Adrian Mirza

GitHub | LinkedIn | adrianmirza.com | mail@adrianmirza.com

Summary

PhD researcher with experience in AI for chemical sciences. I have worked on diverse projects—from low-complexity ML applied to homogeneous catalysis, to building models and systems from scratch for molecular chemistry, spectroscopy, and materials science. Additionally, I have experience with molecular simulations and data management from roles at Software for Chemistry and Materials and Stability AI.

Work Experience

Data Research Engineer – Part Time August 2023 – July 2024
`Stability AI`

Contributed to Open Science and ChemNLP by collecting and curating chemistry datasets and performing data research.

Intern in Computational Chemistry and Software Engineering April 2023 – July 2023
`Software for Chemistry and Materials`

- Conducted literature reviews to identify relevant KPIs for conformer simulation tools.
- Applied graph theory to chemical problems related to conformers.
- Benchmarked internal conformer generation tools against literature.

Teaching Assistant – Part Time April 2023 – July 2023
`Delft University of Technology`

- Restructured a Master-level course on computational, numerical, and data science methods.
- Prepared tutorials for 140 graduate students.
- Prepared and graded assignments.

Teaching Assistant – Part Time April 2023 – July 2023
`TotalEnergies`

- Analyzed and cleaned employee data.
- Built a Microsoft Office pipeline for data streamlining and analysis.
- Provided IT and HR support.

Education

2023 – present **PhD in Digital Chemistry**, Helmholtz Zentrum Berlin
Advisor: Kevin Jablonka

2021 – 2023 **MSc Chemical Engineering**, Delft University of Technology
Advisor: Evgeny Pidko

2018 – 2021 **BSc Chemical Engineering & Biotechnology**, Aalborg University
Advisor: Marco Maschietti

Languages

Native – Romanian, **Professional proficiency** – English, Russian, **Limited proficiency** – German

Selected Publications

Mirza, Adrian and Kevin Maik Jablonka (2024). "Elucidating structures from spectra using multimodal embeddings and discrete optimization". In: *chemrxiv preprint*. DOI: 10.26434/chemrxiv-2024-f3b18-v2.

Mirza, Adrian, Nawaf Alampara, Anagha Aneesh, et al. (2025). "General purpose models for the chemical sciences". In: *arXiv preprint arXiv:2507.07456*.

Mirza, Adrian, Nawaf Alampara, Sreekanth Kunchapu, et al. (2025). "A framework for evaluating the chemical knowledge and reasoning abilities of large language models against the expertise of chemists". In: *Nature Chemistry*, pp. 1–8.

Mirza, Adrian, Nawaf Alampara, Martiño Ríos-García, et al. (2025). "ChemPile: A 250GB Diverse and Curated Dataset for Chemical Foundation Models". In: *Advances in Neural Information Processing Systems 39: Datasets and Benchmarks Track*. URL: https://openreview.net/forum?id=l8Mz3dcK1S.

Skills

Programming Languages	Python, Bash, MatLab, Rust
Scientific Computing and ML	SciPy, NumPy, PyTorch, PyTorch Lightning, Modal, Amsterdam Modelling Suite, Gaussian16, HuggingFace
Other Tools and Frameworks	ChemDraw, Adobe Illustrator, LaTeX, Git

Conferences — Summits — Meetings

2025 NeurIPS 2025: **Poster presentation**
2025 AI4Mat ICLR Singapore: **Spotlight talk**
2024 AI4Mat Vienna: **Spotlight talk**
2024 L2M3 CECAM: **Invited talk**
2024 HAICON24: **Poster presentation**
2024 IWOMI: **Invited talk**
2024 Llama Community Summit: Invited to Meta's HQ for discussions about applications of AI
2023 The Netherlands' Catalysis and Chemistry Conference (NCCC): **Poster presentation**
2020 DTU Student Plastic Project Contest: **Invited talk (top-5 across Denmark)**