

Pranjul Gupta

Pittsburgh, PA | +1 (412) 689-2795 | pranjulg@andrew.cmu.edu
pranjulgupta.com | linkedin.com/in/pranjulgupta
github.com/Pranjulcr7 | huggingface.co/PranjulGupta

Summary

Software engineer with nearly five years at Disney+ Hotstar and HashedIn, spanning distributed systems, personalization, and ML infrastructure. CMU AI engineering master's candidate; available January 2027.

Education

Carnegie Mellon University

Aug 2025 - Dec 2026 (expected)

MS in Artificial Intelligence Engineering - Information Security

GPA: 3.74/4.0

Coursework: Deep Learning; Machine Learning with Adversaries; Systems and Tool Chains for AI;

Deep Learning Systems and Generative AI (in progress).

M. S. Ramaiah University of Applied Sciences

Aug 2016 - Aug 2020

BTech in Computer Science and Engineering

CGPA: 8.26/10

Coursework: Distributed Systems; Data Structures and Algorithms; Database Systems.

Professional Experience

Disney+ Hotstar | Software Engineer II

Aug 2021 - Aug 2025

- Owned Continue Watching end to end across Go, Kafka, ScyllaDB, and Protobuf; cut post-match traffic 72.5% (8M to 2.2M requests/min) through caching and event-driven user-state processing.
- Helped build HotstarX's shared backend-for-frontend and widget architecture using Java/Spring Boot and federated GraphQL; supported mobile, web, and TV launches across a platform spanning 70+ countries.
- Developed XGBoost-based tray ranking with beam search and impression discounting; scored 200+ candidate trays per session through Go microservices at 50 ms p99 latency.
- Built Kafka feature pipelines processing 500K events/second and an ALS/PySpark recommendation engine that increased average watch time 1.18% and entertainment video views 1.12%.
- Improved live-event resilience with Kubernetes autoscaling, Akamai caching, fallback pages, load testing, and on-call support; the platform reached a then-record 59M concurrent viewers in 2023.
- Led feature delivery and A/B experiments; mentored interns and junior engineers.

HashedIn by Deloitte | Software Engineer

Nov 2020 - Aug 2021

- Built Python data pipelines for Eli Lilly with AWS Lambda, S3, and Aurora, processing 10TB+ daily and reducing end-to-end pipeline latency 40%.

Research and Projects

WAVLab / ESPnet - Carnegie Mellon University

Summer 2026

Research Collaborator

ESPnet PR #6491 | Model and evaluation

- Contributed a merged Spanish speech recognition recipe for the 105-speaker HEROICO corpus; trained PyTorch Conformer models on PSC Bridges-2 with Slurm and resolved training-instability failures.
- Identified shared-prompt data leakage, reduced train/test text overlap from 89% to under 6%, and published a corrected model and evaluation on Hugging Face.

Code-LLM Fine-Tuning

CMU

- Fine-tuned Llama-3-8B with QLoRA on curated NumPy vectorization patterns to generate optimized Python code.

DeepShelf - Computer Vision and Depth Estimation

CMU, 2026

- Co-developed a 100-image retail benchmark with 10K+ product boxes and 1.3K+ gaps; evaluated detector/depth-fusion models and exposed ranking reversals between synthetic and real data.

Technical Skills

Languages: Python, Go, Java, SQL, C/C++ | ML: PyTorch, TensorFlow, XGBoost, PySpark, QLoRA

Systems: Kafka, GraphQL, gRPC, Protobuf, Spring Boot, Redis, ScyllaDB, DynamoDB, PostgreSQL

Infrastructure: AWS, Kubernetes, Docker, Slurm, Linux, Akamai CDN, Grafana, Honeycomb, CI/CD

Leadership and Recognition

GSA President for INI, CMU; Teaching Assistant, Applied Machine Learning and Engineering Data-Intensive Scalable Systems. Hotstar Excellence Award (2024); AI mascot hackathon, first runner-up (2023).