Sourin Dey

sourin@email.sc.edu 3077613781 LinkedIn Web

Research & Work Experience

05/2025 – 08/2025 **Data Science Intern**
Hexagon Manufacturing Intelligence
- **Developed a Graph Retrieval-Augmented Generation (Graph-RAG) pipeline using Open Source and Azure Models**– Built a Q/A system to process and query 20 years of software manuals. Using LangChain framework, I converted document knowledge into a neo4j-based graph database to retrieve information that is contextually rich and useful enough to assist Applications Engineers.
- **Delivered scalable knowledge access** – Enabled engineers to efficiently search and extract insights of multi-fold difficulties (temporal, vague, reasoning) from extensive technical documentation.

05/2024 – 08/2024 **Data Science Intern**
Dow, Midland, Michigan
- **Developed a Graph Neural Network (GNN) framework** – Built a heterogeneous network graph for product recommendation using node classification. I ensured effective information flow in non-homogeneous graphs to improve recommendation accuracy.
- **Researched Variational Autoencoder (VAE) applications** – I explored generative modeling for novel product formulation based on graph structures. I also identified how large, dense graphs bias the latent space and hinder stable VAE training.

08/2021 – present **Graduate Research Assistant**
University of South Carolina
- **Unified Atomic Diffusion Transformers** Energy-Aligned latent fine-tuning toward a foundational model for generalizable generation and property prediction in molecules and materials: After pretraining a VAE, I fine-tune its encoder with a lightweight regression head to predict formation energies, aligning the latent space with the underlying energy landscape. This energy-aware alignment regularizes the latent manifold, guiding the diffusion model toward sampling thermodynamically stable and physically realistic structures.
- **Created a topology based mapping algorithm for polymorphic crystal structures** – Leveraged polyhedron connectivity to cluster materials across diverse space groups, advancing materials discovery by improving identification of structural similarities beyond symmetry-based methods.
- **Developed a variant of Atomistic Line Graph Neural Network (ALIGNN) model** by Δ-learning electronic structure of crystals to predict HSE eigenvalues, a key opto-electronic property. By leveraging inexpensive PBE calculations and orbital projections, I could build highly accurate ML model as surrogate for costly DFT calculation.

08/2019 – 12/2021 **Graduate Research Assistant**
Artificially Intelligent Manufacturing Center, University of Wyoming
- I automated the AI powered Laser-Induced Graphene Process(LIG) manufacturing using Bayesian Optimization. The automated system is generalized and can be deployed to manufacture other materials.

Projects

08/2021 – 12/2021 **Benchmarking Graph Neural Network(GNN) architectures for material property prediction.**
- Investigated the impact of node level featurization in rotationally equivariant and atomistic line graph neural networks for material property prediction.
- Conducted a benchmark study to analyze the impact of different message passing blocks to solve over-smoothing problems in graph neural networks.

01/2023 – 05/2023 **Benchmarking Deep Neural and Spiking Neural Network Performances in Edge devices and Neuromorphic devices**
- Becnhmarked DNN latency , throughput in Intel Movidius Neural Computing Stick edge device. This helps to identify and deploy efficient neural network models for real time applications.
- Converted object detection based CNN models to Spiking Neural Networks (SNN).

08/2022 – 12/2022 **Solar Cell based Chemical Synthesis Information Extraction by Fine-tuning Large Language Models Trained on Chemical Data**
- I have designed a NLP pipeline using language model Bio-BERT and Spacy that summarizes solar cell chemical synthesis information from a given paragraph.

Education

08/2021 – present **Ph.D. in Computer Science**
University of South Carolina
Selected Coursework: Computer Processing of Natural Language, Neuromorphic Computing, Data Mining.

08/2019 – 12/2021 **MS in Computer Science**
University of Wyoming
Selected Coursework: Intro to AI, Randomness in Computation, Deep Reinforcement Learning.

05/2014 – 05/2018 **B.Sc. in Electrical and Electronic Engineering**
Khulna, Bangladesh *Khulna University of Engineering & Technology*

Skills

- **Python** (PyTorch, Pytorch-Geometric,Tensorflow, Deep Graph Library, GenSim, SpaCy), Pydantic AI, LangChain, LangGraph, Neo4j.
- SQL, Github Copilot, Jupyter Notebook, VS Code, Github, High-Performance Computing.
- Algorithm & Data Structure coding in Leetcode

Publications

My Google Scholar profile

Extracurriculars

Instructor, Dept. of Chemistry & Biochemistry, University of South Carolina
Worked as an instructor for Python Programming Summer Camp Workshop