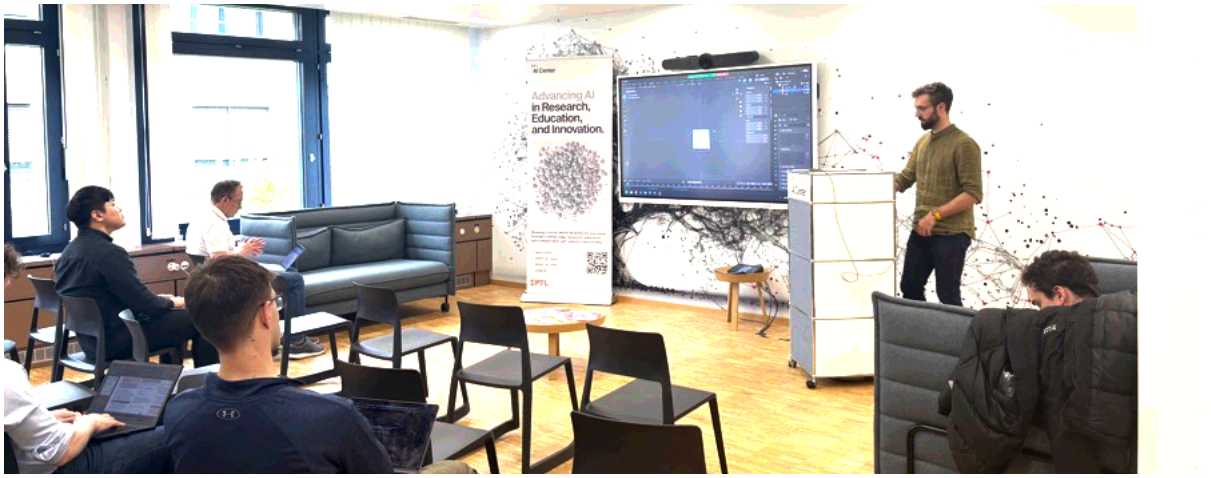
Third AiChemist School – EPFL/CECAM, Lausanne, Switzerland

The third AiChemist school was held from the 23rd of April until the 2nd of May 2025 at EPFL/CECAM in Lausanne, Switzerland. The second week of the school (29th April until the 2nd of May) was dedicated to the AIChemist CECAM Flagship School (https://www.cecam.org/workshop-details/aichemist-1374) which was organised by members of the AiChemist consortium with the assistance of CECAM (Centre Européen de Calcul Atomique et Moléculaire, https://www.cecam.org) and open to in-person and online participants outside of the AiChemist consortium. The first week was hosted by the EPFL AI Centre and was open only to AiChemist students. The school opened on April 23 with a welcome address by the organisers, setting the tone for a week of exploration and hands-on learning. The first day was dedicated to foundational concepts in atomistic machine learning, led by Dr. Davide Tisi, postdoc in the lab of Prof. Michele Ceriotti (COSMO) at the Insitute of Materials, EPFL. His series of lectures introduced atom-centred representations, culminating in a hands-on session that allowed participants to apply these concepts directly.

