

Empowering Experiential Learning through an AI-Driven Agentic Design Framework for Engineering Labs

CSU-AIEIC Final Report

Author: Franz Kurfess, Cal Poly San Luis Obispo

Project Summary, Findings, and Lessons Learned

Project Rationale and Mission

This project was established to bridge the pedagogical gap between passive Generative AI (GenAI) consumption and the active system-level reasoning skills required in modern engineering. While large language models (LLMs) have become ubiquitous, their integration into the curriculum often lacks the depth needed for architectural design. Based on Cal Poly's "Learn by Doing" philosophy, this project developed an Agentic Design Lab Framework, with the Computer Science and Software Engineering (CSSE) and Industrial and Manufacturing Engineering (IME) departments as initial users, empowering students to transition from users to architects, designing autonomous "agentic" systems capable of perceiving environments and executing independent, goal-oriented decision-making.

Summary of Technical Accomplishments

By the end of the funding period in late July 2026, the project achieved several technical milestones. We deployed an LLM-based grading system to evaluate the logic and architectural coherence of student-developed agents. A primary technical accomplishment was the consolidation of fragmented project databases into a unified repository. This consolidation was a response to student feedback during early pilots, where broken links and disconnected directories hindered agent performance. By providing a centralized repository for all course materials, we significantly increased data pipeline throughput and enabled more stable multi-agent orchestration using AutoGen, CrewAI, and LangGraph. Another technical accomplishment is a design revision to replace a complex orchestrator with a much simpler three-phase harness model reflecting the activities before, during, and after a lab session. The main pre-lab activities involve curriculum generation and environment setup. During the lab, the lab companion offers real-time support for students, the participant agents track student progress and observe their thinking patterns, and the integrity agent enforces policies and student behavior expectations. If needed, a safety monitor provides resource oversight to make sure that participants do not endanger themselves or others. The post-lab phase includes assessment, analytics, and reflection on the overall lab. This separation is not strict, and agents will be invoked as needed. For example, the assessment agent may perform some grading activities during the lab in order to give students quick feedback on their progress. Our work on this project was inspired by the use of agentic design principles in the AI for Search and Rescue project. As a consequence, our initial design for this project included an Orchestrator component, whose responsibility is to determine the allocation of tasks to specific agents. As our project evolved, we realized that in this case, this orchestrator agent had very limited responsibilities. Simultaneously, agentic design principles evolved from the use of complex architectures and prompts towards the use of "skills"

and "harnesses" to allow for simpler designs. This led us to the use of the "three-phase harness" reflecting the time periods during which related activities and tasks are typically done. One key observation by the developers and test users is that the harness version behaves in a more predictable manner than the original complex orchestrator.

Key Research Findings

Pilot implementations in CSC 480 (Introduction to AI) and CSC 580 (Graduate AI) provided initial data on student engagement. Findings indicate that students successfully moved beyond single-prompt interactions to orchestrate complex multi-agent systems within multiple domains, in particular the Search and Rescue (SAR) project some students were already familiar with. Specific successes were noted in the development of specialized roles, including the ClueMeister and MissingPersonProfiler. Evaluation of student work showed a 40% improvement in self-reported skills related to distributed systems. Furthermore, students demonstrated a sophisticated ability to utilize the Model Context Protocol (MCP) to connect autonomous agents with real-world tools, achieving high levels of system coherence. We also encouraged students to stress-test the system, in particular the lab companion. This agent contains rudimentary integrity checks, but students quickly found gaps and ways to overcome the built-in constraints. The integrity agent replaces these checks and includes more sophisticated analytical functions, such as a 15-query throttle or 85% similarity flag to identify suspicious behavior. It also includes a more graduated response to students, ranging from mild hints over warnings to alerts requesting instructor intervention.

Lessons Learned

- **Infrastructure Reliability and Data Consolidation:** Technical friction, specifically regarding incomplete information and fragmented data sources, serves as a significant barrier to learning. The July database consolidation was essential to move students from debugging data-access errors to focusing on high-level agent logic and reasoning loops.
- **Scaffolded Complexity for "Black Box" Systems:** Transitioning to agentic design requires intentional scaffolding as a didactical foundation. Students initially struggle with the non-deterministic nature of LLM reasoning; providing structured role-based tables and clear key performance indicator definitions is vital to ensure students can effectively debug and refine agent behaviors.
- **Ethics as a Practical Design Constraint:** Integrating the ETHICAL Principles AI Framework transformed ethics from an abstract concept into a functional design constraint. When designing for high-stakes domains like SAR, students recognized that data privacy and bias in victim location are technical parameters that must be addressed within the agent's architecture to ensure a robust and responsible system.

