

VEDAANG CHOPRA

vedaangchopra1009@gmail.com — [+14047409905](tel:+14047409905) — Georgia Institute of Technology, Atlanta, GA
linkedin.com/in/vedaang-chopra/ — github.com/Vedaang-Chopra

TECHNICAL SKILLS

- Data & Pre-Modelling:** Python (NumPy, Pandas), OpenCV, Feature Engineering, Data Augmentation, Plotly/Matplotlib, Vector Databases (FAISS, Redis), Elasticsearch
- Representation & Modelling:** PyTorch, TensorFlow, Hugging Face Transformers/Diffusers, LangChain, LangGraph, PyTorch Geometric, scikit-learn, CNNs, RNNs/LSTMs, Transformers, GNNs, RL, Generative & Agentic AI
- Optimization & Efficiency:** Quantization, Pruning, Knowledge Distillation, LoRA/QLoRA, ONNX Runtime, OpenVINO, Mixed Precision, Distributed Training (PyTorch Distributed)
- Inference & Serving:** vLLM, MLflow, FastAPI, Docker, Kubernetes, REST APIs, Latency Profiling, Model Monitoring
- Systems & Infrastructure:** Go, C/C++, PostgreSQL, Redis, RabbitMQ, Airflow, Bash, Git, Linux, High-Performance Computing (Slurm, tmux), Edge/Cloud Deployment

EXPERIENCE

Fortinet Technologies Inc. (AIOps, R&D)

Software Development Engineer I & II (ML/AI)

Bengaluru, India

Feb. 2021 – Jul. 2025

- Led development of an **Agentic RAG chatbot** (LLMs + tool calling) — a **Fortinet Hackathon 2023 (5th place)** prototype that evolved into a production feature, reducing issue-resolution time by over 70%.
- Built an **unsupervised connectivity-threshold model (PCT/172022/958026)** to detect wireless anomalies, cutting manual troubleshooting efforts by more than 75%.
- Implemented **DBSCAN-based anomaly detection** for SD-WAN telemetry, proactively preventing over 50% of potential network outages.
- Re-architected **OpenSearch ingestion pipelines** using async I/O and Golang, scaling event throughput from under 50 to over 2000 events per second.
- Deployed **edge-optimized ML models** via ONNX Runtime, lowering inference latency by 40% across distributed FortiGate devices.
- Automated **multi-domain data collection and SLA forecasting** for 60+ ML classifiers and 4 SLA categories, enabling 7-day AI-based performance prediction.

EDUCATION

Georgia Institute of Technology

M.S. in Computer Science (Machine Learning Specialization), GPA: 4.0/4.0

Atlanta, GA

Aug. 2024 – Dec 2026

Maharaja Surajmal Institute of Technology (GGSIPU)

B.Tech. in Information Technology, CGPA: 8.8/10.0

New Delhi, India

Aug. 2016 – Aug. 2020

RESEARCH & OTHER TECHNICAL PROJECTS

Which-VLM Router

— CS 8803 Systems for AI (Advisor: Dr. Anand Iyer) Aug. 2025 – Present

- Designing a **semantic router** that dynamically directs multimodal (text, vision, audio) queries across multiple VLM/LLM endpoints using **budget-aware** and **retrieval-augmented** policies; expected to achieve **30% lower inference cost** with comparable accuracy to Mixture-of-Experts baselines such as *FrugalGPT* and *METIS*.

Edge Glass Assistant

— CS 8803 VLM & LLM (Advisors: Dr. Zsolt Kira, Dr. Alan Ritter) Aug. 2025 – Present

- Building a **lightweight multimodal assistant** aligning vision, audio, and text embeddings through frozen encoders and quantized projectors for **on-device reasoning**; aims to deliver **2× faster inference** with minimal accuracy degradation compared to existing edge-ready frameworks like *PaLM-E* and *MovieChat*.

ATHENA

— CS 8903 Agentic AI (Advisor: Dr. Vijay Madisetti) Jan. 2025 – May 2025

- Developed **ATHENA**, a multi-agent generative AI framework unifying LLMs, diffusion models, and memory orchestration for automated screenplay-to-video generation, outperforming prior single-agent baselines (**BLEU-4 +18%, CLIPScore +22%**).

Ontology-based Text Classification

— Undergraduate Research (Advisor: Dr. Sonika Malik) 2019 – 2020

- Proposed a **semantic ontology framework** using the **Human Disease Ontology (DOID)** for biomedical text classification, improving accuracy by **10%** over classical ML baselines and published in **CEUR Workshop Proceedings (Vol. 2786)**.