MAHADEV MAHESH MAITRI

+1 (302) 685-6915 | mmaitri@udel.edu | github.com/mahadev9 | maitri.pro

TECHNICAL SKILLS

Programming Languages: Python, JavaScript, TypeScript, C/C++, Kotlin, HTML, CSS, SQL, Golang, GraphQL Frameworks: Node.js, React, Bootstrap, Redux, Next.js, TailwindCSS, Flutter, FastAPI Testing Frameworks: JUnit, Jest, Robot Framework, Selenium Cloud & Databases: Amazon Web Services (EC2, ECR, EKS, Bedrock, MemoryDB, EBS, EFS, S3, SQS, Lambda, DynamoDB, RDS), Google Cloud, MongoDB, PostgreSQL, Google Firebase, Docker, Kubernetes ML Libraries: PyTorch, TensorFlow, HuggingFace (Transformers), LangChain, LangGraph, Keras, Scikit-Learn, Gymnasium Developer Tools: Git, Postman, SonarQube, Jenkins, JMeter, Chrome Developer Tools PROFESSIONAL EXPERIENCE Fulcrum Digital - Allied World Python Al Developer • Engineered real-time question answering for insurance operations by developing a FastAPI application integrating Ollama (Ilama3.1) with PGVector for contextually relevant document retrieval supporting underwriting, claims processing, and risk analysis workflows.

- Streamlined automated insurance document processing as measured by 15% reduction in manual policy review time, by building LangGraph **agent-based systems** that execute policy extraction, risk assessment summarization, and intelligent claim routing workflows.
- Achieved 87% accuracy in insurance risk assessment as measured by precision metrics across 2,000+ synthetic scenarios, by optimizing RAG systems with evaluation pipelines and implementing retrieval agents for structured claim pattern analysis.
- Enhanced question answering performance as measured by 20% improvement in response relevance scores, by integrating specialized AI agents with enterprise database tooling that autonomously refine user prompts and execute multi-step reasoning chains.
- Improved document retrieval precision as measured by 25% increase in search relevance scores, by engineering a custom vector database retriever with **LLM-based reranking** capabilities using Hugging Face transformer models.
- Delivered **99.8% system availability** as measured by service uptime metrics, by deploying and managing AWS EKS with horizontal pod autoscaling architecture handling 300+ concurrent requests during peak usage periods.
- Reduced operational costs as measured by 20% decrease in monthly infrastructure spend, by developing a self-hosted OCR service using Vision Transformer (ViT with Phi3.5 vision) deployed in EKS with **Ray clusters** for distributed computing.
- Optimized GPU infrastructure costs as measured by 60% reduction in monthly EC2 instance spend, by migrating the entire codebase from local Ollama deployment to AWS Bedrock API integration, eliminating on-premise GPU dependencies.

Department of Mechanical Engineering - University of Delaware

Research Assistant

- Compiled a cross-platform application utilizing the Google Maps API to provide timely alerts about approaching intersections based on live signal timings, enhancing navigation **decision-making**.
- Designed an algorithm to identify approaching intersections within a specified radius under 50ms, displaying signal data from an MQTT server corresponding to the upcoming signal phase using map data.
- Implemented an alert system using a **dilemma zone algorithm** to warn users 8 seconds before approaching intersections, utilizing real-time traffic data to avoid crossing red signals, accidents and enhance road safety.

