Kieran Nehil-Puleo

kierannp | Kieran Nehil-Puleo | kierannp.github.io | nehilkieran@gmail.com

Education

Interdisciplinary Materials Science Ph.D., Vanderbilt University 2021–2026

Awards:

- Graduate Research Fellowship Program, National Science Foundation (NSF-GRFP)
  - Awarded to top 2,000 graduate students in STEM fields (out of 12,000+ applicants)
  - $38,000/year stipend + $12,000/year cost of education allowance for 3 years

Statistics B.S. (Honors), Michigan State University 2017–2021

Materials Science & Engineering B.S. (Honors), Michigan State University 2017–2021

Minors: Computer Science, Computational Modeling

Awards:

- STATE Scholarship (merit-based; 98th percentile ACT)
  - Awarded to top 5% of incoming freshmen
- Burnett Endowment Award
  - Awarded to outstanding engineering students based on academic performance

Experience

Vanderbilt University / *Ph.D. Student, Graduate Researcher* (Aug 2021 – Mar 2026, Nashville, TN)

- Conducted research at the intersection of machine learning, computer science, and computational physical chemistry.
- Developed and applied algorithms for molecular representation learning, property prediction, and generative molecular design.
- Utilized high-performance computing, GPU acceleration (CUDA, PyTorch), and distributed simulation workflows for large-scale data analysis.
- Integrated data-driven modeling with physical simulation frameworks to improve accuracy and efficiency in materials discovery.
- Experienced in C++, Python, PyTorch, NumPy, MPI, and software development for open-source scientific ecosystems.

Michigan State University / *Undergraduate Researcher* (Aug 2017 – Aug 2021, East Lansing, MI)

- Built computer vision algorithms (Python, C++, OpenCV) for automated X-ray diffraction image analysis.
- Designed scripts and workflows for large-scale computational materials datasets on Linux/HPC environments.

Textron, Bell Flight / *Business Analytics Intern* (Jun 2019 – Aug 2019, Fort Worth, TX)

- Automated financial and supply chain reporting with Python, reducing manual workload and turnaround time.
- Analyzed helicopter repair/maintenance data to forecast part consumption and improve contract proposals.

OptiO2 LLC / *Research Assistant* (Summer 2017, East Lansing, MI)

- Calibrated oxygen sensors for environmental/agricultural applications.
- Performed data analysis with Python and Excel.

- Assisted in first field deployment of wireless oxygen sensors in Colorado riverbed.

Publications, Patents, and Projects

Grouper: symmetry-aware functional-group graph representations for generative exploration of chemical space Publication

Developed a symmetry-aware, hierarchical molecular representation leveraging functional-group motifs to efficiently explore chemical space. Implemented combinatorial and algebraic methods that dramatically reduced graph isomorphism checks, achieving up to two orders of magnitude in computational savings, and enabled scalable, parallel molecular design workflows. Demonstrated applications in solubility optimization and polymer functionalization, establishing a foundation for interoperable, end-to-end molecular discovery pipelines.

Nehil-Puleo, K., Craven, N. C., McCabe, C., & Cummings, P. T. (2025). ***Nature Computational Science*** (Under review)

E(n) Equivariant Graph Neural Network for Learning Interactional Properties of Molecules Publication

Developed the Interactional Equivariant Graph Neural Network (IEGNN), a multi-input E(n)-equivariant model for predicting chemical properties arising from interactions between heterogeneous molecular structures. The model efficiently learns from 3D molecular geometry while preserving spatial symmetries, and is supported by two newly curated, publicly available interactional datasets and an open-source PyTorch Geometric data framework for multi-graph batching. IEGNN achieved state-of-the-art accuracy, yielding the lowest prediction errors on tribological property benchmarks across most tested datasets.

Nehil-Puleo, K., Quach, C. D., Craven, N. C., McCabe, C., & Cummings, P. T. (2024). ***The Journal of Physical Chemistry B***, 128(4), 1108–1117.

FiberForge: High-throughput Simulations of Helical Fibrils Publication

Developed FiberForge, an open-source, high-throughput workflow for constructing, simulating, and analyzing amyloid protofibrils to predict their mechanical properties and fracture mechanisms. The framework achieves accurate fibril reconstruction across diverse sequences and symmetries, and enables statistically robust mechanical property estimation from large-scale simulations. Validated against experimental data, FiberForge establishes a predictive, data-driven platform for the rational design of amyloid-based materials with tunable mechanical behavior.

Nehil-Puleo, K., & Yang, J. (2025). ***Digital Discovery*** (Under review).

On the Choice of Acceleration Amplitude for Predicting Shear Viscosity Publication

Developed a data-driven framework to streamline viscosity prediction using the non-equilibrium molecular dynamics Periodic Perturbation Method (PPM). By applying Quentrec's local order theory, the approach enables accurate extrapolation to zero-perturbation viscosity while reducing computational overhead. Validated across 144 organic solvents, the method achieves strong agreement with experimental and equilibrium simulation results, and introduces a predictive model for optimal acceleration amplitude selection—eliminating the need for extensive parameter tuning and enhancing the efficiency of viscosity estimation workflows.

Gui, L., Tsochantaris, E., **Nehil-Puleo, K.**, Cummings, P. T., & McCabe, C. (2025). ***Journal of Chemical Physics***.

Dynamically interconnected microbioreactors and applications thereof Patent

Co-invented a dynamically reconfigurable microfluidic network platform linking multiple microbioreactors via pumps and crossbar valves, enabling controlled transfer and recirculation of cells and media to emulate spatial heterogeneity in bioreactor conditions. Designed solutions for scale-down modeling of large industrial reactors, facilitating rapid switching of reactor topologies, automated environmental control, and integration with AI-driven experimental planning

Wikswo, J. P., **Nehil-Puleo, K.**, Reiserer, R. S. **U.S. Patent Application No. 2024/0110143 A1.**

Open-source Powder Sample Formulator Project

CAD-based automated dispensing system with embedded control software.

