# **Technical Research Report: A Simplified Two-Agent, Teacher-Student Architecture for Robust, Auditable Research**

Version: 17.0

Date: 15 June 2025

### 1. Abstract

This document outlines a fundamental and decisive pivot in the agentic system's architecture, moving from a complex multi-agent committee to a simplified, more robust Teacher-Student model. Previous iterations (v8-v16) attempted to simulate collaborative refinement using a "constitutional debate" between two instances of the same LLM family. This approach proved to be computationally expensive, logically flawed, and produced noisy, unpredictable outputs. It fundamentally misinterpreted the nature of scientific and technical generation, which **requires precision and verifiability over creative debate.**

The v17 architecture corrects this by simplifying the agent structure and clarifying their roles. The system now employs two distinct, single agents: a Student (e.g., Gemini), responsible for the primary generation tasks of Planning and Synthesis; and a Teacher/Auditor (e.g., DeepSeek), responsible for the independent Evaluation. **The noisy, high-temperature "creative" agent is eliminated entirely.** This architecture reframes the process not as a debate, **but as a formal audit**. The Student's work is checked against the ground truth by an independent party, the Teacher, who then provides structured feedback. This model, combined with the established Plan-Synthesize-Evaluate loop and user supervision, creates a system that is simpler to implement, more efficient to run, and produces a final artifact that is significantly more trustworthy and accountable.

### 2. Research: The Failure of the Constitutional Committee

The core assumption of the "constitutional committee" was that a debate between two similar AIs would lead to a higher-quality synthesis. This proved to be incorrect for non-creative, fact-based tasks. The research identified several key failures of this model:

* Noise Amplification: Using a high-temperature "creative" agent for critique introduced stochastic, often irrelevant, and "chatty" feedback. This degraded the quality of the final artifact rather than improving it.
* Redundant Computation: Running two parallel, complex graphs for every stage (Planning, Synthesis) was highly inefficient and doubled the time-to-result without a proportional increase in quality.
* Lack of True Independence: While using two instances, they were still from the same model family (Gemini). This created a cognitive echo chamber, where the models shared the same inherent biases, training data limitations, and architectural blind spots. The "critique" was not a true independent audit.
* Misaligned Incentives: The adversarial nature of the "debate" was ill-suited for a collaborative goal. The system should be framed as a cooperative process between specialized agents, not a competition.

### 3. The v17 Architecture: A Teacher-Student Framework

The v17 architecture addresses these failures by radically simplifying the agentic structure and clarifying the relationship between the components. It fully embraces the three-graph Plan-Synthesize-Evaluate loop defined in v16, but executes it with single, specialized agents.

* The Student (Planner/Synthesizer): A single, low-temperature LLM (e.g., Gemini) is responsible for all generative tasks. It executes the planning\_graph to create the initial plan and the synthesis\_graph to generate the cited artifact. Its behavior is governed by a System Prompt that instructs it on its role and its collaborative relationship with the Teacher.
* The Teacher (Auditor/Evaluator): A single, low-temperature LLM from a *different model family* (e.g., DeepSeek) is responsible for all evaluation tasks. It executes the evaluation\_graph. Its role is not to debate, **but to perform a formal, dispassionate audit of the Student's work against three criteria: logical consistency, factual accuracy (citation checking), and goal alignment.**
* The Shared Constitution (System Prompt): The concept of a constitution is retained but repurposed. Instead of governing a debate, it now defines the roles and collaborative protocol for the Teacher-Student interaction. It is provided as a system prompt to both LLMs, ensuring they understand the overall workflow and their specific place within it. It explicitly frames the relationship as cooperative, not adversarial.

This simplified model eliminates redundant code and computation, enhances robustness by using low-temperature agents, and introduces true independent verification. This shift moves the project's complexity away from inefficient agent orchestration **and towards the more powerful and flexible domain of prompt engineering.**