Mrinaal Dogra

3869 Miramar Street, La Jolla, CA 92092

mdogra@ucsd.edu | mrinaal.dogra19@gmail.com | +1 (858) 518 3496 | mrinaald | mrinaald | mrinaald.github.io

Experience

Kognitos Inc. *San Jose, CA*

Machine Learning Engineer Intern *Jun. 2025 - Sep. 2025*

- Researched and prototyped an **agentic AI pipeline** for **SOP generation** from desktop recordings, integrating **key-frame extraction** (CV-based), **OCR**, and **multi-agent reasoning**; benchmarked **GPT-4o**, **GPT-4.1**, **GPT-5**, and **Gemini-2.5**.
- Built and deployed a lightweight **multimodal web app** (video, images, documents) for automated **SOP and flow-diagram generation**, **containerized with Docker** and deployed on **AWS EC2**; enabled internal testing across diverse inputs.
- Developed and productionized a **PII masking solution** using **GLiNER** model (supports 40+ entity types), evolving from a **REST API** on **EC2** (PoC) to a scalable **AWS Lambda** service with **Docker packaging** and **role-based access control**.

Samsung R&D Institute India - Bangalore (SRI-B) *Bengaluru, India*

Lead Engineer, Machine Learning *Mar. 2023 - Aug. 2024*

- Led the development of an **automated Python-based Android profiling tool** to benchmark rendering uniformly across devices, identifying bottlenecks (scroll janks) and enhancing rendering performance through improvements in the Android framework source.
- Led the development of four **edge-based personalization LLM solutions**, deriving actionable insights to enhance user experience
- Fine-tuned the **FLAN-T5 LLM model** for our specific use-case, optimizing performance & achieving a ∼**92%** evaluation score.

Senior Software Engineer, Machine Learning *Mar. 2021 - Feb. 2023*

- Designed an **edge ML solution** to analyze phone usage data and detect boredom with ∼**80% accuracy**, enhancing user experience. Built an Android app for **real-time inference with under 50ms latency**, showcasing the model's effectiveness.
- Pioneered a **Federated Learning (FL)**-based solution to predict gender and demographic age, **enhancing privacy** for ∼**10k users**. Explored innovative distributed learning techniques and tested diverse FL algorithms across 20+ input and model configurations, laying the groundwork for future privacy-preserving AI advancements.

Software Engineer, Machine Learning *Jun. 2019 - Feb. 2021*

- Engineered a **privacy-preserving edge ML solution** for **Next App Recommendation**, published in **IEEE ICSC 2022**, by designing a memory-efficient model (**99% size reduction**) to minimize FL bandwidth costs. Trained and deployed the model in Java using **DL4J**, integrating it on Android edge devices with a User Trial (UT) app for **training and inference across 500+ devices**.

Technical Skills

Programming Languages: Python, C/C++, Java, Kotlin, SQL, Golang, Shell Script (Bash)
Frameworks, Libraries & Misc.: TensorFlow, PyTorch, PydanticAI, Scikit-learn, FastAPI, GitHub, DeepLearning4Java (DL4J)

Education

University of California, San Diego (UCSD) *La Jolla, CA*
Master of Science in Computer Science; GPA: **3.95** / **4.0** *Sep. 2024 - Present*

Indian Institute of Technology, Kanpur (IIT Kanpur) *Kanpur, India*
Bachelor of Technology in Computer Science and Engineering; GPA: **9.0** / **10.0** *Jul. 2015 - Jun. 2019*

Patents and Publications

Methods and Electronic Devices for Behavior Detection using Federated Learning *2023*
US 18/191403

- Novel methodology to identify new behavior trends across users and provide behavioral recommendations using Federated Learning

System and Method for Distributed Learning of Universal Vector Representations on Edge Devices *2023*
US 17/946349

- Novel framework for learning user behavior embeddings directly on edge devices using Federated Learning

Memory Efficient Federated Recommendation Model *2022*
IEEE 16th International Conference on Semantic Computing (ICSC)

- Proposed novel model framework for Federated Learning-based recommendation systems which uses a fixed-size encoding scheme for items and users, thus not affected by the number of items/users, hence memory efficient and well suited for large-scale applications.

Relevant Projects

Multi-task Learning with ToolkenGPT Framework | *Course Project, UCSD* *Sep. 2024 - Dec. 2024*

- Re-implemented ToolkenGPT, optimizing efficiency for smaller Llama models and testing multi-task performance.

Awards and Achievements

Samsung Excellence Award | *SRI-B* *2024*

- Recognized as **Star of the Quarter** for excellent contributions in projects

Key Talent Recognition Program Award | *SRI-B* *2023*

- **Globally recognized** for exemplary teamwork and significant contributions to projects and organizational goals