

Harsh Kumar Jha

+91-9315956137 | contactharshjh@gmail.com | LinkedIn | GitHub | Portfolio

Skills

Languages	Python, SQL, Java
AI/ML	Machine Learning, NLP,RAG, Prompt Engineering,Scikit-learn, XGBoost, Hugging Face Transformers, Sentence Transformers
Deep Learning	Neural Networks, NLI, DeBERTa, Embeddings
Data Science	NumPy, Pandas, Matplotlib, Exploratory Data Analysis, Data Cleaning
AI Tools	OpenAI API, Ollama, Jupyter Notebook, Google Colab
Development	FastAPI, REST APIs, Git, GitHub

Experience

DomainLabs

AI/ML Intern

Remote

June 2026 – Aug 2026

- Conducted research in **AI safety**, adversarial testing, and **LLM evaluation** to assess model behavior, robustness, and reliability across diverse AI workflows.
- Designed and implemented **machine learning experiments, evaluation pipelines, and benchmarks** to measure LLM performance, robustness, and safety.
- Developed **adversarial threat taxonomies and red teaming workflows** for large language models, covering prompt-based attacks, failure modes, and model vulnerabilities.
- Contributed to **Bayora**, DomAIyn Labs' automated adversarial red teaming platform, through **LLM integration, model testing, and AI pipeline development**.
- Explored **model alignment, robustness, interpretability, and agentic AI workflows**, translating research findings into technical documentation, experiments, and actionable improvements.

MathonGo

Research & Development Intern

Remote

Jun 2025 – Sep 2025

- Validated **2,000+** Physics questions and structured/tagged educational datasets for JEE/NEET question banks, using AI-assisted quality checks, categorization, and data validation to improve dataset consistency, accuracy, and downstream usability.

Projects

Hallucination Hunter | *Python, FastAPI, DeBERTa-v3, Sentence-Transformers, spaCy, Docker* GitHub

- Engineered a claim-level **LLM evaluation and hallucination-detection pipeline** using DeBERTa-v3 NLI and Sentence-Transformers to verify generated outputs against source context.
- Built **production-ready REST APIs** with explainable confidence scoring, alignment heatmaps, health checks, and Dockerized deployment for direct integration with **RAG pipelines**.

AppForge AI | *Python, FastAPI, Next.js, Pydantic, SQLAlchemy, Cerebras, Groq, OpenAI* GitHub

- Built an **LLM-powered multi-stage agent pipeline** that converts natural-language prompts into applications through intent extraction, system design, schema generation, validation/repair, and execution checks.
- Integrated **OpenAI, Groq, and Cerebras** through a primary-fallback routing layer with automated repair loops, token-budget controls, and end-to-end **AI workflow orchestration** deployed on Railway.

FrameOS | *Next.js, Node.js, PostgreSQL, Supabase, Redis, Docker* GitHub

- Engineered an **AI-native workflow orchestration platform** automating research, script generation, rendering, publishing, and analytics through distributed, event-driven pipelines.
- Implemented **Redis-based job queueing, JWT authentication, and Dockerized deployment** to support reliable multi-step execution and scalable autonomous content workflows.

Education

Maharaja Surajmal Institute of Technology, New Delhi

B.Tech, Electronics and Communication Engineering

2025 – 2029

CGPA: 8.36/ 10