Human-in-the-loop (HITL) in reinforcement learning for de novo molecular design

Research project in MACADAMIA (CS-E4875)

Nguyen Xuan Binh - 887799
Advisors: Yasmine Nahal, Prof. Samuel Kaski
Aalto University – 09/08/2024

Table of Contents

1. State of the art
HITL ML for de novo drug design

De-Novo Molecular Design

- The term "de novo" is Latin for "from the beginning"
- It is the process of creating new molecular structures from scratch using computational approaches that also satisfy a desired molecular profile (Meyers 2021)
- The process typically involves the use of algorithms to explore chemical space, generating and optimizing new molecules based on predefined criteria such as biological activity, drug-likeness, and synthesizability.
- Researchers can find molecules with the best therapeutic qualities by searching the huge chemical space via atom-based, fragment-based or reaction-based paradigms (Meyers 2021)

De-Novo Molecular design
Source: DALL-E

Human-in-the-loop De-Novo molecular design

The 'human-in-the-loop' framework allows humans to use their domain expertise into modeling process, connecting computer and cognitive science (Tharwat et al. 2023)

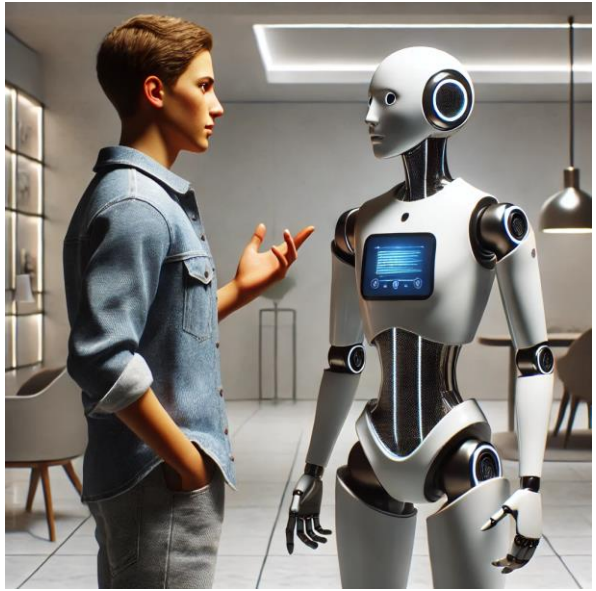
Reinforcement learning from human feedback (RLHF) introduces a critical human-in-the-loop component to let humans define the objective (Kaufmann et al. 2023)

Some notable works of HITL de-novo molecular design for motivation

Winter et al. (2020) presents *grünifai*, an interactive in silico platform for optimizing small molecules in drug discovery, integrating adjustable models, a continuous chemical space representation, and a scalable optimization algorithm with user feedback to balance multiple molecular properties.

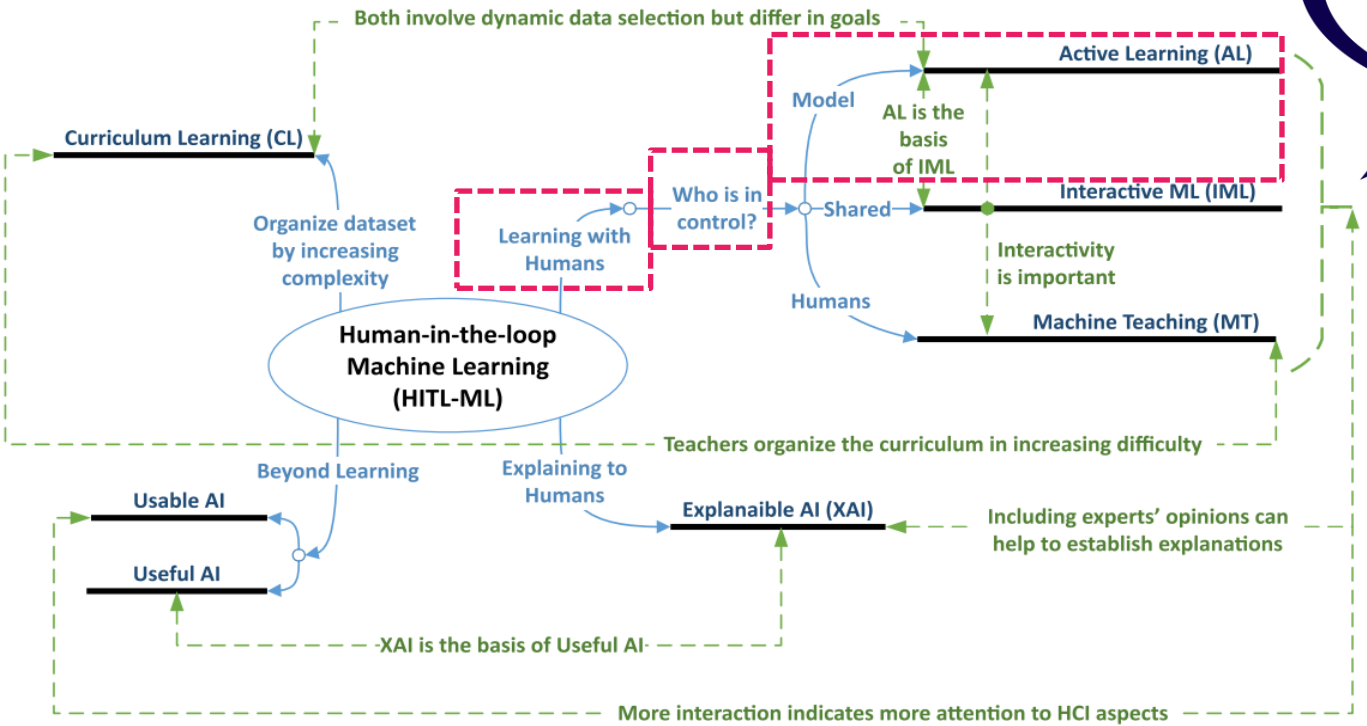
Sundin et al. (2022) presents a HITL ML approach for de novo molecular design, using a probabilistic model and active learning to integrate user feedback into the multi-parameter optimization (MPO) scoring function.

