# Metis - A Python-Based User Interface to Collect Expert Feedback for Generative Chemistry Models

Janosch Menke\* Department of Computer Science and Engineering Chalmers University of Technology 41296 Gothenburg, Sweden janosch.menke.research@proton.me Yasmine Nahal Department of Computer Science Aalto University 02150 Espoo, Finland Esben Jannik Bjerrum

Cheminformania Consulting 43138 Mölndal, Sweden

Mikhail Kabeshov Molecular AI, Discovery Sciences AstraZeneca R&D 43183 Mölndal, Sweden

Department of Computer Science Aalto University 02150 Espoo, Finland

Department of Computer Science University of Manchester Manchester M13 9PL, United Kingdom

#### Ola Engkvist

Samuel Kaski

Molecular AI, Discovery Sciences AstraZeneca R&D 43183 Mölndal, Sweden Department of Computer Science and Engineering Chalmers University of Technology 41296 Gothenburg, Sweden

## Abstract

Current *de novo* drug design models face one crucial challenge: a disparity between the user's expectations and the actual output of the model in practical applications. Tailoring models to better align with chemists' preferences is key to overcoming this obstacle effectively.

While interest in preference-based and human-in-the-loop machine learning in chemistry is continuously increasing, no tool currently exists that enables the collection of standardized and chemistry-specific feedback. Metis is a Python-based opensource graphical user interface (GUI), designed to solve this and enable the collection of chemists' detailed feedback on molecular structures. The GUI enables chemists to explore and evaluate molecules, offering a user-friendly interface for annotating preferences and specifying desired or undesired structural features. By providing chemists the opportunity to give detailed feedback, allows researchers to capture more efficiently the chemist's implicit knowledge and preferences. This knowledge is crucial to align the chemist's idea with the *de novo* design agents. The GUI aims to enhance this collaboration between the human and the "machine" by providing an intuitive platform where chemists can interactively provide feedback on molecular structures, aiding in preference learning and refining *de novo* design strategies. Metis integrates with the existing *de novo* framework REINVENT, creating a closed-loop system where human expertise can continuously inform and refine the generative models.

# 1 Introduction

De novo drug design, a process of creating novel molecular structures with desired biological properties, stands as a cornerstone in the automation of the drug discovery process.[1] It often makes use of reinforcement learning to iteratively optimize molecules to achieve predefined objectives, such as efficacy and safety profiles.[2, 3] Reinforcement learning (RL) is a paradigm in machine learning that involves training agents to make sequences of decisions in an environment to maximize a cumulative reward. RL has demonstrated remarkable success across a wide range of applications outside of chemistry, enabling agents to learn highly complex behaviors. [4-6] A key to the successful training of an RL agent lies in the design of a well-defined reward function, which serves as a guide for the agent to achieve desirable behaviors. Without a carefully crafted reward function, the agents might not converge to desired behaviors. Additionally, there is a risk of reward exploitation and hacking, where the agent may find unintended shortcuts to maximize its rewards, leading to an agent maximizing the reward in an undesirable fashion.[7, 8]

Creating a well-specified reward function is not only difficult as it requires a deep understanding of the task at hand, but the translation of that domain expertise into a parametric function used to compute the reward can pose a challenge. In many cases, domain experts struggle with the translation part, as they have a good idea of what an acceptable solution looks like, but they are not able to translate this into an explicit function.[9] This leads to the researcher having to spend extensive time on reward engineering, to create a reward function that enables the agent to learn the desired behavior.

