Sijie Fu

sijiefu.github.io
fusijie@pku.edu.cn
linkedin.com/in/fusijie

Education

- **Ph.D.** in **Chemistry** | **Ph.D. minor** in **Machine Learning** *Aug'21 – Oct'25*
  Carnegie Mellon University, Pittsburgh, PA Advisor: **Dr. Newell Washburn**
  Thesis: ML for Chemistry and Formulation: Generative Discovery, Data-Driven Modeling, and Autonomous Experimentation
- **B.Eng.** in **Materials Science and Engineering** *Sep'16 – Jul'20*
  Peking University, Beijing, China Advisor: **Dr. Qiang Sun**
  Global Exchange at the University of Toronto, Toronto, Canada *Jan'18 – May'18*

Experience

1. R&D Data Scientist (PhD Intern), Procter & Gamble, Mason, OH *Jul'24 – Oct'24*
   a. Modeling formulation data to accelerate product development with automated ML modeling.
   b. Modeling large consumer data to provide business insight with automated analysis and ML modeling.

Publications

(*equal contribution)

1. Shen, Y.*; Zhang, C.*; **Fu, S.***; Zhou, C.; Washburn, N.; Póczos, B. Chemistry-Inspired Diffusion with Non-Differentiable Guidance. The Thirteenth International Conference on Learning Representations. *ICLR*, **2025**. [openreview][GitHub]
2. **Fu, S.**; Wang, D.; Henderiks, H.; Assis, A.; Charron, J.; Washburn, N. Autonomous Determination of Hansen Solubility Parameters via Active Learning. *Under review*, **2025**. [preprint][GitHub]
3. Roguski, M.*; **Fu, S.***; Walker, L.; Tilton, R.; Washburn, N.; Cochran, B.; Stone, C.; Barker, M.; Jamadagni, S.; Johnson, E. Predictive Modeling of Emulsion Stability via Hierarchical Machine Learning. *Under review*, **2025**.
4. **Fu, S.**; Bruno, C.; Chua, A.; Póczos, B.; Washburn, N. RheoNet: A Physics-Inspired Neural Network for Predicting the Viscosity of Complex Fluids. *Manuscript ready*, **2025**.
5. Qie, Y.; Wang, S.; **Fu, S.**; Xie, H.; Sun, Q.; Jena, P. Yttrium-Sodium Halides as Promising Solid-State Electrolytes with High Ionic Conductivity and Stability for Na-Ion Batteries. *J. Phys. Chem. Lett.* **2020**, *11* (9), 3376-3383. [link]

Selected Projects

1. **ChemGuide**: A Chemistry-Inspired Diffusion Model with Non-Differentiable Guidance
   a. Developed an equivariant diffusion model that integrates guidance from a non-differentiable quantum chemistry oracle.
   b. Implemented novel gradient estimation techniques using zeroth-order optimization and bilevel optimization.
   c. Outperformed existing methods in generating molecules with desired properties while enhancing stability.
2. **RheoNet**: A Physics-Inspired Neural Network for Predicting the Viscosity of Complex Fluids
   a. Constructed a benchmark dataset of 79 multicomponent polymeric fluids with measured viscosities from shear rate sweeps.
   b. Developed parametric symbolic regression w/ genetic algorithms to discover viscosity models. [`pip install parametric-sr`]
   c. Designed a hybrid neural network architecture that integrates constitutive models as a physics prior to predict viscosity.
3. **FormulationAI**: A Hierarchical Machine Learning (ML) Framework for Predicting Emulsion Stability
   a. Constructed a benchmark dataset of 256 emulsions from industry-relevant materials with measured stability.
   b. Developed a robust ML pipeline with various featurization methods and models for real-world formulation applications.
   c. Proposed a hierarchical ML framework that integrates domain knowledge to enhance model performance and interpretability.
4. **autoHSP**: An Autonomous Experimentation Framework for Determining Hansen Solubility Parameters via Active Learning
   a. Built a closed-loop, autonomous experimentation framework that integrates ML-driven decision-making with automation.
   b. Implemented a batch-mode active learning algorithm for experiment design and a computer vision pipeline for data analysis.
   c. Proposed a scalable and accessible blueprint for autonomous research and self-driving laboratories.

Selected Presentations & Posters

1. Closed-loop and Autonomous: CMU Cloud Lab for Measuring Hansen Solubility Parameters *June 27, 2024*
   at the Chemistry Department Summer Seminar (CMU), Pittsburgh, PA
2. Formulation AI for the Guided Design of Complex Fluids *Feb 12, 2024*
   at the annual Chemistry Department Poster Session (CMU), Pittsburgh, PA *washburnlab.chem.cmu.edu/sijie-poster-chem/*
3. Molecular Representation for Property Prediction: Wanders in the Chemical Space *Dec 2, 2022*
   at the Chemistry Department Graduate Seminar (CMU), Pittsburgh, PA

Professional Skills

1. Programming languages: Python (most experienced with), JavaScript; [basics] C/C++, SQL, Mathematica.
2. Libraries and tools: [ML] PyTorch, Scikit-learn; [computational chemistry] xTB, ORCA, Gaussian16.
3. Web/software development: Nginx, Flask, Streamlit, React (basics). *full-stack example: cavall.chem.cmu.edu/demo/equus/*