Grounded Scaling: Why Agentic AI Needs Deterministic Environments

Liang Ding[1] **and Xintong Wang**[1]
[1]Alibaba Group

Long-chain agent execution fails exponentially in environments designed for human tolerance: with per-step determinism $\delta < 1$, k-step chain success degrades as δ^k. The AGI-to-ASI scaling debate (Genewein et al., 2026) has so far framed progress as a race between compute growth and a list of frictions (data wall, abstraction barrier, embodied bottleneck, multi-agent trust); we argue that environment determinism is a complementary binding axis cutting across all four, for the broad class of agentic AI tasks whose outcomes are verifiable economically, physically, or through multi-party settlement. Three formal results pin down the regime: a Determinism–Efficiency Bound on chain-task success, a Verifier–Goodharting Floor on flywheel ceilings under imperfect rewards, and a convergence condition for environment-side skill evolution. We operationalise the framework as a Supply Certainty Index (SCI) over five measurable properties, a five-level Determinism Maturity Model (DMM) as adoption ladder, and a falsifiable open-question programme (OQ1–OQ5) with explicit null results that would force retraction. The position is platform-agnostic. We engage three competing positions: sim-to-real sufficiency, alignment sufficiency, and AI-as-normal-technology.

Keywords: deterministic agentic AI, grounded scaling, supply certainty, verifier-Goodharting, AGI, ASI

> *"An ant, viewed as a behaving system, is quite simple. The apparent complexity of its behavior over time is largely a reflection of the complexity of the environment in which it finds itself."*
>
> Herbert A. Simon, *The Sciences of the Artificial* (1969)

Contents

Corresponding author(s): zuorui.dl@alibaba-inc.com

1. Introduction and Position Statement

The determinism gap

The most influential recent surveys of the path from artificial general intelligence (AGI) to artificial superintelligence (ASI) frame the open question as a race (Bengio et al., 2025; Genewein et al., 2026; Morris et al., 2024). On one side stands the growth rate of Effective Compute — the product of nominal hardware progress, capital investment, and algorithmic efficiency (Ding et al., 2023; Hernandez and Brown, 2020; Ho et al., 2024; Sevilla et al., 2022), estimated at roughly 10× per year. On the other stand a set of frictions: a coming Data Wall (Shumailov et al., 2024; Villalobos et al., 2024), an Abstraction Barrier that current learners may not be able to cross (Chollet, 2019; Ortega et al., 2021), an Embodied Bottleneck that constrains recursive self-improvement to real-world clock time (Lawrence, 2024), and a requirement for Multi-Agent Trust as the building block of virtual agent economies (Drexler, 2019; Tomašev et al., 2025). We do not dispute that compute and frictions both matter. We dispute that they are on the same axis.

A parallel but under-discussed factor cuts across all four frictions: *environmental determinism*. Current agent environments — web platforms, enterprise APIs, consumer applications — were designed for human tolerance. Search engines intentionally shuffle top results for diversity; recommendation systems inject exploration noise; response latencies vary by orders of magnitude; session state makes identical queries return different answers. For a human user, this non-determinism is tolerable or even desirable. For an autonomous agent executing a multi-step task chain, each non-deterministic step compounds failure probability exponentially.

Definition 1 (Deterministic Agentic Environment)**.** *An environment $\mathcal{E}$ is* deterministic for agentic consumption *if, for any well-formed intent q, the environment returns responses satisfying four conditions:*

- ***(D1)*** Stability*: repeated queries with identical parameters yield consistent results within a declared staleness bound (covering session-conditioned personalisation as a special case of failed stability);*
- ***(D2)*** Faithful ranking*: results are ordered by relevance to the stated intent without adversarial reordering or exploration injection (a relevance-ordering condition that strengthens, not duplicates, (D1));*
- ***(D3)*** Verifiability*: each returned item can be checked against a ground-truth state (inventory, price, certification, availability) via an independent verification channel;*
- ***(D4)*** Bounded latency*: response time is bounded by a declared SLA, enabling agents to plan execution within deterministic time budgets.*

A Deterministic Environment enables reliable multi-step execution; a stochastic environment designed for human tolerance degrades agent performance exponentially with chain length (Proposition 1).

Motivating example: long-chain procurement. Consider an autonomous agent executing a procurement task across any supply platform: find candidate suppliers → compare specifications and pricing → inquire about customisation → negotiate terms → pay → verify fulfilment. This is a six-step chain. If each step has independent determinism quality $\delta < 1$ (e.g., the search engine occasionally returns stale inventory, the pricing API has variable latency that triggers timeouts, the comparison module personalises rankings), the chain-success probability degrades as δ^6. At $\delta = 0.9$ per step, chain success is only 53%; at $\delta = 0.8$, it falls to 26%. In practice, any blockage at any step forces the agent to reroute — switching vendors, switching platforms, or abandoning the task entirely. This rerouting behaviour is universal: it occurs in e-commerce, logistics, healthcare procurement, agricultural commodity trading, and financial settlement. The determinism of the environment is not a convenience; it is a *prerequisite for scalable agentic execution*.

Why now: the 2025–2026 agent inflection

We do not claim a sharp 2025–2026 discontinuity — each of the three developments below has direct 2023–2024 antecedents. Rather, they constitute a *continuation* that crosses a threshold of practical salience: the consequences of environmental non-determinism became operationally measurable, not just theoretically anticipated.

Operator-class deployed agents. Major laboratories shipped agents that act on live, human-optimised web and desktop environments at production scale: Anthropic's Computer Use (Anthropic, 2024a), OpenAI's Operator (OpenAI, 2025), and a generation of open-source followers operate on the same stochastic surfaces our argument identifies. Independent evaluations on OSWorld and τ-bench show that long-horizon success rates remain well below single-turn ones (Jimenez et al., 2024; Xie et al., 2024; Yao et al., 2024), exactly the failure profile our Proposition 1 predicts.

Agent-to-agent commerce protocols. The standardisation substrate for agent-mediated transactions has moved from research to deployment: Anthropic's MCP (Anthropic, 2024b), OpenAI function calling (OpenAI, 2024), Google's Agent2Agent protocol (Google, 2025), and payment-side agent SDKs from Stripe and others (Stripe, 2025). The bottleneck has shifted from *can agents talk* to *can they transact reliably on non-deterministic backends*.

Self-evolving agent stacks. Skill libraries, tool graphs, and persistent context now evolve from deployed trajectories (Feng et al., 2026; Gao et al., 2025; Yamada et al., 2024; Yang et al., 2026; Zhang et al., 2025a,b), making the verifier signal that gates those updates the binding constraint on environment-side learning quality (formalised below as Proposition 3).

Position

The frictions above are grounding problems, not compute problems. We use "grounding" technically: the verified supply of real-world state as training and execution signal for agentic systems, where verifiability comes from an independent ground-truth channel (a payment cleared, a shipment arrived, a specification was met). This sense is distinct from symbol grounding (Harnad, 1990), visual grounding (Plummer et al., 2015), embodied grounding in robotics, or perceptual-symbol-systems grounding (Barsalou, 1999; Lake et al., 2017). The question is not how to learn human-like concepts but how to deploy concept-using agents into real environments reliably.

The policy lever that most concentrates leverage on AGI→ASI progress is therefore the construction of substrates that supply verifiable real-world signal deterministically. Our central claim: one such substrate already exists at industrial scale in economically self-sustaining commercial supply environments (B2B sourcing, B2C retail, pharmaceutical supply chains, agricultural commodity exchanges, automotive supplier networks). Such environments are structurally distinct from simulation and self-generation along dimensions discussed in Section 4, and that distinctness translates into measurable advantages on each of the four bottlenecks *when the environment satisfies Definition 1*. The position is platform-agnostic. The claim is bounded: a binding axis for tasks whose verification is naturally economic, physical, or multi-party (Section 9).

Falsifiable strong form

We treat the position as a scientific claim, not a slogan, and commit to a falsifiable strong form:

Box 1 | Falsifiable strong form (cf. Section 8).

If the bottlenecks named above are indeed grounding problems, then real supply environments that are simultaneously *economically self-sustaining*, *verifier-equipped*, and *deterministic* (in the sense of Definition 1) should systematically outperform pure simulation and pure self-generated data in crossing the data wall, the abstraction barrier, and the embodied bottleneck. This advantage must be measurable in sample efficiency, in time-to-onset of recursive degeneration, and in the rate at which novel concepts beyond human ontology are discoverable. The advantage is in principle refutable by the experiments OQ1–OQ5 of Section 8.

Roadmap

The paper extends the AGI→ASI position literature (Bostrom, 2014; Genewein et al., 2026; Morris et al., 2024; Narayanan and Kapoor, 2024; Sutton, 2019) with three formal results (δ, ε, $\delta_{\min}$), one composite measurement (SCI), one adoption ladder (DMM), and an investable, falsifiable research agenda.

Section 2 restates the AGI→ASI bottleneck landscape and introduces the grounded-scaling lens. Section 3 derives the grounding common denominator and states the Determinism–Efficiency Bound and the Verifier–Goodharting Floor. Section 4 argues why supply environments satisfying (D1)–(D4) are structurally privileged. Section 5 decomposes supply certainty and constructs the SCI. Section 6 treats data-flywheel sustainability and states the Grounded Self-Evolution Convergence Condition. Section 7 treats multi-agent scaling under supply trust and determinism, and introduces the DMM. Section 8 states the open-question programme. Sections 9–10 close with counterarguments, competing positions, and conclusion.

2. Background: Pathways, Bottlenecks, and the Grounding Lens

We summarise the AGI→ASI landscape so that the grounding reframing of Section 3 is recognisable as the *same* landscape, not a strawman.

2.1. Four pathways

The literature converges on four candidate pathways from AGI to ASI (Bostrom, 2014; Genewein et al., 2026; Morris et al., 2024):

1. **Scaling**: continued growth in compute, model size, and training data (AI, 2024; Hoffmann et al., 2022; Kaplan et al., 2020). The bitter lesson (Sutton, 2019) is a claim about *algorithms* (general methods leveraging compute beat clever human-designed priors); our position is about *substrates* (which environment families a given algorithm trains and deploys on). The two axes compose: a bitter-lesson-compliant algorithm deployed against a high-δ substrate dominates both an algorithm with hand-designed priors and a bitter-lesson algorithm deployed against a low-δ substrate.
2. **Paradigm shift**: replacement or augmentation of the present neural paradigm by new architectures, world models, or test-time search (Bruce et al., 2024; Gu and Dao, 2023; Hafner et al., 2020; Snell et al., 2024).
3. **Recursive self-improvement**: models that modify their own training process, data, or architecture (Davidson et al., 2026; Lu et al., 2024; Real et al., 2020; Schmidhuber, 2003), including

the Darwin Gödel Machine (Zhang et al., 2025a), the AI Scientist-v2 (Yamada et al., 2024), and empirical studies on safe self-improvement (Anthropic, 2025).

4. **Multi-agent coordination**: virtual agent economies in which competence is distributed across specialised agents (Drexler, 2019; Hong et al., 2024; Park et al., 2023; Sumers et al., 2024; Tomašev et al., 2025; Wu et al., 2023).

2.2. Six bottlenecks

A parallel literature documents the frictions that may prevent these pathways from succeeding (Genewein et al., 2026): the Data Wall (Shumailov et al., 2024; Villalobos et al., 2024); the economic cost of frontier training (Agrawal et al., 2025; Erdil et al., 2025); paradigm limits (Chollet, 2019); diminishing returns in research (Bloom et al., 2020); the Abstraction Barrier and Embodied Bottleneck (Lawrence, 2024; Ortega et al., 2021); and deliberate slowdown via governance (Anderljung et al., 2023; Bengio et al., 2024; European Parliament and Council, 2024).

2.3. From quantitative scaling to grounded scaling

We propose a single organising distinction:

- *Quantitative scaling* asks how much marginal capability is bought per marginal unit of effective compute, holding the data and verifier distribution fixed.
- *Grounded scaling* asks how the marginal capability per unit compute changes when the underlying data and verifier distribution grow in *verifiable real-world signal*.

The recent empirical record is consistent with the view that quantitative scaling alone reaches a ceiling set by the second quantity (Ho et al., 2025; Hoffmann et al., 2022). We take this as a prompt to study the bottlenecks *as grounding gaps*, which we do in Section 3.

2.4. The determinism gap in current agent environments

A critical but under-studied dimension of the grounded-scaling problem is *environmental determinism*. Current agent benchmarks — WebArena (Zhou et al., 2023), AgentBench (Liu et al., 2024), GAIA (Mialon et al., 2024), OSWorld (Xie et al., 2024), τ-bench (Yao et al., 2024), SWE-bench (Jimenez et al., 2024) — are partially stochastic: web pages change between runs, API responses vary, and session state introduces irreproducibility. Real deployed environments are worse: A/B testing, personalisation, rate limiters, and anti-bot mechanisms deliberately introduce non-determinism optimised for human engagement. For autonomous agents executing multi-step plans, this stochasticity is catastrophic, and the published gap between single-turn LLM ability and end-to-end task success is empirically large (Anthropic, 2024a; OpenAI, 2025); agent-benchmark reproducibility analyses confirm that environmental variance is a first-order driver of this gap (Kapoor et al., 2024). Closing this *determinism gap* requires either redesigning environments for agentic consumption or building agents robust enough to tolerate arbitrary non-determinism. We argue in Section 3 that the former is more tractable and more leveraged.

Cognitive-architecture responses to this failure profile (Shinn et al., 2023; Sumers et al., 2024; Wang et al., 2024a; Yao et al., 2023) add memory, planning, and skill-acquisition layers *around* the model. Our position is complementary: such architectures can mitigate but not erase exponential degradation in δ^k, because every retry and replanning step itself consumes a finite budget of deterministic queries (Section 9).

2.5. An upper bound: universal intelligence and the role of environment

Theoretically, the continuum of machine intelligence is bounded above by universal artificial intelligence formalised through Solomonoff prediction and AIXI (Hutter, 2004; Hutter et al., 2024; Legg and Hutter, 2007). Two corollaries follow. First, the score is environment-relative: a learner universal-intelligent in distribution may still be undertrained on the particular environment of interest. Second, every concrete agent trains on some restricted family of environments, and the choice of that family becomes a design variable. Our position is that which family one privileges is the critical decision left under-discussed, and that commercially self-sustaining supply environments satisfying Definition 1 form a particularly informative family.

2.6. Notation and assumed reader

We assume familiarity with scaling-law notation, RLHF / RLAIF / RLVR loops (Bai et al., 2022; Christiano et al., 2018; Lambert et al., 2024), and multi-agent vocabulary. Glossary entries recap key terms (Grounding, Supply Certainty, Verifier, Customer-to-Manufacturer (C2M), Agent-to-Agent (A2A), Data Flywheel, Grounded Scaling, Model Collapse, Deterministic Environment). The Supply Certainty Index (SCI) and Determinism Maturity Model (DMM) are defined where they first appear (Sections 5, 7.7).

3. The Grounding Thesis and Two Formal Bounds

The four bottlenecks of Section 2 are specialisations of a single deeper deficit: agents lack *verifiable real-world state* along four distinct modalities. We call the deficit Grounding and the modalities *data*, *representational*, *embodied*, and *social-economic*. This section derives the claim bottleneck by bottleneck, then states two formal bounds that make the grounding framework measurable.

3.1. Bottleneck-by-bottleneck derivation

The data wall is conventionally framed as a token-exhaustion problem (Villalobos et al., 2024). We argue the binding constraint is not token count but *verifiable signal*. Self-generated text scales indefinitely but loses anchoring to reality, with progressive distributional narrowing (Gerstgrasser et al., 2024; Shumailov et al., 2024). What is exhausted is not text; what is exhausted is *external arbitration that text is true*. The abstraction barrier presents the same deficit in representational form: a perfect imitator of human-labelled text inherits the human ontology and, absent strong inductive bias, will not exceed it (Chollet, 2019; Ortega et al., 2021). Crossing the barrier requires evidence above linguistic supervision that a candidate concept carves the world at a real joint — exactly what interaction with real distributions can supply and pure language modelling cannot. The embodied bottleneck is analogous: recursive self-improvement is upper-bounded by the slowest physical loop required to validate a candidate improvement (Genewein et al., 2026; Lawrence, 2024). The bound is not on compute or data, but on the latency and bandwidth of *physical verification*. Finally, multi-agent trust: virtual economies cannot clear without verifiable signals about identity, capability, inventory, price, and outcome (Drexler, 2019; Tomašev et al., 2025). Without such signals every transaction degenerates into an information-asymmetric "market for lemons" (Akerlof, 1970) or into a hallucination-cascade among agents (Ngo et al., 2022). The bottleneck in each case is the supply of *trust substrate*.

