Aayush Jannumahanti

aayushkjvs@gmail.com | +1(667)-369-8533 | LinkedIn | San Jose, CA

PROFESSIONAL SUMMARY:

- Skilled in developing high-performance web applications using Python (3.x/2.7) and frameworks such as Django, OpenAI API, LangChain, and Flask, with a strong foundation in MVC architecture.
- Over 7 years of extensive experience in software development, including backend, frontend, business intelligence, database management, and automation frameworks.
- Proficient in designing and integrating Restful API and GraphQL endpoints, employing advanced OAuth authentication and microservices architecture.
- Expertise in AI and machine learning, utilizing libraries like scikit-learn, TensorFlow, and PyTorch to drive data science initiatives and deliver business value.
- Hands-on experience with cloud services, including AWS (Lambda, S3, EC2, RDS, Glue) and Azure (Azure Functions, Blob Storage, Cosmos DB). Adept in cloud orchestration and containerization with Kubernetes, Docker, and Azure Kubernetes Service (AKS) for scalable, cloud-native solutions.
- Experienced in automating deployments using CI/CD tools such as Azure DevOps, GitHub Actions, and Jenkins to create efficient and reliable delivery pipelines. Skilled in building and optimizing ETL pipelines and workflow orchestration with Apache Airflow, PySpark, PostgreSQL, and Redis for large-scale data processing.
- Proficient in TDD and BDD practices, creating reliable and maintainable test cases with tools like PyTest to ensure code quality and reliability. Strong foundation in Machine Learning and Artificial Intelligence.
- Experienced in Natural Language Processing and AI language services.
- Knowledgeable in Large Language Models, OpenAI API, and prompt engineering.
- Robust experience in MLOps, LLMOps and AI Agents.
- Strong Python and SQL programming skills with experience in Huggingface datasets, LMStudio, LLMs, Jupyter notebooks and Git-based source control.
- Ability to coach and guide teams, translating ill-defined data needs into concrete deliverables.
- Proficient in collaborating with cross-functional teams to meet business and technical goals.

TECHNICAL SKILLS:

Programming Languages	Python, Java, JavaScript, HTML & CSS
Database Management	PostgreSQL, MySQL, MongoDB, Redis, pgvector, ChromaDB
Frameworks & API's	LangChain, OpenAI, FastAPI, Angular JS, Restful API, Next JS, React JS, Databricks and GraphQL
Libraries	Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, PySpark, Seaborn, Tableau, Streamlit, Matplotlib, Plotly, NLTK, SpaCy
Devops, Cloud & CI/CD	AWS (Lambda, S3, EC2, RDS), GCP, Jenkins, GitHub Actions, Docker, Kubernetes, AZURE
Testing and Debugging	PyTest, TDD, BDD, PDB, SAML, logging modules
Security	OAuth, JWT, OWASP guidelines
Version Control & Tools	Git, GitHub, GitLab, Bitbucket, Swagger, Postman

PROFESSIONAL EXPERIENCE:

Bristol Myers Squibb, MD   **April 2023 - Present**
Sr. Python Developer

Project: AI/ML Framework for Generative AI and Conversational Systems
Responsibilities:

- In my current role at Bristol-Myers Squibb, I contribute to the development of advanced AI/ML solutions that enhance decision-making, user interaction, and data-driven research. Specifically, my responsibilities include building scalable generative AI frameworks that improve knowledge retrieval and conversational intelligence in healthcare-centric applications.

- **AI Framework Development:** Designed and implemented retrieval-augmented generation (RAG) systems and AI agents to improve engagement across BMS's omni-channel platforms, ensuring high-quality interaction in clinical and operational contexts.

- **Natural Language Processing (NLP):** Leveraged NLP capabilities to enhance understanding of complex biomedical data, integrating pre-trained language models from Huggingface and utilizing LMStudio for quantized LLM aand Retrival Augmented Generation (RAG) performance.

- **Scalable AI/ML Pipelines:** Developed robust AI/ML pipelines using TensorFlow, PyTorch, and Scikit-learn, ensuring rapid model iteration and deployment in research environments.