One solution to the problem can be found in Human-in-theloop (HITL) Reinforcement Learning.[10] Here somewhere in the training loop, human behavior or feedback is used to better align the agent's behavior with the human's expectations. An effective solution to that problem involves learning the policy implicitly through methodologies such as imitation learning and behavioral cloning.[11, 12] In these approaches, the agent learns by imitating the actions of an expert, allowing it to grasp the nuances of the task without explicitly defining a reward function. An alternative strategy is inverse reinforcement learning, by inferring the underlying reward structure from observed expert behavior, the system learns a reward model that should match more closely with the expectation.[13] Lastly, preference learning can be used to actively incorporate human feedback into the RL training loop.[9] This integration can occur directly, as demonstrated by methods like Deep Preference Optimization (DPO)[14], or indirectly through the creation of a reward model based on human feedback.[15] In most applications, the user ranks two or more outputs by their preferences and iteratively the model aligns with the expectations of the user

In many popular *de novo* drug design frameworks, the chemist must also express his preferences in the form of a parametric reward function, that describes the properties that chemists expect the generated molecules to have.[16] Chemists can struggle in defining well-specified reward functions. Organic and medicinal chemists are often not overly familiar with potential molecular descriptors that can be used to create a reward function. Additionally, they are not trained to think about molecules as a sum of individual properties. Rather they evaluate the quality of molecules more holistically by looking at the structural formula. This leads to a situation in which the reward functions produce molecules that are not aligned with the ideas of the chemists, and as a result, extensive manual cleaning and filtering of the generated molecules is necessary.

In generative chemistry, preference learning has been applied to mitigate the underspecification of reward functions. For instance, projects like MolSkill leverage human preferences to guide the generation of molecules.[17] A different study uses the liking or disliking of molecules to extract which property ranges are acceptable to chemists.[18] However, chemists often possess nuanced opinions about molecules, extending beyond the binary decisions of liking or disliking. They can provide valuable and specific feedback on properties and substructures, enabling a more nuanced understanding of the desired molecular characteristics. Collecting such specific feedback cannot only align the *de novo* design agent with human preferences, but in the long-term one can elucidate the implicit knowledge and experience of the chemist.

To capture this nuanced feedback from chemists, we have developed Metis, a user interface that facilitates the integration of specific and detailed human feedback into the RL process. Metis enables chemists to communicate their preferences, concerns, and insights, thereby enhancing the RL agent's ability to generate molecules that align more closely with the desired properties and characteristics. Through Metis, we aim to provide an interface that enables practical Human-in-the-loop *de novo* drug Design, ensuring a more effective and collaborative approach to molecular generation.

With Metis, our objective is to provide a first-of-its-kind interface that enables practical Human-in-the-loop (HITL) *de novo* drug design. Additionally, it should serve as an initial step for research surrounding the methodology of HITL drug design, ultimately ensuring a more effective and interactive approach to molecular generation and closing the gap between the chemist's expectation and the generative model.

# 2 Application Overview

Metis is designed to allow (medicinal) chemists to provide feedback on small molecular structures. In particular, it is focused on collecting feedback on De novo-generated molecules. While chemists have general pre-disposition towards specific substructures, in practice the molecules are not evaluated in a vacuum. Rather, the chemists work in the context of a specific project. Typically, these projects involve targeting a specific protein for which an active molecule needs to be developed. Additional constraints such as solubility, selectivity, and toxicity may be specified. Given the dynamic nature of projects, a chemist's preferences may vary significantly from one project to another. Hence, it is essential to collect and interpret feedback within the context of the project. To account for this Metis, does not only allow the Chemist to give feedback but can also provide project-relevant information to the chemists. However, Metis can not only be used to collect feedback but can also directly integrate this feedback into a *de novo* drug design run, which, if done in an iterative manner should align the generated molecules with the preferences of the user.

In the following the different components of Metis will be introduced in more detail.

#### 2.1 Feedback to *de novo* design

Metis can seamlessly re-integrate feedback back into a *de novo* design loop. The newly generated molecules using the feedback will then automatically be loaded into Metis and further feedback can be given. This loop allows the chemist to iterativly fine-tune their feedback as well as the molecular generator. Over multiple feedback rounds, the *de novo* design model should more closely align with the preferences of the chemist. Currently, Metis offers two methods for integrating user feedback into the *de novo* design



Figure 1 View of the Metis GUI.

model. One approach involves utilizing a *Reward Model*, a machine learning algorithm trained on the chemist's feedback. This model predicts whether a given molecule will be favored or disfavored by the chemist, thus contributing to the refinement of the *de novo* model. Presently, Metis supports all scikit-learn models utilizing an RDKit Morgan fingerprint. In the future, we aim to support scikit-mol[19] models, providing more customizability in the input features.



