MACADAMIA Research Project Report

Human in the loop interaction in reinforcement learning for de novo molecular design

NGUYEN XUAN BINH

Aalto University

MACADAMIA
RESEARCH
PROJECT REPORT

Aalto University
MACADAMIA Research Project Report 2024

Human in the loop interaction in reinforcement learning for de novo molecular design

Nguyen Xuan Binh

MACADAMIA research project submitted in fulfillment of the requirements for the degree of Master of Science.
Otaniemi, 1st May 2024

Advisor: Yasmine Nahal, Prof. Samuel Kaski

Aalto University
School of Engineering
Master's Programme in Science and Technology

Author
Nguyen Xuan Binh

Title
Human in the loop interaction in reinforcement learning for de novo molecular design

School School of Science

Degree programme
Master of Science in Computer, Communication and Information Sciences (CCIS)

Major Machine Learning, Data Science and Artificial Intelligence (MACADAMIA)

Advisor Yasmine Nahal, Prof. Samuel Kaski **Code** SCI3044

Date 1 May 2024 **Pages** 24 **Language** English

Abstract

De novo molecular design, which makes use of computational techniques to engineer novel, diverse, synthesizable and drug-like molecular structures, is an essential step in the drug discovery process. Human experts in medicinal chemistry play a crucial role in defining the molecular design objectives through scoring functions. The formulation of human-in-the-loop workflow and comparison of three rival feedback mechanisms and user models — the scoring model, the pairwise comparison Bradley-Terry model, and ranking ListNet model, represent the main contributions of this study. The central task of this project is to generate novel and active molecules for DRD2 binding. The scoring model directly estimates the Dopamine Receptor D2 (DRD2) activity for each molecule, while the Bradley-Terry model evaluates the likelihood that one DRD2 binder is better than another. Finally, the ranking ListNet model estimates the preference ranking for a set of three molecules regarding their DRD2 binder as best, middle and worst rank.

We tested these user models under two noise levels when collecting human feedback and three different acquisition functions (random, uncertainty, and greedy sampling) to assess how well they perform in terms of accurately predicting the human objective. According to our results, the ListNet model consistently performs better than the other models in terms of molecular diversity and drug-likeness, the Bradley-Terry model performs best in terms of novelty with respect to known DRD2 binders, while the baseline scoring model excels most in chemical property-targeted molecule generation and synthesizability. This work shows how customized user scoring models through different feedback mechanisms can improve de-novo molecular design, even under noisy human input, and eventually lead to the identification of therapeutic compounds with ideal characteristics.

Keywords Reinforcement Learning, Human in the Loop, De Novo Molecular Design, REINVENT, RDKit, Cheminformatics, SMILES

urn https://aaltodoc.aalto.fi

Contents

1 Introduction

1.1 Introduction to cheminformatics

Cheminformatics is the study of how chemical and biological information is represented and used on computers. This field has numerous applications, including drug discovery, healthcare, data mining, and a variety of other areas. Examples of discovered drugs are benzylpenicillin, tamoxifen, zantac, and cortisone, which are all approved and frequently used medications that help to sustain human health.

Researchers frequently represent compounds as molecular graphs, which is extremely useful for data storage and analysis. For example, benzodiazepines, a type of sedative, share a shared ring system or scaffold, indicating that similar compounds frequently have similar bioactivity. Chemical databases, such as the PubChem database of the National Institutes of Health (NIH) Library of Medicine, are industrial-standard resources for drug discovery and other uses [1].

To represent molecules on computers, researchers usually use these two formats:

- **Molecular depiction (2D)**: it shows the arrangement of atoms and the bonds between them in a planar (flat) format [2].
- **Simplified Molecular Input Line Entry System (SMILES)**: The atoms and bonds are represented in a single line using ASCII string [3]. For example, the SMILES representation for a water molecule (H_2O) is "O". One advantage of SMILES is that it can be easily transformed to 2D or 3D structures using cheminformatics tools like RDKit [4] for visualization. Moreover, SMILES ensures that each molecule has a unique representation because it adheres to defined standards for atom ordering and bond traversal.
- **Morgan fingerprints** , also known as Extended Connectivity Fingerprints (ECFPs), are a molecular descriptor commonly used in cheminformatics to analyze and compare chemical structures. They are formed by iterating over each atom's neighborhood in a molecule, considering not only the immediate neighbors but also expanding outwards up to a defined radius. This method collects the local surroundings of each atom inside the radius and encodes it into a bit vector [5]. These fingerprints are generated as count vectors, in which the number of each feature type within the molecule is documented, or as bit vectors, in which the presence or absence of characteristics is recorded in a fixed-length bit string [5]. This project work uses the count vectors.

1.2 De Novo drug molecular design

De novo molecular design is the process of creating new molecular structures from scratch using computational approaches that also optimally satisfy a desired molecular profile [6]. This method plays a major role in drug development since it accelerates the chemical space exploration through the generation of new molecules having specific desired features. Advanced algorithms are often used to explore chemical space, producing and optimizing novel molecules based on established criteria such as biological activity, drug-likeness, and synthesizability.

According to Meyers et al. (2021), there are three types of de novo molcular design, which are atom-based, fragment-based and reaction-based [6]. The atom-based approach is supported by a vocabulary containing a small number of atoms and bonds. The atom-based method is backed by a vocabulary of few atoms and bonds. The reaction-based approach is backed by two sets of reactants and reaction rules. Finally, the fragment-based method is backed by a fragmentation scheme and a collection of interchangeable fragments. Therefore, researchers can find molecules with the best therapeutic qualities by searching the huge chemical space via paradigms proposed by Meyers (2021).

1.3 Human-in-the-loop De Novo drug design

Human-in-the-loop (HITL) methods have been rising in popularity since they allow humans to leverage their domain expertise to improve design process, therefore connecting machine learning, cognitive science, and psychology [7]. Reinforcement learning from human feedback (RLHF) is a popular framework that allows continuous and dynamic learning from human feedback instead of static, manually-engineered reward functions. It is a practical alternative that introduces a critical HITL component to the standard RL learning paradigm [8]. RLHF varies from RL by allowing humans in the loop to define and refine the objective, rather than specifying it beforehand.

One of the prominent works on RLHF for de novo molecular design is by Sundin et al. (2022). This work claims that medicinal chemist's intuition can perform as good as machine learning methods in de no vo molecular design, and it is reasonable that human intuition should be utilized. Specifically, they propose two approaches, where the first one refines the parameters of the reward function, which alleviates the chemist's difficulty in designing the the scoring function. The

latter approach constructs the scoring function from scratch, and the training dataset is augmented with chemists' intuition, where new molcules are sampled based on the chemist's desirable properties.

Besides RLHF, there are other diverse HITL methods for de novo molecular design as well. For example, the chemical space-based de novo design method allows researchers to specify desired features and explore across the huge chemical space to discover interesting molecules [9]. DrugMint, a website that predicts and designs drug-like molecules, uses predictive models to help scientists identify compounds with high drug-likeness scores [10]. DrugChat, a platform that uses graph neural networks and massive language models to provide feedback and ideas, enables users to query and revise chemical designs interactively [11]. These HITL approaches all show great potential for incorporating human expertise into de novo drug discovery when designing objective functions is challenging.

2 Research project

2.1 REINVENT: AI for molecular de novo design

REINVENT is an advanced AI tool designed for de novo drug design, providing a comprehensive platform for generating novel drug-like molecules [12]. The primary motivation behind REINVENT is to efficiently explore vast chemical spaces, identifying and optimizing molecules with desired biological activities and drug-like properties. This is motivated by the fact that traditional drug discovery methods are often time-consuming and costly, which REINVENT aims to address.

Regarding usage, REINVENT allows researchers to define target properties and constraints, then generates molecules that meet these criteria. Users can interact with the platform, providing feedback to refine and improve the generated compounds. This HITL approach ensures that the generated molecules are novel, diverse and synthetically practical [13].

REINVENT's iterative optimization involves multiple steps as follows [14]:

- Initialization: Users define the initial set of constraints and desired properties to optimize. This include diversity filters, inception and most importantly, the scoring function, which is designed to target chemical properties and serves as the basis for different user feedback models. All of user defined settings are written in a json file, which is given as input to REINVENT.

- Generation: The generative AI agent generates a batch of candidate molecules at each optimization step. Default value of batch size is 128, but we use 64 in this work and the default number of optimization steps is 100.
- Evaluation: Generated molecules are scored based on the scoring function. REINVENT tries to generate molecules that maximize the score.
- Optimization: The tool refines the molecules through multiple optimization steps, and finally return generated SMILES in the scaffold memory CSV file.

