

Peter Bucci

me@peterbucci.com ♦ (631) 456-7295 ♦ Boston, MA ♦ [GitHub](#) ♦ [LinkedIn](#)

EDUCATION

University of Massachusetts - Boston

Bachelor of Science - Computer Science | GPA: 3.9

Boston, MA

December 2026

Bunker Hill Community College

Associate of Science - Computer Science | GPA: 3.98

Cambridge, MA

December 2024

TECHNICAL SKILLS

Languages: Python, Java, TypeScript/JavaScript, C++, SQL

AI Systems / ML: LLM & Agent Evaluation, RAG, Long-Context Systems, PyTorch, TensorFlow/Keras, scikit-learn, NumPy, pandas, OpenCV, MediaPipe

Frontend / Backend: React, FastAPI, Spring Boot, Node.js, REST APIs, PostgreSQL, MongoDB, Spring Data JPA

Tools / Cloud: Git/GitHub, Docker & Docker Compose, AWS, Flyway, Maven, pytest, JUnit, Testcontainers, Linux, Jupyter

RELEVANT EXPERIENCE

AI/ML Research Intern, UMass Boston

June 2026 - Present

- Expanded an AI context-management research system to realistic software-engineering environments by integrating five public AI-agent benchmarks—including SWE-bench Verified, SWE-Gym, Long Code Arena, Terminal-Bench, and AgentDojo—covering code repair, repository reasoning, terminal interaction, and adversarial tool use, with reproducible task execution and automated validation.
- Built an automated evaluation framework to test how reliably AI agents preserve critical information when context is removed, corrupted, or altered, generating 1,860 controlled test cases across 31 tasks and comparing performance against 23,880 matched baseline evaluations.
- Improved the reliability of the evaluation system by independently verifying that stressed tasks still measured the same underlying objective, rejecting invalid test cases, and adding failure tests that deliberately corrupted protected evaluation data; the final robustness check passed 45 of 45 required comparisons, with 44 ties and one improvement over matched baselines.
- Designed controlled experiments to determine which parts of the AI context-management system actually drive performance by selectively removing mechanisms such as counterfactual reasoning, dependency tracking, and hard constraints while keeping tasks, budgets, and evaluation conditions consistent.

PROJECTS

MAGE Platform | Scrum Lead / Full-Stack Developer

[Link](#) | [GitHub](#)

Software Engineering Capstone, UMass Boston

- Led a multi-developer Agile/Scrum team in turning a real-time, audio-reactive rendering engine into a full-stack web platform for creating, saving, organizing, and launching interactive visual scenes; established the GitHub repositories, project board, sprint workflow, backlog, user stories, and code-review process.
- Built core backend infrastructure with Java, Spring Boot, PostgreSQL, and Flyway, including REST APIs, authentication, scene persistence, tagging, and engagement features, with JUnit and Testcontainers integration testing against containerized PostgreSQL.
- Developed and integrated React/TypeScript product features around the rendering engine, including scene creation and playback, authentication, discovery, user libraries, media controls, and reusable frontend architecture; authored 110 merged PRs across the project's frontend, backend, and engine repositories.
- Containerized and deployed the application to AWS, implementing production configuration, reverse-proxy routing, object-storage workflows, and deployment fixes to move the project from local development to a working hosted application.