Joe Lanzi

AI Engineer| Palatine IL, USA | Email: JoeLanzi@ymail.com | Phone: (347) 207-9726 | Portfolio: joelanzi.vercel.app

Summary

Combined 5+ years of experience developing and leading AI and ML projects in both private industry and research, specializing in multimodal generative AI, cybersecurity, health care, finance, transportation, geospatial, and object recognition. Expertise in end-to-end AI and ML systems design, leveraging both cloud platforms and edge devices to deliver customized, scalable, and efficient solutions to complex business software problems.

Professional Experience

Guild Mortgage, San Diego, CA

AI Engineer Jan 2025 – Current

- Built and deployed GuildIQ, a multi-agentic framework revolutionizing knowledge access and marketing review.
- Developed a TypeScript/React web app front-end and optimized a Python API backend to manage multi-agentic workflows with custom agent tools to expand business solution capabilities.
- Enhanced document processing speeds, reduced response times, and minimized operational overhead with full-stack optimizations.

GovernmentGPT, Mountain View, CA

Lead AI Engineer Jan 2024 – Dec 2024

- Built multi-modal feedback solutions (visual, auditory, haptic) that dynamically adapted to environmental inputs, enhancing situational awareness and decision-making in real-time applications for high-stakes environments.
- Designed, trained, and deployed AI models for cloud and edge environments, optimizing real-time performance while balancing hardware constraints, latency requirements, and scalability across diverse use cases.
- Fine-tuned large language models (LLMs), Visual LLMs, and implemented Retrieval-Augmented Generation (RAG) for domain-specific civilian, government, and military applications, including live image classification, threat detection, and real-time operational analysis.
- Developed and optimized auditory transcription and translation pipelines, achieving 10× faster inference speeds compared to OpenAI's audio models, with hardware-specific performance enhancements on NVIDIA Jetson devices and cloud platforms, enabling real-time multilingual processing.
- Led end-to-end pipeline development, integrating AI models with scalable AWS and edge platforms, streamlining workflows, and collaborating with hardware engineers to ensure seamless software-hardware integration for robust, real-world deployments.

Object Computing Inc., St. Louis, MO

Data Scientist - Machine Learning and AI Aug 2022 – Dec 2023

- Led the development of advanced geospatial ML solutions for crop prediction, terrain forecasting, and natural disaster analysis using ArcGIS, Google Earth Engine, and Google cloud services. Pioneered methods for accurate damage assessment and flood risk modeling, significantly accelerating response times in collaborations with Google and Planet Data Labs. Built custom model pipelines in GCP and utilized REST APIs for integration into client systems.
- Developed AI solutions for automated data analysis and customer service, reducing engineering intervention and operational costs; included chatbots, predictive models, summarization tools, transcription tools, and other custom tool development tailored to client needs and requirements.
- Developed and deployed production-grade ML systems on AWS for precise malware categorization, enhancing threat detection across diverse file types. Utilized advanced techniques for the accurate classification of ambiguous files, resulting in a robust, automated solution for real-time security analysis. Containerized models using Docker and Kubernetes.
- Utilized cloud services (AWS and Azure) to develop and deploy ML models to categorize and uncover fraudulent activities within the Missouri Medicaid system.

Saint Louis University, St. Louis, MO

Research Associate May 2022 – Dec 2022

- Investigate and develop new methods to improve the visual categorization of deep learning models used for an image search system used at the National Center for Missing and Exploited Children in human trafficking investigations.
- Developed a reinforcement learning model to simulate a virtual economy, enabling non-player characters to follow survival and thriving strategies while analyzing their behavior under various economic conditions.

New York Stem Cell Foundation, New York, NY

Software Engineer June 2020 – Feb 2021

- Develop a convolutional neural network model for cell segmentation from nuclear and non-nuclear stains.
- Develop a deep learning object detection algorithm to identify nuclei with specific biological characteristics from nuclear stained images.
- Develop a series of deep learning techniques capable of classifying tumor cells in a 3D stack and forecast the cell growth from benign cell to malignant cell categories.

Manhattan College, Riverdale, NY

Research Assistant | Deep Learning Programmer January 2019 – May 2020

- Worked with neuromorphic vision sensors (Polarimetric Dynamic Vision Sensor) that creates super resolution images of moving or semi-obstructed targets and designed bioinspired-based vision systems for classification, identification, motion detection, and tracking.
- Improve contrast sensitivity and created efficient strategies for rapid scene analysis with the use of deep learning.
- Developed deep learning techniques to discriminate moving targets based on motion presented.
- Presented methods that accurately classify motion pattern-based targets using limited data, low storage, low power consumptions, and high-processing speed.
- Developed a partial facial recognition system capable of 99% accuracy for identity recognition of persons with facial coverings including masks, hats, and glasses.

United States Marine Corps Camp Lejeune, NC

Intelligence Support & Instructor June 2013 – June 2018

- Analyzed operational elements and provided strategic briefs to support future operations.
- Instructed 20,000+ personnel, designing and delivering training programs with 400+ hours of teaching and public speaking.

Skills

- **Machine Learning & AI:** Deep Learning, Convolutional Neural Networks, Reinforcement Learning, Retrieval-Augmented Generation (RAG), Large Language Models (LLMs), Agentic Orchestration
- **Programming Languages:** Python, C, C++, Typescript, JavaScript, SQL
- **Cloud Services:** GCP, AWS, MS Azure
- **GIS Softwares:** Google Earth Engine, ArcGIS
- **Databases:** SQL, BigQuery, RDS, CosmosDB
- **Version Control:** Git, GitHub, Gitlab, Replit, Docker
- **Systems Design**: Dataflow, Glue, Lambda, Cloud Functions, API endpoints, EC2, Compute, VertexAI, Sagemaker
- **Framework**: FastAPI, Flask, REST API
- **Edge Devices:** NVIDIA Jetson deployment, ARM GPU local servers
- **Soft Skills:** Strategic Planning, Negotiation, Team Leadership, Organizational Skills, Classroom Instruction, System Integration
- **Collaborative Development:** Agile, Scrum, Kanban

Education

Saint Louis University | St. Louis, MO MS in Artificial Intelligence August 2021 – Dec 2022

Manhattan College | New York, NY BS in Computer Engineering August 2018 – May 2021

Publications

- A Cognitive Radar for Classification of Resident Space Objects (RSO) operating on Polarimetric Retina Vision Sensors and Deep Learning IEEE
- Super-Resolution Spike Event-based Polarimetric Dynamic Vision Sensor p(DVS) Cognitive Imaging IEEE