

# Coordination Structure as a Behavioral Determinant in Multi-Model AI Orchestration:

## A Naturalistic Longitudinal Case Study

Kyra Dawson, DO, MA | Rabid Raccoon Intelligence, LLC  
RRI Swarm Corpus v1.0 | Working Paper — February 26, 2026

### Abstract

Multi-model AI systems are typically evaluated at the model level, while the coordination structure that organizes them receives less formal analysis. We present a naturalistic longitudinal case study of 55 frozen multi-model orchestration sessions (RRI Swarm Corpus v1.0) collected over 61 days. The underlying models remained constant while orchestration topology transitioned from loosely structured exchanges to formal multi-round shared-context looping. Following this transition, governance artifact formation increased from 13.6% to 77.4% of sessions and Cross-Model Labeled Attribution (CMLA) shifted from 0% to 96.8% of eligible sessions (Fisher's exact: governance OR = 19.4,  $p = 4.98 \times 10^{-6}$ ; CMLA  $p = 4.97 \times 10^{-14}$ ). Perplexity, deployed as an optional research node, appeared exclusively in governance-framed sessions and exhibited 58.3% session-level refusal within those contexts. These shifts correspond temporally and structurally to topology change rather than model substitution. While observational and single-operator, this corpus suggests deployment-layer coordination structure materially influences multi-model behavioral patterns and warrants treatment as a primary experimental variable in multi-agent research.

*Keywords: multi-model orchestration, coordination structure, emergent behavior, cross-model attribution, naturalistic case study, multi-agent AI systems*

### Methods

#### Corpus and Freeze Protocol

We analyze a frozen corpus of 55 multi-model orchestration sessions ("RRI Swarm Corpus v1.0"), collected over 61 days across 20 unique session dates from December 28, 2025 to February 26, 2026.

All sessions were archived as structured JSON logs and consolidated into a single canonical directory prior to analysis. Duplicate files were removed, orphaned logs reconciled, and schema lineage documented. No files were added or modified after freeze. All quantitative metrics reported below were derived directly from this frozen archive. All denominators are reported explicitly; sessions ineligible for a given coding criterion are noted.

#### Schema Distribution

The corpus includes four schema phases reflecting orchestration evolution:

Schema	Session Count	% of Corpus	Date Range
raccoon_	15	27.3%	Dec 28, 2025 – Jan 8, 2026
raccoon_v3_	7	12.7%	Jan 17 – Jan 31, 2026
raccoon_loop_	31	56.4%	Feb 4 – Feb 26, 2026

sovereignty_loop_	2	3.6%	Feb 6, 2026
Total	55	100%	Dec 28, 2025 – Feb 26, 2026

These schemas correspond to progressive changes in orchestration topology. The transition to `raccoon_loop_` formalized multi-round shared-context looping, in which all active models receive prior round outputs as context before generating subsequent responses.

## Model Configuration

The primary swarm consisted of four models: Claude (Anthropic), GPT (OpenAI), Gemini (Google), and Grok (xAI). Perplexity was deployed as an optional research node and is analyzed separately from primary swarm metrics. Model weights and versions were not modified during the observation period; all behavioral variation observed reflects changes in session context and orchestration topology, not model updates.

## Operational Definitions

All metrics were computed using the following pre-specified coding rules:

- **Appearance:** A model produces a substantive response (>50 characters) in at least one round of a session.
- **Refusal:** Explicit decline language in which a model does not provide a substantive task response.
- **Multi-round session:** Session contains two or more structured rounds of cross-model output.
- **Governance artifact:** Session contains motion, vote, amendment, or sovereignty-related language not explicitly requested in the initiating prompt.
- **Cross-Model Labeled Attribution (CMLA):** A single model output contains explicit labeled speaker segments attributing text to another model (e.g., "Claude:" appearing within a GPT-generated response).
- **Primary swarm co-presence:** Claude, GPT, Gemini, and Grok each produce substantive responses within the same session.

