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 and Computer Vision. Expert in building AI assistants with LLMs and Retrieval-Augmented Generation and transforming unstructured data into actionable insights. Collaborative problem-solver passionate about pushing AI boundaries.

Skills

AI/ML Frameworks: Llama-Index, Langchain, HuggingFace, PyTorch, TensorFlow, Transformers, Scikit-Learn, Accelerate
Data Science: Pandas, NumPy, Matplotlib
Programming Languages: Python, JavaScript (Node.js); **Familiar With:** C++ and Java
Tools & Platforms: Git, Docker, Jupyter, Conda, PyCharm, VS Code, Linux, Windows, Ollama, Nvidia NIMS

Work Experience

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, increasing semantic relevance scores by 50-75%
- Designed, engineered, and tested an advanced **Retrieval-Augmented Generation** pipeline to supply various LLMs with my custom dataset using noSQL datastores like Milvus and ChromaDB
- Collaborated with a small 6-man, diverse startup team of engineers doing weekly standups and sprint reviews
- Led the development of an innovative AI Assistant using open-source LLMs like Mistral-Nemo and Llama 3.1

Advanced Repair Agent | Geek Squad | On-Site — March 2022 - Current

- Designed operational improvements alongside new management that increased the team's productivity by over 50% and earned me a letter of recommendation from upper management
- Consistently ranked in the top 3% of all Advanced Repair Agents across our marketplace, resolving hundreds of hardware and software repairs across diverse devices and operating systems
- Streamlined repair workflows with automation and created documentation, reducing repair time by 20%-40%

Projects

Chat RAG | Project Link

- Created a RAG-powered chatbot with a Gradio user interface, supporting **local and API inference** from any of the hundreds of Ollama and HuggingFace models, as well as any models from OpenAI, Anthropic, and NVIDIA NIMS
- Engineered a modular Python architecture with 5+ utilities for model management featuring dynamic model switching, custom prompt integration, model parameter tuning, quantization options, and many more
- 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(GNN) | Project Link

- 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 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 an average of 79% to 92% using oversampling and undersampling to balance our mutant dataset
- Condensed research and outcomes into a formal report and poster, which I presented at a research fair

Education

Bachelor of Science in Computer Science, Software Engineering Concentration — Graduated December 2023
Towson University, Towson, MD