HARSHIT SANDILYA

harshitsandilya@gmail.com | linkedin.com/in/harshit-sandilya | github.com/harshit-sandilya

EXPERIENCE

Machine Learning Engineer June 2025 – Present
Shodh AI

- Designed and implemented CI/CD pipelines for LLM API infrastructure, automating deployments and reducing fetch time to under 30 seconds using AWS Lambda.
- Developed a visualization pipeline to improve question-answer interactions, enhancing the effectiveness of AI-generated responses delivered in under 12 seconds.

Machine Learning Intern *Jan 2024 – June 2025*

- Led a team of 5 engineers to develop and deploy an AI-powered Teaching Assistant for a distance learning MBA course, aimed to serve over 700 students.
- Architected 'Neev,' a modular and scalable LLM training framework, optimizing data preprocessing speed, enabling distributed training, and enhancing hyperparameter tuning, reducing perplexity by 75%.
- Engineered 'Ved,' a high-performance pipeline for automated extraction, cleaning, and transforming unstructured web and PDF data. Validated on 50+ Texas Instruments datasheets and 25+ webpages for data extraction.
- Optimized model architectures and fine-tuning strategies for the 'Livia' LLM suite, contributing to models ranging from 31M to 1B parameters and improving dataset accuracy by up to 20% over vanilla transformer.

EDUCATION

Malaviya National Institute of Technology Jaipur, Rajasthan
Bachelor of Technology in Computer Science *Aug. 2021 – May 2025*

PROJECTS

Railguard | *Unity, Python, Sockets* Jan 2025 – Present

- Pioneered a GPS-based train localization and simulation framework tailored for learning-based controllers, ensuring latency below 50ms for real-time locomotive, control center, and trackside communication.
- Constructed a physics-based simulation using Unity's PhysX engine, modelling frictional and resistive forces, enhancing train movement precision and ensuring smooth speed transitions via acceleration clamping.
- Developed and deployed a PPO-based decision-making controller for real-time train collision avoidance, leveraging reward optimisation and state-action policy modelling to improve operational safety and efficiency.
- Repository available on GitHub: github.com/harshit-sandilya/Railguard.

RNA Secondary Structure Prediction | *Pytorch Lightning, Sklearn, Stable-Baselines* August 2024 – December 2024

- Formulated a computational model for RNA secondary structure prediction, leveraging dot-bracket notation for classification into four classes.
- Evaluated model performance across multiple algorithms, achieving 95.82% accuracy with Random Forest, 93% with Decision Trees, 83% with deep learning embedding layers, and 70% with Deep Q-Networks.
- Published source code on GitHub for reproducibility: github.com/harshit-sandilya/RNA-Secondary-Structure.

PUBLICATIONS

Generating Topic Agnostic Conversation with LLM | *IEEE Explore [First Author]* 2024

- Introduced a framework that leverages three Large Language Model instances to generate high-quality, low-toxicity, generalised synthetic conversational data achieving 0.27% for Toxicity, 0.04% for Obscene, 0.02% for Threat, 0.05% for Insult.
- Access the publication here: ieeexplore.ieee.org/abstract/document/10704625.

PATENTS

Offline Road Safety Educational System | *202311049973* 2023

- Developed a bilingual (Hindi and English) conversational chatbot using BERT embeddings and cosine similarity for answer retrieval, providing real-time educational and road safety assistance in collaboration with Sahayta Jaipur (NGO).
- Deployed on Raspberry Pi 4B with an integrated CPR training mannequin for offline functionality, handling 100+ predefined queries. Source code available at: github.com/harshit-sandilya/Road-Safety.

TECHNICAL SKILLS

Languages: Python, C/C++, JavaScript, SQL
Frameworks: React, Node, Flutter ,FastAPI
Developer Tools: Git, Docker, AWS, Figma, Unity
Libraries: pandas, NumPy, Matplotlib, PyTorch, TensorFlow, PyTorch Lightning, Scikit-learn, Stable-Baselines