote: Self-knowledge notes for evolvable physical reasoning in vision-language model. *arXiv preprint arXiv:2604.24443*, 2026n.

Wenqi Zhang, Ke Tang, Hai Wu, Mengna Wang, Yongliang Shen, Guiyang Hou, Zeqi Tan, Peng Li, Yueting Zhuang, and Weiming Lu. Agent-Pro: Learning to evolve via policy-level reflection and optimization. In *ACL*, 2024b.

Wentao Zhang, Zhe Zhao, Haibin Wen, Yingcheng Wu, Cankun Guo, Ming Yin, Bo An, and Mengdi Wang. Autogenesis: A self-evolving agent protocol. *arXiv preprint arXiv:2604.15034*, 2026o.

Xiaoying Zhang, Zichen Liu, Yipeng Zhang, Xia Hu, and Wenqi Shao. RetroAgent: From solving to evolving via retrospective dual intrinsic feedback. *arXiv preprint arXiv:2603.08561*, 2026p.

Yabo Zhang, Yihan Zeng, Qingyun Li, Zhen Hu, Kavin Han, and Wangmeng Zuo. Tool-r1: Sample-efficient reinforcement learning for agentic tool use. *arXiv preprint arXiv:2509.12867*, 2025j.

Yaolun Zhang, Ruohui Wang, Jiahao Wang, Yepeng Tang, Xuanyu Zheng, Haonan Duan, Hao Lu, Hanming Deng, and Lewei Lu. EVA: Efficient reinforcement learning for end-to-end video agent. In *CVPR*, 2026q.

Yaolun Zhang, Yiran Wu, Yijiong Yu, Qingyun Wu, and Huazheng Wang. Live-evo: Online evolution of agentic memory from continuous feedback. *arXiv preprint arXiv:2602.02369*, 2026r.

Zhang Zhang, Shuqi Lu, Hongjin Qian, Di He, and Zheng Liu. AgentFactory: A self-evolving framework through executable subagent accumulation and reuse. *arXiv preprint arXiv:2603.18000*, 2026s.

Zhaoxi Zhang, Yitong Duan, Yanzhi Zhang, Yiming Xu, Zhixiang Wang, Kun Liang, Yang Li, Jiahui Liang, Deguo Xia, Jizhou Huang, Jiyan He, and Yunfang Wu. One tool is enough: Reinforcement learning for repository-level LLM agents. *arXiv preprint arXiv:2512.20957*, 2025k.

Zijian Zhang, Aiwei Yin, Amaan Baweja, Jiaru Bai, Ignacio Gustin, Varinia Bernales, and Alán Aspuru-Guzik. El agente forjador: Task-driven agent generation for quantum simulation. *arXiv preprint arXiv:2604.14609*, 2026t.

Andrew Zhao, Yiran Wu, Yang Yue, Tong Wu, Quentin Xu, Yang Yue, Matthieu Lin, Shenzhi Wang, Qingyun Wu, Zilong Zheng, and Gao Huang. Absolute zero: Reinforced self-play reasoning with zero data. *arXiv preprint arXiv:2505.03335*, 2025.

Junhao Zheng, Chengming Shi, Xidi Cai, Qiuke Li, Duzhen Zhang, Chenxing Li, Dong Yu, and Qianli Ma. Lifelong learning of large language model based agents: A roadmap. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Lei Zheng, Weinan Song, Daili Li, and Yanming Yang. To know is to construct: Schema-constrained generation for agent memory. *arXiv preprint arXiv:2604.20117*, 2026.

Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. Memorybank: Enhancing large language models with long-term memory. In *AAAI*, 2024.

Han Zhou, Xingchen Wan, Ivan Vulić, and Anna Korhonen. Agentic policy optimization via instruction-policy co-evolution. *arXiv preprint arXiv:2512.01945*, 2025a.

Huichi Zhou, Yihang Chen, Siyuan Guo, Xue Yan, Kin Hei Lee, Zihan Wang, Ka Yiu Lee, Guchun Zhang, Kun Shao, Linyi Yang, and Jun Wang. Memento: Fine-tuning LLM agents without fine-tuning LLMs. *arXiv preprint arXiv:2508.16153*, 2025b.

