Raj Bharat Sangani

+1 530-601-8374 | sanganir@uci.edu | LinkedIn | Google Scholar | GitHub | Blogs | Portfolio

EDUCATION

University of California, Irvine　　　　　　　　　　　　　　　　　　　　　Irvine, CA
*Master of Science in Computer Science (**GPA: 4.0/4.0**)*　　　　　*September 2023 – May 2025*
Advisor: Dr. Sameer Singh

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

Vellore Institute of Technology　　　　　　　　　　　　　　　　　　　　Chennai, India
*Bachelor of Technology in Computer Science and Engineering (**GPA: 9.13/10.00**)*　*June 2018 – June 2022*
Coursework: Artificial Intelligence, Natural Language Processing

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

EXPERIENCE

Graduate Student Researcher　　　　　　　　　　　　　　　　　　October 2023 – Present
University of California Irvine - NLP Group　　　　　　　　　　　　　　　　*Irvine, CA*
- Exploring the effects of positional embeddings on utilization of context in long-context Large-Language Models (LLM) for Multi-Document Question-Answering

Graduate Student Researcher　　　　　　　　　　　　　　　September 2022 – June 2023
Prasant Mohapatra Lab, University of California Davis　　　　　　　　　　　　*Davis, CA*
- Investigated the effects of targeted paraphrasing using various paraphrasing strategies on ROUGUE scores for text-summarization models and explored methods to mitigate these effects
- Investigated structure bias in text-summarization datasets, encoder-decoder models and LLMs

Machine Learning Engineer　　　　　　　　　　　　　　　March 2022 – September 2022
Techsquadeg　　　　　　　　　　　　　　　　　　　　　　　　　　　*Cairo, Egypt*
- 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.

Applied Scientist Intern　　　　　　　　　　　　　　　　　January 2022 – June 2022
Quantiphi Inc　　　　　　　　　　　　　　　　　　　　　　　　　　*Mumbai, India*
- 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

PUBLICATIONS

Do People Mirror Emotion Differently with a Human or Text-to-Speech Voice?
Submitted to ACM CHI Human Factors in Computing Systems, 2024

Large-Scale Knowledge Synthesis and Complex Information Retrieval from Biomedical Documents　PDF
Published in 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan

Comparing deep sentiment models using quantified local explanations　PDF
Published in International Conference on Smart Technologies, Communication and Robotics 2021 (ICSTCR)

PROJECTS

Dreambooth - Personalized Diffusion Model | *Python, Hugging Face, PyTorch*　　　DOCS
- 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
- Merged LoRA adapters to combine image styles to generate pixelated images and animated images

Shoppster - Multimodal Search System | *Python, Hugging Face, PyTorch*　　　DOCS
- 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