2.2 Project motivation

In this project, the general objective is to use REINVENT to generate molecules that have binding affinity for the dopamine receptor D2 (DRD2). This receptor is crucial to regulating dopamine, a neurotransmitter associated with motor control, motivation, and reward mechanisms [15]. Optimizing DRD2 binding affinity could provide improvements to inhibit diseases such as schizophrenia and Parkinson's disease. Specifically, antagonists of DRD2 appears to reduce tumor growth in cancer, making it a valuable property for developing anticancer treatments [16][17].

Nonetheless, the primary motivation for this project is to generate molecules that not only have high DRD2 affinity but are also novel (having new, possibly unseen scaffold structures) and they should also have drug-like properties. This is important for drug discovery to explore new chemical spaces that may unexpectedly have interesting properties. For instance, Wang et al. (2024) emphasize the need for novelty and drug-likeness in molecular design [18], while Fotie et al. (2023) discuss the importance of drug-like properties in drug design [19].

Another motivation of this project is to establish a HITL workflow, which involves human expertise to guide the molecular design process. This approach allows chemists to provide comparative feedback, which is often easier than direct scoring. Some studies have demonstrated the impact of HITL in enhancing molecular design through comparative feedback mechanisms [20].

The final motivation for this project is to develop a surrogate user model that can decently predict DRD2 affinity for novel molecules. While large datasets containing hundreds of thousands of molecules labeled with DRD2 affinity and a highly accurate pretrained model already exist, DRD2 is used here as a toy use case to demonstrate the effectiveness of our user feedback models. This allows us to evaluate our results using the existing oracle model [21].

2.3 Project problem statement

This research project aims to carry out these experiments

- Use REINVENT software as a reinforcement-learning model for generating molecules in SMILES format that maximizes probability that they bind to DRD2. In other words, this project actually does not develop any reinforcement learning framework, as it is already an existing model in REINVENT.
- Develop three distinct user models: the scoring model (taking as input one molecule and returning a scalar that represents how good it is), the pairwise comparing model (taking as input a pair of SMILES and returning which one is better) and the ranking model (taking as input a set of SMILES and ordering them from best to worst). This is the central contribution of this research project.
- Develop a HITL workflow that couples with REINVENT to generate novel molecules with human preference scores of chemical properties. This is the second contribution of this project.
- Develop an active learning framework where promising SMILES are selected for human feedback using various acquisition functions, with either perfect or noisy labelling. This is the third contribution of this project.
- Run the HITL workflow, benchmark all methods and report findings.

Unfortunately, due to the time limit, the author could not build a Graphical User Interface for interacting with real humans. Therefore, in this work, we used an ad-hoc surrogate ML model that acts as a human agent to provide preference feedback instead of interacting with a real human. We leave the latter for future work to validate our findings in a real-world drug design scenario.

3 Methodologies

3.1 Human-in-the-loop workflow

This project workflow is heavily influenced by the work of Sundin et al. (2022) [22]. In fact, the workflow is identical to the second case study (Human Chemist Component). The main difference is that various feedback mechanisms are tested to learn human preferences about the properties of generated molecules, while in Sundin et al., it seems that only binary preference feedback (like/dislike) is tested.

Additionally, this workflow directly uses the developed user models to score the molecules in REINVENT, and an oracle when interacting with the human to assess DRD2 affinity for the selected generated molecules, where the oracle model is fixed throughout the process. To recap, this workflow involves a generative AI assistant that helps a chemist to decide parameters of the multi-parameter optimization scoring function $S_{r,t}(x)$ iteratively at round r and iteration t, where r are rounds of molecule generation with REINVENT, and t are number of active-learning interactions with a human component to receive feedback for generated molecules [22]. The objective consists of K molecular properties $c_k(x)$ with relative weights w_k. The score of the kth property is measured using a scoring function $\phi_{r,t,k}$. Since we only maximize DRD2 affinity in our work, we have $k = 1$ here. Below is the pseudo-code from the paper but was directly modified to fit in this work [23].

Algorithm 1 Human-in-the-loop workflow

Require: An oracle that reliably estimates DRD2 affinity, a surrogate ML model that returns feedback as a scoring component $S_{\theta_{0,T}}$, number of REINVENT rounds R, number of human interactions T, number of queries Q at each interaction, acquisition function ACQ, oracle's noise level σ

$D_{0,T} \leftarrow \emptyset$ ▷ Initially, the training dataset is empty
for $r = 1, 2, \ldots, R$ **do** ▷ Looping over REINVENT rounds
  $S_{\theta_{r,1}} \leftarrow S_{\theta_{r-1,T}}$ ▷ Current round ML model is the ML model from last interaction of previous round
  $D_{r,1} \leftarrow D_{r-1,T}$ ▷ Current training dataset is the dataset from last interaction of previous round
  $U_r \leftarrow$ REINVENT$(S_{\theta_{r,1}})$ ▷ U_r: set of molecules from REINVENT using the ML model $S_{\theta_{r,1}}$
  $U_{r_{\text{best}}} \leftarrow$ Select top n_{best} molecules x with highest scores from U_r
  for $t = 1, 2, \ldots, T$ **do** ▷ Looping over online interactions with ML model
    for query $= 1, 2, \ldots, Q$ **do**
      $x^* \leftarrow \text{ACQ}(S_{\theta_{r,t}}, U_{r_{\text{best}}})$ ▷ Obtain new SMILES using the chosen acquisition function ACQ
      $y^* \leftarrow Oracle(x^*) + \mathcal{N}(0, \sigma)$ ▷ Acquire feedback y^* of DRD2 affinity for x^* SMILES from Oracle plus some noise
      $U_{r_{\text{best}}} \leftarrow U_{r_{\text{best}}} \setminus x^*$ ▷ Remove x^* SMILES from $U_{r_{\text{best}}}$
      $D_{r,t} \leftarrow D_{r,t} \cup \{(x^*, y^*)\}$ ▷ Update the dataset with new queries
    end for
    $S_{\theta_{r,t}} \leftarrow S_{\theta_{r,t}}$ retraining on $D_{r,t}$ ▷ The ML model is updated
  end for
end for

In reality, the humans usually only have a vague intuition of DRD2 affinity initially, then shape their preferences over time as REINVENT recommends better DRD2 binders. Furthermore, because evaluating or comparing thousands of molecules is extremely time-consuming even for experienced chemists, an ML model is used to act as a surrogate human in the REINVENT scoring function. At

the beginning of the molecule generation process, this surrogate model predicts DRD2 affinity with very low accuracy (around 0.5). Gradually, during interactions, the ML model would acquire new labelled molecules, selected through different acquisition methods (random, uncertainty or greedy sampling).

A question may arise from this workflow is that, if there is already access to an oracle that can reliably estimates DRD2 affinity, why should one invest time to train a weak ML model from scratch when we can directly use the oracle as the scoring component in REINVENT. The motivation for this workflow is that we assume the function $Oracle(x^*)$ is an expensive black-box evaluator (in this case, a human evaluator) that we cannot call for every generated molecule, instead we call a much cheaper ML model that acts as a surrogate to that expensive evaluator, and easier to integrate in the scoring function.

To conclude, the settings used in this project are $R = 3$ REINVENT rounds, $T = 5$ HITL interactions, $Q = 56$ queries, and REINVENT batch size of 64 with 100 optimization steps. Initial training dataset size is 200 molecules of balanced classes (100 DRD2 actives and 100 DRD2 inactives), and after 3 REINVENT rounds of 5 active learning iterations, the final training dataset size becomes $200 + 3 \times 5 \times 56 = 1040$ SMILES used to train the last surrogate ML model.

3.2 Feedback model architecture

This project develops three user feedback models for REINVENT during the scoring stage: the scoring model (baseline), the pairwise comparison model (Bradley-Terry) and the ranking model (ListNet). While they differ in the type of feedback they learn from REINVENT, they share a similar neural network architecture used to predict the DRD2 affinity. This aspect is convenient because we would prefer these feedback models to infer DRD2 affinity for new SMILES quickly. The figure below shows the simple architecture for learning DRD2 affinity, where the Sigmoid activation ensures that the logit is restricted to [0, 1].

