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 Kanpur on developing LLM-based solutions for 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

Project Executive Officer Nov 2025 – Present
IIT Kanpur *Kanpur, India*

- Designed a domain-specific tokenizer for Indian Knowledge Systems text to optimize sequence efficiency, and trained a 1B-parameter language model from scratch on structured word-definition mappings, verses and translations, and chunked corpora from the Dhārā knowledge base, with ongoing fine-tuning for question-answering.
- Conceptualized, designed, and deployed **Prashna**, an LLM-based agent for Indian heritage Q&A, integrating Dhārā APIs with Agno-compatible tools, and orchestrating multi-tool reasoning via Ollama and GPT-OSS to enable context-aware, reference-backed, and verifiable agentic RAG workflows.
- **Tools used**: PyTorch, Transformers, Tokenizers, Agno, LangGraph, Weaviate, Ollama, vLLM.

Junior Research Fellow | System Engineer Sept 2024 – Oct 2025
IIT Hyderabad | BHERI (IITH-incubated Startup) *Hyderabad, India*

- Developed the backend for the Dhārā platform, comprising the following tools: **Dictionary** (lexical meanings with references), **Verse Finder** (textual verse search), **Chunk Server** (embedding-based semantic search), and **Sodh** (common query platform), enabling easy access to Indian heritage knowledge from authentic sources.
- Designed and implemented optimized REST APIs to support frontend consumption and external integrations.
- **Tools used**: Python, Django, Django REST Framework, PostgreSQL, Weaviate, Swagger, Google OAuth.

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*)

- Developed a custom Lean 4 tokenizer for efficient language modeling in formal theorem proving.
- Fine-tuned LLMs (ByT5) to generate formal mathematical proofs in the Lean theorem prover.
- Introduced a dynamic sampling strategy combined with data augmentation, demonstrating strong performance on the MiniF2F benchmark (paper: LeanProver).
- Built a Flask-based web interface and public API for goal-conditioned Lean proof generation.
- Developed and maintaining an open-source Lean 3 to Lean 4 port of the ProofNet dataset, widely used and cited by the NLP community.
- **Tools used**: Python, PyTorch, ByT5, LeanDojo, Lean, Flask.

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

- Designed **IndoorGNN**, a Graph Neural Network model that represents WiFi RSSI fingerprints as graphs to capture spatial signal relationships.
- Demonstrated improved indoor localization accuracy compared to classical baselines such as kNN, SVM, and MLP on benchmark datasets.
- Published at 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.

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

EDUCATION

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

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

CERTIFICATES & ACHIEVEMENTS

- Completed the **Deep Learning** Specialization by Andrew Ng on Coursera — Certificate.
- Completed the **IBM Data Science** Specialization on Coursera — Certificate.
- 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.