

PRIYANKA MOORTHY

408-480-8833 | priyankamoorthycit@gmail.com | linkedin.com/in/priyanka-moorthy | github.com/priyanka-moorthy | priyankamoorthy.github.io

EXPERIENCE

AI Engineer

Jul 2023 – Present

Esperecare LLC

San Jose, CA

- **Gathered current workflows across four operations teams**, mapped how each used the source documents, and translated them into a single cohesive tool serving all four.
- Architected the tool as a **router-orchestrated multi-agent system** on Gemini with function calling, deployed on **Cloud Run** with async execution and production observability: an intent-classification router directs queries to five sub-agents (FAQ, exclusions, outreach, letter templates, appeals), blocks out-of-scope requests, and routes low-confidence cases to human review; hit **92% routing accuracy** at ~2s latency and 2–3K queries/day.
- Shipped a **revenue-code assignment service** for itemized-bill line items, using grounded retrieval over a **1.8M-document corpus** of revenue-code references and historical bills: handles ~14K line items per day across multiple claims, replacing ~8 hours of daily manual coding; **98% automated**, with the 2% low-confidence cases routed to human review.
- Built a **data-extraction pipeline** integrated into the application that extracts structured fields from UB-04, HCFA-1500, and itemized bills (previously done by claims assistants), **auto-extracting 96% of records at high confidence** with the rest routed to claims-assistant review; fine-tuned domain-specific models for a **27% accuracy lift**.
- Built a continuous evaluation harness on Vertex AI with pairwise preference scoring, human feedback, and **regression gating across agents and pipelines**, plus an LLM-as-judge harness for automated output QA.

Software Engineer

Jan 2019 – May 2020

RedBlackTree Tech

India

- Shipped an OCR + CNN classification pipeline to extract provisions from lease documents, **cutting processing time by 82%** at 0.92 AUC.
- Built an async processing system on RabbitMQ and Celery to distribute model inference across workers and control concurrency under production load.
- Designed Django backend services and APIs for high-throughput document processing and model serving.

PROJECTS

Multi-Task Robot Learning

Feb 2026

Vision-Language-Action Models, PyTorch, MuJoCo

- Built a **77M-parameter Vision-Language-Action model** with an EfficientNet-B0 vision encoder, DistilBERT text encoder, and cross-attention fusion that maps camera frames and natural-language instructions to robot manipulation actions.
- Implemented a behavior-cloning training pipeline using scripted expert policies for pick, push, and place tasks; cut training loss 22% with multi-task learning.

Continuous Sign Language Recognition

May 2023

SAM-SLR, PyTorch, OpenCV

- Extended the SAM-SLR framework for continuous, sentence-level sign-language recognition, fusing multi-cue visual streams (appearance and skeleton keypoints) to strengthen temporal modeling of signed sequences.
- Implemented the real-time recognition pipeline and evaluation setup.

TECHNICAL SKILLS

AI Systems: Multi-agent orchestration, RAG / grounded retrieval, function & tool calling, LLM evaluation (pairwise preference, LLM-as-judge, human feedback), prompt & system design, model fine-tuning

ML & Modeling: Deep learning, NLP, reinforcement learning, computer vision, Vision-Language-Action models, behavior cloning

Infrastructure & MLOps: GCP (Vertex AI, Cloud Run), AWS, Docker, Kubernetes, RabbitMQ, Celery, async job systems, model monitoring

Frameworks & Libraries: PyTorch, TensorFlow, Transformers, LangChain, LlamaIndex, Scikit-learn, OpenCV

Languages: Python, SQL, Java, C++, JavaScript, Shell

EDUCATION

San Jose State University · M.S., Artificial Intelligence · San Jose, CA

Aug 2021 – May 2023

Anna University · Integrated M.Sc., Software Systems · India

Aug 2014 – May 2019