# **Technical Design Specification: Professional-Grade Teacher-Student Agent**

Version: 3.3 (Revised)

Date: 15 June 2025

Status: Proposed (Based on Technical Research Report v18)

### 1. Architectural Overview

This specification details a professional-grade, tri-phasic agentic system based on a Teacher-Student model. The architecture is defined by two core upgrades over v3.2: 1) The Evaluation graph is now tool-augmented, and 2) Key state artifacts are exchanged as JSON objects.

* Workflow: Planning (Student, outputs JSON) -> HITL Plan Approval -> Synthesis (Student, outputs Markdown with citations) -> Evaluation (Teacher, uses tools, outputs JSON) -> HITL Evaluation Review & Optional Restart.
* Supervisor (supervisor.py): Orchestrates the workflow. It now reads and writes JSON artifacts from the agent graphs, improving reliability. It will provide the citation\_retriever tool to the Evaluation graph.
* Agent Core (agent\_core.py): Contains three simplified, single-agent graph builders. The build\_evaluation\_graph function will now be configured to use tools.
* Configuration (prompts.yaml): Prompts will be updated to instruct agents to read and write JSON where appropriate.

### 2. Artifact Specification: JSON Schemas

#### 2.1. plan.json

The planning\_graph will output a JSON object with the following schema:

{

"title": "Proposed Report Title",

"plan\_items": [

{

"section\_id": "1.0",

"title": "Introduction",

"description": "Outline the core problem and the system's approach."

},

{

"section\_id": "2.0",

"title": "Foundational Concepts",

"description": "Explain the key theoretical underpinnings."

}

]

}

#### 2.2. evaluation.json

The evaluation\_graph will output a JSON object with the following schema:

{

"overall\_consistency\_score": 0.9,

"consistency\_notes": "The document flows well, but section 3 contradicts a point made in the introduction.",

"goal\_alignment\_check": "PASS",

"goal\_alignment\_notes": "The artifact successfully addresses the user's primary request.",

"citation\_audit": [

{

"claim": "The system uses a tri-phasic workflow.",

"citation\_tag": "[Source: TRR-v16.md]",

"retrieved\_source\_text": "This document details the evolution... to a complete, tri-phasic Plan-Synthesize-Evaluate architecture.",

"verification\_status": "PASS"

},

{

"claim": "The system uses a particle filter.",

"citation\_tag": "[Source: TRR-v15.md]",

"retrieved\_source\_text": "The agent uses a stateful, hierarchical model.",

"verification\_status": "FAIL"

}

]

}

### 3. Computational Graphs

#### 3.1. build\_planning\_graph() (Student)

* Prompting: The create\_plan\_prompt will now explicitly instruct the LLM to generate its output in the specified JSON format.
* Output: A plan.json artifact.

#### 3.2. build\_synthesis\_graph() (Student)

* Input: The plan.json object.
* Logic: The graph will iterate through the plan\_items list from the JSON object to generate the report section by section. The requirement to embed [Source: ...] tags in the Markdown output remains.
* Output: A report\_with\_citations.md artifact.

#### 3.3. build\_evaluation\_graph() (Teacher with Tools)

* Input: The report\_with\_citations.md and a citation\_retriever tool provided by the Supervisor.
* Tools:
  1. citation\_retriever(doc\_name: str) -> str: A tool that takes a document name from a citation tag and returns its full text content from the knowledge base.
* Logic:
  1. Parse the report to extract claims and citation tags.
  2. For each citation, the LLM must invoke the citation\_retriever tool with the doc\_name.
  3. The LLM then compares the claim to the text returned by the tool to determine PASS or FAIL.
  4. The graph assembles the final evaluation.json object.
* Output: An evaluation.json artifact.

### 4. Supervisor and UI Logic

* The Supervisor will load the plan.json and evaluation.json artifacts and parse them directly.
* The UI will be updated to render these JSON objects in a clean, human-readable format (e.g., using st.json or by formatting them into tables and lists). This ensures the user sees a clear, structured breakdown of the plan and the audit results, enhancing their ability to provide targeted feedback for the next iteration.