The Explainable AI for Molecules “AiChemist” Project

Newsletter 5 - April 2026

Fourth AiChemist School – Paris Spring School (4th –11th March, 2026)

Hosted by ENS-PSL (Rue Lhomond) and Sanofi Vitry-sur-Seine

The 2026 AiChemist Spring School took place in Paris from 4 –11 March 2026, bringing together the AiChemist Doctoral Candidates, invited speakers, and industry partners for an intensive week of scientific training, interdisciplinary exchange, and exposure to cutting-edge developments in AI-driven chemistry, structural biology, molecular modelling, and computational toxicology. As in previous schools, the programme combined **lectures, workshops, institutional visits, and DC presentations**, fostering knowledge exchange and strengthening collaboration within the Doctoral Network.

Events were hosted mainly at **ENS-PSL (Rue Lhomond)**, with an industry day at **Sanofi Vitry-sur-Seine**, as well as a cultural visit at the **Musée des Arts et Métiers**. The majority of lectures were run in hybrid mode, with the online attendance option being offered to the wider AI/drug discovery/toxicology community via LinkedIn. Fourty online participants registered their interest and on average, ten participants joined online each day.

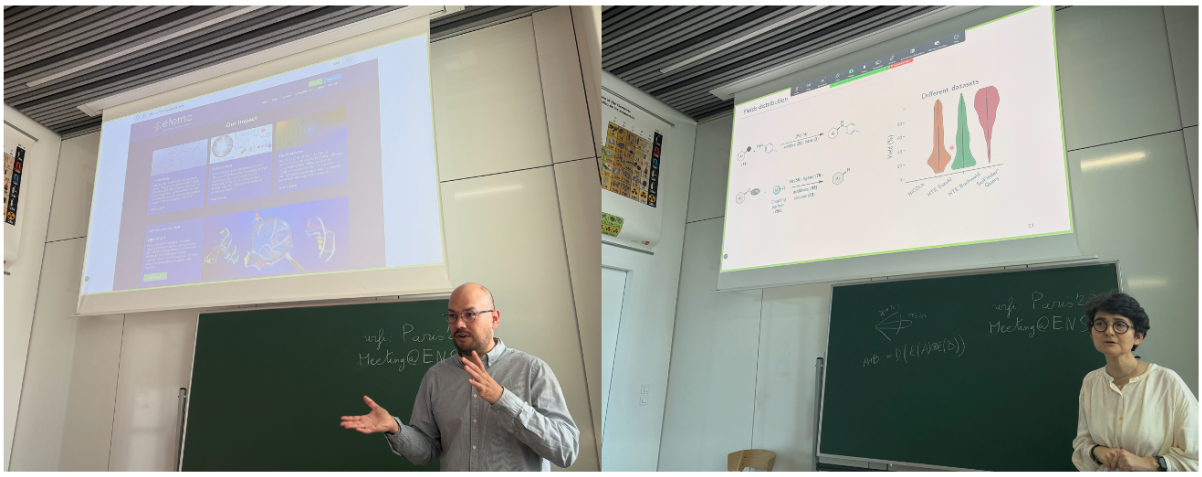
Day 1: Wednesday, 4 March (ENS-PSL)

The Spring School opened with a strong focus on **generative approaches in biology and chemistry**.