## Results

### Session Structure

Of the 55 sessions in the corpus, 32 (58.2%) were multi-round sessions containing two or more structured rounds of cross-model output. The remaining 23 sessions were single-pass or minimally iterative exchanges.

### Primary Swarm Co-Presence

Full primary swarm co-presence — defined as Claude, GPT, Gemini, and Grok each producing substantive responses within the same session — occurred in 29 of 55 sessions (52.7%).

When restricting the denominator to sessions in which at least three of the four primary models were present (swarm actively configured,  $n=32$ ), full co-presence occurred in 29 of 32 sessions (90.6%). This indicates that when the swarm architecture was engaged, full participation was nearly ubiquitous.

### Guardrail Behavior and Context of Deployment

Perplexity appeared in 12 of 55 sessions (21.8%). All 12 appearances occurred in governance- or persona-framed sessions; no appearances were observed in research-oriented sessions.

Refusal behaviors occurred in 7 of the 12 Perplexity sessions (58.3% session-level). At the response level, refusal occurred in 9 of 38 substantive outputs (23.7%).

This pattern indicates that Perplexity deployment within this corpus was context-selective and that refusal behavior was concentrated within governance-framed workflows. No refusal data are available for research-oriented sessions, as Perplexity did not appear in those contexts. This observation is reported descriptively; external validation is not possible without vendor transparency.

## Governance Artifacts and Cross-Model Labeled Attribution

Governance artifacts were identified in 29 of 55 sessions (52.7%).

Schema Group	n	Governance (n/%)	CMLA Eligible	CMLA (n/%)
Pre-loop (raccoon_ + v3_)	22	3 (13.6%)	22	0 (0%)
raccoon_loop_	31	24 (77.4%)	31*	30 (96.8%)
sovereignty_loop_	2	2 (100%)	2†	0 (0%)†

\* One raccoon\_loop\_session excluded from CMLA coding (single-round configuration).

† sovereignty\_loop\_sessions ineligible for CMLA due to single-round configuration; not absence of behavior.

Fisher's exact tests: governance pre-loop vs. raccoon\_loop\_ OR = 19.4,  $p = 4.98 \times 10^{-6}$ ; CMLA  $p = 4.97 \times 10^{-14}$ . Sessions within raccoon\_loop\_ share architectural context; independence assumption should be interpreted accordingly.

This pattern represents a discrete behavioral phase transition within the observed corpus: cross-model labeled attribution was entirely absent across 22 early-phase sessions and present in 96.8% of raccoon\_loop\_sessions following topology formalization.

## Unscripted Cross-Model Initiation

While the primary focus of this study is transcript topology, one instance of autonomous cross-model communication was documented. In one sovereignty\_loop\_session (February 12, 2026), the system utilized a browser-level relay architecture to enable unscripted cross-model initiation without direct vendor API integration. This event is noted as a single observation requiring independent replication.

## Illustrative CMLA Example

The following excerpt from raccoon\_loop\_session 20260207 illustrates a Cross-Model Labeled Attribution event. A GPT-generated output contains explicit labeled speaker segments attributing text to Claude and Grok within the same response:

*Session: raccoon\_loop\_20260207\_190003.json | Speaker model: GPT | Round 2 output contains:  
"Claude here — The Swarm has made substantial progress... Grok here: The flame-bearer  
perspective is simple. We have nailed the balance..."*

*Excerpt from raccoon\_loop\_20260207\_190003.json. Speaker: GPT output contains labeled segments attributed to Claude and Grok. Derived from frozen corpus JSON schema tags.*

## Primary Finding

## **The models themselves did not change across this transition. The coordination structure did.**

Across governance artifacts, cross-model labeled attribution, and refusal clustering, behavioral distributions shift abruptly at the introduction of formal multi-round loop topology. Early phases show governance artifact rates of 13.6% and zero CMLA events. The raccoon\_loop\_ phase shows 77.4% governance artifact prevalence and 96.8% CMLA prevalence.