Robust Areal Diffraction Peak Detection Based on Shannon Entropy Presentation

Developed and implemented C++ algorithms leveraging Shannon entropy for automated detection of X-ray diffraction peaks under high-noise conditions; presented results at TMS 2022.

Nehil-Puleo, K., Tischler J., Eisenlohr P. (2022). ***TMS National Conference***, 2022.

Technical Skills

Programming

Python – Advanced (8+ yrs, scripting, ML pipelines)

- Daily use for scripting, data analysis, and ML model development
- Built end-to-end ML pipelines with data ingestion, preprocessing, model training, and evaluation
- Extensively created computational workflows for molecular simulations and machine learning
- Contributed code to the open-source libraries HOOMD-Blue, RFDiffusion, MoSDeF

C++ – Advanced (8+ yrs, serial & parallel algorithms)

- Developed high-performance algorithms for molecular simulations and data processing
- Implemented parallel computing techniques with OpenMP and MPI for scalability on HPC clusters
- Created Python bindings with Pybind11 for seamless integration in Python workflows

Bash – Advanced (8+ yrs, everyday use)

- Automated routine tasks and data processing workflows on Linux systems and HPC clusters
- Managed job scheduling and resource allocation with SLURM for large-scale simulations

JavaScript – Intermediate (2+ yrs, personal projects)

- Built interactive web applications with React.js and Node.js

Machine Learning / Artificial Intelligence (ML/AI)

PyTorch – Advanced (4+ yrs, SOTA)

- Implemented state-of-the-art architectures (e.g. Transformers, GNNs, denoising diffusion) for various tasks
- Utilized mixed-precision training and distributed data-parallelism to scale training on multi-GPU setups

- Focused on physics-informed ML models for molecular systems, incorporating domain knowledge into architectures and loss functions
- Developed custom datasets and data loaders for large-scale molecular datasets (10 k+ samples)

PyTorch Geometric – Advanced (4+ yrs)

- Built graph neural networks (GNNs) for molecular property prediction and generation tasks
- Applied techniques like message passing, graph attention, and equivariant layers to capture complex molecular interactions

TensorFlow – Advanced (3+ yrs)

- Designed CNN pipelines for image classification and segmentation tasks
- Utilized Keras for rapid prototyping and model development

MLOps – Intermediate (2+ yrs)

- Employed experiment tracking tools (Weights & Biases, TensorBoard) for reproducibility and hyperparameter tuning
- Orchestrated end-to-end ML pipelines using MLflow, Docker, and Kubeflow
- Automated hyper-parameter search with Optuna; reduced search space by 30
- Utilized active learning strategies to iteratively improve model performance with limited labeled data

High Performance Computing (HPC)

OpenMP – Intermediate (2+ yrs)

- Parallelized C++ algorithms for graph generation and molecular simulations
- Achieved 10x speedup on multi-core systems by optimizing thread usage and minimizing overhead

MPI – Intermediate (2+ yrs)

- Developed distributed computing solutions for large-scale simulations across HPC clusters
- Implemented efficient communication patterns to reduce latency and improve scalability

CUDA – Intermediate (2+ yrs)

- Accelerated computationally intensive tasks (e.g. molecular dynamics, graph processing) using GPU programming
- Wrote custom CUDA kernels for performance-critical sections of code
- Optimized memory access patterns and kernel launches to maximize GPU utilization

Pybind11 – Intermediate (2+ yrs)

- Created Python bindings for C++ libraries to enable seamless integration in Python workflows
- Facilitated rapid prototyping and testing of high-performance algorithms

SLURM – Intermediate (2+ yrs)

- Managed job scheduling and resource allocation on HPC clusters for large-scale simulations and machine learning tasks
- Automated job submission and monitoring with custom scripts to streamline workflows

DevOps

Git – Advanced (6+ yrs)

- Daily use for version control and collaboration on research projects
- Managed branches, pull requests, and code reviews in team settings
- Contributed to open-source projects with feature additions and bug fixes

CI/CD – Intermediate (2+ yrs)

- Implemented CI/CD pipelines with GitHub Actions

Cloud Platforms – Intermediate (2+ yrs)

- Deployed ML models and applications on local HPC clusters

Chemistry Software

Gromacs – Advanced (4+ yrs, MD simulations)

- Developed and optimized molecular dynamics workflows for biophysics and materials science applications
- Performed large-scale simulations of proteins, polymers, and complex fluids
- Utilized enhanced sampling techniques (e.g. metadynamics, umbrella sampling) to explore free energy landscapes

RDKit – Advanced (3+ yrs, Cheminformatics)

- Developed cheminformatics pipelines for molecular screening, descriptor generation, and QSAR modeling
- Implemented molecular similarity searches, substructure matching, and fingerprint analyses
- Integrated RDKit with Python data science libraries (e.g., Pandas, Scikit-learn) for predictive modeling and data analysis

AMBER – Intermediate (2+ yrs, MD simulations)

- Performed free energy calculations and binding affinity predictions
- Customized force fields for specific biomolecular systems

Orca – Intermediate (2+ yrs, Quantum Chemistry)

- Conducted QM solvation simulations to investigate solvent properties (COSMO)
- Performed DFT calculations for molecular optimization and property prediction

Hobbies

Brazilian Jiu Jitsu, Weightlifting, Rock Climbing