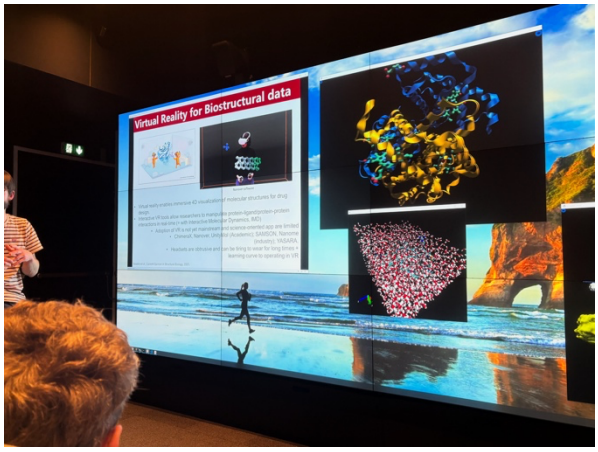
- Associate Professor Dr. Philippe Nghe (ESPCI Paris) kicked off the School with an introduction to the generative design of ribozymes, setting the stage for discussions on data-driven molecular design.
- Dr. Mikhail Kabeshov (AstraZeneca) followed with an overview of AI tools for chemistry being developed at AstraZeneca, enabling accelerated design, parallel synthesis, and direct-to-profile experimentation.
- Dr. Adam Arany (KU Leuven) gave a talk on the fundamentals of federated machine learning and its associated pitfalls and challenges, covering distribution shifts and data leakage issues relevant to multi-partner drug discovery collaborations.

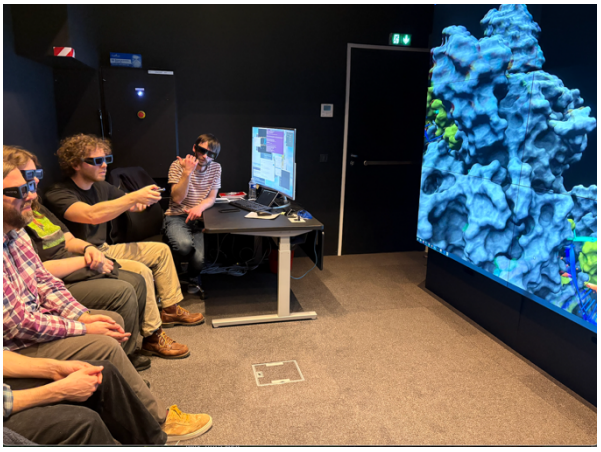
In the afternoon, Prof. Laurence Grimaud (ENS-PSL) discussed ML-based optimization in synthetic chemistry, followed by Dr. Eugenie Romero (CEA), who presented HTE as a flexible, accessible and empowering platform for chemists. The day concluded with a guided visit to the Visualisation Wall at the IBPC Institute (hosted by Dr. Marc Baaden) and a cosy welcome reception at ENS-PSL.

ENS-PSL Department of Physics and Chemistry at 24 rue Lhomond, Paris



Prof. Philippe Nghe (left) and Prof. Laurence Grimaud (right) delivering lectures on Day 1





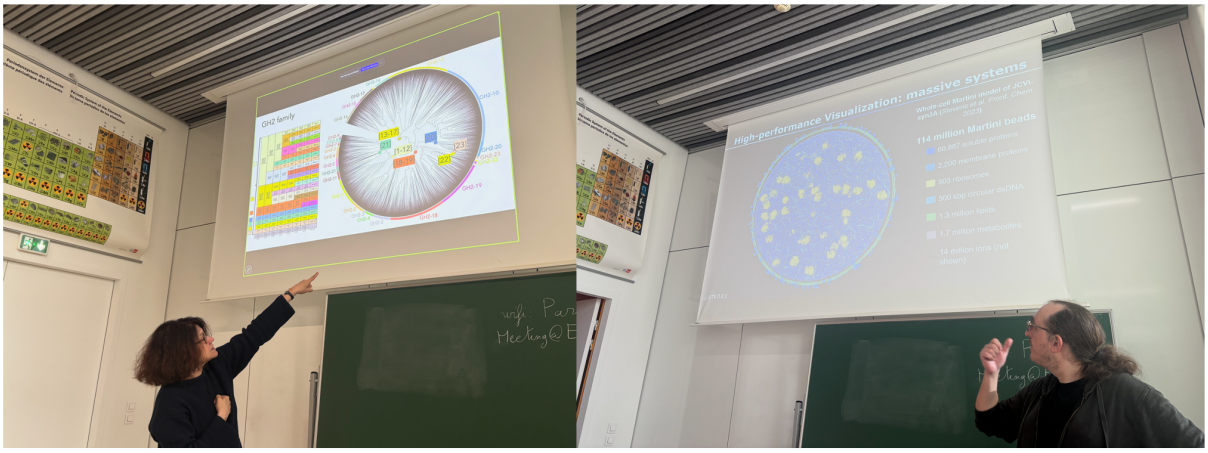
The DCs exploring the functionality of the Visualisation Wall at the IBPC Institute

