

VAISHNAVI VISWESWARAIAH

vvaishnavi2021@outlook.com, [LinkedIn](#), [GitHub](#), Dayton -Ohio-USA

PROFESSIONAL BACKGROUND AND EXPERTISE:

I have over **10 years of experience in the tech industry**. I have worked in various roles such as **Data Engineer, Machine Learning Engineer, and Software Developer**. Throughout my career, I have built a wide range of solutions, from **machine learning models** and **cloud-based AI systems** to **large-scale data pipelines**—using technologies like **Python, PySpark, AWS, Azure, and Databricks**.

TECHNICAL EXPERTISE:

My expertise spans across **cloud computing, artificial intelligence (AI), machine learning (ML), Generative AI (GenAI), and large language models (LLMs)**. I have helped companies improve their data systems, automate workflows, and gain valuable business insights through data-driven solutions. I have worked in a consulting environment, supporting clients across industries, including **healthcare, retail, and public services**, consistently delivering value through end-to-end data and AI engineering using **Databricks**.

I have also explored agentic GenAI architectures using tools like LangChain and LlamaIndex to build RAG-based agents for knowledge retrieval and structured reasoning. Additionally, I have knowledge of Agent-2-Agent and Model context protocol architectures.

UNIQUE STRENGTHS:

One of my unique strengths is my ability to implement data science algorithms from scratch, without relying on libraries like scikit-learn. During my graduate program, I implemented various data science algorithms, such as **Logistic Regression, Linear regression, Neural Network, and clustering algorithms**, from scratch. This experience provided me with a profound understanding of how these algorithms function and enabled me to establish a solid foundation in data science. This expertise has been invaluable in my career, enabling me to develop custom solutions that meet the specific needs of my clients.

LEADERSHIP AND COLLABORATION:

I'm also experienced in **leading teams, mentoring engineers**, and continuously learning new tools and technologies. My expertise and experience make me well-suited to deliver high-quality solutions that meet the needs of my clients. By combining my technical expertise, unique strengths, and experience with traditional ML, Cloud technologies, and GenAI, I am well-positioned to deliver innovative and effective solutions that drive business value.

TECHNICAL SKILLS

- **Programming:** Proficient in Python, SQL, PySpark, Java, and R.
- **Web Technologies:** Knowledgeable in HTML, CSS, JavaScript, and Django
- **Machine Learning Libraries:** Sklearn, Tweepy, NLTK, TensorFlow (Keras), spaCy, and BERT
- **Database Management:** SQL server, MySQL, and Snowflake
- **Cloud Computing:** AWS Sagemaker, Azure, **Databricks**
- **LLM:** Open AI, Langchain, Vector dB, Prompt Engineering
- **Dev Tools:** PyCharm, Eclipse, Jupyter, Azure ML Studio, AWS Sagemaker, Azure Databricks
- **Algorithm Implementation from scratch:** Fuzzy-C, K-Means, Logistic and Linear Regression, Neural network.
- **Palantir Foundry:** Pipeline builder, Code workbook, Code repositories, Ontology, Modeling Objective, Graph Vertex, Workshop, Slate, Contour, AIP.

PERSONAL PROJECTS:

PLANFINDER AI: HEALTH INSURANCE RATE FINDER AGENT: A conversational Streamlit web app that helps users find and compare health insurance plans using inputs like age, state, and tobacco use.

- Built using **LangChain, OpenAI**, Prompt engineering, **Pydantic**, agentic tools and functions, and secure login flows, the assistant utilizes LLM-powered extraction to fetch plan details and generate structured plan summaries from large datasets.(can be found at [gitbub link](#))

SQLGEN: TEXT-TO-SQL GENAI APPLICATION: A Productivity tool built on data stored in **Databricks** that enables product managers and business users to generate SQL queries using natural language.

- Includes features like ERD visualization, query feedback refinement, and Streamlit-based user interface with secure login. Built with **LangChain, Streamlit, Databricks**, Open AI, and Prompt Templates. (can be found at [github link](#))

PROFESSIONAL WORK HISTORY:

Savancys Inc – Data Engineer– Remote

March 2024 – Present

GENENTECH/ROCHE: Built LLM-powered applications for contract assessment, insights, and analysis using RAG pipelines, ontology, and agentic frameworks.

