Vaibhav Deokar

vadeokar@syr.edu +1 (832) 332-7919 portfolio/vaibhavdeokar linkedin/vaibhav-deokar

Education

Syracuse University May 2027

Master of Science (M.S.) in Applied Human-Centered Artificial Intelligence

- **Relevant Coursework:** Generative AI Models, Deep Learning, NLP, Human-AI Interaction

Mumbai University May 2022

Bachelor of Engineering (B.E.) in Electronics Engineering

- **Relevant Coursework:** Object-Oriented Programming, Database Management Systems with SQL, Computer Network, Cyber Security

Experience

Associate Data Scientist – GenAI July 2024 – July 2025

Cognizant Technology Solutions *Mumbai, India*

- Built and deployed an autonomous AI Assistant for Google gTech Ads utilizing LLMs with grounding techniques on internal data sources like Plx to recommend Ads solutions to Ad agencies and patners; operated within Google's XWF development ecosystem using Bazel and Piper; delivered targeted Ad solutions to 150+ agencies and partners, reducing manual query resolution time by 45%.
- Designed internal researcher-facing tool with RAG, Neo4j Knowledge Graphs & Langchain, accelerating knowledge retrieval from unstructured logs by 60% and enhancing solution accuracy
- Developed and secured a GenAI Interview Bot (Azure GPT-4, Bedrock Claude 3.5) for HR; automated screening across 20+ roles, reducing interview scheduling effort by 50 hours/month; led red-teaming (jailbreaks, prompt injection) & MLOps (Kubernetes, CI/CD), validated with 1350 employees.

Programmer Analyst - AI ML May 2022 – June 2024

Cognizant Technology Solutions *Mumbai, India*

- Led document retrieval system for a Johnson & Johnson client, leveraging RAG, GPT-3, and optimized Faiss vector search, boosting solution discovery rates by 61% across internal medical literature.
- Engineered an AI-driven system for Johnson & Johnson automating the categorization and semantic tagging of complex medical information requests using embeddings and t-SNE classification with Random Forest and K-Means (reducing response times by 57%) and accurately routing inquiries to subject matter experts with 86% precision.
- Leveraged advanced NLP techniques with tools like TensorFlow, PyTorch, Transformers, BERT, and T5 for complex data processing and contextual understanding across projects.
- Utilized models including GPT-2 and Salesforce CodeGen, optimizing inference speed with CUDA.
- Optimized data processing pipelines (Python, Sklearn, Docker) & performed analysis (AWS S3/SageMaker, Pandas, NumPy) on 50+ datasets, enhancing efficiency.

Skills

Programming & Data Science: Python, Java, JavaScript, C/C++, NumPy, Pandas, Scikit-learn, SQL and Bash scripting
AI/ML Research & Theory: CNNs, RNNs, LSTMs, Transformers, GANs, Variational Autoencoders, Diffusion Models, Reinforcement Learning, Optimization Algorithms, Benchmarking & Evaluation Metrics
Generative AI & LLMs: Expertise with LLMs (OpenAI GPT series, Claude, Gemini, Llama, BERT, T5), Diffusion Models, Multimodal Models, Prompt Engineering, RAG, Agentic Systems, Safety Evaluation, Hugging Face and Langchain
MLOps & Cloud Infrastructure: Docker, Kubernetes, CI/CD pipelines, Git, AWS, Azure, GCP, Terraform, CloudFormation
Data Technologies & Web: Proficiency with SQL and NoSQL databases, Faiss, Pinecone, Neo4j, data visualization, Jupyter Notebooks, API design (gRPC, RESTful, GraphQL), React, Node.js, Django, Flask

Projects

Knowledge Graphs for RAG with Neo4j | *Neo4j, LangChain, Python* Link to Project

- Developed a Neo4j knowledge graph enhancing Retrieval Augmented Generation (RAG) accuracy; designed schema, extracted entities/relationships, implemented graph queries, and integrated with LangChain enabling precise hybrid retrieval improving factual grounding accuracy by 32% in internal benchmarks from 500+ structured data.

Deploy ML model on REST API & Web App | *Python, TensorFlow, Django REST API, React* Link to Project

- Built and deployed a handwritten digit recognition CNN using TensorFlow, served via a Django REST API integrated with a React web application; explored deep learning image recognition using a dataset of 70,000 images.