

SIDDHARTH RAIPAL

Los Angeles, CA | (323) 926-8503 | siddharthraipal2001@gmail.com | LinkedIn | GitHub | Portfolio | F-1 STEM OPT

WORK EXPERIENCE

Graduate Researcher | InfoLab, USC — Los Angeles, CA

December 2025 – Present

Dementia Precision Risk Tool — UK Biobank

- Migrated UK Biobank feature-engineering pipelines to Polars, reducing memory overhead by 40%
- Benchmarked 7 survival models and built an ensemble powering a clinician-facing risk prediction and intervention dashboard
- Built an Explainable AI pipeline surfacing top risk predictors and personalized risk-reduction plans for clinicians using scikit-learn
- Selected to present 2 research posters at USC CPBH Symposium on dementia risk prediction and clinical dementia classification

Machine Learning Engineer | Rise11 Technologies, India

July 2023 – July 2024

- Scaled a consumer web product to 10,000+ DAU by architecting distributed backend services on AWS with Docker, CI/CD pipelines, and 90%+ test coverage
- Led development of a doctor-facing LLM chatbot over 100K clinical notes by constructing a 5M-node medical knowledge graph in Neo4j with PySpark on Databricks
- Tripled email click-through rate by A/B testing LLM-personalized emails on an internal analytics dashboard

Data Science Intern | Rise11 Technologies, India

January 2023 – June 2023

- Replaced manual reporting with automated metric tracking workflows using Mixpanel and Python
- Improved bounce rate by 37% by identifying user behavioral bottlenecks through funnel analysis and event tracking on Mixpanel

PROJECTS & PUBLICATIONS

GlaucomaGuide: A RAG-Enhanced Medical Chatbot

August 2024 – Present

Under final review at Knowledge-Based Systems (Elsevier, IF ~8.0)

- Built a multi-turn RAG chatbot for glaucoma patient education using LlamaIndex, LangChain, and Streamlit, validated by 2 specialists and 100 lay users
- Outperformed ChatGPT-4 (+14%) and Gemini (+7%) with 75% task success on 20 complex scenarios ($p < 0.001$)

Graph RAG Nutrition Intelligence

May 2025 – August 2025

- Built a 6-stage Graph RAG pipeline converting unstructured nutrition research into a queryable knowledge graph using Pinecone, NetworkX, and LlamaIndex
- Improved retrieval relevance and reduced knowledge-graph fragmentation by 60% using graph-enriched entity embeddings and entity normalization

ICGRNN – Input Convex Graph Recurrent Neural Networks

January 2025 – May 2025

- Implemented a GNN-RNN architecture in PyTorch Geometric with end-to-end convexity guarantees, achieving 78% on Cora and 73% on PubMed (3% trade-off vs GCN)
- Generated physics-informed synthetic training data on graphs and optimized dual-objective loss for heat diffusion control, achieving 0.000101 MSE

EDUCATION

University of Southern California — Los Angeles, CA

August 2024 – May 2026

M.S. in Computer Science (GPA: 4.0/4.0) | Phi Kappa Phi Scholar | USC IEEE Corporate Ambassador

Thapar Institute of Engineering & Technology — Patiala, India

August 2019 – June 2023

B.Tech in Computer Science (CGPA: 9.05/10)

SKILLS

- **Languages:** Python, SQL
- **ML / AI:** PyTorch, PyTorch Geometric, scikit-learn, GNNs, LLMs, RAG, Knowledge Graphs, LlamaIndex, LangChain, NetworkX
- **Data:** PySpark, Databricks, Polars, Neo4j, Pinecone
- **Cloud & Infra:** AWS, Docker, Streamlit