Sijie Fu

Email: fusijie@cmu.edu
Website: sijiefu.github.io
LinkedIn: linkedin.com/in/fusijie

EDUCATION

- **Ph.D.** in **Chemistry** | **Ph.D. minor** in **Machine Learning** *Aug'21 – Sep'25 (expected)*
  Carnegie Mellon University, Pittsburgh, US Advisor: **Dr. Newell Washburn**
  Thesis (proposed): ML for Chemistry and Formulation: Theory, Modeling, and Autonomous Research
- **B. Eng.** in **Materials Science & Engineering** *Sep'16 – Jul'20*
  Peking University, Beijing, China Advisor: **Dr. Qiang Sun**
- Global Undergraduate Exchange at the Faculty of Applied Science & Engineering *Jan'18 – May'18*
  University of Toronto, Toronto, Canada

EXPERIENCE

1. R&D Data Scientist, PhD Intern, at Procter & Gamble, Mason, OH *July'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. *Manuscript 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. *Manuscript under review*, 2025.
4. **Fu, S.**; Bruno, C.; Chua, A.; Póczos, B.; Washburn, N. Machine Learning for Predicting the Viscosity of Complex Fluids and Aiding Formulation Design. *In preparation*, 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. Machine Learning for Predicting the Viscosity of Complex Fluids and Aiding Formulation Design *May'23 - Present*
   a) Benchmark dataset development for complex fluids. b) Genetic algorithm for symbolic regression of viscosity models.
   c) Physics-inspired layer integrated with neural networks for more reliable and interpretable predictions.
   d) Molecular feature engineering for formulation design with unseen novel chemicals.
2. Autonomous Determination of Hansen Solubility Parameters (HSPs) in an Automated Lab *Jul'23 - Present*
   a) *AutoHSP* for autonomously measuring HSPs with on-the-fly experiment analysis and design. One click is all you need.
   b) Computer vision for result analysis: YOLO for object detection and edge detection for analyzing liquid interfaces.
   c) Batch-mode active learning for experiment design with optimal efficiency by explore+exploit strategies.
3. ML-Based Predictive Modeling of Emulsion Stability for Accelerated Formulation Design *Oct'23 - Present*
   a) Moderate-throughput experiment with automated image analysis for monitoring emulsion stability.
   b) A multitude of feature engineering techniques with quantum chemistry and experimental characterizations.
   c) Pipelined ML for feature selection, cross validation, hyperparameter tuning, model selection, and performance evaluation.

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 *washburnlab.chem.cmu.edu/sijie-poster-chem/*
   Feb 12, 2024, at the annual Chemistry Department Poster Session, CMU, Pittsburgh, PA
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++, Mathematica, SQL.
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/*