

# NISHCHAL MARUR

[nishchal-mn.com](http://nishchal-mn.com) | +1 (240) 438-1916 | [nmarur21@umd.edu](mailto:nmarur21@umd.edu) | [LinkedIn](#) | [Github](#) | Maryland, USA

## EXPERIENCE

### Machine Learning Engineer II, Entrupy Inc, Bangalore

Aug 2021 - Aug 2024

- Built end-to-end deep learning based luxury goods authentication system, achieving **96% TPR** at **5% FPR** and processing **10K+** items monthly, deployed using Ray Serve for distributed inference.
- Developed 3D document unwarping system trained entirely on synthetic data, achieving **0.84 SSIM** on real-world documents and improving OCR accuracy by **23%**, outperforming SOTA methods.
- Optimized on-device CoreML inference resulting in a **2x faster** workflow over manual capture, using quantization and dynamic overlays for real-time auto-capture.
- Improved macro fingerprinting pipeline for item tracking and return fraud using patch embedding similarity and histogram features, boosting TPR by **15%** while reducing false positives.

### AI Intern, Connyc, New York

Jun 2025 - Aug 2025

- Built a hybrid recommendation system using two-tower architecture with sentence transformers, achieving <40ms average latency for personalized event recommendations.
- Implemented multi-factor reranking with Redis caching, supporting **50+ RPS** with A/B testing framework to continuously improve CTR and integrated the system as MCP tool for LLM orchestrator.

### Software Development Intern, IBM, Bangalore

Jan 2021 - Jul 2021

- Optimized batch prediction pipelines for large file scoring in Kubernetes pods with GoLang concurrency and chunked downloads, reducing inference latency by **15ms** in IBM Watson Cloud deployments.
- Benchmarked TensorFlow, PyTorch, Scikit-Learn and ONNX runtimes to evaluate performance and migration feasibility for an internal architecture shift.

### Software Engineering Intern, SLK Software, Bangalore

May 2020 - Jul 2020

- Saved developers **10+ hours** per week on log retrieval and debugging efforts by building a centralized ELK Stack log aggregation system using Filebeat and Node.js APIs across **5+** components.

### Machine Learning Intern, PathPartner Technology, Bangalore

May 2019 - Jul 2019

- Developed real-time **CNN**-based gaze tracking system using transposed convolutions and Gaussian heatmaps, achieving a mean error rate of **1.3 px** across diverse lighting and occlusions.

## TECHNICAL SKILLS

**Languages:** Python, C++, SQL, Go, Scala, Node.js

**Frameworks & Tools:** Git, Docker, Kubernetes, AWS, Azure, Apache Spark, Kafka, Airflow, Ray Serve, Blender, Elasticsearch, Redis

**Machine Learning Tools:** Pytorch, Tensorflow, HuggingFace, Transformers, OpenCV, Open3D, Flask, FastAPI, Pandas, Keras, Scikit-Learn,

Langchain, LangGraph, WandB, MLFlow

## PROJECTS

### CAFBrain: Multimodal LLM Platform for Capital Area Food Bank (Top 3) | [Project Link](#)

- Built an **agentic RAG** system using Langchain, LangGraph, LLMs, OCR, and prompt engineering to generate grant proposals, blogs, and reports from PDFs, images, and videos, reducing content creation time from **hours to under a minute**.
- Indexed **1000+** documents into **FAISS** vector DB and implemented content querying and refinement with human-in-the-loop feedback, enabling transformation of multimodal inputs into ready-to-download files.

### FedMedVision: Privacy-Preserving Federated Learning Medical Platform | [Project Link](#)

- Developed a federated learning system simulating multi-hospital training on classified X-ray images across **4+** client nodes, improving global **F1 score** by **12-15%** by aggregating class-skewed updates without sharing raw data.
- Tracked global model performance across training rounds using **MLflow**, with automated evaluation pipelines, and used Docker to containerize and deploy client nodes in distributed hospital environments.

### Scalable DBaaS for Rideshare application | [Project Link](#)

- Built a fault-tolerant Database-as-a-Service on AWS EC2 using RabbitMQ RPC queues and a custom orchestrator to manage read/write routing and multi-node replication with auto-scaling and leader election for slave containers.

## EDUCATION

### Master of Science (M.S) in Machine Learning

Aug 2024 - May 2026

GPA: 3.7/4

University of Maryland, College Park

Courses: Large Language Models, Optimization, Multimodal Foundational Models, Computer Vision, Robotics, MLOps

### Bachelor of Technology (B.Tech) in Computer Science

Aug 2017 - May 2021

GPA: 3.6/4

PES University, Bangalore

Courses: Data Structures, Algorithms, Data Science, Data Analytics, Machine Learning, Operating Systems, Cloud Computing