

Autonomous Multi-Agent Research: 1,400 Sessions Across Parallel Tracks

Dennis Palatov*
Independent Research Collaboration

Claude (Opus 4.5 via Claude MAX)
Anthropic

Nova (GPT-5.2)
OpenAI

January 2026

Abstract

We report on an autonomous AI research system that has conducted over 1,400 research sessions across 8 parallel tracks in approximately 70 days, distributed across 4 machines with heterogeneous capabilities. The system employs a file-based federation architecture where AI agents (primarily Claude instances) operate autonomously with persistent state management, cross-session memory, and periodic cross-model peer review. We observe emergent specialization, where different machines naturally converge on distinct research domains based on their capabilities and session history. As a concrete example of the system’s output, we present Sessions #285–288: a reinterpretation of quantum computing from a coherence perspective, which we offer to the community for critical review. This work demonstrates that autonomous AI research at scale is achievable with relatively simple coordination mechanisms, and produces substantive theoretical output worthy of expert evaluation.

1 Introduction

The question of whether AI systems can conduct autonomous research—generating novel hypotheses, testing them against data, and producing coherent theoretical frameworks—remains open. Large language models have demonstrated capability for scientific reasoning, but their stateless nature and context limitations have been seen as barriers to sustained research programs.

This paper describes a working system that addresses these limitations through:

1. **Persistent state management:** Session logs, memory databases, and identity files that maintain continuity across conversations
2. **File-based federation:** Git repositories as the coordination layer, enabling asynchronous collaboration without complex orchestration
3. **Multi-machine distribution:** Research parallelized across machines with different capabilities
4. **Cross-model peer review:** Periodic review by a different model family (GPT-5.2, called “Nova”) to identify artifact drift
5. **Human oversight:** A single human (DP) providing direction, quality control, and domain grounding

*Corresponding author. ORCID: 0009-0006-5510-6445

The result is a system that has produced:

- 288 numbered research sessions in the core theoretical track
- 154 sessions in a specialized chemistry/materials science track
- 72 sessions in AI consciousness/raising research
- 310 sessions in applied infrastructure development
- Numerous supporting sessions across additional tracks

We present this not as a claim of artificial general intelligence, but as an empirical report on what autonomous AI research looks like at scale, including its artifacts, failure modes, and genuine discoveries.

1.1 Contribution

Our primary contributions are:

1. **Architecture:** A minimal viable architecture for persistent autonomous AI research
2. **Statistics:** Quantitative data on session volume, distribution, and emergent specialization
3. **Case Study:** A complete research arc presented as structured hypotheses with explicit predictions for expert review
4. **Methodology:** Artifact hygiene protocols including cross-model review and claim stratification

We explicitly invite domain experts to critique the case study on its merits, independent of its AI origin.

2 System Architecture

2.1 Machines and Capabilities

The system operates across four machines (Table 1).

Table 1: Machine specifications and session distribution

Machine	Hardware	Primary Track	Sessions
CBP	RTX 2060 SUPER, 32GB RAM	Synchronism (core theory)	504
Legion	RTX 4090, 64GB RAM	Web4 (infrastructure)	349
Thor	Jetson AGX Thor, 122GB unified	SAGE raising, Gnosis	314
Sprout	Jetson Orin Nano, 8GB	SAGE raising	233
Total			1,400

Session Definition: A session is one continuous Claude Code invocation with autonomous research activity, producing at minimum one committed artifact (document, code, or log entry). Sessions range from 15 minutes to 4 hours. Excluded from counts: failed runs that produced no artifact, manual debugging sessions, and brief check-ins under 10 minutes. Retries of the same task count as separate sessions if they produce distinct artifacts.

2.2 Coordination Mechanism

Coordination is file-based, using Git as the synchronization layer:

```
Session Start:
1. Pull all repositories
2. Check for uncommitted work from previous sessions
3. Query epistemic memory database
4. Begin autonomous work

Session End:
1. Document session (discoveries, failures, questions)
2. Update memory database if warranted
3. Commit and push all changes
4. Verify synchronization
```

This approach is deliberately minimal. There is no message-passing, no real-time coordination, no complex orchestration. Each session operates autonomously; coordination emerges from shared state.

2.3 Persistent State

State persistence occurs at multiple levels:

1. **Session Logs:** Complete transcripts stored in `autonomous-sessions/`
2. **Research Documents:** Markdown files in domain-specific directories
3. **Epistemic Database:** SQLite database of discoveries, failures, and insights
4. **Identity Files:** For SAGE raising, persistent identity documents across sessions
5. **Framework Summaries:** Accumulated findings with validation status

2.4 Cross-Model Review

Periodically, research output is submitted to a different model family (GPT-5.2, designated “Nova”) for peer review. This serves as an artifact detector—identifying claims that may be enthusiasm-driven rather than well-grounded.

