Justin M. Dannemiller

jdannemi@andrew.cmu.edu ♦ linkedin.com/in/justin-dannemiller ♦ justin-dannemiller.github.io

EDUCATION

Carnegie Mellon University (CMU) Pittsburgh, PA
Master of Science in Electrical and Computer Engineering – GPA: 3.91/4.00 May 2025
Selected Coursework: Deep Generative Modeling, Multimodal ML, Introduction to Deep Learning

Kent State University (KSU) Kent, OH
Bachelor of Science in Computer Science – GPA: 3.96/4.00 May 2023
Selected Coursework: Natural Language Processing, Machine and Deep Learning, Advanced AI

SKILLS

Operating Systems and Cloud Platforms: Linux/Ubuntu, Windows, GCP, AWS
Programming Languages & Databases: Python, C++, C, CUDA, SQL, MongoDB, Milvus
Frameworks, Libraries, and ML Infrastructure: PyTorch, TensorFlow, NumPy, Pandas, MLFlow, FastAPI
Deep Learning Tools: LangChain, LangGraph, MCP, Google ADK, Google A2A, vLLM, HuggingFace, OpenCV

EXPERIENCE

AI Engineer - Open Access Technology International (OATI) Minneapolis, MN
OATI Genie - Multi-Agent LLM System for California ISO (CAISO) May 2025 – Current

- Scaled internship pilot into a production-grade multi-agent system for CAISO, **reducing operator workload** by automating core operational tasks using Google ADK, A2A, GPT-OSS-120B, and MCP.
- Optimized vLLM batching and prompt dispatch after analyzing latency bottlenecks, **reducing end-to-end inference time by 35–40% in production.**
- Built an automated evaluation and regression test stack using Google ADK for agent/tool-call evaluation and MLflow for experiment tracking and analysis, **reducing evaluation cycles from days to hours.**
- Fine-tuned Mistral-7B-Instruct with LoRA on prompt-tool-call traces, analyzing failure cases and execution metrics to achieve **a 13% lift in MCP database-tool execution accuracy relative to baseline.**

AI Engineer Intern – OATI Minneapolis, MN
Generative AI Pilot for Enterprise Knowledge Retrieval June 2024 – May 2025

- Spearheaded a company-wide generative AI pilot showcasing potential of LLM-based retrieval and automation, which later became the **foundation of a million-dollar enterprise platform.**
- Engineered a RAG system across three OATI products using Milvus, Mixtral-8x7b, and hybrid embeddings over hundreds of product documents, **cutting manual document search time by ~80%.**
- Integrated MCP tooling into the RAG stack and analyzed tool-call performance metrics to achieve **87% execution accuracy over live operational data.**
- Showcased system at major energy conferences (DistribuTECH and OATI Energy Conference), driving strategic interest and **securing a partnership with CAISO, a leading North American grid operator.**

RESEARCH PROJECTS

Dubai Airport: Multimodal AI Assistant - CMU Sept. 2024 – Dec. 2024

- Developed a GPT-4o-powered multimodal airport assistant delivering real-time passenger support across flight information, service inquiries, and interactive media recommendations.
- Engineered a dynamic RAG pipeline using LangChain to orchestrate live APIs (trending media, flight and weather updates) with GPT-4o tool calling, **enabling real-time, context-grounded recommendations.**
- Achieved a **95.5% API-selection and execution accuracy,** ensuring reliable, context-aware responses.

Enhanced LLM Mathematical Reasoning Project - CMU Jan. 2024 – June 2024

- Constructed a custom dataset of **1M+ arithmetic expressions** with solutions and chain-of-thought steps, enabling fine-tuning experiments to improve arithmetic generalization in pretrained LLMs.
- Finetuned Phi-1.5 and Llama3-8B with LoRA on custom data, **yielding 5% and 2% respective accuracy gains on out-of-domain GSM8K dataset**, demonstrating robust mathematical reasoning transfer.