Dr. Amol Thakkar

Science | AI | Engineering | Product

Google Scholar • LinkedIn • GitHub • ORCID

Nationality: British

AI practitioner with experience taking research systems into production. Background spanning lab science, software engineering, and product-facing AI in industry.

Skills

Technical: Machine Learning, Deep Learning, Large Language Models, PyTorch, Hugging face, Tensorflow/Keras, Sklearn, Cloud (AWS, IBM Cloud), FastAPI, Docker, Kubernetes, Ray, SQL and NoSQL Databases (Postgres, SQLite, Mongo, Redis), Dapr, CI/CD, Git, OpenShift, SLURM, LSF, RDKit

Soft: Product Management, Client Relationships, Agreements (SOW, JDA, Licensing), Communication, Stakeholder Management, Adaptability, Agile, Market Research, Technical Specifications, Problem Solving, User Requirements, Roadmaps

Work Experience

Ellison Institute of Technology, Oxford

Applied AI Scientist — July 2025 – Present

- Automated Laboratories Leading the data science and engineering for a large-scale automated laboratory build-out across synthetic biology and plant science. Designed and implemented a natural language system that abstracts complex robotics hardware into high-level code libraries, enabling automated workflow and code generation for internal users. As the project's technical lead, transitioned the initiative from a solo project to a cross-functional team by defining the engineering vision, identifying resource gaps, and advocating for departmental support within a matrix organization. Scaled the team's capabilities by defining technical personas, authoring job descriptions, and serving as the technical interviewer, while overseeing the selection and integration of all sensors, software tooling, and data architecture.

IBM Research Europe, Switzerland

Staff Research Scientist — March 2022 – June 2024

- NLP platform for chemical reaction prediction *(Software, Data, ML, Product, and Consulting)*
  Led client engagements and contributed to the research and development of a cloud-based Natural Language Processing-driven platform for chemical reaction prediction and synthesis route planning used by 30k+ users. Researched and built new features, data pipelines, APIs, trained and deployed language models, integrated 3rd party models. Successfully deployed across multiple client sites to enhance R&D workflows. Led product management for Generative AI in Chemistry SaaS solutions, cross functionally across engineering, payments, legal, and research, working with a team of 4 contractors, and internally across 5 continents.

- Event-driven system combining multi-modal AI and IOT *(Software, Data, and Consulting)*

Collaborated on the design and implementation of an event-driven system for the automated documentation of operational workflows, integrating multi-modal AI (vision and language) and IoT across multiple cloud environments (AWS, IBM Cloud, on premise). Sourced clients for pilot studies and scoped initial work.

- Conversational text-to-SQL *(Software, Data, ML, and Consulting)*
  Contributed to a conversational text-to-SQL system, built an evaluation framework to automatically test on academic and client databases. Led client engagements, fine-tuned LLMs, and augmented data. Deployed across on premise and cloud solutions (AWS, IBM cloud) Topped academic leaderboards (2024).

- Additional Contributions: development of transformers for neuromorphic chip architectures, advocated for good development practices, and mentored junior colleagues.

AstraZeneca, Sweden

PhD Research Scientist — January 2019 – December 2021

- Deep learning guided Monte-Carlo tree search *(Software, Data, ML, Product, Consulting)*
  Co-led the development of one of the first open-source AI-based synthesis planning tools based on deep learning and Monte-Carlo tree search. AiZynthFinder is deployed across AstraZeneca's portfolio and integrated into the generative design platform, impacting ongoing projects. Used widely across industry and academia. Built and curated data pipelines from public and proprietary datasets, data augmentation, data licencing, experimental data acquisition, deep learning spanning feed forward neural networks to transformers, machine learning such as random forest and support vector machines.

- *Generative AI for neglected diseases (ML, Consulting)*
  Achieved sub-micromolar (in vitro) efficacy using AstraZeneca's generative design and synthesis platform for neglected diseases with external partner.

- Additional Contributions: Managed students/interns, wrote and contributed to new project proposals, advised on strategic initiatives, provided computational support to departments across the organisation, and promptly communicated research findings to stakeholders.

Pfizer, UK

Research Intern — September 2015 – September 2016

- Developed and scaled novel synthetic methodology to the pilot plant facility.
- Implemented alternative chemistry to overcome IP bottlenecks in route development.

