Vivek Modi

vivekvm84001@gmail.com | +12679285072 | github.com/viper-vm | in/vivek-modi1 | viper-vm.github.io

Summary:

- Machine Learning Engineer specializing in **Agentic Workflows, Generative AI, and Medical LLMs**.
- Experienced in **low-latency AI solutions**, LLM fine-tuning, RAG architectures, and MLOps pipelines.
- Strong track record in **deep learning, NLP, and bioinformatics**, translating research into real-world impact.

Skills:

- **Languages:** Python, C, C++, JavaScript, React.js, Node.js, Java, ML, NLP, MLOps
- **Framework:** Langchain, PyTorch, TensorFlow, PySpark, Hugging Face, Scikit-Learn, Pandas, Numpy, Matplotlib
- **Technology:** LLM, Generative AI, RAG, Fine-tuning, PEFT, LoRA, Kafka, Argo, Jenkins, evaluations
- **Tools:** SQL, Django, Flask, LaTeX, AWS, MongoDB, Docker, Jira, Git, CI/CD, A/B testing

Achievements:

- Ranked **701/1.3M** in the JEE ADVANCED 2018 (India).
- **Vice President,** Graduate Student Organization, Rutgers University.
- **4th place**, INTER **IIT Tech meet** - IIT Bombay for Campus Sustainability Challenge.
- **Teaching Assistant** at LEAP Academy University Charter School, New Jersey.

Professional Experience:

- **Machine Learning Engineer (GenAI/ML), Optimoz, Rockville, MD [May, 2024 - Present]**
  - Built **Agentic Workflows** for enterprise automation.
  - Implemented **citation extraction** using LLMs to improve **knowledge base indexing and retrieval** with accurate references in responses.
  - Fine-tuned **medical-domain LLMs** to enhance performance in clinical information extraction and evaluation tasks.
  - Built an **automated eval harness** for clinical IE **(factuality, retrieval precision@5 increased to 0.84)**.
  - Optimized p95 inference latency **85%** by restructuring inference pipeline (quantization, caching, async batching).
  - Developed and deployed **Custom Medical LLM solution** for **Optalk.ai** for low-latency chat interactions.
- **Graduate Research Assistant, Machine Learning & Bioinformatics Lab, Rutgers University [Jan, 2023 - May,2024]**
  - Initiated a comprehensive deep learning pipeline to **identify human activities**, enhancing machine's understanding of complex movements.
  - Streamlined a ML learning-based pipeline to identify **protein subcellular sequences** working under the guidance of Dr. Iman Dehzangi.
- **Machine Learning Intern, Capgemini, Gandhinagar, India [May, 2020-Aug, 2020]**
  - Built **'Priority Mailbox'** and Sentiment Analysis **COM add-in** for **MS Outlook**, streamlining email management and improving email prioritization.
  - Utilized **Django, sentiment analysis models, and SQLite** to classify and rank emails based on engagement metrics.
  - Led key contributions in **optimizing mail organization** and enhancing productivity within Outlook.

Projects:

- **ComicBot: ChatBot Generating Jokes along with GIF [Aug, 2020-Jan, 2021]**
  - Authored an innovative research paper focused on creating jokes paired with GIFs utilizing the **knowBERT** model.
  - Built **sarcasm recognition** and emotion classification to improve **contextually aligned GIFs**.
  - Introduced "**EmotionGIF**," a novel dataset curated to categorize GIFs based on emotive labels and proposed a unique **style transfer** methodology for producing humorous content.
- **Protein Subcellular Localization Prediction Using Machine Learning [Sep, 2023-Jan, 2024]**
  - Predicted **protein subcellular localization** in Gram-positive and negative bacteria using ML models.
  - Extracted and analyzed protein sequence features and attributes, generating **87.2% accuracy**.
  - Engineered sequence features for improved biological insight, contributing to advancements in bioinformatics and computational biology.
- **Human Activity Recognition (HAR) Using Machine Learning under Prof. Iman Dehzangi [Nov, 2022-May, 2023]**
  - Applied **CNN, LSTM, Multimodal Transformer, and Action Transformer** architectures on mPOSE-21 and UCI HAR datasets.
  - Achieved **Macro-F1 84.1 - LOSO** (Multimodal) and 88.4% accuracy (Action Transformer) on UCI HAR.

Education:

MS in Computer Science - **Rutgers University** *(2022 - 2024)*
B.Tech in Computer Science - **IIT, Gandhinagar, India** *(2018 - 2022)*