(408) 581-3097
samyak24jain@gmail.com

Samyak Rajesh Jain

San Francisco Bay Area

github.com/samyak24jain
linkedin.com/in/samyak24jain

WORK EXPERIENCE

PowerSchool - *Senior Software Engineer* — Menlo Park, CA | Jan 2025 - Present

- **Agentic Text-to-SQL Data Platform**
  - Shipped a 2x faster version of PowerBuddy for Data Analysis (PBDA) to 50+ school districts, enabling non-technical educators to "talk to their data" via a Chain-of-Thought LLM pipeline, delivering 90% of responses within 20s.
  - Grounded this pipeline in customer data by developing a high-recall RAG system utilizing ANN Search and n-gram decomposition, accurately mapping ambiguous user terms to precise database entries, eliminating false-negative queries.
  - Led the transition to a multi-agent autonomous workflow using AWS Strands in Python (Claude Opus 4.6/GPT-5.4), defining complex agent orchestration (personas, roles, cross-agent communication, tools) to deliver a resilient, conversational experience capable of natural error recovery and native scaling across custom tables.
  - Optimized the pre-production agentic system, slashing inference costs by 80% ($1.70 to $0.33/query) and latency by 55% via prompt caching, reasoning-effort tuning, and preventing context bloat by routing large tool outputs to a shared-state memory accessible via dedicated exploration tools.
  - Established an automated LLM-as-a-Judge evaluation framework using reasoning LLMs (o1, o3-mini) to continuously benchmark Text-to-SQL correctness and consistency against a Gold Dataset of 150+ expert-verified queries

Samsung Research - *Lead Machine Learning Engineer* — Bengaluru, India | Jun 2018 - Jun 2023

- **Real-Time Low-Light Video Restoration**
  - Developed an on-device multi-frame CNN in TensorFlow for the Galaxy S23, achieving 4x better temporal consistency, 6% higher SSIM, and 200x lower complexity than state-of-the-art baselines.
  - Curated a 50K video training dataset incorporating real noise and synthetic motion, heavily reducing motion blur and enhancing low-light details.
- **Optimizing Image Processing Pipeline**
  - Mentored 3 engineers to achieve a 10% performance boost on the multi-frame image processing pipeline of the Galaxy S23 FE.
  - Optimized CPU core utilization and reduced inter-thread dependencies by analyzing Android system traces and identifying idle and under-utilized cores.

EDUCATION

University of California, Santa Cruz (UCSC) — Santa Clara, CA
- Master of Science in Natural Language Processing — *Sept 2023 - Dec 2024*

National Institute of Technology Karnataka (NITK) — Surathkal, India
- Bachelor of Technology in Information Technology — *Aug 2014 - May 2018*

PROJECTS

- **Graph-based Planning System for LLMs**
  Trained a Graph Neural Network (GNN) based encoder model in PyTorch to retrieve graph-structured plans, enhancing LLMs' complex reasoning and long-term planning capabilities for multi-hop Question Answering (QA).
- **Closed Domain Multimodal QA**
  Fine-tuned Vision Language Models (VLMs) like Phi-3-vision and LLaVA using LoRA for a multimodal QA system, achieving a 12.4% improvement in multi-hop QA performance on technical graphs and tables.

PUBLICATIONS (Scholar)

- Reasoning Graph-Structured Question Answering: Datasets and Insights from LLM Benchmarking *(LREC 2026)*
- Legal Answer Validation using Few-Shot Multi-Choice QA *(SemEval 2024)*

TECHNICAL SKILLS

- Python, C, C++, SQL, PyTorch, TensorFlow, Strands Agents, HuggingFace transformers, LangChain, pandas, numpy, scikit-learn, FastAPI, Cursor, GitHub Copilot, Git, Snowflake, AWS