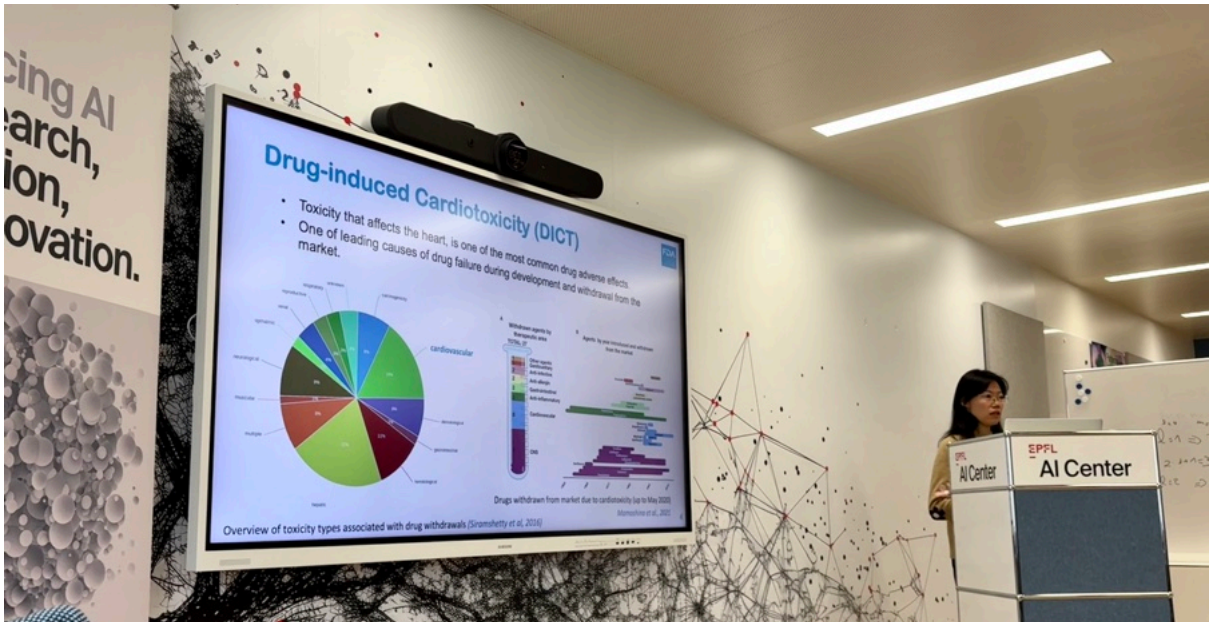
Dr. Davide Tisi giving a lecture on atom-centred representations

In the afternoon, Dr. Tuan Le from Pfizer gave a lecture on structure-based drug design using 3D equivariant generative models, showcasing cutting-edge approaches in molecular modeling. The day concluded with a practical skills workshop by Dr. Edvin Fako, who guided attendees through the use of Blender for creating scientific figures and graphics step-by-step.



Dr. Edvin Fako demonstrating how Blender can be used for producing publication-grade scientific graphics

The second day, April 24 focused on applied AI in toxicology and autonomous systems. Dongying Li from the U.S. FDA delivered a lecture on New Approach Methodologies (NAMs) for predicting liver and cardiac toxicity. This was followed by a hands-on session led by Hyun-Kil Shin from KIT, where participants explored the use of small language models (sLLMs) to build autonomous agents.



Dr. Dongying Li of the FDA discusses her team's work on predicting drug toxicity

The afternoon was dedicated to scientific communication, with scientific writing expert Ms. Ann Bless conducting a workshop on effective scientific writing tailored for PhD students. The day ended with a group dinner at Restaurant Chalet Suisse, offering a relaxed setting for networking and informal discussions.

The final day of the first week, April 25, featured a series of lectures addressing real-world challenges and advanced modeling techniques. Prof. Artem Cherkasov from VPC discussed AI strategies for CACHE hit-finding experiments, followed by Igor Tetko's overview of the Tox24 Challenge. Dr. Julian Cremer from Pfizer concluded the lecture series with insights into diffusion models and geometric deep learning.The afternoon was reserved for DC presentations, where each DC showcased recent progress made within their projects. This provided a platform for peer feedback and highlighted the diverse approaches being explored within the AiChemist consortium.

The second week – the AIChemist CECAM Flagship School, brought together a diverse group of researchers and students to explore the intersection of artificial intelligence (AI), machine learning (ML), and chemical sciences. The Flagship School attracted a total of 180 participants (63 onsite, 117 online, 24 speakers) from all over the world, and the majority of the lectures were run in hybrid mode and/or recorded.

