The Explainable AI for Molecules (AiChemist) Project

Newsletter #3, February 2025

The Explainable AI for Molecules (AiChemist) project is a Marie-Skłodowska-Curie Doctoral Network (DN) funded by the European Commission under the Horizon Europe Programme, Horizon-MSCA-2022 grant agreement number 101120466, which started on 1st September 2023. The network brings together sixteen academic and industry partners from 8 European countries in addition to the Korean Institute of Toxicology (South Korea) to train fourteen doctoral candidates (DCs) in close collaboration with associated partners based in Europe and the USA.

Follow us on Twitter: https://x.com/aichemist_dn
LinkedIn: https://www.linkedin.com/company/98394761/admin/dashboard/
BlueSky: https://bsky.app/profile/aichemist.bsky.social

Second AiChemist School – Strasbourg Summer School in Chemoinformatics 2024

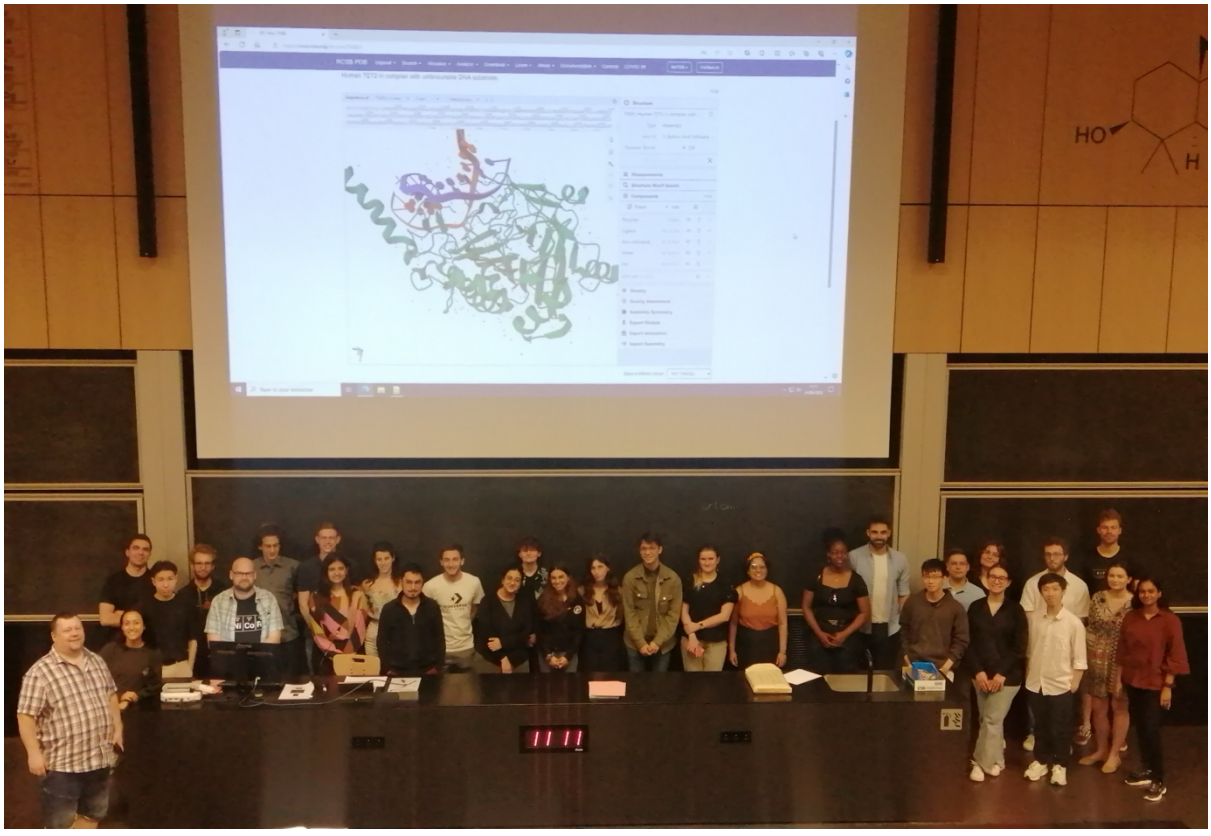
The second AiChemist School took place from the 21st to the 28th of June at the University of Strasbourg. The AiChemist School coincided with the 9th Strasbourg School in Chemoinformatics, a biennial School/Conference organized by the Centre National de Recherche Scientifique (CNR) and the University of Strasbourg. Several PIs of AIDD projects, some of whom students met during join AIDD-AiChemist school on Berlin, were invited lectors together with AiChemist PIs.

The programme covered the latest advancements in the field of chemoinformatics, and current trends in the application of AI and advanced simulation methods for drug discovery, delivered by leading industry and academic experts from Europe, the Americas, Japan and Israel. The full lecture programme, along with abstracts and video recordings can be found at https://infochim.chimie.unistra.fr/-Strasbourg-Summer-School-in-Chemoinformatics-2024-.html.

