HAN MENG

𝔾 Google Scholar | miss.menghan@gmail.com | https://dearhannah.github.io

EDUCATION

The College of William & Mary, USA
Ph.D. in Applied Science (Data Science track) *Apr 2024 – Present*
Michigan State University, USA
M.S. in Computer Science and Engineering (GPA: 3.89/4.0) *Jan 2022 – Apr 2024*
M.S. in Materials Science and Engineering (GPA: 3.91/4.0) *Aug 2019 – Apr 2021*
University of Science and Technology Beijing, China
B.E. in Nanomaterials and Nanotechnology (GPA: 3.85/4.0) *Sep 2014 – Jun 2018*
National Scholarship (5/360 for 4 years); *Chinese Academy of Sciences Scholarship for Undergraduates* (2015)

SKILLS

Programming	Python, C, R, MATLAB, Bash, LaTeX
Data Science Toolkits	PyTorch, TensorFlow, PyG, RDKit, Scikit-learn, SciPy, Numpy, Pandas, Matplotlib, Seaborn
Science Tools	Molecular Dynamics (LAMMPS), DFT (Materials Studio), Chemical Synthesis

SELECTED RESEARCH EXPERIENCES

Data-Driven Decision Intelligence Lab College of William & Mary
Research Assistant (Advisor: Prof. Chen, Haipeng)

Actively conducting research related to Reinforcement Learning and Generative AI, particularly in Maximum Entropy RL, diffusion models, and LLM-based reasoning, with applications spanning medical and physical sciences.

Reinforcement Learning for Accelerator Control *Sep 2024 – Present*

- Collaborated with J-Lab experts to formulate the accelerator control task as a constraint satisfaction problem;
- Leading the exploration to apply reinforcement learning with diffusion policies for generating feasible control solutions;
- Investigating MaxEnt RL and diffusion models to improve solution diversity and progressively capture the full solution space.

Bin Chen Lab & **Illidan Lab** Michigan State University
Research Assistant (Advisor: Prof. Zhou, Jiayu & Prof. Chen, Bin)

My responsibilities included 1)maintaining the drug discovery machine learning code developed by lab members, running the initial virtual screen of potential drugs in the whole drug discovery cycle; 2) providing tech support to lab members with non-CS backgrounds on HPC operation, code running, etc. Meanwhile, I was actively involved in the following research projects.

Predictive Model for Drug-induced Gene Expression Jan 2022 - Apr 2024

- For the specific task, explored the potential of graph neural networks to learn better drug representation compared to ECFP.
- Researched meta-learning and pre-training approaches to address the label scarcity problem in emerging cell lines.
- Applied active learning to optimize training data selection and improve labeling efficiency. Developed algorithms that leverage gene expression profiles of landmark genes to identify anchor drugs. Models trained on anchor drug data outperformed those using knowledge-based and cluster-based selection strategies. **First author paper**[1] was accepted to **SDM 2025**.

LLM for Cancer Diagnosis Sep 2023 - Apr 2024

- Investigated few-shot prompting (BARD API) and fine-tuning (Falcon, an open-source LLM) methods for the cancer diagnosis task on the MIMIC-IV dataset. One collaborative paper[5] was accepted to **AMIA 2024 Informatics Summit**.

Generative Model for Drug Design Jan 2022 - Feb 2022

- Implemented several molecule optimization baselines ranging from genetic algorithm, reinforcement learning, and Markov sampling, and delivered results in ten days. A collaborative paper[4] was accepted to KDD Health Day 2022 as **Best Paper**.

SELECED PUBLICATIONS

[1] **H Meng***, R Chen*, J Zhou, B Chen. AnchorDrug: A Pipeline for Drug-induced Gene Expression Prediction in New Contexts through Active Learning.(SDM 2025)

[2] Y Li, **H Meng**, Z Bi, I Urnes, H Chen. Population Aware Diffusion for Time Series Generation.(AAAI 2025)

[3] J Hong*, W Zheng*, **H Meng**, S Liang, A Chen, H Dodge, J Zhou, Z Wang. A-CONECT: Designing AI-based Conversational Chatbot for Early Dementia Intervention. (ICLR 2024 Workshop on LLM Agents)

[4] M Sun, J Xing, **H Meng**, H Wang, B Chen, J Zhou. Molsearch: Search-based Multi-objective Molecular Generation and Property Optimization. (KDD 2022)

[5] X Zhang, S Vemulapalli, N Talukdar, S Ahn, J Wang, **H Meng**, etc. Large Language Models in Medical Term Classification and Unexpected Misalignment Between Response and Reasoning. (AMIA 2024)