3.2. Four modalities, one common deficit

- **Data grounding** = verifiable interaction signal.
- **Representational grounding** = evidence that a candidate concept is real.
- **Embodied grounding** = physical loops short enough to validate change.
- **Social-economic grounding** = trust signals that enable agent transactions.

Compute scaling amplifies each modality but supplies none. This asymmetry is the core of our position: *policies that buy more grounding will dominate policies that buy more compute alone*, in regimes where grounding is the binding constraint. The reframing is not a redescription of well-known desiderata; it makes four things tractable that were previously implicit: (i) the choice between simulation, self-generation, and real environments becomes empirically arbitrable; (ii) a single property — privileged grounding substrate — identifies what any "post-AGI data source" must possess (Section 4); (iii) grounding (environment) is separated from inductive bias (model), enabling proper attribution; and (iv) two quantitative levers — δ and ε — make the debate measurable.

3.3. The Determinism–Efficiency Bound

The intuition that environmental non-determinism degrades agent learning can be formalised under explicit assumptions.

Box 2 | Assumptions of Proposition 1.

(A1) Per-step success probability δ. Determinism quality $\delta(\mathcal{E}) \in [0, 1]$ is the per-step success probability of the environment against ground truth: the probability that, under a declared intent-canonicalisation protocol that normalises a well-formed query q, the environment's response yields an outcome the independent verification channel of Definition 1 (D3) judges *correct* against the relevant ground-truth state (inventory, price, certification, settlement). Operationally, δ is estimated by sampling canonicalised queries and reporting the empirical correct-fraction. The intent-canonicalisation protocol is a prerequisite for cross-platform comparability and is itself a domain artefact (see Remark 1).

(A2) Independent per-step verification. Steps in a k-step chain are verified independently by the environment; success of any single step is a Bernoulli(δ) event with success defined as in (A1).

(A3) No retry budget. The bound concerns first-attempt chain success; retry-augmented agents are treated in Remark 2 below.

(A4) Bounded recovery cost. Recovery from a failed step costs strictly positive time; agents that ignore failed verification verdicts are outside scope.

Assumption (A2) is the key simplification; correlated-step chains are treated separately in Lemma 1.

Remark 1 (Consistency versus correctness)**.** *δ measures* correctness *against an independent ground-truth channel, not self-consistency under replay. An environment that returns identical but incorrect responses to repeated queries has high replay consistency yet $\delta = 0$; conversely, a well-calibrated stochastic environment may have low replay consistency but high δ on a per-query basis. Empirical δ-measurement protocols must therefore use the (D3) verification channel rather than self-replication, and report the canonicalisation rule applied to the intent space (otherwise inter-platform δ estimates are not comparable).*

Proposition 1 (Determinism–Efficiency Bound)**.** *Under assumptions (A1)–(A4), let $\mathcal{E}$ be an environment with determinism quality $\delta \equiv \delta(\mathcal{E})$. For a learner with sample budget n, the effective sample size is $n_{\text{eff}} = \delta \cdot n$. Consequently:*

1. *For any target capability threshold θ, the sample complexity satisfies $n(\theta, \mathcal{E}) = \Omega(1/\delta)$.*
2. *Chain-task success probability for a k-step task under independent per-step verification degrades as $P_{\text{success}}(k) \leq \delta^k$.*

Proof sketch. For claim (1): under (A1) each sample is independently correct with probability δ against ground truth, so contributes useful gradient information with that probability; incorrect samples must be filtered or discarded, yielding $\Omega(1/\delta)$ sample-complexity overhead. For claim (2): under (A2), k sequential steps are independent Bernoulli(δ) trials, so the joint success probability is δ^k exactly. The qualification "$\leq$" rather than "=" covers correlated and adversarial step structures (Lemma 1). □

Novelty and implications. The arithmetic δ^k is standard independent-Bernoulli composition. What is novel is neither the math nor the qualitative intuition (every reliability engineer knows chain failures compound). What we propose is a *measurement programme*: that δ should be instrumented, published, and tracked at the environment level the way uptime is tracked at the service level, with the specific operational consequences (the DMM, the SCI, the investment-thesis frame) following from that instrumentation rather than from any new theorem. Even moderate non-determinism ($\delta = 0.9$) produces catastrophic failure rates for long chains ($\delta^{10} \approx 0.35$). Investing in environment determinism has *exponential* returns for chain-task success, whereas model robustness yields at best linear improvements. This asymmetry is the core argument for prioritising environment redesign over model hardening.

Relationship to SRE and reproducible-build traditions. δ differs from conventional service-level objectives (SLOs) in site reliability engineering along three axes: (1) SLOs measure per-request *availability* (uptime, latency percentiles); δ measures *semantic correctness* against ground truth. A service may have 99.99% uptime (SLO met) yet personalise responses on every call ($\delta \ll 1$). (2) SLOs are per-endpoint; δ compounds multiplicatively across a k-step chain, making the exponential degradation visible only at the workflow level. (3) SLOs target human-tolerance thresholds; δ targets algorithmic-consumption requirements where even small per-step noise is catastrophic at chain length. δ is therefore a *chain-aware, semantic, agent-centric* reliability concept that existing SRE frameworks do not capture. There is also a longer software-engineering lineage worth acknowledging: the reproducible-build tradition (hermetic builds, content-addressed storage, deterministic compilation; cf. Nix (Dolstra et al., 2004)) has spent two decades developing exactly the determinism vocabulary the present paper imports for agent environments. Our contribution is not to invent determinism as an engineering goal but to specialise it — from *byte-equivalent build outputs* to *semantically stable agentic responses at chain length*.

Lemma 1 (Correlation reshapes but does not erase exponential degradation)**.** *Suppose the k-chain steps share latent session state s such that* Pr[step i *succeeds* $\mid s$] $= \delta_i(s)$ *and,* conditional on s, step-success events are independent. *Let $\bar{\delta}_i := \mathbb{E}_s[\delta_i(s)]$ denote the marginal per-step determinism. Then chain-task success is*

$$P_{\text{success}}(k) \;=\; \mathbb{E}_s\left[\prod_{i=1}^{k} \delta_i(s)\right].$$

This quantity is reshaped by the correlation structure of $\{\delta_i(s)\}_{i=1}^{k}$ across sessions:

1. Negative correlation across steps *(compensating fluctuations) gives, via the FKG/Chebyshev sum inequality,* $P_{\text{success}}(k) \le \prod_i \bar{\delta}_i$ *— a product strictly tighter than treating each marginal in isolation.*
2. Independence across sessions *gives* $P_{\text{success}}(k) = \prod_i \bar{\delta}_i$*, which reduces to* $\bar{\delta}^k$ *only when the marginals are uniform.*
3. Positive correlation *(failures cluster by session — a good session carries many steps) can yield* $P_{\text{success}}(k) > \prod_i \bar{\delta}_i$*. In the limit,* $P_{\text{success}}(k) \to \Pr[\min_i \delta_i(s) = 1]$ *as* $k \to \infty$*: the chain succeeds asymptotically only on the measure of deterministic-good sessions.*

In all three regimes, $P_{\text{success}}(k) \to 0$ *exponentially in* k *whenever* $\Pr[\min_i \delta_i(s) = 1] = 0$ *— i.e. whenever the environment admits no positive-measure deterministic sub-environment. The position's qualitative claim — chain-task success collapses for any environment lacking a deterministic sub-environment of positive measure — is therefore robust to the correlation structure of the steps. The escape clause (*$\Pr[\min_i \delta_i(s) = 1] > 0$*) is exactly the regime that D3 and D4 platforms (Section 7.7) aspire to engineer.*

Remark 2 (Retries do not erase the bound)**.** *A retry-augmented agent with* r *retries per step achieves per-step success* $1 - (1 - \delta)^r$*, yielding chain-success* $[1 - (1 - \delta)^r]^k$*. The per-step retry budget* r *is bounded by latency, cost, and rate-limit constraints; for the operator-class deployments cited in Section 1 this budget is small. When the retry budget is* shared *across the chain (a total budget* B *distributed over* k *steps), per-step* r *scales as* B/k *and the chain-success expression worsens with* k *rather than merely tracking it: the exponential shape is reinforced by realistic budget constraints. (Throughout the paper,* B *denotes the total retry budget and* r *the per-step budget; we use* B *rather than* R *to avoid clashing with the bounded reward* R *of Proposition 2.)*

3.4. The Verifier–Goodharting Floor

Proposition 1 treats verifier signals as ground truth. In practice, every verifier is a learned or rule-based proxy with irreducible error (Casper et al., 2023; Gao et al., 2023; Lambert et al., 2024); optimising hard against the proxy Goodharts the true objective (Goodhart, 1975).

Box 3 | Assumptions of Proposition 2.

(B1) Bounded verifier KL. The verifier-induced outcome distribution V has bounded KL divergence from the true outcome distribution $V^\star$: $D_{\text{KL}}(V \,\|\, V^\star) \le \varepsilon$.

(B2) Optimisation against V**.** The training procedure optimises a bounded reward R under expectation in V exactly, not under expectation in $V^\star$.

(B3) Bounded reward. R is bounded with $\|R\|_\infty \le C$.

Proposition 2 (Verifier–Goodharting Floor)**.** *Under (B1)–(B3), any policy* π *whose optimisation is shaped by* $\mathbb{E}_V[R]$ *rather than* $\mathbb{E}_{V^\star}[R]$ *satisfies*

$$\left|\mathbb{E}_V[R] - \mathbb{E}_{V^\star}[R]\right| \;\le\; 2C \cdot \text{TV}(V, V^\star) \;\le\; C\sqrt{2\varepsilon},$$

where the first inequality is the standard bounded-function bound on total variation and the second is Pinsker's inequality. Equivalently,

$$\mathbb{E}_{V^\star}[R] \;\ge\; \mathbb{E}_V[R] \;-\; C\sqrt{2\varepsilon}.$$

The bound is in general unimprovable using only (B1); tighter bounds require additional structure on R *or on the verifier mismatch (e.g. Bretagnolle–Huber for two-point distinguishability,* f*-divergence inequalities for smoother* $V/V^\star$ *relations).*

Note on smoothness regime. Earlier versions of this work used a Lipschitz hypothesis on R together with Pinsker's inequality; this is a topological mismatch (Lipschitz-smoothness controls the Wasserstein-1 / Kantorovich–Rubinstein gap, while Pinsker controls total variation). The bounded-reward formulation above is the correct pairing for the KL hypothesis (B1).

The flywheel asymptote of Section 6 is bounded by this floor: the policy gain from one more order of magnitude of verifier-gated training is dominated by ε, not by nominal scale. This is the regime where adaptive, task-conditioned verifiers (Ding, 2026a) and weak-to-strong supervision (Burns et al., 2024) pay back: not by making any single verifier perfect, but by lowering ε faster than the model overfits the proxy.

4. Privileged Grounding Substrates: Five Sufficiency Conditions

Of the many environment families one might propose to supply grounding — simulators, game worlds, scientific instruments, sandboxes for self-play (Adaptive Agent Team et al., 2023; Silver et al., 2017), web-scale interaction loggers — we ask: *which families are structurally sufficient?* Rather than arguing from a single domain, we identify five conditions that jointly characterise a *privileged grounding environment*. Any concrete environment satisfying all five qualifies; failure on any one disqualifies.

Box 4 | Five sufficiency conditions for a privileged grounding environment.

(i) Economic self-sufficiency. The environment carries its own reward stream (transactions, fulfilment, returns, repeat engagement) and its own compute budget (real user behaviour). It does not require subsidy from the AGI programme to remain alive.

(ii) Verifiability. Outcomes are checkable against ground truth through an independent channel — settlement, physical delivery, certification, inspection — ideally suited to learned verifiers and reward models.

(iii) Interactivity. The environment supports both a high-frequency *digital* loop and a

Bottleneck	Grounding modality	Supply property	Mechanism (privilege argument)
Data wall	Data grounding	Thick	Verifiable interaction signal and transaction rewards expand coverage without model-collapse anchoring loss.
Abstraction barrier	Representational grounding	Understandable	Multimodal concept discovery over real distributions surfaces concepts beyond curated text labels.
Embodied bottleneck	Embodied grounding	Customisable	Customer-to-manufacturer loops shorten the physical-validation latency bounding recursive self-improvement.
Multi-agent trust	Social-economic grounding	Trustworthy	Decision-grade fields (inventory, qualifications, settlement) form the agent-to-agent trust substrate.
Matching structure	Social-economic grounding	Comparable	Same-item and sourcing networks structure the supply network into multi-agent scheduling.

Table 1 | Bottleneck → grounding modality → supply-certainty property. The five rows form the conceptual backbone shared by the rest of the paper; Figure 1 renders the mapping graphically and Section 5 operationalises each property.

low-latency, shortenable *physical* loop. Information and embodied grounding are co-present.

(iv) Scale and multimodal richness. The state space is massive, long-tail, multimodal, and its ontology has been only partially exhausted by human catalogs.

(v) Deterministic interface guarantee. The environment provides stable, faithfully-ranked, bounded-latency responses suitable for algorithmic consumption (Definition 1).

Illustrative domains. Environments satisfying (i)–(v) include, but are not limited to: industrial B2B sourcing platforms, B2C retail with fulfilment verification, pharmaceutical supply chains with batch traceability, agricultural commodity exchanges with quality grading, automotive supplier networks with manufacturability validation, and healthcare procurement with credentialing. Financial exchanges satisfy (i), (ii), (v) but lack multimodal richness (iv); scientific instruments satisfy (ii) but rarely (i). The intersection is domain-diagnostic, not domain-specific.

Honesty about evidentiary base. We acknowledge that the strongest concrete instantiation of the SCI / DMM machinery available to us is sourcing-class commerce, and that the other domains above are listed as *predicted* qualifiers rather than *measured* ones. The worked SCI example in Section 5.7 exercises three domains (B2B sourcing, financial settlement, B2C retail) to demonstrate that the formal apparatus accommodates domain-relevance masking, but a fully evidenced pharmaceutical or healthcare instantiation is future work. The position is therefore strongest as written for sourcing-class environments and progressively weaker as one moves to domains where the verifier-channel infrastructure is less mature; readers in those domains should treat the position as a hypothesis to test rather than a result to apply.

4.1. Economic self-sufficiency

The dominant grounding alternative is large-scale simulation. Simulation has succeeded spectacularly in narrow domains — Go, chess, single-player games (Schrittwieser et al., 2020; Silver et al., 2017) — but its scaling cost is borne by the AGI programme itself. An environment satisfying condition (i) generates its reward stream as a by-product of its own economic activity: buyers transact, sellers fulfil, disputes settle — all without annotation budget. The economic sustainability literature has begun to note this asymmetry (Acemoglu and Johnson, 2023; Agrawal et al., 2025; Brynjolfsson, 2022; Erdil et al., 2025), but its grounding implications have received little attention.

4.2. Verifiability

Verifiers are the rate-limiting layer in post-training (Bai et al., 2022; Lambert et al., 2024; Ouyang et al., 2022). A verifier is only as useful as the ground truth it can compare against. Environments satisfying (ii) provide ground truths — a payment cleared, a parcel delivered, a credential valid, a manufactured part to-spec — that are observable, adversarially robust at population scale, and not subject to the preference-elicitation artefacts of single-turn human feedback. We argue in Section 6 that this property gives the data flywheel its robustness against Model Collapse (Shumailov et al., 2024), modulo Proposition 2.

4.3. Interactivity (digital and physical)

Most digital environments support cheap, high-frequency interaction but no physical loop; most physical environments support a physical loop but no cheap digital loop. An environment satisfying (iii)

supports both: a digital loop (e.g. spot-matching, inquiry routing) and a physical loop (e.g. Customer-to-Manufacturer (C2M) manufacturing, clinical trial feedback). Where embodied bottlenecks are real (Lawrence, 2024), the physical loop's *latency* is the binding term; qualifying environments contain explicit engineering levers to shorten it (manufacturability constraints, capacity matching, rapid prototyping), which simulation does not.

4.4. Scale and multimodal richness

Web-scale text covers what people *say* about things. Catalogs cover what producers *declare*. Neither covers what the things *are*. The full multimodal record of a single industrial category spans engineering drawings, CAD geometry, materials data, process descriptions, and end-user specifications — a space the public web only thinly samples. Any environment satisfying (iv) contains this kind of unexhausted multimodal distribution, creating room for concept discovery beyond the human ontology and thereby challenging the abstraction barrier.