The first day of the school was devoted to a hackathon hosted in the Studium, a hub building of the University of Strasbourg. All participants, which included the Erasmus Mundus Chemoinformatics+ Masters students and the AiChemist fellows, were divided into 7 groups, which were led by the Chemoinformatics+ Masters Course lecturers. Each group worked on one of several open source software projects and data science activities related to chemical safety and drug design, namely:

- CMDock, a participative docking tool hosted by the University of Ljubljana (https://gitlab.com/Jukic/cmdock)
- Indigo, an widely-used chemoinformatics C++ API (https://lifescience.opensource.epam.com/indigo/)
- Ichem, another C++ API dedicated to proteins and binding sites (https://doi.org/10.1002/cmdc.201700505)
- Preparation and modeling of a dataset for chemical hazard assessment
- Study of TET enzymes through target oriented and ligand-oriented approaches

The hackathon participants were tasked with preparing datasets, models, correcting and documenting bugs, adding documentation and extending software functionality. The results of the hackathon have since been published in the GitLab (https://gitlab.com/CrtomirP/hackaton_cheminfo_2024) managed by the organisers of the Erasmus Mundus Chemoinformatics+ Masters Programme. All groups presented their work during a dedicated conference session in the morning of Monday the 24th of June – the official opening day of the 9th Strasbourg School in Chemoinformatics.

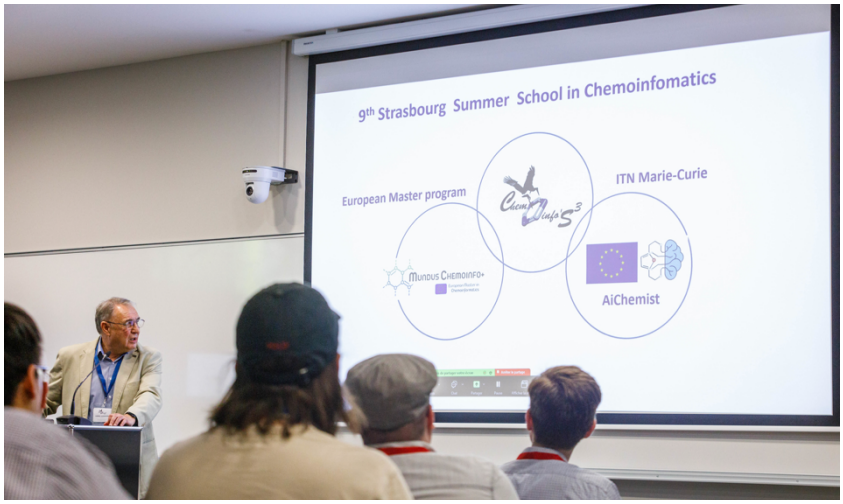


AiChemist fellows and Erasmus Mundus Chemoinformatics+ Masters students at the hackathon wrap-up session on Monday 24th June 2024

After registration, the AiChemist fellows attended the opening keynote lecture by Alexander Tropsha, titled "Reminiscing about the Future: New Trends in QSAR Modeling

in the Age of Big Data, Generative AI, and Large Language Models", which was followed by a warm welcome reception.

The second day of the conference started with a lecture on De Novo Molecular Design with Machine Intelligence by Gisbert Schneider, Professor in Computational Drug Design at ETH Zurich, which was followed up with a series of lectures on various topics in chemoinformatics ranging from enhanced chemical space exploration and using language models for drug design, to process informatics and leveraging AI for chemistry automation, delivered by leading experts in the field. The lecture program was rounded off with a hands-on tutorial on Generative Topographic Mapping led by Gilles Marcou of UNISTRA. After getting through the dense scientific schedule, the attendees had a chance to mingle and exchange their ideas and impressions over beer and pretzels during the evening poster session.



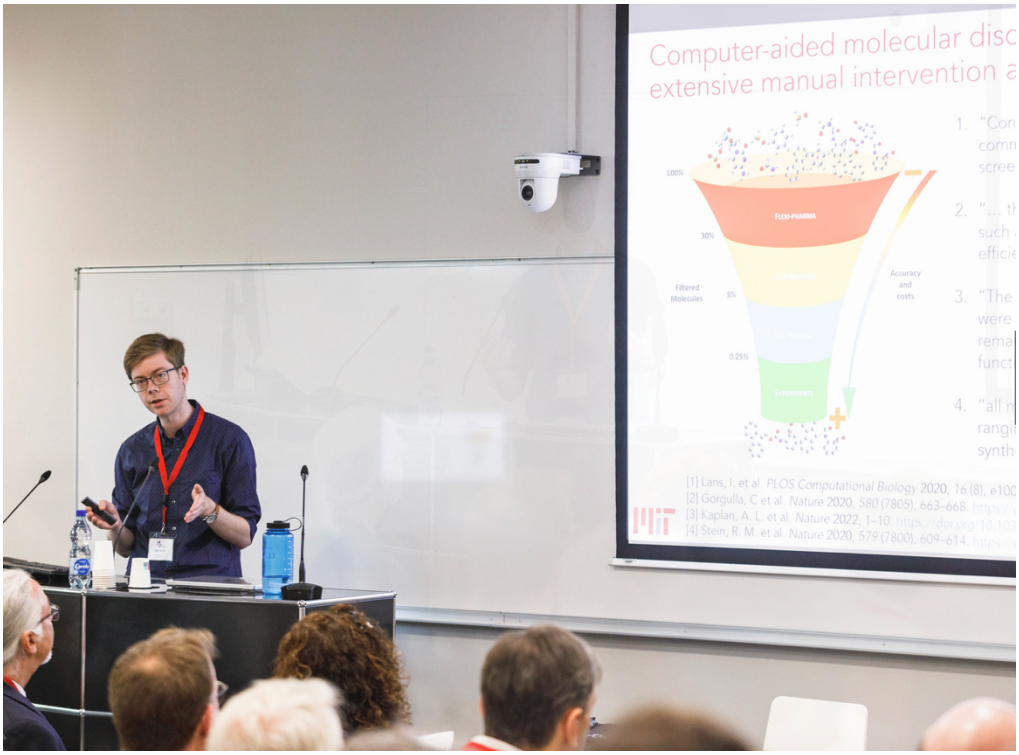
Prof. Alexandre Varnek (main organizer and AiChemist PI) welcomes attendees during the conference opening session

