Saksham Dhull

+91-9718227010 | sakshamdhull442@gmail.com | dhull442 | Dhull442

Work Experience

Samsung Research, South Korea | *Software Engineer, AI & ML Team* Sep, 2021 - Jan, 2024

- Led a 3-person team to design and ship a multimodal Speech LLM, reducing pipeline complexity and latency by 19%.
- Built an under 1s latency video summarization microservice using oneShot-TTS and Llama, with expression cues.
- Developed an ML packet scheduling model to optimize packet allocation across an 8-core vRAN.
- Automated end-to-end deployment with Kubeflow MLOps, enabling iterative model improvements and boosting throughput by 3.7% on a 12-node cluster with 100+ Gbps traffic.
- Directed a 4-person team to build and deploy a lightweight Rust-based dynamic memory profiler, adopted by 900+ engineers as the company's in-house alternative to Valgrind.
- Co-authored technical specification and developed the 6G network tech stack with a team of 12 to reach 30+ Gbps throughput.

Spotnana, Palo Alto, CA | *Software Engineer, Backend Team* Jun, 2021 - Aug, 2021

- Developed an automated REST API generation service (300+ endpoints) to reduce the release cycle by 7%.
- Presented technical progress reports to investors and VCs, effectively communicating platform capabilities and supporting funding discussions.
- Implemented and managed CI/CD automation for backend APIs, maintaining deployment times through regression testing.

Samsung Research, South Korea | *Software Engineer Intern, 5G Research Team* May, 2020 - Jul, 2020

- Investigated and optimized throughput bottlenecks in Linux kernel SCTP for 4+ core 5G system with 1+Gbps throughput.
- Built a lightweight throughput benchmarking tool in Go for technical analysis of 3 SCTP frameworks under 10Gbps load.

Education

Indian Institute of Technology, Delhi – *Bachelor of Technology, Computer Science*
Awards & Achievements:

- Recipient of the IIT Delhi Semester Merit Award, Top 7% of 930 students in bachelors cohort.
- Joint Entrance Examination (JEE) Advanced, 2017: secured **All India Rank 66** among 1.6M students.
- International Olympiads, 2017: Selected among the top 30 students from India for IPhO, IChO and IOAA OCSC.

Skills

Machine Learning: PyTorch, Tensorflow, HuggingFace, LangChain, LlamaIndex, OpenCV

Programming: Python, Rust, C++, Java, Go, SQL

Systems & Infra: Docker, Kubernetes, AWS, GCP, Kubeflow

Specializations: LLM, RAG, GenAI, Computer Vision, Software Design

Projects

Mitigating Prompt Injection in Retrieval-Augmented Generation | *Univesrity of Washington* May, 2025 - present

- Constructed and analyzed a 50,000-query benchmark dataset, exposing ≈20% prompt injection vulnerability in RAG systems.
- Designed a security layer against prompt injection attacks using LangChain, reducing attack success rate to 23.4%, reducing overall prompt attack vulnerability to ≈4.7%.
- Optimized and evaluated retrieval quality with LlamaIndex, ensuring <30ms latency overhead in production-like workloads.

Lattecoin - Rust Blockchain Cryptocurrency Jan, 2025 - Jun, 2025

- Designed and implemented a blockchain with Bitcoin-style PoW consensus with on-line difficulty adjustment to maintain block time of approximately 2 minutes.
- Deployed a local network of 50 miners, peers, and ledgers, achieving near real-time ledger synchronization under a normally distributed probabilistic load with an average throughput of 10 transactions per second.
- Built a Web dashboard to visualize live blockchain stats, block height, transaction throughput, and peer connectivity.

Rust Ray Tracer Dec, 2024

- Built a feature-rich enterprise-grade Ray Tracer in Rust.
- Integrated BVH Acceleration and Parallel processing to optimize performance and slash the rendering times by 10×.
- Published a technical blog detailing system architecture, software design choices and open-source contributions.

Fairness in Computer Vision | *IBM Research* Sep, 2020 - Feb, 2021

- Engineered a Causality-Based VAE ML model to extract dependencies from a schematic model with 100+ features.
- Published a technical report detailing the model structure, design choices and the outcomes on CIFAR.

Incremental BFS (LNCS Journal) | *Prof. Sandeep Sen, Summer Research Fellowship* May, 2019 - Aug, 2019

- Invented a theoretical technique with novel bounds for maintaining the BFS tree of a graph in incremental scenarios.
- Authored a journal paper for Springer LNCS, detailing methodology, mathematical proofs, and experiments.