Sartaaj Khan

Contact Information

Email: sartaajk123@gmail.com | Phone: +1-647-607-5608

Education

University of Toronto Sept. 2023 – Present
PhD Chemical Engineering and Applied Chemistry
Principal Investigator: Dr. Seyed Mohamad Moosavi
cGPA: 4.00/4.00

University of Toronto Sept. 2018 – May 2023
BASc Chemical Engineering and Applied Chemistry
Minor in Bioengineering
cGPA: 3.73/4.00

Research Experience

Doctoral Researcher Sept. 2023 – Present
AI4ChemS Laboratory, University of Toronto Toronto, ON

Thesis: Learning the geometry of metal-organic frameworks (MOFs) using data-driven approaches

- Designed and implemented a multimodal machine learning framework that integrates powder X-ray diffraction (PXRD) patterns and textual descriptions of chemical precursors to link MOF synthesis pathways with targeted applications.
- Improved predictive performance through self-supervised learning and representation learning techniques, with fine-tuning on material properties including band gap, gas uptake (at multiple pressures), pore volume, and Henry's coefficient.
- Contributed to developing a novel graph-based representation (Labelled Quotient Graphs) to capture the bond topology of nanoporous materials, and benchmarked performance against leading graph neural network (GNN) architectures.
- Aided in the construction of a physics-informed machine learning model to predict dynamic viscosity of thermal fluids from small datasets and conducted uncertainty assessments of predictions.
- Leveraged insights from a large language model (LLM) to construct MOF-ChemUnity, a curated database linking experimental and computational MOF properties, to build ML models to predict water stability and discover MOFs for new applications.
- Mentored undergraduate and graduate students in topics related to machine learning in material science, informatics and representation learning.

Undergraduate Researcher July 2021 – Sept. 2022
Advanced Membranes Laboratory, University of Toronto Toronto, ON

- Developed a model showing the concentration profiles of common ions found in brackish water in a reverse osmosis process.
- Created a model involving different process modules in series such as a reverse osmosis membrane and a custom membrane specified.
- Used elemental-based modelling to compute the concentration profiles, flux of water and the flux of solute with respect to position and recovery ratio while including more complexities.

- Created and published a Python solution chemistry package called pySolution for the purpose of calculating parameters that are involved in non-ideal solutions from the Pitzer model.
    - GitHub repository is available here.

Undergraduate Researcher June 2021 – Sept. 2021
Ping Lab, University of California, Los Angeles Los Angeles, CA

- Developed a bioinformatics tool for generating a more reliable protein sequence database for sus scrofa cardiac proteomic studies.
    - GitHub repository is available here.
- Researched and developed features relevant in proteomics such as sequence length, PE score and number of confirmed isoforms.
- Created a classifier trained against homo sapien protein data from UniProt and tested against mus musculus and sus scrofa protein data to determine "reliable" and "unreliable" proteins in those respective organisms.
- Constructed cardiovascular knowledge graphs and created link prediction algorithms to find unknown connections between proteins in different organisms – specifically in cardiac conduction and muscle contraction.
    - GitHub repository is available here.

Undergraduate Researcher May 2020 – Sept. 2020
Laboratory of Metabolic Systems Engineering, University of Toronto Toronto, ON

- Created algorithms that genetically modify E. coli models and performed constrained optimization to optimize the flux of around 70 different industrially important chemicals through flux balance analysis.
- Modified mcPECASO simulator to obtain results with two-stage bioprocesses to analyze the impact of phenotypic switches on produced flux in iJO1366 and iML1515 E. coli models.
    - GitHub repository is available here.

Publications

Updated as of 2025/07/20. For full list, please refer to my Google Scholar.

- **Khan, S. T.**, Moosavi, S. M. (2025). Connecting metal-organic framework synthesis to applications using multimodal machine learning. Nature Communications, 16(1), 5642.
- Ai, Q.*, **Khan, S. T.***, Barthel, S., Moosavi, S. M (2025). "Capturing Global Features of Crystals from Their Bond Networks." AI4Mat Workshop ICLR-2025
- Kochi, M. R., Rezaei, H., **Khan, S. T.**, Mamillapalli, B. T., Ebrahimiazar, M., Ye, H., ... & Moosavi, S. M. (2025). Thermodynamics-informed machine learning for predicting temperature-dependent chemical properties. Preprint available on ChemRxiv.
- Pruyn, T. M., Aswad, A., **Khan, S. T.**, Black, R., & Moosavi, S. M. (2025). MOF-ChemUnity: Unifying metal-organic framework data using large language models. Preprint available on ChemRxiv.
- Zimmermann, Y., ..., **Khan, S. T.**, ... & Blaiszik, B. (2025). 34 Examples of LLM Applications in Materials Science and Chemistry: Towards Automation, Assistants, Agents, and Accelerated Scientific Discovery. arXiv preprint arXiv:2505.03049 and submitted to Machine Learning: Science and Technology (MLST).
- Zimmermann, Y., …, Khan, S. T., ... & Blaiszik, B. (2024) "Reflections from the 2024 Large Language Model (LLM) Hackathon for Applications in Materials Science and Chemistry" arXiv preprint arXiv:2411.15221.