University of St Andrews, UK

Research Intern — 2013-2015

- Explored the synthesis of bioactive peptides from sustainable chemical feedstocks.
- Development and characterisation of iron doped zeolites for catalysis.
- Synthesis and structural determination of compounds for organic light emitting diodes.

Education

University of Bern, Switzerland

Doctor of Chemistry and Molecular Sciences (AI for Chemistry) 2018-2022
Summa cum Laude

Thesis: Exploration of Chemical Space Through Data-Driven and Artificial Intelligence Guided Synthesis Planning
Supervisors: Professor Jean-Louis Reymond (University of Bern), Dr Ola Engkvist (AstraZeneca)

- Applied artificial intelligence techniques spanning deep learning and search strategies to predict the manufacturing routes to chemicals.
- Predicted synthesis for a novel neurotransmitter in *in vitro* studies.

University of St Andrews, UK

Master in Chemistry (Honours) 2012-2017
First Class

Thesis: Modelling Uranyl Chemistry in Liquid Ammonia from Density Functional Theory
Supervisor: Professor Michael Bühl

Languages

Python (Main)
Javascript (Basic)
HTML (basic)
CSS (basic)

English (Native)
Gujarati (A2)
German (A1)
Norwegian (Basic)

Volunteering

Swimming Club

Organizer 2022-2024

Zurich Graduate Consulting Club, Switzerland

Consultant September 2021 – June 2021

- Advised a start-up focused on digital identity verification using computer vision on integrating with large-scale IT system providers, conducting competitive and market analysis.
- Findings resulted in an overhaul of the customer onboarding process and web interface leading to 60% less ID verification abandonment.

University of St Andrews, UK

Student-Staff Representative September 2013 – June 2015

Combined Cadet Force, UK

Company Sergeant Major September 2005 – August 2012

Last update: January 1st, 2026

Awards

- SCS Fall Meeting 2021 – Best Oral Presentation (computational chemistry)
- Marie Skłodowska-Curie PhD Fellowship
- The University Scholarship for Research and Leadership
- The Laidlaw Undergraduate Internship Programme for Research and Leadership
- Dean's List 2014-2016, awarded to students averaging first class marks throughout the year.

Publications

1. **Thakkar, A.**, and Reymond, J.-L. Computer Aided Synthesis Prediction to Enable Augmented Chemical Discovery and Chemical Space Exploration. (Thesis)

2. Sieffert, N., **Thakkar, A.**, and Bühl, M. (2018). Modelling uranyl chemistry in liquid ammonia from density functional theory. Chemical Communications *54*, 10431–10434. https://doi.org/10.1039/C8CC05382K.

3. Johansson, S., **Thakkar, A.**, Kogej, T., Bjerrum, E., Genheden, S., Bastys, T., Kannas, C., Schliep, A., Chen, H., and Engkvist, O. (2019). AI-assisted synthesis prediction. Drug Discovery Today: Technologies *32*, 65–72. https://doi.org/10.1016/j.ddtec.2020.06.002.

4. **Thakkar, A.**, Kogej, T., Reymond, J.-L., Engkvist, O., and Bjerrum, E.J. (2020). Datasets and their influence on the development of computer assisted synthesis planning tools in the pharmaceutical domain. Chemical science *11*, 154–168. https://doi.org/10.1039/C9SC04944D.

5. David, L., **Thakkar, A.**, Mercado, R., and Engkvist, O. (2020). Molecular representations in AI-driven drug discovery: a review and practical guide. Journal of Cheminformatics *12*, 56. https://doi.org/10.1186/s13321-020-00460-5.

6. Genheden, S., **Thakkar, A.**, Chadimová, V., Reymond, J.-L., Engkvist, O., and Bjerrum, E. (2020). AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning. Journal of cheminformatics *12*, 70. https://doi.org/10.1186/s13321-020-00472-1.

7. Bjerrum, E.J., **Thakkar, A.**, and Engkvist, O. (2020). Artificial Applicability Labels for Improving Policies in Retrosynthesis Prediction. Machine Learning: Science and Technology *2*. https://doi.org/10.1088/2632-2153/abcf90.

