

Archana Pradeep

archieprad@gmail.com| arpradeep@ucsd.edu| +1 (424) 380-0979

EDUCATION

UNIVERSITY OF CALIFORNIA SAN DIEGO

MS IN COMPUTER SCIENCE
Grad. March 2025

NATIONAL UNIVERSITY OF SINGAPORE

BENG IN COMPUTER ENGINEERING
Grad. June 2020

SKILLS

Languages:

Python • Go • Java • C

AI Systems:

LLM Agents • Agent Orchestration •
Prompt Caching • AI Evaluation • RAG
• PyTorch

Infrastructure:

Distributed Systems • RPC •
PostgreSQL • Kafka • React • SQL

LINKS

Website:// [archanapradeep](#)

Github:// [archanapradeep](#)

LinkedIn:// [archanapradeep](#)

TEACHING

Machine Learning Fundamentals
(DSC255R)

Winter 2025

Python for Data Science (DSC207R)

Fall 2023

Programming Methodology (CS1010)

Fall 2018

Embedded Systems Design Project

(CG3002, CG4002)

Fall 2019, Spring 2020

COURSEWORK

MLSys

Recommender Systems and Web Mining

Probabilistic AI and Reasoning

Advanced Computer Vision

Natural Language Processing

Graduate Networked Systems

Intro to Artificial Intelligence

Computer Architecture

Data Structures and Algorithms

Software Engineering

Real Time Operating Systems

ACHIEVEMENTS

2022 **Scholarship Recipient** AWS

Machine Learning Engineer Scholarship
Program

2022 **Bronze Medal** 15th National
Taekwondo Poomsae Championships

2021 **Spot Award** Grab

2019 **Third place** 14th STePS under
CS3243 Artificial Intelligence

EXPERIENCE

META

| SOFTWARE ENGINEER, AI FOR CODING / DEVELOPER INFRASTRUCTURE
Apr 2025 - present | Menlo Park

- Built production **LLM-powered coding agent infrastructure** for developer workflows, including agent orchestration, task planning, generated-code review, and developer-facing systems.
- Developed **AI-assisted code review** workflows for generated code using structured plans, risk signals, scope checks, diffs, comments, and reusable review templates.
- Shipped prompt-cache fixes for high-volume LLM agents with 0% cache hits, reducing annualized LLM spend by **\$2.3M**.
- Built persistent **codebase knowledge** and multi-agent task infrastructure using typed schemas, RPC events, subscriptions, backend services, and React UI.
- Built reusable XR/ML modules and release pipelines for **XRPA/VEET**; co-authored **SWE-Protege**, an AI-for-code research paper.

NVIDIA

| SOFTWARE INTERN - AI INFRASTRUCTURE
Jun 2024 - Sep 2024 | Santa Clara

- Intern on the **Autonomous Vehicle** team. Migrated 12K+ lines of functional tests from **SQLite** to **PostgreSQL** for production-faithful test emulation.
- Improved test fidelity and production parity for AI infrastructure validation workflows.

NATIONAL UNIVERSITY OF SINGAPORE

| RESEARCH ASSISTANT
Jul 2022 - Jul 2023 | Singapore

- Worked with Prof **Roger Zimmerman** and Prof **Ooi Wei Tsang** on multi-camera computer vision in video surveillance in collaboration with **Seagate** and funded by **AI Singapore**
- Increased object detection speed by 7x by changing model from **Faster R-CNN** to **YOLOv3**
- Integrated **YOLOv7** for faster and more accurate tracking
- Built end-to-end pipeline** to transfer bounding-boxes and person ID features from multiple Jetson Nanos (multi-cam) to a central server

GRAB

| SENIOR SOFTWARE ENGINEER
Aug 2020 - Jul 2022 | Singapore

- Backend engineer on the **Digibank** team, **Spot award winner**.
- Authored & maintained **microservices** for onboarding, user verification in **Go**.
- Built a bootstrap framework for microservices, supporting **logging, middleware, HTTP routing, graceful shutdown**.
- Designed KYC document upload flow with **Kafka, AV scanning, AWS S3**, and in-house RTC.
- Developed Customer ID generation using **RDS, Redis**

SHOPEE

| DATA ENGINEER
Jun 2020 - Aug 2020 | Singapore

- Worked on the internal version of **Apache Airflow**
- Used **Docker and Python** to improve efficiency of task schedulers by reducing the work done by each scheduler to 50%

VISENZE

| SOFTWARE ENGINEER INTERN
Dec 2019 - Jan 2020 | Singapore

- Implemented gRPC endpoints to replicate existing REST APIs, using **Python, Java** for the **Data Platform Team**

AMAZON

| SOFTWARE DEVELOPMENT ENGINEER INTERN
May 2019 - Aug 2019 | Berlin, Germany

- Built auditors and automated billing processes using **Java, Node.js, AWS Lambdas, SQS, SNS** on the **AWS Bill Generation Team**.

AUTODESK

| SOFTWARE DEVELOPMENT INTERN
May 2018 - Oct 2018 | Singapore

- Worked on building a **Chaos Engineering** tool with **AWS** (EC2, S3, IAM, SimpleDB, DynamoDB), **Java**, **Bash Script**
- Inner sourced** within the company, used by 6 teams

PROJECTS

MEDINQUIRE AI

| BIOMEDICAL NLP PROJECT

- Medical chatbot using **Retrieval-Augmented Generation (RAG)** with contextual query enhancement, combining **FAISS-based** document retrieval and **GPT-4-powered** response generation. Repo :// [here](#)