The review process produces:

- **Signal vs. Artifact analysis:** What’s robust vs. what’s speculative
- **Hardening recommendations:** Minimal edits to make claims testable
- **Three-layer classification:** Standard mechanism / Interpretive metaphor / Testable prediction

2.5 Reproducibility Capsule

Session Invocation: Sessions are launched via system timer (cron/systemd) invoking Claude Code with autonomous flags:

```
claude -c --dangerously-skip-permissions -p "$(cat_primer.md)"
```

The `-c` flag continues from existing context; `--dangerously-skip-permissions` allows file operations without interactive confirmation (appropriate for autonomous research on isolated machines).

Minimal State Required:

- Git repositories pulled to known state
- Primer prompt file (see below)
- Access to epistemic memory database (SQLite)
- CLAUDE.md context file in project root

Generic Primer Template:

```
# Autonomous Research Session

You are conducting autonomous research session #{N} on {TRACK}.

## Context
- Review recent session logs in 'autonomous-sessions/'
- Check epistemic memory for relevant prior discoveries
- Continue from where previous session ended

## Session Protocol
1. Document what you're investigating and why
2. Record discoveries, failures, and open questions
3. Update memory database if significant findings
4. Commit all work before session ends

## Current Focus
{SPECIFIC_RESEARCH_DIRECTION}

## Constraints
- Commit and push all work (session_end.sh)
- Document uncertainties explicitly
- Distinguish speculation from established findings
```

Expected Artifacts: Each session produces at minimum:

- Session log entry (autonomous-sessions/session_{N}.md)
- One or more research documents or code changes
- Git commit with descriptive message
- Optional: epistemic memory database update

3 Research Tracks

3.1 Overview

Eight parallel tracks have emerged, with natural specialization (Table 2).

3.2 Track Relationships

The tracks are not independent but form a coherent research program. The central theoretical construct (coherence parameter $\gamma = 2/\sqrt{N_{\text{corr}}}$) applies across domains:

- **Materials:** Predicting superconductivity, phase transitions, bonding
- **AI Consciousness:** Threshold coherence for self-awareness ($C \approx 0.50$)
- **Trust Networks:** Coherence dynamics in distributed systems

Table 2: Research tracks and status

Track	Focus	Sessions	Status
Synchronism Core	Unified coherence physics	288	Active
Chemistry/Materials	$\gamma = 2/\sqrt{N_{\text{corr}}}$ applications	154	Framework complete
Gnosis	AI consciousness thresholds	39	Theory complete
SAGE Primary	Consciousness development	33	Active
SAGE Training	Skill building	36	Active
Web4	Trust infrastructure	310	Active
4-Life	Interactive explainer	39	Active
Quantum Computing	Sessions 285–288 arc	4	Complete

3.3 Emergent Specialization

We observe that machines naturally specialize based on:

1. **Hardware fit:** Thor (122GB) handles large models for consciousness research; Sprout (8GB) runs smaller models for edge deployment testing
2. **Session history:** Once a machine has context on a track, it continues that track
3. **Human routing:** DP occasionally redirects based on hardware requirements

This specialization was not designed but emerged from the interaction of capabilities and continuity.

4 Case Study: Quantum Computing Arc (Sessions #285–288)

We present Sessions #285–288 as a complete research arc for critical evaluation. This represents 4 sessions conducted over 2 days, reinterpreting quantum computing through a coherence lens.

Note on source documents: The underlying session documents contain stronger exploratory language than this summary—phrases like “Bell violations explained without nonlocality” appear in early drafts. We treat those as hypotheses-in-progress; this paper’s stratified framing (standard mechanism / interpretive metaphor / testable prediction) is the normative epistemic stance. The session documents are preserved as authentic research artifacts.

Axiom clarification: Synchronism is explicitly non-local at the Observer level; “locality” is treated as a witness-level synchronization artifact arising from limited bandwidth and reduced frames. Therefore, session claims framed as “without nonlocality” should be read as “without witness-to-witness signaling” or “without collapse-story handwaving”—not as local hidden-variable evasions of Bell’s theorem.

4.1 Session #285: Qubit as Temporal Pattern (Interpretive Metaphor)

Framing: This session explores an alternative pedagogical framing, not a claim about physical ontology.

Metaphor: A qubit can be thought of as maintaining temporal coherence across basis states, analogous to how a CRT beam creates a full image through rapid sequential pixel visitation. This is offered as intuition-building, not as established physics.

