# SHANMUKHA JWALITH KRISTAM

San Francisco (934) 451-9567 jwalithkristam@gmail.com github.com/jwalith in/jwalith-kristam

## WORK EXPERIENCE

**AI Engineer Intern** **Jul 2025 – Sep 2025**
Schizophrenia & Psychosis Action Alliance Alexandria, Virginia

- Developed data extraction using Playwright, JavaScript, and Python to collect and normalize housing data into JSON.
- Engineered a distributed scraping pipeline with Celery and Redis, enabling parallel data extraction across all 50 U.S. states.
- Built an agentic AI chatbot using LangChain RAG, OpenAI API with FAISS vector indexing for knowledge retrieval.
- Optimized retrieval latency and performance through adaptive text chunking, embedding caching, and MMR re-ranking.

**Data Science Intern** **May 2024 – Aug 2024**
Welspot Miami, Florida

- Conducted offline experimentation on AWS SageMaker, developing and comparing XGBoost and RNN for financial risk forecasting, leveraged S3 for data storage and achieved 88% validation accuracy with XGB as the final model.
- Applied Bayesian Optimization (Optuna) to tune XGBoost and used SHAP to interpret and regularize model predictions.
- Improved the F1 score from 0.78 to 0.89 by conducting causal analysis and statistical hypothesis testing (t-tests, ANOVA) on 20+ financial features and implementing time series feature engineering.

**Software Engineer** **Oct 2020 – Jun 2023**
Tata Consultancy Services Bengaluru, India

- Developed scalable full-stack solutions with React-Redux, Angular, and Vue for frontend and ASP.NET for backend APIs, while implementing CI/CD pipelines using Azure DevOps within an Agile-Scrum workflow for iterative delivery.
- Leveraged Lighthouse reports and A/B testing insights to identify performance bottlenecks and user interaction patterns, driving UI optimizations that increased Lighthouse scores from 57.6 to 92.3 and improved responsiveness.
- Integrated SignalR for real-time dashboard updates by enabling instant client-server communication and optimized data caching with Azure Redis, reducing page load time from 10s to 1.5s.
- Optimized database performance by refactoring complex SQL queries and implementing stored procedures, indexing, and query caching, reducing API response times by 35% and improving scalability under high load.

**Machine Learning Engineer Intern** **Mar 2019 – Jun 2019**
SmartBridge Hyderabad, India

- Engineered and Dockerized a computer vision model with TensorFlow using OpenCV, YOLO and MobileNet-V2, achieving 96% accuracy in identifying the pet feeding on an IoT-enabled food bowl.
- Deployed on IBM Watson Studio and integrated with Node-RED, enabling automated workflows for real-time inference.

## EDUCATION

**Master of Science in Data Science**, Stony Brook University, New York (GPA: 3.61/4) **Aug 2023 – May 2025**
**Relevant Courses:** Large Language Models, Statistical Computing, Machine Learning, Big Data Analytics, Probability

**Bachelor of Technology in Computer Science**, GITAM University, India (GPA: 8.1/10) **Jun 2016 – Apr 2020**
**Relevant Courses:** Data Structures & Algorithms, Operating Systems, Object-Oriented Programming, Data Mining, Calculus

## TECHNICAL SKILLS

- **Programming Languages**: R, Python(NumPy, pandas, scikit-learn, PyTorch, PySpark), SQL, JavaScript, C++, C#
- **Data Science & Machine Learning**: A/B Testing, Statistical Analysis, Time Series Forecasting, Causal Inference, Supervised & Unsupervised Learning, Feature Engineering, Model Evaluation, Interpretability, Dimensionality Reduction
- **Cloud & MLOps**: Azure DevOps, GCP, AWS (S3, SageMaker, Lambda, Redshift), Docker, Airflow, Flask APIs, W&B
- **Generative AI**: RAG, PEFT Fine Tuning, DeepSpeed, Langchain, CLIP, Stable Diffusion, GANs, Hugging Face, FAISS
- **Distributed & Backend Systems**: ASP.NET Core, Spring Boot, GraphQL, Flask, Node.js, Express.js, React.js, Celery, JWT

## PROJECTS

**Mental Health AI Coach** **Oct 2024 – Nov 2024**

- Fine-tuned **DeepSeek-7B** using **LoRA adapters** (Q, V layers) on mental health conversations to enhance empathy and contextual accuracy, integrating the model into a RAG pipeline with Gemini as an LLM judge for response reranking and quality evaluation.

**Multimodal Sentiment Analysis System** **Mar 2025 – Mar 2025**

- Built a multimodal sentiment analysis system using the MVSA-Single dataset, combining **CLIP image embeddings with BERT text embeddings**, implementing fusion strategies and serving predictions via API, achieving 87% accuracy.

**Speech Emotion Classification System** **Nov 2024 – Nov 2024**

- Developed a speech emotion classification system using a **CNN-BiLSTM-Attention** model with FastAPI, enabling real-time prediction of emotions from audio with 80% accuracy on benchmark datasets.