Choung et al. (2023) applies AI learning-to-rank techniques to feedback from human chemists to replicate the lead optimization process in drug discovery.



An expert human provides his expertise domain to the AI agent
Source: DALL-E

Where is this research project's domain on this chart?



The ML model is in control, actively querying new molecules for labelling by humans

The scoring ML model directly learns DRD2 affinity and provides feedback to generative chemistry models

Human-in-the-loop machine learning (HITL-ML-relations) mind map (Mosqueira-Rey et al. 2022)

2. Research project motivation and objectives

Research project motivation

🏆 **PROJECT MOTIVATION**: we aim to demonstrate that we can capture chemist intuition via preference ML and enable state-of-the-art HITL de novo molecular design to support a broader range of feedback mechanisms (binary, pairwise comparison and ranking molecules)

- General objective: use generative model to generate molecules that bind to the dopamine receptor D2 (DRD2). Optimizing DRD2 binding could help inhibit diseases such as schizophrenia and Parkinson's disease => Targeting DRD2 affinity is important
- DRD2 is used as a toy use case in this project to demonstrate the effectiveness of our user feedback models. This allows us to evaluate our results using the existing Oracle model on DRD2 affinity classification

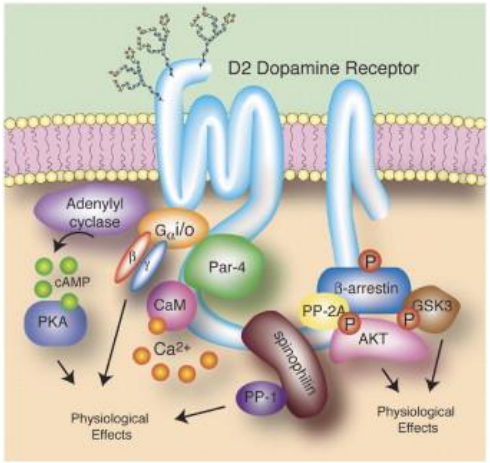


Image Source: Decoding Dopamine Signaling

Research project objectives

Step 1: De Novo molcular design

Generate novel molecules for DRD2 receptor using REINVENT to generate molecules that not only have high DRD2 affinity but are also novel, diverse and have drug-like properties

Step 2: Human-in-the-loop workflow

Establish a HITL active learning workflow to learn user models from human feedback collected via three different user feedback mechanisms and three different acquisition strategies

Step 3: Build ML model for learning DRD2 affinity

Develop a surrogate ML model that predicts DRD2 affinity for novel molecules. We also benchmark learning performance of the different user models against the Oracle

3. REINVENT: generative RL model

REINVENT4 : a brief look

- REINVENT is an advanced AI tool designed for de novo drug design, providing a comprehensive platform for generating novel drug-like molecules (Blaschke 2020)

- Primary motivation: efficiently explore vast chemical spaces, identifying molecules with desired biological activities and drug-like properties.

- Users define the initial set of constraints and desired properties to optimize. This may include diversity filter, inception, chemical properties, configurations and most importantly, the scoring function. All of these settings are written to a json file, which would be read by REINVENT input.py

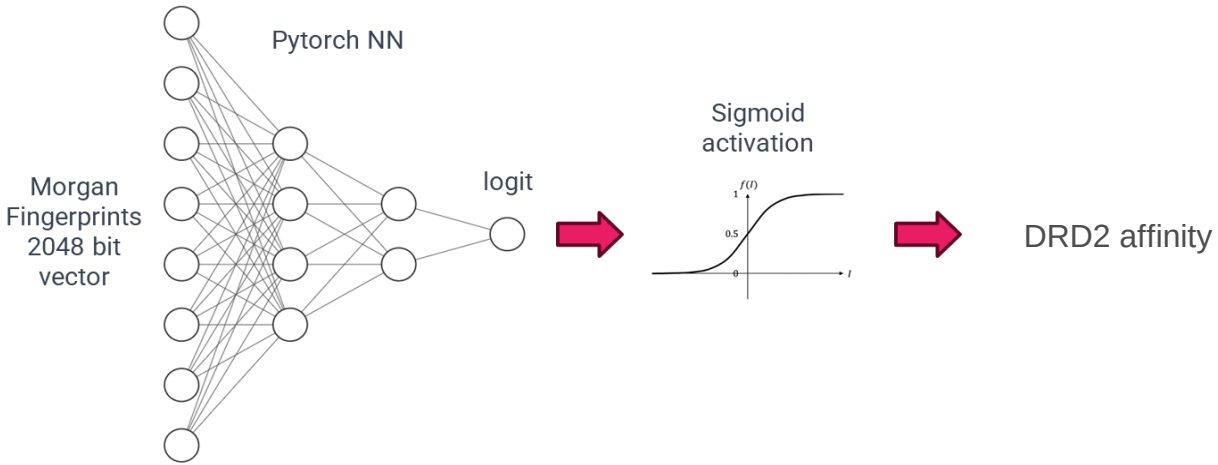
- Generated molecules are scored based on the scoring function (or the feedback function). REINVENT tries to generate molecules that maximize the scores