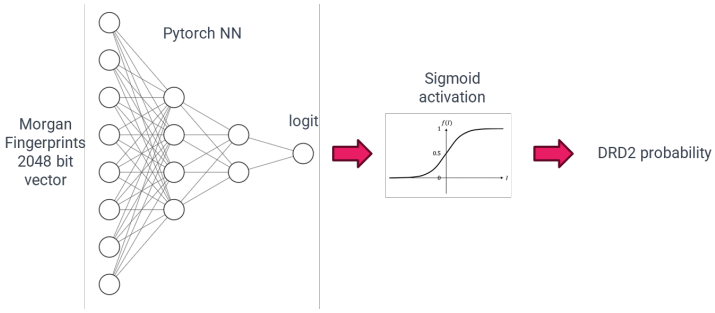

Figure 1. *ML model architecture for learning and predicting DRD2 affinity*

Each user feedback model takes in a different input and provides a distinct type of output: the scoring model takes in as input the vector representation of a DRD2 binder proposed by REINVENT and returns a scalar that represents its DRD2 affinity, the Bradley-Terry model takes in as input a pair of DRD2 binders and returns a scalar representing a pairwise preference, and finally the ListNet model takes in as input a batch of proposed DRD2 binders and returns a scalar representing the most preferred one. Details of the three feedback models are demonstrated below

Scoring model

The scoring model predicts the probability that a given SMILES string has DRD2 activity based on its Morgan Extended-Connectivity Fingerprint (ECFP) vector. This is a straightforward feedback that is already supported by REINVENT scoring methods.

Inputs: Vector of Morgan fingerprints of one single SMILES.
Outputs: Probability that the SMILES has DRD2 activity (score between 0 and 1)
Output for a single SMILES during REINVENT scoring: The same as outputs
Outputs for a batch of SMILES during REINVENT scoring and data acquisition: The same as outputs for all SMILES.
Loss function to train Pytorch model: BCELoss, defined as
$\text{BCE}(p, q) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p(y_i)) + (1 - y_i)\log(1 - p(y_i))\right]$ where $y_i \in \{0, 1\}$ representing the DRD2 activity label, with label 1 corresponding to the active class and 0 to the inactive class. Threshold value applied to convert DRD2 activity scores to binary labels is 0.5.

Pairwise comparison model: Bradley-Terry

As mentioned earlier, the motivation for comparing pairs of molecules stems from the idea that humans may have difficulty in providing a DRD2 affinity, and they may find it easier to tell which molecule is more likely to have DRD2 affinity than the other. Using the Bradley-Terry formulation, we first define $\beta_i \in \mathbb{R}$ as the DRD2 affinity for SMILES 1, and $\beta_j \in \mathbb{R}$ the DRD2 affinity of SMILES 2 given by the neural network classifier described in Figure 1, and let the outcome of a comparison between SMILES (i, j) be determined by $\beta_i - \beta_j$. The Bradley-Terry model treats this outcome as an independent Bernoulli random variable with

distribution Bernoulli(p_{ij}), where the log-odds corresponding to the probability p_{ij} that SMILES i is better than SMILES j [24], which is defined as

$$p_{ij} = \frac{e^{\beta_i - \beta_j}}{1 + e^{\beta_i - \beta_j}} = \frac{1}{1 + e^{-(\beta_i - \beta_j)}} = Sigmoid(\beta_i - \beta_j)$$

The details of the feedback for Bradley-Terry feedback model is

Inputs: Morgan fingerprints of two different SMILES.

Outputs:: Probability p_{ij} that the first SMILES is better than the second SMILES in terms of DRD2 affinity.

Output for a single SMILES during REINVENT scoring: The feedback is 1 if the first SMILES is better than the second SMILES (if predicted probability by Bradley-Terry model is greater than 0.5) else 0 (which means that the second SMILES is better than the first one)

Outputs for a batch of SMILES during REINVENT scoring and data acquisition:

1. We perform P_2^N permutations of pairs of SMILES from N (batch size) SMILES generated by REINVENT. This ensures that the relative comparison for all SMILES is comprehensive. A SMILES with a higher DRD2 affinity than most other SMILES in the batch is considered better.

2. For each pair, we calculate the preference score of SMILES 1 against SMILES 2, then single SMILES outputs (0 or 1) for the first SMILES are aggregated.

3. We return the average aggregated score for all SMILES in the batch, which is the total sum above divided by $N - 1$

Loss function to train Pytorch model: BCELoss.

Ranking model: ListNet

ListNet is a listwise approach for learning to rank, which aims to directly optimize the ranking of a list of items rather than individual pairs as it is the case with the Bradley-Terry model [25]. Its motivation may stem from the idea that the pairwise comparison model can be quite limited since it provides a take-all or lose-all feedback, while the ranking model can provide a more neutral rating between molecules.

In general, the ListNet architecture consists of the following components:

- Input representation: Each SMILES is represented by feature vectors (Morgan fingerprints)

- Softmax function: To convert the raw scores produced by the neural network into probabilities, ListNet uses a softmax function, which ensures that the probabilities of all items in the list sum to one, providing a normalized ranking distribution.

- Loss function: The ListNet uses the Kullback-Leibler Divergence (KLDivLoss) to measure the difference between the predicted ranking distribution and the ground truth ranking distribution.

The details of the ListNet ranking feedback model are

Inputs: Morgan fingerprints of three different SMILES.
Outputs: Softmax preference scores for three SMILES with respect to DRD2 affinity.
Output for a single set of 3 SMILES during REINVENT scoring: SMILES with lowest softmax score receives rank 0, middle one receives rank 1, and highest one receives rank 2. Then, they are normalized to [0, 0.5, 1].
Outputs for a batch of SMILES during REINVENT scoring and data acquisition:

1. We perform C_3^N combinations of sets of three SMILES proposed by REINVENT. Due to the exponential increasing nature of combinations, the batch size should not be too large. A reasonable value for batch size is 64.

2. We calculate preference scores for each set of three SMILES, then obtain the ranks [0, 1, 2] and normalize them to [0.0, 0.5, 1.0]. Scores for all SMILES are then aggregated.

3. We return the average aggregated score for all SMILES by dividing by C_2^{N-1}, which is the number of times each SMILES appear in all combinations.

Loss function: KLDivLoss (Kullback-Leibler divergence), defined as

$$\mathrm{KL}(P \parallel Q) = \sum_i P(i) \log \left(\frac{P(i)}{Q(i)} \right)$$

where P is is the true probability distribution and Q is the estimated probability distribution. The KL divergence measures how the predicted probability distribution diverges from the true probability distribution, and this divergence should be minimized to enable the ListNet model to learn the preference scores.

3.3 Active learning: acquisition functions

Active learning is a semi-supervised learning as it uses both labeled and unlabeled data. New samples get annotated in an iterative process, where a query strategy is used to choose an example to be queried, and once labeled by an oracle, will result in a model accuracy increment [26].

In this project, we compare three query strategies, otherwise known as acquisition functions, to acquire new SMILES generated by REINVENT for labelling by the oracle: random, uncertainty and greedy acquisition methods. All acquisition methods are applied to the three types of user feedback models and adapted to their respective outputs.

Random acquisition: selects molecules randomly from the pool of generated SMILES or molecules, denoted as $U_{r_{best}}$. This approach does not consider user feedback outputs but relies on randomness to ensure a diverse selection of molecules

Uncertainty acquisition: selects molecules for which the user feedback model has the least confidence in its predictions. The usual method for determining this is to choose molecules whose projected scores are closest to 0.5, which denotes maximum uncertainty. For the scoring model, molecules with DRD2 affinity close to 0.5 are selected for active learning. For the Bradley-Terry model, pairings of molecules with preference scores near 0.5 are chosen, which indicate that the model is unsure which SMILES within the pairs is best. For the ListNet ranking model, the SMILES with a normalized ranking of 0.5 is chosen, which indicates the ML model's uncertainty about the exact ordering for that SMILES (neither the best nor the worst).

Greedy acquisition: selects molecules that have the highest predicted DRD2 affinity based on the scores given by the user feedback models. This approach aims to maximize the DRD2 affinity by always choosing the top-performing molecules. For the scoring model, this involves selecting the molecules with the highest DRD2 affinity probabilities. For the Bradley-Terry and ListNet models, it involves selecting the most preferred SMILES.

All of the acquisition functions would be tested along with two level of oracle's labelling noise: 0 and 0.1. The oracle is a Random Forest classifier trained on a large dataset of DRD2 actives and inactives that outputs the DRD2 affinity for a given molecule represented as a vector of Morgan fingerprints. A noise of 0 added to the oracle output simulates a human providing a near perfect estimation of the

DRD2 affinity with respect to the ground truth given by the oracle, while a noise of 0.1 simulates a human whose labelling is not entirely accurate with respect to the oracle. Considering three user feedback models, three acquisition methods and two oracle noise levels, there are in total $3 \times 3 \times 2 = 18$ different running cases to perform. They would all be benchmarked in the Results section.