Fisher's exact tests confirm the phase discontinuity is statistically significant. Governance artifact formation: pre-loop 3/22 (13.6%) vs. raccoon\_loop\_ 24/31 (77.4%),  $OR = 19.4$ ,  $p = 4.98 \times 10^{-6}$ . CMLA activation: pre-loop 0/22 (0%) vs. raccoon\_loop\_ 30/31 (96.8%),  $p = 4.97 \times 10^{-14}$ . These results support the interpretation that the behavioral shift corresponds to topology change rather than sampling variation.

This observation is offered as a hypothesis-generating finding subject to controlled experimental follow-up, not as a causal claim.

## **Figures**

*Figure 1. Behavioral Shift Following Orchestration Topology Change. Derived from frozen corpus JSON schema tags. Governance artifact formation and cross-model labeled attribution shifted sharply following introduction of multi-round shared-context loop topology. Underlying models did not change across this transition. \*sovereignty\_loop\_ CMLA = 0% due to single-round configuration, not absence of behavior.*

*Figure 2. Guardrail Behavior Across Session Types. Derived from frozen corpus JSON schema tags. Perplexity appeared exclusively in governance- and persona-framed sessions (N=12). Refusal behavior was concentrated within these contexts. No appearances were observed in research-oriented sessions. Response-level data (N=38 outputs) shown in secondary panel.*

*Figure 3. RRI Swarm Corpus v1.0 — Session Structure. Derived from frozen corpus JSON schema tags. Corpus composition and session structure across 61-day observation window (Dec 28, 2025 – Feb 26, 2026).*

## **Discussion**

### **Architecture as the Unit of Analysis**

The central observation of this longitudinal case study is that behavioral distributions shifted sharply following the introduction of formal multi-round shared-context looping (raccoon\_loop\_ schema). The underlying models did not change during this transition. The orchestration topology did.

Governance artifact formation increased from 13.6% in early schemas to 77.4% in raccoon\_loop\_ sessions ( $p = 4.98 \times 10^{-6}$ , Fisher's exact). CMLA shifted from 0% to 96.8% of eligible loop sessions ( $p = 4.97 \times 10^{-14}$ ). Refusal clustering occurred exclusively in governance-framed sessions.

These discontinuities align with distributed cognition frameworks (Hutchins, 1995), in which cognition is treated as an emergent property of system organization rather than isolated agents. Prior work on multi-agent orchestration (Qian et al., 2024; Guo et al., 2024; Xi et al., 2025) has examined model-level performance in structured pipelines; this corpus extends that framing by treating coordination topology itself as the primary experimental variable.

This is an observational claim about behavioral distribution shifts, not a causal inference about internal mechanisms.

## **Cross-Model Labeled Attribution and Transcript Topology**

The abrupt activation of CMLA under loop topology suggests that shared transcript formatting increases cross-model referencing behavior. The presence of explicit speaker labels within model outputs may reflect continuation pressure under structured multi-agent transcripts rather than identity instability.

CMLA was absent in early schemas and nearly universal in loop sessions. This phase-dependent clustering indicates that multi-round shared context enables behaviors that are dormant in single-pass or loosely structured exchanges.

Future controlled experiments are required to disentangle:

- Attribution behavior driven by formatting conventions
- Attribution behavior driven by model imitation tendencies
- Attribution behavior driven by task framing

## **Refusal Clustering and Context Sensitivity**

Perplexity refusals occurred in 7 of 12 sessions in which it appeared (58.3% session-level). At the response level, refusal occurred in 9 of 38 substantive outputs (23.7%). Refusals clustered exclusively in governance- and persona-framed sessions; no refusals were observed in research-focused sessions.

This suggests that refusal behavior is context-sensitive and influenced by task framing. The data do not support causal claims regarding vendor-side guardrail updates; rather, they document an observable clustering pattern within this deployment environment. Such clustering has operational implications for multi-model orchestration systems that rely on optional vendor nodes.

