Diya Pandey

Bethlehem, PA diya.pandey@lehigh.edu (717) 461-3207 diya-pandey

Education

Lehigh University, Bethlehem, PA *May 2027*
BS in Computer Science, Minor in Applied Mathematics
Relevant Coursework: Mathematical Models & Simulations, Machine Learning, Data Mining, Probability Theory, Computer Architecture, Design and Analysis of Algorithms, Computing Ethics, Computer Vision

Technical & Analytical Skills

Programming: Python, Java, C++, C, Assembly, Linux, Javascript, Go, HPC/linux
Frameworks/Tools: Pandas, PyTorch, AWS S3, YOLO models, scikit-learn, Matplotlib, HuggingFace, TensorFlow, Keras, NumPy, LangChain, Reinforcement Learning, Computer Vision, LLM

Experience

Mechanistic Interpretability on Chemical Large Language Models *April 2026 – Present*
Math Researcher, Lehigh University

- Trained Top-K sparse autoencoders on 1.7M polymer embeddings (768-d CLLM from Poly4mer/SMI-TED) in PyTorch/CUDA, sweeping dictionary expansion factor ($\alpha \in \{4, 8, 16, 32\}$) and sparsity ($k \in \{80, 160, 500, 1000\}$) to extract interpretable features
- Developed multinomial logistic regression classifiers (BCEWithLogitsLoss) on SAE features to predict molar group; validated feature interpretability via zero-ablation, confirming specific features encode molar property information.
- First author — Poly4mer-SAE: A Controllable Polymer LLM for Understanding Molar Structures and Beyond. In preparation for submission.

Student Software Developer, Alexa Project Lead *Sep 2023 – Present*
Nano Human Interface Lab, Lehigh University

- Self-taught Python and built an AI-powered voice application (Amazon Alexa) delivering real-time procedural guidance on material science and using lab equipments working end-to-end via Amazon Developer Console
- Gathered functional requirements from scientists, conducted user testing and iterative improvement across beta and live releases; developed onboarding documentation and mentored new lab programmers

Undergraduate Researcher
Diffusion Models for Atmospheric Turbulence Super-Resolution *Summer 2026*

- Benchmarked five deep learning architectures — VAE, CNN U-Net, adversarial neural operator(adv-NO),DDPM, and DDIM for super-resolution from the Johns Hopkins Turbulence Database
- Extended the turbulence SR benchmark to include generative and diffusion models, showing that the adv-NO approaches the state-of-the-art with fewer training samples and diffusion models outperform others

Speckle Analysis of Actin Dynamics at the Leading Edge of Lamellipodium *Summer 2024*

- Developed ML models to classify single particle trajectories across 60+ excel sheets, improving classification accuracy and enabling scalable analysis for cellular physics
- Automated data processing; applied mean square displacement analysis; presented findings to research team in weekly colloquium, and presented poster at annual Research Symposium

Projects

Skillion Capstone Project *Jan 2026 - Dec 2026*

- Developed a facial and object recognization model end-to-end architectures YOLO, FaceNet, MTCNN
- Created, and tested datasets under varied real-world conditions (group/individual settings) to support product reliability and conducted iterative user testing
- Partnered with sponsors to define product requirements, delivering a functional MVP on a semester timeline

Leadership & Organizations

President *August 2024 – May 2026*
Girls Who Code(GWC), Lehigh University Chapter

- Recruited 50+ new members, spearheaded weekly Eboard meetings, allocated tasks among Exec, led 15+ events and cross-institution collaborations; Awarded Student Life and Leadership Award
- Secured $9,000+ in sponsorships across five departments for Harvard WeCode conference and increased participation from Lehigh by 80% as Tech Fellow for two consecutive years. Distinguished Tech Fellow awardee