- The tool refines the molecules through multiple steps, and finally return SMILES in the scaffold_memory.csv file, where the number of SMILES in this file is roughly equal to number of optimization steps x batch size

4. The three feedback mechanisms

Model architecture

- This project develops three feedback models for REINVENT during the scoring stage: the scoring (baseline) model, the comparing (Bradley-Terry) model and the ranking (ListNet) model
- While they differ in how they return the feedback to REINVENT, they have essentially the same neural network architecture, which is to learn the DRD2 affinity.
- The figure below shows the simple architecture for learning DRD2 affinity, where the Sigmoid activation ensures that the logit is restricted to range [0, 1]



Scoring (baseline) 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 supported by REINVENT scoring methods.

Model input: Vector of Morgan fingerprints of one single SMILES.

Model output: Model output is the direct quantity returned by the neural network. The model output of scoring model is the probability that the SMILES has DRD2 activity (score between 0 and 1)

Individual feedback: Individual feedback is the feedback that the user feedback model offers to de novo generative software (REINVENT) for scoring a single SMILES. Individual feedback is the same as model output for the scoring model

Batch feedback: Batch feedback is defined as the array of feedback for a batch of SMILES during REINVENT scoring and data acquisition. For scoring model, the batch feedback is simply the model outputs for all SMILES.

Loss function to train Pytorch model: Binary Cross Entropy Loss (BCELoss), defined as

$$BCE(p,q) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i\ log\left(p(y_i)\right) + (1-y_i)\ log\left(1-p(y_i)\right)\right]$$

where 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 comparing (Bradley-Terry) model

Motivation for comparing pairs of molecules: human chemists may have difficulty in providing a DRD2 affinity, but 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 define $\beta_i \in R$ as the DRD2 affinity for SMILES 1, and $\beta_i \in R$ the DRD2 affinity of SMILES 2 given by the neural network classifier, 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 probability p_{ij} that SMILES i is better than SMILES j, which is defined as

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

Model input: Morgan fingerprints of two different SMILES

Model output: probability that the first SMILES is better than the second SMILES in terms of DRD2 affinity

Individual feedback: Feedback is 1 if SMILES 1 better than SMILES 2 (Model output > 0.5), else feedback is 0.

Batch feedback: Step 1: Obtaining P_2^N permutations of pairs of SMILES from REINVENT output (N is batch size)

Step 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.

Step 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. Label 1 if SMILES 1 better than SMILES 2 and 0 otherwise.

Ranking (ListNet) model

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 like the pairwise comparing model (Cao et al .2007).

Motivation: 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.

Model input: Morgan fingerprints of three different SMILES.

Model output: softmax scores of 3 SMILES from original scores of DRD2 affinity

Individual feedback: SMILES with lowest score receive rank 0, second highest being rank 1 and highest one being rank 2. Then the ranks are normalized to [0, 0.5, 1]

Batch feedback: Step 1: Obtaining C_3^N combinations of sets of 3 SMILES from REINVENT output.

Step 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.

Step 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

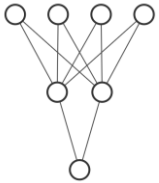
Loss function to train Pytorch model: **Kullback-Leibler divergence loss (KLDivLoss)**, which is defined as $\mathrm{KL}(P \parallel Q) = \sum_i P(i) \log \left(\frac{P(i)}{Q(i)} \right)$

The KL divergence measures how the predicted probability distribution P diverges from the true probability distribution Q. In our case, P is the softmax scores from the neural networks and Q is the true softmax scores of DRD2 affinity.

Scoring Model

This model directly returns DRD2 affinity as feedback

SMILES 1

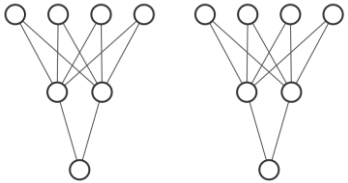


DRD2 score of SMILES 1

Pairwise Comparing Model

This model uses the Bradley-Terry formula

SMILES 1 SMILES 2



DRD2 score of SMILES 1 DRD2 score of SMILES 2

Bradley-Terry formula

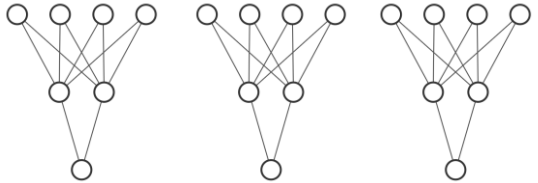
$$Sigmoid(DRD2\ \ SMILES\ 1 - DRD2\ \ SMILES\ 2)$$

Probability of SMILES 1 better than SMILES 2

Ranking Model

This model uses the ListNet architecture

SMILES 1 SMILES 2 SMILES 3



DRD2 score of SMILES 1 DRD2 score of SMILES 2 DRD2 score of SMILES 3

SOFTMAX FUNCTION

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_k e^{z_k}}$$

Vector of 3 preference scores

5. HITL workflow for De-Novo Drug Design

Human-in-the-loop workflow

Algorithm 1 Human-in-the-loop workflow

Require: An Oracle that reliably estimates DRD2 probability, a weak 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
  $U_r \leftarrow \text{REINVENT}(S_{\theta_{r,1}})$ ▷ 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 interations 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 probability 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

