Seungmin (Henry) Lee

Computational Chemist | Chemical Process Engineer

South Bend, IN | 2327332@gmail.com | (608) 471-9085 | linkedin.com/in/slee2327332/ | github.com/henry-min98

PROFESSIONAL SUMMARY

M.S. Chemical Engineering researcher with deep expertise in building and deploying ML workflows for molecular and materials science applications. Designed end-to-end pipelines from data curation through PyTorch model training on GPU/CUDA-accelerated HPC systems to large-scale deployment, reducing compute cost by 98.3% across 10,000+ MD trajectories. Experienced in molecular property prediction, Bayesian optimisation with Gaussian Process surrogate models, and scientific computing with Python. Native domain knowledge in computational chemistry and atomistic simulation - the science underpinning Schrödinger's LiveDesign platform. Published in ACS Omega. Graduating August 2026.

AREAS OF EXPERTIS

ML Pipeline Development & Deployment | PyTorch | scikit-learn | GPU-Accelerated Computing | CUDA-Aware Workflows | Molecular Property Prediction | Bayesian Optimization & Gaussian Process Surrogate Models | Data Preprocessing & Feature Engineering | Model Training, Validation & Evaluation | Large-Scale Scientific Data Processing | HPC & Workflow Automation | Python & Scientific Computing | Atomistic Simulation (VASP, LAMMPS) | NumPy | pandas | SciPy | Git

KEY ACHIEVEMENTS

ML Training Pipeline at Scale: Built an end-to-end ML workflow for molecular systems - curating 8,094 AIMD training configurations, training PyTorch-based MLIPs on GPU/CUDA-accelerated HPC systems, and deploying across 10,000+ LAMMPS trajectories, reducing compute time by 98.3% while maintaining physical fidelity.

Molecular Property Prediction: Built scikit-learn classification pipelines to predict photonic band gaps from high-dimensional crystallographic descriptors across 800+ crystal structures; identified key predictive structural features through importance analysis. Published in ACS Omega (2026).

Bayesian Optimization for Scientific Workflows: Applied Bayesian Optimization with Gaussian Process surrogate models to optimize kinetic parameters in molecular simulations, improving agreement with experimental behavior across multiple system compositions.

Large-Scale Scientific Data Pipeline: Designed and automated Python workflows for high-volume simulation data curation, feature extraction, model training, and large-scale deployment on Linux/HPC systems.

PROFESSIONAL EXPERIENCE

University of Notre Dame — Graduate Research Assistant Notre Dame, IN

Zeolite Dealumination Modeling | Advisor: Prof. William F. Schneider Aug 2025 – Present

- Developed and solved global second-order kinetic models in Python using SciPy ODE solvers and non-linear regression fitting to analyze experimental dealumination kinetics across Si/Al ratios (16, 22, 45), demonstrating failure of mean-field assumptions with MPE = +18.15% and RMSPE = +106.89%.
- Applied Bayesian Optimization with Gaussian Process surrogate models to optimize kinetic parameters in kinetic Monte Carlo (kMC) simulations, improving agreement with experimental dealumination behavior across multiple Si/Al ratios.
- Computed free energy landscapes using umbrella sampling and WHAM analysis to calculate water binding free energy into structural dependent systems

Plasma-Surface Reaction Modeling | Advisor: Prof. William F. Schneider Aug 2024 – Aug 2025

- Generated and curated 8,094 AIMD training configurations from VASP calculations and trained Machine Learning Interatomic Potentials (MLIP/REANN) using Python-based scientific computing workflows.
- Deployed trained MLIPs into LAMMPS Deep Potential Molecular Dynamics (DPMD) simulations across 10,000 trajectories per vibrational state (n = 0–4), reducing computational cost by 98.3% compared to AIMD (630 min → 2 min per trajectory)
- Analyzed statistically converged N_2 dissociation probabilities on Pd(111) using NumPy, pandas, and matplotlib to derive Boltzmann-based reaction rate expressions for plasma-surface catalytic systems.

University of Wisconsin–Madison — Undergraduate Research Assistant Madison, WI

ML for Photonic Band Gap Prediction | Advisor: Prof. Rose Cersonsky Dec 2022 – May 2024

- Built machine learning classification pipelines in Python using scikit-learn, pandas, and NumPy to predict structure-property relationships from high-dimensional molecular and crystallographic descriptors
- Trained and evaluated Random Forest, Support Vector Classifier (SVC), and Logistic Regression models; identified Steinhardt Order Parameters as the strongest predictive descriptor through feature-importance analysis.
- Contributed to a peer-reviewed publication in ACS Omega (2026). DOI: 10.1021/acsomega.6c00256.

PROJECTS

STRAP Process Design — Industrial Capstone Project University of Wisconsin–Madison | 2024

Solvent-Targeted Recovery and Precipitation (STRAP) plant design for recycling multilayer plastic waste into 500 kg/hr polyethylene resin.

- Designed and simulated the complete STRAP recycling process in Aspen Plus, including dissolution, multi-stage filtration, precipitation, heat exchanger networks, dryer-condenser systems, and solvent recovery achieving 99.9% dodecane recycle.
- Used Aspen Economic Analyzer to size and cost 25+ unit operations and Aspen Energy Analyzer to perform heat integration, combining multiple heat exchangers into a single integrated unit to reduce utility demand and operating costs.
- Performed full material and energy balances, thermodynamic analysis, and DCFROR economic evaluation for plant capacities ranging from 500–2,000 kg/hr, obtaining baseline IRR of 5.18% at polyethylene selling price of $1.584/kg.
- Evaluated process sensitivity to polyethylene selling price, solvent recovery efficiency, and plant scale using Aspen-based process simulations and engineering economic analysis.

Business Improvement Design University of Wisconsin–Madison | 2024

Madison Museum of Contemporary Art (MOCA) audience analytics project

- Analyzed 1,000+ attendance and event datasets using Python, pandas, and scikit-learn to identify major drivers influencing museum attendance and audience engagement.
- Trained Random Forest machine learning models and applied feature-importance analysis to quantify effects of weather conditions, movie themes, and event types on audience turnout.
- Developed data-driven forecasting insights for attendance optimization, identifying movie theme selection as the strongest predictor of achieving target audience size.

EDUCATION

University of Notre Dame Notre Dame, IN

M.S. in Chemical Engineering (GPA: 3.83/4.00) Expected August 2026

Thesis: A Mechanism-Based Prediction of Irreversible Dealumination

University of Wisconsin–Madison Madison, WI

B.S. in Chemical Engineering (GPA: 3.43/4.00) May 2024

PUBLICATIONS

Cersonsky, R. K.; Nayak, S. K.; **Lee, S. H.** Data-Driven Design Rules for Three-Dimensional Photonic Crystals. ACS Omega (2026). DOI: 10.1021/acsomega.6c00256.

ADDITIONAL EXPERIENCE

University of Wisconsin–Madison — Course Grader & Teaching Assistant Jan 2022 – May 2024

- Graded assignments and led lab sections for sophomore chemical engineering students covering batch, continuous, and semi-batch process operations.

LEADERSHIP & SERVICE

Notre Dame International — International Ambassador Dec 2024 – May 2026

- Guided 1,400+ incoming international students through orientation and provided 1-on-1 mentoring on academic and cultural transition.

Korean Business Student Association (UW–Madison) Dec 2023 – May 2024

- Applied ML methods to identify profitability drivers and delivered data-driven recommendations for a non-profit client.

HONORS & AWARDS

Ellis C. Riplinger Scholarship Fund