

# Jacob Furtaw

[jfurtaw97@gmail.com](mailto:jfurtaw97@gmail.com) | 410-533-7663 | [My LinkedIn](#) | [My Website](#)

---

## Professional Summary

Innovative Machine Learning Research Engineer with expertise in Natural Language Processing (NLP) and Generative AI. Skilled in developing cutting-edge AI solutions, including large language models (LLMs) and multi-modal systems. Experienced in leading technical teams, collaborating across functions, and driving projects from ideation to deployment. Passionate about solving complex problems and pushing AI boundaries to deliver business value. Please explore my website to view my full [virtual resume](#) and [recent projects](#).

## Skills

**AI/ML Technologies:** PyTorch, TensorFlow, Transformers, Hugging Face, Langchain, Llama-Index, Ollama,

**NLP & Generative AI:** Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Prompt Engineering, Agentic RAG

**Data Science:** Scikit-Learn, Pandas, NumPy, Data Engineering, Synthetic Data Creation

**Cloud:** AWS, Azure

**Programming:** Python, JavaScript (Node.js), C++, Java, CSS

**Tools & Platforms:** Git, Docker, Linux, Jupyter, PyCharm, VS Code, Nvidia AI Workbench

## Professional Experience

**Machine Learning Research Engineer** | SurgePoint Software

Aug 2023-Present

- Developed an innovative AI Assistant using open-source LLMs like Mistral 7B, Mistral-Nemo 12B, and Llama 3.1 8B
- Utilized self-taught **data engineering** skills to craft a 13-million-line dataset of web-scraped data into a structured JSON file
- Implemented **Retrieval-Augmented Generation** to increase semantic relevance scores and combat model hallucinations
- Leveraged **prompt engineering** skills to ensure the prompt was well-structured and the model could answer any user's query
- Worked with a small, diverse, and time constrained group of individuals at a startup in a remote setting

**Advanced Repair Agent** | Geek Squad

Mar 2022-Present

- Recognized with a Letter of Recommendation for designing and implementing operational improvements, resulting in a 150% increase in productivity and customer satisfaction. Consistently ranked in the top 3% of performers in the District and Region
- Completed tens of thousands of hardware and software repairs on various devices and operating systems

## Projects

### Chat RAG

- Developing an advanced, chat-based coding assistant using Retrieval-Augmented Generation technology, integrating multiple language model providers (HuggingFace, Anthropic, 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 using an ASTNN**

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 for finetuning, 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, Requirements Analysis and Modeling

**Associate of Science in Computer Science**

Aug 2019-Dec 2021

