# **Technical Research Report: A Hierarchical Agentic System with Supervised Intent-Planning**

Version: 15.0

Date: 15 June 2025

### 1. Abstract

This document details a paradigm shift in the agent architecture, moving from a monolithic generation process to a hierarchical, two-stage framework incorporating Supervised Intent-Planning. Previous versions (v12-v14), which relied on implicit, single-turn planning, consistently failed to produce high-quality, grounded artifacts. These failures stemmed from a fundamental flaw: the agent would begin synthesis without a validated, coherent, and user-aligned plan. This resulted in ungrounded, generic outputs and a complete lack of user control over the agent's strategic direction. The v15 architecture corrects this by formally separating the cognitive tasks of Planning and Synthesis into two distinct, constitutionally-governed computational graphs. The initial "Planning Graph" is responsible for generating a detailed, grounded table of contents or strategic plan. Crucially, this plan is treated as a primary artifact that is presented to the human user for review and approval. Only after explicit user supervision and consent does the system proceed to the "Synthesis Graph," which executes the user-approved plan. This Human-in-the-Loop (HITL) approach ensures agent actions are aligned with user intent from the outset, dramatically improving the relevance, quality, and trustworthiness of the final generated artifact.

### 2. Research: The Failure of Implicit Planning

The core failure of the previous "Plan, Synthesize, Review" (PSR) architecture was the conflation of planning and synthesis within a single, uninterrupted agentic process. The agent was instructed to *create a plan and then immediately execute it*. This revealed several critical research problems:

* Unaligned Intent: The agent's internal, un-reviewed plan often diverged from the user's true intent, even if it seemed plausible to the LLM. The agent would then spend significant resources generating a high-quality version of the *wrong artifact*.
* Contextual Detachment: Without a validated plan to anchor its reasoning, the agent's synthesis process would often "drift" from the provided knowledge base, resulting in the generation of generic, hallucinatory content (e.g., the fictitious "Experimental Evaluation" section).
* Irrecoverable Errors: A flaw in the initial, un-reviewed plan would cascade through the entire synthesis process, making the final output unsalvageable and wasting all subsequent computational effort.
* Lack of User Agency: The user, acting as the project lead, was relegated to a passive observer, unable to steer the agent's high-level strategy or correct its course before it was too late.

These findings make it clear that for complex, high-stakes generation tasks, planning is not merely the first step of execution; it is a distinct cognitive process that requires its own validation and supervision.

### 3. The v15 Architecture: Supervised, Two-Stage Generation

The v15 architecture remedies these failures by introducing a hierarchical, two-stage workflow with a human-in-the-loop validation step. The system is now composed of two independent, constitutionally-governed agentic workflows, orchestrated by the Supervisor.

Stage 1: Constitutional Planning Workflow

The user's initial prompt triggers this first stage.

* Objective: To produce a high-quality, grounded, and debated Plan Artifact (e.g., a plan.md file containing a detailed table of contents).
* Process:
  1. The supervisor invokes a dedicated Planning Graph.
  2. Two constitutional agents (A and B) independently generate a proposed plan based on the user's prompt and the retrieved knowledge base.
  3. The agents engage in a constitutional debate, critiquing each other's plans for logical structure, completeness, and grounding in the source material.
  4. The agents refine their plans based on peer critique until a converged, high-quality plan is produced.
* Output: The final, debated plan is saved and presented to the user. The entire system then halts and waits for user input.

Stage 2: Supervised Synthesis Workflow

This stage only begins after the user explicitly approves the plan generated in Stage 1.

* Objective: To execute the user-approved plan and generate the final artifact.
* Process:
  1. The supervisor, upon receiving user approval, invokes a separate Synthesis Graph.
  2. The user-approved Plan Artifact is loaded as the primary "ground truth" instruction in the GraphState.
  3. Constitutional agents execute the plan, generating the content for each section.
  4. The system performs a final constitutional debate on the assembled artifact to ensure coherence and quality.
* Output: The final, high-quality artifact that is verifiably aligned with both the knowledge base and the user's approved strategic direction.

This explicit separation provides the necessary control, grounding, and user agency that was critically missing from all previous versions. It re-establishes a logical project workflow where strategy is confirmed before execution begins.