ATANU DAHARI

7530 Brompton St, Apt 790, Houston, TX, 77025 | atanudahari@gmail.com| LinkedIn | (713)-(501)-0136

EDUCATION

Master of Computer Science (MCS) August 2021 – December 2022
Rice University, Houston, Texas CGPA: 3.43 / 4.0
Relevant coursework: Deep Learning, Computational Statistics, R, Database management Systems, Web development

Bachelor of Technology in Electronics and Communication Technology (B. Tech in ECE) July 2017 - June 2021
Vellore Institute of Technology, Vellore, India CGPA: 8.62 / 10.0
Relevant coursework: Data Structures and Algorithms, Cloud Computing, Object oriented programming, Advanced Java

WORK EXPERIENCE

Machine Learning Engineer February 2024 – Present
Circle.ooo Houston, TX

- Created an AI assistant using RAG for an event hosting platform to streamline the process of attending and hosting events on the website.
- Integrated data from various APIs and vector databases, with Gemini models using function and tool calling.
- Automated event creation, finding sponsors and audience management, resulting in a 25% increase in user interaction.

AI Engineer May 2023 – January 2024
WarrantyMe Houston, TX

- Developed and implemented a warranty information extraction system, automating data collection from Gmail accounts, Amazon orders, and invoice images.
- Extracted warranty data from user emails using transformer models in Spacy and NER algorithms with an accuracy of 97.5%.
- Extracted warranty data from invoice images using Power Automate's AI Builder model and engineered a zero-shot classifier to identify products from Amazon orders.
- Developed an RPA agent to automatically file warranty claims for damaged products, using GPT-4 Vision and Selenium, increasing user convenience.

Machine Learning Engineer June 2023 – December 2023
Affekta LLC Houston, TX

- Deployed computer vision models to interpret real-time emotions of students attending online lectures, on Azure Machine Learning Studio, enabling monitoring and feedback.
- Optimized deployment efficiency by testing and validating models locally with the Azure Machine Learning inference HTTP server, and configuring environments using Anaconda, reducing setup time by 30% and ensuring consistency across project stages.

PROJECT EXPERIENCE

Green Connex AI Recycling Project, Westlake Dimex

- Collaborated with industry sponsors and academic advisors to develop a computer vision AI model using PyTorch, classifying industrial scrap images by quality standards.
- Adopted data augmentation and loss adjustment techniques to deal with an imbalanced label distribution dataset.
- Achieved an accuracy of 81.1% on the validation set with the best performing model.

Detecting harmful social media memes using a multi-modal model:

- Detected multimodal hate speech by engineering OpenAI's CLIP (Contrastive Language–Image Pre-training) model, and compared its results to other state-of-the-art models presented by Facebook AI, achieving an accuracy of 57.8%.

Graduate Research Assistant, Rice University

- Optimized search algorithms in IoT applications using SIMD (Single Instruction Multiple Data) accelerated software.
- Developed vectorized algorithms using Intel SSE and AVX instruction sets, and benchmarked their execution speeds against the standard Memchr library in C and various state of the art scalar searching algorithms.
- Achieved a 15x performance boost over scalar algorithms, significantly improving data processing speed in IoT applications.

Flight delay and experience analysis using R:

- Conducted a comprehensive analysis of delays and travel experiences for major U.S. commercial airlines, utilizing R libraries such as ggplot and tidyverse for data visualization and statistical modelling.
- Developed interactive dashboards using Shiny to provide actionable insights for strategic decision-making, through detailed visualizations and regression plots.

SKILLS

- **Skills:** Retrieval Augmented Generation (RAG), Computer Vision, Natural Language Processing (NLP), Robotic Process Automation (RPA), Single Instruction Multiple Data (SIMD)
- **Databases:** PostgreSQL, SQLite, MongoDB, Neo4j
- **Languages:** C/C++, Java, Python, MATLAB, HTML5, CSS, JavaScript, React, Node.js
- **Tools:** IntelliJ, Visual Studio, R Studio, Azure Machine Learning Studio, Google Cloud Platform, Microsoft Power Automate