RAGHAV GAGGAR

Los Angeles, USA | gaggar@usc.edu | +1 (213) 255-6042 | linkedin.com/in/raghav-gaggar | github.com/27rg5 | 27rg5.github.io

EDUCATION

University of Southern California, Los Angeles, USA **December 2023**
Master of Science in Computer Science **(GPA: 3.5/4.0)**
Coursework: ML for Data Science, Database Systems, AI, Data Structures & Algorithms, Deep Learning, NLP, Autonomous Systems

MIT – World Peace University, Pune, India **July 2021**
Bachelor of Technology in Computer Science & Engineering **(GPA: 3.7/4.0)**

TECHNICAL SKILLS

- **Programming Languages:** Python, C, C++, Java, Bash
- **Machine Learning:** PyTorch, Keras, TensorFlow, HuggingFace Transformers, Generative AI, Large Language Models, LangChain, RAG, PySpark, Scikit-learn, XGBoost, OpenCV, NLTK, spaCy, TensorRT, ONNX, NumPy, Pandas, Scipy, Matplotlib, Seaborn, Plotly
- **Database/Web Technologies:** SQL, NoSQL, HTML, CSS, JavaScript, Bootstrap, D3.js, Node.js, PHP, Flask, Gradio
- **Others:** Linux, VMware, Git, Docker, MLflow, Hadoop, Hive, Microsoft Azure Databricks, Tableau

PROFESSIONAL EXPERIENCE

Zilkha Neurogenetic Institute, Los Angeles, USA **June 2022 - Present**
Data Scientist - *Python, Keras/TensorFlow, Docker, Computer Vision, Deep Learning, Statistics, Data Analysis, Time Series*
- Analyzed **100 GB** of human brain **MRI scans** for **Alzheimer's disease detection** using deep learning (research papers in progress).
- Developed, optimized, and deployed **3D computer vision models** for brain region segmentation, enhancing analysis.
- Conducted **regression** and **time series analysis** for pattern recognition in large-scale biomedical datasets.
- Generated Python scripts to efficiently process and statistically examine datasets of images exceeding **300 GB** in size.
- Engineered a Python application reducing noise in MRI scans by over **70%**, improving image clarity.

NVIDIA, Bengaluru, India **August 2021 - April 2022**
Machine Learning Research Intern - *Python, PyTorch, OpenCV, NVIDIA DALI, Docker, Computer Vision, Generative AI, Deep Learning*
- Investigated and trained deep learning algorithms on **32 GB** of data to generate human faces through their voice samples.
- Designed and implemented a **novel GAN** to generate colored facial images, paired with a **CNN** for quality assessment.
- Streamlined preprocessing pipeline using OpenCV and NVIDIA DALI, boosting model speed by **15%**.
- Increased training speed by **4.5x** using multi-GPU setups and mixed precision training to enhance computational efficiency.

AlgoAsylum, Pune, India **June 2020 - July 2021**
Data Science Intern - *Python, SQL, Scikit-learn, Machine Learning, Statistics, Signal Processing, Data Analysis, Time Series*
- Analyzed night-light satellite images of India to correlate luminosity with economic prosperity and presented findings at **PyCon India 2020**.
- Conducted signal processing and statistical experiments in Python to study rainfall patterns of the Indian Monsoon using **21 years of satellite data**, employing **Gaussian Mixture Models**, **Wasserstein distance**, and **Jaccard distance**.

PUBLICATIONS

- **Extraction and Summarization of Explicit Video Content using Multi-Modal Deep Learning**, arXiv (link)
- **Machine-Generated Text Detection using Deep Learning**, arXiv (link)
- **Electroencephalogram Based Depression Assessment Using Machine and Deep Learning Techniques: A Survey**, International Journal of Creative Research Thoughts (IJCRT), November 2020. DOI - http://doi.one/10.1729/Journal.24918 (link)

PROJECTS

Text Summarization - *Python, PyTorch, LLMs, Generative AI, Reinforcement Learning, RAG, LangChain, Gradio, Prompt Engineering*
- Created an application for generating concise summaries of human conversations and deployed it on **HuggingFace spaces**.
- Leveraged the SAMSum dataset to **fine-tune** the **FLAN-T5** model with **LoRA**, boosting **ROUGE** score by **11%** and curbing harmful language in summaries by **55%** through **RLHF** with a **PPO** model.
- Applied **Retrieval-Augmented Generation (RAG)** to further improve results, refining the quality and relevance of outputs.

Intelligent bot for GO game - *Python, Artificial Intelligence, Reinforcement Learning*
- Created an intelligent bot to play GO on a 5x5 board as part of an AI course during my master's degree.
- Applied **minimax algorithm with alpha-beta pruning** to defeat sophisticated bots, including those based on the Q-learning algorithm, achieving a **95.2%** win rate and ranking **5th** in a class of **300+** students.

Sequence Labeling - *Python, PyTorch, Deep Learning, NLP*
- Built deep learning models trained on the CoNLL-2003 dataset for **Named Entity Recognition**.
- Achieved an F1 score of **0.80** with a **BiLSTM** model using custom embeddings, improving it to **0.85** with GloVe embeddings.

Hidden Markov Model from scratch - *Python, NLP, Machine Learning, Data Structures and Algorithms*
- Built an **HMM** with a greedy decoding algorithm from scratch to predict part-of-speech tags.
- Elevated test accuracy from **92.8%** to **93.1%** by implementing the **Viterbi algorithm** for optimal decoding.

Automatic Colorization of Images - *Python, Keras/TensorFlow, OpenCV, Computer Vision, Generative AI*
- Developed a **convolutional autoencoder** to colorize grayscale images, achieving a Similarity Index (SSIM) score of **0.85**.
- Integrated **VGG-16** as an encoder and designed a decoder using convolutional and upsampling layers for image colorization.

ACHIEVEMENTS

- Completed the '**Humanoid Robotics**' course at my university in 2019, and then became its instructor for **70+** students. Partnered with 3 colleagues to program the **'NAO' robot** to aid the elderly, which received appreciation from a **Union Minister of India**.