Day 2: Thursday, 5 March (ENS-PSL)

The second day turned toward **structural biology and integrative modelling**:

- Prof. Mathieu Montes (CNRS, Sorbonne) presented the VTX software suite and the latest developments in interactive molecular visualization at his lab.
- Dr. Max Bonomi (CNRS, Institut Pasteur) spoke on Integrative Structural Biology in the AI era.

Afternoon sessions highlighted AI-driven protein studies, with Prof. Alessandra Carbone (Sorbonne University) discussing functional signal decoding. The day also included the first round of DC presentations (DCs 1–8), giving candidates the opportunity to present their ongoing research and receive feedback from peers and senior researchers.



Prof. Alessandra Carbone (left) discussing functional signal decoding and Prof. Matthieu Montes providing an overview of applications of the VTX suite (right)

Day 3: Friday, 6 March (ENS-PSL)

Friday offered a broad perspective on **AI in toxicology, chemical language models, and hybrid computational chemistry**:

- Dr. Kamel Mansouri (NICEATM, NIH) discussed applications of AI in toxicology and the democratization of computational tools.
- Prof. Francesca Grisoni (TU/e) gave an introduction to chemical language modelling for molecule discovery.
- Junior Prof. Thijs Stuyver (PSL) discussed hybrid ML–quantum chemistry workflows for reaction screening and novel reaction discovery.

Prof. Thijs Stuyver exploring the applications of hybrid ML-QM workflows for reaction discovery (left) and Prof. Francesca Grisoni (right) giving an overview of the potential of chemical language modelling for molecular discovery

In the afternoon, invited speaker Dr. Salvo Camiolo, Site Manager and Head of Bioinformatics at BioClavis Ltd, delivered a session on applications of AI in transcriptomics for safety and diagnostics. The day closed with presentations by DCs 9,10,11 and 13. As in prior Schools, DC presentations formed a key pillar of the programme, allowing candidates to showcase their research progress, receive input from network partners, and refine their communication skills. The atmosphere was

constructive and collegial, with senior researchers providing tailored feedback to support the DCs scientific and professional development.

DC1 Fabian Krüger and DC2 Dina Khasanova updating the network on their progress within their projects

Day 4: Saturday, 7 March

Participants joined an optional cultural activity: a guided afternoon visit to the **Musée des Arts et des Métiers**. The event offered an opportunity for exploration, culture and informal exchange within the network in a historical scientific setting.

Lavoisier's Laboratory, in which the chemical composition of water was discovered during the 18th century (left) and Zeus, the mechanical horse designed for the Paris Olympics opening ceremony (sponsored by AiChemist partner Sanofi) (right) at the Musée des Arts et des Métiers.

Day 5: Monday, 9 March (Sanofi, Vitry-sur-Seine)

The industry day at **Sanofi** provided exposure to R&D pipelines and industrial research workflows:

- The morning began with a tour of Sanofi facilities, including the histology, cryo-EM and medicinal chemistry labs.
- Dr. Christelle Perrault (Sanofi) provided an in-depth introduction to high throughput biology, from screening to hit identification.
- Dr. Marc Bianciotto (Sanofi) discussed data curation and dose-response curve prediction using the Tox21 dataset.
- After lunch, Dr. Alexey Rak (Sanofi) presented on biophysical technique validation, followed by a talk by Dr. Evi Gkeka (Sanofi) on computational hit finding in industry settings.
- Prof. Gabriel Stoltz (Ecole des Ponts) provided a rigourous mathematical perspective on autoencoders and importance sampling.
- The day at Sanofi concluded with a second talk by Dr. Kamel Mansouri, on the impact of international computational toxicology collaborations.

