

Praveen Kumar

+91-8302789850 | praveenthakur180109@gmail.com | [LinkedIn](#) | [GitHub](#)

EDUCATION

Central University of Haryana

Bachelor of Technology in Computer Science & Engineering

Mahendragarh, HR

Aug. 2023 – July 2027

EXPERIENCE

Data Analyst Intern

June 2025 – July 2025

TrendalyTix

Remote

- Architected predictive models using **Scikit-learn** and **Pandas** to analyze real-world datasets, improving data-driven decision accuracy by optimizing automated data cleaning pipelines.
- Synthesized complex technical findings into actionable business insights, presenting stakeholders with interactive visualizations to drive strategic decision-making.

Artificial Intelligence Trainer (Freelance)

Dec. 2024 – Feb. 2025

Outlier

Remote

- Optimized Large Language Model (LLM) performance through **Advanced Prompt Engineering** and Chain-of-Thought reasoning across STEM and coding domains.
- Executed **RLHF (Reinforcement Learning from Human Feedback)** workflows to evaluate and refine model outputs, ensuring technical accuracy and ethical alignment.

ACHIEVEMENTS

National Finalist - 5G Lab Innovation Hackathon (2025): Developed **AR Smart Glasses** for the speech and hearing impaired using **5G URLLC** for real-time communication; secured government funding for product prototyping.

Open Source Alpha: Recognized for high-quality code contributions during the GSSoC term.

PROJECTS

Agentic Data Analyst | [Live Demo](#) | [Code](#) | *LangGraph, Llama-3.3, Docker, Streamlit*

Feb. 2026

- Developed an autonomous **Cognitive Agent** using **LangGraph** that plans analysis steps, executes Python code, and autonomously resolves runtime errors via recursive reflection loops (Self-Healing).
- Deployed a secure, containerized application using **Docker** on **Render** with a **Streamlit** interface, implementing sandbox guardrails to prevent malicious code execution and ensure schema consistency.

BioGraph-RAG | Clinical Trial Reasoning Engine | *Llama-3, Neo4j, LlamaIndex*

Jan. 2026

- Engineered a **GraphRAG architecture** combining **Neo4j Knowledge Graphs** with Vector Search to resolve complex multi-hop medical queries in unstructured clinical PDFs.
- Orchestrated **Llama-3 (70B)** via Groq for high-precision Named Entity Recognition (NER), mapping complex Drug-Protein-Disease relationships into a structured graph schema.

Image Captioning | *Python, TensorFlow/Keras, NLP, CNN, LSTM*

Aug. 2025

- Designed a multimodal deep learning pipeline utilizing **CNNs (Encoder)** for spatial feature extraction and **LSTMs (Decoder)** for sequential natural language generation.
- Enhanced model accuracy on the Flickr8k dataset using **Transfer Learning** with pre-trained vision models, validating performance via BLEU metrics.

TECHNICAL SKILLS

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

AI/ML: RAG, GraphRAG, GenAI, LLMs, NLP, Agentic Workflows

Frameworks/Tools: Streamlit, Flask, Docker, LlamaIndex, LangChain, LangGraph, Neo4j, FastAPI

Libraries: TensorFlow, PyTorch, Pandas, Numpy, Scikit-learn

Developer Tools: Git, GitHub, VS Code, Google Colab, Linux