# **Technical Research Report: A Tri-Phasic Agentic System with Verifiable Citation and Iterative Feedback**

**Version:** 16.0 **Date:** 15 June 2025

### **1. Abstract**

This document details the evolution from a two-stage hierarchical agent to a complete, tri-phasic **Plan-Synthesize-Evaluate** architecture. The v15 model, while an improvement, terminated after the synthesis phase, leaving the final artifact un-audited and providing no mechanism for iterative improvement based on holistic feedback. This final architecture addresses these omissions by introducing two critical innovations: **Grounded-Source Citation (GSC)** and a **Supervised Evaluation & Refinement Loop**.

The synthesis phase is now mandated to produce in-line citations that trace every claim back to the source documents in the knowledge base. Following synthesis, a dedicated **Evaluation Graph**, powered by a constitutionally-governed committee of independent LLMs (e.g., DeepSeek), performs a rigorous audit. This audit checks for logical coherence, verifies the accuracy of the citations, and assesses the artifact's alignment with the user's strategic goals. The resulting evaluation report is presented to the user, who can then provide their own feedback and choose to either accept the artifact or trigger a stateful, full-cycle restart. This new loop feeds the combined evaluation and user feedback into a new planning phase, allowing the agent to learn from its mistakes. This tri-phasic design with a human-in-the-loop feedback cycle transforms the agent from a linear generator into a truly iterative and accountable research partner.

### **2. Research: The Need for Verifiability and Iteration**

The Supervised Intent-Planning of v15 solved the problem of *strategic* alignment but did not address two subsequent challenges:

* **Factual Grounding:** An agent could flawlessly execute a user-approved plan but still "hallucinate" facts or misrepresent sources during the synthesis phase. Without a mechanism to verify claims against the knowledge base, the final artifact, while structurally correct, could not be trusted. This is the problem of verifiability.
* **Single-Shot Limitation:** The v15 workflow was a "single-shot" process. If the final output contained subtle flaws or failed to meet the user's higher-level goals, the only recourse was to start the entire process over with a manually re-engineered prompt. The system itself had no capacity to learn from its final performance. This is the problem of iteration.

These limitations prevent the system from emulating a true deep research process, which inherently involves cycles of drafting, feedback, and revision.

### **3. The v16 Architecture: Plan, Synthesize, Evaluate**

The v16 architecture completes the cycle by introducing a distinct third phase, creating a stateful, iterative loop that prioritizes verifiability and continuous improvement.

**Phase 1: Supervised Planning (Unchanged from v15)**

* **Objective:** Produce a user-approved plan.md.

**Phase 2: Grounded-Source Synthesis**

* **Objective:** Execute the approved plan to produce an artifact with verifiable, in-line citations.
* **Innovation: Grounded-Source Citation (GSC):** The synthesis agent is now prompted with a strict requirement to cite its sources for every major claim or piece of synthesized information. Citations will be in a machine-readable format (e.g., [Source: document\_name, chunk\_id]). This makes the agent's reasoning process transparent and its outputs auditable.

**Phase 3: Independent Constitutional Evaluation**

* **Objective:** To perform a multi-faceted audit of the synthesized artifact and produce a detailed evaluation report for the user.
* **Innovation: Independent Evaluator Committee:** This phase is executed by a separate committee of agents using a different LLM family (e.g., DeepSeek) to prevent self-confirmation bias. This Evaluation Graph performs three tasks:
  1. **Logical Consistency Check:** Assesses the internal coherence and narrative flow of the document.
  2. **Citation Fact-Check:** Parses all GSC tags, retrieves the corresponding source text, and verifies that the synthesized statement is factually supported by the cited source.
  3. **Goal-Alignment Review:** Assesses whether the final artifact successfully meets the strategic, higher-level goals outlined in the original user prompt.

**The State-ful Feedback Loop** The output of the Evaluation Graph is not the end of the process. It is presented to the user alongside the synthesized artifact. The user is then given two choices:

1. **Accept Artifact:** The process concludes.
2. **Restart with Feedback:** The user can add their own comments to the evaluation report. This combined feedback is then injected back into the GraphState, and the **Planning Graph** is invoked again. The planning agents are now tasked with creating a *new, improved plan* that explicitly addresses the critiques from the previous cycle.

This architecture directly solves the problems of verifiability and iteration, creating a robust, accountable system that learns and improves through a combination of automated auditing and human supervision.