Azaan Barlas

(909) 610-0066 | azaanbarlas@gmail.com | linkedin.com/in/a-barlas | github.com/kingazaan

Education

University of Illinois, Urbana-Champaign — Champaign, IL
Master's Degree, Computer Science; Graduate Teaching Assistant for CS 437 — *May 2024*
Courses: Natural Language Processing (NLP), Machine Learning, Deep Learning for Healthcare, Bayesian Statistics, Computer Vision, Cloud Computing

California State University, Fullerton — Fullerton, CA
Bachelor's Degree: Computer Information Science, Minor in Computer Science — *May 2021*
Courses: Database Management, Data Science and Big Data in R, Data Structures, Algorithms, Data Mining, Statistics I & II, Business Intelligence, Distributed Systems, Data Visualization

Experience

Freelance Software Engineer — *August 2024 – Present*

Lead Software Engineer — *November 2024 – Present*

- Developed a full-stack AI Multi-Agent System (MAS) with a director agent and role-specific agents for calendar, marketing, and financial planning tasks using CrewAI, Flask, and Next.js; deployed frontend on Vercel and backend on Render for a startup entering its seed round
- Led a team of three developers and a project manager to create an AI assistant for the music industry, ensuring integration of financial, calendar, and venue modules while maintaining structural consistency and performance
- Integrated OpenAI embeddings for a custom memory class enabling context-aware agent collaboration through Retrieval-Augmented Generation (RAG), preserving agent independence in multi-tasking scenarios
- Deployed and tested agent interactions via Flask-based API, connecting to Cloudflare APIs for LLMs; integrated Google Calendar API for scheduling and Cloudflare's NoSQL database for financial data updates
- Collaborated with clients in the music industry to customize a touring agent for venue booking, competitor analytics via Spotify API, and real-life scenario training, supporting the startup's mission to empower solopreneurs and musicians as small businesses by providing tailored financial and operational tools

Software Engineer — *August 2024 – December 2024*

- Developed a full-stack RAG-based LLM application for FAIR Latter-Day Saints, utilizing embeddings and vector databases to enable semantic search and integrate FAIR information and links into responses
- Developed a Generative AI healthcare application: an LLM-based multi-label classifier aimed at predicting labels from 20,000+ research articles using LLaMA 3; achieved 0.67 F1 score after LoRA fine-tuning
- Prompt Engineered and used Multi-Shot learning enhanced with semantic search capabilities with NVIDIA V100 GPUs onto LLaMA models using Pytorch and HuggingFace to extract important entities from healthcare research papers, rivaling the accuracy and precision of the industry benchmark model PubMedBERT

University of Illinois Urbana-Champaign | Data Scientist — *June 2023 – July 2024*

- Engineered and deployed an LLM featuring semantic search capabilities using transformers across 1,600 UIUC policy documents for internal use regarding university policy information, reducing search times by $\sim 85\%$
- Conducted research on legislative voting patterns and built LSTM model through PyTorch that predicts bill approvals with a 78% AUC score; laid foundation of an alert system for highly likely-to-pass bills
- Compiled legislation data by web scraping legislative websites, preprocessing data into PostgreSQL database, and categorized 200,000+ Illinois bills via LDA Topic Modeling into KNIME ETL platform; enabled administrative team to quickly identify bill type and increasing productivity by $\sim 40\%$
- Constructed student graduation-likelihood prediction ensemble model using a random forest classifier and ARIMA for UIC; achieved an 83% accuracy, enabling admissions committee to estimate tuition budgets for the next year
- Designed and deployed an Azure-based pipeline utilizing GPT-4 for automated entity extraction from legislative text, integrating OpenAI APIs with cloud-based data aggregation workflows for efficient and scalable analytics

Department of Computer Science at UIUC | Graduate Teaching Assistant — *August 2024*

- Refactored existing Raspberry Pi and camera autonomous vehicle lab by introducing Arduino Zumo and Pixy2 module as new hardware to decrease lab costs by 50%; programmed car to drive through an obstacle course I designed in a Mario Kart-like fashion with power-ups and boosts through Arduino Code, as a fun lab experience for student engagement
- Designed 3 new demo "live" labs: AWS architecture for a content delivery network (Netflix), Wireless spectrum analysis with access point deployment and Packet Tracer network configuration, and oscilloscope data visualization