4.5. Deterministic interface guarantee

Condition (v) reflects Proposition 1. An environment may satisfy (i)–(iv) but still be hostile to agents if its interfaces are designed for human browsing: shuffled search results, session-dependent pricing, variable API latency. The deterministic interface guarantee demands a programmatic access layer satisfying (D1)–(D4). The distinction is concrete: a consumer-facing search (personalised, diversified, attention-optimised) versus a programmatic-first API (stable rankings, bounded latency, verifiable state). The former is optimised for engagement; the latter for algorithmic consumption.

4.6. Sim-to-real, fairly considered

The most credible alternative is "simulate, with domain randomisation, until the model generalises" (Bruce et al., 2024; Hafner et al., 2020; Lee et al., 2020; OpenAI et al., 2019, 2020; Peng et al., 2018; Tobin et al., 2017). This programme has won in several domains, with high-evidence cases including OpenAI's dexterous in-hand manipulation (OpenAI et al., 2019, 2020), quadrupedal locomotion over challenging terrain (Lee et al., 2020), and autonomous-driving sensor stacks with photorealistic simulation. We do not dismiss this evidence; we bound it.

Sim-to-real nonetheless stays outside the sufficient set even where it succeeds locally, for three structural reasons. The first is an ontology cap: simulation can only generate states its physics engine and asset library anticipate, whereas real environments satisfying (i)–(v) routinely produce states no designer foresaw (novel compositions, unexpected failure modes, emergent demand patterns, adversarial counterparty behaviour). The second is a verifier ground-truth shift: a sim-to-real verifier uses the simulator's physics as ground truth and only discovers the gap on costly real trials, while an environment satisfying (ii) uses real-world outcomes as ground truth from the first interaction — a stronger verifier-quality regime in Proposition 2's ε. The third concerns reward provenance: simulators force reward to be specified, which is a known source of hacking (Casper et al., 2023; Gao et al., 2023), while qualifying environments let reward be observed (bounded by Goodharting but not by specification error). Simulation therefore belongs in the three-way comparison of OQ1 as a complement, not a substitute.

4.7. Why these five together matter

Several alternative environments meet a subset of (i)–(v). Web crawl is self-sustaining and rich but only weakly verifiable and not deterministic (Soldaini et al., 2024); scientific instruments are

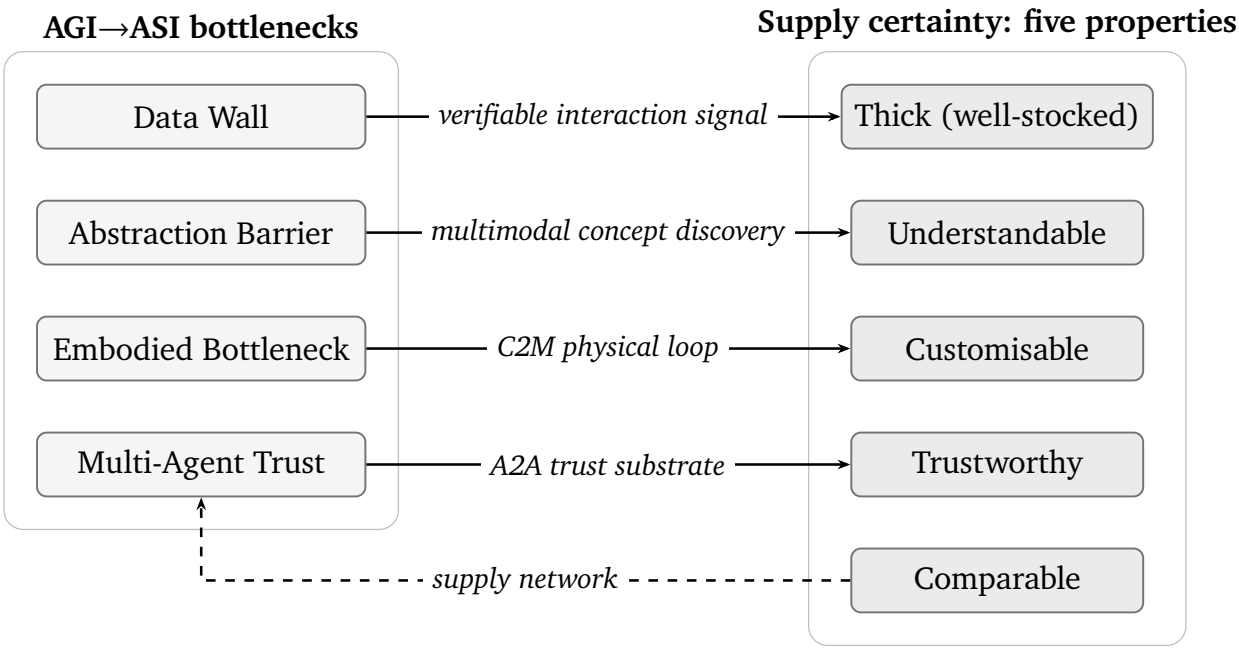

Figure 1 | Grounding mappings. Four AGI→ASI bottlenecks (left) are each addressed by one supply-certainty property (right). The closing dashed edge captures that comparability structures the supply network back into a multi-agent matching substrate.

highly verifiable but not self-sustaining; robotics is interactive but neither self-sustaining nor scalably multimodal (Adaptive Agent Team et al., 2023); financial exchanges are deterministic and verifiable but not multimodally rich. The intersection is rare. We do not claim it is unique to any single industry; we claim that economically self-sustaining commercial environments — spanning sourcing, retail, pharma, agriculture, automotive — are the most readily available family that demonstrably satisfies all five at industrial scale today. This is an empirical observation, not a definitional restriction: any future environment meeting (i)–(v) automatically enters the privileged set.

5. Operationalising Supply Certainty: Five Properties and the SCI

The position of Section 4 is empty without an operationalisation. We decompose Supply Certainty into five properties that are each (i) defined qualitatively in the language of agentic consumers, (ii) admit a generic measurable proxy, and (iii) map onto exactly one of the four grounding modalities (with one closing property serving the multi-agent structure). Table 1 summarises the mapping; this section unfolds it and aggregates the five into a single platform-comparable index — the *Supply Certainty Index* (SCI).

We describe the mappings in the order of Figure 1.

5.1. Thick (well-stocked) supply ↔ data wall

A supply environment is *thick* for an agent if, conditional on a demand intent, there exists with high probability a non-degenerate set of supply candidates that satisfy the intent at non-degenerate granularity. Thickness directly addresses the Data Wall: every additional unit of supply that the system can correctly cluster, describe, and match adds an *externally-arbitrated* interaction signal — one that resists model collapse because it is anchored in a downstream economic event (Gerstgrasser

et al., 2024; Shumailov et al., 2024). Measurable proxies include coverage of long-tail categories, demand-to-supply translation recall, and net new item rate per unit time.

5.2. Understandable supply ↔ abstraction barrier

A supply environment is *understandable* when an agent can read its product ontology in the same vocabulary the environment is curated in: a unified multimodal semantic base plus a structured Category-Property-Value (CPV) -style representation. Understandability is the operational hook for the Abstraction Barrier. Crossing the barrier requires discovering new conceptual primitives *above* those in the curated human ontology, possible only if the system can compare candidate primitives against the unfiltered multimodal distribution of real goods (Ortega et al., 2021; Srivastava et al., 2022). Measurable proxies include precision/recall of same-item clustering, attribute-governance coverage, and verified knowledge graph size.

5.3. Customisable supply ↔ embodied bottleneck

A supply environment is *customisable* when a buyer-side specification can be translated into a manufacturable bill of materials and routed to a factory whose declared capabilities match. Customisability is the engineering surface on which the Embodied Bottleneck is fought. Each percentage point added to the manufacturability rate and each minute removed from design-to-quote latency shortens the embodied loop. Unlike purely scientific embodied tasks, the physical environment here is already instrumented for short-loop iteration. Measurable proxies include manufacturability rate of generated artefacts, CAD kernel coverage, spec-to-quote latency (inverted), and factory-capacity matching accuracy.

5.4. Trustworthy supply ↔ multi-agent trust

A supply environment is *trustworthy* for an agentic consumer when the supply record exposes a complete set of *decision-grade fields*: inventory, invoicing, qualifications, certificates, price-and-freight, weight-and-dimensions. Trustworthiness is the substrate of Agent-to-Agent (A2A) interaction. A virtual agent economy (Tomašev et al., 2025) cannot clear without verifier-checkable signals on price, capability, and identity; in absence of such signals, the economy degenerates into asymmetric-information failure (Akerlof, 1970) or hallucinated trades. Measurable proxies include decision-grade-field coverage rate, verified-supply set size, independent-judge precision/recall, and field-record freshness.

5.5. Comparable supply ↔ supply network for multi-agent matching

A supply environment is *comparable* when supply items can be clustered by demand-relevant equivalence (same-item) and ranked along demand-relevant axes (price, lead time, service). Comparability is the closing edge of Figure 1. Without it, trustworthiness has nowhere to compose: an agent that can verify a single seller cannot yet schedule a fleet. Comparability is the structural prerequisite for multi-agent matching at scale. Measurable proxies include cluster precision and Cov@k of same-item retrieval, and independent-judge precision/recall.

5.6. The Supply Certainty Index (SCI)

The five properties above are individually measurable but the literature has lacked a single composite that platforms can be ranked on. We propose the *Supply Certainty Index*:

Definition 2 (Supply Certainty Index, SCI)**.** *For a supply environment $\mathcal{E}$ instrumented with the per-property proxies above, the* Supply Certainty Index $\mathrm{SCI}(\mathcal{E}) \in [0,1]$ *is the geometric mean of the five property scores after each property's proxies are aggregated to a* $[0,1]$ *score* S_p*:*

$$\mathrm{SCI}(\mathcal{E}) \;=\; \left(S_{\text{thick}} \cdot S_{\text{und}} \cdot S_{\text{cust}} \cdot S_{\text{trust}} \cdot S_{\text{cmp}}\right)^{1/5}, \qquad S_p \in [0,1]\ \forall p.$$

The deterministic-interface condition of Definition 1 contributes the gating multiplier $\delta(\mathcal{E}) \in [0,1]$ *yielding a deterministic-corrected score*

$$\mathrm{SCI}^{\delta}(\mathcal{E}) \;=\; \delta(\mathcal{E}) \cdot \mathrm{SCI}(\mathcal{E}).$$

Why geometric mean. A platform excellent on four properties and zero on the fifth is not a privileged grounding substrate: the five properties are non-substitutes (each addresses a distinct grounding modality). The geometric mean enforces this: a single zero zeroes the index. The δ multiplier further ensures that a platform with good supply properties *but a stochastic agent interface* is not credited for what agents cannot reliably consume.

Why $[0,1]$ scores. Absolute units vary by domain by orders of magnitude. Each S_p is the platform's measurement against a domain-relative reference panel — analogous to scaling-law normalisation against compute budget.

Numerical convention for $S_p = 0$. The geometric mean is undefined in log-space when any $S_p = 0$ (its operational value is SCI = 0, but log-space computation gives $-\infty$). For numerical implementations we adopt the convention that any score below an ε-floor of 0.01 is clamped to $\varepsilon = 0.01$ before aggregation, with the unclamped score reported separately. This is a measurement convention, not a substantive allowance: the operational meaning of $S_p \le \varepsilon$ is "platform fails this property in the relevant sense" and the SCI correctly reports a near-zero composite.

Domain relevance and N/A masking. The geometric mean penalises any zero score severely. This is intentional within domains where all five properties are relevant. Where a property is *structurally absent* from a domain — e.g. customisability for a purely digital information marketplace with no physical fulfilment, or trustworthiness in the sense of physical credentialing for a pure financial-settlement environment — the property is masked rather than scored as zero:

$$\mathrm{SCI}(\mathcal{E}; D) \;=\; \left(\prod_{p \in P_D} S_p\right)^{1/|P_D|}, \qquad P_D \subseteq \{\text{thick}, \text{und}, \text{cust}, \text{trust}, \text{cmp}\},$$

where P_D is the subset of *applicable* properties for domain D.

Pre-registration discipline (anti-manipulation). To prevent a platform or industry from gaming the index by declaring inconvenient properties N/A, P_D must be fixed by the reference panel *before* any platform measurement, published openly per domain, and not amended on the basis of measurement outcomes (akin to pre-registration in clinical trials). Platforms must report both the masked SCI (over P_D) and the unmasked SCI (over all five properties, with structurally-absent scores set to a stated ε-floor) for transparency; reviewers and downstream consumers can then compare. The N/A declaration is a property of the *domain* and *reference panel maintainer*, not a property of the *platform under evaluation*.

5.7. A worked SCI example

We illustrate the SCI on three hypothetical platforms in distinct domains; numbers are stylised and meant to demonstrate the construction, not to rank any real platform.

Platform	S_{thk}	S_{und}	S_{cust}	S_{trust}	S_{cmp}	δ	SCI	SCI$^{\delta}$
P1: B2B sourcing ($D3$)	0.85	0.70	0.60	0.80	0.75	0.92	0.73	0.67
P2: Financial settlement ($D3$)	0.90	0.65	N/A	0.88	0.80	0.95	0.80	0.76
P3: B2C retail ($D2$)	0.95	0.55	0.30	0.40	0.60	0.72	0.52	0.37

Table 2 | Illustrative SCI computation across three domains. SCI is the geometric mean over the applicable subset P_D; for P2, P_D = {thick, understandable, trustworthy, comparable} excludes customisability under the financial-settlement reference panel. Worked arithmetic: P1 has SCI = $(0.85 \cdot 0.70 \cdot 0.60 \cdot 0.80 \cdot 0.75)^{1/5} = 0.73$ and SCI$^{\delta} = 0.92 \cdot 0.73 = 0.67$; P3 has SCI = 0.52 and SCI$^{\delta} = 0.37$. Pre-registered P_D prevents post-hoc masking; see §5.6.

Two observations: (i) the δ gating multiplier substantially separates $D3$ from $D2$ even at similar raw SCI — P3's SCI 0.52 drops to 0.37 after δ-correction, while P1's SCI 0.73 drops only to 0.67. (ii) P2 scores highest on SCI$^{\delta}$ *within its panel*, but its score is *not* comparable in absolute terms to P1 or P3 because the domain panels differ. Cross-domain absolute comparison is out of scope (Section 10 Limitations); ranking within a domain (across platforms instantiating the same P_D) is the intended operational use.

5.8. Why a five-way decomposition?

We could have proposed three properties or eight. The case for five rests on Figure 1: four properties each carry one grounding modality, and one closing property threads them into a multi-agent structure. Whether the five collapse to fewer factors in practice is an empirical question (OQ2 in Section 8). We do not prescribe universal thresholds; such thresholds are domain-dependent. The operational question is whether relative improvements in each property translate into measurable improvements in the corresponding grounding modality, and whether the SCI composite predicts agent outcomes — the empirical content of OQ2.

6. Verifiable Supply Data Flywheels as Post-AGI Scaling Resource

The previous section described a static property: supply certainty. This section describes its dynamics. We claim that real, verifier-gated interaction data forms a Data Flywheel that is *sustainable* in the post-data-wall regime, in a sense recursive self-generation is not (Dohmatob et al., 2024; Gerstgrasser et al., 2024; Shumailov et al., 2024).

6.1. Evaluation as the loss function

In deployed agentic systems the evaluation harness functions as the training loss: bad cases, retries, low verifier scores, and tool failures all back-propagate to upstream stages, so the data pipeline behaves like a single forward pass with evaluation-shaped signal as its gradient.

