

# Raj Bharat Sangani

+1 530-601-8374 | [sanganir@uci.edu](mailto:sanganir@uci.edu) | [LinkedIn](#) | [Google Scholar](#) | [GitHub](#) | [Blogs](#) | [Portfolio](#)

## EDUCATION

### University of California, Irvine

Master of Science in Computer Science (GPA: 4.0/4.0)

Advisor: Prof. Sameer Singh

Irvine, CA

September 2023 – May 2025

Coursework: Algorithms, Convex Optimization, Machine Learning, Generative AI, Image Understanding, Reinforcement Learning, Scientific Computing

### Vellore Institute of Technology

Bachelor of Technology in Computer Science and Engineering (GPA: 9.13/10.00)

Coursework: Artificial Intelligence, Natural Language Processing

Chennai, India

June 2018 – June 2022

## EXPERIENCE

### Machine Learning Engineer Intern

Keysight Technologies

- Designed a hybrid search powered **RAG** system from scratch for hardware device manuals and improved on the previous search system by **20 points for NDCG@2 and 24 points for Recall@2**
- Designed and jointly trained a retriever and **re-ranker** with shared bodies to improve RAG performance while maintaining a small memory footprint and search latency of **64ms (for approximately 40 token queries) on AMD EPYC CPUs**
- Scaled batch size for contrastive learning by a **factor of 1000** using gradient caching and cross GPU negative sharing which led to an improvement in NDCG@1 by an additional 1.5 points
- Applied methods from the GISTEmbed paper to guide in-batch sampling of negatives, suppressing false negatives and improving **NDCG@1 by 1 point**

June 2024 – Present

Santa Rosa, CA

### Graduate Student Researcher

University of California Irvine - NLP Group

- Designing novel representations learning techniques for improving long document retrieval and **RAG** systems
- Investigating attention attributions in retrieval models(**LLMs**)

October 2023 – Present

Irvine, CA

### Machine Learning Engineer

Techsquadeg

- Created a Title Generator and Ranker for sports news articles tuning a Pegasus model on scraped news articles which achieved a **ROUGE-L of 31.70** and averages **1000** monthly downloads on the HuggingFace Hub
- Deployed the pipeline using Ray Serve and FASTAPI and optimized the pipeline for CPU inference, reaching **24 queries per second** for models over **2.2 GB** in size.

March 2022 – September 2022

Cairo, Egypt

### Applied Scientist Intern

Quantiphi Inc

- Improved the embedding model used for entity and relation canonicalization through **Generative Pseudo Labeling** to reduce noise in a biomedical knowledge-graph by **7%**
- Built an open-domain Question-Answering system using Multi-Hop Dense Retriever (from FAIR) and a fine-tuned MiniLM Reader, and improved the end-to-end **F1-score by 9 points**
- Co-Authored *Large-Scale Knowledge Synthesis and Complex Information Retrieval from Biomedical Documents* in IEEE International Conference on Big Data, Osaka, Japan, 2022

January 2022 – June 2022

Mumbai, India

## PUBLICATIONS

### Do People Mirror Emotion Differently with a Human or Text-to-Speech Voice?

Published in *ACM CHI Human Factors in Computing Systems, 2024, Hawaii, USA*

[PDF](#)

### Large-Scale Knowledge Synthesis and Complex Information Retrieval from Biomedical Documents

Published in *IEEE International Conference on Big Data (Big Data) 2022, Osaka, Japan*

[PDF](#)

### Comparing deep sentiment models using quantified local explanations

Published in *International Conference on Smart Technologies, Communication and Robotics 2021 (ICSTCR)*

[PDF](#)

## PROJECTS

### Dreambooth - Personalized Diffusion Model | Python, Hugging Face, PyTorch

- Implemented a diffusion model for personalized face images from text prompts and achieved a **CLIP-I score of 0.97**
- Used **LoRA** and various memory optimizations to train a model over 10GB in size on a 16GB NVIDIA T4 GPU
- Interpolated between LoRA adapters to combine image styles to generate pixelated images and animated images

[DOCS](#)

### Shoppster - Multimodal Search System | Python, Hugging Face, PyTorch

- Created a transformer-based multimodal system to support text to image and text plus image to image search
- Trained a network consisting of ViT Base and MPNet using a contrastive margin loss to achieve a **Recall@100 of 41**
- Pre-trained the Vision Encoder using *SimMIM* and *VitMAE* on fashion images to improve recall of the system

[DOCS](#)

## TECHNICAL SKILLS

**Languages:** Python, C++, R, C, Java, MATLAB

**Frameworks/Databases:** Flask, Node.js, FastAPI, Ray, Tableau, Hadoop, Spark, MySQL, PostgreSQL, MongoDB

**Libraries:** PyTorch, JAX, Keras, Tensorflow, HuggingFace, SpaCy, NLTK, Pandas, NumPy, Matplotlib, Plotly