

Aditya Raghuvanshi

Undergraduate Researcher | B.Tech Computer Science + MS Computational Linguistics | IIIT Hyderabad

✉ tanalpha.aditya@gmail.com

☎ +91 89896 94908

🌐 Portfolio Website

📄 aditya-raghuvanshi

🔗 tanlapha-aditya

WORK EXPERIENCE

IBM Research

May 2025 – Aug 2025

Research Intern - AI | Bangalore, India

- Designed a **static-analysis-augmented LLM framework** for **global-to-local variable localization** in legacy codebases; introduced the **Super30 dataset** (30 C programs with globalized/original pairs) and achieved **F1-scores up to 83.3** by integrating control-flow and read/write dependencies.
- Co-led research on AST-guided chunking for LLM-based COBOL-to-Java translation; designed and evaluated scalable algorithms that improved translation quality by **+70 points** and **2× variable retention** on real-world legacy codebases using **Granite-20B**, **LLaMA-3 8B**, and **Mistral 7B**.

AIMA

Oct 2024 – Feb 2025

Machine Learning Intern (Remote) | Dublin

- Built a **RAG** system for industrial manuals, optimizing metadata structuring and chunking, saving engineers **20+ hours/week**.
- Engineered an OCR pipeline with **PyMuPDF**, **Tesseract**, and Gemini APIs, improving table extraction accuracy by **30%**.
- Launched a **Firestore app** (Auth, Firestore, Functions, Hosting) with secure login, SOP pages, PDF integration, and real-time sync.

TCS Research

Jun 2024 – Aug 2024

Research Intern - AI | Pune, India

- Implemented a multi-stage pipeline for personalized music generation by combining user profiling, preference modeling, and AI-driven synthesis; employed Spleeter, Librosa, and integrated Facebook MusicGen APIs for genre-specific and diverse composition.
- System adapts to user preferences and environmental context to enhance personalization in music experience.

EDUCATION

International Institute of Information Technology, Hyderabad

Nov 2021 – May 2026

B. Tech (Hons.) in Computer Science and MS (By Research) in Computational Linguistics

8.01/10

PROJECTS

SPARK: Step-by-step Proof Assistant for Reasoning and Knowledge

- Developed a step-by-step proof assistant for mathematical reasoning, integrating small LLMs with formal verification.
- Engineered **DPO/GRPO** fine-tuning pipelines (**4-bit LoRA**) for **Qwen2.5-MATH** and **Phi-mini-4k**, achieving **+14.7 pp** on MATH500, surpassing Gemini-2.0 on GSM8K, with custom data prep and error-driven reward design.

Amazon ML Challenge 2024

All India Rank 14 / 75,000+ participants

- Built an **OCR pipeline** with **DocTr** and **PaddlePaddle** for text extraction, paired with a prompt-tuned **Llama 3.1 (7B)** to retrieve entity values.
- Leveraged **Qwen-VL (2B/7B)** for direct image-based entity extraction, with rule-based post-processing for higher accuracy.

Hybrid Model Classification

Speech and Linguistics - AI

- Collaborated with **Prof. Chiranjeevi Yarra** to design a **hybrid topic classifier** combining acoustic and textual features.
- Achieved **+40% accuracy** over text-only and audio-only baselines, demonstrating applicability in **low-resource settings**.

PUBLICATIONS & PATENTS

Efficient Translation of Long Code Blocks using Large Language Models

IBM Research

Under Review, 2025

- Co-developed an **AST-guided chunking algorithm** for COBOL-to-Java translation, achieving up to **2× variable retention** and **+70 point quality gains** with models like **Granite-20B**, **LLaMA 3-8B**, and **Mistral 7B**.

More than words: Music, Not Lyrics or Vocals at the heart of Emotional Expression

CogSci Lab, IIITH

ICMPC 2025

- Analyzed emotional perception across 2,700 songs via **source separation** and **VA modeling**, quantifying contributions of music, vocals, and lyrics.

Method & System for creating an Integrated Platform for Set-List and Loop Generation

TCS Research

Indian Patent Application No. 202421088529.

SKILLS

Languages: Python, C, C++, SQL, JavaScript, HTML/CSS

ML & Data Science: PyTorch, TensorFlow, Scikit-Learn, Pandas, NumPy, OpenCV, LangChain, LangGraph

Web & Databases: React JS, FastAPI, MongoDB, MySQL

Cloud & Tools: Firebase (Auth, Firestore, Functions), IBM CCC Clusters, , Docker, Git, PySpark, Render, Tree-sitter

ACHIEVEMENTS

- Recipient of **Deans Academic Award**, **Merit List Academic Award** and **Research Award** for the spring semesters of 2024 and 2025 for academic performance and research capabilities.
- Technical Content Creator:** Authored 20+ technical articles on AI/ML, amassing over **50,000 views** on Medium
- Student Mentor** and **Teaching Assistant** at IIIT-H for DSA, CSO, BRED and ESGME. Mentored 180+ students in these courses
- French A2 Diploma** awarded by the French Ministry of Education via Alliance Française.