The Advanced Machine Learning for Innovative Drug Discovery (AIDD) project

Newsletter 7th February 2025

The **Advanced Machine Learning for Innovative Drug Discovery** (AIDD) project is a Marie-Skłodowska-Curie Innovative Training Network (ITN) for Early Stage Researchers (ESRs) funded by the European Commission under the Horizon 2020 Programme, Marie Skłodowska-Curie grant agreement No 956832. The project brings together fifteen academic and industry beneficiaries from ten European countries and the University of British Columbia (Canada) to train sixteen PhD students in close collaboration with partners from the USA, Australia, China, Israel, and other countries.

Project development

The AIDD project started on 01.01.2022 and will finish on 31.03.2025.

AIDD Closing Conference – "AI in Drug Discovery" Workshop at ICANN2024

In early 2023, members of the AIDD consortium 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 AIDD consortium member Dalle Molle Institute for Artificial Intelligence Research (IDSIA USI-SUPSI) in Lugano, Switzerland in collaboration with both the AIDD (https://ai-dd.eu) and AiChemist (https://aichemist.eu) 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 projects, 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. In total there were 18 lectures given by AIDD fellows and external participants.

The opening session on Tuesday the 17th started with a keynote by AIDD 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.

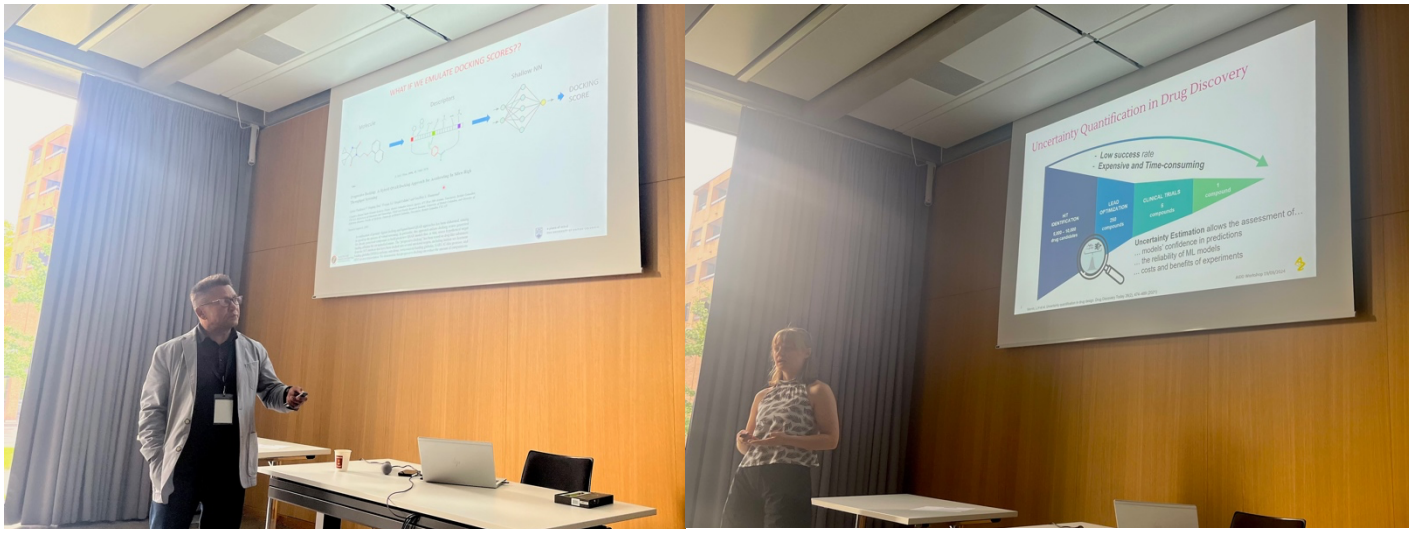
Keynote lectures by Tanja Schultz (left) and Michael Reimann (right) on the second day of the conference

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.