- Designed and developed backend applications on Foundry using **Workshop, Contour, pipeline builder, data lineage, and Code Repository**, creating PySpark pipelines for scalable ETL, data transformations, and writeback functionality.
- Implemented **RAG-based workflows and LLM agents** by leveraging Foundry ontology and data pipelines to deliver actionable insights for contract analysis and Agent evaluations.

BLUE HEALTH INTELLIGENCE: Migrated legacy healthcare data systems to a cloud-based platform using **Databricks** and Snowflake to enable GenAI analytics and enhance cost transparency.

- Migrated and modernized legacy healthcare data pipelines using **Databricks**, converting **Java logic to PySpark notebooks**, and implemented **Airflow DAGs** with YAML to automate scalable, distributed workflows.
- Built robust data ingestion and transformation pipelines to load curated data into **Snowflake**, developed **Snowflake procedures**, applied data quality and cleaning techniques, and optimized **Spark job performance** for efficient downstream analytics.

Accenture – AI/ML Computational Science Engineering Manager– Remote

April 2021 – Feb 2024

Gen AI/LLM CoE - ACCENTURE INTERNAL: Built an AI-powered platform as part of Accenture’s internal COE to support M&A and pharmaceutical insights by modeling data structures, integrating intelligent systems, and applying LLM-based automation for strategic decision-making. Built and maintained ML pipelines using **Databricks, PySpark, and MLflow** for scalable big data processing.

- Designed and deployed **LLM-based applications** using **LangChain, Vector DB, and Cosmos DB** for intelligent search and chatbot functionalities; implemented **vector and cognitive search** for dynamic retrieval.
- Developed, tuned, and monitored **LLM prompts** and model performance using **Power BI dashboards**, with focus on **drift detection, reproducibility, and domain-specific accuracy**.
- Architected scalable infrastructure on **Azure ML Studio** and **Databricks**, and deployed **LLMs using AWS SageMaker and Bedrock**, including pharmaceutical-focused applications.
- Developed an **LLM cost estimator** to compare and forecast cloud deployment costs (FLAN, Cohere, AI21) and Built **serverless GenAI architectures** using **AWS Lambda and API Gateway**.
- Explored **agentic GenAI architectures** by prototyping **RAG-based AI agents** using **LlamaIndex, LangChain, and Pydantic** to generate structured responses from enterprise knowledge bases; applied these learnings to develop internal demos for smart document retrieval, chatbot agents, and context-aware search.
- AI system evaluation context precision, context recall, context relevancy, Response precision, and factual verification using ROGUE, BLUE scores.

PALANTIR TECHNOLOGIES INC: Developed and integrated analytics-driven applications using Palantir to deliver actionable business insights and support diverse system and data environments.

- Developed analytics-based solutions using **Palantir Foundry** to deliver quantitative and qualitative business insights; partnered with cross-functional teams for seamless system and data integration; created and deployed ML use cases with interactive **Vertex apps, digital twins, and APIs**; built a **time-series data generator library** to support forecasting.
- Built **end-to-end data engineering and ML pipelines** using **PySpark, Pipeline Builder, Code Workbooks, Code Workspace, and Code Repositories**; created ontology models and business applications using **Workshop** and **Slate**; used **CLI tools** and **Object Gateway APIs** to automate data operations and maintain data quality for downstream analytics. Performed initial data analysis using Contour dashboard and shared results with end clients for quick overview of their data.

ROOM2Go – Product catalog feature engineering and ML classification Databricks PoC:

- Collaborated on a PoC for **Room To Go**, a furniture retailer, aimed at **automatically classifying product catalog data** based on features such as **category, color, and attributes** using existing metadata and textual descriptions. Performed **data ingestion and preprocessing** in **Databricks** using **PySpark**, followed by **feature extraction and text cleaning** for unstructured product descriptions.
- Built **ML models** to classify products into feature-based tags, applying NLP techniques such as **TF-IDF, word embeddings, and classification algorithms**. Used **MLflow** for experiment tracking and model versioning.

MCDONALDS CORPORATION: Architected and analyzed large-scale data to deliver high-impact insights and models, driving value across the data lifecycle and supporting strategic decision-making through thought leadership and domain expertise.

