Aswin Kumar Janakiraman

Baltimore, MD | +1 667-802-9383 | aswinkj1@umbc.edu | LinkedIn | GitHub | Medium

Education

University of Maryland, Baltimore County **Baltimore, MD**
MS in Information Systems *(specializing in Artificial Intelligence & Data)* Aug 2023 - May 2025 (expected)
Relevant Courseworks: Deep learning with PyTorch, Information Extraction, Data mining & Analytics using Tableau, Machine learning using PyTorch, Multi-Cloud Computing and Resource Management using Python & Terraform

Skills

Languages: Python3, JavaScript, PowerShell, TypeScript, RUST
Databases: SQL, DynamoDB, PostgreSQL, MongoDB, PL/SQL, CosmosDB, Oracle Database 23ai, Pinecone
Technologies: Flask, Django, FastAPI, GraphQL, React.js, Redux toolkit, Next.js, Nuxt.js, REST, Microservices, Kafka, PySpark, ROS, JSON
Cloud: AWS EC2, S3, IAM, SageMaker, Bedrock, Lambda, Route 53, ElasticCache, ECS, GCP Vertex AI Studio
AI/ML: PyTorch, Langchain, TensorFlow, HuggingFace, Transformers, SNN, LLM, Agents, MCP
Tools: Git, VSCode, Docker, Azure DevOps, Jira, Jenkins, Terraform, OAuth/2, Jupyter Notebook, Google Colab, Tableau, OpenCV, PyLint, JsLint
Testing Frameworks: Unit Testing, Integration Testing, and A/B Testing

Work Experience

University of Maryland, Baltimore County **Baltimore, MD**
Graduate Research Assistant - CARDS Nov 2023 - Present
- Engineered an end-to-end navigation system of swarm robotics for the fleet of 3 Boston Dynamics SPOT and 2 Clearpath Husky using Python3 and ROS, enhancing the cross-platform navigation system by 70%
- Optimized voice-based Llava-1.5b VQA model using RAG and Prompt Engineering, improving image interpretation accuracy by 90% and reducing processing time to 1.2 seconds per response
- Developed voice-enabled, context-aware navigation systems for robotics, utilizing advanced NLP and open-source voice models like Coqui-XTTS and Kyutai's Moshiko to enhance robot-human interaction
- Architected and developed a multi-agent coordination system using Langchain AI agents and Model Context Protocol to execute high-level voice commands, and improved autonomous mission success by 85%

University of Maryland, Baltimore County **Baltimore, MD**
Full Stack Developer/ML Engineer Dec 2024 - Jan 2025
- Developed front-end for educational web app using Nuxt.js, improving user experience for 2 K+ daily users across multiple US universities with a 90% conversion rate using A/B Testing
- Designed the application using AWS EC2 backend, MongoDB, and deployed it with an auto-scaling setup and Load-Balancer, and Web Application Firewall to improve security and availability by 85%
- Integrated RAG trained Meta 3.2 3B Instruct model from AWS Bedrock with backend using REST, increasing student engagement with the caselet tool by 80%
- Created and pipelined an adaptive student performance predictive model using collaborative filtering algorithms with the backend, improving audience engagement by 95%

Vagus Technologies Inc. **Trichy, India**
Lead Software Engineer Jul 2018 - Aug 2023
- Developed scalable web applications for an applicant tracking system, increasing system reliability by 85% and reducing downtime by 60%
- Designed and configured a caching database for backend applications using AWS ElastiCache, improving application efficiency by 70%, as measured by reduced API response times during peak loads
- Consumed APIs while utilizing Python requests to read numerous JSON reports and file automatic bugs for intermittent bugs during the testing phase
- Implemented RESTful APIs for data migration to AWS cloud services, reducing migration time by 65% and ensuring 99.9% data integrity
- Utilized Docker for containerization and AWS services (EC2, S3, Route 53), streamlining data transfer and

reducing latency by 60%

Internships

Headstarter AI **Remote, US**
SWE Fellow Jul 2024 - Aug 2024

- Deployed a chatbot solution using AWS Bedrock, Pinecone, and Next.js, improving the airline's customer support efficiency by 85% based on the customer satisfaction report
- Enhanced chatbot model with multi-language support and user authentication using Firebase and Clerk, expanding functionality to assist broader customer sections
- Built a web scraper pipeline using BeautifulSoup and integrated it with a RAG system using LlamaIndex and GPT-3, enhancing response accuracy and relevance for RateMyProfessor 2.0
- Orchestrated coding, design, and deployment best practices using EC2, while collaborating with Y Combinator investors and engineers to implement industry-leading methodologies.

Google Summer of Codes ’16 - Forced Alignment of words (RedHen Labs) **Remote, India**
Summer Intern Mar 2016 - Jul 2016

- Engineered a forced alignment system using Kaldi ASR, SRILM, and IRSTLM, optimized for HPC clusters, reducing alignment time by 60% for large-scale news video datasets
- Developed Python scripts to automate the alignment workflow, increasing processing speed by 75% and enabling the system to handle 500+ hours of video content daily
- Integrated Edinburgh Speech Tools for advanced phonetic analysis and feature extraction, significantly improving word-level alignment precision
- Collaborated with Red Hen Lab to integrate the system into their framework, resulting in a 40% increase in research output for multimodal communication studies

Projects

Smart Code Refactoring Application Feb 2025 - May 2025

Developed a smart code refactoring application using TypeScript, Terraform, GraphQL, AWS, and transformers

- Designed and implemented the core logic using the CodeBERT transformer model deployed in AWS Sagemaker to analyze and translate code with 90% of semantic equivalence
- Designed scalable infrastructure with Terraform to manage AWS Lambda, Bedrock, S3, and AppSync services in private and public subnets, and reduced the manual overhead by 75%

NSF HDR ML Challenge - Sea Level Anomaly Detection Nov 2024 - Jan 2025

- Developed an anomaly detection model based on 20 years of real-time sea level data from 20 coastal regions of the US using AWS Sagemaker, S3, Kinesis, and Athena, with a precision of 89%

AI-based Customer Chatbot for Delta Airlines Aug 2024 - Oct 2024

- Developed an AI-powered chatbot for customer support using a custom RAG pipeline built with Llama 3.1 8b model and LangChain, Next.js, OAuth, AWS Bedrock, and Pinecone
- Developed multi-language input queries, collected user feedback to train the model and user authentication, and designed for scalability and user engagement using Next.js, React, and Boto3

Publications & Achievements

- AAAI 2025 Spring Symposium: Edge LLMs for Real-time Contextual Understanding for Robots
- Awarded 3rd runner-up at nationwide NSF HDR Machine Learning Challenge - 2025

Certificates

- AWS Solution Architect (Associate)
- Neo4j Certified Professional
- Oracle Cloud Infrastructure 2024 Generative AI Professional
- AWS Educate Machine Learning
- OpenCV TensorFlow-Keras Developer
- AWS Certified Cloud Practitioner