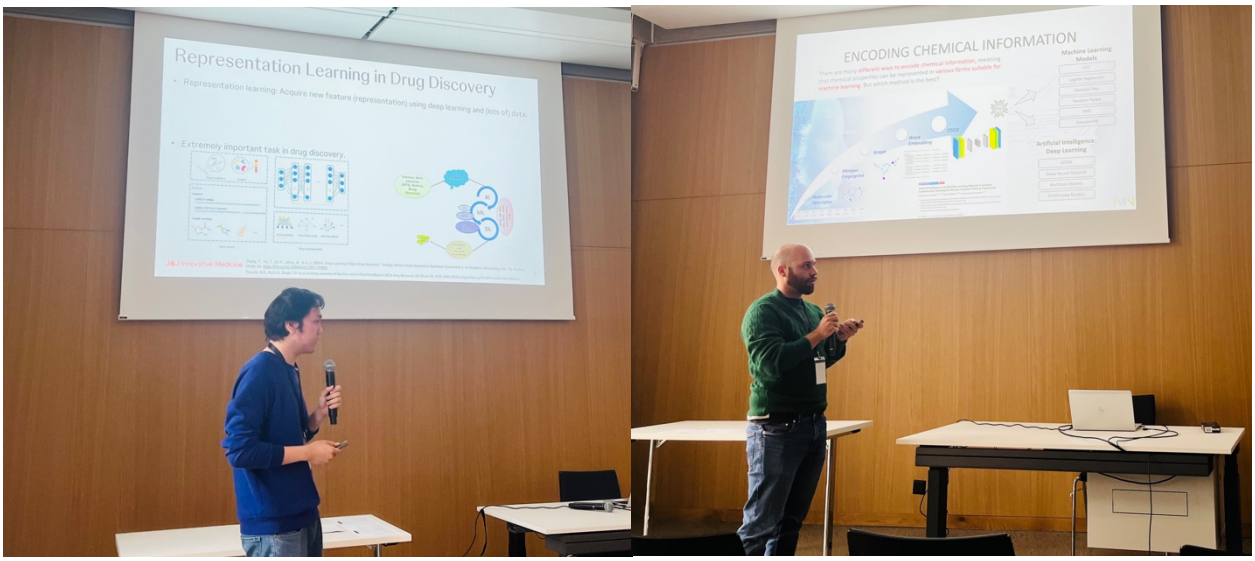
The second session began shortly after lunch, and featured talks on retrosynthesis planning, human-in-the-loop assisted drug discovery and *de novo* ligand generation from AIDD fellows Paula Torren-Peraire, Yasmine Nahal and Julian Cremer, in addition to talks on a range of topics from speakers outside 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 AIDD project.

Congratulations to our first set of PhD graduates!

Three AIDD students successfully defended their PhD theses in the second half of 2024!

Dr. Mariia Radaeva (centre) defended her PhD on the 30th of August at the University of British Columbia, followed by Dr. Varvara Voinarovska (right of centre), who defended her thesis at the Technical University of Munich on the 17th of October. Dr. Julian Cremer (second from left) defended his PhD on the 28th of November 2024 at Universitat Pompeu Fabra.

Dr. Radaeva joined McKinsey and Company as an Associate shortly after defending her thesis, while Dr. Voinarovska and Dr. Cremer are staying on at their AIDD industrial host organisations (AstraZeneca and Pfizer, respectively) as postdoctoral researchers. Wishing them all the best for what we are sure will be fantastic careers!

Publications from the AIDD Consortium

(1) Simm, J.; Arany, A.; De Brouwer, E.; Moreau, Y. Expressive Graph Informer Networks. In Machine Learning, Optimization, and Data Science; Nicosia, G., Ojha, V., La Malfa, E., La Malfa, G., Jansen, G., Pardalos, P. M., Giuffrida, G., Umeton, R., Eds.; Springer International Publishing: Cham, 2022; pp 198–212. https://www.springerprofessional.de/expressive-graph-informer-networks/20089896

(2) Sanchez-Fernandez, A.; Rumetshofer, E.; Hochreiter, S.; Klambauer, G. Contrastive Learning of Image- and Structure-Based Representations in Drug Discovery; 2022. https://openreview.net/forum?id=OdXKRtg1OG

(3) Roland, T.; Böck, C.; Tschoellitsch, T.; Maletzky, A.; Hochreiter, S.; Meier, J.; Klambauer, G. Domain Shifts in Machine Learning Based Covid-19 Diagnosis From Blood Tests. J. Med. Syst. 2022, 46 (5), 23. https://link.springer.com/article/10.1007/s10916-022-01807-1.