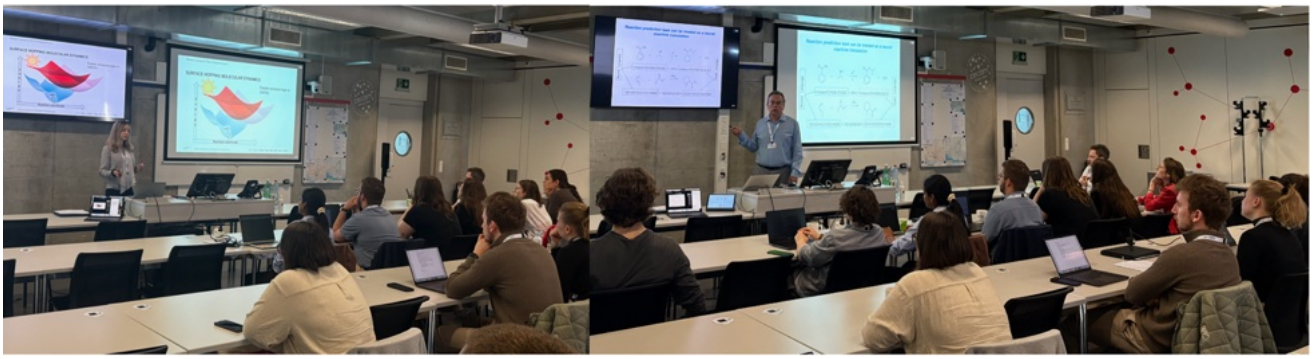
A central theme of the school was the **prediction and optimization of chemical reactions** using machine learning. Lectures covered a wide range of topics, from retrosynthesis and reaction condition optimization to the use of large-scale reaction databases and generative models. Tools such as AiZynthFinder and web-based modeling and visualisation applications were introduced, enabling participants to explore multistep synthesis pathways interactively. Another major focus was on **machine learning potentials (MLPs)** and their application in molecular simulations.

Several talks highlighted how MLPs can be trained on quantum mechanical data to simulate complex chemical systems with high accuracy and efficiency. These methods were applied to study reaction mechanisms, solvation effects, and excited-state dynamics. **Large language models (LLMs)** and **transformers** were another highlight, with presentations demonstrating their use in encoding chemical knowledge, predicting molecular properties, and even controlling robotic synthesis platforms. The integration of LLMs with cheminformatics tools showcased the potential of AI to automate and accelerate the drug discovery pipeline. Several lectures also touched on improving **model interpretability and explainable AI (XAI),** a critical area for ensuring the interpretability and trustworthiness of AI models in chemistry. Hands-on sessions complemented the lectures, offering practical experience with retrosynthesis tools, transformer models, and multi-task learning for compound profiling using QSAR.

Day 1: Prof. Heather Kulik (left) discusses how experimental data can be leveraged with ML for material/catalyst discovery and Prof. Jean-Louis Reymond (right) gives a lecture on expanding chemical reaction space with generated databases.

Day 2: Dr. Paula Torren Peraire (left) gives a talk on novel convergent synthetic route prediction strategies and Prof. Fernanda Duarte (right) speaks about the status of ML interaction potentials for modelling chemical reactions.

Day 3: Dr. Marwin Segler (left) gives participants an overview of concepts in synthesis planning and generative molecular design and Dr. Janani Durairaj (right) delivers a lecture on deep learning for protein-ligand structure prediction.

