

Jacob Furtaw

jfurtaw97@gmail.com | 410-533-7663 | www.linkedin.com/in/jacob-furtaw/ | www.jfcoded.com/

Baltimore, MD

Professional Summary

Innovative Machine Learning Research Engineer specializing in Natural Language Processing, Data Engineering, and Computer Vision. Expert in building AI assistants with LLMs and Retrieval-Augmented Generation, transforming unstructured data into actionable insights. Collaborative problem-solver passionate about pushing AI boundaries.

Skills

AI/ML & Data Science: PyTorch, TensorFlow, Transformers, HuggingFace, Langchain, Llama-Index, Scikit-Learn, Pandas, NumPy, Matplotlib, Ollama, Cuda, Accelerate, **Retrieval-Augmented Generation (RAG)**

Programming Languages: Python, JavaScript (Node.js); **Familiar With:** C++ and Java

Tools & Platforms: Git, Docker, Jupyter, Conda, PyCharm, VS Code, Linux, Windows, Nvidia NIMS

Professional Experience

Advanced Repair Agent | Geek Squad | On-Site

March 2022 - Current

- Recognized with a Letter of Recommendation for designing and implementing operational improvements, resulting in a 50% increase in productivity and improved customer satisfaction.
- Consistently ranked in the top 3% of all Advanced Repair Agents in the District and Region
- Completed tens of thousands of hardware and software repairs on various devices and operating systems

Machine Learning Research Engineer | SurgePoint Software | Hybrid

August 2023 - October 2024

- Utilized **data engineering** and **software engineering** skills to reduce 200 million lines of unstructured data into a 13-million-line structured dataset. This was achieved using custom Python parsers and LLMs, which increased the model's ability to pull relevant context from the dataset and increase semantic relevance scores by 50-75%
- Designed, engineered, and tested an advanced Python **Retrieval-Augmented Generation** pipeline that supplied various LLMs with my custom dataset, using noSQL datastores like Milvus and ChromaDB, created from the university website to reduce model hallucinations
- Collaborated with a small, diverse startup team of engineers doing weekly standups and sprint reviews
- Led the development of an innovative AI Assistant (Chatbot) using open-source LLMs like Mistral-Nemo 12B, Mistral 7B, and Llama 3.1 8B

Projects

Chat RAG

- Created a RAG-powered chatbot with a Gradio user interface, supporting 10+ local models from Ollama and any Hugging Face model, as well as cloud-based models from OpenAI, Anthropic, and NVIDIA NIMS
- Engineered a modular Python architecture with 3 to 5 utilities for model management, chat functionality, and UI interactions, featuring dynamic model switching, custom prompt integration, model parameter tuning, and quantization options
- Designed flexible data ingestion from three diverse sources (local files, GitHub repositories, and vector databases)
- 20+ Stars on GitHub with active users of the software*

Automatic Identification of Equivalent Mutants using an ASTNN

- Excelled as a member of a five-man Scrum Team, engaged in sprint planning, daily standups, and sprint reviews
- Investigated and implemented the use of a transformer-based model (CodeBERT) for the task of binary classification
- Optimized data preprocessing by creating a custom Python parser, added new and tuned existing hyperparameters, and customized Python training scripts
- Increased the model's F1 and accuracy scores from a 79% average to a 92% average using oversampling and undersampling to balance our mutant dataset

- Condensed research and outcomes into a comprehensive formal report and poster, which my team and I presented at a campus research fair

Education

Bachelor of Science in Computer Science, Software Engineering Concentration

Graduated December 2023

Towson University, Towson, MD

• **Coursework:** Artificial Intelligence, Database Management Systems, Software Quality Testing & Assurance, Data Structures and Algorithms, Requirements Analysis and Modeling