Figure 2 Overview of iteratively aligning the  $de \ novo$  model with the vision of the chemist using Metis

An alternative method is to directly build a reward function from the feedback of the user. This reward function constitutes a sum of multiple equally weighted properties. At its core, it tries to minimize the presence of substructures flagged by the chemist as liabilities and tries to maximize the presence of favorable substructures. Additionally, it seeks to enhance similarity to liked molecules up to a specified threshold while minimizing similarity to disliked ones. In comparison to the reward model, the use of a reward function enables the integration of more fine-grained feedback that goes beyond liking versus disliking a molecule

For now, only REINVENT[20] is supported. In order to make use of that feature, a working REINVENT Installation needs to be present on a remote machine, to which the users have access through an SSH key.

2.2 Molecular Display

The molecular display provides an image of the generated molecule, for which feedback should be collected. Users can click on atoms to highlight substructures within the molecule. Additional tabs offer users more detailed information about the molecule. The "Most Similar Active" tab allows users to see known active molecules most similar to the generated one, allowing chemists to judge whether the generated molecule is a sensible extension based on already known information. Additionally, the "Explainability" tab provides insights into why a scikit-learn[21] QSAR model suggests a generated molecule as potentially active, empowering chemists to make informed decisions based on the model's reasoning. Currently, only the RDKits[22] native explainibility function developed by Riniker and Landrum[23] is available. But an extension to other methods should be easy to implement.



Figure 3 Showcase of the three windows the user can cycle through.
(a) is the default window in which the user can select atoms to highlight.
(b) the explanation window shows the per atom contribution to the prediction of an ML model and (c) is the window that shows the most similar active compounds from the training set

## 2.3 Target Product Profile

The Target Product Profile window can be used to present information on the project at hand. Here are relevant properties and their relevancy can be explained. Descriptors that relate to these properties can also be shown for each molecule displayed.

## 2.4 Global Liabilities

The Global Liability window collects feedback on the overarching properties of a molecule. These global liabilities encompass characteristics not tied to specific substructures but rather arise from the molecule's overall structure. For instance, the molecule's size falls under this category. Additionally, it accommodates liabilities challenging to assign to specific atoms, such as synthetic accessibility, which can be easier to evaluate globally. Chemists can assess each property showcased and determine if the molecule aligns with their conceptualization of it or not. Crucially, within this window, chemists can provide feedback on their overall preference, whether they like or how much they like this molecule.

## 2.5 Local Liabilities

Local Liabilities refer to liabilities of molecules that can be directly mapped to specific local substructures. Users have the flexibility to toggle between different liabilities they wish to highlight within the molecule. By selecting atoms in the molecular display, these substructures can be associated with the corresponding liability, each distinguished by a unique color. Additionally, users can create new labels, not predefined, by specifying their concerns for a particular substructure in a text field. Not technically a liability, but by default Metis also provides the chemist the option to highlight substructures that they like. The chemist can also make a distinction of whether the feedback he provides is feedback that is relevant only to the current project, or whether the feedback is generally valid across many projects. The highlighted substructures are stored and saved by recording the atom indices of highlighted atoms, additionally, the substructures are directly mapped to a SMARTS pattern that is also saved. Next to the atom and bond type, the SMARTS pattern also recorded ring membership and the number of attached hydrogens for each atom. To ensure that adequate information is recorded, the SMARTS pattern is expanded to also include all atoms that are directly connected to the highlighted substructure. This way the feedback of the chemist can more adequately be saved. The significance of this approach becomes evident when considering examples such as distinguishing between amides and ketones. If a chemist is dissatisfied with a ketone, they are likely to flag the atom and the double-bonded carbon but may overlook the two additional carbons. Further down the line, this can lead to desirable amides being flagged as liabilities. By recording the expanded environment of a highlighted substructure, such oversight is mitigated, ensuring that chemists' feedback is accurately represented.

