Pratik Bhangale

3948 E River RD, Monroe County, West Henrietta, NY 14586
+15859107690 pratik.bhangale@icloud.com pratik–bhangale PratikBhangale

Technical Skills

Languages: Python, R, C++, C, Java, Javascript, Kotlin, SQL
Developer Tools: Jupyter Notebooks, VS Code, Microsoft Azure MLOps, Google Cloud Platform, Tableau, GitHub
Technologies/Frameworks: Tensorflow, PyTorch, Retrival Augmented Generation(RAG), LangGraph, LLM Inference, FAISS, Pinecone, Hugging Face Transformers, Knowledge Graphs, Chatbot, NLP, Scikit-learn, Docker, Kubernetes.

Education

Rochester Institute of Technology **Aug. 2023 – May 2025**
Masters of Science in Artificial Intelligence (3.79 - 4.00 GPA) *Rochester, NY, USA*

MIT ADT University **Aug. 2018 – May 2022**
Bachelors of Technology in Computer Science and Engineering (7.53 - 10.00 CGPA) *Pune, India*

Relevant Coursework

- Natural Language Processing
- Deep Learning
- Machine Learning for Difficult Data
- Visual Analytics
- Ethics of AI
- Research Methods for AI
- Mathematical Methods for AI
- Fundamentals of AI

Experience

LTI Mindtree **Jan 2021 – Feb 2021**
Data Science Intern | Python, TensorFlow, Scikit-Learn, Pandas, and Seaborn. *Mumbai, India*

- Executed customer segmentation using 10 clustering algorithms on data gathered from 1,00,000 people.
- Evaluated the performance and efficiency of these algorithms and determined the 3 best algorithms.
- Implemented advanced descriptive statistics, data visualization, and dimensionality reduction techniques and boosted data analysis accuracy by 10 percent.

Projects

Generative AI-Driven Medical Imaging and Diagnosis Support | *In-Progress* **Jan 2025**

- *Python, LangGraph, Image Segmentation, Azure MLStudio*
- Developed and integrated a medical image segmentation model with Graph RAG and Similarity Search RAG into clinical workflows, enhancing diagnostic accuracy by 10 percent and reducing knowledge retrieval time by 20 percent, expediting early disease detection and patient care.
- Leveraged Azure Machine Learning Studio to streamline model deployment, reducing deployment time by 15 percent and enhancing scalability. Github Repo 1.

SAP Installation Search Automation Using Agentic AI and Multi-Modal RAG | *In-Progress* **Jan 2025**

- *Transformers, Torch, Datasets, PEFT*
- Built an automated search system for SAP installation guides using Agentic AI and Multi-Modal RAG, utilizing Python, PyTorch, and Hugging Face Transformers to streamline user queries.
- Integrated Pinecone vector search and LangChain orchestration with OpenAI embeddings, reducing retrieval latency. Github Repo 3.

Fine-Tuning, Knowledge Distillation, and Pruning of LLM's | *Transformers, Torch, Datasets, PEFT* **Mar 2024**

- Engineered model optimization like knowledge distillation and pruning on LLM Models resulting in a 87.5 percent reduction in model size while maintaining performance.
- Designed a data preprocessing pipeline that handled over 200,000 conversation records improving data quality and model training efficiency. Github Repo 2.

Object Detection with Android App Deployment. | *Python, Tensorflow, Numpy, Java, Android Studio* **Jan 2022**

- Trained and fine-tuned a neural network model achieving a final accuracy of 97 percent. Quantized and reduced the size of the model for use with edge devices from 150 megabytes to 40 megabytes.
- Engineered an Android application for the deployment of the previously trained neural network model. Model Training 4 Android App 4.

Publications

"Fake Review Identification using Various Kinds of Neural Networks – A Systematic Review." **Sep 2022**

- DOI - R. Sajjan, P. Bhangale, P. Desai