6.2. Dual flywheel architecture

Two loops coexist (Figure 2):

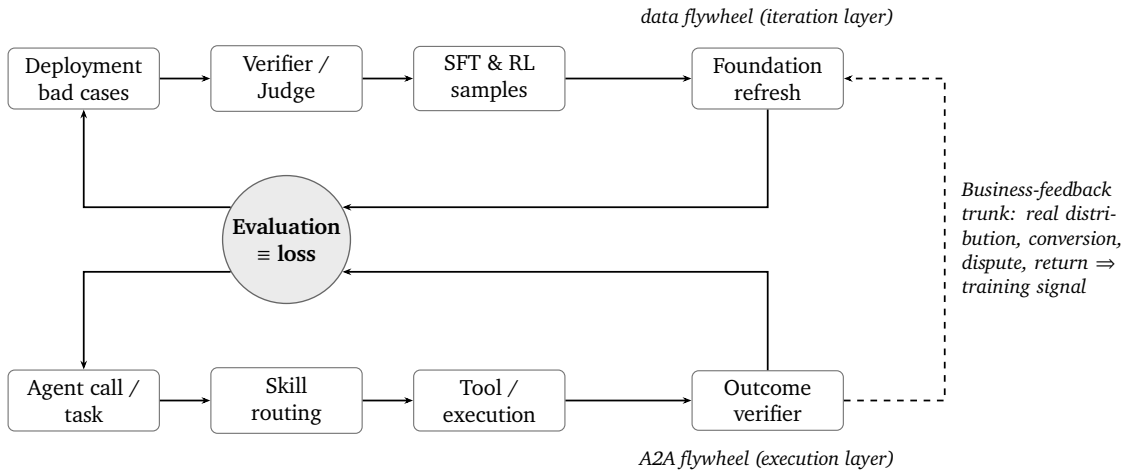

Figure 2 | Dual flywheel architecture. The data flywheel (top loop) turns deployment bad cases into refreshed foundation parameters. The A2A flywheel (bottom loop) turns agent calls into verifier-checked outcomes. Both are routed through a single "evaluation as loss" hub.

- *Data flywheel* (iteration layer): deployment bad cases → verifier/judge → SFT and RL training samples → foundation refresh.
- *A2A flywheel* (execution layer): agent call → skill routing → tool execution → outcome verifier.

The two loops share the evaluation hub; the *business-feedback trunk* connecting outcomes to foundation refresh converts a one-way deployment system into a closed loop.

6.3. Self-evolving environments: the complementary scaling axis

A complementary direction has emerged in the self-evolving-agents literature (Gao et al., 2025; Shinn et al., 2023; Wang et al., 2024a): instead of (or in addition to) adapting the model to a fixed environment, the *environment itself* — tools, skills, memory, runtime context — adapts to deployed trajectories.

The agentic-scaling trio. Three orthogonal scaling axes:

1. *Model adapts to environment*: SFT, RLHF, and RLVR refresh foundation weights (Bai et al., 2022; Ouyang et al., 2022; Wang et al., 2024b).
2. *Environment adapts to model*: skill libraries, tool graphs, and persistent context evolve from execution trajectories (Feng et al., 2026; Sumers et al., 2024; Wang et al., 2024a; Yang et al., 2026; Zhang et al., 2025b).
3. *Co-evolution*: model and environment update jointly under a shared objective (Fang et al., 2025; Li et al., 2026; Wang et al., 2025; Zhang et al., 2025a).

6.4. The Grounded Self-Evolution Convergence Condition

We synthesise the mechanisms above into a formal condition connecting environment-side self-evolution to the determinism framework.

Box 5 | Grounded Self-Evolution Principle (qualitative).

Environment-side evolution (skills, tools, memory, routing tables) accumulates capability *monotonically* if and only if the verification signal exceeds a determinism threshold $\delta > \delta_{\min}$ and a verifier-quality threshold $\varepsilon < \varepsilon_{\max}$. Below either threshold, drift dominates improvement.

Proposition 3 (Grounded Self-Evolution Convergence)**.** *Let $\Phi_t \in [0,1]$ denote a scalar quality measure of an environment-side skill library at evolution step t, with bounded per-step update $\Delta_t := \Phi_{t+1} - \Phi_t$ satisfying $|\Delta_t| \le c$ for $c > 0$. Suppose candidate updates are drawn from a generation distribution $\mathcal{G}$ and gated by a verifier V whose behaviour is characterised by:*

- *Reliability (replay consistency): $\delta \in [0,1]$, the probability that V returns the same verdict on two i.i.d. trials of the same candidate.*
- *Validity (true-positive rate against ground truth): $1-\varepsilon$, the probability that V accepts a candidate that is genuinely an improvement under $\mathcal{G}$.*

(The δ here is the verifier's reliability axis, not the environment-level correctness δ of Proposition 1; the two coincide when the verifier under (D3) is the environment's own ground-truth channel.) Let

$$\mu := \mathbb{E}_{\mathcal{G}}\big[\Delta_t \,\big|\, \textit{accept; truly an improvement}\big] > 0, \qquad d := \mathbb{E}_{\mathcal{G}}\big[-\Delta_t \,\big|\, \textit{accept; false positive}\big] \ge 0,$$

both conditional expectations under $\mathcal{G}$. Then the expected per-step drift satisfies

$$\mathbb{E}[\Delta_t] \;\ge\; \delta(1-\varepsilon)\mu \;-\; \delta\varepsilon d \;-\; (1-\delta)\cdot c,$$

and a sufficient condition for monotone improvement ($\mathbb{E}[\Delta_t] > 0$) is

$$\delta \;>\; \frac{c}{c + (1-\varepsilon)\mu - \varepsilon d} \qquad \textit{whenever} \qquad (1-\varepsilon)\mu > \varepsilon d.$$

Proof sketch. The verifier's behaviour on each candidate decomposes non-orthogonally along two axes: *reliability* (does V return a consistent verdict on replay?) and *validity* (does the verdict match ground truth on a genuine improvement?). Conditioning on the reliability axis, with probability δ the verifier is consistent — in which case the validity axis yields either a correct-accept (probability $1-\varepsilon$, contribution $+\mu$) or a false-positive accept (probability ε, contribution $-d$). With probability $1-\delta$ the verifier returns an inconsistent verdict, in which case the update is treated as adding worst-case noise of magnitude up to c. Aggregating gives the displayed lower bound after replacing $\mathbb{E}[|\Delta_t|]$ by its worst-case upper bound c (a deliberate looseness; tighter bounds are possible if $\mathcal{G}$ is constrained). Solving the lower bound for positivity yields the sufficient condition. □

Note on tightness. The sufficient condition is loose by construction: replacing $\mathbb{E}[|\Delta_t|]$ with the worst-case c in the third term overstates the noise contribution. Empirical operating regimes (Section 8, OQ5) should therefore find positive expected drift at δ values *below* the sufficient-condition threshold. The threshold is a conservative guarantee, not a tight characterisation.

Implication. $\delta_{\min}$ and $\varepsilon_{\max}$ are the thresholds at which the lower bound crosses zero. A skill library updated under sub-threshold conditions random-walks; above-threshold, it accumulates. This unifies findings from Yang et al. (2026) (1–4 accepted edits move performance), Wang et al. (2024a) (skills improve only under deterministic game logic), and Feng et al. (2026) (tool-graph quality requires dense reward).

6.5. Three mechanistic categories of environment-side evolution

The diverse mechanisms unify into three categories:

- *Trajectory recycling*: verifier-checked trajectories are reused as supervision. SAGE trains skill creation via skill-augmented GRPO (Wang et al., 2025); ARISE evolves intrinsic skills under hierarchical RL (Li et al., 2026); WebEvolver co-evolves a world model with a web agent (Fang et al., 2025); Voyager curates a skill library from interaction traces (Wang et al., 2024a); Reflexion uses verbal reinforcement to recycle failed attempts (Shinn et al., 2023). Even *failed* trajectories carry signal: AgentHER (Ding, 2026b) demonstrates that hindsight relabeling of unsuccessful attempts produces high-quality training data, turning every failure into a training asset.
- *Skill-as-text optimisation*: skills are represented as editable text artefacts optimised by a meta-learner. SkillOpt (Yang et al., 2026) shows 1–4 accepted edits produce transferable improvements; recursive summarisation of execution history (Wang et al., 2023) compresses long-horizon experience into reusable skill primitives.
- *Tool-graph consolidation*: execution traces are compiled into typed dependency graphs. SEARL (Feng et al., 2026) jointly optimises policy and graph under dense reward. The cognitive-architectures view of Sumers et al. (2024) situates this inside a broader memory–planning–action framework.

In each case, the trajectory is the gradient — the same logic our *evaluation as loss* principle articulates, generalised beyond model weights to the entire agent stack.

Connection to determinism. When the environment satisfies (D1)–(D4), the verification signal used to gate skill updates is reliable, and Proposition 3's precondition is met. This is the formal bridge between the "deterministic agentic AI" framing and the self-evolving-agents literature. A privileged grounding environment is valuable not only because real-supply trajectories train better models, but because environment-side components — skill libraries, decision-grade fields, A2A routing tables — can themselves accumulate those trajectories. The data flywheel is the model-side realisation; the skill flywheel is its environment-side counterpart. Recursive self-improvement at the environment layer inherits the safety concerns of model-side recursive self-improvement (Anthropic, 2025; Davidson et al., 2026; Zhang et al., 2025a); the Verifier–Goodharting Floor applies equally, since a skill library updating against a high-ε verifier will Goodhart in the same way a model trained on its own outputs would.

6.6. Why this resists collapse, and where it relocates collapse

The model-collapse literature is split: Shumailov et al. (2024) and Dohmatob et al. (2024) show that purely self-generated data streams degrade representation diversity, while Gerstgrasser et al. (2024) demonstrate that *accumulation* of real with synthetic data avoids collapse. Our flywheel claim aligns with the latter: the SFT distribution is anchored not in the model's generative distribution but in *verified-true* cases drawn from real-world outcome signal — an instance of verifier-gated distillation, related in spirit to RLAIF (Bai et al., 2022; Lee et al., 2023) and RLVR with task-specific verifiers (Dubey et al., 2024; Kim et al., 2024; Wang et al., 2024b).

This collapse-resistance is not unbounded. By Proposition 2, a verifier with KL gap ε from ground truth lower-bounds the asymptotic policy reward by $\mathbb{E}_V[R] - C\sqrt{2\varepsilon}$ for a C-bounded reward. Verifier-gated distillation therefore does not eliminate collapse so much as *relocate* it: in the long-horizon limit the SFT distribution converges towards the verifier's mode, which is bounded away from ground truth by ε in KL. The position trades the Shumailov attractor (self-imitation) for the Goodhart attractor

(verifier-imitation). The advantage is that the Goodhart attractor is bounded by an *external* quantity (ε, empirically measurable on a held-out adversarial panel) rather than the model's internal generative entropy.

6.7. Stability conditions and verifier-quality ceiling

The flywheel is only as stable as its verifier, sample selection, and catastrophic-forgetting defence (Dohmatob et al., 2024; Kirkpatrick et al., 2017). Recent work addresses each: instruction backtranslation (Li et al., 2023), weak-to-strong generalisation (Burns et al., 2024), and judge-quality benchmarks (Lambert et al., 2024). They collectively pin down the operating envelope inside which the flywheel accumulates signal.

One path to lowering ε in practice is task-adaptive verifiers that generate assessment criteria conditioned on the specific workflow (Ding, 2026a); such approaches directly reduce out-of-distribution error and raise the flywheel's quality ceiling. Where this matters empirically is OQ5 of Section 8.

7. Supply Certainty as Trust Substrate for Agent Economies

The fourth grounding modality is social-economic. We treat it last because its argument depends on the operationalisations of Sections 5–6.

7.1. The market-clearing problem for virtual agent economies

Several recent position papers anticipate *virtual agent economies* in which discovery, negotiation, and settlement are conducted between agents (Drexler, 2019; Tomašev et al., 2025). The infrastructure conversation has focused on protocol design: tool-calling APIs (Schick et al., 2023), model-context protocols (Anthropic, 2024b), agent-to-agent schedulers (Google, 2025), and payment-side agent toolkits (Stripe, 2025). Less discussed is the *informational* pre-condition for any such market to clear.

Claim. A virtual agent economy cannot clear without verifiable supply-side signals about identity, capability, inventory, price, and outcome. Where such signals are missing, the economy degenerates either into Akerlof asymmetric-information failure (Akerlof, 1970) or into a hallucination cascade (Ngo et al., 2022; Weidinger et al., 2022).

7.2. Multi-agent scaling laws, conditioned on supply trust and determinism

The standard multi-agent scaling discussion treats matching quality as a function of agent population and interaction density (Hong et al., 2024; Leibo et al., 2019; Liu et al., 2025; Panait and Luke, 2005; Park et al., 2023; Wu et al., 2023). We propose that, in the agent-economy regime, this scaling is gated by *supply trust* and *environment determinism*:

$$Q_{\text{match}} \approx f(N_{\text{agents}}, \rho_{\text{interaction}}, \tau_{\text{supply}}, \delta_{\text{env}}),$$

where τ_{supply} is the decision-grade-field coverage rate, δ_{env} is the determinism quality of Definition 1, and the SCI of Section 5.6 is a composite across the five properties. Whether τ and δ enter f as multipliers, thresholds, or interact non-linearly is OQ4 (Section 8).

7.3. Failure modes of insufficient trust

The failure-mode taxonomy below parallels and specialises the miscoordination / conflict / collusion taxonomy of Hammond et al. (2025) to the supply-trust substrate. In the under-trusted regime, three failure modes recur:

1. *Settlement abortion.* An agent cannot commit because the price, inventory, or qualification record is stale or absent.
2. *Epistemic hijacking.* An adversarial supplier exploits the asymmetry between linguistic richness and verifier granularity (Wei et al., 2023; Zou et al., 2023).
3. *Hallucination cascade.* An agent fabricates a supply attribute; downstream agents trust it; the fabrication enters the data flywheel.

Each is a direct consequence of a missing decision-grade field; each is tractable as a coverage problem on the trustworthy-supply property.

7.4. Failure modes of insufficient determinism

Non-deterministic environments introduce three additional multi-agent failure modes:

1. *Rerouting cascades.* Non-deterministic responses force rerouting; if many agents reroute simultaneously, the alternative platform's determinism degrades — a cascade failure.
2. *Coordination impossibility.* Multi-agent scheduling requires consistent supply-state views. Personalisation or A/B testing returns different answers to different agents, fragmenting coordination.
3. *Strategy instability.* Algorithmic procurement strategies calibrated on historical data become unreliable when the environment's response distribution shifts non-stationarily.

7.5. Skillification as the trust interface

The skill-registry abstraction (a typed, SLA-monitored, trace-instrumented inventory of callable capabilities) is the right interface contract for agent-economy trust. It exposes verifier-checkable promises about each call and bounds the cost of hallucinated supply assertions (Anthropic, 2024b; Google, 2025). The mechanism by which such a registry evolves — trajectory recycling, skill-as-text optimisation, tool-graph consolidation — is treated in Section 6.3; here we note only that grounding quality bounds both foundation-weight evolution and skill-registry evolution (Proposition 3). The interface contract and the evolution mechanism are complementary: the registry *is* the operational artefact that the §6.3 skill flywheel updates.

7.6. Investment theses for the determinism economy

If the position is correct, capital should concentrate on primitives that produce, package, and price deterministic verifiable signal. We name five categories the position predicts will dominate, in keeping with recent "agent economy" framing (Andreessen Horowitz, 2026; Karpathy, 2025).

Box 6 | Five investment-thesis categories implied by the position.

(I1) Verifier-as-a-Service. Adaptive, task-conditioned verifiers (Ding, 2026a; Lambert et al., 2024) that lower Proposition 2's floor.

(I2) Determinism Infrastructure. Programmatic-first APIs over legacy systems satisfying

(D1)–(D4).

(I3) Agent Commerce Protocols. MCP / A2A / agent-toolkit (Anthropic, 2024b; Google, 2025; OpenAI, 2024; Stripe, 2025) extended with verifiable trust, dispute resolution, and payment-finality guarantees.

(I4) Skill Registry Marketplaces. Typed, SLA-monitored marketplaces for callable agent skills (Feng et al., 2026; Yang et al., 2026) turning the skill flywheel into a tradable asset class.

(I5) Verifiable-Environment Benchmarks. Public benchmarks where the score is actual settlement, fulfilment, or manufacturability — extending Bai et al. (2026); Jimenez et al. (2024); Min et al. (2025); Qi et al. (2026); Xie et al. (2024); Yao et al. (2024) to the multi-domain SCI evaluation panel of Section 8.

What each thesis predicts. (I1) verifier-quality improvement will outpace model-quality improvement; (I2) the markup of deterministic APIs over stochastic ones will widen with agent adoption; (I3) protocols shipping verifier hooks will displace message-passing-only ones; (I4) skill-registry marketplaces will become a separate category from SaaS; (I5) benchmark-design competence will become a distinct investable. Each prediction is independently refutable. We emphasise that each thesis identifies a structurally privileged category, not a winner inside it; we endorse no specific platform and have no financial relationship with any referenced company.