Optum - UnitedHealth Group

Software Engineer

- Accomplished seamless data migration, moving **100,000+ records with 98%** accuracy, by developing **ETL pipelines** using Python and AWS Glue to transfer eligibility data from legacy systems to a centralized data warehouse, AWS Redshift.
- Improved data extraction accuracy by 95%, processing 10,000+ forms automatically, by using NLP (SpaCy, NLTK) to parse eligibility criteria and applicant details, reducing manual effort and speeding up processing with predictive models deployed on AWS SageMaker.
- Increased document processing efficiency by **85%**, extracting key details from 5,000+ scanned forms, by **integrating OCR** using Tesseract, automating data entry and accelerating workflows.
- Achieved real-time processing with **95% accuracy**, by integrating **NLP and OCR models via FastAPI** to handle incoming applications, while ensuring low-latency responses with PyTorch deep learning models for eligibility assessments.
- Enhanced front-end microservice performance by eliminating API dependency for generating UI elements, improving UI design, and utilizing Redux for **better data management**, resulting in faster page loading and contributing to the project's success over a year.
- Optimized application efficiency by removing unnecessary API calls and implementing generalized and appropriate UI validations, leading to better usability and accessibility.
- Executed functional regression testing with Robot Framework and **performance testing** with JMeter to meet 97% SLA of RESTful APIs in the OMMS project, ensuring system robustness and reliability by executing and analyzing test suites bi-weekly.

Electronics and Radar Development Establishment (LRDE), DRDO

Trainee

- Led the development of an advanced Automatic Target Recognition solution for ships using Deep Learning algorithms, improving accuracy and efficiency in ship identification.
- Engineered a GAN with CNNs to enhance ship recognition accuracy by expanding the ISAR image dataset.
- Implemented transfer learning with MobileNetV2 for ship classification, achieving 89.6% accuracy and showcasing cutting-edge proficiency.
- Created an authentic and reliable ISAR image dataset with ANSYS Electromagnetics SBR+ and 3-D ship models.

Newark, DE

Bengaluru, India

Bengaluru, India Jan 2020 – May 2020

Jul 2020 – Jul 2022

Jan 2024 – Feb 2025

Wipro IISc Research Innovation Network (WIRIN), IISc

Project Intern

- Enhanced system efficiency by implementing sequence-by-sequence batch augmentation for video frames, resulting in accelerated inference timing and increased frames per second for input video streams.
- Improved the performance and reliability of video segmentation processes by optimizing the existing segmentation algorithm and connecting with above algorithm. Reduced processing time and enhanced overall algorithm accuracy, contributing to overall system performance.

EDUCATION

University of Delaware Newark, DE Master of Science in Computer Science - GPA: 4.0 Aug 2022 - May 2024 Relevant Courses: Algorithm Design, Advanced Deep Learning, Cybersecurity, Network Analysis, Compiler Construction **R. V. College of Engineering** Bengaluru, India Bachelor of Engineering in Electronics and Communication Engineering Aug 2016 - Jul 2020

PROJECTS

Formula 1 Race Strategy Optimization with Deep Reinforcement Learning

Python, PyTorch, Gymnasium

- Designed and trained a Deep Q-Learning agent to optimize race strategies in Formula 1 by creating a data-driven environment using real-world racing data from FastF1.
- Developed a Markov Decision Process (MDP) with a reward system that considers pit stop penalties, tire wear, and lap times, enabling the agent to make strategic decisions about pit stops and tire selection to achieve optimal race performance.

Sudoku using Backtrack Search

Python, Flask, HTML, CSS

- Developed a Flask application for Sudoku puzzles, integrating AC3 and Backtracking algorithms for solving predefined and custom puzzles.
- Enhanced user experience with step-by-step, color-coded solutions and efficiently managed failures using binary constraints in AC-3.

Base Language Compiler

Kotlin, Antlr, Intellij IDEA

- Developed an Object System by defining class structures, modifying the lexer and parser to add reserved keywords and object identifiers, implementing method calls, and ensuring semantic analysis, resulting in enhanced functionality and code maintainability.
- Implemented First-Class Functions by modifying the lexer and parser to support function types as parameters and adding new AST nodes, enabling the creation and utilization of functions as first-class entities, facilitating more flexible and powerful coding paradigms.

