Xianqi Deng

Linkedin | Google Scholar | xdeng5@albany.edu | +1-518-528-1496

Education

PhD Student, Computer Science, University at Albany – SUNY, Albany, NY 2024 – present
Research Area: *Natural Language Processing, Multi-modal Models, Generative Models, AI for Quantum Chemistry and Nanomaterial*
Bachelor of Computer Science and Engineering, North Minzu University, Yinchuan, China 2020 – 2024

Key Skills & Domain Expertise

Highly motivated and collaborative AI researcher with deep expertise in developing multi-modal and generative models, to solve complex challenges in scientific discovery.

- **Core Competencies:** Multi-modal AI Systems, Generative Models, Language Model Fine-Tuning, Machine Learning, Computer Vision, AI for Scientific Discovery
- **Programming:** Python, JavaScript, Java, C/C++, MATLAB, SQL, HTML/CSS
- **Tools:** Git/GitHub, Linux, Jupyter, SQL Database

Research Experience

Research Assistant, University at Albany – SUNY Advisor: Prof. Petko Bogdanov

Generative Models for Nanomaterial Design; Multi-modal AI 2024 – present

- Built a **BART-based masked generator** that learns **sequence–property** relations from short DNA strands, then **zero-shot** generates longer sequences with **target fluorescence**.
- Validated effectiveness: **47% reconstruction accuracy** on short sequences; generated longer sequences with up to **46%** and **40%** accuracy in **zero-shot longer sequence extrapolation**.
- Enhanced the model by developing a novel adaptive masking strategy to learn the **positional importance** of critical DNA bases and integrating a **length-generalization module** to improve extrapolation from short training examples.

Property-Isometric Variational Autoencoder for Sequence Design 2024 – 2025

- Co-developed a **geometry-preserving VAE** that integrates **GNN-based smoothing** and an **isometric regularizer** for controllable generation of biological sequences with complex properties.
- Validated the model's real-world utility through the design of de novo DNA for fluorescent nanoclusters, achieving up to a **16.1-fold enrichment** of rare NIR emitters in wet-lab synthesis.

Visiting Student, University of Massachusetts Amherst Advisor: Prof. Zhou Lin

Molecular Structures Derivation from IR Spectra with Seq2Seq Model 2023 – 2025

- Engineered **sequence-to-sequence models** to identify molecular structures from IR spectra, achieving **61.1% top-1 accuracy** and **>3.5×** prior benchmarks.
- Achieved **97.2% accuracy** in identifying **13 common functional groups** from spectral data alone.
- Extended to **mixtures molecular**, achieving **up to 89.4%** accuracy in correctly identifying at least one molecule from **two-component spectra**.

Publications

2025 French, E., **Deng, X.**, Chen, S., Ju, C., Cheng, X., Zhang, L., Liu, X., Guan, H., & Lin, Z. *Revolutionizing Spectroscopic Analysis Using Sequence-to-Sequence Models I: From Infrared Spectra to Molecular Structures*. Under revision at ***Journal of the American Chemical Society***. ChemRxiv: 2025-n4q84.

2025 Sadeghi, E., **Deng, X.**, Lin, I., Copp, S. M., & Bogdanov, P. *Property-Isometric Variational Autoencoders for Sequence Modeling and Design*. Submitted to ***ICASSP 2026***. arXiv: 2509.14287.

2024 Chen, S.#, Wang, Z.#, **Deng, X.#**, Shen, Y.#, Ju, C., Yi, J., Xiong, L., Ling, G., Alhmoud, D., Guan, H., & Lin, Z. *Integrating Graph Neural Networks and Many-Body Expansion Theory for Potential Energy Surfaces*. ***NeurIPS 2024 AI4Mat Workshop (Spotlight)***. arXiv: 2411.01578.