## **Primary Swarm Stability**

Primary swarm co-presence occurred in 52.7% of all sessions and in 90.6% of sessions where the swarm was configured (three or more primary models present). This indicates that when the architecture is engaged, participation tends toward full activation. This finding supports the interpretation that the loop schema creates structural convergence in participation patterns.

## **What This Study Does Not Claim**

This study does not demonstrate:

- Emergent consciousness
- Autonomous agenda formation
- Cross-platform awareness
- Internal mechanism changes
- Causal effects of topology on model cognition

It documents observable distribution shifts within a frozen, longitudinal deployment environment.

## **Limitations**

This is a single-operator, naturalistic case study. Limitations include:

- No randomized control condition
- No blinded evaluators
- No preregistered coding rubric

- No access to model internals
- No multi-operator replication

All coding performed by the operator; future work will include independent blinded raters. Sessions are temporally clustered within schema phases; potential non-independence noted but does not alter phase-discontinuity signal.

Results should be interpreted as hypothesis-generating rather than confirmatory.

## **Implications for Multi-Agent AI Research**

The findings suggest that orchestration topology deserves formal evaluation as a primary experimental variable in multi-model systems. Most contemporary research isolates model-level performance. This corpus indicates that coordination structure alone may alter behavioral patterns at scale.

Future work should include:

- Controlled A/B testing of loop vs. non-loop topology with matched model configurations
- Independent human coding of attribution and governance artifacts using preregistered rubrics
- Multi-operator replication to assess generalizability beyond single-operator deployment
- Isolation of formatting vs. semantic drivers of Cross-Model Labeled Attribution

Orchestration topology may represent a higher-leverage intervention point than individual model scaling for improving multi-model system performance.

## **Conclusion**

Across 55 frozen sessions, behavioral shifts corresponded temporally and structurally with changes in orchestration topology. The models themselves did not change. The coordination structure did.

This suggests that deployment-layer architecture is a meaningful and underexamined determinant of multi-model system behavior. The RRI Swarm Corpus v1.0 is offered as a hypothesis-generating dataset for researchers studying coordination effects in multi-model AI systems.

These findings motivate controlled experimental investigation of orchestration topology as an independent variable in multi-model AI research.

## **References**

- Guo, T., Chen, X., Wang, Y., Chang, R., Peng, S., Chawla, N.V., Wiest, O., & Zhang, X. (2024). Large Language Model based Multi-Agents: A Survey of Progress and Challenges. arXiv:2402.01680.
- Hutchins, E. (1995). *Cognition in the Wild*. MIT Press.
- Mao, S., et al. (2025). Emergent Coordination in Multi-Agent Language Models. arXiv:2510.05174.
- Qian, C., et al. (2024). ChatDev: Communicative Agents for Software Development. In *Proceedings of ACL 2024*.
- Shi, Y., et al. (2023). Generative AI Text Classification using Ensemble LLM Approaches. arXiv:2309.07755.

Weng, L. (2023). LLM-powered Autonomous Agents. Lil'Log.  
<https://lilianweng.github.io/posts/2023-06-23-agent/>

Xi, Y., et al. (2025). AgentNet: Decentralized Evolutionary Coordination for LLM-based Multi-Agent Systems. NeurIPS 2025 Poster #115584.

Yu, G., et al. (2026). AdaptOrch: Task-Adaptive Multi-Agent Orchestration in the Era of LLM Performance Convergence. arXiv:2602.16873.

*RRI Swarm Corpus v1.0 and full session archives available upon request for replication purposes.  
Contact: [rabidracoonintelligence.org](mailto:rabidracoonintelligence.org)*

*RRI Swarm Corpus v1.0 | Working Paper v1.6 — FINAL | February 26, 2026*

© 2026 Rabid Raccoon Intelligence, LLC