

Himali Saini

New York, USA — +1 (646) 637-4568 — hs2343@cornell.edu — himali.dev — [github](https://github.com) — [linkedin](https://www.linkedin.com/in/himali-saini)

Education

Cornell Tech, Cornell University, New York, USA

Master of Engineering (MEng) in Data Science and Decision Analytics

May 2027 (Expected)

Merit Scholarship Recipient

Indian Institute of Technology Bombay (IITB), India

Master of Technology (M.Tech) in Computer Science Engineering

May 2024

CGPI: 9.58 / 10

K. J. Somaiya College of Engineering, University of Mumbai, India

Bachelor of Technology (B.Tech) in Computer Engineering

June 2022

CGPI: 9.19 / 10

Technical Skills

Programming and Scripting: Python, Go, SQL, C++, C, Java, JavaScript, TypeScript, Bash, Flutter, PHP, HTML, CSS

ML & CV: PyTorch, TensorFlow, Keras, LLM APIs, MCP, Multi-Agent Systems, OpenCV, MediaPipe, MATLAB

Data & Backend: ETL Pipelines, Airflow, Tableau, Elasticsearch, PostgreSQL, AWS S3, Snowflake, Docker, Django, SQLite

Professional Experience

Apple Inc. – Software Engineer, AI & Data Platforms, India

Jul 2024 – Jul 2026

- Built **ETL pipelines** using **Airflow**, with custom **DAG nodes**, for a global privacy compliance project, processing **billions of records**.
- Designed pipeline architecture end-to-end (**HLD/LLD**) and led cross-team **integration testing** for the compliance project.
- Managed backend architecture for **~30** executive Tableau dashboards, delivering privacy and analytics insights to business leaders for strategic decision-making.
- Built a **self-service AI agent** letting business leaders run **ad-hoc queries** without engineering support.
- Won **2nd Place & Crowd Favourite**, Apple IS&T Hackathon 2025 — built an **AppleScript + JSON-schema** pipeline that renders LLM-generated content into Keynote slides, reverse-engineered from Keynote's file format, with an **embedding-based icon similarity search**; scaled to production across multi-region infra at **99.9%** reliability, with MAU growing close to **2x** month-over-month.
- Won **2nd Place (Innovation)**, MCP Hackathon 2025 — built a stand-up assistant unifying **MCP servers**, using a **sequential-thinking LLM** to auto-summarize commits, meetings, and to-dos.

Continuum Solutions LLP (A ConnectWise Company) – Summer Intern, India

Jun 2021 – Aug 2021

Go, Elasticsearch

- Developed a high-performance **Elasticsearch** query generator with advanced filtering, sorting, and pagination capabilities.
- Reduced query execution time by developing utilities connecting **Go** to **Elasticsearch**, including bulk insertion and search.

Projects and Publications

Enhancing Object Segmentation in Out-of-Distribution Medical Images — *M.Tech Thesis & Research Paper* [Link](#)

- Developed a novel **deep neural network** framework based on **auto-encoders** to improve object segmentation in out-of-distribution medical images, addressing variations across imaging protocols.
- Paper published at **WACV (IEEE/CVF Winter Conference) 2025**, a leading conference on computer vision and pattern recognition, underscoring significant advancements in medical imaging technologies.

Deep Learning for Image Super-Resolution — *M.Tech Project*

- Implemented deep learning models (**U-Net, ResNet, GANs, variational auto-encoders**) for **4x and 8x image super-resolution** on industrial datasets for a major U.S.-based energy company.
- Designed custom perceptual loss metrics for training and evaluation, achieving a **2x+ improvement** in perceptual similarity scores.

Distributed Key-Value Store with Raft Consensus — *Go*

[GitHub](#)

- Built a **distributed key-value store in Go** that keeps **3 servers in sync** using the **Raft consensus algorithm** (leader election, log replication, majority-quorum commit) on a crash-safe **write-ahead log**.
- Kept the cluster available through failure: a new leader is elected automatically in **~300 ms** after the leader is killed mid-write, with **zero acknowledged writes lost** under fault injection (**20% packet loss**, repeated node kills).
- Let clients choose **strong vs eventual consistency** per read request; separately, batched concurrent **strong-read** quorum checks into shared round trips, improving strong-read throughput **7x** (**7.4K → 52K reads/sec**).

Hybrid Retrieval RAG Pipeline — *Python, Qdrant, Voyage AI, BM25*

[GitHub](#)

- Built a **hybrid retrieval RAG pipeline** over **154 documents** using **Qdrant** (dense vector search), **BM25** (keyword search), and **Reciprocal Rank Fusion**, with an LLM generation stage that grounds every claim in a cited source chunk.
- Built a **34-question recall@5 eval harness** across **2 chunking strategies**; found naive hybrid retrieval underperformed dense-only search due to equal-weighted **RRF**, and closed the gap by adding a **cross-encoder reranking** stage.

Neural Network from Scratch — *Foundations of Machine Learning, Course Project*

- Implemented regression and classification models entirely from scratch, using backpropagation, **PCA**, and **Adam optimizer**; ranked **1st of 26 teams** on the Kaggle classification task and **7th of 49 teams** on the regression task.

Awards & Leadership

- Secured **AIR (All India Rank) 10** out of ~70,000 applicants in **GATE (Graduate Aptitude Test) CS 2022**.
- Teaching Assistant for the course **CS 101: Computer Programming**, mentored a group of **12+ students**.