#### 2.6 Additional Features

In the navigation bar at the bottom of the GUI, multiple additional helpful buttons are provided. Most importantly, the "Next" and "Back" button allow the users to switch between molecules that are supposed to be evaluated. The "Edit" button will open a molecular editor in a separate window. The editor can be used by the user to suggest an alternative molecule to the one that is currently to be evaluated. The editor will open with the current molecule already loaded. The molecular editor that is used is the rdEditor[24] Changes made to the molecule in the editor will then be stored separately in the backend. The "History" button will also open a separate window, in which the chemist can scroll through the already evaluated molecules. Lastly, the "Send" button, will start a new *de novo* run on a remote machine using the feedback provided by the Chemist. A more detailed description is found in the following section.

# 3 Customizability

Metis is designed to be customizable through a yaml file, in which the users can specify which information to show to the chemist, and what kind of feedback can be given by the chemist. The exact liabilities can be changed, but also complete GUI elements can be removed if needed. A complete list of all settings and their use, together with some examples are provided with the GitHub Repository. The examples are designed around a fictitious drug design project around designing a MAPK10 (JNK3) kinase inhibitor. For this, initial molecules were generated with REINVENT. The generated molecules as well as the models are provided with examples. The examples provided three different setup files to cover different use cases and complexities in setting up the GUI.

- 1. **UI Only Example** In this example, Metis is only used to collect feedback for generated structures. No models are retrained and no *de novo* run can be started.
- 2. **Reward Model Example** In addition to setting up the GUI, here the feedback is used to directly train the reward model. However, still, no *de novo* run can automatically be started. This setup can be useful, in scenarios where one is only interested in building a reward model or the reward model shall be used in a different *de novo* environment.
- 3. De Novo Loop Example This example showcases all the functionalities of Metis. The user feedback is collected, a reward model is trained and subsequently used to start a *de novo* run using REINVENT on a remote machine. The newly generated structures are then copied and loaded into Metis. While the other to examples can be started immediately. This example requires REINVENT installation on a remote machine and some files need to be transferred.

As Metis is written in Python changes to modules not "exposed" through the yaml settings file, can also be changed by adding additional classes that follow the design of already implemented classes. Examples of such are classes that take care of the sampling of molecules, or how the reward models are trained.

For the iterative re-training of the *de novo* models, one could in theory use any *de novo* model. However, Metis does soft-lock users to use REINVENT. The limiting factor is that no unified standard for *de novo* design tools has been proposed or adopted. Thus, most

models require different setups with different configuration files, which then return their results in different file formats. This makes it difficult to ensure operability between different models. While it is possible to use an alternative to REINVENT, it would require significant modifications by the user to the existing code of Metis.

# 4 Implementation

Metis is written in Python and relies at its core on three libraries. PySide2[25] is a Python implementation of the Qt Framework and is used to create the Graphical User Interface (GUI) a user can interact with. The molecular drawing, highlighting and editing capabilities are provided by rdEditor[24]. Additionally, RDKit is used to manipulate, and save molecular structures and information. Further core libraries in use are pandas, numpy, and scikitlearn.

Figure 4 gives an overview of how the different parts of Metis come together. The interface the user can interact with is created using the aforementioned PySide2 and rdEditor. Molecules and their associated feedback reside within a custom extension of a pandas data frame. This specialized data frame efficiently stores diverse forms of feedback and translates between atom indices and their corresponding SMARTS structure representations. The molecules presented to the user are sampled from an initial set of molecules stored in a separate file.

The "Reward Model Trainer" class handles both the training of the reward model and the creation of the reward function. If a QSAR Model exists that needs fine-tuning with user feedback, the trainer loads the QSAR model along with its original training data and merges it with the obtained feedback. Subsequently, the model undergoes re-training using the combined dataset. In the absence of an initial model, training commences from scratch.

For a *de novo* run, the process involves initializing a "De Novo Runner" class instance on a separate core. This runner then generates input files for REINVENT and transfers them, along with the updated model in a "pickle" format, to the remote location via SSH. A remote run is then initiated using SLURM. The "De Novo Runner" remains in a waiting state until the REINVENT run concludes, after which it retrieves the current state of the Agent and the newly generated molecules back to the local machine. From here, new molecules are selected to be evaluated by the user.