Response to CSU Community Learning Prompt: *"What is one thing you would most like the CSU community to learn from your project?"*

Using LLMs to enhance the capabilities of agents greatly enhanced their relevance for practical applications. Our framework demonstrates that agentic design provides a solid basis for lab assignments in the Engineering disciplines that follow Cal Poly's "Learn by Doing" philosophy and foster analytical and problem-solving skills in students.

Evaluation Findings

The main findings for this report are based on data collected during the Fall 2025 and Winter 2026 quarters. In the CSC 480-F25 Introduction to AI course, we redesigned eight out of ten labs to use agentic

design principles. The first two labs familiarized students with AutoGen, the development environment used for the remaining labs. Students then implemented agents to apply systematic search methods like A* and Monte-Carlo Tree Search to solve the [NY Times Spelling Bee](#) puzzles. One interesting observation about the textbook approach to define a generic search function and then modify it through parameters emerged through the use of agents: For simple search algorithms (depth/breadth-first and similar), an agentic approach is impractical since the basic computational steps are very simple. In more complex situations, like Monte-Carlo Tree Search, an agentic approach appears very "natural" and is well suited. There are also intermediate scenarios, where the suitability depends on factors such as the complexity of the heuristic in A*. Later, the students developed individual agents to assist with the search for missing persons and coordinated work among multiple agents as teams. In the CSC 580W26 graduate AI class, students tested some of the revised CSC 480 labs in combination with a "lab companion" agent. This agent has access to the course material provided to the students (including excerpts from the textbook) and uses scaffolding to offer hints, but not solutions, to students. Especially students not familiar with AutoGen appreciated the assistance from the lab companion, as it saved them the effort to investigate related questions on their own. Some students with gaps in the prerequisite knowledge also relied on the lab companion for help and overall were happy with its performance.

Objective Metric Performance

While these measures depend on the course topic and particular lab activities, the table below gives an example of objective metrics in CSSE labs with coding as the main activity. The initial results reported here are from selected CSC480-F25 labs. Due to the small sample size, a large variation in the familiarity of students with agentic design principles, and significant support from the teaching assistant and instructor, the number may not be reflective of a broad computer science user group.

<i>Metric Category</i>	<i>Measurement Method</i>	<i>Observed Outcome/Target</i>
Task Completion Rate	Automated test suites with predefined success criteria	92% (Target: ≥90%)
Code Quality	Static analysis scores via Pylint/modularity metrics	8.8/10 (Target: >8.0)
Robustness	Error recovery success rate under stress-test conditions	85% Recovery (Target: ≥80%)
API Compatibility	% of lab interfaces and external tools supported	98% (Target: ≥95%)
Scalability	Support for concurrent agent instances	60+ Parallel Agents (Target: ≥50)

Subjective Educational Outcomes

1. **Perceived Learning Outcomes:** 88% of students reported high confidence in their ability to design autonomous agents on a Likert scale.
2. **Developer Experience (DevEx):** Survey results indicated that while the abstraction levels of AutoGen and CrewAI were appropriate, the initial learning curve remains steep; however, the

consolidation of data sources significantly improved the perceived effectiveness of debugging tools.

3. **Instructional Alignment:** Initial feedback from faculty confirmed that the framework successfully deepened conceptual understanding of system-level reasoning and autonomous reasoning loops in students. While some of the labs turned out to be unsuited for agent-based implementations, students saw that as a learning experience for situations where agents are not the best approach.

Instructional Assessment and ABET Alignment

Initial faculty assessments focused on the "Architectural Elegance" and "Error Handling Creativity" of student-designed agents. Crucially, the evaluation criteria for the SAR domain were mapped directly to ABET student outcomes, specifically those addressing "system-level reasoning" and "complex problem solving" in engineering contexts. Rubrics indicated that students who were working on the PathExplorer and ClueMeister roles in the Search and Rescue domains demonstrated a strong ability to manage multi-objective constraints, satisfying the rigorous standards for engineering accreditation. While all students succeeded in the creation of basic agent configurations, some of them struggled with more challenging tasks, such as the effective use of Model Context Protocol (MCP), Agent-to-Agent Communication (A2A), and Redis publish-subscribe message queues.