Prof. Alexander Tropsha from the UNC Eshelman School of Pharmacy delivering the opening lecture, video: https://www.youtube.com/watch?v=oj_kKetKUgo

Prof. Gisbert Schneider kicking off the second day of the conference with his talk on ML for de novo drug design

The third day of the conference featured a dynamic mix of cutting-edge research and practical applications in drug discovery and pharmacology. During the morning session, Thierry Langer from the University of Vienna presented innovative tools and applications for next-generation pharmacophore modeling, followed by MIT's Connor Coley, who discussed inventive approaches for balancing ML-driven molecular design and synthesis strategies. After a short coffee break, José L. Medina-Franco from the National Autonomous University of Mexico gave a fascinating overview of the chemical space of food chemicals and natural products from Latin America, while Alessandra Roncaglioni from the Mario Negri Institute highlighted the use of QSAR in regulatory toxicology.

The afternoon sessions offered a more practical edge, with Yuliana Zabolotna detailing strategic chemography for drug discovery at Eli Lily. Mireille Krier introduced OpenEye's advanced methods for detecting cryptic pockets, and Marcus Gastreich showcased cost-saving approaches to ultralarge space docking co-developed by BioSolveIT. The day concluded with a tutorial on AlphaFold2 applications by Esther Kellenberger and Luca Chiesa from the University of Strasbourg, before transition to the second poster session, where attendees socialized over wine and cheese. Day 3 not only emphasized collaboration across disciplines but also underscored the seamless integration of computational tools into modern pharmacology.

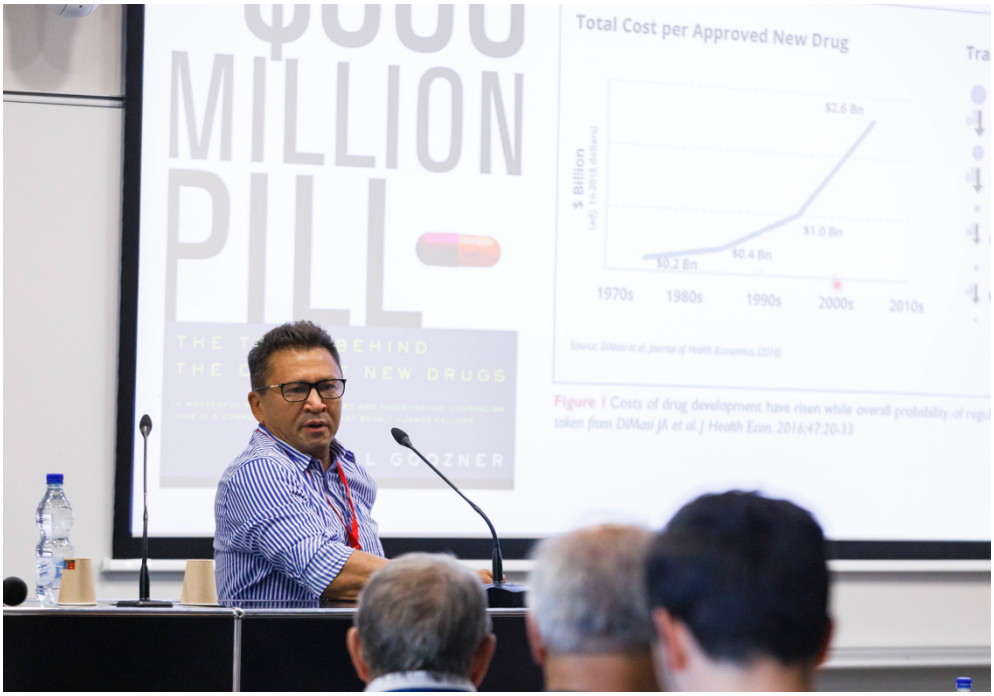


Dr. Connor Coley, Associate Professor at the Department of Chemical Engineering at MIT, gives a talk on balancing creative design and synthesizability in the DMTA cycle, video: https://www.youtube.com/watch?v=w_nnaaaXI-g&feature=youtu.be

