Gerardo Montemayor

gerardom1226@gmail.com • (443) 538-0729 • linkedin.com/in/gerry-montemayor • github.com/gerry-montemayor

EDUCATION

Cornell University | **Ithaca, NY**

B.S. in Computer Science, Minor in Electrical & Computer Engineering | **GPA:** 3.5/4.0   **May 2026**

Relevant Courses: Generative AI, Deep Learning, Machine Learning, Algorithms, Data Science, OOP & Data Structures

EXPERIENCE

Machine Learning Engineering Intern | **Seam AI** | **San Francisco, CA**   **June 2025 - Aug 2025**

- Built a conversational Sales Persona generation agent with FastAPI backend, evaluation and monitoring on Langsmith, and Typescript/React frontend, reducing customer time-to-search by 1 hour per Persona
- Trained a NN classifier using feature-based transfer learning to align research queries to retrievers, maintaining accuracy parity with Gemini2.0-Flash router and reducing p50 latency from 1.5s to 50ms
- Implemented a string utilities module for broad use in the application unifying logging, error handling, and normalization. Normalization reduced our job title FAISS index disk space by 9MB without impacting accuracy

Application Developer | **AI Innovation Lab** | **IT@Cornell**   **Sep 2024 - Mar 2025**

- Developer for AI initiative at Cornell building GenAI-powered apps to support campus organizations
- Building application integrating OCR technologies, VLMs, and LLMs to automate expense processing, ensure compliance with state/school policies, and reduce manual workload for financial team

Software Developer Intern | **Technology Advancement Center** | **Columbia, MD**   **May 2024 - Aug 2024**

- Developed a virtual assistant to support business operations, including front-desk interactions and company events
- Deployed a secure web application with Docker and AWS EC2, establishing a CI/CD pipeline to streamline delivery of weekly MVPs and improvements based on user feedback

Robotics Software Engineer | **Cornell Mars Rover**   **Aug 2023 - Present**

- Programmer for project team that builds semi-autonomous rover to compete in University Rover Challenge (URC)

PROJECTS

GenAI Research Project | [Paper](#) | [Code](#) | CS 6784   **Aug 2025 - Dec 2025**

- Co-authored research paper on LuKA, our proposed KV cache design that compresses long contexts for more efficient LLM inference while preserving information
- Implemented sparse attention mechanisms and developed the evaluation framework for benchmarking performance, achieving 27% accuracy improvement over SOTA methods on multi-topic QA benchmarks with LuKA

AR Tag Detector | **Cornell Mars Rover**   **Aug 2024 - May 2025**

- Implemented a ROS2 package with OpenCV to process camera input to detecting tags in field, determine relative positions and orientations of tags, and navigate difficult terrain for Autonomous Challenge at URC

Rover Interface | **Cornell Mars Rover**   **Aug 2023 - May 2024**

- Implemented a high-performance interface to operate and access rover capabilities, including camera feeds, life-detection systems, robotic arm simulations, and a central ROS2 node launch hub
- Created specialized GUI components using React and Foxglove Studio, a robot visualization platform, enhancing rover control through seamless integration with ROS2

McGraw Tower Defense | [Github](#)   **Sep 2023 - Dec 2023**

- Developed a Cornell-themed tower defense game in OCaml with player upgrades and bullet-tracking physics

SKILLS & OTHER

Programming Languages: Python, C/C++, JavaScript, Java, OCaml , HTML/CSS   **Languages:** English, Spanish

Frameworks/Libraries: Pytorch, Scikit, Numpy, Langsmith, ReactJS, ExpressJS, NodeJS, ROS

Developer Tools: Git, Docker, Google Cloud Platform, AWS ECS, Linux, MongoDB, Foxglove Studio