Haote Li

Email : haote.li@yale.edu

Education

- **Yale University** — New Haven, CT
  PhD in Chemistry — *Sep. 2019 - Current*
- **University of Massachusetts Amherst** — Amherst, MA
  Master of Science in Physics — *Sep. 2018 - Aug. 2019*
- **University of Massachusetts Amherst** — Amherst, MA
  Bachelor of Science in Physics; GPA: 3.98/4.00 — *Sep. 2016 – May. 2018*
  - **Honor:**: Summa Cum Laude

Research & Internship

- **Boehringer Ingelheim** — Danbury, CT
  Internship Chemical Development — *June.2023 - August.2023*
  - **High-Throughput Investigation of Key Pharmaceutical Reaction**: Developing an automatic pipeline integrating GFN-xTB method for high-throughput catalyst screening. Resulting lead candidate was experimentally validated and showed desired high enantio-selectivity and yield.
- **Yale University** — New Haven, CT
  Theoretical Chemistry*; PhD Student; Mentor: Victor Batista* — *September.2019 - Current*
  - **Continizing Chemical Problems by Machine Learning**: LLM and Generative models for drug, antibody, and catalysis design; Site-specific retrosynthesis with sequence-to-sequence models. Gradient Descent algorithms for protein secondary structure content determination from unassigned NMR spectrum.
- **Umass Amherst** — Amherst, MA
  Biophysics*; Research Assistant; Mentor: Shuang Zhou* — *August. 2018 - August. 2019*
  - **Living System inside Liquid Crystal**: Investigating homeotropic alingment in lyotropic liquid crystals with etched surfaces. Experimental study of Microtubules motion patterns in the nematic phase liquid crystal.
- **Umass Amherst** — Amherst, MA
  Physical Chemistry*; Research Assistant; Mentor: Dhandapani Venkataraman* — *Dec. 2016 - Jan. 2018*
  - **Electron Transport on PEDOT**: Exploring electron transport mechanisms of polystyrene sulfonate (PEDOT). Optimizing polymerization process of PEDOT by studying the electrical and thermal conductivity under various physical and chemical treatments.
- **Tsinghua University** — Haidian, Beijing
  Material Engineering*; Research Student; Mentor: JinQuan Wei* — *June. 2015 - Aug. 2015*
  - **Double-junction SWCNT-Silicon interface solar cell**: Leveraged the synergy of single-walled carbon nanotubes and traditional monocrystalline silicon to engineered dual-PN-junction solar cells. Achieved new designs with enhanced energy conversion efficiency.

Manuscripts

- **Graph Protein Nuclear Magnetic Resonance Shifts Prediction (2020)**: Extracted both sequential and spatial features from protein data banks. Applied a combination of muti-adjacency-GNN and Transformer architecture for spatial-feature-based prediction; Predictions outperformed those made by the most commonly used BLAST-based algorithms by 26%.
- **Kernel-Elastic Autoencoder (KAE) (2023)**: KAE is a generative model which enhances the performance of traditional Variational Autoencoder by designing both modified maximum mean discrepancy and weighted reconstruction loss functions. KAE addresses the long-standing challenge of achieving superior generation and reconstruction performances at the same time. Applications include molecule generation, similarity-constraint property optimization, docking candidate searching and retrosynthesis planning.
- **Protein Secondary Structure Content Determination from 1D Spectrum (2022)**: Apply Gradient Descent to obtain the secondary structure content of the given protein sample using a 1D-Unassigned NMR C13 spectrum. The result rivals with the state of the art program that uses Circular Dichroism (CD). The method was the first of its kind and enabled secondary structure content studies in systems that are otherwise impossible by CD.
- **Site-Specific Retrosynthesis with Large Language Model (2023)**: Designed Transformer-based architecture which allows atom-specific reaction predictions. The results was benchmarked on USPTO-Full (1M) dataset and achieve state-of-the-art predictions site-specific reactions.
- **Monodentate Catalyst Design with KAE (2023)**: Integration of Monodentate ligand dataset for novel catalyst proposal. Conditional generation based on catalyst physical properties resulting in the new ligand MachinePhos A its class of Alkynyl-Phoshines.

Presentations

- **American Physical Society National Meeting 2023**: Rapid Protein Secondary Structure Determination from a Single Unassigned 1D 13C NMR Spectrum.
- **Yale Innovation Summit 2023**: Pitch: Solving Problems by Generation.
- **American Chemical Society National Meeting 2024**: Kernel-Elastic Autoencoder for Molecular Design

Awards

- **Jack Ragle Research Fellowship 2017**: Chemistry departmental award for academic excellence
- **Chancellor's Award 2016&2017**: Undergraduate Excellence Award
- **L'Oreal Product Development Intensive Award 2021**: Award to first place team for developing de novo sunscreen product formulae

Programming Languages

- **Python**: Proficient in using Python for exascale high-throughput computational projects. Building large language model based libraries for ML applications with packages such as PyTorch, Scikit-learn, RDKit, Numpy and Pandas.
- **Command-line-interface (Linux)**: SLURM system-relevant commands for high performance computing job queue managements.
- **Java**: Basic level Java. Fluent in reading and modifying codes. Built Java applications for Gauss-Jordan Elimination calculator, Gram-Schmidt process calculator for linear systems.

Teaching and Mentoring

- **General Chemistry Laboratory, 2019 Fall**: Undergraduate Laboratory
- **Comprehensive University Chemistry II, 2020 Spring**: Undergraduate Chemistry course
- **Introduction to Statistical Mechanics, 2020 Fall**: Graduate Statistical Mechanics course
- **Physical Chemistry With Applications II, 2021 Spring**: Undergraduate Physical Chemistry course
- **Introduction to Statistical Mechanics, 2021 Fall**: Graduate Statistical Mechanics course