- Workflow inspired by Sundin et al. [2023]
- Evaluating score is time-consuming
  => use ML model to act as a surrogate human.
- Initially, we start from a low accuracy surrogate ML model because it had not observed enough human preference data => ML models should classify DRD2 affinity with low accuracy (around 0.5) at the beginning.
- Settings used in this project are
  - R = 3 REINVENT rounds
  - T = 5 HITL interactions
  - Q = 56 queries
  - REINVENT's batch size = 64
  - REINVENT's optimization steps = 100
  - Initial training dataset size is 100 SMILES with DRD2 and 100 without DRD2 affinity
- After 3 REINVENT rounds of 5 iterations, the final training dataset size becomes
  200 + 3 x 5 x 56 = 1040 SMILES
  for the last surrogate ML model.

Active learning (AL)

Active learning is semi-supervised learning as it uses both labeled and unlabeled data for training ML model (Tharwat et al. 2023)

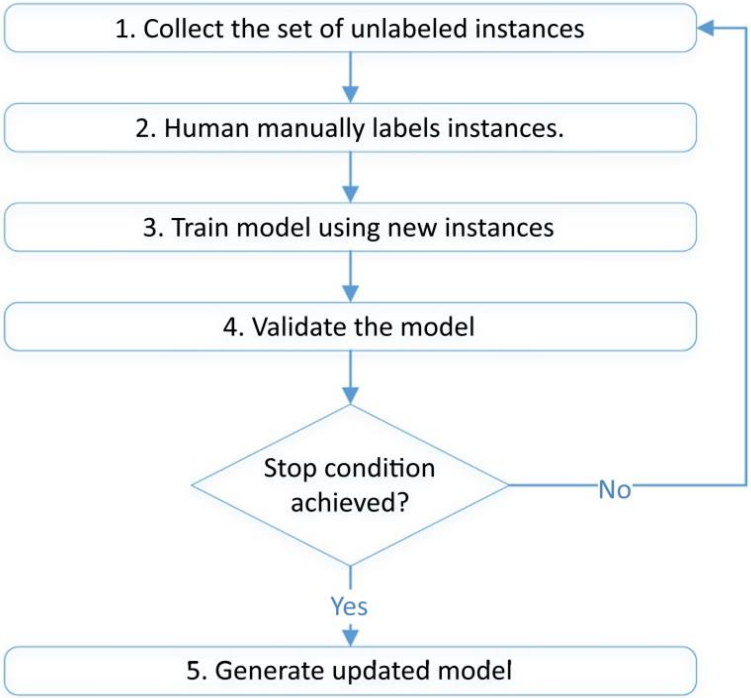
New samples get annotated in an iterative process, where an acquisition function choose an unlabelled data, and once labeled by an Oracle, will result in a model accuracy increment

In this project, random, uncertainty and greedy acquisition functions are used to choose most promising generated SMILES

1. Random acquisition selects randomly from the pool of unselected SMILES. This helps introduce diversity in new training molecules.

2. Uncertainty acquisition selects molecules for which the ML model has the least confidence in its prediction

3. Greedy acquisition selects molecules that have the highest predicted DRD2 probabilities based on feedback from ML models



Molecular descriptor filtering for REINVENT's generated SMILES

- Properties that affect pharmacokinetic (PK) parameters are the molecular descriptors, which are used to characterize the physical, chemical, and structural properties of molecules.
- Molecules that bypass these molecular descriptor filters' threshold are considered to be good candidates as bioactive, efficient drugs

	Description	Lower threshold	Higher threshold	Known rules
DRD2 affinity	Objective to maximize	0.75	1.0	0.75 is average score of SMILES actually having DRD2 affinity predicted by Oracle
logP	Filtering	1.0	5.0	Lipinski's rule of five
Molecule weight	Filtering	0	500.0	Lipinski's rule of five
Hydrogen bond donors number	Filtering	0	5	Lipinski's rule of five
Hydrogen bond acceptors number	Filtering	0	10	Lipinski's rule of five
TPSA	Filtering	0.0	140.0	Veber et al.
Number of rotatable bonds	Filtering	0	10	Veber et al.
Number of rings	Filtering	0	7	Muegge et al.

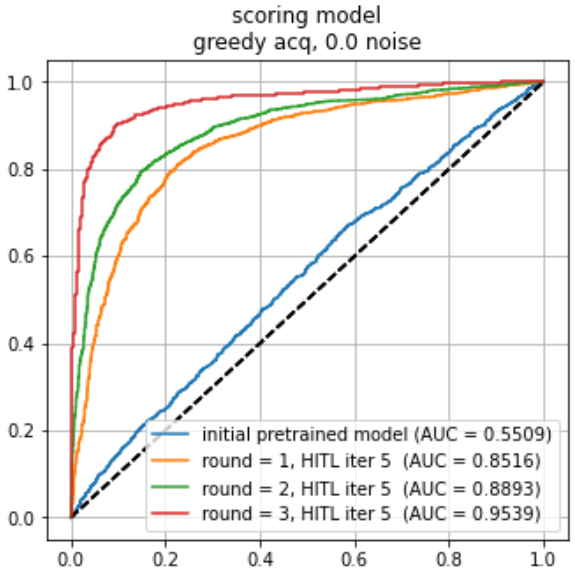
Other metrics for REINVENT generated SMILES besides DRD2 affinity

