KSHITIJ ALWADHI

Mountain View, California
linkedin.com/in/kshitijalwadhi/

Email : kshitijalwadhi@gmail.com
Website : kshitijalwadhi.github.io/
Mobile : +1 765-694-5118

Education

❖ **Purdue University** — West Lafayette, IN
Master of Science in Computer Science (GPA: 4.0/4.0) — *08/2023 - 05/2025 (Expected)*

- **Coursework:** Computer Networks, Software Engineering (Compilers Level Program Testing), Information Security, Algorithm Design and Analysis, Distributed Systems, Compilers for GPUs†, Reasoning with LLMs†

❖ **Indian Institute of Technology, Delhi** — New Delhi, India
Bachelor of Technology in Electrical Engineering (GPA: 3.93/4.0) — *07/2019 - 05/2023*
Minor in Computer Science (GPA: 4.0/4.0): [Dean's List]

- **Publication:** A deep learning framework for the detection of tropical cyclones from satellite images. *IEEE GRSS*
- **Research Thesis:** Demographic Prediction from Satellite Imagery using Deep Learning
- **CS coursework:** Operating Systems, Natural Language Processing, Deep Learning for Computer Vision, System Design Practices, Computer Architecture, Information Retrieval and Web Search, Machine Learning, Convex Optimization for ML

Technical Skills

Programming	C, C++, Java, Python, Go, Javascript, Scala, Rego
Development	NodeJS, React, MySQL, MongoDB, FastAPI, gRPC, Neo4J, Akka, Kafka
DevOps	Docker, Kubernetes, Git, CI/CD (CircleCI), AWS, GCP, BigQuery, Prometheus, Grafana
ML	TensorFlow, Pytorch, OpenCV, LangChain, AutoML, Sagemaker, LangSmith

Experience

❖ **Deductive AI | MLE Internship** — Mountain View, CA
Topic: Low Latency Ingestion and Code Reasoning — *05/2024 - ongoing*

- Reduced latency of ingestion pipeline from hours to seconds using **Akka** streams parallelization and DBMS optimizations.
- Owning the E2E pipeline for ingestion, retrieval and reasoning of code and its interaction with telemetry data.
- Implemented an `agent-less` approach which outperforms SOTA (SWE-bench) and cuts down on LLM costs by 500%.

❖ **DevRev.ai | MLE Internship** — Bangalore, India
Topic: Developing a Retrieval Augmented Generation Chatbot — *05/2023 - 07/2023*

- Led the development of a retrieval augmented conversational agent RPC from scratch. Implementation fetches neighbors from a vector DB, data from **S3** bucket and memory from **Redis**; currently the **top selling feature** of DevRev.
- Implemented a custom *LangChain*-like solution for chaining LLM calls and added support for function calling. Created an RPC for converting natural language queries into API calls; integrated into a general purpose search bar.
- Managed the migration of a microservice from Golang to Python to enhance the development of ML serving pipelines.
- Established a comprehensive encoder benchmarking pipeline using **AWS SageMaker** for proprietary datasets.

❖ **DevRev.ai | SDE Internship** — Bangalore, India
Topic: Adding support for third party integrations to the platform — *05/2022 - 08/2022*

- Contributed to backend in **Golang** to support 2-way communication with other SaaS apps like Slack and GitHub.
- Exposed internal workflows through RPCs to enable users to write automations using OPA policy **Rego**.
- Worked on multiple integrations for Slack, leading to DevRev's **initial breakthrough** with their first customers.

❖ **Sharechat - Data Science | MLE Internship** — Remote (Part-time)
Topic: Rule based modelling of DL models for Ads CTR prediction — *12/2021 - 05/2022*

- Trained a DNN model feeding in a large number of continuous & categorical features for ads CTR prediction; achieved test AUC score of 0.76 on Sharechat's proprietary dataset (3.5% improvement over existing implementation).
- Formulated RuleNet for distilling rules based on features with historically consistent correlation into models prediction.
- Productionized model using **GCP** for serving predictions; setup **Airflow** job for regular re-training from **BigQuery**.

Project Highlights

- **Distributed Systems:** Implemented a linearizable sharded key/value storage system using Paxos for fault tolerance and scalability to support cross-group transactions while ensuring robustness against system failures and network partitions.
- **Program Analysis:** Developed a Valgrind tool for dynamic analysis, detecting data dependencies in C code. Also implemented an LLVM module for static analysis, identifying memory leaks in C programs.
- **Vulnerabilities and Attacks:** Investigated vulnerable C codes susceptible to stack-smashing attacks. Attacked diverse vulnerabilities including buffer overflow and DEP bypass while using GDB for debugging and analyzing memory locations.
- **Network Bandwidth Allocation:** Developed and implemented a linear programming solution for optimal bandwidth allocation in multi-stream video analytics (using YOLOv8) within edge computing.
- **Natural Language Inference:** Implemented Few-Shot Cross-lingual transfer learning approaches using Adapter modules and fine-tuned XLM-R models for transferring knowledge from high-resource languages to low-resource languages.
- **Severe Thunderstorm Prediction using DL:** Created a pipeline using Mask R-CNN for segmentation and wind speed filter to identify cyclone formation visual signatures in satellite imagery, enabling short-term prediction.