NISHCHAL MARUR

+1 (240) 438-1916 | nmarur21@terpmail.umd.edu | nishchal-mn.com | LinkedIn | GitHub | College Park, MD

SUMMARY

Machine Learning Engineer with **4+ years** of experience specializing in **Production ML Systems** and **Multimodal Deep Learning**. Proven record in architecting end-to-end ML pipelines from synthetic data generation to scalable real-time inference, automating **95% of manual verification workflows** in production.

EXPERIENCE

AI Engineer Intern, Connyct Inc, New York, **Jun 2025 - Aug 2025**

- Built a Two-Tower hybrid recommendation system using **sentence transformers** and **Elasticsearch**, achieving **Recall@10** of **0.85** for personalized recommendations at **<40 ms** latency.
- Implemented multi-factor reranking with **Redis caching** and an A/B testing framework to continuously improve CTR, exposing it as an MCP tool for the LLM orchestrator.

Machine Learning Engineer II, Entrupy Inc, Bangalore, **Aug 2021 - Aug 2024**

- Built a deep learning authentication pipeline achieving **96% TPR** at **5% FPR** for **50K+** luxury items monthly, automating 95% of the verification volume, reducing manual expert reviews.
- Led the end-to-end R&D of a **3D document unwarping** system that outperformed SOTA methods, resulting in **0.84 SSIM** score and **23% increase** in OCR accuracy, trained exclusively on synthetic data.
- Optimized **on-device** CoreML inference resulting in a **2x faster** workflow over manual capture, using quantization and dynamic overlays for real-time auto-capture.
- Reduced cloud infrastructure costs by **40%** by implementing automated monitoring and alerts on DataCrunch to identify and shut down idle GPU resources.
- Designed **synthetic data** pipelines using Blender Python to simulate camera intrinsics, lighting conditions, and material textures, where real-world data was limited.

MLOps Intern, IBM, Bangalore, **Jan 2021 - Jul 2021**

- Reduced inference latency by **15ms** in IBM Watson Cloud deployments by optimizing batch prediction pipelines using Go concurrency and chunked downloads on Kubernetes.
- Benchmarked TensorFlow, PyTorch, and ONNX runtimes to evaluate performance trade-offs, and contributed to the design of a new internal architecture.

Software Engineering Intern, SLK Software, Bangalore, **May 2020 - Jul 2020**

- Saved developers **10+ hours weekly** in debugging time by building a centralized log aggregation system using **ELK Stack**, Filebeat, Node.js across **5+** distributed components.

TECHNICAL SKILLS

Languages: Python, C++, SQL, Go, Scala, JavaScript (Node.js).
ML Frameworks: PyTorch, TensorFlow, Keras, Scikit-Learn, Transformers, HuggingFace, LangChain, LangGraph, OpenCV, CLIP, LoRA, ONNX
MLOps & Cloud: AWS, Azure, Docker, Kubernetes, Ray, MLflow, WandB, Airflow, Triton, CI/CD
Data & Tools: Redis, Elasticsearch, FAISS, Pinecone, MongoDB, PostgreSQL, Apache Spark, Pandas

PROJECTS

CAFBrain: Multimodal LLM Platform for Capital Area Food Bank (Agentic RAG | LLM)

- Built a LangGraph-based **Agentic RAG** workflow handling 5000+ multimodal documents (PDFs, Videos) via FAISS, reducing grant proposal and report creation time from **hours to under a minute**.

Temporal Change Retrieval (Computer Vision | Multimodal)

- Achieved **64% Recall@10** on satellite imagery by adapting RemoteCLIP with LoRA, multi-scale frequency analysis, and difference attention mechanisms to handle hard-negative mining for vision-language alignment.

Scalable DBaaS for RideShare (Distributed Systems | Cloud Infra)

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

EDUCATION

Master of Science (M.S) in Machine Learning **Aug 2024 - May 2026**
University of Maryland, College Park *GPA: 3.8/4*
Courses: Deep Learning, Large Language Models, Multimodal Foundational Models, Computer Vision, Robotics, MLOps

Bachelor of Technology (B.Tech) in Computer Science **Aug 2017 - May 2021**
PES University, Bangalore *GPA: 3.6/4*
Courses: Data Science, Data Analytics, Data Structures, Algorithms, Machine Learning, Operating Systems, Cloud Computing