

# Vedaang Chopra

Student (present), AI/ML Engineer (past), based in Atlanta (GA), USA.

[+1 404-740-9905](#) · [vedaangchopra1009@gmail.com](mailto:vedaangchopra1009@gmail.com) · [linkedin.com/vedaang-chopra](https://linkedin.com/vedaang-chopra)

## EDUCATION

<b>Masters in Computer Science</b> , Machine Learning (specialization)	Aug 2024-Dec 2026
Georgia Institute of Technology (Atlanta, GA)	GPA: - 4.0/4.0
<b>Bachelor of Technology (B.Tech)</b> , Information Technology (specialization)	2016-2020
Maharaja Surajmal Institute of Technology (GGSIPU) (New Delhi, India)	CGPA:- 3.52/4.0

## TECHNICAL SKILLSET

Pre Modelling:	NumPy, Pandas, Scikit-learn, Plotly, Matplotlib, Seaborn, OpenCV, NLTK, SpaCy
Modelling:	PyTorch, TensorFlow, Hugging Face (Transformers / Diffusers), LangChain, LangGraph,
Inference:	MLflow, ONNX Runtime, OpenVINO, Weights & Biases, Model Quantization & Optimization
Systems:	Python, Go, SQL, Bash, Docker, ElasticSearch, Azure, ML Security, Pentesting, RabbitMQ, Multiprocessing, Git, Distributed Training

## RESEARCH EXPERIENCE

<b>Which VLM</b> , CS 8803 SAI (Dr. Anand Iyer)	Aug 2025- Present
Designing an <b>adaptive semantic router</b> that directs multimodal (text, image, audio) queries to the most efficient VLM/LLM using budget-aware and retrieval-augmented routing; expected to cut inference cost by ~30%.	
<b>Edge Glass Assistant</b> , CS 8803 VLM and LLM (Dr. Zsolt Kira and Dr. Alan Ritter)	Aug 2025- Present
Developing a <b>lightweight, privacy-preserving multimodal assistant</b> using frozen encoders and quantized projectors for on-device inference; targeting 2x faster runtime with minimal accuracy loss	
<b>Connectivity ML Threshold</b> , Fortinet Research; ( <a href="#">PCT/172022/958026</a> )	Jun 2021- Nov 2022
Co-authored a <b>patent</b> on modeling optimal wireless connectivity thresholds across enterprise devices using unsupervised ML, <b>reducing manual troubleshooting work by over 75%</b> .	
<b>Ontology-based Text Classification</b> , Undergrad Research (Dr. Sonika Malik)	Aug 2019- May 2020
Developed an ontology-driven text classification model using DOID, improving accuracy by <b>10%</b> and F1 by <b>6%</b> over classical ML methods; published in <b>CEUR Workshop Proceedings (Vol. 2786)</b>	

## WORK EXPERIENCE

<b>Fortinet (AIOps, R&amp;D)</b> , Bengaluru	
<b>Software Development Engineer 2</b>	Feb 2024 - July 2025
● Built an <b>Agentic RAG-based enterprise diagnostics chatbot</b> (top-5 at Fortinet Hackathon 2023) integrating LLMs, tool-calling, and telemetry signals to reduce network issue-resolution time by ~70%.	
● Deployed edge-optimized ML models via ONNX, lowering inference latency by 40% across distributed FortiGate devices	
<b>Software Development Engineer 1</b>	Feb 2021 - Feb 2024
● <b>Redesigned OpenSearch ingestion architecture</b> , scaling throughput from ~50 to over <b>2,000 events/sec</b> for production ML workloads.	
● Implemented <b>unsupervised SD-WAN anomaly detection</b> using DBSCAN, preventing over 50% of potential outages across enterprise networks.	

## PROJECTS

[github.com/Vedaang-Chopra](https://github.com/Vedaang-Chopra)

<a href="#">nlp-work</a> ; <a href="#">cv-work</a>	Author	A collection of projects, learning material related to <b>CV and NLP</b> .
<a href="#">neural-malware-shield</a>	Co-Author	A comparison study of <b>DL architectures</b> (CNN, GNN, RNN, etc.) on Malnet.
<a href="#">ATHENA</a>	Author	A modular <b>multi-agent AI</b> framework combining for content creation(CS 8903)