Cryptocurrency Price Forecasting

Python, TensorFlow, Numpy, Pandas, Matplotlib

- Collaborated with a team to develop neural networks for cryptocurrency price prediction, employing ablation studies for model optimization.
- Utilized Binance Data Vision for data analysis, determining that a single-layer LSTM with 128 units yielded the most accurate predictions.
- Validated 1D-CNN with LSTM for integrating supplementary data streams, proving its effectiveness for better trading strategies.

ACHIEVEMENTS

- Won the Best Artificial Intelligence/Machine Learning Hack category at HenHacks 2024 for the innovation in Yoga Pose Detection and Pose Correction with an Android app. (https://devpost.com/software/yoga-pose-detection-and-pose-correction)
- Won the Creativity Award at the DS+AI Hackathon 2023 for the work towards the "Generating High-quality, Fine-scale Precipitation Dataset for the Great Lakes Region Building upon Existing Dataset". (https://sites.udel.edu/ai/dsai-summer-hackathon-2023/)
- Published a paper on "Face Aging Through Uniqueness Conserving by cGAN with Separable Convolution" in Sustainable Communication Networks and Application, Proceedings of ICSCN 2020.
- Secured the second place out of 30 participating teams in Hackathon (Blockchain Hackathon 2019) organized by NextGrids and powered by JUINCUBATOR, a TBI supported by DST, Government of India and JAIN(Deemed-to-be-University).
 - Built a smart contract using Ethereum for verification of employee background check and deployed on to IBM Blockchain platform. To demonstrate, we developed a sample Angular website to perform the transactions.
- Secured the first place out of 50 participating teams in Hackathon (Stay Late And Code, 2019) organized by Amrita School of Engineering sponsored by General Electricals Healthcare.
 - Built a Machine Learning model for Speech Recognition (offline) with help of Recurrent Neural Network and implemented an Android Application using the same model's lite version for efficient computation as demo which could be implemented in medical appliances.
- A finalist (from 70 teams) in India Police Hackathon 2019 organized by Karnataka State Police at R. V. College of Engineering.
- A finalist (from 30 teams) in Hackathon (TECHATHLON 2018) organized by Curl Analytics in association with Coding Club of RVCE.

VOLUNTARY WORK

• Served as the Website Manager for IGSA, overseeing the comprehensive redesign of the WordPress. Implemented enhancements to showcase upcoming events, student information and news, and curated photos and videos captured during events. Additionally, took a leadership role in planning and executing various events hosted by IGSA. sites.udel.edu/igsa/

Bengaluru, India Jun 2019 - Dec 2019

Feb 2024 - May 2024

Feb 2023 - May 2023

Feb 2023 - May 2023

Sept 2022 - Dec 2022

TRAINING AND CERTIFICATIONS

- Natural Language Processing Specialization from DeepLearning.AI. (2024) (Link: <u>9JE9SXH8ZXMJ</u>)
- Natural Language Processing with Attention Models from DeepLearning.AI. (2024) (Link: <u>7WQ8Z3KGVWDY</u>)
- Natural Language Processing with Sequence Models from DeepLearning.AI. (2024) (Link: <u>JBDUBKN7XB75</u>)
- Natural Language Processing with Probabilistic Models from DeepLearning.AI. (2024) (Link: 2SULGURPLQTG)
- Natural Language Processing with Classification and Vector Spaces from DeepLearning.AI. (2024) (Link: TV9269RX7L7M)
- Structuring Machine Learning Projects from DeepLearning.AI. (2020) (Link: <u>2JGT88YLP7AR</u>)
- Improving DNN: Hyperparameter Tuning, Regularization and Optimization from DeepLearning.AI. (2020) (Link: Z4L6ZPBD5XQT)
- Neural Networks and Deep Learning from DeepLearning.AI. (2020) (Link: <u>XCLPC88BC5N3</u>)
- Machine Learning Basic Nanodegree from Udacity India. (2018) (Link: https://confirm.udacity.com/RSXCMEGL)
- Android Foundation Nanodegree II from Udacity India. (2018) (Link: https://confirm.udacity.com/HJP5C35N)