In each iteration, the molecules, their feedback, the reward models, and the *de novo* model are saved.

## 4.1 Installation

Metis is an open-source software that can be downloaded from https://github.com/JanoschMenke/metis. After setting up the environment either manually or through the use of setup.py the software can be used. A more detailed description of the setup and the settings is provided on the GitHub repository, together with examples that should let the user get started directly.

#### 4.2 License

Metis is published under the permissive MIT license.

#### 4.3 Limitations

The sole reliance on Python, Pyside2, and RDKit makes Metis very adaptable for all researchers in cheminformatics and its adjacent fields. Most researchers code in Python and are able to make their desired changes. However, this choice for PySide/Python, also makes Metis not currently hostable on the web. This can be attributed to the fact that PySide2 at its core uses C++ and currently does not have WebAssembly support. The second limitation, previously mentioned, is that by default only REINVENT as

a *de novo* model is supported. While it is not difficult for users to adapt the code to their model of choice, many small changes need to be made, as Metis is written with REINVENTs file formatting in mind.

# 5 Conclusion

Here we introduce Metis, a Graphical User Interface, that enables researchers to collect feedback on generated molecules that go beyond simple like or dislike. Chemists can assign substructures specific liabilities, flag concerning properties, and can suggest alternative molecules to the generated ones. Metis also serves as a platform to provide chemists with sufficient information on the task to make informed decisions on the generated compounds. As the feedback can be directly integrated within existing de novo Design loops, the GUI has its practical application and can help endusers to fine-tune and align the generative model with their ideas and preferences. To our knowledge, no other application exists that provides such functionality and Metis can serve as a starting point for the community to develop and test ideas on how elaborate chemical knowledge and the feedback it gives rise to, can adequately be modeled and integrated into existing deep learning models.

# 6 Acknowledgements

This study was partially funded by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie Innovative Training Network European Industrial Doctorate grant agreement No. 956832 "Advanced Machine Learning for Innovative Drug Discovery" as well as the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. Further, this work was supported by the Academy of Finland Flagship program: the Finnish Center for Artificial Intelligence FCAI. Samuel Kaski was supported by the UKRI Turing AI World-Leading Researcher Fellowship, [EP/W002973/1]

The computations were supported by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725."



Figure 4 Schematic overview of Metis. Yellow squares indicate modules that are only optional and only needed if Reward Models should be trained and/or *de novo* model run should be started.

## References

- [1] Schneider, G. Automating Drug Discovery. *Nature Reviews* Drug Discovery **2018**, *17*, 97–113.
- [2] Korshunova, M.; Huang, N.; Capuzzi, S.; Radchenko, D. S.; Savych, O.; Moroz, Y. S.; Wells, C. I.; Willson, T. M.; Tropsha, A.; Isayev, O. Generative and Reinforcement Learning Approaches for the Automated De Novo Design of Bioactive Compounds. *Communications Chemistry* **2022**, *5*, 129.
- [3] Svensson, H. G.; Tyrchan, C.; Engkvist, O.; Chehreghani, M. H. Utilizing Reinforcement Learning for De Novo Drug Design. arXiv preprint 2023,
- [4] Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; Riedmiller, M. A. Playing Atari with Deep Reinforcement Learning. *arXiv preprint* 2013, *arxiv*:1312.5602.
- [5] Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; others Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature* 2016, 529, 484–489.
- [6] Vinyals, O.; Babuschkin, I.; Czarnecki, W. M.; Mathieu, M.; Dudzik, A.; Chung, J.; Choi, D. H.; Powell, R.; Ewalds, T.; Georgiev, P.; others Grandmaster Level in StarCraft II Using Multi-agent Reinforcement Learning. *Nature* 2019, *575*, 350–354.
- [7] Amodei, D.; Olah, C.; Steinhardt, J.; Christiano, P.; Schulman, J.; Mané, D. Concrete Problems in AI Safety. arXiv preprint 2016, arXiv:1606.06565.
- [8] Skalse, J.; Howe, N.; Krasheninnikov, D.; Krueger, D. Defining and Characterizing Reward Gaming. *Advances in Neural Information Processing Systems* **2022**, *35*, 9460–9471.
- [9] Lee, K.; Smith, L.; Abbeel, P. Pebble: Feedback-efficient Interactive Reinforcement Learning Via Relabeling Experience and Unsupervised Pre-training. arXiv preprint 2021, arXiv:2106.05091.