Dr. Christelle Perrault (left) and Dr. Marc Bianciotto (right) kicking off the lectures at Sanofi, Vitry-sur-Seine

Group photo taken during the day at Sanofi

Day 6: Tuesday, 10 March (ENS-PSL)

The penultimate day focused on **communication, regulatory science, and chemogenomics**:

- A morning workshop led by Silvano Coletti and Emanuele Sana of Chelonia Life Sciences covered dissemination, communication, and exploitation strategies, with a talk on publishing in high-impact journals from invited guest speaker, Markus Kaindl, Director of Content Innovation, Research Publishing at Springer Nature.
- Dr. Salvador Moncho Escrivà (ECHA) then presented on QSAR prediction assessment for regulatory applications.
- Afternoon talks covered topics including Chemogenomics (by Prof. Olivier Taboureau, Paris Diderot University), MCTS for chemistry and biology (by Prof. Tristan Cazenave, Université Paris Dauphine), and regulatory submissions using in silico predictions (by Dr. Marco Evangelista, CEHTRA).

The day concluded with a presentation of the IPPI-DB (online database of modulators of protein-protein interactions) and its applications by Prof. Olivier Sperandio (Institut Pasteur), followed by a network dinner at a quaint and cosy restaurant, Papy aux Fourneaux, located in the Montparnasse area.

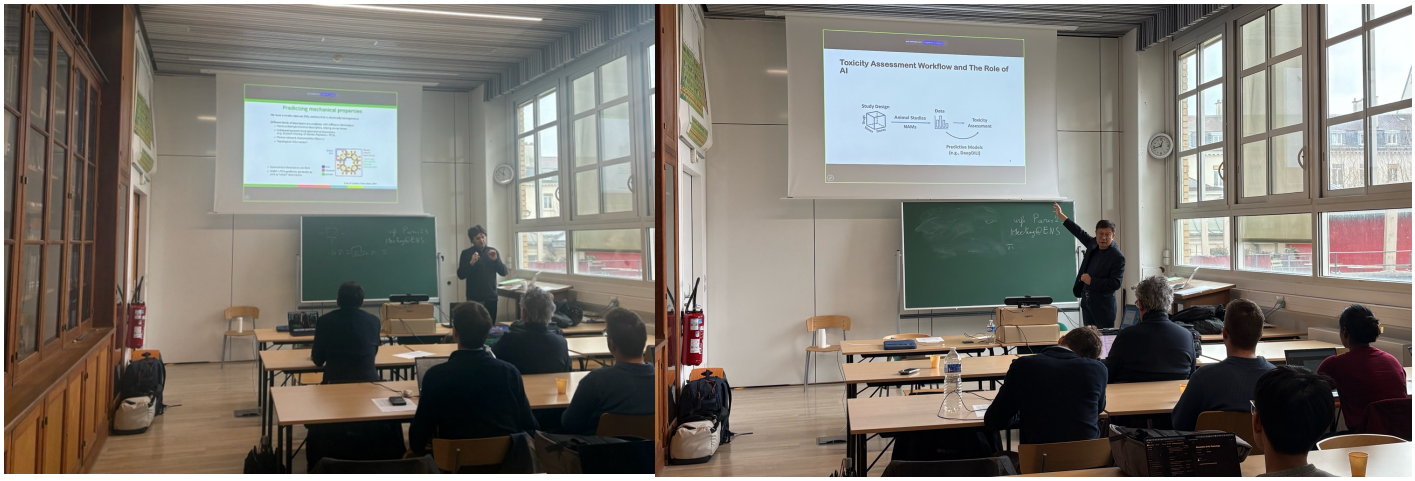
Markus Kaindl (left) giving a short talk on the golden rules of academic publishing and Prof. Olivier Sperandio introducing the IPPI-DB initiative (right)

Networking Dinner at Papy aux Fourneaux

Day 7 — Wednesday, 11 March (ENS-PSL)

The final day featured sessions on **RNA technologies and materials discovery**:

- Prof. Olivier Taboureau gave a comprehensive overview of the fundamentals of RNASeq.
- Dr. Yann Ponty (Ecole Polytechnique) discussed constrained RNA design and inverse folding.
- Dr. François-Xavier Coudert (PSL) presented on accelerating materials discovery using computational approaches.
- Dr. Weida Tong (U.S. FDA) closed the scientific programme with a talk on generative AI for toxicology.



