JEFF GUO

Google Scholar | LinkedIn | GitHub | jeff.guo@outlook.com

I am a final year PhD student at EPFL (**September 2025**) advised by Professor Philippe Schwaller, developing **molecular generative models** for **accelerated drug discovery**. Previously, I was an intern at Microsoft Research AI4Science with Jose Garrido Torres, Elise van der Pol, and Marwin Segler. Before that, I was a **cheminformatics researcher** in the Molecular AI department at AstraZeneca where I **developed tools for molecular design** and contributed to the **identification of experimentally validated candidate drug molecules**. My undergraduate degree was in chemistry, inspiring my research to **harness domain knowledge to augment machine learning methods** for chemical discovery.

RESEARCH EXPERIENCE

PhD Researcher, EPFL **Sep. 2022 –**

Developing **sample-efficient molecular generative models** for drug and catalyst design with **synthesizability constraints**. Prospective experimental validation of generated molecules for binding. Applying Bayesian optimization for chemical reaction optimization with on-going collaboration with experimental chemists.

Research Intern, Microsoft AI4Science **Oct. 2023 – Jan. 2024**

Machine learning for chemistry. Supervised by Jose Garrido Torres, Elise van der Pol, and Marwin Segler.

Graduate Scientist Researcher, Molecular AI AstraZeneca **Sep. 2020 – July 2022**

Expanded the capabilities of the generative reinforcement learning molecular design tool, `REINVENT`, to accelerate early-phase drug discovery. Contributed to the identification of candidate drug molecules by coupling `REINVENT` with structure-based molecular docking, ligand-based pharmacophore matching, and fragment linking functionalities. Devised and co-advised a research project to integrate active learning with `REINVENT` to improve sample efficiency.

Research Assistant, Imperial College London **Nov. 2019 – Aug. 2020**

Led small molecular inhibitor design for an oncology target by applying fragment-based strategies with molecular docking. Designed and synthesized candidate molecules and demonstrated binding via X-ray crystallography and NMR. Set up a virtual screening `Glide` docking pipeline on an AWS cloud computing server, increasing throughput from the order of 10^4 to 10^7 molecules.

Wet-lab Research, McGill University and Alnylam Pharmaceuticals **May 2017 – Sep. 2019**

Three research internships in organic synthesis for green and nucleic acid chemistry.

SELECTED OPEN-SOURCE PROJECTS

`Saturn` – GitHub, Paper

Sample-efficient generative molecular design frame-work. Includes `TANGO` (Constrained synthesizability), `General Synthesizability` and, `Steerable and Granular Synthesizability Control` implementations.

`Augmented Memory` – GitHub

Fork of `REINVENT` 3.2 with `Augmented Memory` and `Beam Enumeration` implementations for sample efficiency and explainability in generative molecular design.

`REINVENT` Version 3.2 – Repository Contributor, GitHub

Contributed to the expansion of `REINVENT`` capabilities with Curriculum Learning and Linker Generation. Wrote tutorial notebooks (GitHub).

EDUCATION

PhD Chemistry and Chemical Engineering, EPFL — **Sep. 2022 –**
NSERC Post-Graduate Doctoral Scholarship – $120,000
MRes Drug Discovery and Development, Imperial College London — **Oct. 2019 – Sep. 2020**
BSc Honours Chemistry, McGill University — **2015 – 2019**

SKILLS

Molecular design: Generative models with transfer learning/fine-tuning and goal-directed generation (structure-based and ligand-based). Virtual screening and with active learning for large libraries.
Molecular docking: Glide, **gnina**, **AutoDock Vina**, **QuickVina2-GPU**, MOE, GOLD, OpenEye
Ligand-based tools: ROCS (OpenEye), ESP-sim
Chemical reactions: Bayesian optimization, hands-on wet-lab experience (synthesis and characterization)
Quantum chemistry: xTB (semi-empirical), DFT (ORCA)
Chemical visualization: ChemDraw, PyMOL
Programming: Python, PyTorch, Linux, HPC, scikit-learn, RDKit

SUPERVISION

Victoire Lang '25 BSc, EPFL (Internship – **NeurIPS 2024 AI4Mat**) — **July 2024 – Aug. 2024**
Rémi Schlama '24 MSc, EPFL (Internship) — **Mar. 2024 – May 2024**
Sacha Raffaud '23 MSc, ICL (MSc Thesis – **NeurIPS 2023 Diffusion Workshop**) — **June 2023 – Sep. 2023**
Christian Josefson & Clara Nyman '22 MSc, Chalmers (MSc Thesis) — **Jan. 2022 – May 2022**

SELECTED PUBLICATIONS (See Google Scholar for all publications)

§ **denotes equal author contribution**

[1] **J. Guo**§, V. Sabanza§, Z. Jončev, J. Luterbacher, P. Schwaller "Generative Molecular Design with Steerable and Granular Synthesizability Control", ***arXiv*, 2025.** Link

[2] H. Wang, **J. Guo**, L. Kong, R. Ramprasad, P. Schwaller, Y. Du, C. Zhang "LLM-Augmented Chemical Synthesis and Design Decision Programs", **ICML 2025.** Link

[3] **J. Guo**, P. Schwaller "It Takes Two to Tango: Directly Optimizing for Constrained Synthesizability in Generative Molecular Design", ***Nat. Comput. Sci*** **(accepted), 2026.** Link

[4] **J. Guo**, P. Schwaller "Directly Optimizing for Synthesizability in Generative Molecular Design using Retrosynthesis Models", ***Chem. Sci.*, 2025.** Link

[5] **Review Paper**: Y. Du§, A. R. Jamasb§, **J. Guo**§, T. Fu, C. Harris, Y. Wang, C. Duan, P. Liò, P. Schwaller, T. L. Blundell "Machine Learning-Aided Generative Molecular Design", ***Nat Mach Intell*, 2024.** Link

[6] **J. Guo**, P. Schwaller "Saturn: Sample-efficient Generative Molecular Design using Memory Manipulation", ***Nat Mach Intell*** **(accepted).** Link

[7] **J. Guo**, P. Schwaller "Beam Enumeration: Probabilistic Explainability For Sample Efficient Self-conditioned Molecular Design", **ICLR 2024.** Link

[8] **J. Guo**, P. Schwaller "Augmented Memory: Sample-Efficient Generative Molecular Design with Reinforcement Learning", ***JACS Au*, 2024.** Link

[9] M. Dodds, **J. Guo**, T. Löhr, A. Tibo, O. Engkvist, J. P. Janet "Sample Efficient Reinforcement Learning with Active Learning for Molecular Design", ***Chem. Sci.*, 2024.** Link

[10] **J. Guo**§, F. Knuth§, C. Margreitter, J. P. Janet, K. Papadopoulos, O. Engkvist, A. Patronov "Link-INVENT: Generative Linker Design with Reinforcement Learning", ***Digital Discovery*, 2023.** Link

[11] **J. Guo**§, V. Fialková§, J. D. Arango, C. Margreitter, J. P. Janet, K. Papadopoulos, O. Engkvist, A. Patronov "Improving De Novo Molecular Design with Curriculum Learning", ***Nat Mach Intell*, 2022.** Link

[12] **J. Guo**, J. P. Janet, M. R. Bauer, E. Nittinger, K. A. Giblin, K. Papadopoulos, A. Voronov, A. Patronov, O. Engkvist, C. Margreitter "DockStream: A Docking Wrapper to Enhance De Novo Molecular Design", ***J Cheminform*, 2021.** Link