8. **Thakkar, A.**, Selmi, N., Reymond, J.-L., Engkvist, O., and Bjerrum, E.J. (2020). “Ring breaker”: neural network driven synthesis prediction of the ring system chemical space. Journal of medicinal chemistry *63*, 8791–8808. https://doi.org/10.1021/acs.jmedchem.9b01919.

9. Sumner, D., He, J., **Thakkar, A.**, Engkvist, O., and Bjerrum, E.J. (2020). Levenshtein augmentation improves performance of smiles based deep-learning synthesis prediction. ChemRxiv. https://doi.org/10.26434/chemrxiv.12562121.v2.

10. **Thakkar, A.**, Johansson, S., Jorner, K., Buttar, D., Reymond, J.-L., and Engkvist, O. (2021). Artificial intelligence and automation in computer aided synthesis planning. Reaction chemistry & engineering *6*, 27–51. https://doi.org/10.1039/D0RE00340A.

11. **Thakkar, A.**, and Schwaller, P. (2021). How AI for synthesis can help tackle challenges in molecular discovery: Medicinal chemistry and chemical biology highlights. Chimia *75*, 677–677. https://doi.org/10.2533/chimia.2021.677.

12. **Thakkar, A.**, Chadimová, V., Bjerrum, E.J., Engkvist, O., and Reymond, J.-L. (2021). Retrosynthetic accessibility score (RAscore)–rapid machine learned synthesizability classification from AI driven retrosynthetic planning. Chemical science *12*, 3339–3349. https://doi.org/10.1039/D0SC05401A.

13. **Thakkar, A.**, and Reymond, J.-L. (2022). Automatic Extraction of Reaction Templates for Synthesis Prediction. Chimia *76*, 294–297. https://doi.org/10.2533/chimia.2022.294.

14. Kannas, C., **Thakkar, A.**, Bjerrum, E., and Genheden, S. (2022). Rxnutils–a cheminformatics python library for manipulating chemical reaction data. https://doi.org/10.26434/chemrxiv-2022-wt440-v2.

15. Chakraborty, A., **Thakkar, A**., Vaucher, A.C., and Laino, T. (2022). Data-driven Reaction Template Fingerprints. ChemRxiv. https://doi.org/10.26434/chemrxiv-2022-4kzp1.

16. Cretu, M., Alberts, M., Chakraborty, A., Leonov, A., **Thakkar, A.**, and Laino, T. (2023). Tools for Synthesis Planning, Automation, and Analytical Data Analysis. Chimia *77*, 17–21. https://doi.org/10.2533/chimia.2023.17.

17. Cretu, M.T., Toniato, A., **Thakkar, A.**, Debabeche, A.A., Laino, T., and Vaucher, A.C. (2023). Standardizing chemical compounds with language models. Machine Learning: Science and Technology *4*, 035014. https://doi.org/10.1088/2632-2153/ace878.

18. **Thakkar, A.**, Vaucher, A.C., Byekwaso, A., Schwaller, P., Toniato, A., and Laino, T. (2023). Unbiasing retrosynthesis language models with disconnection prompts. ACS Central Science *9*, 1488–1498. https://doi.org/10.1021/acscentsci.3c00372.

20. **Thakkar, A.**, and Laino, T. (2024). Neural Template Extraction-Learning the Language of SMIRKS. ChemRxiv. https://doi.org/10.26434/chemrxiv-2024-bmk3w.

21. Gabrieli, G., Morales, I.E., Christofidellis, D., Graziani, M., Giovannini, A., Zipoli, F., **Thakkar, A.**, Foncubierta, A., Manica, M., Ruch. P.W. (2024) Activity recognition in scientific experimentation using multimodal visual encoding. Digital Discovery, 4, 393. https://doi.org/10.1039/d4dd00287c.

22. Carrel, A., Yiannakas, A., Roukens, JJ., Reynoso-Moreno, I., Orsi, M., **Thakkar, A.,** Arus-Pous J., Pellegata, D., Gertsch, J., Reymond JL., (2025) Exploring Simple Drug Scaffolds from the Generated Database Chemical Space Reveals a Chiral Bicyclic Azepane with Potent Neuropharmacology. Journal of Medicinal Chemistry. https://doi.org/10.1021/acs.jmedchem.4c02549.