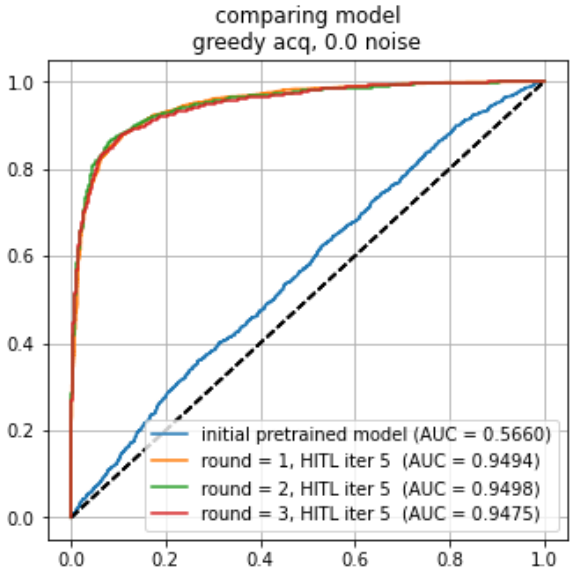
- As per objectives of de novo molecular design, the novelty score, diversity score, synthetic accessibility (SA) score, and quantitative estimate of drug-likeness (QED) score are important metrics to assess the quality of generated molecules.

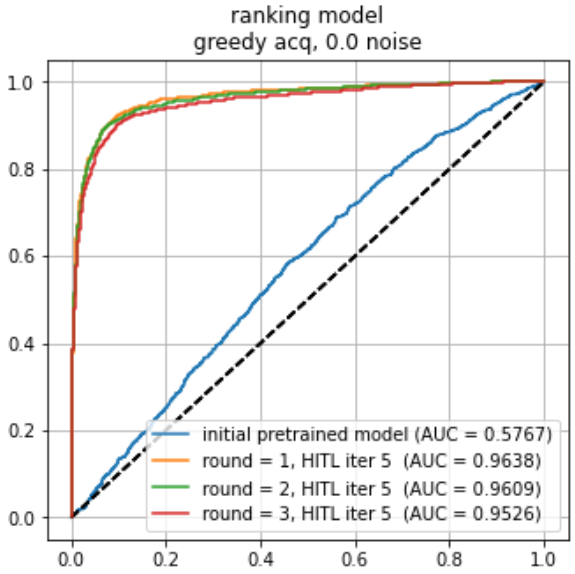
	Definition	Good lower threshold	Good upper threshold	Known rules
Novelty score [1]	Fraction of the generated molecules not present in the training set. Range [0, 1]. The higher the better	Possibly more than 0.95	1.00	Polykovskiy et al.
Diversity score [2]	Internal chemical diversity within the generated molecules set. Range [0, 1]. The higher the better.	Possibly more than 0.7	1.00	Benhenda et al.
Synthetic assessibility (SA) score	How easily a molecule can be synthesized. Range [1, 10]. The lower, the better	1.0	3.0	Ertl & Schuffenhauer
Quantitative estimate of drug-likeness (QED) score	Composite metric that evaluates the drug-likeness of a compound based on several molecular properties. Range [0, 1]. The higher the better	0.5	1.0	Bickerton et al.