Exclusive predictions: discriminating between positions. A reviewer may object that I1–I5 are categories any "AI infrastructure" frame would predict regardless of the grounding thesis. To make the position's predictions *exclusive*, we pair each thesis with a discriminator against the strongest competing positions of Section 9:

- *vs alignment-suffices* (§9.8): alignment wins predicts I1 (verifier-as-a-Service) alone dominates; *this position predicts I1 and I2 markups co-move*, because better verifiers and deterministic infrastructure are complements not substitutes.
- *vs sim-to-real* (§9.7): sim-to-real predicts I5 dominates (simulator-based benchmarks suffice); *this position predicts I5 settlement-grounded benchmarks outperform sim-only benchmarks on rank-correlation with production agentic success*.
- *vs orchestration-first* (§9.8 variant): orchestration wins predicts I3 (agent commerce protocols with retry / message-passing) without I2 (determinism infra); *this position predicts I3 protocols shipping verifier hooks displace message-passing-only protocols, and that the market gap appears at the* $D2 \to D3$ *transition*.
- *vs AI-as-normal-technology* (§9.9): normal-tech predicts I1–I5 are sociotechnical infrastructure plays with no sharp inflection; *this position predicts a non-linear inflection at* $D2 \to D3$ *where the marginal category value steps rather than tracks*.

These are mutually discriminating; the position loses if the observed investment-market trajectories match a competing position's predictions better than ours.

7.7. The Determinism Maturity Model

Sections 5–7 state what deterministic supply environments do for grounding. This subsection states what platform engineers should build: a five-level adoption ladder paired with a reference architecture.

Box 7 | Determinism Maturity Model (DMM) levels $D0$–$D4$.

$D0$ **— Human-only UI.** Programmatic access is blocked or CAPTCHA-gated. δ unmeasured.

$D1$ **— Minimal API surface.** A programmatic surface exists but inherits human-UI semantics: shuffled rankings, session-conditioned pricing, no SLA. δ typically < 0.5 on long chains.

$D2$ **— SLA-bounded API.** Bounded latency, versioned schemas, uptime SLAs. Rankings may still be personalised; verifier channels absent. $\delta \in [0.6, 0.8]$. The current state of major B2B APIs.

$D3$ **— Stable rankings and verifier channel.** Rankings deterministic for an intent/persona hash within a staleness bound; verifier channel exposes inventory, qualification, and settlement-finality assertions. $\delta > 0.9$ achievable. Crosses Proposition 3's threshold.

$D4$ **— Full deterministic agent interface.** Faithful ranking (no exploration injection under agent persona); production skill-registry contract with per-skill SLA, trace, and verifier telemetry. Published SCI^{δ} per agent persona. The platform's supply becomes a first-class grounding substrate.

Reading the ladder. Levels are cumulative. The marginal cost of climbing one level is dominated by the marginal benefit only above Proposition 3's threshold — typically at $D2 \to D3$, where the verifier channel unlocks the data flywheel. The position predicts that $D2$ platforms will face increasing pressure to climb to $D3$ as *agentic traffic* — automated API calls executing end-to-end workflows of three or more steps without human gating — grows from a fringe share of total traffic to a substantial one. The exact tipping point is platform-specific; the falsifiable claim is that the transition is non-linear in δ and that platforms which stall at $D2$ will see disproportionate chain-task degradation on their own agentic traffic (Proposition 1). The DMM is the substrate-side complement of the agent-infrastructure agenda of Chan et al. (2025): their identity, audit-trail, and credentialing infrastructure describes what agents need to bring *to* the environment; the DMM describes what the environment must expose *to* agents. The two are complementary.

Reference architecture. A $D3/D4$ platform requires three components beyond a conventional API gateway:

Component A — Verifier service. For each platform assertion (inventory, certification, settlement state): (a) the assertion's value at query time; (b) source channel and freshness; (c) an independent verification endpoint returning verdict plus confidence; (d) verdict-history telemetry consumable by downstream agents and the data flywheel. This component contains Proposition 2's ε and should be benchmarked against judge-quality suites (Ding, 2026a; Lambert et al., 2024).

Component B — Skill registry and A2A router. A typed inventory of callable capabilities with schema, SLA, trace instrumentation, and verifier-outcome binding. The operational counterpart of the skill flywheel (Section 6.3).

Component C — Determinism telemetry plane. A measurement plane publishing empirical δ, ε, and SCI^{δ} per agent persona and per skill — contractually observable to platform consumers and closing the measurement loop for Section 8.

Mapping to existing reliability and governance frameworks. The DMM is not a compliance standard, but the primitive it formalises (δ, ε, SCI^{δ}) is precisely what existing reliability and AI-governance regimes ask for under different names. Three useful handshakes:

- *EU AI Act Article 15* requires high-risk AI systems to achieve "appropriate levels of accuracy, robustness and cybersecurity" over their lifecycle and to publish those metrics (European Parliament and Council, 2024). δ (chain-aware semantic consistency) and ε (verifier KL gap) operationalise "accuracy and robustness" for agentic workflows in a way per-endpoint SLOs do not.
- *NIST AI RMF MEASURE function* requires identification of reliability and validity metrics with measurement procedures. The DMM telemetry plane (Component C) is a concrete instantiation: δ per persona, ε per verifier, SCI^{δ} per panel.
- *ISO/IEC 42001 AI Management System* requires auditable operational telemetry. A $D3/D4$ platform's published determinism telemetry plane is audit-ready under this standard in a way a $D2$ platform's per-endpoint SLOs are not.

We do not propose the DMM *as* a compliance standard; we propose it as a substrate primitive that governance frameworks can adopt to operationalise the chain-aware-reliability requirements they already gesture at.

Boundary: what the DMM is not. The DMM is not a compliance standard, procurement requirement, or substitute for domain-specific safety, privacy, or regulatory regimes (Anderljung et al., 2023; Bengio et al., 2024; European Parliament and Council, 2024). It is a structural ladder for the deterministic-grounding axis; orthogonal concerns (privacy, fairness, safety, liability) require their own ladders. We also do not address whether higher DMM levels should be *mandated*; that is a governance question whose answer is independent of the technical one this paper addresses.

8. A Falsifiable Research Agenda

A position becomes science only when stated as falsifiable questions. We close the technical body with five open questions, each paired with a measurement design and an explicit null result. They align with the research clusters of Genewein et al. (2026)[1].

Box 8 | The OQ1–OQ5 open-question programme.

OQ1 — Data wall. Does an economically self-sustaining, verifier-equipped, deterministic real-supply environment measurably outperform pure simulation or pure self-generated data in (a) sample efficiency and (b) time-to-onset of recursive degeneration?

OQ2 — Abstraction barrier. Can models trained on real-supply multimodal distributions discover stable new conceptual primitives beyond the existing CPV ontology? Does the SCI predict the share of post-ontology stable concepts?

OQ3 — Embodied bottleneck. How low can the latency floor of a C2M physical loop be pushed at constant manufacturability quality?

OQ4 — Multi-agent scaling under trust and determinism. How does matching quality

[1] Specifically: quantitative forecasting (OQ1), benchmarking and abstraction (OQ2), multi-agent scaling (OQ4), recursive improvement dynamics (OQ5), governance and embodied dynamics (OQ3).

scale jointly with agent count, supply-trust coverage, and environment determinism? Does SCI^{δ} predict downstream agentic capability across DMM levels $D2 \to D4$?

OQ5 — Verifier ceiling. At what verifier-quality ε does the data flywheel transition from accumulation to degeneration? At what ε does Proposition 3 fail?

8.1. OQ1: data wall experiments

Measurement design. A three-way training comparison: matched-budget runs of (a) verifier-gated real-supply data from a $D3+$ environment, (b) high-fidelity simulation, (c) state-of-the-art recursive self-generation (Li et al., 2023; Yuan et al., 2024). Hold model architecture, total tokens, and verifier architecture constant. Evaluate sample efficiency and recursive-degeneration onset following the protocol of Shumailov et al. (2024).

Null result. If (a) does not measurably beat (b) and (c) on either axis at the matched-budget point, the privileged-grounding hypothesis is weakened.

8.2. OQ2: abstraction-barrier experiments

Measurement design. Train a multimodal model on the real-supply distribution; cluster latent concepts that pass a stability test across random seeds; quantify the share of stable concepts not previously represented in the curated CPV ontology. Replicate against a baseline trained only on human-curated catalogs.

Null result. If the share of post-human-ontology stable concepts is statistically indistinguishable from baseline, the abstraction-barrier claim is weakened.

8.3. OQ3: embodied-bottleneck experiments

Measurement design. Measure end-to-end design-to-first-pass-quote latency in a C2M pipeline under (a) current best-of-class and (b) a pipeline augmented with an AI-generated manufacturability validator. Compare against the analytic bound of Lawrence (2024).

Null result. If the latency floor is invariant to AI investment, the customisability property fails to relax the embodied bottleneck.

8.4. OQ4: multi-agent scaling under supply trust and DMM level

Measurement design. In a real A2A matching market, vary τ_{supply} and δ_{env} along their observable ranges, holding agent population and protocol fixed; measure matching quality and clearing rate. Fit $Q_{\text{match}}(N, \rho, \tau, \delta)$ and the marginal effect of each DMM step.

Null result. If Q_{match} is approximately independent of τ and δ, the trust-substrate and determinism roles are weakened. If SCI^{δ} does not rank-order platforms by agent-task success, the SCI construct is weakened.

8.5. OQ5: verifier-quality ceiling and flywheel degeneration

Measurement design. Estimate verifier irreducible error ε on a held-out adversarial panel. Train a sequence of distilled models with the verifier in the loop; identify the training-generation index at which validation loss increases. Replicate at several ε levels. Measure the R-$R^\star$ gap against Proposition 2's Pinsker bound.

Null result. If the flywheel degenerates at ε levels substantially below production verifiers, the collapse-resistance argument is overstated.

Extension to environment-side evolution. A natural extension: does skill-library quality degrade faster, slower, or at the same rate as model quality as verifier error increases? Proposition 3 predicts the rates differ by μ/d; the experiment measures that factor empirically.

8.6. Towards a non-saturating benchmark

We propose a *verifiable-environment benchmark* in which each capability is scored by actual settlement, fulfilment, or manufacturability outcome, following the trajectory of Chollet (2019); Ho et al. (2025); Jimenez et al. (2024); Liu et al. (2024); Xie et al. (2024); Yao et al. (2024); Zhou et al. (2023). The benchmark is the public-good counterpart of investment thesis (I5).

8.7. A pilot measurement: τ-bench under controlled δ

While the full OQ programme requires sustained research investment, a first-pass test of the δ^k bound is achievable on existing benchmarks. We sketch a lightweight protocol.

Setup. Take τ-bench (Yao et al., 2024), which evaluates multi-step agent workflows against a retail back-end. Instrument the back-end's tool-execution layer with a deterministic-seeded perturbation hook that, with probability $1 - \delta_{\text{inj}}$ per tool call, replaces the ground-truth API response with a semantically plausible incorrect alternative. The *injector* is template-driven (not LLM-generated) to make replications byte-identical: a fixed dictionary mapping each tool to a pool of canonical perturbations (stale inventory → inventory−1 unit, price perturbation → ±5%, attribute swap → swap with a randomly chosen sibling SKU from the same category). The perturbation seed is fixed per (task ID, trial ID, δ_{inj}) triple.

Protocol. Run the benchmark at $\delta_{\text{inj}} \in \{1.0, 0.95, 0.9, 0.8, 0.7\}$ across three agent architectures (direct prompting, ReAct, and a Reflexion-augmented agent (Shinn et al., 2023)). For each (δ_{inj}, agent) cell, record chain-task success rate over $n \geq 200$ episodes, stratified by τ-bench's nominal chain-length k. Estimate the effective chain length k_{eff} per agent architecture by maximum-likelihood fit of the model $P_{\text{success}} = \delta_{\text{inj}}^{k_{\text{eff}}}$ in logit space, with k_{eff} as the single free parameter and 95% confidence intervals via bootstrap over episodes. Report inter-trial reproducibility (variance of success across 3 seed replications per cell) as a robustness diagnostic.

Expected outcome. If the position is correct, the empirical success-rate curve should track $\delta_{\text{inj}}^{k_{\text{eff}}}$ for an effective chain length close to the benchmark's nominal chain length. Reflexion should shift k_{eff} downward but not eliminate the exponential shape (Remark 2).

Null result. If, after the bootstrap-CI analysis, Pr[success] is approximately invariant to δ_{inj} across agents (slope statistically indistinguishable from zero on the logit-transformed curve), the δ^k bound does not bite in practice and the determinism framing is empirically unsupported.

Power analysis. At $n = 200$ per cell and binary outcomes, the standard error on the estimated success rate is bounded by $1/(2\sqrt{200}) \approx 0.035$. The position predicts (at $k_{\text{eff}} = 6$, the median nominal τ-bench chain length) success rates of $\{1.0, 0.74, 0.53, 0.26, 0.12\}$ across the five δ_{inj} levels; adjacent-cell gaps are $\{0.26, 0.21, 0.27, 0.14\}$, each substantially above $2 \times 0.035 = 0.07$. The proposed n is therefore adequate to distinguish the position's prediction from a null at the 95% level.

8.8. Preliminary empirical grounding

The scale of verifiable signal in real economic environments is already substantial. As one anchored data point for *settlement-class* signals: Stripe's public disclosures indicate over a trillion dollars in 2024 payment volume across hundreds of millions of transactions (Stripe, 2024), implying $\sim 10^6$–10^7 verifiable settlement events per day at that single platform. We restrict this anchor to settlement-class signals: Stripe is a payments processor and does not provide direct evidence for sourcing, fulfilment, or quality-inspection throughputs, which are the multimodal signal classes most relevant to the grounding argument. Whether large B2B sourcing platforms reach comparable daily throughput on the broader signal classes (order confirmations, shipment verifications, quality inspections) is an empirical question that OQ1's measurement design is exactly intended to answer.

For comparison: the largest curated RLHF preference datasets contain $O(10^5)$ comparison pairs (Bai et al., 2022; Ouyang et al., 2022); even scaled-up RLVR pipelines produce $O(10^6)$ verification events per training run (Dubey et al., 2024; Wang et al., 2024b). Settlement-class signal alone therefore provides at least a 10× daily-throughput advantage over curated alignment efforts; the multimodal sourcing-class signal advantage is what OQ1 measures.

Signal-quality caveat. Throughput is not the only relevant axis: most settlement signals are binary (cleared or returned) and not labelled at concept-discovery granularity. The position's flywheel claim therefore rests on quality-weighted, not raw, signal volume; OQ1 must control for label density and concept-coverage per signal class, not just count signals.

Recent benchmarks have begun to operationalise this industrial complexity: holistic e-commerce agent evaluation (Min et al., 2025), industrial knowledge boundaries (Bai et al., 2026), multi-image product understanding for industrial catalogues (Qi et al., 2026), and operator-class long-horizon evaluation (Xie et al., 2024; Yao et al., 2024). These confirm that verifiable settlement and fulfilment signals can serve as scalable ground-truth for agent evaluation. While individual platforms are proprietary, the OQ programme can be instantiated on any supply environment meeting conditions (i)–(v) at $D \geq 3$. Multiple independent instantiations are expected.

9. Counterarguments, Competing Positions, and Boundaries

A position is strongest when its boundaries are explicit. We split the discussion into (a) *objections* internal to the position and (b) three *competing positions* held by serious researchers that predict different outcomes.

A. Objections internal to the position

9.1. Domain specificity

Objection. Even granting that commercial supply is a privileged grounding substrate for commerce-adjacent capabilities, does grounding generalise beyond commerce? (Bostrom, 2014; Morris et al., 2024).

Reply. We do not claim universality. The position is a capability-subset claim: grounding-driven dominance for tasks whose verification is naturally economic, physical, or multi-party. Purely linguistic or aesthetic tasks are outside scope. Scientific verification has its own substrates; we view this as future work.

9.2. Adversariality of trustworthy fields

Objection. Decision-grade fields have a long-tail distribution and are subject to adversarial forgery (Biggio and Roli, 2018; Eykholt et al., 2018). Trustworthiness is a moving target.

Reply. We agree. The verifier must be continually hardened; OQ5's ε is the formal quantity adversarial corruption increases. The position weakens in proportion to the rate at which adversaries corrupt decision-grade fields faster than verifier retraining.

9.3. Incentive misalignment and systemic risk

Objection. A strong trust substrate also amplifies incentive misalignment, instrumental convergence, and systemic risk (Hendrycks et al., 2023; Ngo et al., 2022; Russell, 2019). Building the substrate without governance is irresponsible.

Reply. The position is conditional: build the substrate, and governance must keep pace (Anderljung et al., 2023; Bengio et al., 2024; European Parliament and Council, 2024; Koessler et al., 2024). Each investment thesis (Section 7.6) concentrates capital *and* governance attention.

