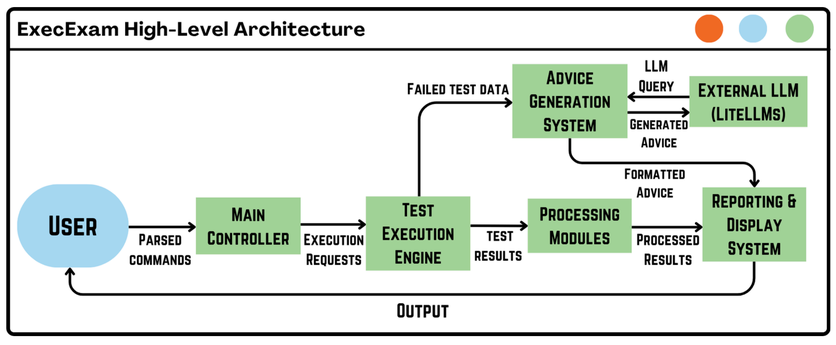
Hemani Alaparthi
Pallas-Athena Cain
Gregory Kapfhammer
Allegheny College

ExecExam: Streamlining Python Assessments with Automation and Personalized Feedback

Project Goals

- **Executable examinations** invite students to complete realistic and feasible **programming tasks** with **industry-standard** tools
- **Test suites** and **linters** analyze the student's project submission
- ExecExam **streamlines** the **assessment** of programming **tasks**
- Tool provides **automated & personalized feedback** for students
- It improves the **efficiency of grading** executable examinations

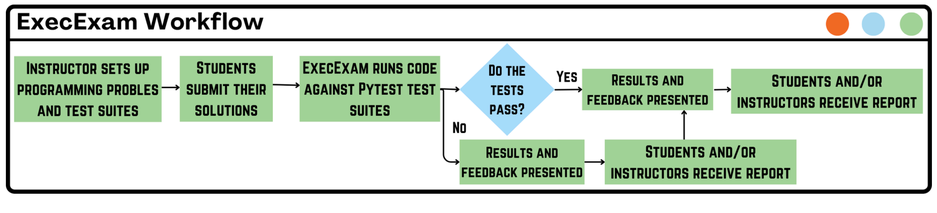
Executable exams with ExecExam streamline grading and enhance learning with automated tools to assess realistic programming tasks and provide personalized feedback



Coding Mentor

- Integrates **large language models** (LLMs) via LiteLLM's unified API
- LiteLLM's **web proxy** enables **democratized access** to setup LLMs
- **Instructors** or **students** can provide **access tokens** and/or API **keys**
- Automatically offers **step-by-step suggestions** for **fixing** code errors
- Offers **context-aware feedback beyond** only **test case failure** details

LiteLLM integrates multiple LLMs to provide context-aware, step-by-step programming feedback that helps students to better understand programming and algorithmic concepts



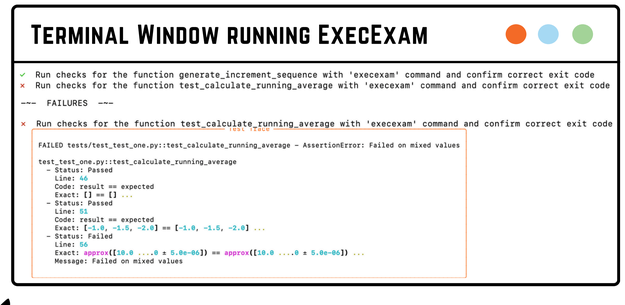
Automated Tools

- ExecExam runs **provided Pytest** tests to **verify** student solutions
- Generates comprehensive test reports that clearly **summarize** both successful **outcomes** and specific **points of failure** in student code
- Adds **command-line options** to ensure **best use** of Pytest features
- Ensures **consistent** and **fair evaluation** of student's projects
- Shows **all failures**, not just the first one, **unlike typical Pytest** run
- **Integrates** with GatorGrade, GitHub Classroom, and GitHub Actions

ExecExam uses Pytest to auto-verify student code, generate detailed reports, and ensure consistent grading while giving all failures to support instructor assessment

Actionable Insights

- Moves **beyond pass/fail grading** with **tailored explanations**
- Highlights **specific errors** and **suggests alternative solutions**
- Encourages students to **iteratively improve** their code
- Fosters **deeper understanding** of programming principles



Broad Applicability

- **User-friendly** and **easy to integrate** into Python assessments
- **Scalable** for classrooms, online courses, and development
- Uses i**ndustry-standard** automated **testing** and **debugging**
- Runs in **CI/CD pipelines** as an introduction to **best practices**

ExecExam is a user-friendly tool that automates testing, enhances debugging, improves code quality, and supports CI/CD integration

Lessons Learned

- Detailed feedback **increases student engagement**
- Automated assessment **reduces instructor workload**
- **Context-aware suggestions** improve debugging
- Leveraging **industry-standard tools** paves the way for students to effectively engage in follow-on projects
- **Over-reliance on LLMs can hinder learning**; need options to disable and restrict coding mentor's advice

Automated feedback boosts engagement, reduces workload, but balancing automation and limiting LLM over-reliance is effectiveness key

Future Work

- **Hold test cases** for instructor and advanced grading use
- Develop **analytics tools** to track student progress on exam
- **Log LLM interactions** to monitor **usage** and **effectiveness**
- Student **reviews** on the quality of **coding mentor's advice**
- Give **offline feedback** with **local** machine learning **models**
- Conduct **full experiments** to confirm **anecdotal evidence**
- Support **more programming languages** beyond Python
- Develop a **Pytest plugin** to offer ExecExam's detailed feedback and LLM-based suggestions in **stand-alone tool**

Future work will enhance ExecExam so it offers (a) more useful feedback to learners and (b) stand-alone functionality for software developers

Learn More

- **Prior research**: Chris Bourke, Yael Erez, and Orit Hazzan. 2023. Executable Exams: Taxonomy, Implementation and Prospects. In Proceedings of the 54th SIGCSE conference.

Try out ExecExam and contribute to the project!

- **https://github.com/GatorEducator/execexam**
- **https://pypi.org/project/execexam/**