

EDUCATION

- **University of California, Davis** Davis, CA
Bachelor of Science in Computer Science; GPA: 3.8 Dec 2024
 - **Dean’s Honors List:** Multiple semesters
 - **Relevant Coursework:** Artificial Intelligence, Machine Learning, Probability and Statistics

EXPERIENCE

- **Oracle** Austin, TX
Software Engineer Jun 2025 – Present
 - **AI Infrastructure for Real-Time GPU Monitoring:** Developing an AI infrastructure platform using Prometheus, Oracle Kubernetes Engine (OKE), and Terraform to monitor GPU metrics in real-time across distributed cloud instances. Designed alerting, dashboarding, and logging systems to support large-scale LLM and AI workloads.
- **Oracle** Austin, TX
Software Engineer Intern Jun 2024 – Sep 2024
 - **Chatbot Development with Generative AI and RAG:** Implemented a Generative AI-driven chatbot integrated with Oracle Identity Governance (OIG) using Retrieval-Augmented Generation (RAG) techniques. Reduced information retrieval time by 3x through real-time, AI-powered responses, streamlining identity workflows and enhancing user experience.
 - **Agentic AI Automation for Identity Workflows:** Built an agentic AI system combining a RAG-based chatbot and Cohere’s multi-step tool use to automate Oracle Identity Manager (OIM) tasks like user data retrieval and policy updates. Enhanced system interoperability and reduced manual workload by 50%.
- **Center for Mind and Brain** Davis, CA
AI Research Assistant Apr 2023 – May 2025
 - **Lead Developer for AR Assistive Device:** Led development of a wearable AI-powered AR assistive device, leveraging ASR and diarization models to achieve a 15% improvement in transcription accuracy. Demonstrated scalable AI deployment for accessibility applications, aligning with real-world impact goals.
 - **Cloud Migration and Dockerization of ASR Systems:** Spearheaded the migration of ASR systems from offline models to cloud-based solutions utilizing AWS Transcribe and Google Cloud Speech-to-Text, incorporating Docker pipelines to streamline deployment and scalability by 50%.

PROJECTS

- **FAQ Chatbot with Retrieval-Augmented Generation (RAG):** Built an FAQ chatbot leveraging OpenAI GPT models and RAG to retrieve and summarize data from a company’s FAQ section. Created a FAISS-based vectorized database, deployed via Streamlit, and logged interactions with AWS S3 for continuous improvement. Designed a scalable NLP pipeline to preprocess and vectorize data for efficient retrieval.
- **Trash Classification Model with ResNet50:** Trained a ResNet50-based deep learning model for multi-class waste classification, achieving 94% accuracy. Optimized performance through data augmentation and hyperparameter tuning. Deployed the model using Docker, demonstrating scalability and applicability to real-world use cases.

SKILLS

- **Languages:** Python, SQL, C++, Java
- **Machine Learning Frameworks:** PyTorch, TensorFlow, Scikit-Learn, Hugging Face, Keras
- **Tools and Platforms:** Docker, FAISS, AWS (S3, Transcribe, SageMaker), Google Cloud, OCI (23ai), Git, Linux
- **Core Skills:** NLP, Transfer Learning, Data Preprocessing, Feature Engineering, Data Visualization, Model Deployment

CERTIFICATIONS

- **AWS Certified Machine Learning – Specialty** Oct 2024
- **Oracle Machine Learning using Autonomous Database 2024 Certified Associate** Sep 2024
- **OCI 2024 Certified Data Science Professional** Aug 2024
- **OCI 2024 Generative AI Certified Professional** May 2024
- **AWS Certified Cloud Practitioner** Dec 2023