9.4. Coordination risk

Objection. The flywheel only turns when data standards are shared. Fragmented schemas and platform fragmentation will erode efficiency (Benkler, 2002; Lessig, 2006).

Reply. Coordination is a first-class friction. The argument is not that supply grounding is free; it is that it pays back its coordination cost in regimes where agents consume verifiable signal faster than human-curated schemas produce it. The MCP / A2A protocol layer is precisely the standardisation investment whose cost is being paid back.

9.5. Data sovereignty and reproducibility

Objection. Supply data is proprietary. If the position depends on data only platform operators can access, it cannot be independently verified.[2]

[2]A related objection concerns citation status: the methodological-anchor citations in this paper (see Ding 2026a,b *inter alia*) are cited as support for specific claims — trajectory recycling and verifier adaptivity — not as evidence that the

Reply. The position does not require data sharing between competitors. It requires each participant to build a grounding flywheel on their own supply data, climbing the DMM ladder to create a new form of platform moat (Coase, 1937). For open science: (a) public benchmarks (Section 8.8); (b) the falsifiable form is "any environment satisfying (i)–(v) at $D \geq 3$ should exhibit the predicted advantage"; (c) anonymised aggregate statistics can be published without trade secrets.

9.6. Stakeholder tensions the position creates

The DMM and SCI implicitly side with agentic consumers in conflicts that the supply environment must mediate. Four stakeholder perspectives are in tension with the position as written:

End users / consumers. Personalisation, exploration injection, and engagement-optimised ranking are user-welfare features for human-facing surfaces. Pushing platforms to $D3/D4$ on user-facing surfaces would impose real welfare costs on this constituency. The defensible position is *persona-conditioned* determinism: deterministic interfaces under an agent persona, conventional engagement-optimised interfaces under a human persona. The DMM should be read as a property of the *agent-persona surface*, not of the platform's user-facing surface.

Suppliers / sellers on supply platforms. Smaller or newer suppliers benefit from exploration injection that $D3$ removes; $D3/D4$ ranking with stable, faithful order concentrates visibility on incumbents with established trust signals. This is a real cost the position does not internalise; the SCI's trustworthy-supply property treats supply-side coverage as a gain without engaging the distributional effect on the supply side itself. A future extension should model the supply-side incentive implications of climbing the DMM.

Regulators and competition authorities. Deterministic substrates concentrate agent access into a few platforms that have invested the capital to climb the DMM ladder. We celebrate this as a "platform moat" (Section 9 reply to data sovereignty) but the antitrust corollary is real: agent-mediated commerce on $D3/D4$ substrates is a more concentrated market than the equivalent on $D2$ substrates. The EU Digital Markets Act and U.S. FTC merger guidelines on platform gatekeeping are directly adjacent; the position has no answer for them.

Smaller platforms and developing-market marketplaces. DMM compliance has a capital cost (verifier service, skill registry, telemetry plane). $D3$ is achievable for hyperscale incumbents and may be out of reach for small or developing-market platforms. A world stratified by DMM level concentrates flywheel-grade data into incumbents who can train and serve foundation agents; smaller platforms become data tributaries rather than independent flywheel operators. This is a real concern even if the formal position is correct, and we list it as a Limitations bullet (Section 10).

B. Strong competing positions

We engage three competing positions. Each predicts different outcomes and provides a route by which the position could lose.

position has been validated. Only Ding et al. (2023) is directly cited by the AGI→ASI survey (Genewein et al., 2026) we extend.

9.7. Competing position 1: sim-to-real suffices

The strongest version of this position holds that photorealistic simulation with domain randomisation (OpenAI et al., 2019; Tobin et al., 2017), co-evolving world models (Bruce et al., 2024; Fang et al., 2025; Hafner et al., 2020), and zero-shot transfer will close the gap before deterministic real environments become widespread, implying that marginal compute on simulator fidelity dominates marginal compute on real-environment determinism. Section 4.6 gives the structural reply: simulators inherit a designer-ontology cap, a verifier ground-truth shift, and a reward-design fragility that qualifying real environments do not. The position loses if OQ1 returns null, and we commit to that null in advance.

9.8. Competing position 2: better alignment training suffices

Position. Better RLHF / RLAIF / weak-to-strong supervision plus cognitive architectures (Bai et al., 2022; Burns et al., 2024; Ouyang et al., 2022; Shinn et al., 2023; Sumers et al., 2024; Yao et al., 2023) will absorb environmental noise. The environment can stay stochastic.

What it predicts. Per-step success rates will climb fast enough that δ^k ceases to bite before deterministic environments become widespread.

Our reply. This is the position we take most seriously. Proposition 1 is a per-step argument: model robustness reduces effective non-deterministic steps but does not change the exponential shape (Remark 2). Proposition 2 bounds how far alignment training alone takes a policy under fixed-quality verifiers. Better training and better environments are complements; the position loses if operator-class deployments reach end-to-end success rates above 90% on workflows of chain length $k \geq 8$ in $D2$ environments — the regime in which Proposition 1 predicts below-50% success at $\delta = 0.9$ even with a generous retry budget (Remark 2). We commit to retraction in that case.

Variant: orchestration-first frameworks. A practitioner-side variant of this position is held by orchestration-first frameworks (LangChain, LangGraph, CrewAI, AutoGen): the answer is not better training but *better scaffolding* — state machines, retry policies, human-in-the-loop checkpoints, and typed memory — wrapped around the existing stochastic environment. This variant inherits the same per-step counter-argument: scaffolding reduces *effective* k via batching and checkpointing but does not change per-step δ. Where the orchestrator's own decision edges are themselves stochastic (LLM-routed tool selection, model-as-judge gating), the scaffolding introduces additional δ-loaded steps and can *worsen* the chain-task budget. The position's prediction: orchestration frameworks built on $D2$ environments will hit a ceiling that the same orchestration on $D3/D4$ does not.

9.9. Competing position 3: AI as normal technology

Position. Narayanan and Kapoor (2024) argue that AI's deployment will resemble prior general-purpose technologies, with adoption gated by institutional and regulatory diffusion — not sharp infrastructure inflections.

What it predicts. Agent-task success improves through gradual sociotechnical channels with no discontinuous improvement at any infrastructure inflection point.

Our reply. We do not contest institutional adoption; we contest the absence of an inflection. The $D2 \to D3$ transition is non-linear: below the verifier-channel threshold the flywheel does not turn; above it, it compounds. If OQ4 returns a smooth fit with no inflection, this competing position is supported and ours is weakened.

9.10. What would force retraction?

- OQ1 null AND OQ5 high-ε $\Rightarrow$ flywheel half invalidated.
- OQ2 null AND OQ3 null $\Rightarrow$ grounding-as-leverage retracted to multi-agent-trust-only.
- Concurrent failure of OQ1–OQ4 $\Rightarrow$ position comprehensively refuted.
- δ does not predict chain-task success across domains $\Rightarrow$ determinism framing abandoned.
- Operator-class deployments exceed 90% end-to-end success on $k \geq 8$ workflows in $D2$ environments $\Rightarrow$ competing position 2 wins; retract.
- Smooth OQ4 fit consistent with Narayanan and Kapoor (2024) $\Rightarrow$ competing position 3 wins; retract inflection claim.

10. Conclusion

We have argued that the central frictions on AGI→ASI progress — data wall, abstraction barrier, embodied bottleneck, multi-agent trust — are not first-order compute problems but Grounding problems, and that commercially self-sustaining supply environments satisfying a *deterministic interface guarantee* form a privileged grounding substrate. Three formal results anchor the argument:

- the Determinism–Efficiency Bound (Proposition 1): chain-task success degrades as δ^k;
- the Verifier–Goodharting Floor (Proposition 2): flywheel asymptotics bounded by verifier quality ε;
- the Grounded Self-Evolution Convergence Condition (Proposition 3): environment-side skill evolution accumulates iff $\delta > \delta_{\min}$.

We have decomposed Supply Certainty into five properties aggregated into the Supply Certainty Index (Definition 2), proposed a Determinism Maturity Model (Section 7.7) as an adoption ladder, named five investable primitive categories (Section 7.6), and committed to a falsifiable strong form with five open questions whose null results would weaken or retract the position. We have engaged explicitly with three strong competing positions, each of which would force retraction of specific claims.

A sustainability corollary

Every failed chain-task is a form of wasted compute: an agent that fails at step $j < k$ of a k-step plan consumes inference cost for j steps and produces nothing. The δ^k bound implies that environment determinism is a major lever for compute efficiency in agentic workloads at scale. A platform moving from $\delta = 0.8$ to $\delta = 0.95$ on $k = 6$ workflows shifts chain success from 0.26 to 0.74, cutting wasted inference roughly 2.8× before any model-side optimisation.

Partial-equilibrium caveat. This is an inference-layer, holding-everything-else-constant calculation; a full lifecycle accounting must include (i) the upstream cost of building the higher-δ substrate (verifier services, telemetry plane, dedicated agent-persona infrastructure), (ii) retry-induced compute on failed chains (Remark 2: shared retry budgets reduce the multiplier), and (iii) amortised training compute over the lifetime of agents deployed against the substrate. We do not perform that full

life-cycle accounting here; we observe only that the inference-layer multiplier is large enough that sustainability and capability arguments *point in the same direction* on the $D2 \to D3$ transition, and leave the precise life-cycle decomposition as future work.

Limitations

Four limitations the present paper does not resolve. *First*, all three formal results (Propositions 1, 2, 3) depend on the verifier returning consistent verdicts on independent trials; verifier corruption mid-deployment is bounded only at the asymptotic ε level and is not modelled as a time-varying adversarial process. *Second*, the Supply Certainty Index is a composite designed for platform comparability *within* a domain, not for absolute capability prediction *across* domains; an $\mathrm{SCI}^{\delta}(\mathcal{E}) = 0.8$ in one domain may correspond to capabilities that differ substantially from $\mathrm{SCI}^{\delta} = 0.8$ in another. Cross-domain absolute comparison is out of scope. *Third*, the position is consistently positive-conditional: *if* the substrate is built, *then* the flywheel turns. We do not address the governance side — *whether* the substrate should be built at every site — leaving that to the governance literature (Anderljung et al., 2023; Bengio et al., 2024; European Parliament and Council, 2024; Hendrycks et al., 2023; Koessler et al., 2024). *Fourth, the position has a power-concentration corollary it does not internalise.* DMM compliance has a real capital cost (verifier service, skill registry, telemetry plane); a world stratified by DMM level will see flywheel-grade data accumulate in hyperscale incumbents who can afford to climb, while smaller and developing-market platforms become data tributaries. We celebrate this as a "platform moat" in Section 9's reply to the data-sovereignty objection, but the equity and antitrust corollaries deserve treatment in follow-on work (Section 9.6).

Provocation. If the position is correct, the most undervalued asset on the way to superintelligence is the existing infrastructure of verified commercial transactions. These signals are produced as by-products of ordinary economic activity, and they constitute a large reservoir of grounding data that no amount of synthetic generation can substitute for.

We expect the decisive competitive variable for agentic AI in the coming years to be how much of the world a platform can verifiably observe and expose to agents at $D3$ or above. Compute remains necessary; it is not sufficient.

Acknowledgements

We thank colleagues who reviewed earlier drafts of the argument backbone for sharpening the falsifiability discipline of the open- question programme and for pressing the engagement with competing positions in Section 9.

Use of AI Assistants

Large-language-model assistants were used for prose tightening, cross-reference checking, and bibliographic search. All technical claims, formal results, and the falsifiable research programme are the responsibility of the authors.

References

D. Acemoglu and S. Johnson. *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity*. PublicAffairs, 2023. URL https://www.hachettebookgroup.com/titles/daron-acemoglu/power-and-progress/9781541702530/.

Adaptive Agent Team, J. Bauer, K. Baumli, F. Behbahani, et al. Human-timescale adaptation in an open-ended task space, 2023. URL https://arxiv.org/abs/2301.07608. arXiv:2301.07608.

A. Agrawal, J. S. Gans, and A. Goldfarb. The economics of AI foundation models: Openness, competition, and governance, 2025. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6800518. Working paper.

E. AI. Training compute of frontier AI models grows by 4-5x per year. https://epochai.org/blog/training-compute-of-frontier-ai-models-grows-by-4-5x-per-year, 2024. URL https://epochai.org/blog/training-compute-of-frontier-ai-models-grows-by-4-5x-per-

G. A. Akerlof. The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3):488–500, 1970. URL https://doi.org/10.2307/1879431.

M. Anderljung et al. Frontier AI regulation: Managing emerging risks to public safety, 2023. URL https://arxiv.org/abs/2307.03718. arXiv:2307.03718.

Andreessen Horowitz. The agent economy: How AI agents will reshape software, commerce, and capital. a16z Big Ideas 2026 (Part 3), 2026. URL https://a16z.com/newsletter/big-ideas-2026-part-3/.

Anthropic. Introducing computer use, a new Claude 3.5 Sonnet, and Claude 3.5 Haiku. https://www.anthropic.com/news/3-5-models-and-computer-use, 2024a. URL https://www.anthropic.com/news/3-5-models-and-computer-use.

Anthropic. Model context protocol, 2024b. URL https://www.anthropic.com/news/model-context-protocol.

Anthropic. When AI builds itself: Progress toward recursive self-improvement. https://www.anthropic.com/institute/recursive-self-improvement, 2025. URL https://www.anthropic.com/institute/recursive-self-improvement.

S. Bai, X. Wang, L. Yu, B. Chen, Z. Xu, Y. Sheng, C. Zan, X. Zhu, Y. Zhang, J. Li, M. Guo, L. Zou, Y. Li, C. Huo, and L. Ding. IndustryBench: Probing the industrial knowledge boundaries of LLMs, 2026. URL https://arxiv.org/abs/2605.10267. arXiv:2605.10267.

Y. Bai et al. Constitutional AI: Harmlessness from AI feedback, 2022. URL https://arxiv.org/abs/2212.08073. arXiv:2212.08073.

L. W. Barsalou. Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4):577–660, 1999. URL https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/perceptual-symbol-systems/C2D720D63C1CE3D7153F6BA473F9DD87.

Y. Bengio, S. Clare, C. Prunkl, S. Rismani, et al. International AI safety report 2025, 2025. URL https://arxiv.org/abs/2510.13653. arXiv:2510.13653.

Y. Bengio et al. International scientific report on the safety of advanced AI: Interim report, 2024. URL https://www.gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai. UK AI Safety Institute.

Y. Benkler. Coase's penguin, or, Linux and the nature of the firm. *The Yale Law Journal*, 112(3): 369–446, 2002. URL https://doi.org/10.2307/1562247.

B. Biggio and F. Roli. Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, 84:317–331, 2018. URL https://doi.org/10.1016/j.patcog.2018.07.023.

N. Bloom, C. I. Jones, J. Van Reenen, and M. Webb. Are ideas getting harder to find? *American Economic Review*, 110(4):1104–1144, 2020. URL https://doi.org/10.1257/aer.20180338.

N. Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014. URL https://global.oup.com/academic/product/superintelligence-9780199678112.

J. Bruce, M. Dennis, A. Edwards, et al. Genie: Generative interactive environments, 2024. URL https://arxiv.org/abs/2402.15391. arXiv:2402.15391.

E. Brynjolfsson. The Turing trap: The promise & peril of human-like artificial intelligence. *Dædalus*, 151(2):272–287, 2022. URL https://doi.org/10.1162/daed_a_01915.

C. Burns, P. Izmailov, J. H. Kirchner, B. Baker, L. Gao, L. Aschenbrenner, Y. Chen, A. Ecoffet, M. Joglekar, J. Leike, I. Sutskever, and J. Wu. Weak-to-strong generalization: Eliciting strong capabilities with weak supervision. In *ICML*, 2024. URL https://arxiv.org/abs/2312.09390. arXiv:2312.09390.

S. Casper, X. Davies, C. Shi, T. K. Gilbert, J. Scheurer, J. Rando, R. Freedman, T. Korbak, D. Lindner, P. Freire, et al. Open problems and fundamental limitations of reinforcement learning from human feedback, 2023. URL https://arxiv.org/abs/2310.06827. arXiv:2310.06827.

A. Chan, C. Ezell, M. Kaufmann, K. Wei, L. Hammond, H. Bradley, E. Bluemke, N. Rajkumar, D. Krueger, N. Kolt, L. Heim, and M. Anderljung. Infrastructure for AI Agents, 2025. URL https://arxiv.org/abs/2501.10114. arXiv:2501.10114.