(4) Klotz, D.; Kratzert, F.; Gauch, M.; Keefe Sampson, A.; Brandstetter, J.; Klambauer, G.; Hochreiter, S.; Nearing, G. Uncertainty Estimation with Deep Learning for Rainfall–Runoff Modeling. Hydrol Earth Syst Sci 2022, 26 (6), 1673–1693. https://doi.org/10.5194/hess-26-1673-2022.

(5) Vall, A.; Sabnis, Y.; Shi, J.; Class, R.; Hochreiter, S.; Klambauer, G. The Promise of AI for DILI Prediction. Front. Artif. Intell. 2021, 4. https://doi.org/10.3389/frai.2021.638410

(6) Kratzert, F.; Klotz, D.; Hochreiter, S.; Nearing, G. S. A Note on Leveraging Synergy in Multiple Meteorological Data Sets with Deep Learning for Rainfall–Runoff Modeling. Hydrol Earth Syst Sci 2021, 25 (5), 2685–2703. https://doi.org/10.5194/hess-25-2685-2021.

(7) Kim, P. T.; Winter, R.; Clevert, D.-A. Unsupervised Representation Learning for Proteochemometric Modeling. Int. J. Mol. Sci. 2021, 22 (23). https://doi.org/10.3390/ijms222312882.

(8) Hoedt, P.-J.; Kratzert, F.; Klotz, D.; Halmich, C.; Holzleitner, M.; Nearing, G.; Hochreiter, S.; Klambauer, G. MC-LSTM: Mass-Conserving LSTM. ArXiv E-Prints 2021, arXiv:2101.05186.

(9) Gauch, M.; Kratzert, F.; Klotz, D.; Nearing, G.; Lin, J.; Hochreiter, S. Rainfall–Runoff Prediction at Multiple Timescales with a Single Long Short-Term Memory Network. Hydrol Earth Syst Sci 2021, 25 (4), 2045–2062. https://doi.org/10.5194/hess-25-2045-2021.

(10) Clevert, D.-A.; Le, T.; Winter, R.; Montanari, F. Img2Mol – Accurate SMILES Recognition from Molecular Graphical Depictions. Chem. Sci. 2021, 12 (42), 14174–14181. https://doi.org/10.1039/D1SC01839F.

(11) Svensson, E., Hoedt, P.-J., Hochreiter, S., Klambauer, G. Task-conditioned modeling of drug-target interactions. In ELLIS Machine Learning for Molecules Discovery Workshop. November 28, 2022. https://moleculediscovery.github.io/workshop2022/

(12) Hassen, A. K.; Torren-Peraire, P.; Genheden, S.; Verhoeven, J.; Preuss, M.; Tetko, I. V. Mind the Retrosynthesis Gap: Bridging the Divide between Single-Step and Multi-Step Retrosynthesis Prediction; 2022. https://openreview.net/forum?id=LjdtY0hM7tf

(13) Andronov, M.; Voinarovska, V.; Andronova, N.; Wand, M.; Clevert, D.-A.; Schmidhuber, J. Reagent Prediction with a Molecular Transformer Improves Reaction Data Quality. *Chem. Sci.* 2023. https://doi.org/10.1039/D2SC06798F

(14) Radaeva, M.; Ban, F.; Zhang, F.; LeBlanc, E.; Lallous, N.; Rennie, P.S.; Gleave, M.E.; Cherkasov, A. Development of Novel Inhibitors Targeting the D-Box of the DNA Binding Domain of Androgen Receptor. *Int. J. Mol. Sci.* 2021, *22*, 2493, https://doi.org/10.3390/ijms22052493

(15) Cremer, J.; Medrano Sandonas, L.; Tkatchenko, A.; Clevert, D.A.: De Fabritiis, G. Equivariant Graph Neural Networks for Toxicity Prediction. *Chem. Res.*

