Yashvardhan Gupta

+14085910621 ◇ gupta.yashv@northeastern.edu ◇ San Jose, California, United States ◇ LinkedIn ◇ BrutalCaeser

SUMMARY

I specialize in robot learning, developing diffusion-based policies, VLM-driven controllers, and reinforcement learning agents for embodied tasks. I have experience working with physical robotic hardware as well as simulation platforms for large-scale data generation. I focus on building robust training pipelines and model architectures that translate to robotic systems.

SKILLS

Programming Languages: Python, TypeScript, Dart, C, Bash
Audio & Real Time : Audio & Real-time: WebRTC, gRPC, WebSockets, AEC, VAD, sampling/latency/buffering
ML & Inference: PyTorch, CUDA, FAISS, ONNX, TensorRT, CI/CD, torchaudio, ONNX Runtime
NLP : ASR (wav2vec2), TTS, speech enhancement, denoising networks
Systems & Devops : Docker, FastAPI, Flask, AWS, CI/CD, multithreading/asyncio

EDUCATION

MS in Artificial Intelligence, Northeastern University — Sep '25 — Present, San Jose, United States

- **Current** : Foundations of AI, Algorithms
  Future : Machine Learning, Deep Learning, Human Computer Interaction, Pattern Recognition

B.Tech in Mechanical Engineering, Delhi Technological University (GPA: 7.84 (out of 10)) — Aug '19 — Jul '23

EXPERIENCE

Digital Solutions & Technology Engineer — Apr '23 — Jan '25
BIOWOLK HEALTHCARE

- Built and deployed Python microservices (Docker, FastAPI) for automated sales & inventory analytics; integrated model inference endpoints and scheduling pipelines that saved ~7 hrs/week for operations.
- Designed monitoring and logging for data pipelines (latency, failure alerts), improving pipeline reliability.

Machine Learning Research Intern — Dec '21 — May '22
TVISHTRYON SOLUTIONS PVT. LTD.

- Built a Virtual Teacher MVP with Flutter front end and Python back end enabling interactive lessons and dynamic content; integrated real-time video/audio playback and low-latency chat/response flow.
- Implemented server pipelines for on-the-fly media generation and deployment of pre-trained models for inference.

PROJECTS

REAL TIME SPEECH, Northeastern University Link — Sep '25 — Oct '25

- Built a low-latency demo using WebRTC (browser → Python server) for full-duplex audio.
- Server runs VAD + small speech-enhancement model exported to ONNX Runtime; audio forwarded to a lightweight ASR (wav2vec2) returning transcripts with ~200–300 ms processing latency (measured).
- Implemented monitoring: per-stream latency, packet loss, and SNR dashboard; containerized with Docker.

AI POWERED ADAPTIVE LED MATRIX, DTU Link — Jul '22 — Aug '23

- Curated and labelled a dataset of 500+ night-driving images with LabelImg (80/20 train/test) and fine-tuned TensorFlow's SSD Inception V2, achieving ~95% accuracy in detecting car lights under low-light conditions.
- Built a Python pipeline that ingests webcam video, runs TensorFlow inference to extract up to two vehicle X-coordinates, encodes detection states into a 3-char posX string, and sends it via serial to an Arduino.
- Developed Arduino firmware to drive a 4×8 LED high-beam matrix, dynamically switching off columns aligned with detected vehicles-ensuring glare-free illumination and boosting nighttime road safety.

AI LAWYER Link — Mar '25 — Jun '25

- Google Gemini embeddings + FAISS for semantic retrieval to surface case-relevant documents.
- Built a secure document-vault microservice with versioning, end-to-end encryption, column-level access controls, RBAC and audit logging to protect and manage legal drafts.
- Optimized retrieval & UX (chunking, batching, hybrid search) to cut search latency by 50%, boost answer relevance by 30%, and power a "super" chat generating structured arguments & counter-arguments accelerating legal research.

PUBLICATIONS

VISION LANGUAGE ACTION MODELS : A COMPLETE SURVEY
Currently Working
This survey reviews modern Vision–Language–Action models that integrate visual perception, language grounding, and action generation for robotics. It analyzes leading architectures from research labs and industry. The paper identifies strengths, gaps, and opportunities for building next-generation embodied AI systems.

CERTIFICATIONS & COURSES

Machine Learning Specialization - Stanford University (Coursera), Coursera — Dec '24
TensorFlow: Advanced Techniques Specialization (Coursera), Coursera — Aug '21