Our initial goal was to conduct experiments with students in multiple CSSE and IME classes during Cal Poly's Spring 2026 quarter. Administrative delays, in particular with the availability of funding to pay students during the 2025 summer, and stipends to instructors disbursed as late as June 2026, led to the postponement of experiments to the Fall 2026 semester.

Showcase Artifacts and Assets

The results of the project were demonstrated at multiple showcase events.

- **Cal Poly AI Showcase on November 6, 2025.** The team had two posters on display, one for the project "AI-Driven Agentic Framework for Engineering Lab Learning" and the other for "AI in Search and Rescue".
- **CSU AIEIC Teaching with GenAI on April 23, 2026:** Six team members gave a presentation and demoed the project "AI-Driven Agentic Framework for Engineering Lab Learning".
- **CSU AI in July Showcase Overview on July 29, 2026.** The event featured a presentation by three team members on the project "AI-Driven Agentic Framework for Engineering Lab Learning".

Featured Artifacts

All assets are hosted on the GitHub repository <https://github.com/PKalathas/CSUAIEI> and developed under open-source licenses. In its current state, the system is not suitable for direct reuse. Instructors with Python skills can clone the GitHub repository and update the .env file with their respective OpenAI or AWS Bedrock API credentials. Interested parties are encouraged to contact the team for further details. The system will be under ongoing development for the foreseeable future. Our plan is to make a stable version available through the CSU AI Commons. To do so, however, we need to secure resources to support scaling beyond the current set of courses, and ultimately across the 23-campus system. We have reached out to Cal Poly's DxHub, the AWS Trainium Educational Support program, Microsoft's Azure Educational Grants team, and other contacts to help with this.

- **Agentic Design Framework Library:** A curated set of open-source Python modules and configuration templates. It is currently under development and intended for systemwide replication via the CSU AI Commons.

- **Agentic Design Framework Prototype:** A multi-agent system featuring instructor and participant dashboards, an orchestration manager, and multiple task-specific agents (curriculum design, participant support, assessment, integrity agents).
- **Pedagogical Infographics:** Visual assets of the system architecture, dashboard design, individual agents, and a mapping of the "ETHICAL Principles AI Framework" to specific design steps, ensuring responsible development is embedded in the student workflow. **Collaboration Assets:** While about a dozen students contributed to the overall project, four student research assistants integrated all components to ensure that the library is documented for student-to-student peer learning.

Updated Public-Facing Abstract

Our project, "Empowering Experiential Learning through an AI-Driven Agentic Design Framework," enables students to architect autonomous multi-agent systems for complex domains with a primary focus on Engineering disciplines. It prepares students for the AI-driven workforce by integrating technical proficiency with the "ETHICAL Principles AI Framework," ensuring that the next generation of engineers are both empowered and responsible developers. Inspired by Cal Poly's "Learn by Doing" approach, the project team developed a multi-agent system framework together with a library of open-source modules, enabling multiple agents to work as a team. In the first phase supported by the CSU-AIEIC initiative, the team implemented the core agent functionality plus dashboards for instructors and participants in lab activities. Proof-of-concept testing took place in the Fall 2025 and Winter 2026 quarters; in these classes, students first implemented agents for basic AI concepts, and then expanded these agents into scenarios where multiple agents worked together on Search and Rescue missions to find a missing person. Tests with a wider selection of courses will continue during the Fall 2026 and Spring 2027 terms.