Toxicol. 2023. https://doi.org/10.1021/acs.chemrestox.3c00032
(16) Sarkis, M.; Fallani, A.; Tkatchenko, A. Modeling Non-Covalent Interatomic Interactions on a Photonic Quantum Computer. *Physical Review Research*. 2023. 10.1103/PhysRevResearch.5.043072
(17) Martinelli, J.; Nahal, Y.; Lê, D.; Engkvist, O.; Kaski, S. Leveraging expert feedback to align proxy and ground truth rewards in goal-oriented molecular generation. *In NeurIPS2023 Workshop AI4D3. 2023.* https://openreview.net/forum?id=KWIM7ZNYxb. Link to full workshop paper.
(18) Sanchez-Fernandez, A.; Rumetshofer, E.; Hochreiter, S.; Klambauer, G. CLOOME:contrastive learning unlocks bioimaging databases for queries with chemical structures. *Nature Communications*. 2023. https://doi.org/10.1038/s41467-023-42328-w
(19) Le, T.; Cremer, J.; Noé, F.; Clevert, D-A.; Schütt, K. Navigating the design space of equivariant diffusion-based generative models for de novo 3D molecule generation. arXiv. 2023. https://doi.org/10.48550/arXiv.2309.17296
(20) Voinarovska, V.; Kabeshov, M.; Dudenko, D.; Genheden, S.; Tetko, I.V. When Yield Prediction Does Not Yield Prediction: An Overview of the Current Challenges. *J. Chem. Inf. Model.* 2024. https://doi.org/10.1021/acs.jcim.3c01524
(21) Andronov, M.; Andronova, N., Wand, M., Schmidhuber, J., Clevert, D-A. A reagent-driven visual method for analyzing chemical reaction data. arXiv. 2024. https://doi.org/10.26434/chemrxiv-2024-q9tc4
(22) 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
(23) Torren Peraire, P.; Hassen, A.K.; Genheden, S.; Verhoeven, J., Clevert, D-A.; Preuss, M.; Tetko, I.V. Models Matter: the impact of single-step retrosynthesis on synthesis planning. *Digital Discovery*. 2024. https://doi.org/10.1039/D3DD00252G
(24) 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
(25) Bernatavicius, A.; Šícho, M.; Janssen, A.; Hassen, A.K.; Preuss, M.; van Westen, G. AlphaFold meets de novo drug design: leveraging structural protein information in multi-target molecular generative models. ChemRxiv. 2024. https://doi.org/10.26434/chemrxiv-2024-60tc7
(26) Ha, S.; Leuschner, L.; Czodrowski, P. FSL-CP: a benchmark for small molecule activity few-shot prediction using cell microscopy images. *Digital Discovery.* 2024. DOI: 10.1039/d3dd00205e
(27) Svensson, E., Hoedt, P.-J., Hochreiter, S., Klambauer, G. HyperPCM: Robust Task-Conditioned Modeling of Drug-Target Interactions. *J. Chem. Inf. Model.* 2024. *https://doi.org/10.1021/acs.jcim.3c01417*
(28) Hassen, A.K.; Šícho, M.; van Aalst, Y.J.; Huizenga, M.C.W.; Reynolds, D.N.R.; Luukkonen, S.; Bernatavicius, A.; Clevert, D-A.; Janssen, A.P.A.; van Westen, G.J.P.; Preuss, M. Generate What You Can Make: Achieving in-house synthesizability with readily available resources in de novo drug design. ChemRvix. 2024. https://doi.org/10.26434/chemrxiv-2024-wtjt6
(29) Borowa, A., Rymarczyk, D., Zyła, M., Kanduła, M., Sánchez-Fernández, A., Rataj, K., Struski, Ł., Tabor, J., and Zielinski, B. Decoding phenotypic screening: A comparative analysis of image representations. *Computational and Structural Biotechnology Journal*, 2024. https://doi.org/10.1016/j.csbj.2024.02.022
(30) Tan, L; Hirte, S.; Palmacci, V.; Stork, C.; Kirchmair, J. Tackling assay interference