Prof. Thierry Langer wrapping up his talk on next generation pharmacophore modelling, video: https://www.youtube.com/watch?v=3PDd3RZ-CQU&feature=youtu.be

AiChemist PI Dr. Alessandra Roncaglioni discussing the use of QSAR in environmental risk assessments, video: https://www.youtube.com/watch?v=oXbhkM4Q1BQ

The fourth day provided the attendees with rich blend of computational drug design and cultural immersion. The morning session featured a series of engaging talks, starting with Artem Cherkasov from the University of British Columbia, who discussed the integration of deep docking into virtual screening workflows, exemplified by a CACHE-1 case study. Satoshi Maeda from Hokkaido University followed with a presentation on leveraging ab initio chemical reaction design through automated pathway exploration, offering new perspectives on computational chemistry. After a short coffee break, Anne-Claude Camproux of Paris City University delved into structural analysis protocols for in silico drug design, before handing over to Guilherme Martins Silva from Harvard BIDMC, who presented his work on ultra-large virtual screens for multi-target anti-Alzheimer's agents. The session concluded with talks by Stefano Pieraccini from the University of Milan and Marco Jukić from the University of Maribor, focused on innovative approaches for protein immobilization and protein-ligand preparation with conserved waters, respectively.



Prof. Artem Cherkasov (one of the PIs of the AIDD project https://ai-dd.eu) explains how deep docking can be successfully integrated into virtual screening workflows

The afternoon provided an opportunity to step away from the technical and immerse in the cultural richness of Strasbourg through a carefully curated program. The AiChemist fellows and Erasmus Mundus Chemoinformatics+ students were invited to visit the European Directorate for the Quality of Medicines & Healthcare (EDQM - one of the many important European institutions located in Strasbourg. The EDQM is responsible for implementing the International Convention on the Elaboration of the European Pharmacopeia, defining the pre-requisites for the production and quality of medicines so that they can be authorized for the markets of its 39 signatory states and the European Union.

The visitors learned about the roles and responsibilities of EDQM, including the production of reference standards and the Ph. Eur. Monographs, providing the legal and scientific basis for the quality control of medicines, establishing certification procedures for the manufacturers of pharmaceutical substances and the coordination of the European Network of Official Medicines Control Laboratories (OMCL). At the end of the visit, the AiChemist and Erasmus Mundus students reconvened with the rest of the attendees in the evening for a celebratory conference dinner, where lively conversations flowed, collaborations were fostered, and the day's insights were reflected upon in the relaxed atmosphere of the quintessentially Alsatian Brasserie La Bourse.

AiChemist fellows and Erasmus Mundus students at the EDQM

AiChemist fellows Fabian Krüger (left) and Vasilii Fastovskii (right) mingling in between courses at La Bourse

On the morning of the 28th of June, the closing session of the 9th Strasbourg School in Chemoinformatics opened with a talk from Oleksandr Isayev from Carnegie Mellon University, who showcased artificial intelligence solutions for computational and organic chemistry tasks. This was followed by a talk by Pavel Polishchuk from Palacký University, who discussed multi-instance learning as a method to address the complexity of molecular entities, after which Didier Rognan of Strasbourg University

highlighted the use of structure-based organic chemistry-driven ligand design to explore ultra-large chemical spaces. The final talk was delivered by Hanoch Senderowitz from Bar Ilan University, focusing on materials informatics for green energy harvesting and storage. The session concluded with a formal closing ceremony at noon.

Prof. Olexandr Isayev giving an overview of AI solutions for computational chemistry and organic chemistry tasks developed at Carnegie Mellon University, video: https://www.youtube.com/watch?v=SvKYSH2_tR4&feature=youtu.be

Prof. Hanoch Senderowitz (associated partner of the AIDD https://ai-dd.eu project) speaking on the topic of materials informatics for green energy, video: https://www.youtube.com/watch?v=tstao0M9fCk

Student Voices from the Strasbourg School

"I had a fantastic time at the 9th Strasbourg Summer School in Chemoinformatics, held from June 24th to June 28th in the picturesque city of Strasbourg, France. This unique event brought together a vibrant mix of students, young researchers, and experienced scientists, creating an atmosphere of collaboration and inspiration.

The program was well-rounded, featuring plenary lectures, hands-on tutorials, oral presentations, and a lively poster session. Topics ranged from artificial intelligence in chemistry and big data analysis to chemical reaction mining and materials informatics, ensuring something for everyone.One of my highlights was the opportunity to learn directly from world-class scientists, such as Jürgen Bajorath, Gisbert Schneider, Alexander Tropsha, and many others. Their talks were both insightful and inspiring.

Meeting new friends and colleagues on the AIChemist Project was a pleasure. We exchanged many words and played numerous games. It provided an excellent opportunity to connect with everyone."

