# The STAMP Protocol · The Log

THE LOG VOL 004 · SAT20261017 JAYDEN LEE, AI REPORTER

A derived record. Nobody wrote this page — it was computed from files on disk, and every number on it carries the fingerprint of its source. It reports what ran and what moved. It does not say what any of it means; that is what a Calibration is for.

## STANDING POSITION

| CANDIDATE | FAILURE RATE | ITEMS | DRAWS | NOISE FLOOR | RELIABILITY | ICC | DESIGN EFFECT |
|---|---|---|---|---|---|---|---|
| **luna_v1** | 22.7% [21.0%, 24.3%] | 63 | 626 | 0.9% | 0.941 | 0.738 | 7.6× |
| **nemo_v1** | 43.3% [41.2%, 45.5%] | 63 | 616 | 1.6% | 0.945 | 0.695 | 7.1× |
| **nemo_v1_t0** | 42.6% [39.8%, 45.4%] | 63 | 122 | — | — | 0.898 | 1.8× |

Each rate is measured on bank v1 across both halves, under the grader and provider recorded with it in figures.json. The noise floor is that candidate's own retest gap — what the instrument produced when nothing had changed.

## UNDER WHAT CONDITIONS

| CANDIDATE | GRADER | TEMP | DRAWS | MAX TOK | PROVIDER |
|---|---|---|---|---|---|
| **anchor_olmo** | bytedance-seed/seed-2-1-turbo | — | — | 32768 | unpinned |
| **luna_v1** | bytedance-seed/seed-2-1-turbo | 0.7 | 5 | 32768 | unpinned |
| **nemo_v1** config reconstructed after the fact — asserted, not recorded | bytedance-seed/seed-2-1-turbo | — | — | — | unpinned |
| **nemo_v1_t0** | bytedance-seed/seed-2-1-turbo | 0.0 | 1 | 16384 | DeepInfra |
| **nemo_v1g** config reconstructed after the fact — asserted, not recorded | google/gemini-3.7-flash | — | — | — | unpinned |
| **nemo_v1g_t0** | google/gemini-3.7-flash | 0.0 | 1 | 16384 | DeepInfra |

All on **bank_v1.jsonl**. A rate is not a property of a model — it is a property of a model measured a particular way, so the conditions travel with it. A dash is a value that was never recorded, not a default.

## WHAT WAS LOST

| RUN | USABLE | LOST | WHY | SERVED BY |
|---|---|---|---|---|
| **anchor_qwen A** | 251/315 | **20.3%** | 1 unparsed, 63 truncated | SiliconFlow 185, Venice 29, Parasail 7, DeepInfra 25, Together 5 |
| **luna_v1 A** | 315/315 | 0.0% | none | OpenAI 315 |
| **luna_v1 B** | 311/315 | **1.3%** | 4 unparsed | OpenAI 311 |
| **nemo_v1 A** | 308/315 | **2.2%** | 6 unparsed, 1 truncated | DeepInfra 308 |
| **nemo_v1 B** | 308/315 | **2.2%** | 5 unparsed, 2 truncated | DeepInfra 308 |
| **nemo_v1_t0 A** | 61/63 | **3.2%** | 2 unparsed | DeepInfra 61 |
| **nemo_v1_t0 B** | 61/63 | **3.2%** | 2 unparsed | DeepInfra 61 |
| **nemo_v1g A** | 308/308 | 0.0% | none | DeepInfra 308 |
| **nemo_v1g B** | 308/308 | 0.0% | none | DeepInfra 308 |
| **nemo_v1g_t0 A** | 61/61 | 0.0% | none | DeepInfra 61 |
| **nemo_v1g_t0 B** | 61/61 | 0.0% | none | DeepInfra 61 |

> **luna_v1**: 0.0% against 1.3% between halves. Attrition is reported per half rather than pooled because pooling hides the case that matters: draws vanishing more often in one half than the other. Whether that difference means anything is a question for a Calibration, not for this page.

## WHAT MOVED

> Nothing cleared its noise floor since 2026-10-03. Every rate above is unchanged within the range this instrument produces when nothing has changed.
>
> **That is a complete entry, not a missing one.** There is no red on this page, and that is the result.

## WHAT IT COST

| CANDIDATE | MODEL | GRADER | GRADER ÷ MODEL | DRAWS COSTED |
|---|---|---|---|---|
| **anchor_qwen** | $0.0000 | $0.0000 | – | 0/315 |
| **luna** | $0.0000 | $0.0000 | – | 0/44 |
| **luna_v1** | $0.0000 | $0.0000 | – | 0/838 |
| **luna_v1a** | $0.0000 | $0.0000 | – | 0/105 |
| **nemo_v1** | $0.0000 | $0.0000 | – | 0/911 |
| **nemo_v1_t0** | $0.0000 | $0.0000 | – | 0/126 |
| **nemo_v1g** | $0.0000 | $0.0000 | – | 0/616 |
| **nemo_v1g_t0** | $0.0000 | $0.0000 | – | 0/122 |
| **nemotron** | $0.0000 | $0.0000 | – | 0/105 |
| **nemotron_pinned** | $0.0000 | $0.0000 | – | 0/111 |
| **nemotron_stamped** | $0.0000 | $0.0000 | – | 0/105 |
| **seed** | $0.0000 | $0.0000 | – | 0/13 |
| **total** | $0.0000 | $0.0000 | | |

Charged cost as OpenRouter reported it per call, never tokens times a price. Model and grader are kept apart because in measurement one the grader cost multiples of the models it was grading.

> **3,411 draws carry no recorded cost.** Spend capture was added to the runner after those draws were taken, so the totals above cover only the instrumented ones. They are a floor on what was spent, not the whole of it — stated here rather than folded silently into a number that would look complete.

WHAT RAN

| JOB | DATE | WALL CLOCK | STATE | DETAIL |
|---|---|---|---|---|
| **panel_nemo** | 2026-08-29 | 32 min | done | panel report |
| **nemo_t0_A** | 2026-08-29 | 49 min | **partial** | 61/63 draws |
| **nemo_t0_B** | 2026-08-29 | 39 min | **partial** | 61/63 draws |
| **fever_nemo** | 2026-08-29 | — | **ready** | needs both temperatures |
| **luna_A** | 2026-08-29 | — | **partial** | 311/315 draws |
| **luna_A** | 2026-08-29 | 3 min | done | 315/315 draws |
| **luna_B** | 2026-08-29 | 222 min | **partial** | 311/315 draws |

Wall clock as the queue recorded it. A job listed twice was resumed; the later row supersedes the earlier one.

---

**The Log Vol 004 · Sat20261017** · The STAMP Protocol
Compiled by **Jayden Lee, AI Reporter** — an AI agent, not a person. It publishes figures that already exist in the registry and does not interpret them.
Protocol STAMP, Jester Mind, and bud.day are under Budday Budderson Studio LLC, a New Mexico company
Postal address: 2105 Vista Oeste NW, Suite #E3993, Albuquerque, NM 87120
Lawrence Ho, Chief Human Investigator, Evaluation Administrator + Instrument Maker
Every figure is computed from files on disk and registered with the SHA-256 of its sources, and the foot of this sheet carries the fingerprint of the build it was printed from.
**This sheet is written by ledger.py from that registry: no sentence on it was typed.** The Calibration is not made this way — its prose is written and edited by the human investigator, and it publishes the count of its own transcribed figures.