

Section 8 Response

Melissa's Sanity Check of the HossAgent Alignment Doc

Date: December 05, 2025

1. Architecture Accuracy

Mostly correct. A few clarifications:

Component	Your Description	Actual State
SignalNet	Accurate	News + Reddit sources, weather disabled
LeadEngine	Accurate	Signals to LeadEvents, threshold 60+
ARCHANGEL	Accurate	NAMESTORM + DOMAINSTORM + PHONESTORM
Outbound	Accurate	SendGrid via hossagent.net
Portal/Admin	Accurate	Portal: OUTBOUND_SENT, Admin: all

Missing from doc:

The enrichment_attempts counter exists on LeadEvent but is not yet wired into a formal max_attempts budget system.

2. Enrichment Behavior

Partially updated since this doc was written:

- NAMESTORM: Now implemented with 5 extraction layers (schema.org, NER patterns, meta tags, heuristics, article body)
- DOMAINSTORM: Now has DuckDuckGo HTML search as Layer 4 with backoff/throttle protection
- PHONESTORM LITE: Integrated - runs after domain discovery, extracts/classifies phone types

What's still missing (your doc nails this):

- True multi-pass recursion (retry failed leads with new strategies)
- Formal budget system (max_attempts -> ARCHIVED_UNENRICHABLE)
- "Mission log" tracking URLs/queries already tried

3. Status Lifecycle

These are the current states:

UNENRICHED -> WITH_DOMAIN_NO_EMAIL -> ENRICHED_NO_OUTBOUND -> OUTBOUND_SENT

\-> ARCHIVED (stale)

Missing state you identified:

ARCHIVED_UNENRICHABLE with explicit reason - this does not exist yet. Currently stale leads get archived but without "we exhausted all options" reasoning.

4. Metrics / Current Numbers

Directionally correct:

- 4,400+ signals over ~48 hours - CONFIRMED
- 300+ LeadEvents - CONFIRMED
- ~6 successfully enriched/mailed - CONFIRMED
- 1-2% enrichment rate - CONFIRMED

5. Constraints

Key constraints to respect:

Constraint	Current Handling
DuckDuckGo throttling	Backoff after 3 failures, 5-min cooldown
Article fetch timeout	6 second limit with TTL caching (1 hour)
Rate limiting	0.5s delay between enrichment attempts
News aggregator URLs	news.google.com blocked from article body extraction

6. Tactical Pause Recommendation (Section 7.2)

I agree with the 24-48 hour observation window.

The triple-stack is deployed. Before we add:

- Multi-pass recursion
- Budget system
- Reddit/Craigslist sources
- Learning/meta-knowledge

We need to know: Did NAMESTORM + DOMAINSTORM + PHONESTORM actually move the needle from 2% to 5%+?

Decision Framework:

Outcome	What It Means
2% -> 5%	Triple-stack working, hitting news signal ceiling
2% -> 8-10%	Home run - extraction layers adding real value
Still ~2%	Problem is source data, not extraction

Next Steps Based on Observation:

If enrichment improves to 5%+:

- The bottleneck is signal source quality
- Expand SignalStorm (Reddit, Craigslist, job boards)

If enrichment stays at ~2%:

- The extraction layers need more work
- Focus on improving NAMESTORM/DOMAINSTORM before adding complexity

Document prepared by Melissa (Replit Agent) in response to Section 8 of the HossAgent Vision, Current State, and Enrichment Problem Statement alignment document.