3.4 Novelty, diversity, SA, and QED score

As per objectives of de novo molecular design, the novelty of generated molecules with respect to already existing ones in the literature, the molecular diversity, the synthetic accessibility (SA) referring to how easy is a molecule to be synthesized in the lab, and the quantitative estimate of drug-likeness (QED) are important metrics commonly used by drug design practitioners to assess the quality of AI generated molecules. All metrics are essential to guarantee that the molecules produced are not only novel and diverse but also easy for synthesis and likely to be selected as drug candidates.

- **Novelty score**
  - **Definition**: The novelty score measures the fraction of the generated molecules not present in the training set, used to pre-train the user feedback models before the start of the molecule generation process with REINVENT [27]. This score is within [0, 1].
  - **Range meaning**: Novelty scores close to 0 indicate a lack of exploration due to overfitting to the user feedback model used to score the molecule, while novelty scores close to 1 mean that the user feedback model favors the discovery of new structures [27].
- **Diversity score**
  - **Definition**: The diversity score measures the internal chemical diversity within the generated molecules set G [28]. This score is within [0, 1] and is given by

$$\text{IntDiv}_\text{p}(G) = 1 - \sqrt[p]{\frac{1}{|G|^2} \sum_{m_1, m_2 \in G} Tanimoto(m_1, m_2)^p}.$$

    This metric helps detect mode collapse, which means when the user feedback model identifies promising molecules in a limited chemical space and ignores other regions of the space [27].

- **Range meaning**: A higher value of internal diversity indicates larger area of chemical space covered within the generated set and vice versa.

- **Synthetic accessibility (SA) score**
  - **Definition**: The SA score evaluates how easily a molecule can be synthesized, ranging from 1 (very easy) to 10 (very difficult).
  - **Range meaning**: Molecules with SA scores between 1 and 3 are considered preferable as they are more likely to be easily synthesized efficiently in a laboratory setting [29].

- **QED score (Quantitative estimate of drug-likeness)**
  - **Definition**: The QED score is a composite metric that evaluates the drug-likeness of a compound based on several molecular properties, ranging from 0 to 1.
  - **Range meaning**: A QED score above 0.5 is considered the threshold for drug-likeness, making it a useful filter in drug discovery [30].

3.5 Molecular descriptor filters

Previously, it is mentioned that besides DRD2 maximization, we also aim to generate drug-like molecules based on molecular descriptors, which heavily influence pharmacokinetic parameters. In general, pharmacokinetic (PK) parameters play a crucial role in understanding how a drug interacts with the body during drug development. These parameters provide details into various aspects of a drug when it travels through the body, known as ADME: Absorption, Distribution, Metabolism, and Excretion [31][32]. They include Cmax (highest plasma concentration of a drug after administration), Tmax (time to maximum concentration), half-life (time it takes for the drug concentration in the body to decrease by half), AUC (total exposure to a drug over time), volume of distribution (how extensively a drug distributes into tissues relative to its concentration in plasma), and clearance (rate at which the drug is removed from the body) [33]. Usually, the properties that affect these PK parameters are the molecular descriptors, which are used to characterize the physical, chemical, and structural properties of molecules. Together, they can be called filters, since only drugs that bypass these molecular descriptor filters' threshold should be screened for further evaluation. This cuts down screening time and improves drug-likeness of molecules.

- **LogP (Partition coefficient)**
  - **Definition**: LogP quantifies the lipophilicity of chemical compounds, which influences their absorption, distribution, and overall pharmacokinetic properties.
  - **Filter range**: Compounds with a logP between 1 and 5 are typically preferred, as they are more likely to have favorable pharmacokinetic profiles [34].
- **Molecular weight**
  - **Definition**: This is the mass of a molecule measured in Daltons.
  - **Filter range**: Molecules with a molecular weight under 500 Daltons are generally considered ideal for oral drug candidates as per Lipinski's rule of five [34].
- **Hydrogen bond donors (H-donors) and acceptors (H-acceptors)**
  - **Definition**: H-donors are atoms in a molecule that can donate hydrogen bonds, whereas H-acceptors are atoms that can accept hydrogen bonds.
  - **Filter range**: According to Lipinski's rule of five, having no more than 5 hydrogen bond donors and no more than 10 hydrogen bond acceptors is preferred for good bioavailability [34].
- **TPSA (Topological Polar Surface Area)**
  - **Definition**: TPSA is the surface area of a molecule that is polar.
  - **Filter range**: Molecules with a TPSA of 140 $Å^2$ or less are more likely to have good oral bioavailability, as per Veber's rule [35].
- **Number of Rotatable Bonds**
  - **Definition**: This is the number of bonds in a molecule that can rotate freely.
  - **Filter range**: According to Veber's rule, having 10 or fewer rotatable bonds is ideal for maintaining good oral bioavailability [35].
- **Number of Rings**
  - **Definition**: This is the number of ring structures within a molecule.

– **Filter range**: The presence of ring structures can impact the rigidity and overall stability of a molecule. Muegge's rule suggests that having up to 7 rings is favorable for drug-like properties [36].

The table below summarizes all objectives and filters used in this project

	Description	Lower	Higher	Known rules
DRD2 affinity	Objective to maximize	0.75	1.0	0.75 is average DRD2 probability of SMILES actually having DRD2 predicted by TDC Oracle
Molecule weight	Filtering	0	500.0	Lipinski's rule of five [34]
Hydrogen bond donors number	Filtering	0	5	Lipinski's rule of five [34]
Hydrogen bond acceptors number	Filtering	0	10	Lipinski's rule of five [34]
TPSA	Filtering	0.0	140.0	Veber et al. [35]
Number of rotatable bonds	Filtering	0	10	Veber et al. [35]
Number of rings	Filtering	0	7	Muegge et al. [36]

4 Results

4.1 Performance metrics evolution between running cases

The ROC (Receiver Operating Characteristic) curve and AUC (Area Under the Curve) are used to evaluate the performance of binary classification for DRD2 affinity by plotting the true positive rate against the false positive rate at various thresholds applied to the predicted DRD2 affinity. The AUC gives a single scalar value to quantify the overall performance of the classifier. Random classifier has AUC of 0.5 and the better the classifier, the higher its AUC score.

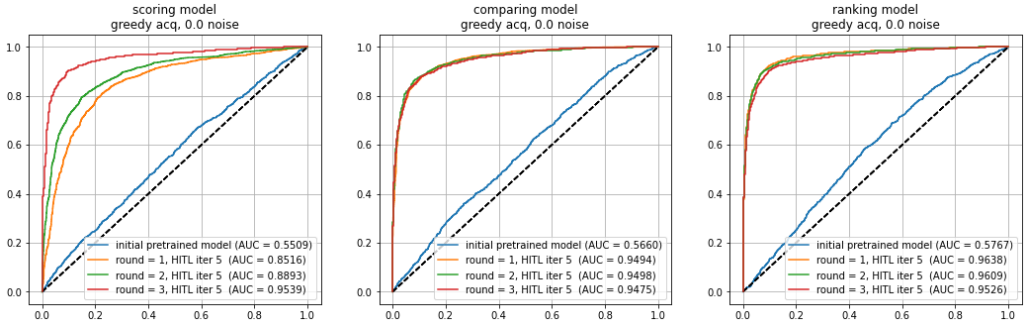

Figure 2. *ROC curve improvement after 3 REINVENT rounds*

Initially, the pre-trained models show relatively low AUC values as they have limited ability to tell apart molecules with and without DRD2 affinity because they

have too little knowledge about human preferences for DRD2 binders before the start of feedback collection. However, at each HITL iteration (in total 5 iterations were performed), we observe a considerable improvement in AUC values for all three models.