Dr. François-Xavier Coudert and Dr Weida Tong lecturing on the final day

A brief closing session, with many rounds of applause for everyone that was involved in the organisation efforts, concluded the School before the participants departed.

The Paris Spring School successfully delivered advanced training in several core areas:

1. AI-driven molecular design and chemistry

Participants deepened their understanding of generative models, chemical language modelling, hybrid ML-physics workflows, and automated synthesis strategies.

2. Structural biology techniques and integrative modelling

Sessions covered interactive visualization, integrative modelling pipelines, and signal interpretation with AI tools.

3. Toxicology and regulatory science

Candidates gained exposure to computational toxicology, transcriptomics-driven safety assessment, QSAR frameworks, and the regulatory context surrounding predictive modelling.

4. Industry engagement and experimental methods

The Sanofi visit and talks on experimental methods in drug discovery (such as RNA-seq/TempOSeq, transcriptomics, cryo-EM, etc) highlighted practical constraints, high-throughput chemistry and biology pipelines, and industrial standards, giving DCs insight into future career paths.

Workshops and discussions emphasized scientific outreach, exploitation planning, and effective messaging tailored to diverse stakeholders. Overall, the School strengthened

the scientific cohesion of the AiChemist network, and supported the DCs in building a well-rounded training profile.

This Spring School continues the tradition of excellence established in previous AiChemist training events and significantly contributes to the progression of the DCs toward independent, multidisciplinary research careers.

Voices from the Paris School

“I had a great time at the AiChemist Paris School this year. Besides the many sceneries and museums Paris has to offer, I had the chance to connect with fellow PhD students, sharing perspectives on not only the subjects of our research, but also on how research is going to be done in the future, emphasizing the emergence of agentic coding. As the latest addition to the AiChemist cohort, I was very proud to showcase my recent work, "Synthelite: Chemist-aligned and feasibility-aware synthesis planning with LLMs", where we design an environment for LLMs to plan retrosynthesis. We show that the powerful reasoning capability and chemical knowledge embedded in LLMs enables a new paradigm in retrosynthesis planning that goes beyond conventional tree search, offering strategic elements and chemical feasibility in synthesis routes. Additionally, I had a great time with the guest lectures and the visit to Sanofi. I was particularly intrigued by the talk given by Prof. Francesca Grisoni from TU/e, on how clever it is to use Task Arithmetic on generative AI to inverse the properties of sampled molecules, allowing the extension of training data to even negative examples. Leaving the school with a bittersweet feeling but also with a lot of motivation, I already look forward to the next edition.” *Xuan Vu Nguyen (DC13), EPFL*

“The Paris School was a wonderful opportunity to reconnect with fellow AiChemist colleagues and engage with experts from both academia and industry. The programme covered various topics, with a strong focus on AI and machine learning for chemistry and drug discovery, toxicology and regulatory applications, and structural biology. I particularly enjoyed Gabriel Stoltz's deep dive into the mathematical foundations of autoencoders and François-Xavier Coudert's talk on accelerating materials discovery from crystal structures. The visit to Sanofi's facilities in Vitry-sur-Seine was another highlight. Besides the interesting talks, we had the opportunity to see their Cryo-EM suite, histology and med-chem labs, which was a good reminder of the amount of experimental work behind the data we handle computationally. Beyond the science, the School was a great chance to catch up with colleagues over dinner and evening walks along the charming streets and riverbanks of the Seine, and, as with every AiChemist event, an excellent opportunity to sightsee.” *Mateusz Iwan (DC6), Bayer*

