Jacob Furtaw

Jfurtaw97@gmail.com | 410-533-7663 | www.linkedin.com/in/jacob-furtaw/ | www.jfcoded.com/projects
Baltimore, MD | Willing and Ready to Relocate Anywhere

Professional Summary

Innovative Machine Learning Research Engineer specializing in Natural Language Processing and Computer Vision. With over 4 years of machine learning research experience, I have become an expert in building AI assistants with LLMs and Retrieval-Augmented Generation, Multimodal Chatbots, Training and Finetuning transformer models, 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 | Full-Time August 2023 - April 2025

- Utilizing **data engineering** skills to reduce 200 million lines of unstructured data into a 13-million-line structured dataset, increasing semantic relevance scores by 50-75%
- Designing and implementing a **Retrieval-Augmented Generation** (RAG) pipeline, integrating vector databases (Milvus, ChromaDB) to supply LLMs with custom datasets, optimizing accuracy and scalability for production use.
- Collaborating with a 6-person cross-functional startup team in weekly standups and sprint reviews, delivering actionable insights, and aligning technical efforts with business goals
- Championing MLOps practices, leveraging Docker and CI/CD workflows to streamline model deployment and lifecycle management, ensuring robust, scalable AI solutions

Advanced Repair Agent | Geek Squad | On-Site | Seasonal March 2022 - Present

- Designing 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
- Streamlining repair workflows with automation and creating documentation, reducing repair time by 20%-40%

Research Projects

Agent Qwen | Project Link

- Architected a cutting-edge **multimodal AI assistant** using the Qwen2.5-Omni-7B, seamlessly integrating text, image, audio, and video inputs to help diagnose device issues faster and enhance technical support precision
- Engineered robust audio processing with OpenAI’s Whisper for audio-to-text transcription, ensuring accurate transcription for actionable diagnostics

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

Education

Bachelor of Science in Computer Science, Software Engineering Concentration December 2023

Towson University, Towson, MD

Clubs: Machine Learning Research Group