Final Budget Report

Expenditure Summary Table

Category	Allocated Funds	Reconciled Spend
Faculty & Staff Salaries (Stipends/Buyouts)	\$35,000	\$28,500
Student Research Assistants (Co-developers)	\$15,000	\$35,271.38
Materials & Supplies (API fees, licenses)	\$5,000	\$0
Operational Costs (Cloud Services/AWS)	\$5,000	\$0
Total	\$60,000	\$63,771.38

Fiscal Narrative

We confirm that all \$60,000 of the awarded CSU-AIEIC funds were fully expended by the June 30, 2026 deadline. The total expenses for our project amounted to \$63,771.38, with a \$3,771.38 overage in student payroll that was covered by subsequent funding sources (Cal Poly Noyce Grant) to ensure no interruption in student work. We didn't incur any expenses for Materials & Supplies or Operational Costs, as we were able to use in-kind donations by companies (Microsoft), research grants (NCSA ACCESS), and access to high-performance computing resources (NRP). We also made some adjustments to the faculty stipends in order to increase funds available for student salaries. Due to availability constraints, we had to reduce the number of instructors involved and used these funds for student support. In the end, we had expenses for the two Co-PIs (Jenny Wang - IME 312; Paris Kalathas - CSC 203), who were also responsible for research guidance and administrative support. The PI (Franz Kurfess - CSC 480/580) is participating in the Faculty Early Retirement Program (FERP) and not eligible for additional compensation; he volunteered his time on the project. There were six instructors: two from CSSE (Maria Pantoja - CSC 469/569; Kirk Duran - CSC 349), and four from IME (Xuan Wang - IME 432; Trian Georgiu - IME 101; Bryan Lutz - IME 335 & IME 336; Martin Koch - IME 101).

While the total expenditure of \$63,771.38 exceeds the Tier 2 funding limit of the CSU-AIEIC, we are using subsequent funding sources for the overage. All funds were utilized in compliance with AIEIC guidelines: no indirect costs were charged, and no funds supported hardware or travel. Faculty salaries supported the extensive curriculum redesign and the development of the Agentic Design Framework, while student assistant funds supported four co-developers. Their responsibility included supervision of students working on this project for credit (e.g., Senior Projects) and volunteers who were in it for the experience. Continued funding for this project has been made possible by a grant from the Cal Poly Noyce School of Applied Computing; in-kind support in the form of computing resources is anticipated from cloud service providers (Amazon AWS, Microsoft Azure).

Infrastructure Reflections: AWS and NVIDIA AI Factory

Cloud Computing Support

The initial implementation utilized Microsoft's Azure cloud infrastructure, supported by a grant from Microsoft Education. Due to administrative and technical reasons, we decided to use Amazon's AWS going forward. The transition to AWS infrastructure is vital for maintaining the scalability required for classroom-wide simulations. AWS supported the simultaneous execution of multiple parallel agent instances, providing the necessary data pipeline throughput for real-time interaction with students from multiple classes. This high-availability environment ensures that students can test even complex setups for lab activities without performance degradation.

NVIDIA AI Factory Integration

The NVIDIA AI Factory offers a powerful computational backbone for the project's heavy LLM processing and reinforcement learning (RL) tasks. Supplemented by an initial NCSA ACCESS computing resources grant and access to the National Research Platform, this infrastructure enables the low-latency reasoning loops required for agents to adapt to dynamic changes in the varying lab activities and environments. During initial testing, student teams used this capability to implement real-time agentic adaptation for Search and Rescue scenarios, e.g. to weather shifts and new clues about a missing person.

Future-Proofing

This computational foundation supports the longer-term objective of developing adaptive agents that learn from previous experiences and observations of student activities. The integration of high-performance

cloud and GPU resources ensures the framework remains compatible with the next generation of autonomous AI systems. Further testing across different disciplines and in a variety of lab settings will serve to enhance its functionality and fine-tune its performance.

References and Citations

- Acharya, D. B., Kuppan, K., & Divya, B. (2025). Agentic AI: Autonomous Intelligence for Complex Goals—A Comprehensive Survey. *IEEE Access*, 13, 10.1109/ACCESS.2025.3532853.
- California State University. (2025). *The CSU Artificial Intelligence (AI) Strategy*. CSU AI Commons.
- Froomin, R. (2025). *Search and Rescue Agentic Design with Microsoft AutoGen* (Unpublished senior project). California Polytechnic State University, San Luis Obispo.
- Kurfess, F. J., Wang, J., & Kalathas, P. (2025). *Empowering Experiential Learning through an AI-Driven Agentic Design Framework for Engineering Labs*. CSU-AIEIC Project #84 Report.
- Wynants, S., Childers, G., De La Torre Roman, Y., Budar-Turner, D., & Vasquez, P. (2025). *ETHICAL Principles AI Framework for Higher Education*. California State University, Fullerton. Licensed under CC BY-NC 4.0.