Esperanto Technologies | Artificial Intelligence/Cloud Intern — *June 2022 – September 2022*

- Implemented Meta's DLRM on a 1TB Criteo dataset for use in Esperanto's ET-SoC-1 (system-on-chip) demo, optimizing recommendation performance; used in Esperanto's ET-SoC-1 demo to semiconductor clients

- Set up an AWS Aurora database, using shell script for data ingestion from S3 bucket containing 1TB data
- Developed an optimized C script to rapidly compute summary statistics (count, min, max, mean) on a 1TB Criteo Advertisement dataset, significantly reducing processing time

City of La Habra | Data Engineer *March 2021 – June 2022*

- Spearheaded the migration of La Habra's legacy data systems to the cloud, optimizing data storage in AWS S3 and DynamoDB
- Developed Python scripts in AWS Sagemaker to establish data ingestion into AWS S3 from local legacy database, utilizing Lambda and DynamoDB for enhanced data processing and distributed storage
- Connected La Habra's proprietary platform with city Databank using APIs to leverage in-house datasets; this enabled the Management Analyst team to utilize specialized datasets for analyzing policies, budgets, contracts, procedures, and services

Projects

AI Engineer Project – Healthcare Chatbot (Github)

- Developed a full-stack Django-based healthcare chatbot using LangChain that enables patients to interact with an AI bot for health-related inquiries, medication management, and appointment rescheduling;
- Integrated a Neo4j knowledge graph and PostgreSQL to store patient data, extracted from user conversations via BioClinicalBERT and regex matching; implemented LLM-agnostic design using Langchain
- Trained a binary classifier to filter health-related queries using MedQuAD and Natural Questions datasets

ICU Readmission Prediction using Deep Learning Architectures (Github)

- Utilized deep learning models (RNNs, attention mechanisms, mixed models) to predict ICU readmissions with the MIMIC-III dataset (45,298 ICU stays, 33,150 patients)
- Evaluated performance using F1 Score, PPV, NPV, AUROC, and Precision, improving results by developing an RNN + Logistic Regression mixed model

HackIllinois John Deere Autonomous Litter Detecting Rover (DevPost)

- Programmed the Raspberry Pi rover with motors and camera to autonomously traverse an area and detect litter using Python's OpenCV for object detection; fine-tuned a lightweight YOLO model with TACO litter dataset of 10,000 images
- Finalist in the competitive Hack Knights path; presented rover in final showcase with top groups

LLM Chatbot for Academic Resource Assistance (Github)

- Developed GPT and BERT Language Models (LLM) integrated in Flask-powered chat-bot to answer academic questions
- Preprocessed text and created GPT model from pure Numpy, such as calculating multi-head attention weights, transformer blocks, and forward methods for GPT decoder; Created CNN to evaluate user responses to chat-bot from scale of 1-10 based on IMDB number and review correlations
- Fine-tuned BERT and GPT-2 LLMs on 96 lecture transcripts to answer student questions, reducing search time by $\sim 80\%$

Wildfire Predictions Showcase (Presentation)

- Analyzed wildfires in major cities in California and established statistical correlations between climate change and fire occurrences
- Split fire types into low, medium, and high intensities; classified these with regression model r2=0.76 and predicted future fires until 2025 with Random Forest + ARIMA model; Received 1st place in competition

Technical Skills

Languages and Frameworks: Python, R, C++, SQL, JavaScript, TypeScript, Next.js, Node.js, Flask, Django, TensorFlow, PyTorch, Scikit-learn, LangChain, CrewAI, Bash, Linux/Unix, Arduino IDE, Raspberry Pi
Developer Tools: AWS (S3, RDS, Redshift, Lambda, EC2, SageMaker, Glue), Azure, Vercel, Render, Cloudflare, Docker, Kubernetes, Git, PostgreSQL, Neo4j, Huggingface
Concepts: Generative AI, Cloud Computing, Machine Learning, Healthcare AI, Large Language Models, Web Development, CI/CD Pipelines and DevOps, RESTful APIs, Database Architecture
Honors: HackIllinois Hack Knights finalist, UIUC Data Science showcase winner, Thomas Siebel Tuition Scholarship