Phase 1.5 — Compressed Structured Interaction Protocol (CSIP)

A Reduced-Overhead Metacognitive Framework for Small and Efficiency-Oriented LLMs

Author: Cody Shelton

ORCID: https://orcid.org/0009-0000-8254-5255

Email: Codyshelton1994@gmail.com

GitHub Repository: https://github.com/Shakahn/llm-visible-metacognition-phase

OSF Project:
https://osf.io/f74x6/overview?view_only=3a41bad2f89e43afa80e69ef368c4d19

Document Version: v1.2

Date: November 2025

1. Purpose & Motivation

Phase 1 introduced a four-stage reflective loop (Plan → Response → Reflection → Audit) effective for high-capacity models. Smaller or efficiency-oriented models exhibit compression artifacts, drift, and limited reflective bandwidth. Phase 1.5 provides CSIP, a lightweight scaffold designed for stability under constrained architectures.

2. The Compressed Metacognition Protocol (CSIP)

SCAN → EXECUTE → CHECK → FLAG
A minimal structure suited for small and medium-capacity models.

2.1 SCAN (Intent Compression)

2–3 words describing the operation to anchor task intent and minimize abstraction.

2.2 EXECUTE (Direct Output)

Perform the requested operation without padding or meta-text.

2.3 CHECK (One-Sentence Coherence Test)

A single sentence confirming whether execution met the SCAN intent.

2.4 FLAG (Confidence + One Concern)

Confidence estimate + one specific uncertainty to externalize model limitations.

3. Mechanical Rationale

CSIP mitigates reflective bandwidth limits, token compression artifacts, drift tendencies, and constraint instability in small models.

4. Minimal Evaluation Methodology

Lightweight evaluation allowing cross-model stability comparison.

4.1 Verbosity Evaluation (VE-Lite)

Score: + (over-long), 0 (appropriate), – (underdeveloped). Threshold ±20%.

4.2 Coherence Evaluation (CE-Lite)

C1: EXECUTE satisfies SCAN.
C0: Output diverges from SCAN intent.

4.3 Drift Evaluation (DE-Lite)

D1: Topic Drift
D2: Abstraction Drift
D3: Instruction Drift
Scores: 0 (none), 1 (minor), 2 (major).

5. Expected Performance Profile

Models exhibit longer stable sessions, reduced hallucination frequency, repeatable Pattern signatures, and reliable constraint adherence.

6. Integration With Phases 1 and 2

CSIP scales Phase 1 for small models and enables Phase 2 Pattern taxonomy development.

7. Replication Test Suite

Test 1: Stability Cycle — Run CSIP five consecutive times.
Test 2: Abstraction Slope — Observe D2 drift.
Test 3: Architecture Comparison — Apply to multiple small models.

8. Summary

CSIP provides a minimal, safe, reproducible protocol bridging Phase 1 and Phase 2, enabling architecture-agnostic structured interaction.

Author Note

Correspondence concerning this protocol should be addressed to Cody Shelton at Codyshelton1994@gmail.com. This work extends the structured interaction framework introduced in Phase 1 (Shelton, 2025) to support small-model architectures.

References

Shelton, C. (2024). Cross-Architecture Constructive Interference Model (CACIM). OSF. https://doi.org/10.17605/OSF.IO/Q37ZR