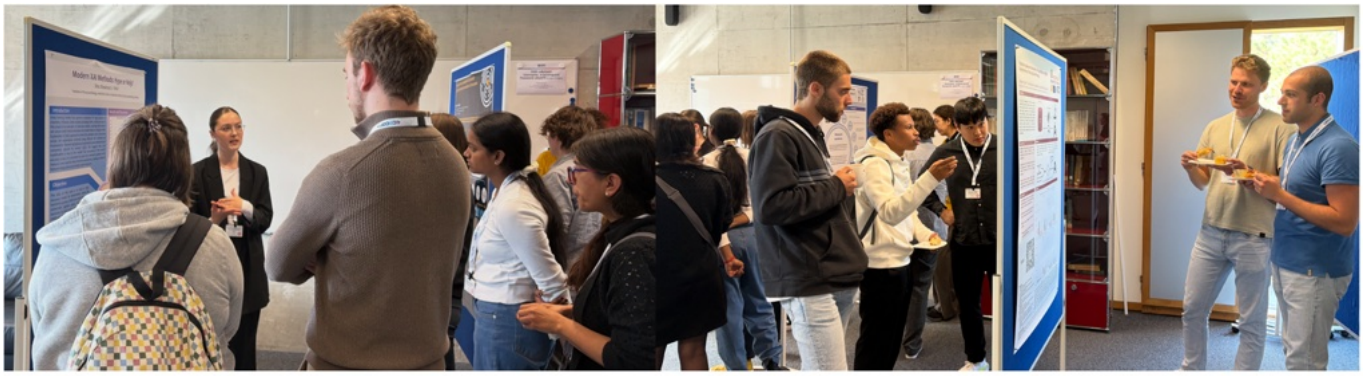


Day 4: Dr. Julia Westermayr (left) discusses applications of ML potentials for modelling excited states and Dr. Igor Tetko (right) gives a talk (jointly with Dr. Luc Patiny) on consensus modelling of reaction pathways.

Day 5: Dr. Kevin Jablonka (left) gives a lecture and hands-on session on the use of transformers for chemistry and materials science and Prof. Philippe Schwaller (right) provides participants with an overview of developments in LLMs for chemistry.

The poster sessions, which took place on Monday the 28th of April and Thursday the 1st of May showcased a diverse, innovative range of research at the intersection of AI and chemistry. Key themes included advancements in **retrosynthetic planning**, as well as the **integration of quantum mechanical data with ML for reaction feasibility prediction** and charge assignment.

Several contributions focused on **XAI**, and others explored topics in drug discovery such as **personalized cardiotoxicity prediction** using pharmacovigilance data and **p**rivacy risks in neural network publication**.** Environmental applications were also featured, including the in silico design of enzymes for PFAS degradation and the development of nano-QSAR models for nanoparticle toxicity prediction.



Poster sessions: AiChemist students took the valuable opporunity to network and present their work to invited speakers (many of whom are world-leading experts in ML/AI for chemistry and materials science), AiChemist PIs and the invited external participants.

The CECAM Flagship School concluded with a lively round-table debate on the experiences of women in ML/AI. The five key take-home messages from the round-table were:

- To increase women's representation in ML/AI, computer science must be promoted in school for young girls. Many talented female computer scientists/ML researchers would have never considered it as an option if it they weren't given the opportunity to explore it early on.
- AI-driven systems must be examined carefully for sources of bias (i.e. in recruiting, in finance, healthcare) to identify more complex causal links so they can be tackled effectively e.g. mobility from countries like Russia, Iran, India, Middle East where women's movement/education might be restricted.
- We must try to initiate discussions about gender gaps and inclusion across all levels of society, with people of all education levels.
- Political action within the field is necessary – make suggestions to management in your institutions, speak to your local government representatives, share your concerns with them i.e. ask to provide childcare, ask to make hiring more transparent etc.
- Men in the field – call out male friends/colleagues if/when they display discriminatory or inappropriate behaviour and push for consequences for such behaviour.

Overall, the scientific take-home message of the AIChemist CECAM Flagship School was that AI/ML is transforming synthesis, reactivity, and molecular design, while redefining tools and methods available to chemists. Participants explored retrosynthesis, MLPs, LLMs, and GNNs, gaining insights into interpretability, data quality, and domain knowledge for building trustworthy AI models, and real-world use cases. Chemistry's future lies in blending data-driven innovation with expert insight, to tackle complex challenges with greater precision, efficiency, and creativity.

The conference booklet, including the full schedule of talks, list of speakers and abstracts (for talks and posters) and the participant list is attached to this deliverable. Video recordings of almost all lectures will be made available on the Lhumos Learning Hub (https://alpha.lhumos.org) in the coming weeks. We are also currently in the process of uploading a selection of lecture recordings to the AiChemist Youtube Channel (https://www.youtube.com/@AiChemist-MSCA-DN). Just 8 out of 24 speakers did not consent to having their lectures/sessions recorded.