Conference Papers

1. **Thakkar, A.**, Bjerrum, E.J., Engkvist, O., and Reymond, J.-L. (2019). Neural Network Guided Tree-Search Policies for Synthesis Planning. In International Conference on Artificial Neural Networks (Springer International Publishing Cham), pp. 721–724. https://doi.org/10.1007/978-3-030-30493-5_64.

2. **Thakkar, A.**, Byekwaso, A.A., Schwaller, P., Vaucher, A., Toniato, A., and Laino, T. (2022). Augmented disconnection aware retrosynthesis to facilitate user interaction. In American Chemical Society (ACS) Fall Meeting.

3. **Thakkar, A.**, Byekwaso, A.A., Vaucher, A., Schwaller, P., Toniato, A., and Laino, T. (2022). Disconnection Aware Steering of Retrosynthesis Transformer to Facilitate Materials Design. In Materials Research Society (MRS) Fall Meeting.

4. **Thakkar, A.**, and Laino, T. (2023). NeuTE-Neural Template Extraction. In American Chemical Society (ACS) Spring Meeting.

5. Ruch, P., Vaucher, A., **Thakkar, A**., Giovannini, A., Foncubierta, A., Christofidellis, D., Zipoli, F., Born, J., Cadow, J., Rapsomaniki, M., et al. (2023). A new era of collaborative experimentation enabled by multi-cloud and AI foundation models. In Future Labs Live.

6. **Thakkar, A.**, Giovannini, A., Manica, M., Vaucher, A., Ruch, P., and Laino, T. (2023). Creating labs that learn through automated data management. In American Chemical Society (ACS) Fall Meeting.

7. **Thakkar, A.**, Giovannini, A., Foncubierta, A., Baldassari, C., Christofidellis, D., Zipoli, F., Gabrieli, G., Born, J., Graziani, M., Alberts, M., et al. (2023). Using Foundation Models to Promote Digitization and Reproducibility in Scientific Experimentation. In NeurIPS 2023 AI for Science Workshop.

8. Ruch, P., Gabrieli, G., Morales, I.E., Christofidellis, D., Graziani, M., Giovannini, A., Zipoli, F., **Thakkar, A.**, Foncubierta, A., Manica, M., et al. (2024). Multimodal foundation models for more reproducible scientific experimentation and data capture. In Future Labs Live.

Public Talks

1. Cheminformatics Strasbourg Summer School, Strasbourg, France, June 2018
2. Machine Learning in Drug Discovery Summer School, KU Leuven, Belgium, August 2018
3. Swiss Chemical Society Fall Meeting, Lausanne, Switzerland, 2018
4. German Cheminformatics Conference, Mainz, Germany, 2018
5. BigChem Spring School, Gothenburg, Sweden, May 2019
6. International Conference on Artificial Neural Networks, Munich, Germany,
7. September 2019

8. Biopharmaceuticals R&D Science Symposium, Gothenburg, Sweden, October 2019
9. German Cheminformatics Conference, Mainz, Germany, November 2019
10. AI-Powered Drug Discovery and Manufacturing, Boston, USA, February 2020
11. Swiss Chemical Society Fall Meeting, Bern, Switzerland, August 2020
12. Reaxys Academic Network, Virtual, October 2020
13. Swiss Chemical Society Fall Meeting, Bern, Switzerland, August 2021
14. RSC AI in Chemistry, London, UK, September 2021
15. Cheminformatics Workshop, Lausanne, Switzerland, September 2021
16. RDKit User Group Meeting, Virtual, October 2021
17. NSF Workshop – Lab Automation and Molecular Innovation, University of Illinois Urbana-Champaign, June 2022
18. American Chemical Society Fall Meeting, Chicago, USA, August 2022
19. ILMAC Symposium on Flow Chemistry and Artificial Intelligence, Lausanne, Switzerland, September 2022
20. Materials Research Society Fall Meeting, Boston, USA, November 2022
21. SCS Spring School on Digital Chemistry, Le Diablerets, Switzerland, April 2023
22. American Chemical Society Spring Meeting, Virtual, April 2023
23. American Chemical Society Fall Meeting, Virtual, April 2023
24. European Conference on Computational and Theoretical Chemistry, Thessaloniki, Greece, August 2023
25. ChemAI, Amsterdam Science Network, Amsterdam, Netherlands, November 2023
26. AIChemist, CECAM-HQ-EPFL, Lausanne, Switzerland, April 2025

References

Available on request.