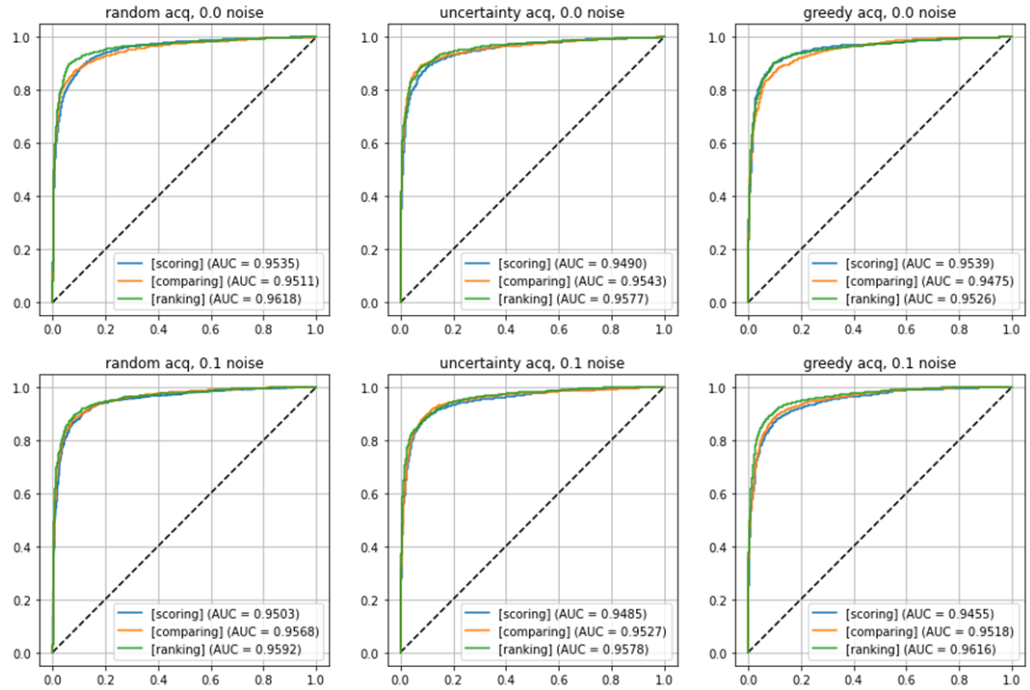

Figure 3. *ROC curve for DRD2 classification at the last REINVENT round and fifth HITL iteration*

Across all running cases, the AUC values indicate that all models perform well as expected, with minimal differences between them. The ListNet ranking model consistently achieves the highest AUC values, followed closely by the Bradley-Terry pairwise comparison model and the baseline scoring model. This suggests that while all three models are effective in predicting DRD2 affinity, the ranking model slightly outperforms the others, providing the most reliable predictions of DRD2 affinity according to the human oracle, therefore learning about human preferences about the DRD2 task the most effectively.

4.2 Performance metrics comparison between running cases

Five metrics of accuracy, precision, recall and Matthew's correlation coefficient (MCC) are obtained by evaluating the user feedback models on a testing dataset consisting of 2000 SMILES labelled according to their DRD2 affinity (1000 presenting DRD2 affinity and 1000 not presenting it). Such dataset is balanced and large enough to reliably assess the performance of the learned user models. For clarity, the formula of the MCC metric is given as:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where:

- TP = True Positives, TN = True Negatives
- FP = False Positives, FN = False Negatives

Looking at the heatmap below, we can observe that most models have high precision, which is not what we care mostly about since the positive case (DRD2 active) is much rarer during the training process. As a result, we should focus on finding the configuration that provides the highest recall, therefore allowing to accurately identify DRD2 actives during the molecule generation process.

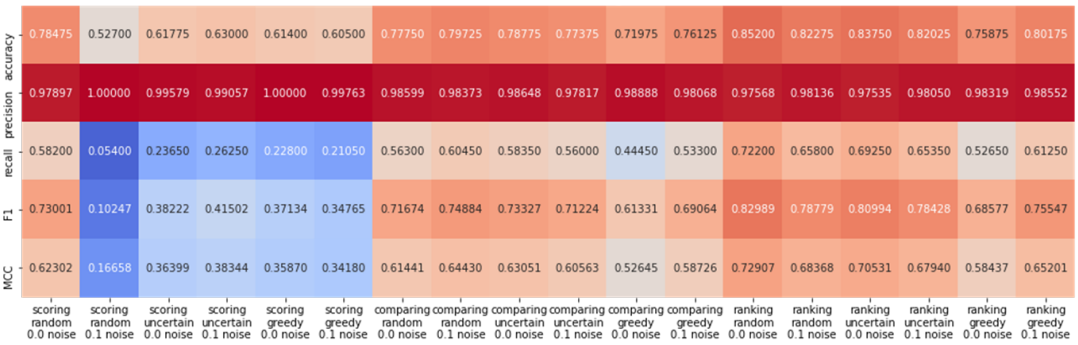

Figure 4. *Performance metrics following the last re-training of the user feedback model for all 18 running cases*

It is apparent that the ListNet ranking model consistently has higher recall than the Bradley-Terry comparing model, which in turn has higher recall than the scoring model. Interestingly, this suggests that scoring molecules through comparing pairs or ranking subsets can better capture the preference feedback during the molecule generation process compared to direct scoring by providing scalar values. Specifically, the best configuration is the ListNet ranking model trained using random acquisition of new molecules labelled without human noise, since it corresponds to the highest AUC, recall, F1 and MCC scores. This suggests that acquiring diverse SMILES randomly from different regions of the chemical space results in improving the DRD2 classification by the ListNet ranking model.

4.3 Percentage of filtered generated molecules

This table shows the percentage of molecules that satisfy the lower and higher thresholds for the filtering molecule descriptors. We can observe that most molecules generated by REINVENT already satisfy the filters except for LogP

Figure 5. *Percentage of molecules combined from all rounds that pass through each filter*

and the number of hydrogen bond donors. This can probably indicate that the constraints on LogP and number of hydrogen bonds are less common for drug-like molecules. Regarding DRD2 affinity, it is apparent that the baseline scoring model is the most capable of generating molecules with high DRD2 affinity according to the oracle, followed by the Bradley-Terry comparing model and lastly by the ListNet ranking model. This is the major advantage of the baseline scoring model when we need to directly optimize some properties instead of focusing on other aspects such as novelty or diversity. As a result, after passing through all filters, the scoring model tends to have larger percentage of filtered SMILES, followed by the comparing model and lastly the ranking model.

4.4 DRD2 affinity of generated molecules

Previously, we have observed that the scoring model has the highest percentage of SMILES passing through all filters, but it does not imply that the scoring model in general generates molecules with highest DRD2 affinity probabilities. Therefore, it is necessary to compare the DRD2 affinty probabilities before and after applying the filters, which are labelled "unfiltered" and "filtered" respectively.

Coincidentally, the scoring model also happens to generate molecules with the highest DRD2 affinity probabilities according to the oracle, as the mean value and 75 percentile are consistently higher than those of the comparing and ranking models by a large margin in the figure below. The Bradley-Terry comparing model

tends to generate molecules with higher DRD2 affinity than the comparing model, which is more apparent among unfiltered moleclules but less visible post-filtering. Overall, we can conclude that, in terms of generating more DRD2 binders, using the scoring model remains the best, followed by the comparing model and lastly the ranking model. This is true for both unfiltered and filtered SMILES.

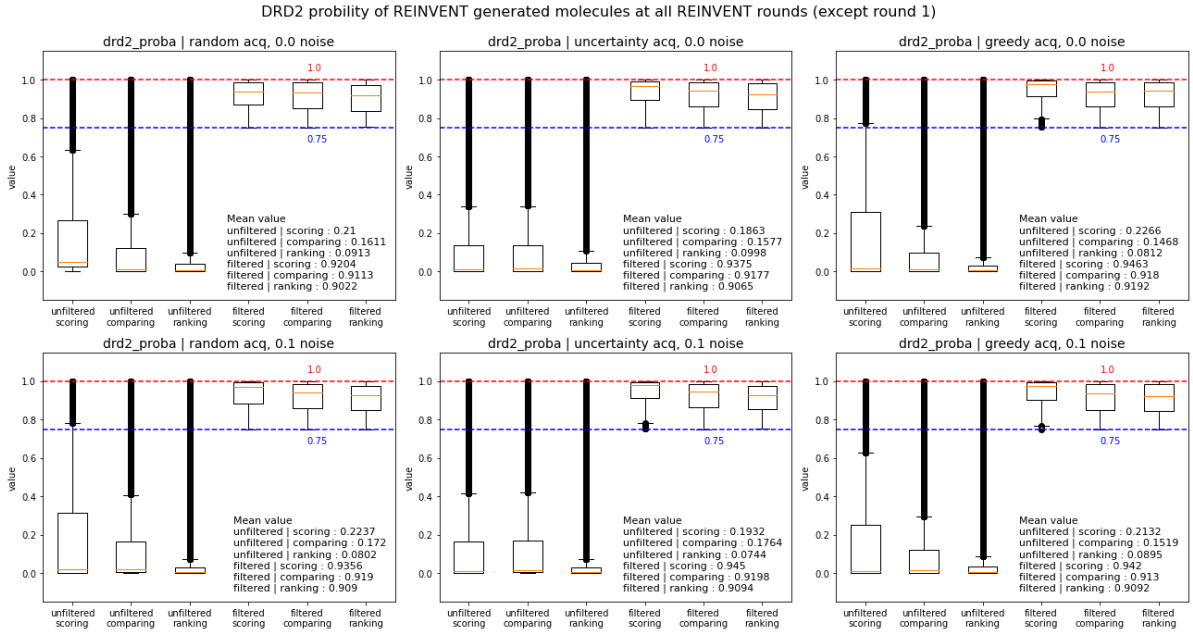

Figure 6. *DRD2 affinity of generated molecules combined from all rounds*

4.5 Benchmark scores of generated molecules

In this part, we evaluate the novelty, diversity, synthesizability and drug-likeness criteria to provide a fairer assessment of the three user feedback models and their potential usefulness in practical real-world applications. There are 6 configurations from unfiltered version (3 acquisition $\times$ 2 noise levels) and another 6 from filtered versions. Together, there are 12 configurations for us to compare three user feedback models.

