Nischit Hosamani

Singapore | phone on request | email on request | github.com/Nischit

EDUCATION

Nanyang Technological University, Singapore Aug 2026 - Dec 2027 (expected)

MSc in Artificial Intelligence. Semester 1 coursework (Aug 2026): Machine Learning, Deep Learning, Computer Vision, Introduction to AI, AI Ethics.

Indian Institute of Technology Indore, India Aug 2020 - May 2024

B.Tech in Computer Science and Engineering, Minor in Astronomy. GPA: 8.6/10. B.Tech Project: Medical Image Segmentation using Diffusion Models. Courses: Computer Vision, Reinforcement Learning, Machine Learning, Computational Intelligence, Optimization Algorithms and Techniques, Linear Algebra and Ordinary Differential Equation-I, Complex Analysis and Differential Equations-II, Data Structures and Algorithms, Design and Analysis of Algorithms

EXPERIENCE

Flipkart Jul 2024 - Jul 2026

Software Development Engineer - I *Bangalore, India*

- Built cross-org fee-adjustment platform across 11 PnP microservices using Java and Dropwizard, coordinating 5 teams to deliver INR 1B estimated annual impact and INR 72M realized bottom-line in first month.
- Migrated and stabilized 6 Apache Storm clusters processing Kafka streams into Aerospike DB, re-allocating topologies by criticality to increase free slots 47%, cut disk utilization 40%, and drop alerts 75% over one month.
- Implemented client-level rate limiting at network filter layer using Guava-based wrapper across 4 PnP services for Big Billion Days readiness, reducing network hogging 15% during peak sale traffic.
- Redesigned coupon application flow across 3 services solo, introducing decorator-chain design pattern for My Coupons API refactor and increasing apply-coupon click-through rate ~50%.
- Automated event scheduler end-to-end workflow on RabbitMQ, testing full pipeline on playground environment to verify message routing and processing reliability.

Athabasca University May 2023 - Aug 2023

MITACS Global Research Intern *Edmonton, Alberta, Canada*

PROJECTS

TrackX (browser-based hand tracking) Aug 2026

- Built real-time hand tracking in browser using JavaScript, MediaPipe HandLandmarker WASM runtime, canvas rendering, custom One Euro filter, Jest unit tests, Playwright end-to-end tests, Docker multi-stage build, nginx CSP headers, and Kubernetes deployment with HPA.
- Implemented per-landmark adaptive jitter filtering, two tracking-anchor modes, live FPS and latency telemetry HUD, targeting-reticle overlay, hand-skeleton rendering, camera enumeration, and offline use after first load.

Fused Online Softmax in CUDA Oct 2026

- Implemented three fp16 row-wise softmax kernels in CUDA C++ with PyTorch C++ extensions: 3-pass safe-softmax baseline, 2-pass online-softmax with warp shuffles, and shared-memory row-cached kernel, profiling all in Nsight Compute.
- Achieved 17% faster online-softmax kernel at S=16,384 moving 6.00 bytes/element vs 7.99 baseline; FlashAttention-inspired kernels reached 94% DRAM bandwidth peak and ran 1.9x faster than torch.softmax at S=4,096.

OmniRec (multi-agent LLM evaluation) Oct 2026

- Built multi-perspective attribution module of 6-agent framework: generated 733-episode fault-injected benchmark corpus with train/calibration/test splits, integrated Ollama, OpenRouter, Gemini providers behind cached request layer.
- Achieved 37.9% step-attribution accuracy on tau retail benchmark and 25.0% on magentic vs AgentRx baselines of 31.0% and 18.2%; LLM selector beat learned ranker on 9 traces to 1 (exact McNemar, p=0.0215).

MCP-Forge (Daytona HackSprint Singapore 2026) Aug 2026

PUBLICATIONS

- Aditya Dixit, Nischit Hosamani, Puneet Gupta, Ankur Garg. "VISIONARY: Novel Spatial-Spectral Attention Mechanism for Hyperspectral Image Denoising." WACV 2025 (submitted 2024); IEEE Xplore Digital Library, Apr 2025.

SKILLS

Languages: Python, C++, CUDA C++, Java, JavaScript, Bash/Shell, PHP, MATLAB

Backend & Systems: Java (Dropwizard, Guava), Microservices, Kafka, RabbitMQ, Aerospike, Redis, MySQL, Kubernetes, Docker, Istio, Maven, CI/CD, Git, Nginx

ML & CUDA: PyTorch, CUDA kernel programming, warp-level primitives, online softmax, FlashAttention mechanics, Nsight Compute profiling, MediaPipe, XGBoost, Monte Carlo simulation

Agentic & Evaluation: Multi-agent LLM systems, LLM-as-judge evaluation, REINFORCE, statistical hypothesis testing (McNemar, Wilson intervals), bootstrap confidence intervals, SPRT