Xiaohan Kuang

Vanderbilt University, Data Science Institute, Nashville, TN
Email: xiaohan.kuang@vanderbilt.edu | Phone: (615) 593-7765 | Web: https://kuangxh9.github.io/

SUMMARY of ESPERTISE

- Development of multibody docking using **Generative AI**
- Proven skills in **software development**, **web application design**, and **object-oriented programming** in computational biology
- **Research experience** in structural biology, epitope prediction and peptide design

RESEARCH EXPERIENCE

Vanderbilt University Medical Center, Meiler Lab Oct 2023 – Present

- Developed a novel generative AI framework that integrates score-based diffusion models with equivariant graph neural networks to predict water or metal-binding sites in protein structures, outperforming state-of-the-art tools. Two manuscripts have resulted from this project: one accepted at NeurIPS MLSB 2024, and the other currently under review.
- Developed a computational approach to detect antigen-binding sites by representing antigen surface information as 2D images. Preliminary results indicate that convolutional neural networks (CNNs) can predict binding sites effectively. Part of this work has been published in a peer-reviewed journal.
- Developed an active learning framework integrating molecular dynamics simulations and machine learning to explore the relationship between peptide sequences and their fibril formation propensity. This manuscript is currently under review.

Vanderbilt University, Data Science Institute Fall 2023

Developing a Web Platform for Converting Academic Papers into Plain Language Summaries

- Developed an AI-powered web application to convert complex academic papers into easily understandable summaries, bridging the gap between scientific research and the general public.
- Implemented the platform using Streamlit, integrating user authentication and content customization features with Python and Deta to provide a seamless and engaging user experience.

EDUCATION

Master of Science, Data Science Expected: May 2025
Vanderbilt University, Nashville, TN

Bachelor of Science, Computer Science, Mathematics, Data Science May 2022
University of Wisconsin – Madison, Madison, WI

PUBLICATIONS

1. **Kuang, X.**, Su, Z., Liu, Y. (Lance), Lin, X., Spencer-Smith, J., Derr, T., Wu, Y., & Meiler, J. SuperWater: Predicting water molecule positions on protein structures by generative AI. *Preprint: https://www.biorxiv.org/content/10.1101/2024.11.18.624208v1* (2024)
2. Lin, X., Su, Z., Liu, Y., Liu, J., **Kuang, X.**, Cummings, P. T., Spencer-Smith, J., & Meiler, J. SuperMetal: A generative AI framework for rapid and precise metal ion location prediction in proteins. *Accepted at NeurIPS MLSB* (2024). *Poster preview: https://neurips.cc/virtual/2024/102527*
3. **Kuang, X.**, Jalali, S., Rahman, T., Michalowski, J., Sheng-Wong, C., Wong-Ekkabut, J., Su, Z., & Dias, C. L. Discovering new amyloid-like peptides using all-atom simulations and artificial

intelligence. *Preprint: https://www.biorxiv.org/content/10.1101/2025.02.12.637911v1* (2025)

4. Zhang, G., **Kuang, X.**, Zhang, Y., Liu, Y., Su, Z., Zhang, T., & Wu, Y. Machine-learning-based structural analysis of interactions between antibodies and antigens. *BioSystems* (2024)

TEACHING EXPERIENCE

Vanderbilt University

Teaching Assistant, Data Science Institute Sep 2024 – Present
Course: DS5220 – Principles of Programming and Simulation
Course: DS5640 – Machine Learning

Head Teaching Assistant, Data Science Minor Sep 2024 – Present
Course: DS 3262 – Applied Machine Learning

University of Wisconsin – Madison

Course Assistant, Department of Mathematics Jan 2022 – May 2022
Course: Math535 – Mathematical Methods in Data Science

PRESENTATIONS

NVIDIA GTC 2025 San Jose, CA | March 2025

Poster Presenter, P73524: SuperWater: Predicting Water Molecule Positions on Protein Structures by Generative AI

Vanderbilt University March 2024

Guest Speaker, DS 3891: Intro to Generative Artificial Intelligence Models

- Invited by Dr. Charreau Bell to deliver a lecture on the applications of diffusion models in the biological domain.
- Presented an overview of diffusion models, their mathematical foundations, and their applications in protein structure prediction and drug discovery.

AWARDS

- Graduate School Travel Grant (2025)
- Tuition Scholarship, Vanderbilt University (2024)
  Awarded for outstanding performance in the Data Science Master's program, including a 3.96 GPA, innovative programming projects, and collaborative research.
- 1st Place, Vanderbilt University Summer AI Showcase (2024)
- Data Science for Social Good Research Scholarship, Vanderbilt University (2024)
- Data Science Institute Award, Vanderbilt University (2023)

RESEARCH INTEREST

- Integrating advanced data science methodologies and AI-driven models, such as diffusion models and equivariant graph neural networks, with biological research and other interdisciplinary fields to improve current methods and develop practical applications.

SKILLS

- Programming Languages: Python, R, Java
- Data Analysis & Visualization: PyMOL, Matplotlib, Seaborn, Plotly, ggplot2
- Web Development: Streamlit, Flask, Django, HTML, CSS, JavaScript
- Machine Learning Frameworks: PyTorch, Keras, Scikit-learn, TensorFlow
- Tools & Technologies: Docker, Git, Jupyter Notebooks, VS Code
- Databases: MongoDB, MySQL, Cassandra, Redis, Neo4j