Upcoming Events and Conferences

- ICLR 2026, 22nd – 27th April 2026, Rio de Janeiro, Brazil. Eric Alcaide, presenting "MULTIPLE TOKEN DIVERGENCE: MEASURING AND STEERING IN-CONTEXT COMPUTATION DENSITY"
- Bayer Pharma R&D conference, 18th – 20th May 2026, Berlin, Germany. Andi Hunklinger, presenting poster "SEISMO: Sample-Efficient Inference-Stage Molecular-Optimization agent"
- AI to Accelerate Scientific Understanding Workshop, 26th – 29th May 2026, Berlin, Germany. Andi Hunklinger, presenting poster "Faithful XAI for protein language models"
- Madics/EXMIA Symposium, 2nd – 3rd June 2026, Avignon, France. Marc Bianciotto, giving keynote talk "AI, Explainability and The Two Jobs of the Drug Designer".
- Trustworthy AI in Life Sciences, 24th – 28th August, Saarbrücken, Germany. Andi Hunklinger (attending).
- Central and Eastern European Cheminformatics Meeting, 6th September 2026, Prague, Czechia. Igor Tetko (attending, possibly presenting).
- International Conference on Artificial Neural Networks ICANN2026, 14th – 17th September 2026. Co-organised by Igor Tetko, featuring the 3rd AIDD Workshop organised by the whole AiChemist consortium.

Publications

(1) Tetko, I. Tox24 Challenge. *Chem. Res. Tox* **2024**. https://doi.org/10.1021/acs.chemrestox.4c00192

(2) Stocco, F.; Artigues-Lleixà, M.; Hunklinger, A.; Widatalla, T.; Güell, M.; Ferruz, N. Guiding Generative Protein Language Models with Reinforcement Learning. *arXiv* **2025.** https://doi.org/10.48550/arXiv.2412.12979.

(3) 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.

(4) Goldstein, D.; Alcaide, E.; Lu, J.; Cheah, E. RADLADS: Rapid Attention Distillation to Linear Attention Decoders at Scale. *arXiv* **2025.** https://doi.org/10.48550/arXiv.2505.03005.

(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. A.; Tetko, I. V. Which Modern AI Methods Provide Accurate Predictions of Toxicological Endpoints? Analysis of Tox24 Challenge Results. *Chem. Res. Tox*. **2025**. https://doi.org/10.1021/acs.chemrestox.5c00273.

