TARUN AJITKUMAR

tarun@vt.edu | (571) 337-1509 | LinkedIn | GitHub | Portfolio

EDUCATION

Virginia Polytechnic Institute and State University **Aug 2024 – May 2026**
Master of Engineering – Computer Science and Applications (CGPA: 3.90/4.00)

Narsee Monjee Institute of Management Studies **Aug 2020 – May 2024**
Bachelor of Technology – Computer Science and Business Systems (CGPA: 3.40/4.00)

SKILLS

Languages: Python, SQL, C++, JavaScript, R
ML/AI: scikit-learn, PyTorch, TensorFlow, Keras, XGBoost, Transformers, OpenCV, Feature Engineering, Ensemble Learning, Model, LLM Integration, Prompt Engineering, LangChain, LangGraph, RAG, MCP
Data & Analytics: NumPy, Pandas, Matplotlib, Seaborn, Tableau, Power BI, Streamlit, EDA, ETL
Backend & Cloud: FastAPI, REST APIs, JWT Auth, Docker, AWS (EC2, Lambda), Azure, Kubernetes
Databases: PostgreSQL, MySQL, Oracle, DynamoDB
Tools: MS Office, Git, GitHub, Jupyter, VS Code, PyCharm, MATLAB, R Studio

TECHNICAL EXPERIENCE

Virginia Tech – Research Assistant (NLP) **Jan 2026 – Present**

- **Evaluated systematic bias and reasoning patterns** in LLMs (GPT-5 mini, Mistral 3 large, Gemini 3 flash, Claude Sonnet 3.5) using a uniform framework to analyze model-based decision logic.
- **Benchmarked model performance** using **Zero-shot**, **Few-shot**, and **Chain-of-Thought (CoT)** prompting to determine the impact of context and reasoning steps on annotation accuracy.
- **Quantified model-human alignment** by calculating **Inter-Rater Agreement (IRA)** scores between model-generated explanations and human-labeled gold-standard datasets.

Solar Secure Solutions – Data Science Intern **Dec 2022 – Feb 2023**

- Built and evaluated **machine learning models** to solve business forecasting problems, improving decision accuracy by **25%.**
- Performed **exploratory data analysis** and **statistical modeling** on structured datasets; created **20+ visualizations** for stakeholder insights.
- Developed end-to-end predictive analytics pipelines using **Python**, leveraging **scikit-learn**, **numpy**, and **pandas** to enable data-driven business decision-making.

ACADEMIC PROJECTS

FoodieSafety – AI-Powered Food Recall & Safety Platform (Capstone Project, Boeing-Mentored) **Aug 2025 – Dec 2025**

- Designed and developed a full-stack web application to deliver real-time food safety and recall alerts for U.S. grocery shoppers.
- Built a **FastAPI**-based backend with **JWT** authentication, secure password hashing, and **REST APIs** for user accounts and products.
- Built an automated recall ingestion pipeline using **AWS Lambda** to fetch and parse **OpenFDA data**, storing results in **DynamoDB**.
- Integrated **LLM**-based features to generate recipe recommendations for soon-to-expire products, helping reduce food waste.
- Deployed **containerized services** using **Docker on AWS EC2**, managing the end-to-end lifecycle of a cloud-native platform.

Human v/s AI Content Classification **Aug 2025 – Dec 2025**

- Distinguished human-written text from AI-generated content by developing a **multi-architecture classification framework** trained on a balanced dataset of **80,000 samples**.
- Engineered and benchmarked a diverse model suite including Statistical baselines **(TF-IDF + MLP)**, Transformer models **(DistilBERT, RoBERTa)**, and deep sequential architectures **(GRU, TCN)**.
- Optimized detection performance by evaluating **Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics.**
- Designed a robust **hybrid ensemble** by **stacking XGBoost over GRU and TCN architectures** to enhance model generalizability against paraphrased and adversarial content.

Healthcare Analytics Dashboard **Sep 2024 – Dec 2024**

- Developed a real-time healthcare analytics platform hosted on **Railway.app** for stakeholders, including administrators, doctors, and medical store owners.
- Integrated and optimized **PostgreSQL** to store and analyze **500+ records**, including patient vitals, prescriptions, and ambulances.
- Implemented **LLMs** for generating query-based insights and visualizations (e.g., charts, heatmaps) to identify trends in disease patterns, medical inventory, and vitals.
- Designed and developed interactive dashboards using **Streamlit**, enabling role-based access and improving data-driven decision- making for **10+** user categories.