Aryan Dua

RESEARCH RESUME

AI research & engineering

aryan.iitd1@gmail.com +91 8830639139 tan90cot0.github.io GitHub / tan90cot0 LinkedIn / aryandua

EDUCATION

IIT Delhi / B.Tech + M.Tech, Computer Science 2020 - 2025

B.Tech GPA 9.00/10 | M.Tech GPA 8.80/10

RESEARCH

IIT Delhi & Wipro / Master's thesis / Gen AI & NLP research Jan 2024 - May 2025

Advisor: Prof. Mausam | Biomedical models, retrieval, and diagnostic agents | Submitted Jun 2025

- Fine-tuned **MEDITRON-7B** with 8-bit LoRA on 200K PubMedQA examples; improved accuracy from **43.8% to 73.8%**. Compared medical models across four benchmarks and clinical cases.
- Built PubMedBERT / FAISS / BM25 hybrid RAG with re-ranking; thesis experiments report **+8.9 percentage points** in average benchmark accuracy and hallucination rates falling from 8.3% to 4.8%.
- Designed DeepSeek R1-8B diagnostic agents; revised safety and urgency specifications to reach **0.490 peak F1 vs. 0.475**. Studied missing-context robustness, agent ablations, and human-aligned evaluation.

Zuse Institute Berlin / Research intern Jun - Aug 2022

- Co-developed asynchronous projective splitting in Julia; evaluated sparse classifier training and image recovery. Reduced runtime **923.4 s to 701.4 s (24%)** on three workers under simulated latency.
- Studied worker scaling, delay bounds, and proximal-parameter schedules. Co-authored an Optimization Online preprint; code: AsyncProx.

APPLIED AI

Lymph Labs / Co-Founder & CTO Jul 2026 - Present

- Led a medical evidence-search MVP used by **2,000+ doctors**, including clinicians from AIIMS and JIPMER; built OpenSearch/Qdrant retrieval with passage-level provenance and citation validation.
- Engineered **LangGraph/LangChain** workflows with parallel tools, structured outputs, and bounded retries; built Langfuse tracing, versioned prompts, evaluation harnesses, and retrieval replay.
- Built durable FastAPI/PostgreSQL operations with Cloud Tasks and replayable SSE; delivered web/mobile clinical workflows and Cloud Run infrastructure provisioned with Terraform.

Abacus.ai / ML Platform Engineer Jul 2025 - Jul 2026

- Built a framework for users to create and use custom skills with their own DeepAgent instances; developed an internal AI software engineer, multi-provider tool calling, speech, and S3 integrations.

SELECTED NLP STUDIES

Graph-to-text / WebNLG: built GPT-2 and GCN pipelines; team BLEU 23.00 overall / 12.91 unseen. Report.

Wikipedia controversy: compared prompting (87%) with classical features (93.1%); 29 test examples. Report.

ADDITIONAL EXPERIENCE

IIT Delhi TA: CUDA/MPI/OpenMP, 150+ students. **Quadeye:** quant research. **JackandJill.ai:** LLM extraction. **Ripik.ai:** graph scheduling.

METHODS & TOOLS

Learning: PyTorch, Transformers, fine-tuning, LoRA/PEFT, quantization, distillation.

Retrieval & agents: RAG, FAISS, BM25, Qdrant, LangGraph, LangChain, tool calling.

Evaluation & compute: ablations, robustness, hallucination analysis, Python, C/C++, Julia, CUDA, HPC.

SELECTED DISTINCTIONS

5th, Dream11 Gameathon ML (2025) | JEE Advanced AIR 516 | IIT Delhi CSE branch change: 11 selected from a batch of 1,200.