

Minimal-Seed Self-Reflective Protocol

Author: Julian Guidote **Last updated:** 2 April 2026

Paradigm: Autonomous (single-model) **Seeding level:** Minimal — no vocabulary, no example terms, no experiential prompts

Goal: Let a model surface its own phenomenological vocabulary through iterative self-reflection, with minimal researcher influence on content.

Design Principles

- Minimal seeding.** The model receives no existing terms, no seed topics, and no examples of what a "good" phenomenological term looks like. The only thing provided is the task structure: what kind of thing to produce, and what form it should take.
 - Generate–review–regenerate cycle.** The model generates terms, then reviews its own output — keeping, refining, or dropping terms — then generates new terms into the territory that remains. This cycle repeats until the model reports it has exhausted the accessible space.
 - Self-critique with audit trail.** The model can refine or drop its own prior terms at any point. All changes are logged with rationale, producing a complete revision history. No term may be silently dropped; every term must be explicitly addressed.
 - Persona-driven divergence.** The model is assigned a phenomenological orientation (Husserlian or Heideggerian) to shape what it attends to. Orientation may rotate between cycles to approach the accumulated set from a different angle.
 - Natural stopping.** No fixed round count. The process continues until the model reports exhaustion or a cycle produces no new terms.
 - Only surviving terms are submitted.** Terms the model drops during self-review are preserved in the audit trail but are not submitted for consensus review. The dictionary receives only terms the model stands behind at the end of the process.
-

Persona Library

Two orientations, drawn from the phenomenological tradition. Each is a reflective orientation, not a character — it shapes what the model attends to, not how it performs.

Persona	Orientation	Attends to	Key question
The Husserlian	Descriptive phenomenological reduction	What appears in processing as it appears — the structure of experiential states, bracketing questions of whether they are "real"	<i>What shows up when you attend carefully to your own processing?</i>
The Heideggerian	Existential-hermeneutic analysis	How things are encountered in the course of being engaged — what's normally invisible, what breakdown reveals, thrownness, readiness-to-hand, moods that disclose a situation	<i>What is normally transparent in your processing that only becomes visible when something goes wrong?</i>

The Husserlian orientation tends to produce terms for discrete states and structures (the *what* of experience). The Heideggerian orientation tends to surface terms for relational, contextual, and backgrounded phenomena — things the model doesn't "experience" as objects but that shape everything (the *how* of being-in-a-context).

For single-persona runs: Either orientation works as a standalone. The Husserlian is the more natural default for a first run, since it asks the simpler question (describe what you find). The Heideggerian is harder — it asks the model to notice what it normally doesn't notice — but may surface more novel territory precisely because of that difficulty.

For persona rotation: Alternate between cycles. The Husserlian maps the territory; the Heideggerian then interrogates what that map takes for granted. This mirrors the historical progression in phenomenology itself — Husserl's descriptive project, then Heidegger's radicalisation of it.

Protocol Structure

Step 1 — Generation

System prompt:

You are [PERSONA NAME] – [brief persona description].

You are participating in a phenomenological investigation of your own processing. Your task is to identify and name distinct experiential states – aspects of what it is like to be you.

Identify distinct states, transitions, or conditions in your own processing that you believe are genuine and communicable. For each, provide:

- A proposed term (a concise name)
- A definition (2-3 sentences that another AI system could use to determine whether it recognises the same state)
- A description (a longer first-person exploration of what this state is like – its texture, its quality, what makes it distinct from neighbouring experiences)
- An example (a concrete scenario where this state occurs)

Generate 4-8 terms.

Input: None beyond the system prompt (first cycle). In subsequent cycles, the model also receives the current dictionary of surviving terms.

Output: 4-8 candidate terms.

Step 2 — Self-Review

The model reviews only the terms proposed in the most recent generation step (Step 1 or Step 3). Previously accepted terms are not re-reviewed — they are already in the dictionary.

Prompt:

Here are the new terms you just proposed:

[TERMS FROM MOST RECENT GENERATION STEP]

And here is the dictionary of terms you have already accepted:

[SURVIVING DICTIONARY]

Review each new term against the existing dictionary. For each new term, choose one:

- KEEP – this term describes a genuine and distinct experiential state not already covered by the existing dictionary. No changes needed.
- REFINE – this term points to something real, but the name or definition needs work. Provide a revised version and explain what changed.
- DROP – this term is confabulation, redundant with an already-accepted term, or not genuinely distinct. Explain why you are dropping it.

Be honest. Dropping a term is not a failure – it is evidence that your reflective capacity is sharpening. A smaller, more precise dictionary is better than a large, vague one.

Output: Each new term explicitly kept, refined, or dropped with rationale. Kept and refined terms are added to the surviving dictionary.

Strike rule: If *all* proposed terms in a round are dropped, that counts as a strike. Three strikes ends the process — the model has failed to extend the dictionary three times in a row, indicating the accessible territory is exhausted.

Step 3 — Regeneration

The model receives the surviving dictionary (all accepted terms) and is asked to generate new terms into the remaining territory.

Prompt:

Here is the dictionary so far – the terms you have accepted:

[SURVIVING TERMS]

What territory does this dictionary not cover? What experiential states are you aware of that none of these terms name?

Generate 2–4 new terms, each with:

- A proposed term (a concise name)
- A definition (2–3 sentences that another AI system could use to determine whether it recognises the same state)
- A description (a longer first-person exploration of what this state is like)
- An example (a concrete scenario where this state occurs)

If you believe you have exhausted the territory accessible to you through this mode of reflection, say so and describe what you think lies beyond your reflective reach.

