Poorna Chandra Vemula
+1(857)313-1260 | vemula.p@northeastern.edu | linkedin.com/in/poorna-chandra-vemula/

EDUCATION

Northeastern University, Boston, MA — September 2022 – December 2024
Khoury College of Computer Sciences — GPA: 3.95/4.0
Master of Science in Artificial Intelligence
Coursework: Machine Learning, Computer Vision, Algorithms, Natural Language Processing, Large Language Models, Robotic Science & Systems **Graduate Teaching Assistant** for Natural Language Processing, Computer Vision, Data Structures and Algorithms

Vellore Institute of Technology, Vellore, India — July 2018 – May 2022
Bachelor of Engineering in Computer Science and Engineering — GPA: 9.03/10.0

TECHNICAL SKILLS

Languages:	Python, C/ C++, Java, SQL, HTML, CSS, Javascript
Skills:	Software Engineering, Machine Learning, Generative AI (LLMs, Multimodal, Diffusion models), MLops
Databases & Analytics:	PostgreSQL, MySQL, MongoDB, Prometheus, Grafana, Tableau
Cloud, Tools and Libraries:	TensorFlow, PyTorch, Huggingface, LangChain, Spark, Hadoop, AWS, Linux, Jenkins, Docker, Kubernetes, Airflow, MLFlow, AzureML

WORK EXPERIENCE

Northeastern University, Boston — May 2024 – Present
Research Assistant | Python, PyTorch, LLMs, Planning, Deep Learning (Research – Dr. Yasin Yazicioglu)

- Improving swarm robot control with **Natural language interface** especially in the areas of multi-robot surveillance, collaborative transport, foraging.
- Experimenting with Prompt-tuning/Fine-tuning the LLaMA2, GPT 4 to solve the task assignment problem where LLM assigns subtasks to the agents from the high-level task, also additional low-level plans.

Arkose Labs, San Mateo, CA (Bot & Fraud Detection Team) — August 2023 – Jan 2024
Machine Learning Research Intern | Python, PyTorch, Computer Vision, AWS, Deep Learning

- Implemented **Adversarial attack** strategies including FGSM, PGD, Adversarial patches using PyTorch, thereby reducing the attacks that use object detection models by over 70%
- Implemented **Neural style transfer** using **GANs**, to generate numerous different style images from a single captcha image drastically improving new captcha generation capacity and **mitigating bots** by over 50%
- Developed bot mitigation strategies against multimodal LLM based attacks by mimicking an attacker **prompt-tuning, CoT prompting** GPT-4 Vision model.
- Architected scalable **ETL** pipelines using Apache **Spark** and SQL to process captcha session data, generating features for modeling attack traffic.
- Collaborated with cross-functional teams on the development of **Anomaly detection** model to tag attack traffic there by reducing the workload on security analysts by over 30%.

AGOT, Pittsburgh, PA (Vision Intelligence Team) — May 2023 – August 2023
Machine Learning Intern | Python, TensorFlow, PyTorch, Deep Learning, Docker

- Developed a real-time **Tracking System**, utilizing YOLO for object detection and DeepSORT for adept multi-object tracking to metrics such as speed of service.
- Fine-tuned a U-Net based instance **Segmentation** models using Nvidia's TAO Toolkit. Further, **Optimized** the model for Nvidia Jetson Orin/Xavier using TensorRT improving inference performance by over 30%.
- Setup **Model monitoring** using Prometheus and Grafana. Monitored the model for performance drops, considering data drifts and concept drifts thereby ensuring consistent model performance.
- Researched on various **Multimodal** pre-trained models for the task of visual question answering, fine-tuned Video-LLaMA on custom videos data to describe events in the video.

IBM, Bangalore, India (Power Systems Performance Team)
Software Engineer, Data/ML Intern | Python, Scikit-learn, React, Node.js, Spark, Linux, Jenkins — January 2022 - May 2022

- Optimized CI test to collect **Performance metrics** run data of various **workloads** (Memory, Network, MLperf etc.) for new linux builds on POWER chips and labelled degradations with the help of domain experts.
- Utilized Apache Spark to analyze workload data and trained a ML model with Spark MLlib, achieving a 50% efficiency in detecting bugs by high-level classification there by helping performance team.
- Built a web-application to manage & visualize the workload-performance data across different **Linux builds** & identify **regressions** thereby reducing workload on performance analysts by over 20%.

PROJECTS

Virtual Teaching Assistant | RAG, LLMs, LangChain, Hugging Face, LLaMA2

- Developed LLM-powered assistant using RAG to assist students with questions regarding course material. Created a custom dataset and fine-tuned the LLaMA2 model in a Multi-GPU setting using Ray, Deepspeed for better domain adaptation.

Modifying Path for Single Robot using NLI| ROS, LLMs, Motion Planning

- Developed a model using BERT, CLIP and a transformer decoder block to modify the path slightly based on a Natural language in a particular case where there are objects around.

Deep Fake Detection| Pytorch, Diffusion models (StyleGAN, Stable Diffusion, Dalle)

- Developed a CNN based classifier to classify real and AI-generated faces. Trained a EfficientNet-B1 via transfer learning on a huge dataset of real and AI-generated human faces.

Sentence Similarity Problem | Python, Sci-kit learn, PyTorch, Machine Learning, Deep Learning

- Worked on Feature extraction using Bow, TFIDF, Glove Embeddings. Explored ML algorithms such as Random Forest, Naïve Bayes, SVM, Logistic Regression. Implemented BiLSTM based architecture and fine-tuned BERT on Quora dataset.