Smyan Sengupta

518-428-0765 | sengupta.sm@northeastern.edu | Boston, MA | LinkedIn | GitHub | Website | Availability: July–December 2026

EDUCATION

Northeastern University — September 2023 – Present
Khoury College of Computer Sciences — *Boston, MA*
Bachelor of Science in Computer Science — *Expected May 2027*
- **GPA:** 3.86/4.00 | **Minor:** Mathematics | **Concentration:** Artificial Intelligence | **Honors:** Dean's List (all semesters)

EXPERIENCE

Machine Learning Engineering Co-op — July – December 2025
Pfizer — *Andover, MA*
- Engineered a spaCy-based Natural Language Processing (NLP) chat assistant with a custom Named Entity Recognition (NER) pipeline from scratch, providing automated insights on experimental data, process documentation, and model optimization strategies for workflows across **5 distinct manufacturing processes**
- Developed, automated, and deployed a data cleaning and predictive modeling pipeline using XGBoost, Random Forest, and SVR on **9+ time-series lab datasets** (**1,000+ datapoints each**) for an **$11 billion** polysaccharide analysis project
- Implemented SHAP explainability analysis and EconML causal inference to identify highest-impact factors across datasets, ensuring model interpretability and consistency
- Researched and implemented MLflow experiment tracking and Data Version Control (DVC) for reproducible ML workflows; presented proof-of-concept to modeling teams across **4 sites**

Teaching Assistant – Foundations of Data Science — September 2024 – April 2025
Khoury College of Computer Sciences — *Boston, MA*
- Facilitated **6+ office hours per week**, directed project meetings, and proctored labs/exams to assist students with assignments and aid them in understanding data analysis, linear algebra, statistics, and machine learning concepts
- Graded **90+ assignments per week**, providing useful feedback

PROJECTS

NewsFactChecker | *Python, JAX, NumPy, Scikit-learn, SentenceTransformer* — February – April 2025
- Designed a probabilistic NLP fact-checking model on **10K+ statements** from the LIAR dataset using a 6-class Bayesian softmax classifier, applying Hamiltonian Monte Carlo with JAX to sample **5,000 posterior weight estimates**
- Implemented semantic feature extraction using SentenceTransformer embeddings; benchmarked posterior-averaged credibility scores against logistic regression baselines using AUC and threshold-optimized F1

Guardrails: Atomic | *TypeScript, Next.js, Z3 Prover, MongoDB, OpenRouter* — October 2025
- Architected an AI-powered full-stack formal verification platform in Next.js/TypeScript that iteratively generates mathematically verified code from natural language inputs, **winning 1st place** at an AI Agent Hackathon
- Engineered YAML-to-Z3 conversion pipeline with automated counterexample generation for program verification

OpenLegislation | *React, TailwindCSS, OpenAI API* — October 2024
- Built a web application that simplifies congressional legislation for general audiences, enabling users to search, read, and understand active bills without legal expertise, **winning 1st place** at the HackHarvard 2024 on the Open Data Track
- Implemented semantic search using OpenAI embeddings to surface relevant bills from natural language queries, and integrated GPT-based summarization to translate legal language into plain English

LEADERSHIP

Founding Member and Vice Chair — September 2024 – Present
Northeastern Association for Computing Machinery — *Boston, MA*
- Directed the planning and execution of two large-scale 24-hour hackathons with **40+ attendees**, coordinating budgeting, sponsorships, judging, workshops, and cross-club collaborations across **6+ student organizations** and **3 local startups**
- Facilitated communications between student software engineers and local startups to develop impactful real-world products

Co-Founder and Vice President — May 2025 – Present
MedCS Lab — *Boston, MA*
- Spearheaded an initiative to establish an interdisciplinary undergraduate research group focused on the intersection of computer science, data science, and the life sciences, facilitating communication between members and advisors

Director of Operations — August 2025 – Present
Northeastern Claude Builder Club — *Boston, MA*
- Secured Anthropic as a sponsor for two large-scale 24-hour hackathons, negotiated Claude API credit allocations for winners
- Coordinated demos and workshops with Claude Student Ambassadors and student developers to drive adoption and showcase capabilities across the Northeastern developer community

SKILLS

Languages/Frameworks: Python, R, SQL, Java, C, JavaScript, TypeScript, React, Next.js, TailwindCSS
Libraries/Tools: Pandas, NumPy, spaCy, Scikit-learn, XGBoost, PyTorch, Keras, Matplotlib, DVC, MLflow, Git/GitHub