

RAJ PRATAP SINGH GURJAR

☎ +91-9340592875 ✉ rajpragur@gmail.com 🔗 linkedin.com/in/rajpragur 🐙 github.com/rajpragur 🌐 rajpragur.in

Education

Indian Institute of Technology, Patna

2024 – 2028 (Expected)

B.Tech in Mathematics and Computing — CGPA: 8.3/10.0

Patna, India

Relevant Coursework

- Machine Learning
- Data Science
- Probability & Statistics
- Linear Algebra
- Discrete Mathematics
- Optimization

Experience

Middlesex University London

May 2026 – Jul 2026

Research Intern

London, United Kingdom

- Researched on **Neurosymbolic AI** and its approaches to combine neural language models with formal logical verification systems (**Z3 SAT solvers**) to improve consistency and interpretability of LLM outputs
- Worked on a hybrid **CoT + symbolic verification** framework for legal appeals judgement predictions, and achieving **0.85 F1** on UK Court of Appeal data — outperforming both heuristic LLM reasoning and pure symbolic pipelines
- Investigated **Quantum Machine Learning** techniques including QNN rerouting for edge inference, quantum error-correcting output codes (ECOC), and **quantum federated learning (QFL)** for distributed model deployment

Uber (AI)

Nov 2025 – Mar 2026

Independent Contractor - Project Sandbox

Remote, India

- Executed large scale LLM evaluation and post-training workflows in Dockerized production environments, analyzing execution traces to debug model behavior on SWE-bench style datasets for an external frontier model

Nurtureme.ai

May 2025 – Oct 2025

Software Engineering Intern

Remote – Palo Alto, CA

- Built an **LLM powered healthcare platform** for patient intake and clinical triage, integrating AI workflows that streamlined onboarding across **5+ clinics**.
- Developed a clinical scoring engine that transformed unstructured patient responses into structured severity metrics, reducing manual screening effort by approximately **40%**.
- Engineered AI-powered recruitment workflows, including a real-time interview assistant for adaptive questioning, candidate evaluation, and automated feedback generation.

Projects

Tewtr | Python, FastAPI, Multi-Agent LLMs, RAG

GitHub

- Built a **multi-agent document understanding system** combining **LLaMA 4 Maverick (VLM)** and **GPT-OSS-120B (LLM)** to transform PDFs into structured educational content with contextual reasoning.
- Designed a **sliding-window inference pipeline** with adaptive context management, reducing document processing latency from **30 minutes to 12 seconds** while preserving long-context reasoning quality.
- Optimized high-throughput inference using **speculative decoding** and **flash attention mechanism** to reach upto 300+ tokens/sec, achieving up to **5× faster** document processing through efficient model routing and cost-aware execution.

disql – LSM Tree Storage Engine | Go, Storage Systems

GitHub

- Implemented an **LSM-tree storage engine** featuring Skip List memtables, immutable SSTables, sparse indexes, and achieving **430 ns writes**, **325 ns reads**, and **6.7 μs** range scans.
- Designed a self-describing on-disk storage format with data blocks, **Bloom filters**, **CRC32 integrity checks**, and magic-number validation, enabling efficient point and range lookups.
- Built durable storage with **Write-Ahead Logging (WAL)**, **fsync**-backed persistence, crash recovery through WAL replay, tombstone-based deletions, and background compaction.

Technical Skills

Languages: Python, C++, Go, JavaScript, TypeScript

ML/DL: PyTorch, Hugging Face, Transformers, vLLM, Scikitlearn, NumPy, Pandas

Specialized: LLMs, Neurosymbolic AI, Knowledge Distillation, Quantization, Model Evaluation

Developer Tools: Git, Linux, Bash, LaTeX

Achievements

- Selected Member**, Cohere For AI Science Community | **Gold Certificate**, WorldQuant Research Consultant Program
- Solved 600+** algorithmic problems across Codeforces, LeetCode & CodeChef | **Rank 1.8k / 45k+** IICPC Codefest 2025
- AIR 32**, AIME 2024 and Selected among **top 3%** globally | **99.7 percentile**, JEE Main 2024