

SANCHIT GOEL

San Fransisco, California 94105 ◊ +1(858) 642-5961 ◊ s5goel@ucsd.edu ◊ linkedin.com/in/-sanchitgoel

EDUCATION

Master of Science in Data Science, University of California San Diego

Sep 2023 - March 2025

Bachelor of Science in Computer Science, Ashoka University

Aug 2016 - Jul 2020

TECHNICAL SKILLS

Programming Languages: Python, SQL, Scala, Java, SQL, R

Frameworks: Spark, Docker, Kubernetes, Ray, CUDA, vLLM, Tensorboard, Apache Superset, MySQL

ML Libraries: Langchain, PyTorch, Tensorflow, Transformers, LlamaIndex, Scikit-Learn, Keras, Causal-Learn

Relevant Coursework: ML Systems, NLP, Algorithms, Deep Learning, Machine Learning, Causal Inference

WORK EXPERIENCE

Frida Jun 2024 - Sep 2024

Machine Learning Engineer

Miami, FL

- Engineered **multimodal RAG knowledge retrieval system** leveraging GPT LLM, Langchain and LLamaIndex to enhance internal knowledge management and **enable agent-driven customer support**.
- Optimized text chunking using LangChain Recursive Splitter and Unstructured library, enabling efficient table and image extraction and achieving a 0.83 context recall as evaluated by the **RAGAS** framework.
- **Improved retrieval accuracy by 14%** by implementing **cross-encoder reranking** and **ReAct AI agents** to dynamically route queries to relevant partitions within the **Pinecone** vector database.

Merck & Co. Dec 2021 - Aug 2023

Machine Learning Engineer

India

- **Improved marketing engagement by 12%** by developing an end-to-end ML pipeline on AWS SageMaker for automated promotional material tagging, optimizing business campaign outreach strategies.
- Led a cross-functional team of 3 to transition from AWS Comprehend to **finetuned BERT** models, achieving a **30% cost reduction and a 14% improvement in F1 score**, validated through **McNemar's A/B test**.
- **Fine-tuned GPT-3** for data-scarce markets, leveraging few-shot learning techniques to improve adaptability.
- Developed CI/CD pipelines with Jenkins, Git, Docker, and AWS ECR for automated testing and deployment.

Fractal.ai Oct 2020 - Dec 2021

Machine Learning Engineer

India

- Saved **650+ manual hours yearly and reduced turnaround time by 75%** by developing an invoice entity extraction solution for Fractal's admin using XGBoost, Selenium, and UiPath.
- Engineered Propay.bot, an NLP-based fraud detection system, integrating POS tagging and NERs to flag problematic health insurance claims; adopted by UnitedHealth for further development.

RESEARCH EXPERIENCE

Research Assistant Lee Lab, Prof. Ellen Lee Sept 2023 - Current

- Secured **\$40K** in research funding and a monthly stipend to advance AI-driven healthcare research.
- Developed **multi-modal models on speech data**, integrating acoustic and linguistic features to detect social isolation and early signs of schizophrenia in older adults.
- Conducted a longitudinal analysis of sensory data, employing **mixed-effects models** to compare biomarker trajectories between schizophrenia and control groups. **Publications in progress.**

PROJECTS

- **LLM-Augmented Information Retrieval:** Presented an ensemble approach integrating finetuned LLama2 via the LoRa framework with BM25, achieving a 15% improvement in retrieval.
- **Graph Embedding for Representation:** Implemented node embedding techniques - Laplacian Eigenmap, DeepWalk, and GCNs for semi-supervised learning to encode both graph structure and node features.

PUBLICATIONS

Decoding loneliness: Can explainable AI help in understanding language differences in lonely older adults? Psychiatry Research, 339, 116078. Wang, N., [Goel, S.](#), Subbalakshmi, K., & Lee, E. (2024).