6. Results

ROC curve evolution after 3 REINVENT rounds

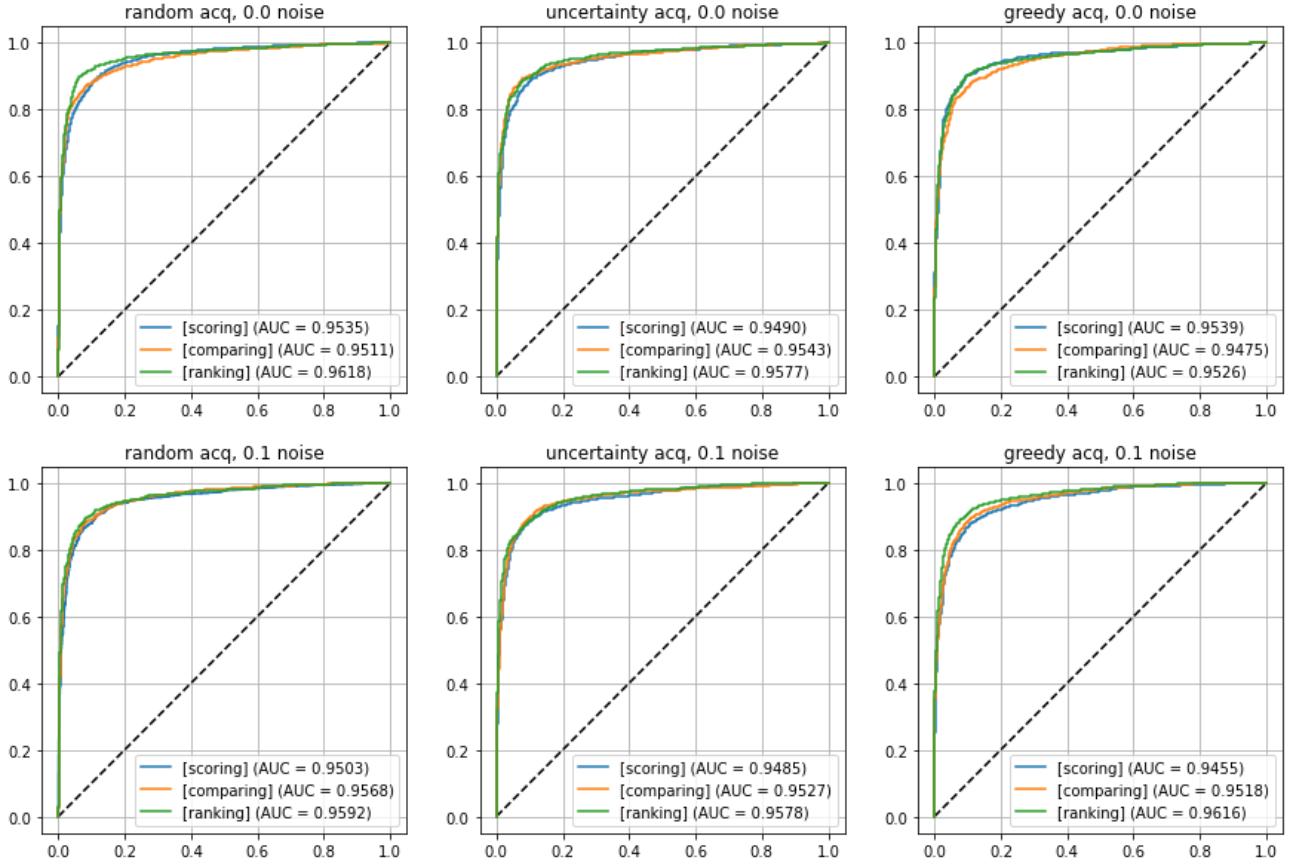






Initially, the pre-trained models show relatively low ROC AUC values as very few chemist knowledge about DRD2 affinity is available. After each round of HITL iterations, we observe a considerable improvement in ROC AUC values for all models

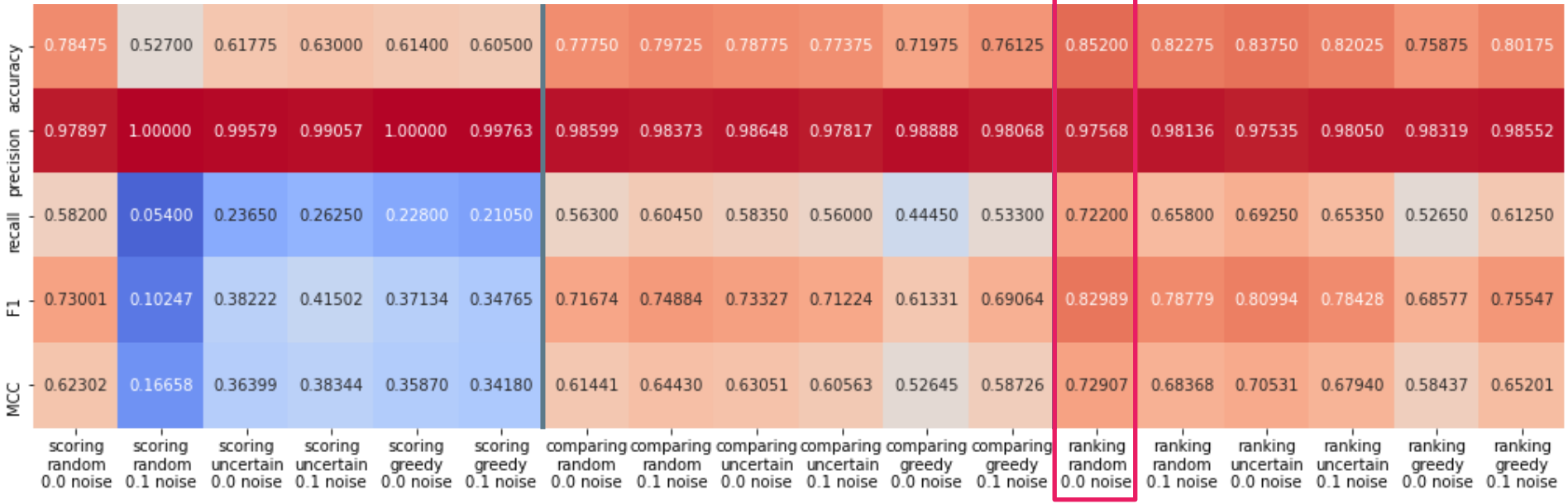
ROC curve comparison between 6 running cases



- There are 6 running cases for all feedback models: 3 acquisition strategies x 2 levels of noise labelling (0.0 and 0.1 level) = 6 running cases
- When molecules are presented to the Oracle for labelling DRD2 affinity, 0.0 noise means Oracle directly returns its prediction, and 0.1 noise means its prediction plus a noise of Normal(0, 0.1).
- Across all running cases, the AUC values indicate that all models perform well in learning the human chemist knowledge or preferences for DRD2 actives.
- The ranking model consistently achieves highest AUC values, followed closely by the comparing model and the scoring model.

