

VARADRAJ BARTAKKE

📞 858-214-6059 ✉️ vbartakke@ucsd.edu [in](#) [LinkedIn](#) [GitHub](#)

EDUCATION

University of California, San Diego

Master of Science in Data Science — GPA: 4.00

Sept 2024 – Present (Expected March 2026)

San Diego, USA

Vishwakarma Institute of Technology, Pune

Bachelor of Technology in Mechanical Engineering — GPA: 3.84(9.63/10)

Aug 2018 – June 2021

Pune, India

EXPERIENCE

University of California, San Diego

Data Assistant

Jan 2026 – Present

San Diego, USA

- Building a taxonomy and search system for **1,800 marine species** with genomic metadata, delivered **Django REST + PostgreSQL** APIs consumed by web and mobile clients.
- Implemented quick species lookup via indexing, query optimization, and **real-time filtering** and a dataset gallery.

Johnson & Johnson

GenAI Data Science Intern

June 2025 – Sept 2025

San Diego, CA

- Built a RAG based **GenAI-assisted clinical document authoring workflow** for healthcare data integrating retrieval, context building, summarization, fact-source attribution and evaluation.
- Developed a multimodal image-captioning system with LangChain & CLIP, achieving **BLEU 0.62**, **METEOR 0.56**, **CLIPScore 0.97**, deployed for scalable inference.

LTIMindtree

Senior Data Engineer

Jul 2021 – Jul 2024

Pune, India

- Architected scalable data pipelines, **improving throughput by 40%**, using AWS to ingest and process 5,000+ structured JSON & XML messages daily integrating with cross-functional teams.
- Migrated and tested 30+ back-end ETL jobs to AWS Lambda & Snowflake by converting to Snowpark Python jobs scheduled via a Scheduler API.
- Built end-to-end CI/CD data architecture, **reducing deployment time by 30%** using Jenkins and Agile practices.

TECHNICAL SKILLS

Programming: Python, C++, NumPy, Pandas, Scikit-learn, FastAPI, Flask

Machine Learning: PyTorch, TensorFlow, GPU, CUDA, Training, Inference, Evaluation, Deep Learning, Neural Networks, CNN, LLM, VLM, Quantization, Pruning, Natural Language Processing, JAX

GenAI & LLMs: LangChain, RAG, MCP, Agentic AI, Multi-modal Inference, Model Evaluation, Prompt Engineering, GPU

Data: SQL, No-SQL, PostgreSQL, Statistics, Spark, Hadoop

Cloud & Other: AWS, Snowflake, Docker, Kubernetes, Jenkins, CI/CD, Git

Certifications: IBM Data Science Certificate, Snowflake SnowPro Core Developer, AWS Certified Cloud Practitioner

PROJECTS

Invisible Image Watermarking Remover | *Watermark Robustness* [[GitHub](#)]

Sept 2025 – Dec 2025

- Achieved **95.7%** watermark removal with minimal perceptual loss (**PSNR 28.1 dB**, **SSIM 0.82**, **LPIPS 0.078**) by implementing VAE and diffusion based inference pipeline.
- Improved image quality by **+6 dB PSNR** and **+0.18 SSIM** by integrating color-contrast transfer in CIELAB space and fine-tuning a VAE-based reconstruction model.

LLM Post-Training against Adversarial Attacks | *Responsible AI* [[GitHub](#)]

Sept 2025 – Dec 2025

- Performed post-training optimization (**LoRA-based SFT + DPO**) on 3–4B parameter LLMs, improving inference robustness without full retraining
- Built a 20k-example preference dataset (WildJailbreak + Alpaca) and evaluated on AdvBench and JailbreakBench, reducing attack success rate on Qwen-4B from **9.0% to 2.7%** while increasing benign helpfulness from **37% to 72%**

Bird Audio Classifier | *Unstructured Data* [[GitHub](#)]

March 2025 – April 2025

- Built an GPU accelerated audio machine learning pipeline including segmentation of recordings into 5-second windows, time masking, **Mel-spectrogram image creation**, **image signal processing**, **color restoration** and model fine-tuning.
- Fine-tuned pre-trained CNN and PANN architectures on a real-world labeled dataset, achieving **ROC AUC of 0.86** across 206 bird species classification.