Md Halim Mondol

3930 Jackson St, Raleigh, NC 27607 | mmondol@ncsu.edu | +1-919-600-2438
LinkedIn | GitHub | Google Scholar | Website

RESEARCH INTERESTS

Agentic AI system development, LLM evaluation and benchmarking, AI model reliability, retrieval-augmented scientific knowledge systems, and generative AI for molecular representation learning.

EDUCATION

- **North Carolina State University**, Raleigh, NC
  – Ph.D. in Electrical Engineering (CGPA: 3.82/4.00) Aug 2022 – Jul 2027 (Expected)
  Advisor: Edgar Lobaton
  – M.S. in Electrical Engineering (CGPA: 3.80/4.00) Aug 2022 – May 2024
- **Rajshahi University of Engineering and Technology**, Bangladesh
  – B.Sc. in Electronics and Telecommunication Engineering (CGPA: 3.68/4.00) Jan 2016 – Mar 2021

TECHNICAL SKILLS

- **Machine Learning:** PyTorch, scikit-learn, HuggingFace Transformers, NumPy, Pandas; Variational Autoencoders, Transformer Architectures, Sequence Representation Learning, Model Training, Hyperparameter Tuning, Experiment Tracking, GPU/Distributed Training
- **LLM, Agentic AI, and Evaluation:** Agentic AI Workflows, Prompt Engineering, RAG, Embedding Models, Automated Report Generation, LLM-as-a-Judge Evaluation, Scientific Question Answering, AI Decision Support Systems
- **Molecular Machine Learning:** SMILES/SELFIES Modeling, Molecular Generation, Latent Space Exploration, Property Prediction, Novel Molecule Discovery, RDKit Validation and Featurization
- **Knowledge Graphs and Biomedical Data Systems:** Neo4j, Cypher, Biomedical Knowledge Graph Construction, Drug Repurposing Pipelines, Multi-Source Biomedical Data Integration, REST API-Based Data Enrichment, Entity Resolution, ChEMBL, PubChem, DrugBank, BindingDB, Open Targets, Therapeutic Target Database
- **Data and Screening Pipelines:** Dataset Cleaning and Canonicalization, Automated Filtering and Deduplication, Validity/Novelty Metrics, Candidate Ranking Workflows, Scientific RAG and Knowledge Systems
- **Software and MLOps:** Python, MATLAB, C/C++, Git/GitHub, Docker, Kubernetes, AWS, GCP, MLflow, JupyterHub, Linux, REST APIs, Streamlit

WORK EXPERIENCE

- **Software Engineering Intern**, *Collaboration Pharmaceuticals, Inc.*, Raleigh, NC *May 2026 – Present*
  – Built a drug repurposing pipeline integrating drug, disease, and target-protein data from TTD, FDA and WHO clinical trials, Open Targets Platform, PubChem, ChEMBL, BindingDB, and related biomedical databases.
  – Automated biomedical data enrichment, download, and verification workflows using REST APIs across multiple public drug discovery and cheminformatics resources.
  – Cleaned, curated, and processed heterogeneous biomedical datasets to construct a Neo4j-based knowledge graph for downstream AI/ML modeling.
  – Developed graph neural network models to prioritize approved-market drugs as potential repurposing candidates for target diseases.
  – Built Streamlit-based tool for model interaction, result visualization, and internal project review; maintained codebase using GitHub/version control.
- **Graduate Research Assistant**, ARoS Lab, NC State University, Raleigh, NC *May 2024 – May 2026*
  – Developed and trained multiple generative molecular models (LTVAE, TVAE, and TransVAE) for molecular representation learning and novel molecule discovery.
  – Built automated molecular data curation and analysis pipelines: SMILES cleaning, canonicalization, invalid

removal, RDKit validation, latent-space traversal, SA scoring, and PCA-based chemical space analysis.
– Designed evaluation framework measuring reconstruction accuracy, chemical validity, novelty, and downstream property prediction performance.
– Deployed molecular inference API (SMILES → embeddings → predictions) using FastAPI and Docker on AWS.

- **Graduate Teaching Assistant (Deep Learning)**, NC State University, Raleigh, NC *Jan 2024 – May 2024*
  – Supported 400+ students by administering course infrastructure: managed GitHub Classroom/Organization, maintained NSF-backed JupyterHub living in JetStream2 VM, assisted in assignment preparation, and developed an automated Discord chatbot for student support.

- **Electrification Research Intern**, ABB Corporate Research, Raleigh, NC *May 2023 – Aug 2023*
  – Prototyped measurement and data-collection systems; delivered tested hardware + software for real-time monitoring.
  – Designed circuits/PCBs and validated performance using lab instrumentation and DAQ workflows.

PROJECTS

- **Knowledge Graph and GNN-Based Drug Repurposing Platform** *Internship Project*
  – Integrated multi-source biomedical data into a Neo4j knowledge graph connecting drugs, diseases, proteins, and therapeutic targets.
  – Automated REST API workflows for data enrichment, download, validation, and cross-source verification of biomedical entities.
  – Developed graph neural network pipeline to predict approved drugs with repurposing potential for disease-specific use cases.
  – Built Streamlit interface for querying model outputs, reviewing candidate drugs, and supporting sponsor-facing project updates.

- **Physics-Aware Synthetic Data and Scientific Knowledge Hub** *NSF Funded Research*
  – Built a Python pipeline generating synthetic phosphorus experiment datasets from user parameters.
  – Automated LLM report generation (multi-page technical summaries with tables) via API integration.
  – Designed a QA system combining RAG retrieval and physics-based feature extraction for reliable scientific question answering.

- **Cloud Deployment of Molecular Generative Model (AWS AppRunner)** *Course Project*
  – Containerized trained molecular model using Docker and deployed on AWS AppRunner.
  – Implemented inference service: SMILES → latent vectors, predicted properties, and 2D molecular structure output.
  – Exposed model inference via REST API and interactive Streamlit dashboard for real-time molecular analysis.

- **LTVAE: Hybrid LSTM–Transformer VAE for Molecular Representation Learning** *NSF Funded Research*
  – Designed BiLSTM encoder + VAE latent space + Transformer decoder trained on ZINC15 SMILES.
  – Achieved ∼99% reconstruction accuracy and ∼98% chemical validity on test molecules.
  – Generated synthesizable novel dyes with low synthetic accessibility scores for VOC sensing applications.

- **Self-Powered Motor Health Monitoring and Data Acquisition System** *ABB Corporate Research*
  – Designed current-transformer power supply harvesting energy from motor primary winding.
  – Delivered ∼5V regulated output powering monitoring electronics (∼500 mW).
  – Integrated DAQ system for real-time RPM, voltage, current, and temperature monitoring.

PUBLICATIONS

- **M. H. Mondol**, Q. Wei, E. Lobaton, "LTVAE: Hybrid LSTM-Transformer VAE Architecture for Molecular SMILES Representation", *Revision 3 Submitted, Journal of Computational Chemistry (Wiley)*, 2026.

- E. Lobaton, et al., **M. H. Mondol**, "AI as a Constrained Guide: Measuring Knowledge Transfer in a Graduate Engineering Course", *Accepted, American Society for Engineering Education (ASEE)*, 2026.

- S. Jamalzadegan, A. Penumudy, B. Mativenga, **M. H. Mondol**, et al., "AI-Powered Colorimetric Sensing with LLMs", *AIChE Annual Meeting*, 2025.