- Led the management and architectural design of the **software delivery lifecycle**, emphasizing Continuous Integration and Continuous Deployment (CI/CD) **methodologies** to streamline development and release processes.
- Developed and executed cost and execution optimization strategies for Amazon Web Services (**AWS**) and **Pyspark** query workloads, ensuring efficient resource utilization and cost-effectiveness without compromising performance or scalability.

WALGREENS BOOTS ALLIANCE: Collaborated with teams at Walgreens to validate and preprocess large volumes of data using the IDF framework, enabling robust data quality checks.

- Built scalable ML workflows in **Databricks** and collaborated with teams to align on **MLOps requirements**, supporting model development & training.
- Generated large volumes of **mock data** using **Python/PySpark** for testing **data curation rules**, and conducted **volumetric ETL validation** using the **IDF framework** on **Snowflake** and **Azure Synapse** to ensure performance and scalability.
- Managed **database access and roles** in Snowflake, granting analytics teams secure access to data warehouses.

Wright State University

Aug 2019 – Dec 2020

Graduate Research & Graduate Student Teaching Assistant:

- Research on "Nomophobia" using NLP and ML Techniques on text data. Experience in model evaluation and model building, topic modeling on nomophobia tweets (pre-COVID, COVID-19).
- **Big Data Programming:** Running Hadoop jobs, copying files, enhancing MapReduce k-means algorithm, solving relational analytical tasks using **spark**, and Creating AWS instances and other storage access operations using "boto API." Create a docker image by writing a docker file and installing a docker on an AWS instance using Python Boto3 and the Paramiko Python **SSH library**. **Multi-Cloud Application:** A Guest book application that uses GAE for the web interface and stores data on GAE datastore and Amazon Dynamo DB for better availability. Integrate dynamo DB with GAE by developing a RESTful microservice that runs on an EC2 instance that serves as an interface to the dynamo DB. **Information Retrieval and Text mining** using feature engineering techniques like Feature selection (chi-square and mutual information method), and feature extraction methods have been utilized for text classification (**Multinomial, Bernoulli NB, kNN, SVM**) and clustering (k-means, **Agglomerative**) tasks on newsgroup data and evaluate the performance using **CV, f1-score**, Silhouette Coefficient and NMI (Normalized mutual Index).
- Worked as a teaching assistant for courses "Data Science" and "Computer Programming in Python."

Accenture – Senior Software Engineer India

Jan 2014 - Jul 2018

CALHEERS: an online streamlined application for eligibility and enrollment under the Affordable Care Act (ACA), developing automation frameworks and scripts to validate end-to-end health coverage scenarios and improve testing efficiency.

- Developed a comprehensive automation framework for **CalHEERS**, an ACA-compliant health benefit exchange system, to streamline eligibility, enrollment, and retention workflows.
- Built and maintained automated regression test scripts using **Selenium (Java/Python)** and conducted **build validation** and **sanity testing** for daily builds via **Jenkins**.
- Enhanced test execution efficiency by **80%** through modular function design and performed end-to-end **database management and testing** by optimizing **SQL queries in MySQL**.
- Integrated **SQL Server** and **MySQL** database validation within Selenium and UFT automation frameworks, optimizing backend workflows and ensuring robust data integrity across health coverage applications.
- Led team onboarding and training efforts, conducted **root cause analysis**, and executed **cross-browser and compatibility testing**, improving the overall quality and reliability of the automation suite.

EDUCATION & TRAINING:

- Harrisburg University of Science and Technology- Master of Science in ISEM (GPA of 4.0) - **Present, Harrisburg, PA**
- Wright State University - Master of Science in Computer Science (GPA of 4.0 with a Secondary Major in "Big and Smart Data")- **Dec 2020, Fairborn, OH**
- Malla Reddy Institute of Technology and Science (MRITS) – Bachelor of Technology in Computer Science and Engineering (GPA 3.8) – May **2013, Hyderabad, Telangana**.

CERTIFICATIONS

- AHM - 250 Basic Health Industry Certification.
- AWS Certified Cloud Practitioner - 2023
- Partner Training - Developer Foundations – **Databricks & Solutions Architect Essentials – Databricks -2021**
- Palantir Technologies – Foundry certification exam for Data engineers
- Big and Smart Data certification - Wright State University.

PUBLICATIONS: [Nomophobia Research Article](#), [Techtimes Scientific article](#), [Sciencetime article](#) , [everydayllm](#)