# Harsh Sharma

hsharm68@asu.edu | +1 623 285 5824 | LinkedIn | Github

## Education

**Arizona State University** 08/2025 – 05/2027
Master of Science in Computer Science Tempe, AZ
GPA: **3.89/4.00** — Coursework: Machine Learning, Artificial Intelligence, Data Mining, Deep Learning, Knowledge Representation

**Netaji Subhas University of Technology (NSUT)** 11/2021 – 05/2025
Bachelor of Engineering in Electronics and Communication Engineering New Delhi, India
GPA: **8.32/10.0** — Coursework: Data Structures, Probability, Statistics, Machine Learning, Digital Signal Processing

## Technical Skills

- **Programming:** Python, C++, SQL, Bash, Git, Linux
- **ML & AI:** Supervised/Unsupervised Learning, Deep Learning, Representation Learning, Time-Series Modeling, NLP, Computer Vision
- **LLMs & Retrieval:** Retrieval-Augmented Generation (RAG), Agent Architectures, Vector Search, Prompt Orchestration, Embeddings, Evidence Scoring
- **Backend & Systems:** FastAPI, REST APIs, Service-Oriented Design, State Management, Request Validation (Pydantic)
- **Data & Infra:** ChromaDB, GitHub API, PubMed API, Docker, CI Pipelines (GitHub Actions)
- **Frameworks:** PyTorch, PyTorch Geometric, SentenceTransformers, Transformers, scikit-learn, NumPy, Pandas

## Research and Professional Experience

**Sysmat Research Solutions Pvt. Ltd.** — Remote 06/2024 – 08/2024
*Machine Learning Engineer Intern – ML Systems & Signal Pipelines*

- Built modular Python pipelines for large-scale EEG preprocessing, including filtering, artifact removal, normalization, and FFT-based spectral feature extraction across 64-channel signals.
- Integrated spatio-temporal GCN–GRU architectures into reusable training and evaluation workflows for multi-class EEG classification.
- Designed dynamic neural graph construction to generate adaptive adjacency matrices from EEG connectivity, improving validation accuracy by **9%**.
- Standardized experiment execution with structured logging and configuration-driven pipelines, reducing iteration and handoff time by **40%**.

**Bow Creek Financial** — Remote 06/2023 – 08/2023
*Machine Learning Engineer Intern – Time-Series & Anomaly Detection*

- Developed Python-based batch pipelines for financial time-series ingestion, windowed feature engineering, normalization, and inference.
- Integrated ARIMA and autoencoder-based models into unified evaluation workflows for rare-event anomaly detection.
- Implemented automated threshold tuning and window selection across model variants, improving rare-event recall by **22%**.
- Dockerized training and inference pipelines with reproducible artifacts for local and cloud-based execution.

## Selected Projects

### RepoScope — Repository-Aware Code Intelligence System

- Built a repository-aware Retrieval-Augmented Generation backend that indexes entire GitHub repositories for LLM-based code explanation and review.
- Designed deterministic semantic chunking and file-local context window expansion to enable precise cross-file retrieval under strict token budgets.
- Implemented pluggable embedding and LLM backends (OpenAI, Claude, Gemini, Hugging Face, local fallback) with explicit provider control (BYOK).
- Engineered a persistent, rebuildable ChromaDB indexing layer with bounded retrieval, deduplication, and production-oriented error handling.

### Evi-Agent — Agentic Evidence Retrieval System

- Implemented a policy-driven agent for scientific question answering with explicit STOP/FETCH decisions, bounded loops, and deterministic termination.
- Designed a three-tier hybrid vector memory system combining query caching (LRU + TTL), semantic chunk storage, and summary-only conversational memory.
- Built confidence-aware evidence scoring using semantic similarity, document recency, and cross-document diversity with stagnation detection.
- Added full observability with per-run logging of retrieval quality, reuse events, API usage, stop reasons, and evidence counts.