- [10] Mosqueira-Rey, E.; Hernández-Pereira, E.; Alonso-Ríos, D.; Bobes-Bascarán, J.; Fernández-Leal, Á. Human-in-the-loop Machine Learning: A State of the Art. *Artificial Intelligence Review* 2023, 56, 3005–3054.
- [11] Hussein, A.; Gaber, M. M.; Elyan, E.; Jayne, C. Imitation Learning: A Survey of Learning Methods. ACM Computing Surveys (CSUR) 2017, 50, 1–35.
- [12] Torabi, F.; Warnell, G.; Stone, P. Behavioral Cloning from Observation. arXiv preprint 2018, arXiv:1805.01954.
- [13] Arora, S.; Doshi, P. A Survey of Inverse Reinforcement Learning: Challenges, Methods and Progress. *Artificial Intelligence* 2021, 297, 103500.
- [14] Rafailov, R.; Sharma, A.; Mitchell, E.; Manning, C. D.; Ermon, S.; Finn, C. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. *Advances in Neural Information Processing Systems* 2024, 36.
- [15] Christiano, P. F.; Leike, J.; Brown, T.; Martic, M.; Legg, S.; Amodei, D. Deep Reinforcement Learning from Human Preferences. Advances in Neural Information Processing Systems 2017, 30.
- [16] Meyers, J.; Fabian, B.; Brown, N. De Novo Molecular Design and Generative Models. *Drug Discovery Today* 2021, 26, 2707–2715.
- [17] Choung, O.-H.; Vianello, R.; Segler, M.; Stiefl, N.; Jiménez-Luna, J. Extracting Medicinal Chemistry Intuition Via Preference Machine Learning. *Nature Communications* 2023, 14, 6651.
- [18] Sundin, I.; Voronov, A.; Xiao, H.; Papadopoulos, K.; Bjerrum, E. J.; Heinonen, M.; Patronov, A.; Kaski, S.; Engkvist, O. Human-in-the-loop Assisted De Novo Molecular Design. *Journal of Cheminformatics* **2022**, *14*, 86.
- [19] Bjerrum, E. J.; Bachorz, R. A.; Bitton, A.; Choung, O.-h.; Chen, Y.; Esposito, C.; Ha, S. V.; Poehlmann, A. Scikit-Mol

Brings Cheminformatics to Scikit-Learn. *chemRxiv preprint* **2023**,

- [20] Blaschke, T.; Arús-Pous, J.; Chen, H.; Margreitter, C.; Tyrchan, C.; Engkvist, O.; Papadopoulos, K.; Patronov, A. REIN-VENT 2.0: An AI Tool for De Novo Drug Design. *Journal of Chemical Information and Modeling* **2020**, *60*, 5918–5922.
- [21] Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011, 12, 2825–2830.
- [22] RDKit: Open-Source Cheminformatics Software. 2021; https://www.rdkit.org, https://www.rdkit.org.
- [23] Riniker, S.; Landrum, G. A. Similarity Maps A Visualization Strategy for Molecular Fingerprints and Machine-learning Methods. *Journal of Cheminformatics* 2013, 5, 1–7.
- [24] Bjerrum, E. J.; Palunas, K.; Menke, J. Python-Based Interactive RDKit Molecule Editing with rdEditor. *chemRxiv preprint* 2024, 10.26434/chemrxiv-2024-jfhmw.
- [25] Qt for Python Team. PySide2: Python bindings for the Qt cross-platform application and UI framework. https://www. pyside.org, Accessed: 2024-04-08.