

NISHANT SHARMA

+1-646-772-9142 · New York, NY · nishantsharma@nyu.edu

linkedin.com/in/nishantsh20/ · github.com/nishant-ai · nishant-ai.github.io/portfolio

SUMMARY

Graduate Computer Engineer (M.S., NYU Tandon, GPA 3.6) who designs distributed micro-services and MLOps pipelines on AWS and GCP. Scaled an 8-service MERN SaaS to **5K daily active users** (\approx 300K requests/day) with 99.9% monthly uptime, fine-tuned LLMs on TPUs for sub-20 ms inference, and ships production APIs in Python/Java. Seeking SWE or Applied-AI roles building reliable, user-centric systems.

TECHNICAL SKILLS

- **Languages:** Python, Java, C++, TypeScript, SQL
- **Web / API:** React, Node.js (Express), Django, Flask, REST, GraphQL
- **Cloud & DevOps:** AWS (EC2, S3, RDS, EKS), GCP (GKE, Cloud Build, BigQuery), Docker, Kubernetes, Terraform, Helm, Jenkins, GitHub Actions, Istio, Prometheus/Grafana
- **ML & Data:** PyTorch, TensorFlow, Hugging Face, Vertex AI, Transformers, Pandas, scikit-learn, Langchain, Tavily
- **Practices & Tools:** Microservices, CI/CD, Test-Driven Development, Secure SDLC, OAuth2/JWT, Jira, Slack, Git

EDUCATION

NEW YORK UNIVERSITY, TANDON SCHOOL OF ENGINEERING, NEW YORK, NY

Master of Science in Computer Engineering, GPA: 3.6 (Sep 2024 - Expected May 2026)

Relevant Coursework: Machine Learning, Deep Learning, MLOps, Internet Architecture & Protocols, Real Time Embedded Systems

DR. A.P.J. ABDUL KALAM TECHNICAL UNIVERSITY, DELHI, INDIA

Bachelor of Technology in Computer Science & Engineering, 1st Division with Honors (Aug 2020 - Jun 2024)

Leadership & Activities: Technical Head, CSI Student Chapter – Led a 15-member team to run full-stack, DevOps, and Python hackathons/workshops for 200+ students.

EXPERIENCE

Co-Founder & Lead Engineer | Ingelt Board | Dec 2022 - Jul 2024, India

- Architected an 8-service MERN platform (\approx 300 K requests/day) for 5 K DAU and 200 counselors, deployed on EKS via Helm and Terraform.
- Maintained \geq 99.9 % monthly uptime by implementing self-healing, SLO alerts, and blue/green rollouts (Prometheus + Grafana).
- Cut release cycle from 25 min to 7 min through GitHub Actions pipelines and IaC-driven canary tests.
- Optimized data layer with indexed MongoDB + Redis cache, reducing P95 API latency by 35 %.

Software Engineer Intern (Full-Stack + DevOps) | Macverin Technologies | Sep 2022 - Jul 2023, India

- Developed full-stack web apps (Django, Flask, React) and dashboards, improving load performance by 25%.
- Automated CI/CD workflows (Jenkins, GitHub Actions) with cloud infrastructure; increasing deployments.
- Applied secure authentication systems and conducted penetration testing, increasing security by 15%.

PROJECTS

BERT-Finetuned AGNews | github.com/nishant-ai/BERT-tuned-AGNews (2025)

- Fine-tuning BERT models over AGNews Dataset for improved text classification and news categorization. Developed state-of-the-art NLP techniques for better accuracy.

A Comprehensive Study on Generalization in AI | kronosai.vercel.app (2024)

- Engineered and evaluated BERT transformer, K-Means clustering, regression, and time-series analyses on e-commerce datasets to diagnose and propose solutions for model generalization challenges.

LEADERSHIP & TEACHING

Subject Matter Expert – Part Time | Chegg India | Dec 2022 - Jul 2024 Concurrent with Ingelt, India

- Provided 1,000+ detailed computer science solutions, achieving 97% positive ratings.