

Atharvraj Patil

☎ +1 619-213-4020 | ✉ atpatil@ucsd.edu | [in LinkedIn](#) | 📍 San Diego, CA

EDUCATION

University of California San Diego (UCSD)

Sept. 2024 – Dec. 2025 (Expected)

Master of Science in Machine Learning & Data Science | GPA: 3.92/4.0

San Diego, CA

Coursework: Deep Generative Models, Statistical Learning I & II, Visual Learning, Recommender Systems, Optimization and Acceleration of Deep Learning on Various Hardware Platforms, Programming for Data Analysis, Linear Algebra

College of Engineering Pune (COEP)

Aug. 2018 – June. 2022

Bachelor of Technology in Electrical Engineering | GPA: 3.91/4.0

Pune, India

TECHNICAL SKILLS

Languages & Tools: Python, Java, C++, SQL, Git, GitHub, Jira, Kibana

Frameworks & Libraries: PyTorch, TensorFlow, Keras, Scikit-learn, Pandas, NumPy, Matplotlib, OpenCV, LangChain, MLFlow, Django, Spring Boot, FastAPI

Technologies: Machine Learning, Deep Learning, Generative AI, RAG, Microservices, CI/CD, Docker, Kubernetes, AWS

Databases & Platforms: MySQL, PostgreSQL, MongoDB, Databricks, Snowflake

EXPERIENCE

Grainger

June 2025 – Aug. 2025

Applied Machine Learning Intern

Chicago, IL

- Designed and implemented an **prompt optimization pipeline** leveraging the **TextGrad** framework and **LangChain**, resulting in a **25%** improvement in product selection accuracy for Grainger's Digital Assistance.
- Conducted in-depth data analysis using **Databricks** and **Snowflake** to curate a benchmark **dataset**, and **integrated MLflow** to log experiment configurations and evaluation metrics for reproducibility and systematic comparison.

Bajaj Finserv Direct Limited (BFDL)

July 2022 – Aug. 2024

Software Development Engineer II

Pune, India

- Developed the YARA chatbot **backend** using **Django**, **PostgreSQL** and **MongoDB**, enhancing sales team efficiency by automating customer queries and **reducing** response time by **30%**.
- Integrated **Generative AI** capabilities into the chatbot using **Retrieval Augmented Generation (RAG)** and **prompt engineering** through **AWS Bedrock**, enhancing its ability to handle complex customer queries.
- Built a **scalable Spring Boot** microservice to retrieve customer credit scores, optimizing performance with **Redis caching** and asynchronous processing using **AWS SQS and SNS**, which led to a **20%** latency reduction.
- Refactored the esign document generation logic to **HTML-based PDF generation**, reducing PDF generation time by over **50%** and significantly **decreasing file size**.
- Integrated multiple modes of automated income verification into the personal loan journey, leveraging **CI/CD pipelines** for seamless deployment, which led to a notable improvement in user experience and faster customer onboarding.

Autonise

Jan. 2021 – April 2021

Machine Learning Intern

Remote

- Developed a web-based **Quick Draw Recognizer** using **Flask** with an interactive canvas for user sketches and real-time predictions from a **CNN-based PyTorch** model.
- Achieved 92% test accuracy on hand-drawn image classification using **custom-trained CNN** on QuickDraw dataset.

RESEARCH & PROJECTS

Drugchat: Prediction of Drug Mechanism and Properties

Graduate Research Assistant at Prie Lab, UC San Diego

Oct. 2024 – June 2025

- Improved the performance and efficiency of a **large language model** for metabolite property prediction by **fine-tuning** on a curated **multi-modal dataset**, followed by model compression through **quantization and pruning**.
- Collected, processed, and integrated molecular structures, metabolite data, and pharmacological properties to build a high-quality multi-modal dataset, which was used to train and evaluate models on multiple metrics.

KV Cache Implementation for a Transformer Decoder [Repo](#)

PyTorch, HuggingFace

Feb. 2025 – March 2025

- Implemented a custom **Llama Transformer decoder** from scratch with self-attention and **KV cache** for text completion on the Penn Treebank dataset, achieving a **2.2x** speedup in CUDA inference time.

Automatic Number Plate Recognition System (ANPRS) [Repo](#)

Computer Vision, Tensorflow, Keras, OpenCV, Django, MySQL

Dec. 2021 – May 2022

- Developed an application for vehicle detection, license plate extraction, and number recognition using **deep learning models**. Designed and integrated a vehicle authentication system, coordinating API calls to match extracted license plate numbers with a central SQL database to ensure accuracy and efficiency.
- Publication:** "Deep Learning-based Approach for Indian License Plate Recognition using Optical Character Recognition" IEEE GCON 2023. [Link](#)