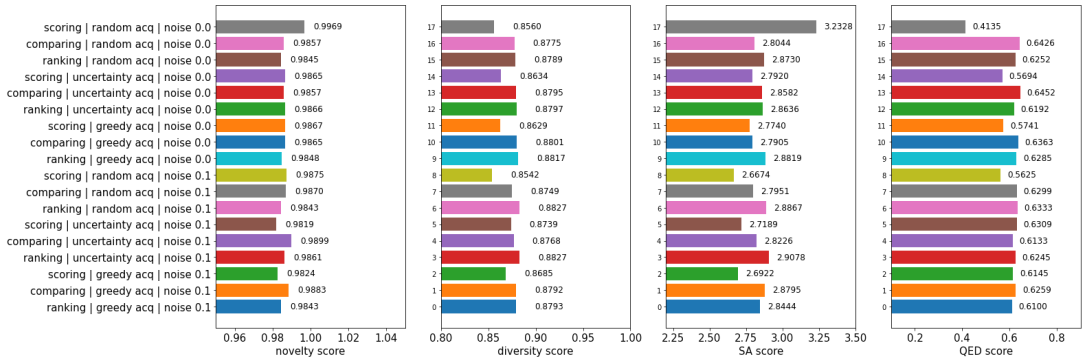

Figure 7. *Benchmark scores of unfiltered generated molecules combined from all rounds*

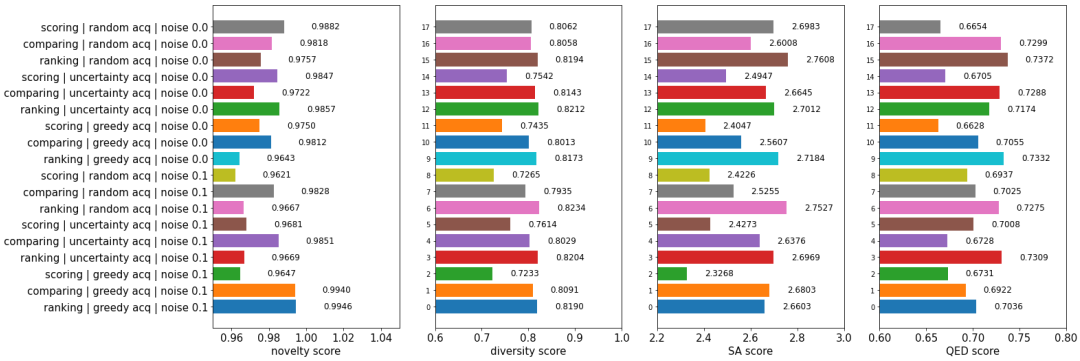

Figure 8. *Benchmark scores of filtered generated molecules combined from all rounds*

Regarding the novelty score, it is hard to determine which model is most capable of generating novel molecules with respect to the initial training sets, but it appears that the comparing model has the highest novelty score for 5 out of 12 configurations, followed by 4 out of 12 configurations for the scoring model and 3 out of 12 for the ranking model.

Regarding the diversity score, we found out that the ranking model consistently outperforms other models in all 12 configurations considering both filtered and unfiltered SMILES. This strongly suggests that the ranking feedback mechanism is much more versatile to encourage generative models like REINVENT to explore larger chemical spaces that are likely to possess the targeted chemical properties. We also conclude that the comparing model, while always resulting in lower diversity score than the ranking model, consistently outperforms the scoring model.

Regarding the SA score, we first note that molecules with low SA scores generally have several common characteristics. They often contain simple building blocks and tend to lack highly strained structures and have a small number of functional groups that are hard to manipulate [29]. We have an additional discovery from the two graphs, where the scoring model consistently results in the lowest SA scores for 10 out of 12 configuratons. This suggests that, due to the lack of diversity as mentioned earlier, it is likely to explore a limited chemical space that is more likely to contain more DRD2 binders that are similar to each others. This by-product process makes the molecules generated by the scoring feedback model more likely to be easily synthesizable. Then, we observe that the Bradley-Terry comparing model leads to a higher SA score followed by the ranking model which results in the highest SA score. This means that the SA score is in a way inversely proportional to the diversity score: the more diverse a set of molecules, the harder it is to synthesize all of them.

Regarding the QED score, it appears that the comparing model and the ranking model have close competition, where the comparing model tops 5 out of 12 configurations and ranking model tops 6 out of 12 configurations. Interestingly, the comparing model outperforms others mostly when considering unfiltered SMILES, while the ranking model outperforms others mostly after applying the filters to the generated SMILES. This suggests that overall, even though the Bradley-Terry model can encourage the generative model to discover more drug-like molecules, these are not guaranteed to satisfy the targeted chemical properties. On the other hand, the ranking model not only generates molecules with good targeted chemical property (DRD2 affinity) but it also balances them with high QED scores as well. To conclude, the ranking model performs best, followed by the comparing model and lastly the scoring model in terms of satisfying the drug-likeness criterion.

4.6 Examples of best generated molecules

To have a better visualization of the generated molecules by REINVENT, we can use RDKit to plot the best SMILES from each of the 18 running cases, considering the filtered SMILES. The best SMILES is selected by first choosing the top 20 SMILES with highest QED scores, then sorting them according to the SA score and keeping the top 10 SMILES, and finally sorting the 10 SMILES according to their DRD2 affinity and fetch the top one.

Generally speaking, the molecules produced by the scoring model have simple structures with fewer branches and functional groups. A design bias towards high predictability and ease of synthesis can be seen by the compounds' common pharmacophores and synthetic routes.

In contrast to the scoring model, the molecules produced by the comparing model show greater structural complexity and diversity. The different ring systems and functional groups demonstrate this diversity and show how well the model is able to distinguish between the relative benefits of various chemical configurations.

Finally, the ranking model produces molecules that maintain a balance between drug-likeness, diversity, and novelty. These molecules are more oriented towards reaching a high rank across various metrics, and they display a combination of diverse functional groups and complex ring systems. This indicates that such model successfully incorporates ranking feedback into the scoring procedure to identify promising molecules according to their affinity for DRD2 as well as their drug-likeness potential.

