RAHUL VISHWAKARMA

rahulvishwakarma3613@gmail.com ◇ +91 86907 26611
GitHub ◇ Google Scholar ◇ LinkedIn ◇ Portfolio

OBJECTIVE

Researcher and developer with an Integrated MSc in *Mathematics & Computer Science* from NISER Bhubaneswar. Currently working at IIT Hyderabad on an *LLM-based agent* providing authentic, source-backed access to Indian heritage knowledge. My primary research interests lie in using *AI to do mathematics*. Experienced in building and training (pretraining+fine-tuning) LLMs, developing LLM-based agents, and backend-focused web development.

EXPERIENCE

System Engineer April 2025 – Present
BHERI Tech Foundation (IIT Hyderabad) *Hyderabad, India*

- Initiated and developed a domain-specialized Large Language Model trained on our structured Verse DB, including dataset curation, tokenizer design, noise-robust pretraining, and custom inference pipeline.
- Conceptualized, designed, and deployed Prashna, an LLM-based agent for Indian heritage Q&A, integrating Dhārā APIs with specialized retrieval modules and Agno-compatible API tools to enable structured tool-call workflows.
- Integrated Agno with Ollama and the GPT-OSS model to orchestrate multi-tool reasoning, context-aware query handling, and reference-backed, verifiable outputs in the agentic RAG pipeline.

Junior Research Fellow (Project Associate) Sept 2024 – Mar 2025
Indian Institute of Technology Hyderabad *Hyderabad, India*

- Developed Dhara platform having three tools: Dictionary, Verse Finder, and Chunk Server as part of the digital Takshila project to enable easy access to ancient Indian knowledge with authentic source references.
- Designed optimized REST APIs to ensure access to our services for both frontend and external integrations.
- Engineered scalable backend services using Django, PostgreSQL, Swagger, and Beautiful Soup to scrape and serve structured heritage data.

PUBLICATIONS

- **IndoorGNN: A Graph Neural Network based approach for Indoor Localization using WiFi RSSI**
  Published at the 11th International Conference on Big Data and Artificial Intelligence (BDA 2023).
  Authors: Rahul Vishwakarma, Rucha Bhalchandra Joshi, and Subhankar Mishra (Springer Link)
- **Enhancing Neural Theorem Proving through Data Augmentation and Dynamic Sampling Method**
  Available as a preprint on arXiv.
  Authors: Rahul Vishwakarma and Subhankar Mishra

PROJECTS

Neural Theorem Proving (Masters Thesis)

- Fine tuned LLMs (ByT5) to generate formal mathematical proofs in Lean prover.
- Introduced a dynamic sampling method and used a data augmentation method to achieve SOTA performance on the MiniF2F dataset. (More details in our **paper**: LeanProver)
- Created a Lean 4 custom tokenizer for developing an efficient LLM for theorem proving in Lean.
- Developed an interactive web interface and public API using Flask, enabling users to input goals and receive generated Lean proofs.
- Earlier experience from an internship at **IISc Bangalore** with *Prof. Siddhartha Gadgil* focused on using ML models to predict premises to be used in the proof of a given theorem in Lean.
- **Tools used**: Python, PyTorch, ByT5, LeanDojo, Lean, Flask.

GNN-Based Indoor Localization with WiFi RSSI (Paper link: IndoorGNN)

- Designed and implemented *IndoorGNN*, a **Graph Neural Network**–based model to improve indoor localization accuracy using WiFi RSSI fingerprints.
- Captured spatial signal patterns from WiFi data and modeled them as graph structures to enhance positional prediction in indoor environments.
- Outperformed traditional models like kNN, SVM, and MLP on benchmark datasets.
- Published the results in BDA 2023 (11th International Conference on Big Data and Artificial Intelligence).
- **Tools used**: Python, PyTorch Geometric, Scikit-learn, Docker.

Recommendation System

- Developed a file access pattern–based **recommendation system** for NISER Archive to suggest study materials.
- **Tools used**: Django, HTML.

NISER Bus Tracker

- Developed a **web app** for sharing the live location of the NISER buses with its members.
- **Tools used**: Django, HTML, CSS, JavaScript.

EDUCATION

Int. MSc in Mathematics and Computer Science, NISER Bhubaneswar 2019 – 2024
CGPA: 7.63 / 10

Class 12, Sainik School Nagrota, Jammu 2018
Percentage: 82.8%

SKILLS

AI & Research	PyTorch, LangGraph, Agno, Neural Theorem Proving, Transformers (HF), LLM Pretraning and Fine-tuning, Ollama, Lean Prover, LeanDojo
Web Development	Django, Flask, PostgreSQL, Swagger, HTML, CSS, JavaScript
Programming & Tools	NumPy, Pandas, Docker, Git, Selenium, Beautiful Soup

CERTIFICATES & ACHIEVEMENTS

- Completed the **Deep Learning** Specialization by Andrew Ng on Coursera — Certificate.
- Completed the **IBM Data Science** Specialization on Coursera — Certificate.
- **Developed** and **maintaining** an *open-source* port of the ProofNet dataset from Lean 3 to Lean 4, actively used and cited by the NLP research community with regular contributions via pull requests.
- **Junior Research Fellow** at IIT Hyderabad, contributing to AI-driven knowledge systems for Indian heritage.
- Awarded the **DISHA Scholarship** by the Department of Atomic Energy (DAE) – a 5-year fellowship for undergraduate science students.
- Qualified **GATE 2024** in **Data Science and AI** (DA) with a score of 418 and 42/100 marks.
- Elected President of Coding Club NISER (Aug 2022 – May 2023); led coding sessions, promoted open-source, and fostered a campus programming community.