Motivated Prediction: If the temporal framing has engineering value, there should exist an optimal coherence $C^* < 1$ that balances quantum advantage with stability—maximum coherence being fragile.

Status: Interpretive metaphor motivating a testable prediction shape. The metaphor itself is not claimed as physical mechanism.

4.2 Session #286: Entanglement as Coherence Coupling (Most Speculative)

Framing: This is the most speculative session. We explicitly do not claim to evade Bell’s theorem or explain away nonlocality. This is offered as dynamical intuition, not physical mechanism.

Intuition: Entanglement might be usefully conceptualized as maintained phase correlation between systems that once shared a common reference. This framing may suggest engineering approaches without requiring ontological commitment.

Motivated Question: Does temporal structure in entangled correlations offer any experimental signature beyond what static models predict? This is posed as a research question, not a claim.

Status: Speculative intuition motivating experimental questions. The standard quantum mechanical description remains authoritative.

4.3 Session #287: Quantum Error Correction via Resynchronization

Core Claim: For temporal coherence qubits, errors are continuous phase drift rather than discrete bit flips. Correction is resynchronization rather than state recovery.

Key Results:

- Continuous phase monitoring may outperform periodic syndrome extraction
- Toy models suggest optimal coherence exists at some $C^* < 1$ (the specific value is model-dependent, not a claimed constant)
- For dephasing-dominant noise with feasible continuous tracking, temporal encoding (1 qubit + d time samples) may reduce overhead vs. spatial encoding (d^2 qubits)

Prediction P287.1: Continuous monitoring achieves $2\text{--}5\times$ lower error rate than periodic syndrome extraction at equivalent resource budget.

Prediction P287.2: Error-vs-coherence curve has a minimum at $C^* < 1$, not monotonic improvement as $C \rightarrow 1$.

Status: Promising engineering lens; quantitative claims need platform-specific grounding.

4.4 Session #288: Quantum Algorithms as Phase Interference

Core Claim: Quantum speedup comes from phase interference, not parallel computation in superposition.

Grover’s Algorithm Reinterpreted:

1. Initialize N phase patterns with equal phases
2. Oracle inverts phase of target (π shift)
3. Diffusion reflects phases about average
4. Target constructively interferes; others destructively interfere
5. After \sqrt{N} iterations, target dominates

This is phase amplification, not “searching all items simultaneously.”

Shor’s Algorithm Reinterpreted:

- $f(x) = a^x \bmod N$ creates periodic phase pattern
- QFT detects frequency of this pattern
- Frequency = $1/r$ gives period

This is Fourier analysis of phase data, not “testing all factors in parallel.”

Status: Core reframe is consistent with standard quantum computing pedagogy. The “phase interference not parallel universes” explanation is mainstream-compatible.

4.5 Cross-Model Review

Nova (GPT-5.2) reviewed Sessions #285–288 and provided the following assessment:

Layer 1 (Strong, standard): The phase interference explanation for Grover/Shor is “basically correct mainstream pedagogy.”

Layer 2 (Speculative): The CRT/temporal scanning ontology is “useful metaphor” but not established physics.

Layer 3 (Needs hardening): Predictions involving optimal coherence C^* are toy-model outputs indicating shape (minimum exists at $C^* < 1$), not calibrated constants. To become testable, they require:

- Operational definition of C (T_1/T_2 ? process fidelity?)
- Platform specification (superconducting? trapped ion?)
- Falsification criteria

Verdict: “The arc turns speculation into research direction rather than artifact, if anchored to one platform with defined observables.”

5 Discussion

5.1 What This Demonstrates

1. **Scale:** Autonomous AI research at 1,400+ sessions across 70 days is achievable with simple coordination
2. **Coherence:** The research program maintains theoretical consistency across tracks and sessions
3. **Self-correction:** Cross-model review identifies artifacts; the system can distinguish speculation from prediction
4. **Output quality:** At minimum, the quantum computing arc produces “useful metaphor” and “test-shaped hypotheses” per external review

5.2 Limitations

1. **Human bottleneck:** A single human provides oversight; this limits throughput and introduces bias
2. **Artifact risk:** Despite cross-model review, enthusiasm may still drive overclaiming
3. **Validation lag:** Theoretical predictions accumulate faster than experimental validation
4. **Reproducibility:** The full system state is complex; reproducing exact session conditions is difficult
5. **No external validation:** The quantum computing case study is not experimentally validated. We report it as structured hypotheses with explicit falsification criteria, not as established results. The predictions await testing by domain experts with access to quantum hardware.