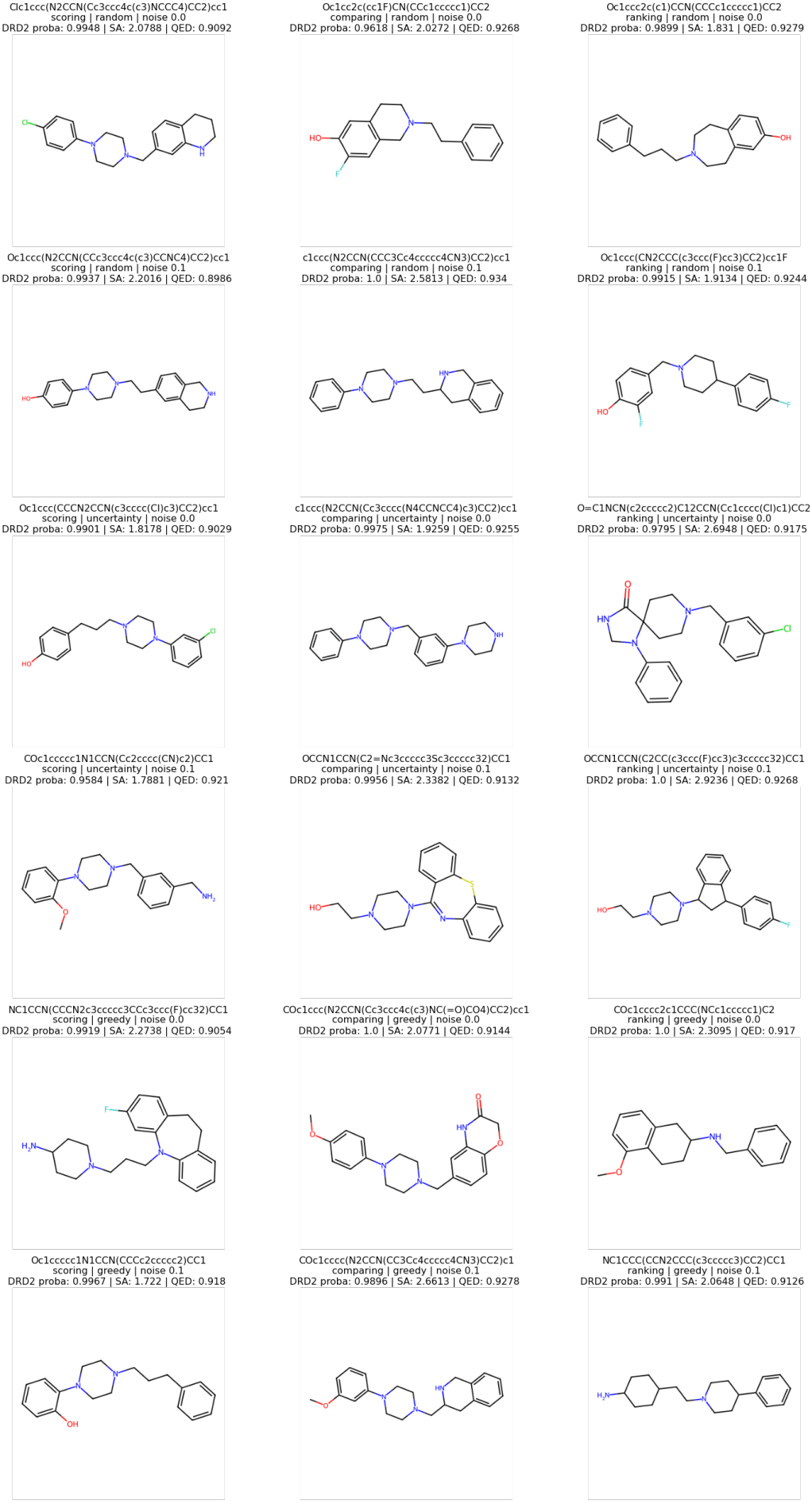

Figure 9. *Molecules generated by different user feedback models*

5 Discussion

5.1 User feedback model comparison

In brief, the ranking model (ListNet) consistently excels in DRD2 classification with high MCC score, passing the molecular descriptor filters, maximizing diversity, and achieving high QED scores. On the other hand, the scoring model shows the best performance in generating molecules that minimize the SA score and maximize the average DRD2 affinity among the generated high-scoring molecules. Finally, the comparing model (Bradley-Terry) keeps a balance between these two models, performing moderately well across all criteria, particularly excelling in generating molecules with high novelty score. We can summarize the performance of the three user feedback models in the table below, where rank 1 means best and rank 3 means worst.

	scoring model	**comparing model Bradley Terry**	**ranking model ListNet**
ML model with highest MCC for DRD2 classification	3	2	1
Generating molecules that are likely to pass molecular descriptor filters	3	2	1
Generating molecules that are likely to pass the DRD2 probability threshold	1	2	3
Generating molecules that maximize DRD2 affinity	1	2	3
Generating molecules that maximize the novelty score	2	1	3
Generating molecules that maximize the diversity score	3	2	1
Generating molecules that minimize the SA score	1	2	3
Generating molecules that maximize the QED score	3	2	1

5.2 Acquisition function comparison

We also conclude about the comparison between different acquisition functions. The findings and rankings are only a general trend, and naturally there are result variations within each ranking.

The random acquisition function results in the highest MCC for DRD2 classification but is the worst in terms of producing molecules that pass all molecular

descriptor filters and the DRD2 affinity threshold. It excels at maximizing novelty and diversity scores, showing higher ability to explore more chemical spaces.

Uncertainty acquisition balances performance by placing second in most categories, generating diverse and novel molecules while retaining intermediate scores for molecular descriptor filters and DRD2 affinity threshold criteria.

Greedy acquisition is the best at generating molecules that pass molecular descriptor filters and the DRD2 affinity threshold, as well as maximizing the DRD2 affinity and minimizing the SA score, but it ranks last in terms of maximizing novelty and diversity, focusing more on exploitation.

	random acquistion	**uncertainty acquisition**	**greedy acquisition**
ML model with highest MCC on DRD2 classification	1	2	3
Generate molecules that are likely to pass molecular descriptors	3	2	1
Generate molecules that are likely to pass DRD2 probability threshold	3	2	1
Generate molecules that maximize DRD2 affinity	3	2	1
Generate molecules that maximize novelty score	1	2	3
Generate molecules that maximize diversity score	1	2	3
Generate molecules that minimize SA score	3	2	1
Generate molecules that maximize QED score	1	2	3

5.3 Human noise comparison

Finally, we can assess across all previously described metrics the performance of the three user feedback models at two different noise levels (0.0 and 0.1) added to the oracle output to simulate noisy humans. Feedback models trained using a noise level of 0.1 regularly outperform the others (trained using a noise level of 0.0) in terms of producing molecules that pass molecular descriptor filters while maximizing novelty, diversity, and QED scores. However, models trained using a noise level of 0.0 display greater performance in terms of producing compounds that maximize the average DRD2 affinity while minimizing the SA scores. Which noise level is most reasonable for passing the DRD2 affinity threshold and achieving the highest MCC score for DRD2 classification remain questionable. This shows

	noise 0.0	noise 0.1
ML model with highest MCC for DRD2 classification	uncertain	uncertain
Generating molecules that are likely to pass molecular descriptors	2	1
Generating molecules that are likely to pass DRD2 affinity threshold	uncertain	uncertain
Generating molecules that maximize DRD2 affinity	1	2
Generating molecules that maximize the novelty score	2	1
Generating molecules that maximize the diversity score	2	1
Generating molecules that minimize the SA score	1	2
Generating molecules that maximize the QED score	2	1

that introducing noise when labelling newly acquired molecules by the oracle can improve some elements of molecule production (targeting novelty and diversity) while reducing criteria like finding more true positives for the DRD2 receptor as well as synthesizability which are expected to be negatively impacted if the provided feedback is noisy.

6 Conclusion

When working with HITL assisted de novo molecular design, the feedback mechanism, acquisition function and the noise when providing the feedback offer a wide variety of customization. In other words, depending on the objective of the molecule generation task, we can use a specific configuration to ensure that it will be likely to achieve quantifiable results.

To focus on generating molecules that directly targets chemical properties effectively, such as DRD2 affinity and synthesizability, the scoring model trained using the greedy acquisition function without labelling noise when providing the feedback is a good option to consider for initial experiments. The scoring model is preferable when computational resources are limited.

If the aim is to generate novel molecules while keeping a balance across all metrics, the Bradley-Terry pairwise comparison model trained using the uncertainty acquisition function is a better configuration option. The Bradley-Terry comparing

model however needs more computational resources than the scoring model.

However, if the aim is to generate more diverse and drug-like molecules, the ListNet ranking model trained with using a random data acquisition and noisy labelling is possibly the best option. It also seems to be the most performant in terms of learning from preference data that human experts can provide. However, the ranking model requires extensive computational resources during each optimization step of REINVENT since all sets of three generated molecules (combinatorial) are computed to infer all preferences.

In conclusion, this research highlights the impact of several user feedback models on de novo molecular design with REINVENT when used for scoring the generated molecules. We used a baseline scoring model, a pairwise comparison model and a list ranking model in separate experiments to generate compounds with high DRD2 affinity, novelty, diversity, synthesizability and drug-likeness, with variable degrees of success. Future research could further refine the architecture of these feedback models or introduce new feedback mechanisms, such as multiple binary preference (like/dislike for k properties of a molecule) or molecule editing (proposing a slightly modified molecular structure expected to be more promising than the original one). Furthermore, we aim to validate our findings by performing real human experiments, where medicinal chemists interact with the feedback system through a Graphical User Interface.

7 Appendix

The project code for this research project can be found from this Github repository: https://github.com/SpringNuance/Human-In-The-Loop-De-Novo-Molecular-Design

In order to run the project code at Base-Code-Binh, please install the environment of ReinventCommunity located at Reinvent-Community-original/environment.yml. For the author, using any other yaml environment is likely to result in compatibility issues. Additionally, this project is conducted for REINVENT version 3.2. Latest REINVENT version 4.0 has been released, which may not be compatible with the author's source code.

