

Shreepoorna Purohit

Boston, MA | purohit.shr@northeastern.edu | +1 (857) 263-1664 | [linkedin.com/in/shreepoorna-dp-870737192](https://www.linkedin.com/in/shreepoorna-dp-870737192)

EDUCATION

Northeastern University

M.S. in Computer Science | CGPA: 4.00 / 4.00

Boston, Massachusetts

Sep. 2025 – Jun. 2027

- **Coursework:** Advanced Algorithms, NLP, Machine Learning (ML), Reinforcement Learning (RL)

PES University

B.Tech in Electronics and Communication Engineering

Bangalore, India

Aug. 2019 – Jun. 2023

- **Coursework:** Data Structures and Algorithms, DBMS, Operating Systems, Computer Networks, Image Processing

TECHNICAL SKILLS

Languages & Machine Learning: Python(**Strong**), C++(**Strong**), Go, C, NumPy, Matplotlib, PyTorch, TensorFlow, Keras, HuggingFace Transformers, Scikit-learn, NLP

Systems, Cloud & DevOps: Docker, Kubernetes, Microservices, gRPC, REST APIs, CI/CD, Linux, Bash, Distributed Systems, SQL, NoSQL, Git, GitHub, Prometheus, Grafana

EXPERIENCE

Software Developer (SDE-2)

Jun. 2023 – Jul. 2025

IBM

Bangalore, India

- Built and deployed Go-based backend microservices exposing REST APIs to manage the full lifecycle of distributed storage resources via Docker and Kubernetes across **15+** regions, and owned the delivery of **5+** high-priority features.
- Implemented a Go-based asynchronous background service to schedule and reconcile distributed resource lifecycle workflows, using idempotent processing, retries, and state validation to achieve **>99%** successful execution across re-runs.
- Devised scheduled cross-region data replication workflows in Go to synchronize distributed storage data, improving availability by **25%** as measured via Prometheus metrics and Grafana dashboards.
- Spearheaded resolution of recurring production reliability issues by leading 3 junior developers to root-cause request handling and lifecycle edge cases that caused APIs to return **500** errors, driving fixes and code reviews to stabilize system behavior.
- Strengthened release quality by adding targeted regression tests after diagnosing production bugs across multiple global regions, reducing recurring API and lifecycle failures by **25%** through improved validation and CI/CD checks.

SWE Intern

Jan. 2023 – Jun. 2023

IBM

Bangalore, India

- Built LSTM and ARIMA-based predictive models to forecast VMHost resource utilization, improving capacity usage by **15%** and reducing false-positive alerts by **25%**, with results visualized via Grafana dashboards for operational monitoring.

SWE Intern

May. 2022 – Aug. 2022

Pthinks Pvt. Ltd

Bangalore, India

- Engineered an AI-driven attendance system using facial and voice recognition, automating employee authentication via a multi-modal biometric pipeline and reducing manual attendance effort by **40%**.
- Implemented FaceNet-based face verification achieving **95%** accuracy on **10K+** biometric samples, combined with GMM-based voice authentication (MFCC, DFT) in a Linux-based web application with database-backed profile storage.

PROJECTS

News Sentiment Analysis for Predicting Next-Day Stock Movements

- Built a scalable NLP pipeline on **1.5M** news articles to predict next-day stock movements using the **FNSPID** dataset.
- Defined deterministic movement labels using a 0.2% return threshold, creating a stable and noise-aware classification target.
- Implemented baseline models using VADER and FinBERT polarity with Logistic Regression and Random Forest.
- Deployed a full-article text retrieval system with long-document processing using Transformers and **512-token** chunking.
- Developed contextual text models utilizing **768-d** mean-pooled FinBERT transformer embeddings via HuggingFace.
- Fine-tuned FinBERT on CUDA achieving **49.3% accuracy** and **32.1% Macro-F1**.

Probing Transformer Representations for Named Entity Recognition (NER)

- Layerwise probing of frozen BERT (**100M**) to analyze NER encoding across **13** transformer layers on **CoNLL-2003**.
- Aligned word-level NER labels to BERT subword tokens by HuggingFace tokenization for token-level supervision.
- Extracted **12K+ 768-d** contextual embeddings using PyTorch and trained Logistic Regression probes with **10-fold** CV.
- Achieved **97.35%** token-level accuracy at Layer 11 vs. **76.5%** majority baseline, indicating deep semantic encoding.
- Designed capitalization control probes showing early-layer feature encoding distinct from semantic representations.

Advanced Melanoma Skin Cancer Detection with Deep Learning

- Built a CNN with **26.3M** parameters, achieving **97%** accuracy across 9 melanoma classes.
- Optimized hyperparameters over 20 epochs, boosting feature learning by **15%**.
- Applied TensorFlow and Keras augmentations to address class imbalance, increasing minority class samples by **30%**.