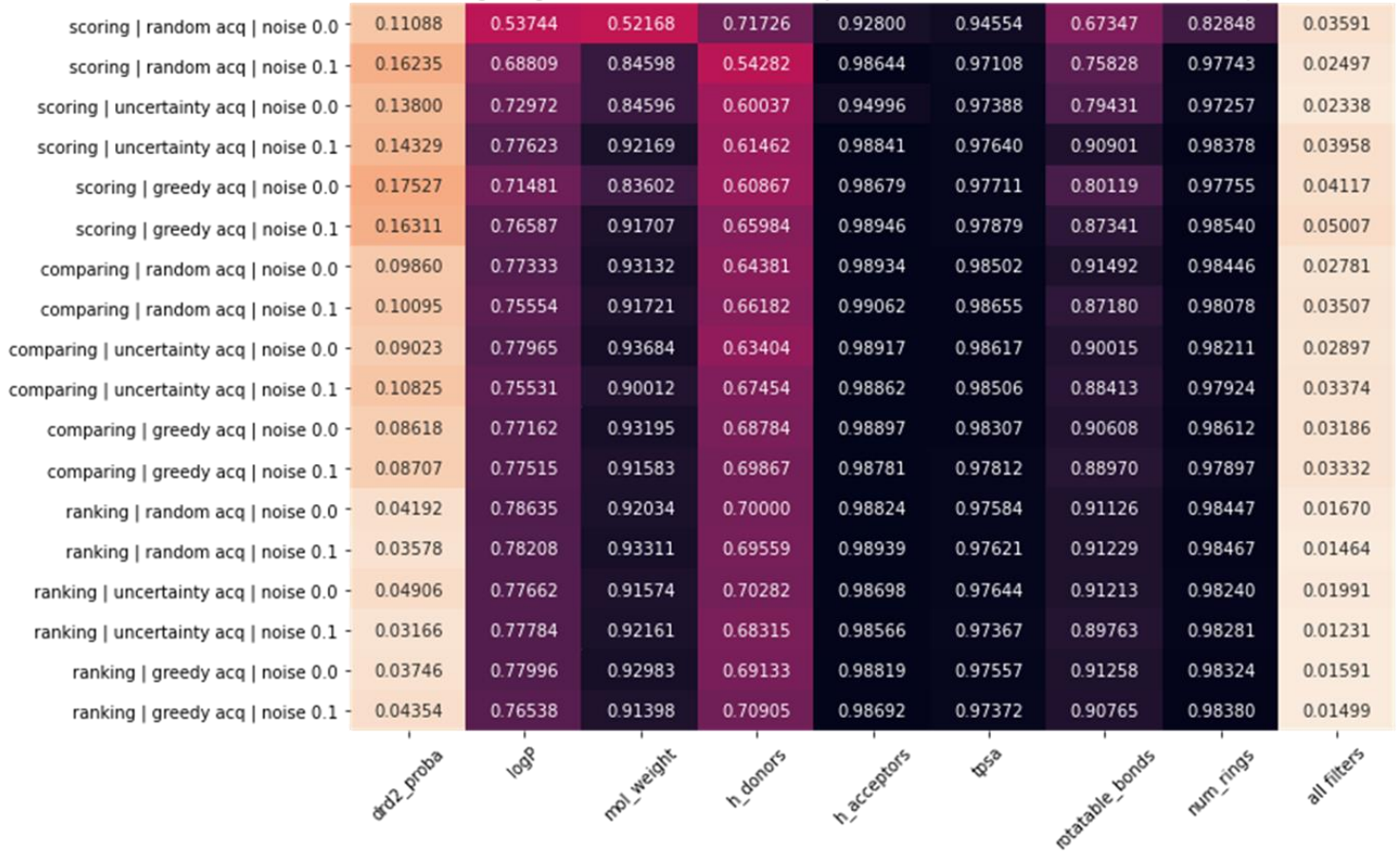
Classification metrics of user feedback models

Best configuration on classifying DRD2 affinity



- ❖ Testing dataset: 1000 SMILES with and without DRD2 affinity for each class.
- ❖ Most models have high precision, but unhelpful because the positive case (DRD2) is much rarer during the training process => We should focus more on recall and Matthews correlation coefficient (MCC) score
- ❖ The ranking model consistently has higher recall than comparing model, which in turn has higher recall than scoring model.
- ❖ Comparing and ranking molecules deliver better classification results than the scoring model.
- ❖ Feedback collected through random sampling help improve the MCC score