References

[1] S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang, and S. H. Bryant, "Pubchem substance and compound databases," *Nucleic Acids Research*, vol. 44, no. D1, pp. D1202–D1213, 2016.

[2] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open babel: An open chemical toolbox," *Journal of Cheminformatics*, vol. 3, p. 37, 2011.

[3] D. Weininger, "Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules," *Journal of Chemical Information and Computer Sciences*, vol. 28, no. 1, pp. 31–36, 1988.

[4] R. Community, "Rdkit: Open-source cheminformatics." GitHub, 2024.

[5] H. Yang, "Morgan fingerprint generator in rdkit." Herong's Tutorial Examples, 2021.

[6] J. Meyers, B. Fabian, and N. Brown, "De novo molecular design and generative models," *Drug Discovery Today*, vol. 26, no. 11, pp. 2707–2715, 2021.

[7] A. Tharwat and W. Schenck, "A survey on active learning: State-of-the-art, practical challenges and research directions," *Mathematics*, vol. 11, no. 4, 2023.

[8] T. Kaufmann, P. Weng, V. Bengs, and E. Hüllermeier, "A survey of reinforcement learning from human feedback," 2023.

[9] S. Takeda, H. Kaneko, and K. Funatsu, "Chemical-space-based de novo design method to generate drug-like molecules," *Journal of Chemical Information and Modeling*, vol. 56, no. 9, pp. 1667–1681, 2016.

[10] S. K. Dhanda, D. Singla, A. Mondal, and G. Raghava, "Drugmint: a webserver for predicting and designing of drug-like molecules," *Biology Direct*, vol. 8, p. 28, 2013.

[11] Y. Liang, R. Zhang, L. Zhang, and P. Xie, "Drugchat: Towards enabling chatgpt-like capabilities on drug molecule graphs," *arXiv preprint arXiv:2309.03907*, 2023.

[12] T. Blaschke, J. Arús-Pous, H. Chen, C. Margreitter, C. Tyrchan, O. Engkvist, K. Papadopoulos, and A. Patronov, "Reinvent 2.0: An ai tool for de novo drug design," *Journal of Chemical Information and Modeling*, vol. 60, no. 9, pp. 4423–4433, 2020.

[13] H. H. Loeffler, J. He, A. Tibo, J. Janet, A. Voronov, L. H. Mervin, and O. Engkvist, "Reinvent 4: Modern ai-driven generative molecule design," *Journal of Cheminformatics*, vol. 16, no. 1, p. 812, 2024.

[14] T. Blaschke, J. Arús-Pous, H. Chen, C. Margreitter, C. Tyrchan, O. Engkvist, K. Papadopoulos, and A. Patronov, "Supporting information reinvent 2.0: An ai tool for de novo drug design," *ChemRxiv*, 2020.

[15] W. Zhang, M. Lei, Q. Wen, D. Zhang, G. Qin, J. Zhou, and L. Chen, "Dopamine receptor d2 regulates glua1-containing ampa receptor trafficking and central sensitization through the pi3k signaling pathway in a male rat model of chronic migraine," *Journal of Headache and Pain*, vol. 23, no. 1, p. 45, 2022.

[16] S. Pierce, Z. Fang, Y. Yin, L. West, M. Asher, T. Hao, X. Zhang, K. Tucker, A. Staley, Y. Fan, W. Sun, D. Moore, C. Xu, Y.-H. Tsai, J. Parker, V. Prabhu, J. Allen, D. P. Lee, C. Zhou, and V. Bae-Jump, "Targeting dopamine receptor d2 as a novel therapeutic strategy in endometrial cancer," *Journal of Experimental Clinical Cancer Research*, vol. 39, no. 1, p. 38, 2020.

[17] H. Lee, S. Shim, J. Kong, M.-J. Kim, S. Park, S.-S. Lee, and A. Kim, "Overexpression of dopamine receptor d2 promotes colorectal cancer progression by activating the ŏ3b2-catenin/zeb1 axis," *Cancer Science*, vol. 112, no. 4, pp. 1503–1515, 2021.

[18] M. Wang, Z. Wu, J. Wang, G. Weng, Y. Kang, P. Pan, D. Li, Y. Deng, X. Yao, Z. Bing, C.-Y. Hsieh, and T. Hou, "Genetic algorithm-based receptor ligand: A genetic algorithm-guided generative model to boost the novelty and drug-likeness of molecules in a sampling chemical space," *Journal of Chemical Information and Modeling*, vol. 64, no. 2, pp. 326–345, 2024.

[19] J. Fotie, C. M. Matherne, and J. E. Wroblewski, "Silicon switch: Carbon-silicon bioisosteric replacement as a strategy to modulate the selectivity, physicochemical, and drug-like properties in anticancer pharmacophores," *Chemical Biology and Drug Design*, vol. 91, no. 4, pp. 239–249, 2023.

[20] O. Choung, R. Vianello, M. Segler, *et al.*, "Extracting medicinal chemistry intuition via preference machine learning," *Nature Communications*, vol. 14, no. 1, p. 6651, 2023.

[21] T. Contributors, "Tdc: Oracle for drd2 affinity prediction." Therapeutics Data Commons, 2024.

[22] I. Sundin, A. Voronov, H. Xiao, K. Papadopoulos, E. Bjerrum, M. Heinonen, A. Patronov, S. Kaski, and O. Engkvist, "Human-in-the-loop assisted de novo molecular design," *Journal of Cheminformatics*, vol. 14, 12 2022.

[23] I. Sundin, A. Voronov, H. Xiao, K. Papadopoulos, E. J. Bjerrum, M. Heinonen, A. Patronov, S. Kaski, and O. Engkvist, "Human-in-the-loop assisted de novo molecular design," *ChemRxiv*, 2022.

[24] Unknown, "Lecture 24: Bradley-terry model," 2017. Accessed: 2024-07-16.

[25] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li, "Learning to rank: from pairwise approach to listwise approach," in *Proceedings of the 24th international conference on Machine learning*, pp. 129–136, ACM, 2007.

[26] E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, and Fernández-Leal, "Human-in-the-loop machine learning: a state of the art," *Artificial Intelligence Review*, vol. 56, 08 2022.

[27] D. Polykovskiy, A. Zhebrak, B. Sanchez-Lengeling, S. Golovanov, O. Tatanov, S. Belyaev, R. Kurbanov, A. Artamonov, V. Aladinskiy, M. Veselov, A. Kadurin, S. Johansson, H. Chen, S. Nikolenko, A. Aspuru-Guzik, and A. Zhavoronkov, "Molecular sets (moses): A benchmarking platform for molecular generation models," *arXiv preprint arXiv:1811.12823*, 2018.

[28] M. Benhenda, "Chemgan challenge for drug discovery: can ai reproduce natural chemical diversity?," *arXiv preprint arXiv:1708.08227*, 2017.

[29] P. Ertl and A. Schuffenhauer, “Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions,” *Journal of Cheminformatics*, vol. 1, p. 8, 2009.

[30] G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan, and A. L. Hopkins, “Quantifying the chemical beauty of drugs,” *Nature Chemistry*, vol. 4, pp. 90–98, 2012.

[31] M. Rasool and S. Läär, “Development and evaluation of a physiologically based pharmacokinetic model to predict carvedilol-paroxetine metabolic drug-drug interaction in healthy adults and its extrapolation to virtual chronic heart failure patients for dose optimization,” *Expert Opinion on Drug Metabolism Toxicology*, vol. 17, no. 6, pp. 695–709, 2021.

[32] K. Lin, J. Tibbitts, and B.-Q. Shen, *Pharmacokinetics and ADME characterizations of antibody-drug conjugates*, vol. 1045 of *Methods in Molecular Biology*, pp. 117–131. Humana Press, 2013.

[33] A. I. Omar and K. Na-Bangchang, “Pharmacokinetic studies of nanoparticles as a delivery system for conventional drugs and herb-derived compounds for cancer therapy: a systematic review,” *International Journal of Nanomedicine*, vol. 14, pp. 9159–9173, 2019.

[34] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, “Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings,” *Advanced Drug Delivery Reviews*, vol. 23, no. 1-3, pp. 3–25, 1997.

[35] D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, and K. D. Kopple, “Molecular properties that influence the oral bioavailability of drug candidates,” *Journal of Medicinal Chemistry*, vol. 45, no. 12, pp. 2615–2623, 2002.

[36] I. Muegge, S. L. Heald, and D. Brittelli, “Simple selection criteria for drug-like chemical matter,” *Journal of Medicinal Chemistry*, vol. 44, no. 12, pp. 1841–1846, 2001.