Vijay Chevireddi

240-9409294 | vijaydevmasters@gmail.com | LinkedIn | GitHub | Website

Education

University of Maryland Aug 2023 – May 2025
Master of Engineering (M.Eng.) – Robotics | GPA – 3.77 *College Park, MD*
Coursework: Multi-Modal Models, Natural Language Processing, Deep Learning, Perception, Localization, Path Planning

Osmania University Aug 2018 – June 2022
Bachelor of Engineering – Mechanical Engineering | GPA – 9.05/10 *Hyderabad, India*

Skills

Deep Learning & Machine Learning – Transformers, CNNs, Object Detection, LLMs, NLP, Multimodal Learning, Generative Models, Fine-Tuning, AI Alignment, AI Distillation, Reinforcement Learning, RLHF
Cloud and DevOps – AWS Certified (Link), Amazon SageMaker, Docker, Git
Programming Languages – Python, C++, MATLAB, SQL
Frameworks & Libraries – PyTorch, TensorFlow, OpenCV, Open3D, Hugging Face Transformers, TIMM, Open AI SDK, LangChain, LangGraph, RAG
Optimization & Tools – CUDA, SciPy, NumPy, Pandas, TensorRT, ONNX, Statistical Methods, ROS 2, SOLIDWORKS
Simulations – CARLA, ROS, Issac Sim, OpenAI Gym, SLAM, Path Planning, Data Visualization

Experience

UMD Fischell Department of Bioengineering | Machine Learning Intern June 2024 – Present

- Implemented YOLOv8 for dredger detection in a University of Maryland initiative to modernize oyster farming, later upgraded to RT-DETR which improved IoU by 6%.
- Developed human pose detection for walking style classification using MMPose and a CNN transformer, enhancing classification accuracy and contributing to improved walking analysis.

Sai Vamsi Industries | Machine Learning Engineer Jan 2022 – May 2023

- Integrated a camera-based system using the YOLO framework to detect visual anomalies in press tool machines, showcasing applied computer vision and image understanding skills.
- Combined vibration data from Fluke 3561 FC Vibration Sensors with real-time image analysis to enhance defect detection by 5%, demonstrating practical software development and distributed training methodologies.

Projects

LLM-Powered Personalized Learning Pipeline – LLMs, RAG, and Speech Synthesis | Link May 2025 Global MIT AI Hackathon
Tools and Technologies – Python, PyTorch, AWS, Amazon SageMaker, Gemini 1.5 Flash, Tavily, ElevenLabs API, Flask

- Built an LLM-driven assistant that transforms user prompts into curated five-minute educational audio snippets using Gemini 1.5, Tavily, ElevenLabs APIs and RAG (Retrieval Augmented Generation).
- Engineered structured prompts and postprocessing logic to extract user intent, infer latent curiosity, and expand learning topics.
- Implemented a fully automated pipeline from natural language to high-quality speech, demonstrating applied prompt engineering, reasoning control, and real-time generation.

Agentic AI Vision-Based Task Planner using DeepSeek's R1-distilled LLaMA 70B Model Mar 2025 UMD
Tools and Technologies – Python, PyTorch, AWS, YOLO v8, GroqCloud, OpenCV, CUDA

- Built an autonomous AI workflow using YOLOv8 and DeepSeek's R1-distilled LLaMA-70B to detect objects, predict possible interactions, and plan steps to organize a messy workspace.
- Implemented dynamic task generation with automatic validation and correction, ensuring reliable and executable plans for intelligent scene organization.

Transformer Based NLP Pipeline for Robot Navigation with LoRA Optimization | Link Dec 2024 UMD
Tools and Technologies – Python, PyTorch,, Hugging Face, ROS2, Gazebo, CUDA

- Designed a custom transformer-based 3D object detection model for LiDAR point clouds, trained using KITTI dataset.
- Integrated Low-Rank Adaptation (LoRA) to reduce training parameters by 99.64% while retaining 92.5% test accuracy.
- Implemented the resulting NLP model with ROS2 and Gazebo to validate autonomous navigation on a TurtleBot3.

Transformers for 3D Object Detection in LiDAR Point Clouds | Link Dec 2024 UMD
Tools and Technologies – Python, PyTorch, Open3D, KITTI, CUDA

- Developed and trained a custom transformer-based 3D object detection model for LiDAR point clouds on the KITTI dataset.
- Combined PointNet++ embeddings with specialized transformer encoders and introduced a novel loss function, achieving a 3D bounding box detection mAP IoU of 0.67—on par with leading benchmarks.

Vision-Augmented Deep-Q Learning for Mario Using SWIN Transformer | Link May 2024 UMD
Tools and Technologies – Python, PyTorch, Open AI Gym, OpenCV, Git, CUDA

- This project combines DQN with a SWIN Transformer to train Mario-playing AI agent.
- Enabling efficient interpretation of complex visuals and achieving a 500-moving average reward of 2700 in just 1250 episodes-far outperforming a model without SWIN.

Advanced Vision Systems for Autonomous Navigation Using YOLOv8 and Homography | Link May 2024 UMD
Tools and Technologies – Python, Pytorch, YOLO v5, OpenCV, ROS2, Gazebo

- Applied homography for dynamic route planning. Integrated optical flow and YOLOv5 to the pipeline to detect obstacles and their velocities in real time, enabling adaptive navigation.
- By leveraging projective geometry to locate vanishing points and horizon lines, the robot achieved over 80% successful navigation.