- **Cloud-Based Deployment:** Implemented and maintained cloud-native solutions on AWS (Sagemaker, Bedrock, Lambda, S3, Connect, Lex, DynamoDB, API Gateway, CloudWatch) and Azure, optimizing operational workflows for scalable AI services.

- **Responsible AI Implementation:** Ensured ethical AI development by integrating Responsible AI frameworks, continuous model evaluation, and real-time monitoring systems to uphold safety, transparency, and fairness in applications involving sensitive biomedical data.

- **Data-Driven Insights:** Built real-time data processing pipelines utilizing Spark, Data Lake, MLflow, and Snowflake, enabling actionable insights and supporting BMS's mission to deliver innovative medicines.

- **Microservices and CI/CD:** Deployed microservices architecture using Docker, Kubernetes, and Swagger, facilitating modular and efficient service delivery in research and clinical applications.

- **Data Visualization:** Developed interactive dashboards and visualizations using Plotly, Seaborn, and Matplotlib, supporting cross-functional teams with real-time metrics for drug discovery and operational enhancements.

- **Agile Methodology:** Actively participated in Agile practices (Scrum, Kanban, SAFe), ensuring timely delivery of key milestones aligned with BMS's strategic goals.

AstraZeneca, MD   **August 2022 to March 2023**
Python Developer
Project: End-to-End Automated Speech Recognition System – Audio Tagger System
Responsibilities:

Conducted a thorough analysis of test cases and contributed to designing detailed test strategies to ensure system robustness and accuracy.

- Improved the End-to-End ASR system's accuracy by fine-tuning pre-trained models, such as Wav2Vec 2.0 and DeepSpeech, and incorporating advanced data augmentation techniques, including noise injection and speed perturbation.

- Automated the entire process of tagging and transcribing audio and video files, significantly reducing manual intervention and increasing operational efficiency by over 40%.

- Integrated speech-to-text APIs and fine-tuned them for domain-specific applications to enhance transcription accuracy.

- Developed custom Python scripts to preprocess and clean large datasets, ensuring better model training and reducing errors caused by noisy data.

- Collaborated with international, cross-functional teams to integrate the Audio Tagger (HAT) system with other multimedia systems, improving its functionality and adoption across diverse use cases.

- Leveraged NVIDIA NeMo Guardrails to create a robust, error-tolerant pipeline that ensured secure and contextually aware speech recognition.

- Designed and implemented a scale able back end architecture using Python, enabling the system to process high volumes of data in real time.
- Enhanced the user interface to display transcription metadata and tagging results using Python frameworks, improving accessibility for end-users.
- Conducted extensive unit testing and created automated test scripts using PyTest, ensuring high reliability of system components.
- Deployed the system on cloud platforms, including AWS and GCP, optimizing performance through auto-scaling and load balancing mechanisms.
- Documented system workflows, architecture, and implementation details to facilitate on-boarding and knowledge sharing among team members.
- Contributed to open-source repositories, sharing innovative solutions and engaging with the broader developer community to improve the project's impact.

Apple Inc, Hyderabad, India   **December 2019 – August 2022**
Data Engineer
Project: Apple Maps Production and Development
Responsibilities:
- **Continuous Integration/Continuous Deployment (CI/CD):** Contributed to the CI/CD processes for Apple Maps across regions, including **Americas** and **Rest of World Countries**, ensuring timely and high-quality delivery of updates.
- **Projects and Training:** Worked on multiple projects as an on-site vendor for Apple Inc., with a primary focus on Apple Maps. Successfully completed **On the Job Training (OJT)** to enhance expertise in Apple Maps products and processes.
- **Geospatial Data Analysis:** Collaborated on geospatial data analysis initiatives, leveraging **machine learning techniques** to improve **map accuracy**, optimize **routing algorithms**, and enhance user experiences.
- **Data Automation and ML Integration:** Automated data validation and anomaly detection workflows using **Python** and **machine learning models**, significantly increasing workflow efficiency. Developed predictive models to enable **real-time updates** and deliver actionable mapping insights.
- **Test Strategy and QA Processes:** Analyzed test cases and actively contributed to the creation of a comprehensive **Test Strategy** for projects. Participated in project planning, coordination, and the implementation of **QA methodologies** to ensure quality delivery.
- **Requirement Gathering and Validation:** Engaged in **requirement gathering** and analysis phases, validating and verifying requirements to ensure successful outcomes. Worked closely with stakeholders to align deliverables with business objectives.
- **Development and Implementation:** Designed, developed, and implemented various modules for the company's **core automation products**, following both **Agile** and **Waterfall** development methodologies throughout the **SDLC process**.
- **User Interface Testing:** Conducted **user evaluation tests** to ensure the website and application interfaces were intuitive, effective, and aligned with user needs.
- Developed Python scripts to manipulate files and update content in databases.
- Utilized **NumPy** for performing complex mathematical and data operations.
- Parsed data in **XML** and **JSON** formats using Python, storing results efficiently in databases.
- Generated actionable reports for business users using **Python ReportLab**, aiding data-driven decision-making.

