NIKHIL CHOWDARY PALETI

nikhil-paleti.github.io ◇ npaleti@ucsd.edu ◇ +1(510) 935-8895 ◇ linkedin.com/in/nikhil-paleti/

EDUCATION

Master of Science in Data Science — **Sep 2024 - Jun 2026**
University of California San Diego — **(Expected)**

B.Tech in Computer Science and Engineering (Artificial Intelligence) — **Oct 2020 - Jun 2024**
Amrita Vishwa Vidyapeetham University, India | **First Class with distinction** — **GPA: 9.15/10**

SKILLS

Programming Languages	Python (Advanced), SQL (Advanced), Scala (Intermediate), C++ (Intermediate)
Libraries	NumPy, Pandas, Matplotlib, Scikit-Learn, PyTorch, TensorFlow, Transformers, PySpark
Machine Learning Specializations	Supervised/Unsupervised Learning, Deep Learning, Reinforcement Learning, Natural Language Processing, Computer Vision, Generative AI, LLMs (Fine-tuning, Prompt Engineering), Time Series Analysis
Tools & Cloud	Docker, Kubernetes, Kafka, Git, AWS, Google Cloud Platform

EXPERIENCE

Machine Learning Engineer Intern — **Jan 2024 - Jun 2024**
TechProfuse | *Python, Gemini LLM, Google Vision API, CV2, ML, REST APIs* — *Hyderabad, India*

- Implemented unstructured data extractor using **Gemini 1.5 Pro LLM**, processing **50k OCR documents in 15 hours** with multiprocessing optimization.
- Led 3-intern team in **image-to-image translation project**, integrating **Google Vision API** and **CV2** to convert **100% of English text to Dutch** in scientific diagrams.
- Developed **Vision API parser** for TIFF images, reducing document processing costs by **95%** compared to AWS solution.
- Designed **mobile app backend** with **Google Vision API and YOLO object detection models**, achieving **99% accuracy** in extracting text, numeric, and OMR data from form images.
- Built customer data dashboard using **REST APIs**, reducing support team's data retrieval time by **70%** and improving response efficiency.
- Automated **email classifier** and **thread summarizer** using **Gemini 1.5 Flash model**, processing **1000+ daily emails** with **95% accuracy**, streamlining communication workflow.

PROJECTS

Indic Verse: Advanced Indic Language LLM System — **Jan 2024 - Apr 2024**
Tech Stack: Python, PyTorch, Hugging Face Transformers, PEFT, NLP — [Hugging Face]

- Engineered English-to-Indic translation and transliteration modules, processing **3 types of datasets** for model training.
- Processed **1M+ sentences** from **6+ datasets** to create a robust training dataset for the Indic language models.
- Fine-tuned **3 LLM models** (Gemma 2B, Gemma 7B, LLaMA 3 8B) using PEFT techniques, improving performance on Indic language tasks.

ChoreBot: Voice-Controlled Robotic Navigation System — **Sep 2023 - Dec 2023**
Tech Stack: Python, Whisper ASR, YOLOv3, DistilBERT, A Algorithm* — [Video Demo]

- Achieved **95% success rate** in autonomous navigation tasks with **AI-powered 2D robotic navigation system**.
- Integrated **Whisper ASR** (98% accuracy) for voice recognition and **DistilBERT-NLU** (95% accuracy) for command understanding.
- Implemented **YOLOv3** for object detection and **A* algorithm** for real-time path planning.

PUBLICATIONS

- *An analysis of data leakage and generalizability in MRI based classification of Parkinson's Disease using explainable 2D Convolutional Neural Networks.* **Digital Signal Processing**, Volume 147. [Link] — **Apr 2024**
- *Enhancing Knee Osteoarthritis severity level classification using diffusion augmented images.* **Proceedings of ICACECS**, pp. 266-274. [Link] — **Dec 2023**
- *Improving Reinforcement Learning Agent Training using Text based Guidance: A study using Commands in Dravidian Languages.* **DravidianLangTech**. [ACL Anthology] — **Sep 2023**

ADDITIONAL INFORMATION

- **Workshop:** Conducted "LLM Applications using LangChain" for 200+ students at Anoka Tech Fest — **Apr 2023**
- **Certification:** Deep Learning Specialization - Coursera, instructed by Andrew Ng — **Feb 2023**