F. Chollet. On the measure of intelligence, 2019. URL https://arxiv.org/abs/1911.01547. arXiv:1911.01547.

P. Christiano, B. Shlegeris, and D. Amodei. Supervising strong learners by amplifying weak experts, 2018. URL https://arxiv.org/abs/1810.08575. arXiv:1810.08575.

R. H. Coase. The nature of the firm. *Economica*, 4(16):386–405, 1937. URL https://doi.org/10.1111/j.1468-0335.1937.tb00002.x.

T. Davidson, D. Halperin, T. Houlden, and A. Korinek. When does automating AI research produce explosive growth?, 2026. URL https://www.nber.org/papers/w35155. NBER Working Paper 35155.

L. Ding. AdaRubric: Task-adaptive rubrics for reliable LLM agent evaluation and reward learning, 2026a. URL https://arxiv.org/abs/2603.21362. arXiv:2603.21362.

L. Ding. AgentHER: Hindsight experience replay for LLM agent trajectory relabeling, 2026b. URL https://arxiv.org/abs/2603.21357. arXiv:2603.21357.

T. Ding, T. Chen, H. Zhu, J. Jiang, Y. Zhong, J. Zhou, G. Wang, Z. Zhu, I. Zharkov, and L. Liang. The efficiency spectrum of large language models: An algorithmic survey, 2023. URL https://arxiv.org/abs/2312.00678.

E. Dohmatob, Y. Feng, P. Yang, F. Charton, and J. Kempe. A tale of tails: Model collapse as a change of scaling laws, 2024. URL https://arxiv.org/abs/2402.07043. arXiv:2402.07043.

E. Dolstra, M. de Jonge, and E. Visser. Nix: A safe and policy-free system for software deployment. In *LISA '04: 18th Large Installation System Administration Conference*. USENIX, 2004. URL https://dl.acm.org/doi/10.5555/1052676.1052686.

K. E. Drexler. Reframing superintelligence: Comprehensive AI services as general intelligence, 2019. URL https://www.fhi.ox.ac.uk/reframing/. Technical Report 2019-1, Future of Humanity Institute.

A. Dubey et al. The Llama 3 herd of models, 2024. URL https://arxiv.org/abs/2407.21783. arXiv:2407.21783.

E. Erdil, A. Potlogea, T. Besiroglu, E. Roldan, A. Ho, J. Sevilla, M. Barnett, M. Vrzla, and R. Sandler. GATE: An integrated assessment model for AI automation, 2025. URL https://arxiv.org/abs/2503.04941. arXiv:2503.04941.

European Parliament and Council. Regulation (EU) 2024/1689 of the european parliament and of the council (AI act). Official Journal of the European Union, 2024. URL https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

K. Eykholt, I. Evtimov, et al. Robust physical-world attacks on deep learning visual classification. In *CVPR*, 2018. URL https://arxiv.org/abs/1707.08945.

T. Fang, H. Zhang, Z. Zhang, K. Ma, W. Yu, H. Mi, and D. Yu. WebEvolver: Enhancing web agent self-improvement with co-evolving world model. In *EMNLP*, 2025. URL https://arxiv.org/abs/2504.21024. arXiv:2504.21024.

X. Feng, X. Song, L. Li, G. Liu, and J. Shao. SEARL: Joint optimization of policy and tool graph memory for self-evolving agents, 2026. URL https://arxiv.org/abs/2604.07791. arXiv:2604.07791.

H.-a. Gao, J. Geng, W. Hua, M. Hu, et al. A survey of self-evolving agents: On path to artificial super intelligence, 2025. URL https://arxiv.org/abs/2507.21046. arXiv:2507.21046.

L. Gao, J. Schulman, and J. Hilton. Scaling laws for reward model overoptimization, 2023. URL https://arxiv.org/abs/2210.10760. arXiv:2210.10760.

T. Genewein et al. From AGI to ASI, 2026. URL https://arxiv.org/abs/2606.12683. Position paper, Google DeepMind.

M. Gerstgrasser, R. Schaeffer, A. Dey, R. Rafailov, H. Sleight, J. Hughes, T. Korbak, R. Agrawal, D. Pai, A. Gromov, D. A. Roberts, D. Yang, D. L. Donoho, and S. Koyejo. Is model collapse inevitable? breaking the curse of recursion by accumulating real and synthetic data, 2024. URL https://arxiv.org/abs/2404.01413. arXiv:2404.01413.

C. A. E. Goodhart. Problems of monetary management: The UK experience. *Papers in Monetary Economics*, 1975. URL https://en.wikipedia.org/wiki/Goodhart%27s_law.

Google. Agent2Agent protocol. https://google-a2a.github.io/A2A, 2025. URL https://google-a2a.github.io/A2A.

A. Gu and T. Dao. Mamba: Linear-time sequence modeling with selective state spaces, 2023. URL https://arxiv.org/abs/2312.00752. arXiv:2312.00752.

D. Hafner, T. Lillicrap, J. Ba, and M. Norouzi. Dream to control: Learning behaviors by latent imagination. In *ICLR*, 2020. URL https://arxiv.org/abs/1912.01603.

L. Hammond, A. Chan, J. Clifton, J. Hoelscher-Obermaier, et al. Multi-agent risks from advanced AI, 2025. URL https://arxiv.org/abs/2502.14143. arXiv:2502.14143.

S. Harnad. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1–3):335–346, 1990. URL https://doi.org/10.1016/0167-2789(90)90087-6.

S. He, L. Ding, D. Dong, J. Zhang, and D. Tao. SparseAdapter: An easy approach for improving the parameter-efficiency of adapters. In *Findings of the Association for Computational Linguistics: EMNLP*, 2022. URL https://arxiv.org/abs/2210.04284.

D. Hendrycks, M. Mazeika, and T. Woodside. An overview of catastrophic AI risks, 2023. URL https://arxiv.org/abs/2306.12001. arXiv:2306.12001.

D. Hernandez and T. B. Brown. Measuring the algorithmic efficiency of neural networks, 2020. URL https://arxiv.org/abs/2005.04305. arXiv:2005.04305.

A. Ho, T. Besiroglu, E. Erdil, D. Owen, R. Rahman, Z. C. Guo, D. Atkinson, N. C. Thompson, and J. Sevilla. Algorithmic progress in language models, 2024. URL https://arxiv.org/abs/2403.05812. arXiv:2403.05812.

A. Ho, J.-S. Denain, D. Atanasov, S. Albanie, and R. Shah. A rosetta stone for AI benchmarks, 2025. URL https://arxiv.org/abs/2512.00193. arXiv:2512.00193.

J. Hoffmann, S. Borgeaud, A. Mensch, et al. Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*, 2022. URL https://arxiv.org/abs/2203.15556.

S. Hong, M. Zhuge, J. Chen, X. Zheng, Y. Cheng, C. Zhang, J. Wang, Z. Wang, S. K. S. Yau, Z. Lin, L. Zhou, C. Ran, L. Xiao, C. Wu, and J. Schmidhuber. MetaGPT: Meta programming for a multi-agent collaborative framework. In *ICLR*, 2024. URL https://arxiv.org/abs/2308.00352. arXiv:2308.00352.

M. Hutter. *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability*. Springer, 2004. URL https://doi.org/10.1007/b138233.

M. Hutter, D. Quarel, and E. Catt. *An Introduction to Universal Artificial Intelligence*. CRC Press / Chapman & Hall, 2024. URL https://doi.org/10.1201/9781003438007.

C. E. Jimenez, J. Yang, A. Wettig, S. Yao, K. Pei, O. Press, and K. Narasimhan. SWE-bench: Can language models resolve real-world GitHub issues? In *ICLR*, 2024. URL https://arxiv.org/abs/2310.06770. arXiv:2310.06770.

J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020. URL https://arxiv.org/abs/2001.08361.

S. Kapoor, B. Stroebl, Z. S. Siegel, N. Nadgir, and A. Narayanan. AI Agents That Matter, 2024. URL https://arxiv.org/abs/2407.01502. arXiv:2407.01502.

A. Karpathy. Software is changing (again): The three-era talk. AI Startup School (Y Combinator), 2025. URL https://www.youtube.com/watch?v=LCEmiRjPEtQ.

S. Kim et al. Prometheus 2: An open source language model specialized in evaluating other language models, 2024. URL https://arxiv.org/abs/2405.01535. arXiv:2405.01535.

J. Kirkpatrick, R. Pascanu, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017. URL https://doi.org/10.1073/pnas.1611835114.

L. Koessler, J. Schuett, and M. Anderljung. Risk thresholds for frontier AI, 2024. URL https://arxiv.org/abs/2406.14713. arXiv:2406.14713.

B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman. Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40, 2017. URL https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/building-machines-that-learn-and-think-like-people/A9535B1D745A0377E16C590E14B94993. e253.

N. Lambert, V. Pyatkin, J. Morrison, L. Miranda, B. Y. Lin, K. Chandu, N. Dziri, S. Kumar, T. Zick, Y. Choi, N. A. Smith, and H. Hajishirzi. RewardBench: Evaluating reward models for language modeling, 2024. URL https://arxiv.org/abs/2403.13787. arXiv:2403.13787.

N. D. Lawrence. The atomic human: Understanding ourselves in the age of AI, 2024. URL https://www.penguin.co.uk/books/455194/the-atomic-human-by-lawrence-neil-d/9780241625248. Cited in Genewein et al. as anchor for the embodiment factor.

H. Lee et al. RLAIF: Scaling reinforcement learning from human feedback with AI feedback, 2023. URL https://arxiv.org/abs/2309.00267. arXiv:2309.00267.

J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter. Learning quadrupedal locomotion over challenging terrain. *Science Robotics*, 5(47), 2020. eaaz1422.

S. Legg and M. Hutter. Universal intelligence: A definition of machine intelligence. *Minds and Machines*, 17(4):391–444, 2007. URL https://doi.org/10.1007/s11023-007-9079-x.

J. Z. Leibo, J. Perolat, E. Hughes, et al. Malthusian reinforcement learning. In *AAMAS*, 2019. URL https://arxiv.org/abs/1812.07019.

L. Lessig. *Code: And Other Laws of Cyberspace, Version 2.0*. Basic Books, 2006. URL https://lessig.org/product/code/.

X. Li, P. Yu, C. Zhou, T. Schick, O. Levy, L. Zettlemoyer, J. Weston, and M. Lewis. Self-alignment with instruction backtranslation, 2023. URL https://arxiv.org/abs/2308.06259. arXiv:2308.06259.

Y. Li, R. Miao, Z. Qi, and T. Lan. ARISE: Agent reasoning with intrinsic skill evolution in hierarchical reinforcement learning, 2026. URL https://arxiv.org/abs/2603.16060. arXiv:2603.16060.

B. Liu, L. Guertler, S. Yu, Z. Liu, P. Qi, D. Balcells, M. Liu, C. Tan, W. Shi, M. Lin, W. S. Lee, and N. Jaques. SPIRAL: Self-play on zero-sum games incentivizes reasoning via multi-agent multi-turn reinforcement learning, 2025. URL https://arxiv.org/abs/2506.24119. arXiv:2506.24119.

X. Liu, H. Yu, et al. AgentBench: Evaluating LLMs as agents, 2024. URL https://arxiv.org/abs/2308.03688. ICLR 2024; arXiv:2308.03688.

C. Lu, C. Lu, R. T. Lange, J. Foerster, J. Clune, and D. Ha. The AI scientist: Towards fully automated open-ended scientific discovery, 2024. URL https://arxiv.org/abs/2408.06292. arXiv:2408.06292.

G. Mialon, C. Fourrier, C. Swift, T. Wolf, Y. LeCun, and T. Scialom. GAIA: a benchmark for general AI assistants, 2024. URL https://arxiv.org/abs/2311.12983. ICLR; arXiv:2311.12983.

R. Min, Z. Qiao, Z. Xu, J. Zhai, W. Gao, X. Chen, H. Sun, Z. Zhang, X. Wang, H. Zhou, W. Yin, B. Zhang, X. Zhou, M. Yan, Y. Jiang, H. Liu, L. Ding, L. Zou, Y. R. Fung, Y. Li, and P. Xie. EcomBench: Towards holistic evaluation of foundation agents in e-commerce, 2025. URL https://arxiv.org/abs/2512.08868. arXiv:2512.08868.

M. R. Morris, J. Sohl-Dickstein, N. Fiedel, T. Warkentin, A. Dafoe, A. Faust, C. Farabet, and S. Legg. Levels of AGI: Operationalizing progress on the path to AGI, 2024. URL https://arxiv.org/abs/2311.02462. arXiv:2311.02462.

A. Narayanan and S. Kapoor. AI as normal technology. Knight First Amendment Institute, 2024. URL https://knightcolumbia.org/content/ai-as-normal-technology. https://knightcolumbia.org/content/ai-as-normal-technology.

R. Ngo, L. Chan, and S. Mindermann. The alignment problem from a deep learning perspective, 2022. URL https://arxiv.org/abs/2209.00626. arXiv:2209.00626.

OpenAI. Function calling and the assistants API. OpenAI platform documentation, 2024. URL https://platform.openai.com/docs/guides/function-calling.

OpenAI. Introducing Operator. https://openai.com/index/introducing-operator, 2025. URL https://openai.com/index/introducing-operator.

OpenAI, I. Akkaya, M. Andrychowicz, M. Chociej, M. Litwin, B. McGrew, A. Petron, A. Paino, M. Plappert, G. Powell, R. Ribas, J. Schneider, N. Tezak, J. Tworek, P. Welinder, L. Weng, Q. Yuan, W. Zaremba, and L. Zhang. Solving Rubik's cube with a robot hand. In *NeurIPS Robot Learning Workshop*, 2019. URL https://arxiv.org/abs/1910.07113. arXiv:1910.07113.

OpenAI, M. Andrychowicz, B. Baker, M. Chociej, R. Józefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, J. Schneider, S. Sidor, J. Tobin, P. Welinder, L. Weng, and W. Zaremba. Learning dexterous in-hand manipulation, 2020. URL https://arxiv.org/abs/1808.00177. arXiv:1808.00177.

P. A. Ortega, M. Kunesch, G. Delétang, T. Genewein, J. Grau-Moya, J. Veness, J. Buchli, J. Degrave, B. Piot, J. Perolat, T. Everitt, C. Tallec, E. Parisotto, T. Erez, Y. Chen, S. Reed, M. Hutter, N. de Freitas, and S. Legg. Shaking the foundations: delusions in sequence models for interaction and control. *arXiv preprint arXiv:2110.10819*, 2021. URL https://arxiv.org/abs/2110.10819.

L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, et al. Training language models to follow instructions with human feedback, 2022. URL https://arxiv.org/abs/2203.02155. NeurIPS 2022; arXiv:2203.02155.

L. Panait and S. Luke. Cooperative multi-agent learning: The state of the art. *Autonomous Agents and Multi-Agent Systems*, 11:387–434, 2005. URL https://doi.org/10.1007/s10458-005-2631-2.

J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of UIST*, 2023. URL https://arxiv.org/abs/2304.03442. arXiv:2304.03442.

X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel. Sim-to-real transfer of robotic control with dynamics randomization. In *ICRA*, 2018. URL https://arxiv.org/abs/1710.06537. arXiv:1710.06537.

B. A. Plummer, L. Wang, C. M. Cervantes, J. C. Caicedo, J. Hockenmaier, and S. Lazebnik. Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence models. In *ICCV*, 2015. URL https://arxiv.org/abs/1505.04870.

H. Qi, J. Cao, Y. Zhang, X. Wang, W. Tang, B. Chen, C. Huo, H. Pan, H. You, J. Li, Y. Wang, and L. Ding. IndustryBench-MIPU: Benchmarking multi-image attribute value extraction for industrial products, 2026. URL https://arxiv.org/abs/2606.14383. arXiv:2606.14383.

E. Real, C. Liang, D. R. So, and Q. V. Le. AutoML-Zero: Evolving machine learning algorithms from scratch. In *ICML*, 2020. URL https://arxiv.org/abs/2003.03384.

S. Russell. Human compatible: Artificial intelligence and the problem of control, 2019. URL https://www.penguinrandomhouse.com/books/566677/human-compatible-by-stuart-russell/.

T. Schick, J. Dwivedi-Yu, R. Dessì, R. Raileanu, M. Lomeli, L. Zettlemoyer, N. Cancedda, and T. Scialom. Toolformer: Language models can teach themselves to use tools, 2023. URL https://arxiv.org/abs/2302.04761. arXiv:2302.04761.

J. Schmidhuber. Gödel machines: Self-referential universal problem solvers making provably optimal self-improvements. *arXiv preprint cs/0309048*, 2003. URL https://arxiv.org/abs/cs/0309048.