- **Performance Monitoring:** Monitored and troubleshot OS (Linux) and performance-related issues to maintain system reliability and uptime.
- **Functional Testing:** Performed **functional testing** using **QTP**, defining various checkpoints for result verification and leveraging **data-driven testing methodologies** for robust validation.

- **Scripting and Unix Expertise:** Developed and executed **Bash scripts** for automation and system management tasks. Worked extensively in **UNIX environments**, demonstrating expertise in Unix commands and shell scripting.
- **Ongoing Support and Maintenance:** Provided continued support, maintenance, and development for bug fixes and patch sets, ensuring the stability and performance of existing web applications.

Elevance Health, Inc. (Previously Anthem)   **June 2018 to October 2019**
Python Developer
Project: Figma and Machine learning pipeline for G-Code Generation
- **Responsibilities:**
- Designed, developed, and optimized **RESTful APIs** to enable seamless integration of the machine learning algorithm with the frontend application, supporting real-time **G-Code generation** and enhancing user experience by minimizing response time.
- Built and refined a comprehensive **machine learning pipeline** for translating design files (e.g., CAD or SVG formats) into precise G-Code instructions, leveraging advanced preprocessing, feature extraction, and postprocessing techniques to improve the accuracy and efficiency of the automated workflows.
- Integrated advanced **data validation and error correction** mechanisms within the pipeline to ensure compatibility with diverse design file inputs and avoid system crashes or inaccuracies during G-Code generation.
- Implemented **hyperparameter tuning and model optimization techniques** to improve machine learning performance, reducing errors in G-Code translation by 25% while accelerating computation speed.
- Developed a robust **version control system** for the pipeline, ensuring seamless rollback and traceability of algorithm updates to maintain system stability and performance.
- Created **interactive Figma prototypes** to visually demonstrate product workflows and gather iterative feedback from stakeholders and end-users, refining the design to align with user expectations and enhancing usability.
- Conducted extensive user research sessions, incorporating insights into the iterative design process to address pain points and improve the overall user interface and functionality of the product.
- Designed a **comprehensive user interface blueprint** that bridged the gap between technical machine learning processes and intuitive user interactions, ensuring accessibility for non-technical users.
- Deployed and monitored the APIs on cloud platforms (e.g., AWS and Azure) with autoscaling and load balancing, ensuring high availability and low latency during peak usage.
- Automated testing pipelines for both machine learning models and API endpoints, using tools like PyTest and Postman, to ensure consistent reliability and performance in production environments.
- Implemented real-time **logging and monitoring tools** (e.g., ELK Stack, Prometheus) to track API usage, identify bottlenecks, and optimize G-Code generation processes dynamically.
- Documented the entire pipeline, including API endpoints, model architecture, and data flow processes, to facilitate collaboration among teams and provide a clear reference for future development.
- Collaborated with DevOps teams to integrate CI/CD pipelines for API deployment and model updates, enabling faster iteration cycles and seamless feature rollouts.

EDUCATION:

Master of Science | Computer Science | Baltimore, USA
The University of Maryland Baltimore County
Master's Thesis: Quantized Large Language Models for Mental Health Applications: A Benchmark study on efficiency, accuracy and resource allocation

Bachelor of Technology |
SRM Institute of Science and Technology