Invited Presentations

- Ai, Q., **Khan, S. T.**, Barthel, S., Moosavi, S. M. "Capturing Global Features of Crystals from Their Bond Networks." [Paper]. AI for Accelerated Materials Design – ICLR, 2025
- **Khan, S.T.**, Moosavi, S.M. "Unveiling insights of metal-organic frameworks using a self-supervised multimodal model with only PXRDs and precursors." [Poster]. American Chemical Society (ACS), 2025, San Diego, California

- **Khan, S.T.**, Moosavi, S.M. "XRayPro: A Self-Supervised Multimodal Model for MOF Application Recommendations Using PXRD and Precursors." [Paper]. AI for Accelerated Materials Design – NeurIPS, 2024, Vancouver, Canada.
- **Khan, S.T.** (2024, October 8). Unveiling Property Insights and Recommending MOF Applications from X-Ray Diffraction Patterns using Self-Supervised Transformer Models. [Presentation]. Canadian Society for Chemical Engineering (CSChE) 2024, Toronto, Canada
- **Khan, S.T.**, Moosavi, S.M. (2024, August 7). "Using a self-supervised model for property predictions and MOF application recommendations with only powder x-ray diffraction patterns." [Poster]. Accelerate Conference 2024, Vancouver, Canada

Work Experience

Artificial Intelligence Engineer Sept. 2021 – Sept. 2022
MGRID Energy Inc. Whitehorse, YK

- Created and tested a variety of object detection models (primarily YOLO) using convolutional neural networks for the purpose of vehicle classification.
- Designed the architecture of a hybrid nonlinear regression-convolutional neural network model for the automatic detection of wildfires.
- Researched and extracted different features to include in the wildfire models through literature done on previous machine learning models in the area.
- Generated a synthetic dataset for the purpose of training a natural language processing (NLP) model to send alerts on new projects for bidding.

Teaching Experience

Machine learning (ML) tutor Feb. 2025 – Present
University of Toronto Toronto, ON

- Served as a Machine Learning Tutor for the Data Science Institute (DSI)'s CrossTALK – a bootcamp focused on applying AI and laboratory validation techniques for drug discovery.
- Provided technical support by troubleshooting a range of issues, including syntax errors, memory constraints, data loading, and documentation, across machine learning frameworks such as XGBoost and LightGBM.
- Delivered conceptual instruction on machine learning models, clustering algorithms, and the selection of molecular fingerprints, helping students understand both theoretical foundations and practical applications.

Research Mentorship

1. **Zakariyya Brewster** (Summer 2025): University of Toronto undergraduate student in Engineering Science; Common pitfalls in applying machine learning in metal-organic frameworks.
2. **Hari Om Chadha** (Summer 2024): University of Toronto undergraduate student in Engineering Science; Pretraining graph convolutional networks using Barlow-Twin for predicting fluid properties.

Awards

Queen Elizabeth II/DuPont Canada Scholarship in Science and Technology May 2025

LLMs in Chemistry and Materials Hackathon May 2024

Awarded 3rd place ($250 prize funded by Radical AI and Anthropic) in the LLM Hackathon for Applications in Materials and Chemistry for the development of **PoreVoyant** – a chemistry-informed AI agent that can generate new linkers to decrease the

band gap in metal-organic frameworks while leveraging MOF literature. This work was featured in a Medium article. The GitHub repository can be found here.

Bayesian Optimization for Chemistry and Materials Hackathon March 2024

Awarded 2nd place ($500 CAD prize funded by Acceleration Consortium) in the Bayesian Optimization for Chemistry and Materials Hackathon hosted by Acceleration Consortium at the University of Toronto. The project involved the application of Bayesian Optimization to accelerate the discovery of fluids with the highest heat transfer coefficients. The repository can be found here.

Dean's List Sept. 2018 – May 2023

Award issued for achieving an average of at least an A- per semester with a full course load. This was achieved for all semesters of undergraduate.

NSERC USRA May 2020

Award of $6000 CAD issued for completion of 16 weeks of research work for the summer 2020 term and high academic achievements.

Skills

Programming: Python, MATLAB and Simulink, Cypher Query Language, Neo4j, HTML, CSS, Bash

Machine learning: PyTorch, Tensorflow, Scikit-Learn

Software: ChemDraw, AutoCAD, Aspen Plus, Aspen Dynamics, VESTA, iRASPA

Applications: Microsoft Word, Microsoft PowerPoint, Microsoft Excel, Microsoft VISIO