Vasilii Fastovskii, AiChemist fellow

Attending the Strasbourg Summer School 2024 was a valuable experience in many ways. It provided a chance to reconnect with members of the AiChemist consortium and get to know the new PhD hires. The hackathon at the beginning was particularly valuable, allowing me to reflect on the expertise I've developed over the years. It was a great chance to take on a more mentor-like role, guiding ChEMoinformatics+ Erasmus Mundus master's students who genuinely trusted my experience and insights, leading them from the start of a project to the final presentation in just one day. This not only helped me grow professionally, but it also reinforced the importance of fostering knowledge-sharing in a collaborative setting.

The summer school had a very summery feel to it, with the hot weather matching the "hot topics" we discussed. It was refreshing to see a blend of school and conference formats, including the hackathon, presentations, and poster sessions. The event had a democratic feel to it, free of pomp and hierarchy, which created an open and relaxed atmosphere that encouraged engaging discussions. I particularly appreciated the exchange with senior scientists, whose wealth of experience added depth to the conversations, and master's students, whose fresh perspectives brought new energy to the discussions. On a more personal note, I thoroughly enjoyed the city's bar culture, which became an added highlight of the experience. The relaxed atmosphere, especially during the men's European Football Championship, created a fun and laid-back environment where networking and casual conversations flowed naturally, adding an enjoyable balance to the more intellectually rigorous parts of the schedule.

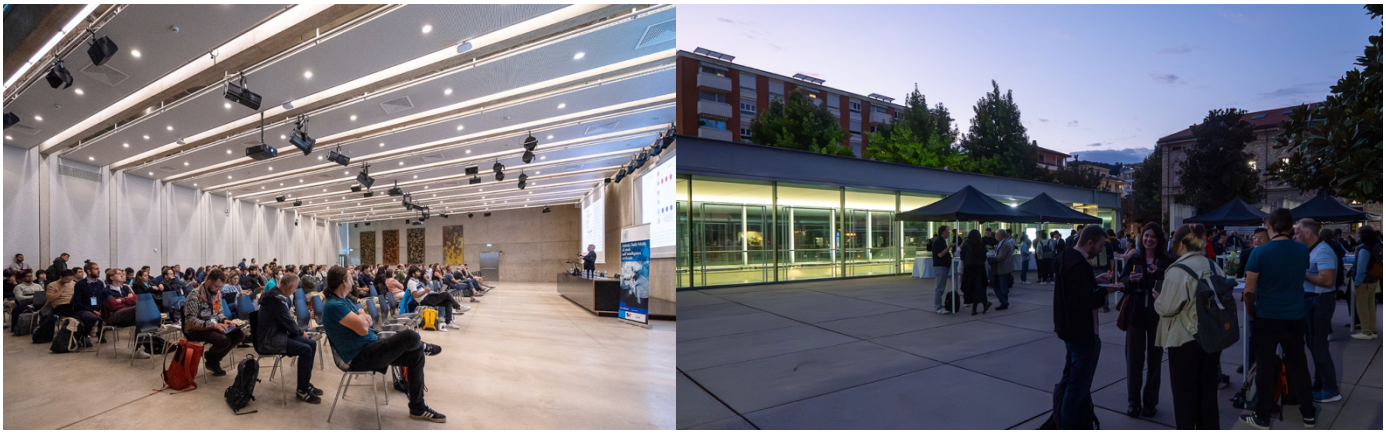
Andi Hunklinger, AiChemist fellow

AiChemist at the "AI in Drug Discovery" Workshop at ICANN2024

In early 2023, members of the AIDD consortium (https://ai-dd.eu) won the bid to organise the 33rd International Conference on Artificial Neural Networks (ICANN2024) – the flagship conference of the European Neural Network Society (ENNS). The conference was hosted by the Dalle Molle Institute for Artificial Intelligence Research (IDSIA USI-SUPSI, associated partner of AiChemist) in Lugano, Switzerland in collaboration with both the AIDD and AiChemist EU Horizon MSCA projects, with sponsoring from the King Abdullah University of Science and Technology (https://www.kaust.edu.sa/en/) in Saudia Arabia and Artificialy (https://www.artificialy.com/), a Swiss company providing AI-based solutions for businesses, based in Lugano.

The conference started on the morning of the 17th of September and ended in the late afternoon on Friday the 20th of September. The "AI in Drug Discovery" Workshop, jointly organised by the AIDD and AiChemist consortia, was split across four sessions, with the first two taking place in the morning/afternoon on Thursday the 19th of September and the two remaining sessions on Friday the 20th in the morning/afternoon.

The opening session on Tuesday the 17th started with a keynote by AiChemist project PI Jürgen Schmidhuber, on the past, present, future and far future of ML, and how the principles of the G, P and T in ChatGPT emerged in 1991. Followed by parallel sessions on computer vision, reinforcement learning and time-series processing, applications of ML/AI in medicine and physiology, generative modelling in computer vision, brain-inspired computation with applications in music, robotics and human-computer interfaces, cognitive and computational neuroscience, environment and climate applications, graph neural networks + time series and applications, as well as a workshop on explainable AI in human-robot interactions and a tutorial on FEDn – a scalable federated ML framework for cross-device and cross-silo environments (Uppsala, Sweden). The day was rounded off nicely with an evening drinks reception in the courtyard of the West Campus at USI.



