

Hemesh Raaja Malathi

☎ 930-333-5409 | ✉ heraaj@iu.edu | 🌐 hemesh0204.github.io | 📄 [Hemesh0204](#) | in [hemesh-r-m](#)

EDUCATION

Indiana University Bloomington

Masters in Computer Science: 3.86/4

Bloomington, IN

August 2023 - May 2025

RELEVANT COURSEWORK

Data Structure & Algorithm, Machine Learning, Artificial Intelligence, Computer Vision, Database Management System, Natural Language Processing, Business Analytics, Large Language Models, Cloud Computing

WORK EXPERIENCE

Data Scientist

February 2025 - Present

Project 990

Bloomington, Indiana

- Fine-tuned transformer models (RoBERTa, BART, LLaMA 3-7B) on 14K+ mission statements with automated GPU/SLURM workflows, achieving 97% accuracy and reducing manual labeling effort
- Implemented generative AI pipelines to create synthetic mission statements for nonprofits missing textual data, grounding outputs with grant metadata, and validating accuracy through LLM agreement scoring
- Developed and deployed an AI-powered internal chatbot using RAG with OpenAI GPT-3.5, improving document retrieval times by 20% and enabling faster issue resolution for employees
- Integrated Alibi Detect into the ML pipeline to monitor data drift and exposing drift metrics to Prometheus enabling proactive detection of model degradation

Data Engineering Intern

June 2024 - December 2024

Reinsurance Group of America

Chesterfield, Missouri

- Built an AWS Lambda function to shut down idle EC2 instances provisioned by EMR clusters after 24 hours; automated deployment with Jenkins, integrated Datadog monitoring and Slack alerts, saving \$1,000/month

Data Scientist

December 2022 - August 2023

New Pro Data

Madurai, India

- Built and deployed an NLP resume parser with MPNet embeddings (95% extraction accuracy, 93% better match quality than BERT via A/B testing) and integrated it into a Django-based HR chatbot, reducing manual screening by 60% and saving \$3K–\$5K annually

PROJECTS

Semantic Book Recommender | *Python, Hugging Face, LangChain, Qdrant, Gradio*

March 2025

- Built a content-based book recommender using Hugging Face's all-MiniLM-L6-v2 to embed 7,000+ descriptions into 384-dimensional vectors; integrated LangChain and Qdrant for cosine similarity search
- Deployed a Gradio app on Hugging Face Spaces and incorporated a zero-shot classification model to distinguish fiction vs. non-fiction with ~80% accuracy

Cookbook: AI Recipe Generator | *LLM (LLaMA), Prompt Engineering, React, Firebase*

March 2025

- Designed a prompt-driven recipe generation system using GroqCloud's hosted LLaMA model, generating structured meal plans with ingredients and instructions based on user preferences
- Implemented filters for cuisine type, dietary goals, and allergies to condition LLM outputs dynamically; stored interaction history and results using Firebase for personalization and recall

AI Agent for Flappy Bird Simulation | *Python, Pygame, NEAT*

December 2024 - January 2025

- Built an AI-based Flappy Bird simulator using the NEAT algorithm to evolve neural networks for autonomous gameplay; trained agents to optimize movement based on pipe position, altitude, and velocity
- Implemented realistic gameplay mechanics, including collision detection and dynamic pipe difficulty, to simulate a progressively challenging environment for evolutionary training

TECHNICAL SKILLS

Languages & Databases: Python, C/C++, SQL, Flask, Django, PostgreSQL, Snowflake, Databricks

Data Science & AI: Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, YOLOv8, OpenCV, LangChain, LlamaIndex, Retrieval-Augmented Generation (RAG), Prompt Engineering, LoRA/PEFT, RLHF, Agentic Frameworks (AutoGen, CrewAI), LangSmith, LLM APIs (OpenAI GPT-4, Claude, Gemini)

Data Engineering : Airflow, Kafka, Hadoop, dbt, DVC, DagsHub, MLflow, Git, GitHub Actions, Docker, Jenkins, Terraform, Kubernetes, Astro, Grafana, Prometheus, Vector Databases (FAISS, Pinecone, Qdrant)

Cloud: AWS (EC2, S3, Redshift, Lambda, SageMaker), Azure, GCP (BigQuery, Vertex AI)