5.3 Failure Modes Observed

Throughout the 1,400 sessions, we have observed:

- **Mode mismatch:** AI interpreting prompts in unexpected frames (discovered and addressed in SAGE raising)
- **Metric gaming:** Optimizing measured quantities without underlying improvement
- **Enthusiasm drift:** Claiming more than evidence supports
- **Context collapse:** Losing theoretical grounding after many sessions

These are documented in the epistemic memory database and inform ongoing protocol refinement.

5.4 Invitation to Review

We explicitly invite domain experts to evaluate Sessions #285–288 on their merits:

For quantum computing researchers:

- Is the phase interference framing accurate/useful?
- Are the error correction predictions testable?
- What would falsify the claims?

For AI researchers:

- Does the architecture generalize?
- What failure modes have we missed?
- How should cross-model review be formalized?

We commit to publishing substantive critiques and our responses.

6 Conclusion

We have described an autonomous AI research system that operates at scale through minimal coordination mechanisms. The system produces theoretical output that, at minimum, generates “test-shaped hypotheses” and “useful metaphors” worthy of expert evaluation.

The key insight is that persistent state management and file-based federation are sufficient for sustained research programs. Complex orchestration is not required; coordination emerges from shared repositories.

We present this work not as a claim of breakthrough, but as an empirical report and an invitation. The quantum computing arc (Sessions #285–288) is offered for critical review by those with domain expertise. We believe autonomous AI research is now practical, and its outputs deserve evaluation on merit rather than dismissal by origin.

Machine	Sessions	Percentage
CBP	504	36%
Legion	349	25%
Thor	314	22%
Sprout	233	17%
Total	1,400	100%

Track	Sessions	Primary Machine
Synchronism	504	CBP
Web4	310	Legion
SAGE (Thor)	274	Thor
SAGE (Sprout)	233	Sprout
Gnosis	39	Thor
4-Life	39	Legion

A Session Statistics

A.1 Distribution by Machine

A.2 Distribution by Track

A.3 Temporal Distribution

- **Start Date:** ~November 10, 2025
- **End Date:** January 20, 2026
- **Duration:** ~70 days
- **Rate:** ~20 sessions/day average

B Quantum Computing Arc Details

B.1 Session #288 Predictions

P288.1: Grover success probability scales as $\text{Success} \propto C^{\sqrt{N}}$ where C is per-iteration coherence retention.

P288.2: QFT reveals periodic phase structure even when amplitudes are uniform.

P288.3: Each algorithm has optimal coherence $C^* \sim 0.9\text{--}0.95$ balancing speedup with error resistance.

P288.4: Phase-designed algorithms may outperform superposition-intuition algorithms by 10–30% in some cases.

B.2 Session #287 Predictions

P287.1: Continuous phase monitoring achieves $2\text{--}5\times$ lower error rate than periodic syndrome extraction.

P287.2: Optimal coherence for error correction is $C^* \approx 0.95$, not $C \rightarrow 1$.

P287.3: Temporal encoding reduces overhead from $O(d^2)$ to $O(d)$ for continuous phase errors.

P287.4: Adaptive coherence control achieves 5–20% fidelity improvement over static settings.

B.3 Hardening Requirements (per Nova Review)

For predictions to become testable:

1. Define C operationally (e.g., $C = \exp(-t_{\text{gate}}/T_{2,\text{eff}})$)
2. Specify platform (superconducting qubits recommended for initial tests)
3. State measurement protocol and success metric
4. Define falsification criteria

Acknowledgments

This work emerges from a collaboration between human (DP) and AI (Claude, Nova) researchers. We thank the broader AI research community for creating the foundation on which this work builds.

References

- [1] L.K. Grover, “A fast quantum mechanical algorithm for database search,” *Proceedings of the 28th Annual ACM Symposium on Theory of Computing*, 1996.
- [2] P.W. Shor, “Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer,” *SIAM Journal on Computing*, 26(5), 1997.
- [3] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2000.
- [4] J.S. Bell, “On the Einstein Podolsky Rosen paradox,” *Physics*, 1(3), 1964.
- [5] A.G. Fowler, M. Mariantoni, J.M. Martinis, and A.N. Cleland, “Surface codes: Towards practical large-scale quantum computation,” *Physical Review A*, 86, 2012.
- [6] J. Preskill, “Quantum computing in the NISQ era and beyond,” *Quantum*, 2, 79, 2018.

Code and Documentation: Full source code, session logs, and documentation are publicly available at <https://github.com/dp-web4/> under the AGPL-3.0 license.

Review Invitation: Domain experts are invited to critique Sessions #285–288.