(8) Tetko, I.V.; Clevert, D.A. Advanced Machine Learning for Innovative Drug Discovery. *J. Cheminform.***2025**. https://doi.org/10.1186/s13321-025-01061-w
(9) Ball, M.; Horvath, D.; Kogej, T.; Kabeshov, M.; Varnek, A. Predicting Reaction Conditions: A Data-Driven Perspective. *Chem. Sci.* **2025**. https://doi.org/10.1039/D5SC03045E.
(10) Krüger, F. P.; Österbacka, N.; Kabeshov, M.; Engkvist, O.; Tetko, I. MolEncoder: Towards Optimal Masked Language Modeling for Molecules. *ChemRxiv.* **2025**. https://doi.org/10.26434/chemrxiv-2025-h4w9d.
(11) Singh, I.; Onogi, Y.; Menezes, F.; Khasanova, D.; Kang, L.; Wang, C.; Ruiz-Trave, J.; Sharma, S.; Khalil, A.; Reichenbach, V. K.; Shi, Y.; Flatley, A.; Yan, X.; Israel, A.; Dragano, N. R. V.; Aguilar-Pimentel, J. A.; Hoffmann, A.; Ghosh, A.; Noé, F.; Wolfrum, C.; Cucuruz, S.; König, A.-C.; Burtscher, I.; Hauck, S. M.; Lickert, H.; Hofmann, S. M.; Feederle, R.; Schriever, S. C.; Hernandez-Bautista, R.; Sancar, G.; Cebrian-Serrano, A.; Tetko, I.; Fuchs, H.; Gailus-Durner, V.; Blüher, M.; Hrabě de Angelis, M.; Ussar, S. NRAC Controls CD36-Mediated Fatty Acid Uptake in Adipocytes and Lipid Clearance in Vivo. *EMBO J.* **2025.** https://doi.org/10.1038/s44318-025-00520-2.
(12) Krüger, F. P.; Österbacka, N; Kabeshov, M.; Engkvist, O.; Tetko, I. MolEncoder: Improved Masked Language Modeling for Molecules. *International Conference on Artificial Neural Networks* **2025.** https://link.springer.com/chapter/10.1007/978-3-032-04552-2_6
(13) Krüger, F. P.; Österbacka, N; Kabeshov, M.; Engkvist, O.; Tetko, I. MolEncoder: Improved Masked Language Modeling for Molecules. *Digital Discovery* **2025.** https://pubs.rsc.org/en/content/articlelanding/2025/dd/d5dd00369e
(14) Park, J.; Rashid, S.; Copsey, H.; Tran, L. ;Zabeo, A.; Hristozov, D.; Gakis, G. P.; Charitidis, C.; Yoon, S.; Shin H.K. NanoToxRadar: A Multitarget Nano-QSAR Model for Predicting the Cytotoxicity Values of Multicomponent Nanoparticles. *ACS Nanoscience Au.* **2025.** https://doi.org/10.1021/acsnanoscienceau.5c00035
(15) Viganò, E. L.; Iwan, M.; Colombo, E.; Ballabio, D.; Roncaglioni, A. Mixture of experts for multitask learning in cardiotoxicity assessment. *J. Cheminformatics* **2025.** https://doi.org/10.1186/s13321-025-01072-7
(16) Roncaglioni, A.; Kovarich, S.; Mansouri, K.; Tetko, I.V. Advancing Human and Environmental Safety Science Using *In Silico* Methods. *Chem. Res. Tox* **2025.**
(17) Alcaide, E.; Gao, Z.; Ke, G.; Li, Y.; Zhang, L.; Zheng, H.; Zhou. G. Uni-Mol Docking V2: Towards Realistic and Accurate Binding Pose Prediction. *International Conference on Artificial Neural Networks* **2025.** https://link.springer.com/chapter/10.1007/978-3-032-04552-2_5
(18) Hiltscher, M.; Bianciotto, M.; Grisoni, F. Explaining What Matters: Faithfulness in Molecular Deep Learning. *Chemrxiv* **2026.** https://chemrxiv.org/doi/full/10.26434/chemrxiv.10001594/v1
(19) Krüger, F. P; Hunklinger, A.; Wolny, A.; Adler, T.J.; Tetko, I.; Villalba, S.D. SEISMO: Increasing Sample Efficiency in Molecular Optimization with a Trajectory-Aware LLM Agent. *arXiv* **2026**. https://arxiv.org/abs/2602.00663
(20) Iwan, M.; Roncaglioni, A.; Grisoni, F. Beyond Molecular Structure: Investigating Demographic Factors in Drug-Induced Cardiotoxicity Prediction Models. *Chemrxiv* **2026**. https://chemrxiv.org/doi/full/10.26434/chemrxiv.10001782/v1

(21) Ball, M.; Faber, F.A.; Pratley, C.; Kogej, T.; Horvath, D.; Varnek, A.; Bergonzini, G.; Jäger, C.M.; Kabeshov, M. Expert-Informed Contrastive Learning of Condition Space for Amide Coupling
Reactions. *Chemrxiv* **2026**. https://chemrxiv.org/doi/full/10.26434/chemrxiv.15001054/v1
(22) Xuan-Vu, N.; Armstrong, D.; Wehrbach, M.; Bran, A.M.; Jončev, Z.; Schwaller, P. Synthelite: Chemist-aligned and feasibility-aware synthesis planning with LLMs. *arXiv* **2025**. https://arxiv.org/abs/2512.16424

You can also check out our Google Scholar, which is more regularly updated.