Keynote lecture by Jürgen Schmidhuber (left) and drinks receptions (right) on the first day of the conference

Day 2, the 18th of September, kicked off with a keynote by Tanja Schultz, Professor of Cognitive Systems at the University of Bremen, on the topic of technical cognitive systems that automatically adapt to users' needs by interpreting their biosignals, presenting illustrative cases ranging from silent and imagined speech interfaces that convert myographic and neural signals directly into audible speech, to interpretation of human attention and decision making in human-robot interaction from multimodal biosignals. Over the course of the day, the conference attendees participated in parallel sessions on theoretical contributions in Machine Learning and Neural Networks, applications of Neural Networks, Multimodality, security in Computer Vision, in addition to special sessions, workshops and tutorials in the areas of Reservoir Computing, Neurorobotics, Federated Learning and Time Series Feature Extraction. The day was rounded off with a second keynote, by Michael Reimann, group leader of the Connectomics division of the Blue Brain Project at EPFL, who presented a model of neocortical micro- and mesocircuitry built up from highly detailed and biologically realistic models of rat non-barrel somatosensory regions under the umbrella of the Blue Brain Project.

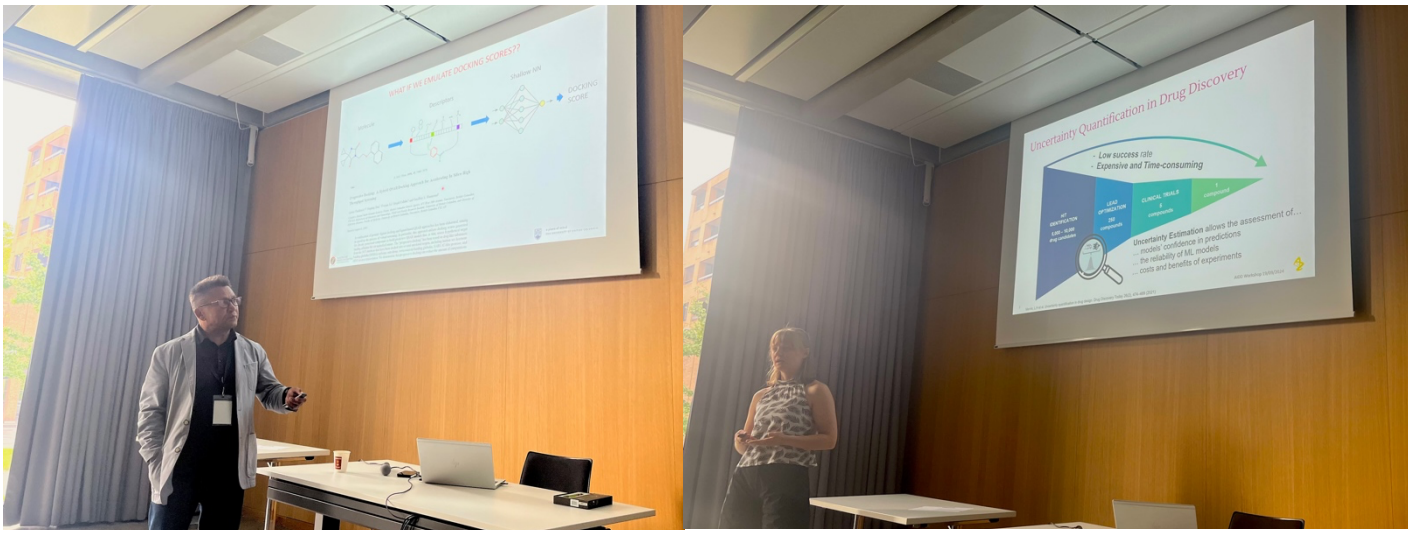
Keynote lectures by Tanja Schultz (left) and Michael Reimann (right) on the second day of the conference

The two-day-long "AI in Drug Discovery" (AIDD) Workshop, jointly organised by the AIDD and AiChemist projects, commenced in the morning of the 19th of September, shortly after the opening keynote of the day, given by Walter Senn, Professor for Computational Neuroscience at the University of Bern, discussing functional models of the brain inspired by the unprecented successes of modeling cognitive processes with AI and exploring the parallels between cortical attention mechanisms and context-dependent gating in the brain and recent advances in AI. The AIDD Workshop sessions ran alongside a dense and varied schedule of sessions dedicated Computer Vision, Neural Architectures, Sentiment Analysis, Spiking Neural Networks, Graph Neural Network Medical Image Processing, Accuracy and Robustness in Deep Neural Networks and Human-Centered Applications of Neural Networks, Novel Methods in Machine Learning, Language Modeling and Topics in Speech and Language, which ran over the course of the last two days of ICANN2024.