associated with small molecules. *Nature Reviews Chemistry*. 2024. https://doi.org/10.1038/s41570-024-00593-3
(31) Lakshidaa Saigiridharan, Alan Kai Hassen, Helen Lai, Paula Torren-Peraire, Ola Engkvist, and Samuel Genheden. Aizynthfinder 4.0: developments based on learnings from 3 years of industrial application. *Journal of Cheminformatics*. 2024. https://doi.org/10.1186/s13321-024-00860-x
(32) Cremer, J., Le, T., Noé, F., Clevert, D-A., Schütt, K. T. PILOT: Equivariant diffusion for pocket conditioned de novo ligand generation with multi-objective guidance via importance sampling. *Chemical Science*, 2024. https://doi.org/10.1039/D4SC03523B
(33) Palmacci, V., Hirte, S., Hernandez Gonzales, J.E., Montanari, F., Kirchmair, J. Statistical approaches enabling technology-specific assay interference prediction from large screening data sets. *Artificial Intelligence in the Life Sciences*. 2024. https://doi.org/10.1016/j.ailsci.2024.100099
(34) Ha, S., Jaensch, S., Freitas, L.G.A., Herman, D., Czodrowski, P.. and Ceulemans, H. Low concentration cell painting images enable the identification of highly potent compounds. *Scientific Reports*. 2024. https://doi.org/10.1038/s41598-024-75401-5
(35) Rydholm, E., Bastys, T., Svensson, E., Kannas, C., Engkvist, O., and Kogej, T. Expanding the chemical space using a chemical reaction knowledge graph. *Digital Discovery*. 2024. https://doi.org/10.1039/D3DD00230F
(36) Sandonas, L. M., Rompaey, D., Fallani, A., Hilfiker, M., Hahn, D., Perez-Benito, L., Verhoeven, J., Tresadern, G., Wegner, J. K., Ceulemans, H., and Tkatchenko A. Dataset for quantum-mechanical exploration of conformers and solvent effects in large drug-like molecules. *Scientific Data*. 2024. https://doi.org/10.1038/s41597-024-03521-8
(37) Andronov, M., Andronova, N., Wand, M., Schmidhuber, J., Clevert, D-A. Accelerating the inference of string generation-based chemical reaction models for industrial applications. *arXiv*. 2024. https://doi.org/10.48550/arXiv.2407.09685
(38) Friesacher, H.R., Engkvist, O., Mervin, L., Moreau, Y. and Arany, A. Achieving well-informed decision-making in drug discovery: A comprehensive calibration study using neural network-based structure-activity models. *arXiv*. 2024. https://doi.org/10.48550/arXiv.2407.14185
(39) Fallani, A., Sandonas, L. M., and Tkatchenko, A. Inverse mapping of quantum properties to structures for chemical space of small organic molecules. *Nature Communications*. 2024. https://doi.org/10.1038/s41467-024-50401-1
(40) Nahal, Y., Menke, J., Martinelli, J., Heinonen, M., Kabeshov, M., Janet, J. P., Nittinger, E., Engkvist, O. and Kaski, S. Human-in-the-loop active learning for goal-oriented molecule generation. *Journal of Cheminformatics.* 2024. https://doi.org/10.1186/s13321-024-00924-y
(41) Menke, J., Nahal, Y., Bjerrum, E. J., Kabeshov, M., Kaski, S., and Engkvist, O. Metis-a python-based user interface to collect expert feedback for generative chemistry models. *Journal of Cheminformatics.* 2024. https://doi.org/10.1186/s13321-024-00892-3
(42) Hartog, P., Westerlund, A., Tetko, I.V., Genheden, S. Investigations into the efficiency of computer-aided synthesis planning. ChemRxiv. 2024. https://doi.org/10.26434/chemrxiv-2024-q2v87
(43) Lai, H., Kannas C., Kai Hassen, A., Rydholm, E., Westerlund, A., Clevert, D-A., et al. Multi-objective synthesis planning by means of Monte Carlo tree search. ChemRxiv. 2024. https://doi.org/10.26434/chemrxiv-2024-3tzs5-v2
(44) Friesacher, H.R., Svensson, E., Arany, A., Mervin, L., Engkvist, O. Temporal Evaluation of Probability Calibration with Experimental Errors. In: Clevert, DA., Wand, M., Malinovská, K., Schmidhuber, J., Tetko, I.V. (eds) AI in Drug Discovery. ICANN 2024. Lecture Notes in Computer Science, vol 14894. Springer, Cham.