Huichi Zhou, Siyuan Guo, Anjie Liu, Zhongwei Yu, Ziqin Gong, Bowen Zhao, Zhixun Chen, Menglong Zhang, Yihang Chen, Jinsong Li, Runyu Yang, Qiangbin Liu, Xinlei Yu, Jianmin Zhou, Na Wang, Chunyang Sun, and Jun Wang. Memento-skills: Let agents design agents. *arXiv preprint arXiv:2603.18743*, 2026.

Wangchunshu Zhou, Yixin Ou, Shengwei Ding, Long Li, Jialong Wu, Tiannan Wang, Jiamin Chen, Shuai Wang, Xiaohua Xu, Ningyu Zhang, Huajun Chen, and Yuchen Eleanor Jiang. Symbolic learning enables self-evolving agents. *AI Open*, 6:314–322, 2025c.

Yongchao Zhou, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, and Jimmy Ba. Large language models are human-level prompt engineers. In *ICLR*, 2023.

Zijian Zhou, Ao Qu, Zhaoxuan Wu, Sunghwan Kim, Alok Prakash, Daniela Rus, Jinhua Zhao, Bryan Kian Hsiang Low, and Paul Pu Liang. Mem1: Learning to synergize memory and reasoning for efficient long-horizon agents. *arXiv preprint arXiv:2506.15841*, 2025d.

Jinchang Zhu, Jindong Li, Cheng Zhang, Jiahong Liu, and Menglin Yang. HeLa-mem: Hebbian learning and associative memory for LLM agents. *arXiv preprint arXiv:2604.16839*, 2026a.

Xinyu Zhu, Yuzhu Cai, Zexi Liu, Cheng Wang, Fengyang Li, Wenkai Jin, Wanxu Liu, Zehao Bing, Bingyang Zheng, Jingyi Chai, Shuo Tang, Rui Ye, Yuwen Du, Xianghe Pang, Yaxin Du, Tingjia Miao, Yuzhi Zhang, Ruoxue Liao, Zhaohan Ding, Linfeng Zhang, Yanfeng Wang, Weinan E, and Siheng Chen. EvoMaster: A foundational evolving agent framework for agentic science at scale. *arXiv preprint arXiv:2604.17406*, 2026b.

Yinghao Zhu, Yifan Qi, Zixiang Wang, Lei Gu, Dehao Sui, Haoran Hu, Xichen Zhang, Ziyi He, Junjun He, Liantao Ma, and Lequan Yu. Healthflow: A self-evolving AI agent with meta planning for autonomous healthcare research. *arXiv preprint arXiv:2508.02621*, 2025.

Yuchen Zhuang, Xiang Chen, Tong Yu, Saayan Mitra, Victor Bursztyn, Ryan A. Rossi, Somdeb Sarkhel, and Chao Zhang. Toolchain*: Efficient action space navigation in large language models with a* search. In *ICLR*, 2024.

Ziqing Zhuang, Linhai Zhang, Jiasheng Si, Deyu Zhou, and Yulan He. Beyond meta-reasoning: Metacognitive consolidation for self-improving LLM reasoning. *arXiv preprint arXiv:2604.17399*, 2026.

Deyu Zou, Yongqiang Chen, Fan Feng, Mufei Li, Pan Li, Yu Gong, and James Cheng. On information self-locking in reinforcement learning for active reasoning of LLM agents. *arXiv preprint arXiv:2603.12109*, 2026a.

Mingxi Zou, Jiaxiang Chen, Aotian Luo, Jingyi Dai, Chi Zhang, Dongning Sun, and Zenglin Xu. Finevo: From isolated backtests to ecological market games for multi-agent financial strategy evolution. *arXiv preprint arXiv:2602.00948*, 2026b.

Adam Zweiger, Jyothish Pari, Han Guo, Ekin Akyürek, Yoon Kim, and Pulkit Agrawal. Self-adapting language models. *arXiv preprint arXiv:2506.10943*, 2025.