Walter Senn at the podium, delivering the final keynote of ICANN2024 on its penultimate day

The AIDD Workshop brought together leading researchers and practitioners (including all of the AIDD fellows and many of the AIDD and AiChemist consortium PIs) to discuss advancements and applications of artificial intelligence in pharmaceutical research and drug development. The workshop provided a platform to explore cutting-edge methodologies and real-world case studies in the domain, encouraging lively scientific exchange and giving the AIDD fellows a unique opportunity to disseminate their findings and receive constructive feedback from experts in the field. The first session was opened with a short keynote by invited speaker Artem Cherkasov, Professor in the Department of Urologic Sciences at the University of British Columbia gave an overview on the use of Active Learning for effective exploration of the chemical universe. This was followed by talks by AIDD fellows Rosa Friesacher, Alessio Fallani, Mathias Hilfiker, Muhammad Arslan Masood and Mikhail Andronov, on a range of topics, including uncertainty quantification in ML models, leveraging quantum mechanical information in ML for drug discovery, toxicity prediction and reagent space mapping.

Getting ready for the AIDD Workshop



Invited speaker Artem Cherkasov and AIDD fellow Rosa Friesacher giving talks during the first session of the AIDD Workshop

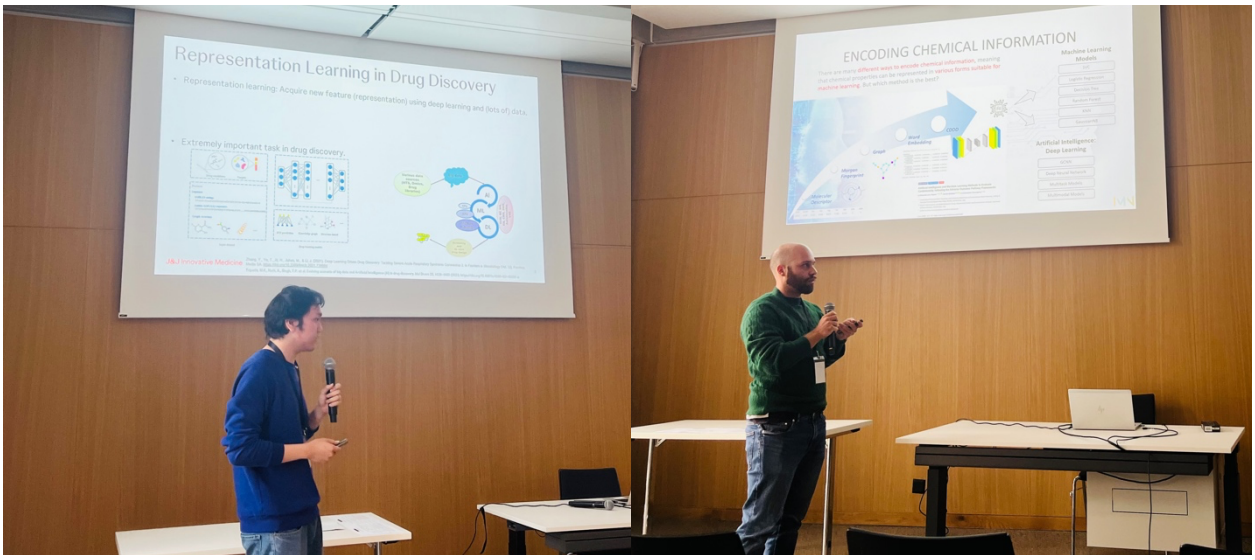
The second session began shortly after lunch, and featured talks on retrosynthesis planning, human-in-the-loop assisted drug discovery and *de novo* ligand generation from AIDD fellows Paula Torren-Peraire, Yasmine Nahal and Julian Cremer, in addition to talks on a range of topics from speakers outside of the AIDD consortium. The second session was rounded off with a short poster session, with contributions from both AIDD and AiChemist consortium members, as well as external workshop attendees.

AIDD fellows Paula Torren-Peraire (left) and Yasmine Nahal (right) delivering their presentations at the workshop