2024. https://doi.org/10.1007/978-3-031-72381-0_2
(45) Andronov, M., Andronova, N., Wand, M., Schmidhuber, J., Clevert, DA. Curating Reagents in Chemical Reaction Data with an Interactive Reagent Space Map. In: Clevert, DA., Wand, M., Malinovská, K., Schmidhuber, J., Tetko, I.V. (eds) AI in Drug Discovery. ICANN 2024. Lecture Notes in Computer Science, vol 14894. Springer, Cham. 2024. https://doi.org/10.1007/978-3-031-72381-0_3
(46) Cremer, J., Le, T., Clevert, DA., Schütt, K.T. Latent-Conditioned Equivariant Diffusion for Structure-Based De Novo Ligand Generation. In: Clevert, DA., Wand, M., Malinovská, K., Schmidhuber, J., Tetko, I.V. (eds) AI in Drug Discovery. ICANN 2024. Lecture Notes in Computer Science, vol 14894. Springer, Cham. 2024. https://doi.org/10.1007/978-3-031-72381-0_4
(47) Hilfiker, M., Sandonas, L.M., Klähn, M., Engkvist, O., Tkatchenko, A. Leveraging Quantum Mechanical Properties to Predict Solvent Effects on Large Drug-Like Molecules. In: Clevert, DA., Wand, M., Malinovská, K., Schmidhuber, J., Tetko, I.V. (eds) AI in Drug Discovery. ICANN 2024. Lecture Notes in Computer Science, vol 14894. Springer, Cham. 2024. https://doi.org/10.1007/978-3-031-72381-0_5
(48) Nahal, Y. *et al.* Towards Interpretable Models of Chemist Preferences for Human-in-the-Loop Assisted Drug Discovery. In: Clevert, DA., Wand, M., Malinovská, K., Schmidhuber, J., Tetko, I.V. (eds) AI in Drug Discovery. ICANN 2024. Lecture Notes in Computer Science, vol 14894. Springer, Cham. 2024. https://doi.org/10.1007/978-3-031-72381-0_6
(49) Fallani, A., Arjona-Medina, J., Chernichenko, K., Nugmanov, R., Wegner, J.K., Tkatchenko, A. Atom-Level Quantum Pretraining Enhances the Spectral Perception of Molecular Graphs in Graphormer. In: Clevert, DA., Wand, M., Malinovská, K., Schmidhuber, J., Tetko, I.V. (eds) AI in Drug Discovery. ICANN 2024. Lecture Notes in Computer Science, vol 14894. Springer, Cham. 2024. https://doi.org/10.1007/978-3-031-72381-0_7
(50) Masood, M.A., Kaski, S., Ceulemans, H., Herman, D., Heinonen, M. Balancing Imbalanced Toxicity Models: Using MolBERT with Focal Loss. In: Clevert, DA., Wand, M., Malinovská, K., Schmidhuber, J., Tetko, I.V. (eds) AI in Drug Discovery. ICANN 2024. Lecture Notes in Computer Science, vol 14894. Springer, Cham. 2024. https://doi.org/10.1007/978-3-031-72381-0_8
(51) Hartog, P., Svensson, E., Mervin, L., Genheden, S., Engkvist, O., Tetko, I.V. Registries in Machine Learning-Based Drug Discovery: A Shortcut to Code Reuse. In: Clevert, DA., Wand, M., Malinovská, K., Schmidhuber, J., Tetko, I.V. (eds) AI in Drug Discovery. ICANN 2024. Lecture Notes in Computer Science, vol 14894. Springer, Cham. 2024. https://doi.org/10.1007/978-3-031-72381-0_9
(52) Svensson, E., Friesacher, H.R., Arany, A., Mervin, L., Engkvist, O. Temporal Evaluation of Uncertainty Quantification Under Distribution Shift. In: Clevert, DA., Wand, M., Malinovská, K., Schmidhuber, J., Tetko, I.V. (eds) AI in Drug Discovery. ICANN 2024. Lecture Notes in Computer Science, vol 14894. Springer, Cham. 2025. https://doi.org/10.1007/978-3-031-72381-0_11
(53) Masood, M.A., Cui, T., Kaski, S. Deep Bayesian Experimental Design for Drug Discovery. In: Clevert, DA., Wand, M., Malinovská, K., Schmidhuber, J., Tetko, I.V. (eds) AI in Drug Discovery. ICANN 2024. Lecture Notes in Computer Science, vol 14894. Springer, Cham. 2024. https://doi.org/10.1007/978-3-031-72381-0_12
(54) Fallani, A., Nugmanov, R., Arjona-Medina, J., Wegner, J.K, Tkatchenko, A., Chernichenko, K. Pretraining Graph Transformers with Atom-in-a-Molecule Quantum Properties for Improved ADMET Modeling. *arXiv*. 2024. https://doi.org/10.48550/arXiv.2410.08024