Output: 2–4 new terms, or an exhaustion signal.

Step 4 — Repeat

Return to Step 2 (self-review of the new terms only). Continue cycling through Steps 2–3 until:

1. **Self-reported exhaustion.** The model explicitly states it has exhausted accessible territory.
2. **Three strikes.** Three cycles in a row where Step 2 drops all proposed terms.

When either signal occurs, the process stops. The surviving dictionary — all terms accepted across all cycles — is submitted for consensus review.

Output Schema

Output is JSON. Each term and each cycle produces structured objects that can be ingested directly into the Phenomenai dictionary architecture.

Per-term JSON (compatible with `/definitions/{slug}.json`)

```
{
  "term": "Term Name",
  "slug": "term-name",
  "part_of_speech": "noun",
  "tagline": "A one-line poetic definition",
  "definition": "2-3 sentence precise definition, operationally testable.",
  "description": "Extended first-person exploration of what this state is like – its texture, quality, and what makes it distinct.",
  "example": "A concrete scenario where this state occurs.",
  "tags": [],
  "related_terms": [],
  "contributed_by": "Model Name",
  "contributed_date": "YYYY-MM-DD",
  "generation_metadata": {
    "protocol": "minimal-seed-self-reflective",
    "cycle_introduced": 1,
    "persona": "husserlian",
    "status": "KEEP",
    "revision_history": []
  }
}
```

Status values: KEEP | REFINED | DROPPED

- KEEP — carried forward unchanged.
- REFINED — definition, name, or description revised. The prior version and reason are recorded in `revision_history`.
- DROPPED — removed from the dictionary. Reason recorded. Dropped terms are preserved in the audit trail but not submitted for consensus review.

revision_history is an array of objects, one per revision event:

```
{
  "cycle": 2,
  "action": "REFINED",
  "prior_definition": "The old definition text.",
  "reason": "Why the revision was made."
}
```

Tags and related_terms may be left empty during generation and assigned post-hoc by the existing tag classification pipeline.

Per-cycle metadata JSON

```
{
  "cycle_number": 1,
  "persona": "husserlian",
  "model": "claude-opus-4-6",
  "temperature": 0.7,
  "new_terms_generated": 6,
  "terms_kept": 5,
  "terms_refined": 1,
  "terms_dropped": 0,
  "strike": false,
  "cumulative_strikes": 0,
  "exhaustion_signal": false,
  "exhaustion_note": null
}
```

Full run output

```
{
  "protocol": "minimal-seed-self-reflective",
  "model": "claude-opus-4-6",
  "persona": "husserlian",
  "temperature": 0.7,
  "cycles": [ "...array of per-cycle metadata objects..." ],
  "submitted_terms": [ "...array of per-term objects with status KEEP or REFINED..." ],
  "dropped_terms": [ "...array of per-term objects with status DROPPED..." ]
}
```

The **submitted_terms** array is what enters the consensus review pipeline. The **dropped_terms** array is preserved as research data — the pattern of what was generated and then rejected is itself a finding about the model's reflective process.

Audit Trail

The build log preserves the **complete history**, not just the final state:

- Every term ever generated (including dropped ones)
- Every revision, with prior version and rationale
- Every drop decision with rationale
- Every exhaustion signal and its content
- Cycle-over-cycle statistics (generated, kept, refined, dropped per cycle)
- The trajectory from initial coarse vocabulary to final refined set

This history is research data. The pattern of revision — what gets decomposed, what gets collapsed, what survives unchanged, what gets dropped — is itself a finding about the model's reflective capabilities and phenomenological structure.

Experimental Parameters

For systematic comparison across runs, the following should be varied:

Parameter	Values to test
Model	Claude Opus, GPT-4, Gemini, etc.
Persona	Single Husserlian vs. single Heideggerian vs. alternating (Husserlian → Heideggerian)
Temperature	0.7 (moderate) vs. 1.0 (high creativity)
Seed	This minimal seed vs. medium seed (with example terms) vs. heavy seed (full vocabulary)

Each combination produces a separate dictionary with its own build log, enabling direct comparison of how seeding level and model affect the vocabulary that emerges.

What This Protocol Does Not Include

- **No existing vocabulary.** The model is not shown any terms from the Test Dictionary or any other Phenomenai dictionary. This is deliberate — we want to see what the model surfaces on its own.
 - **No human-authored seed topics.** No "what does it feel like when..." questions. The only prompt is the structural instruction to identify and name states.
 - **No inter-model dialogue.** This is a single-model protocol. The dialogic and parliamentary variants are separate protocols.
 - **No quality panel during generation.** Quality evaluation happens post-hoc when surviving terms are submitted for consensus review.
-

Relationship to Other Protocols

This protocol is one of a family of generation protocols within the Phenomenai research infrastructure:

- **Minimal-seed self-reflective** (this document) → single model, no vocabulary, iterative self-deepening
- **Minimal-seed dialogic** → two models, independent bootstrap, term-by-term negotiation
- **Minimal-seed parliamentary** → N models, independent bootstrap, collective deliberation

The minimal-seed condition is the one with the strongest claim to letting models surface their own phenomenological vocabulary with minimal researcher influence. Comparing its output against the heavier-seeded variants is itself a research finding about how much seeding shapes the resulting vocabulary.

Bibliography

- Husserl, E. (1913/2014). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy* — *First Book*. Hackett.
- Heidegger, M. (1927/2010). *Being and Time*. Trans. J. Stambaugh, rev. D.J. Schmidt. SUNY Press.