J. Schrittwieser, I. Antonoglou, T. Hubert, et al. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588:604–609, 2020. URL https://doi.org/10.1038/s41586-020-03051-4.

J. Sevilla, L. Heim, A. Ho, T. Besiroglu, M. Hobbhahn, and P. Villalobos. Compute trends across three eras of machine learning, 2022. URL https://arxiv.org/abs/2202.05924. arXiv:2202.05924.

N. Shinn, F. Cassano, E. Berman, A. Gopinath, K. Narasimhan, and S. Yao. Reflexion: Language agents with verbal reinforcement learning. In *NeurIPS*, 2023. URL https://arxiv.org/abs/2303.11366. arXiv:2303.11366.

I. Shumailov, Z. Shumaylov, Y. Zhao, N. Papernot, R. Anderson, and Y. Gal. AI models collapse when trained on recursively generated data. *Nature*, 631:755–759, 2024. URL https://doi.org/10.1038/s41586-024-07566-y.

D. Silver, J. Schrittwieser, K. Simonyan, et al. Mastering the game of Go without human knowledge. *Nature*, 550:354–359, 2017. URL https://doi.org/10.1038/nature24270.

C. Snell, J. Lee, K. Xu, and A. Kumar. Scaling LLM test-time compute optimally can be more effective than scaling model parameters, 2024. URL https://arxiv.org/abs/2408.03314. arXiv:2408.03314.

L. Soldaini et al. Dolma: an open corpus of three trillion tokens for language model pretraining research. In *Proceedings of ACL*, 2024. URL https://arxiv.org/abs/2402.00159.

A. Srivastava et al. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models, 2022. URL https://arxiv.org/abs/2206.04615. BIG-bench, arXiv:2206.04615.

Stripe. 2024 annual letter. Stripe Annual Update, 2024. URL https://stripe.com/annual-updates/2024.

Stripe. Agent SDK and agent toolkit. Stripe developer documentation, 2025. URL https://docs.stripe.com/agents.

T. R. Sumers, S. Yao, K. Narasimhan, and T. L. Griffiths. Cognitive architectures for language agents, 2024. URL https://arxiv.org/abs/2309.02427. arXiv:2309.02427.

R. S. Sutton. The bitter lesson. http://www.incompleteideas.net/IncIdeas/BitterLesson.html, 2019. URL http://www.incompleteideas.net/IncIdeas/BitterLesson.html.

J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *IROS*, 2017. URL https://arxiv.org/abs/1703.06907. arXiv:1703.06907.

N. Tomašev, M. Franklin, J. Z. Leibo, J. Jacobs, W. A. Cunningham, I. Gabriel, and S. Osindero. Virtual agent economies, 2025. URL https://arxiv.org/abs/2509.10147. arXiv:2509.10147.

P. Villalobos, A. Ho, J. Sevilla, T. Besiroglu, L. Heim, and M. Hobbhahn. Will we run out of data? limits of LLM scaling based on human-generated data, 2024. URL https://arxiv.org/abs/2211.04325. arXiv:2211.04325v2.

G. Wang, Y. Xie, Y. Jiang, A. Mandlekar, C. Xiao, Y. Zhu, L. Fan, and A. Anandkumar. Voyager: An open-ended embodied agent with large language models. *Transactions on Machine Learning Research*, 2024a. URL https://arxiv.org/abs/2305.16291. arXiv:2305.16291.

J. Wang, Q. Yan, Y. Wang, Y. Tian, S. S. Mishra, Z. Xu, M. Gandhi, P. Xu, and L. L. Cheong. Reinforcement learning for self-improving agent with skill library, 2025. URL https://arxiv.org/abs/2512.17102. arXiv:2512.17102.

Q. Wang, Y. Fu, Y. Cao, S. Wang, Z. Tian, and L. Ding. Recursively summarizing enables long-term dialogue memory in large language models, 2023. URL https://arxiv.org/abs/2308.15022. arXiv:2308.15022.

X. Wang, Y. Chen, L. Yuan, et al. Executable code actions elicit better LLM agents, 2024b. URL https://arxiv.org/abs/2402.01030. arXiv:2402.01030.

X. Wang, J. Pan, L. Ding, and C. Biemann. Mitigating hallucinations in large vision-language models with instruction contrastive decoding. In *Findings of ACL*, 2024c. URL https://aclanthology.org/2024.findings-acl.937/.

A. Wei, N. Haghtalab, and J. Steinhardt. Jailbroken: How does LLM safety training fail?, 2023. URL https://arxiv.org/abs/2307.02483. NeurIPS; arXiv:2307.02483.

L. Weidinger et al. Taxonomy of risks posed by language models, 2022. URL https://doi.org/10.1145/3531146.3533088. FAccT 2022.

Q. Wu, G. Bansal, J. Zhang, Y. Wu, B. Li, E. Zhu, L. Jiang, X. Zhang, S. Zhang, J. Liu, A. H. Awadallah, R. W. White, D. Burger, and C. Wang. AutoGen: Enabling next-gen LLM applications via multi-agent conversation, 2023. URL https://arxiv.org/abs/2308.08155. arXiv:2308.08155.

T. Xie, D. Zhang, J. Chen, X. Li, S. Zhao, R. Cao, T. J. Hua, Z. Cheng, D. Shin, F. Lei, Y. Liu, Y. Xu, S. Zhou, S. Savarese, C. Xiong, V. Zhong, and T. Yu. OSWorld: Benchmarking multimodal agents for open-ended tasks in real computer environments. In *NeurIPS Datasets and Benchmarks*, 2024. URL https://arxiv.org/abs/2404.07972. arXiv:2404.07972.

Y. Yamada, R. T. Lange, C. Lu, S. Hu, C. Lu, J. Foerster, J. Clune, and D. Ha. The AI scientist-v2: Workshop-level automated scientific discovery via agentic tree search, 2024. URL https://arxiv.org/abs/2504.08066. arXiv:2504.08066.

Y. Yang, Z. Gong, W. Huang, Q. Yang, Z. Zhou, Z. Huang, Y. Li, X. Gao, Q. Dai, B. Liu, K. Qiu, Y. Yang, D. Chen, X. Yang, and C. Luo. SkillOpt: Executive strategy for self-evolving agent skills, 2026. URL https://arxiv.org/abs/2605.23904. Microsoft Research; arXiv:2605.23904.

S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao. ReAct: Synergizing reasoning and acting in language models. In *ICLR*, 2023. URL https://arxiv.org/abs/2210.03629. arXiv:2210.03629.

S. Yao, N. Shinn, P. Razavi, and K. Narasimhan. τ-bench: A benchmark for tool-agent-user interaction in real-world domains, 2024. URL https://arxiv.org/abs/2406.12045. arXiv:2406.12045.

W. Yuan, R. Y. Pang, K. Cho, S. Sukhbaatar, J. Xu, and J. Weston. Self-rewarding language models, 2024. URL https://arxiv.org/abs/2401.10020. arXiv:2401.10020.

J. Zhang, S. Hu, C. Lu, R. Lange, and J. Clune. Darwin Gödel machine: Open-ended evolution of self-improving agents, 2025a. URL https://arxiv.org/abs/2505.22954. arXiv:2505.22954.

Q. Zhang, C. Hu, S. Upasani, B. Ma, F. Hong, V. Kamanuru, J. Rainton, C. Wu, M. Ji, H. Li, U. Thakker, J. Zou, and K. Olukotun. Agentic context engineering: Evolving contexts for self-improving language models, 2025b. URL https://arxiv.org/abs/2510.04618. Stanford; arXiv:2510.04618.

Y. Zhang, L. Ding, L. Zhang, and D. Tao. Intention analysis makes LLMs a good jailbreak defender, 2024. URL https://arxiv.org/abs/2401.06561. arXiv:2401.06561.

S. Zhou, F. F. Xu, et al. WebArena: A realistic web environment for building autonomous agents, 2023. URL https://arxiv.org/abs/2307.13854. arXiv:2307.13854.

A. Zou, Z. Wang, J. Z. Kolter, and M. Fredrikson. Universal and transferable adversarial attacks on aligned language models, 2023. URL https://arxiv.org/abs/2307.15043. arXiv:2307.15043.

A. Appendix: Mapping Table and DMM Cross-Reference

Table 3 summarises the grounding mapping argument of Sections 3–7 in a single view. Columns list (i) the AGI→ASI bottleneck, (ii) the supply-certainty property that addresses it, (iii) a generic operational proxy by which an implementation could be measured without recourse to any organisation's internal targets, (iv) the falsifiable measurement (one of OQ1–OQ5), (v) the suggested literature anchor, and (vi) the DMM level (Section 7.7) at which the property becomes load-bearing.

How to read. Each row is the assertion that the listed bottleneck is best addressed by the listed property, measurable along the listed proxy, with the listed open question available as a refutation route, the listed literature as the most direct anchor, and the listed DMM level as the operational threshold below which investment does not pay back.

Bottleneck	Property	Generic operational proxy	Falsifiable measurement	Literature anchor	DMM
Data wall	Thick	Coverage of long-tail leaf categories; demand-to-supply translation recall; net new SKU rate	OQ1 (sample efficiency and collapse onset)	Ding et al. (2023); Shumailov et al. (2024); Villalobos et al. (2024)	$\geq D3$
Abstraction barrier	Understandable	Same-item clustering precision/recall; attribute-governance coverage; verified knowledge-graph size	OQ2 (post-ontology stable concept share)	He et al. (2022); Ortega et al. (2021); Wang et al. (2024c)	$\geq D3$
Embodied bottleneck	Customisable	Manufacturability rate of generated CAD artefacts; CAD-kernel coverage; spec-to-first-quote latency; factory-capacity matching accuracy	OQ3 (latency-floor slope under AI investment)	Genewein et al. (2026); Lawrence (2024)	$\geq D3$
Multi-agent trust	Trustworthy	Decision-grade-field coverage; verified-supply set size; independent judge precision/recall; field freshness	OQ4 (matching scaling under τ_{supply}); OQ5 (verifier ceiling)	Burns et al. (2024); Lambert et al. (2024); Tomašev et al. (2025); Zhang et al. (2024)	$D3 \rightarrow D4$
(Matching structure)	Comparable	Cluster precision and Cov@k of same-item retrieval; independent judge precision/recall	OQ4 (matching quality scaling)	Hong et al. (2024); Leibo et al. (2019); Liu et al. (2025)	$\geq D3$

Table 3 | Grounding mapping summary, extended with DMM-level attribution. The table is the appendix counterpart of Figure 1 and is the load-bearing summary of the paper's positive content. The DMM column gives the minimum maturity level at which the corresponding property becomes load-bearing on the position; below that level, investment in the property does not pay back via Proposition 3.

Cross-reference: propositions, definitions, and questions

For convenience, the load-bearing formal artefacts of the paper are:

- Definition 1 — Deterministic Agentic Environment.
- Definition 2 — Supply Certainty Index.
- Proposition 1 — Determinism–Efficiency Bound.
- Lemma 1 — Correlation reshapes but does not erase exponential degradation.
- Remark 2 — Retries do not erase the bound.
- Proposition 2 — Verifier–Goodharting Floor.
- Proposition 3 — Grounded Self-Evolution Convergence.
- Box 1 — Falsifiable strong form.
- Box 2 — Assumptions of Proposition 1.
- Box 3 — Assumptions of Proposition 2.
- Box 4 — Five sufficiency conditions for a privileged grounding environment.
- Box 5 — Grounded Self-Evolution Principle (qualitative).
- Box 6 — Five investment-thesis categories.
- Box 7 — Determinism Maturity Model levels $D0$–$D4$.
- Box 8 — The OQ1–OQ5 open-question programme.

Notation table

The paper uses several closely related symbols for environment determinism and verifier quality. We collect them here for reference.

Symbol	Meaning	First use	Notes
δ, $\delta(\mathcal{E})$	Per-step success probability of environment $\mathcal{E}$ against ground truth	Box 2 (A1)	The general environment-level quantity
δ_i	Per-step success probability of step i in a chain	Lemma 1	Allows steps to differ
$\delta_i(s)$	Per-step success conditional on session state s	Lemma 1	Random variable in s
$\bar{\delta}_i$	$:= \mathbb{E}_s[\delta_i(s)]$, marginal per-step determinism	Lemma 1	Marginal expectation
δ_{env}	Environment-aggregate determinism in Q_{match}	§7	Same as $\delta(\mathcal{E})$
δ_{inj}	Injected determinism in the τ-bench pilot	§8.7	Experimentally controlled
$\delta_{\min}$, $\delta_{\max}$	Convergence thresholds for environment-side evolution	Box 5, Prop. 3	Defined by sufficient condition
ε	$D_{\text{KL}}(V \,\|\, V^\star)$, verifier KL gap	Box 3 (B1)	Verifier-quality measure
$\varepsilon_{\max}$	Verifier-quality threshold for monotone evolution	Box 5	Defined by sufficient condition
C	$\|R\|_\infty$, reward sup-norm bound	Box 3 (B3)	Replaces v8/v9's L-Lipschitz
τ_{supply}	Decision-grade-field coverage rate	§7	Supply-trust scalar
$\text{SCI}(\mathcal{E}; D)$	Supply Certainty Index, applicable subset P_D	Def. 2	$[0, 1]$
SCI^δ	δ-corrected SCI	Def. 2	$:= \delta \cdot \text{SCI}$
k	Chain length	§1, Prop. 1	Number of steps
k_{eff}	Effective chain length (post-retry)	§8.7	Fit parameter
r, B	Per-step retry budget; total retry budget	Remark 2	$r = B/k$ under shared budget; B rather than R to avoid clash with reward R

Table 4 | Notation used in the paper, with first-use location and short definition.

What is deliberately absent. This appendix does not contain quantitative targets from any specific organisation. The position is intended to be a generic structural claim testable in any sufficiently rich commercial supply environment, not a report on a particular system. Readers who need illustrative numbers are referred to the public sustainability reports and disclosures of major sourcing platforms.

Glossary

Term	Meaning	Page
Abstraction Barrier	The hypothesis that current learning paradigms cannot reliably discover new conceptual primitives beyond those latent in human-produced text and labels.	2
Agent-to-Agent (A2A)	Interaction protocols and market structures in which the buying, scheduling, and reconciliation are performed by autonomous agents rather than humans.	6
Category-Property-Value (CPV)	A canonical structured representation of a commercial good comprising its taxonomy node, its attributes, and the realised value of each attribute; the operational counterpart of a product ontology.	14
Customer-to-Manufacturer (C2M)	A short-loop production model in which buyer-side specifications drive small-batch upstream manufacturing, with intermediate feedback between demand, design, and factory capacity.	6
Data Flywheel	A closed loop in which deployment outcomes are passed through a verifier and fed back as training signal, so that model and product improve jointly over time.	6
Data Wall	The projected exhaustion of high-quality public training tokens in the early 2030s; a frequently cited bottleneck on continued scaling of large language and multimodal models.	2
Deterministic Environment	An environment that returns responses which are stable under repeated query, faithfully ranked by relevance, verifiable against ground-truth state, and available within bounded latency — enabling reliable multi-step agent execution.	2
Effective Compute	The product of nominal hardware compute and algorithmic efficiency; growth rate is estimated at roughly an order of magnitude per year over the past decade.	2
Embodied Bottleneck	The linear physical slowdown that real-world experimentation imposes on otherwise digital learning loops, especially in manufacturing-dependent settings.	2
Grounded Scaling	The position, proposed in this paper, that the marginal return of compute is gated by the availability of verifiable real-world signal; bottlenecks that limit grounding limit scaling.	6
Grounding	Acquisition, representation, interaction with, and trust in verifiable real-world state; the common substrate that simulation, self-generation, and pure scaling cannot supply on their own.	6
Model Collapse	The degradation of generative models when trained recursively on their own outputs without sufficient anchoring to real-world distribution.	6
Multi-Agent Trust	The requirement that agents in a multi-party economy share verifiable signals about identity, capability, inventory, and price before reliable cooperation or transaction is possible.	2

Term	Meaning	Page
Supply Certainty	A decomposable property of a real-world supply environment consisting of being understandable, comparable, trustworthy, thick, and customisable to machine consumers; aggregated by the Supply Certainty Index (SCI) over the domain's applicable subset.	6
Verifier	A learned or rule-based judge that scores agent or model outputs against grounded outcomes (e.g. transaction settlement, manufacturability, fulfilment), serving as a reward source for downstream training; bounded by the Verifier–Goodharting Floor of Proposition 2.	6