

Sankalp Rajeev

 Dearborn, MI |  (480) 208-2139 |  srajeev@umich.edu |  LinkedIn |  GitHub |  Portfolio

EDUCATION

University of Michigan – Dearborn <i>M.Sc., Artificial Intelligence & M.S.E., Robotics Engineering; GPA: 3.90/4.00</i>	Sep 2024 – Dec 2026 (Expected)
	<i>Dearborn, MI</i>
Arizona State University <i>B.S., Computer Science; Minor: Mathematical Concepts of Engineering; GPA: 3.71/4.00</i>	Aug 2020 – May 2024
	<i>Tempe, AZ</i>

WORK EXPERIENCE

Zoetis <i>AI Engineering Intern</i>	May 2025 – Nov 2025
• Deployed 4 production LLM automation tools (GPT-4, Claude) on Azure Kubernetes, reducing regulated documentation workload by 1,000+ engineer-hours annually.	<i>Kalamazoo, MI</i>
• Built asynchronous ETL and semantic similarity pipelines for ServiceNow workflows, reducing incident misclassification by ~40%.	
• Developed MySQL-backed analytics dashboards for stakeholders, enabling scalable cross-team adoption.	
Mistral Solutions Pvt Ltd <i>AI/ML Engineer Intern</i>	Jun 2024 – Aug 2024
• Deployed YOLO-based real-time UAV object detection on Qualcomm edge hardware with optimized inference for field-tested drone systems.	<i>Bangalore, India</i>
• Implemented LiDAR obstacle avoidance and telemetry scripts for autonomous mission execution.	
Astroseed – Terra Pod Capstone <i>Robotics Software Engineer (Scrum Master)</i>	Aug 2023 – May 2024
• Led development of autonomous robotic arm integrating YOLOv5 (94% accuracy) and depth sensing for precision plant maintenance.	<i>Tempe, AZ</i>
• Built ROS-based perception and SLAM simulation pipeline in Gazebo/RVIZ; coordinated sprints and stakeholder demos as Scrum Master.	
Machani Robotics <i>Software Engineering Intern</i>	Jun 2022 – Aug 2022
• Developed control software for a 6-DOF robotic arm via CAN (C++/Python), integrating vision feedback.	<i>Bangalore, India</i>

TECHNICAL PROJECTS

CodeMind – AI Code Intelligence Platform <i>Python, CrewAI, RAG, MLflow</i>	
• Built a full-stack application with CrewAI-based multi-agent system for large-codebase analysis (code exploration, security review, test generation, and refactoring guidance).	
• Implemented hybrid RAG (Vector + BM25 + RRF fusion) for code Q&A, achieving 15.4ms median retrieval latency on 42k+ indexed chunks.	
• Fine-tuned Gemma 3 (12B) using QLoRA/4-bit quantization for test generation, achieving 74% CodeBLEU improvement and 3.3x inference speedup.	
MeetingMind AI – Multimodal Meeting Intelligence <i>Python, FastAPI, React, GCP</i>	
• Architected a 5-phase multimodal ML pipeline (audio diarization, face tracking, speaker-face matching, LLM summarization, visual intelligence) coordinating 6+ models across audio and video modalities.	
• Architected a speaker-to-face association system with VLM capabilities, reducing expensive API calls by 85% via content-aware pre-filtering that restricts analysis to high-value semantic frames.	
• Deployed a production-ready full-stack system (23 REST endpoints) on Google Cloud Run with RAG-based Q&A using LangChain, ChromaDB, and Vertex AI embeddings.	

SKILLS

Languages	Python, C/C++, SQL, TypeScript, Java
LLMs & Agents	RAG (Hybrid Search/RRF), CrewAI, LangChain, QLoRA, Vertex AI (Gemini)
Machine Learning	PyTorch, Transformers, scikit-learn, Whisper, Pyannote
Cloud & DevOps	Docker, Azure (AKS), GCP (Cloud Run, Vertex AI), FastAPI, Vector DBs (ChromaDB)
Computer Vision	YOLO (v5/v11), OpenCV, InsightFace, EasyOCR, SLAM, ROS