Jacob Furtaw

Jfurtaw97@gmail.com | 410-533-7663 | My LinkedIn | My Website

Intro

I am an innovative Machine Learning Engineer/Researcher specializing in Natural Language Processing (NLP), seeking opportunities to advance AI technology. Experienced in developing cutting-edge AI Assistants and implementing advanced NLP techniques including RAG. Skilled in leveraging large language models for versatile, high-impact solutions. Expertise spans data preprocessing, model fine-tuning, and deploying scalable ML solutions. I am passionate about solving complex problems and pushing AI boundaries, with a focus on NLP but open to exploring new machine learning frontiers.

Skills

Machine Learning/Data Science: PyTorch, TensorFlow, Scikit-Learn, **Transformers**, HuggingFace, **Langchain**, **Llama-Index**, Gradio, Pandas, NumPy, Data Preprocessing, Synthetic Data Creation, **Retrieval-Augmented Generation**
Programming Languages: Python, JavaScript (Node.js), CSS, C++, Java
Tools & Platforms: Git, **Docker**, Linux, Windows, Jupyter, PyCharm, VS Code, Nvidia AI Workbench

Work Experience

Machine Learning Engineer\Researcher   Aug 2023- Present
SurgePoint Software
- Developing an innovative AI Assistant using open-source LLMs like Mistral 7B, NeMo 12B, and Llama 3.1 8B
- Utilizing self-taught **data engineering** and **data preprocessing** skills to craft a 13-million-line dataset of web-scraped data
- Implementing **Retrieval-Augmented Generation** to increase semantic relevance scores and combat model hallucinations
- Leveraging self-taught **prompt engineering** skills to ensure the model has a well-structured prompt and enough context to answer any query a user might input

Advanced Repair Agent   Mar 2022-Present
Geek Squad
- Earned a Letter of Recommendation from Upper Management for designing and helping to implement a plan to revamp our locations day to day operations, increasing overall productivity by 57% and customer satisfaction
- Executing tens of thousands of Software and Hardware fixes on desktops, laptops, tablets, and cell phones
- Receiving an award for Top Performing ARA at our location every quarter and was in the top 3% in our District and Region

Projects

Chat RAG
- Developing an advanced, chat-based coding assistant using Retrieval-Augmented Generation technology, integrating multiple language model providers(HuggingFace, Ollama, OpenAI), and implementing a user-friendly Gradio UI.
- Engineering a flexible system with dynamic model switching, custom embedding integration, and efficient memory management, showcasing proficiency in PyTorch, Llama Index, and state-of-the-art language models.
- Implementing key features including context-aware document retrieval, streamed responses for real-time interaction, and quantization options, demonstrating strong skills in ML engineering and AI application design.

Automatic Identification of Equivalent Mutants
Towson University, Towson, MD
- Excelled as a member of a Scrum Team, engaged in sprint planning, daily standups, and sprint reviews
- Investigated and implemented the use of a transformer-based model for the task of binary classification
- Optimized data preprocessing, added and tuned hyperparameters, and increased overall performance
- Condensed research and outcome into a comprehensive formal report my team presented at a campus research fair

Education

Bachelor of Science in Computer Science, Software Engineering Concentration   Jan 2022-Dec 2023
Towson University, Towson, MD
- **Related Coursework**: Artificial Intelligence, Database Management Systems, Software Quality Testing & Assurance, Data Structures and Algorithms, Requirement Analysis and Modeling

Associate of Science in Computer Science   Aug 2019-Dec 2021