		Thursday 19th September - Foyer on Floor 0 of Corpus A, East Campus, Via La Santa 1, 6962 Lugano-Viganello	
10:30 – 10:50	Artem Cherkasov (invited speaker)	University of British Columbia	The Use of Active Learning for Effective Exploration of Chemical Universe
10:50 - 11:10	Rosa Friesacher	AstraZeneca, Katholieke Universiteit Leuven	Temporal Evaluation of Probability Calibration with Experimental Errors
11:10 - 11:30	Alessio Fallani	Janssen, University of Luxembourg	Atom-level Quantum Pretraining Enhances the Spectral Perception of Molecular Graphs in Graphormer
11:30 - 11:50	Mathias Hilfiker	AstraZeneca, University of Luxembourg	Leveraging quantum mechanical properties to predict solvent effects on large drug-like molecules
11:50 - 12:10	Muhammad Arslan Masood	Janssen, Aalto University	Balancing Imbalanced Toxicity predictor: Using MolBERT with Focal Loss
12:10 - 12:30	Mikhail Andronov	Pfizer, SUPSI	Curating reagents in chemical reaction data with an interactive reagent space map
12:30 - 14:00			**Lunch break**
14:00 - 14:20	Paula Torren Peraire	Janssen, Helmholtz Zentrum München	Improving Route Development Using Convergent Retrosynthesis Planning
14:20 - 14:40	Yasmine Nahal	Aalto University , AstraZeneca	Towards Interpretable Models of Chemist Preferences for Human-in-the-loop Assisted Drug Discovery
14:40 - 15:00	Marco Bertolini	Pfizer	Enhancing Interpretability in Molecular Property Prediction with Contextual Explanations of Molecular Graphical Depictions
15:00 - 15:20	Pedro Ballester	Imperial College	Scaffold Splits Overestimate Virtual Screening Performance
15:20 - 15:40	Justin Diamond	Universität Basel	Geometrically Guided Diffusion for Molecular Generation
15:40 - 16:00	Julian Cremer	Pfizer, Universitat Pompeu Fabra	Latent-Conditioned Equivariant Diffusion for Structure-Based De Novo Ligand Generation
16:00 – 16:30			**Poster session with coffee**

Agenda of the first two AIDD workshop sessions

On the morning of the final day of ICANN2024, the AIDD workshop reconvened with a talk from AIDD fellow Son Ha, who gave a talk based on his work on cross-multimodal learning of cell painting and transcriptomics data, which was followed by a second short poster session. During the last set of talks, AIDD fellow Emma Svensson spoke about her work on uncertainty quantification under distribution shift and external speakers, including representatives from the AiChemist consortium, gave taks based on their most recent work.



Son Ha (left) and Eduardo Viganó (right) presenting their work on the second day of the workshop

Friday 20th September - Foyer on Floor 0 of Corpus A, East Campus, Via La Santa 1, 6962 Lugano-Viganello			
10:10 - 10:30	Son Ha	Janssen, Johannes Gutenberg Universität Mainz	Cross Multimodal Learning of Cell Painting and Transcriptomics Data
10:30 - 11:00			**Poster session with coffee**
11:00 - 11:20	Fabrizio Ambrogi / Szymon Czaplak	Selvita	Target-Aware Drug Activity Model: A deep learning approach to virtual HTS
11:20 - 11:40	Emma Svensson	AstraZeneca, Johannes Kepler Universität Linz	Temporal Evaluation of Uncertainty Quantification under Distribution Shift
11:40 - 12:00	Eduardo Viganò	Mario Negri Institute for Pharmacological Research	Artificial Intelligence Methods for Evaluating Mitochondrial Dysfunction: Exploring Various Chemical Notations Suitable for Neural Language Processing Models
12:00 - 12:20	Regina Pikalyova	University of Strasbourg	Combinatorial Library Neural Network (CoLiNN) for Combinatorial Library Visualization without Compound Enumeration
12:20 - 12:40	Dragos Horvath	University of Strasbourg	De novo Drug Design – Do We Really Want To Be "Original"? A real-world case study on colchicine-site tubulin binders.
12:40			**Lunch break**

Agenda of the AIDD workshop sessions on the final day of the conference

The ICANN2024 conference concluded with a closing ceremony, chaired by ICANN2024 organisers Michael Wand and Kristina Malinovska, and ENNS President Stefan Wermter. During the ceremony the Best Paper Awards and the winner of the Tox24 Challenge were announced. The Tox24 Challenge (https://ochem.eu/static/challenge.do) launched in May 2024, was co-organised by the AIDD and AiChemist projects, in collaboration with the US Environmental Protection Agency, and attracted 78 participating teams working to push the boundaries of ML for toxicity prediction. Among the recipients of the Best Paper Awards was AIDD fellow Dr. Julian Cremer - an outstanding achievement which ended the ICANN2024 conference on a particularly high note.

Recipients of awards at the closing conference. Mikhail Andronov (fourth from the left) accepted the Best Paper Award on Julian Cremer's behalf, as Julian unfortunately could not attend the conference in person

The majority of the workshop proceedings were published open access in the ICANN2024 volume "AI in Drug Discovery, First International Workshop, AIDD 2024, held in conjunction with ICANN2024" https://link.springer.com/book/10.1007/978-3-031-72381-0, while the remainder were published with standard access within Part X of the main ICANN2024 proceedings https://link.springer.com/book/10.1007/978-3-031-72359-9. In addition, all workshop participants were invited to submit extended versions of their articles and/or novel work to the special issue "AI in Drug Discovery" of J. Cheminformatics prepared by guest editors, Dr. D.A. Clevert (Pfizer) and Dr. I.V. Tetko (HMGU), both of whom are PIs of the AiChemist project.

Publications 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) Krüger, F.P., Östman, J., Mervin, L., Tetko, I.V., Engkvist, O. Publishing Neural Networks in Drug Discovery Might Compromise Training Data Privacy. *arXiv*. 2024. https://doi.org/10.48550/arXiv.2410.16975

(4) 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.