

# ARYAAN PESHOTON

Evanston, IL 773-681-2029 [aryaaneshoton2027@u.northwestern.edu](mailto:aryaaneshoton2027@u.northwestern.edu)  
[aryaaneshoton.com](http://aryaaneshoton.com) [github.com/AryaanPe](https://github.com/AryaanPe) [linkedin.com/in/aryaaneshoton](https://linkedin.com/in/aryaaneshoton)

## EDUCATION

---

**Master of Science in Artificial Intelligence**, Northwestern University GPA: 4.0/4.0 **Sep 2025 - Dec 2026**

*Relevant Coursework:* Deep Learning, Generative AI, Advanced Algorithms for ML, Representation Reasoning & Language, Agentic AI

**Bachelor of Technology in Artificial Intelligence**, NMIMS (MPSTME), Mumbai GPA: 3.52/4.0 **2021 - 2025**

*Relevant Coursework:* Intro to AI, Algorithms & Data Structures, Intelligent Systems, Computer Vision, NLP, Reinforcement Learning

## EXPERIENCE

---

**AI Intern, Mphasis.AI, New York, NY** *Financial Knowledge Graph & Reasoning System* **Jun 2026 - Aug 2026**

- Parsed FIBO and MISMO into a **24,000-node, 52,000-edge knowledge graph** in **Python**, joined by 48 cross-standard mappings.
- Built **multi-agent workflows** with routing vector, keyword, 5-signal search, and graph search into two multi-tool **ReAct agents**, with **zero hallucinated citations**, enforced by an automated OWL axiom and citation audit with validation checker.
- **Cut RAG retrieval 11x (29.5s to 2.7s per 20 queries)** by profiling the pipeline and making reranking configurable.
- Deployed it as an **MCP server** (auth, rate limiting, timeouts), with a deterministic rules engine validating against 969 value lists and a **RAG** layer over AWS S3, Google Drive, and SharePoint.

**Data Science Intern, Finarkein Analytics, Mumbai, India** *Bank Statement Processing* **Nov 2024 - Jul 2025**

- Built a 12-step **Python** OCR pipeline (line repair, artifact removal, contour-based cell detection) structuring scanned bank statements across hundreds of layouts, **cutting manual digitization 80%**.
- **Cut repeat-document OCR time 60%** with SIFT/ORB logo detection caching each bank layout template.
- Matched transaction counterparties at **90%+ accuracy across 50,000+ records** via tokenization, regex masking, and KDTree search aided by fine tuning a **RoBERTa** to collapse 7,000+ company name variants, used downstream to build a **RAG** system for natural-language querying.

**ML Intern, Asset Analytix, Mumbai, India** *Predictive Maintenance Platform* **May 2024 - Jul 2024**

- Built an embeddings-based **retrieval chatbot** over **10,000+ technician service records**, replacing an unsearchable legacy database while handling inconsistent per-technician shorthand so the same repair logged different ways still matched.
- Built a **Random Forest and XGBoost** pipeline to detect and predict equipment failure risk at **85% precision**.

**AI Intern, Max Life Insurance, Mumbai, India** *Clinical Imaging Pipeline* **May 2023 - Jul 2023**

- Extracted ECG waveforms from scanned printouts with adaptive thresholding and parallelized segmentation.
- Standardized chest X-rays across **5+ clinical sources** into a single training pipeline, then built a residual-connection **CNN** in **PyTorch** for X-ray pathology classification: **91% accuracy**, 86% clinical.

## RESEARCH

---

### Research Assistant, Northwestern FORGE Program *3D Behavioral Reconstruction* Jun 2026 - Aug 2026

- Built a markerless **3D pose-reconstruction pipeline** for freely-moving mice on a six-camera rig, lifting **DeepLabCut (ResNet-101)** 2D keypoints to 3D via **RANSAC multi-view triangulation: 2.4 mm precision**, 3 px reprojection error.
- Developed hybrid calibration (ChArUco intrinsics, bundle-adjusted extrinsics from a tracked 3D object) that **generalizes to new sessions without recalibration**, holding 2-3 mm across sessions.
- **Cut per-frame paw jitter 4-5x** with reprojection-gated RANSAC and signal-processing smoothing (Hampel and median outlier rejection, Savitzky-Golay filtering) while preserving head precision.

## PROJECTS

---

### Capsule Endoscopy Triage Cascade Inference Optimization, PyTorch [GitHub](#) 2024

- Cut inference cost with a **two-stage cascade**: a light gate screens every frame, only flagged ones reach the full head.
- Fused ResNet-50, DeiT, and MobileNetV3 through learned attention for the ten-class stage.
- Handled a **180x class imbalance** with focal loss and weighted sampling. **92.9% accuracy**.

### Real-Time Exam Proctoring System Flask, SocketIO, YOLOv5 [GitHub](#) 2025 - 2026

- Ran **five models concurrently at 30fps** behind a Flask + SocketIO server streaming a live annotated feed.
- YOLOv5 phone detection, a liveness CNN, identity matching, virtual-camera detection, and gaze prediction.

### Motion Transfer GAN PyTorch, Generative Vision [GitHub](#) 2026

- Implemented a First Order Motion Model from scratch in PyTorch rather than fine-tuning an existing one.
- Multi-scale appearance encoding, soft-heatmap keypoints, dense flow prediction, occlusion masking, PatchGAN.

### U.S. Patent Phrase-to-Phrase Matching DeBERTa, LoRA [GitHub](#) 2026

- Scored **0.8636** on the hidden test set with a LoRA-tuned DeBERTa cross-encoder.
- Found **97% of score variance** sat within anchor groups, then used grouped CV to eliminate leakage.
- Trained only **3.2% of a 435M-parameter model** (14M weights), 5-fold ensembled.

### BookTunes: Cross-Modal Mood Retrieval Multimodal Embeddings, FAISS [GitHub](#) 2026

- Built with two teammates for a Northwestern ML course: mapped audio and text into a **shared emotion embedding space**.
- Trained a ResNet for 512-D audio embeddings, reduced to 128-D via PCA for fast nearest-neighbour lookup.
- Embedded passages with Sentence-BERT; FAISS retrieves each track, librosa matches tempo at transitions.

## TECHNICAL SKILLS

---

**Languages:** Python, SQL

**ML / DL:** PyTorch, TensorFlow, scikit-learn, XGBoost, OpenCV, CNNs, Transformers, Reinforcement Learning

**LLMs & Agents:** RAG (ChromaDB, FAISS), ReAct agents, LangGraph, MCP, LangChain, LoRA / PEFT, knowledge graphs (RDF/OWL)

**Cloud & Deployment:** AWS (S3), Azure, Docker, FastAPI, Flask, SocketIO, REST APIs, Microservices, automated testing, Git