Voices from the Lausanne School

The Lausanne school, combined with the CECAM Flagship School, was a fantastic opportunity to connect with fellow PhD students and top researchers and professors in the field. It fostered valuable discussions, knowledge exchange, and collaborations. The program offered both academic depth and meaningful networking opportunities. One particularly interesting topic was the exploration of how knowledge encoded in large language models can be leveraged to improve predictions in drug discovery, which holds significant promise for future research. It was an inspiring and productive environment for advancing research and building connections.

- Fabian Krüger, AiChemist Fellow

I thoroughly enjoyed the AIChemist CECAM Flagship School – I found the talks on machine learning potentials particularly engaging, as this is the focus of my research. It was really exciting to meet both academics and students working on similar challenges—this was the first time I've felt part of a wider community with shared goals and a collective enthusiasm for the future of the field. This sense of connection and perspective is my main takeaway from the school: a refreshing look at how other research groups approach similar problems, and a clearer view of what the open questions in the field are. I also valued the opportunity to explore broader topics such as retrosynthesis and protein structure prediction—areas I've been curious about for a while. The sessions offered just enough depth to spark further interest, especially as structure prediction, building on methods like AlphaFold, is a rapidly evolving area of research. The EPFL campus was beautiful, and the week was exceptionally well organised; by the end I found myself sad to be leaving this little bubble of likeminded researchers.

- Imogen Daisy Smith, PhD Student at University College London

Upcoming Conferences

- ICANN2025, 9th – 12th September 2025, Kaunas, Lithuania. Igor Tetko, Eric Alcaide and Michael Wand.
- Datafest Yerevan, 12th – 13th September 2025, Yerevan, Armenia. Igor Tetko.
- EUROPIN Summer School on Drug Design, 14th – 18th September, Vienna, Austria. Igor Tetko.
- COLM2025, 17th-10th October 2025, Montreal, Canada. Eric Alcaide.
- The 28th European Conference on Artificial Intelligence, 25th-26th October 2025, Bologna, Italy. Dina Khasanova.
- **Fourth AiChemist School, 4th – 11th March 2026, Paris, France.**

Publications and pre-prints from the AiChemist Consortium

(1) Hartog, P.; Krüger, F .; Genheden, S.; Tetko, I.V. Using test-time augmentation to investigate explainable AI: inconsistencies between method, model and human intuition. Journal of Cheminformatics. 2024. https://doi.org/10.1186/s13321-024-00824-1
(2) Kopp, A.; Hartog, P.; Šícho, M.; Godin, G.; Tetko, I. The openOCHEM consensus model is the best-performing open-source predictive model in the First EUOS/SLAS Joint Compound Solubility Challenge. SLAS Discovery. 2024. https://doi.org/10.1016/j.slasd.2024.01.005
(3) Stocco, F., Artigues-Lleixa, M., Hunklinger, A., Widatalla, T., Guell, M., Ferruz, N. Guiding Generative Protein Language Models with Reinforcement Learning. arXiv. 2024. https://arxiv.org/abs/2412.12979.
(4) Krüger, F.P., Östman, J., Mervin, L., Tetko, I.V., Engkvist, O. Publishing Neural Networks in Drug Discovery Might Compromise Training Data Privacy. *J. Cheminformatics*. 2025. https://doi.org/10.1186/s13321-025-00982-w
(5) Hunklinger, A., Ferruz, N. Toward the explainability of Protein Language Models for Sequence Design. *arXiv*. 2025. https://doi.org/10.48550/arXiv.2506.19532
(6) Cirino, T., Pinto L., Iwan, M. *et al*. Consensus Modeling Strategies for Predicting Transthyretin Binding Affinity from Tox24 Challenge Data. *Chem. Res. Toxicol*. 2025. https://doi.org/10.1021/acs.chemrestox.5c00018
(7) Eytcheson, S. and Tetko, I.V. Which modern AI methods provide accurate predictions of toxicological endpoints? Analysis of Tox24 challenge results. *arXiv*. 2025. https://doi.org/10.26434/chemrxiv-2025-7k7x3-v3