Percentage of generated SMILES that satisfy each filter at all REINVENT rounds

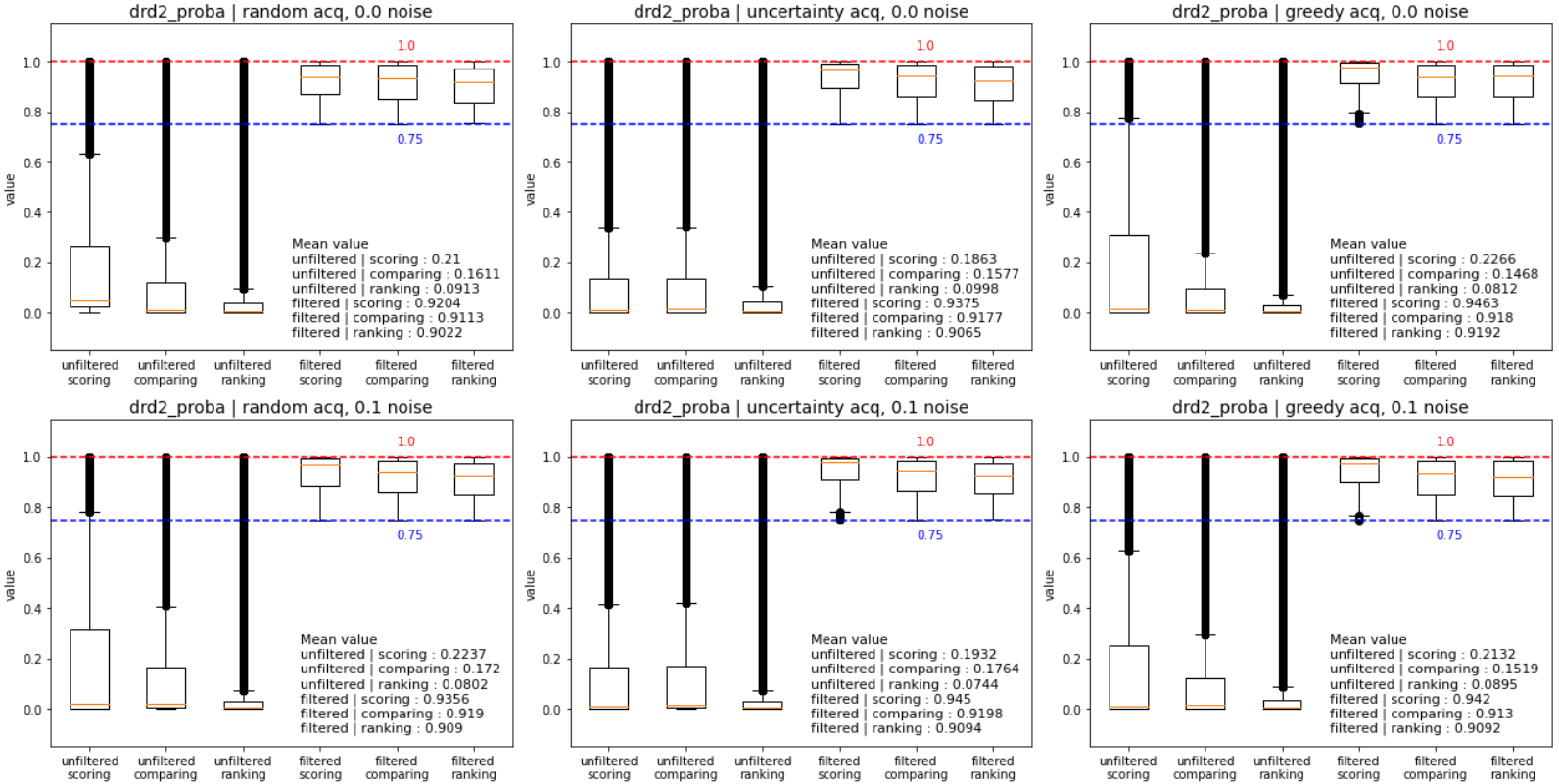


	drd2_proba	logP	mol_weight	h_donors	h_acceptors	tpsa	rotatable_bonds	num_rings	all filters
scoring | random acq | noise 0.0	0.11088	0.53744	0.52168	0.71726	0.92800	0.94554	0.67347	0.82848	0.03591
scoring | random acq | noise 0.1	0.16235	0.68809	0.84598	0.54282	0.98644	0.97108	0.75828	0.97743	0.02497
scoring | uncertainty acq | noise 0.0	0.13800	0.72972	0.84596	0.60037	0.94996	0.97388	0.79431	0.97257	0.02338
scoring | uncertainty acq | noise 0.1	0.14329	0.77623	0.92169	0.61462	0.98841	0.97640	0.90901	0.98378	0.03958
scoring | greedy acq | noise 0.0	0.17527	0.71481	0.83602	0.60867	0.98679	0.97711	0.80119	0.97755	0.04117
scoring | greedy acq | noise 0.1	0.16311	0.76587	0.91707	0.65984	0.98946	0.97879	0.87341	0.98540	0.05007
comparing | random acq | noise 0.0	0.09860	0.77333	0.93132	0.64381	0.98934	0.98502	0.91492	0.98446	0.02781
comparing | random acq | noise 0.1	0.10095	0.75554	0.91721	0.66182	0.99062	0.98655	0.87180	0.98078	0.03507
comparing | uncertainty acq | noise 0.0	0.09023	0.77965	0.93684	0.63404	0.98917	0.98617	0.90015	0.98211	0.02897
comparing | uncertainty acq | noise 0.1	0.10825	0.75531	0.90012	0.67454	0.98862	0.98506	0.88413	0.97924	0.03374
comparing | greedy acq | noise 0.0	0.08618	0.77162	0.93195	0.68784	0.98897	0.98307	0.90608	0.98612	0.03186
comparing | greedy acq | noise 0.1	0.08707	0.77515	0.91583	0.69867	0.98781	0.97812	0.88970	0.97897	0.03332
ranking | random acq | noise 0.0	0.04192	0.78635	0.92034	0.70000	0.98824	0.97584	0.91126	0.98447	0.01670
ranking | random acq | noise 0.1	0.03578	0.78208	0.93311	0.69559	0.98939	0.97621	0.91229	0.98467	0.01464
ranking | uncertainty acq | noise 0.0	0.04906	0.77662	0.91574	0.70282	0.98698	0.97644	0.91213	0.98240	0.01991
ranking | uncertainty acq | noise 0.1	0.03166	0.77784	0.92161	0.68315	0.98566	0.97367	0.89763	0.98281	0.01231
ranking | greedy acq | noise 0.0	0.03746	0.77996	0.92983	0.69133	0.98819	0.97557	0.91258	0.98324	0.01591
ranking | greedy acq | noise 0.1	0.04354	0.76538	0.91398	0.70905	0.98692	0.97372	0.90765	0.98380	0.01499

- Most molecules generated by REINVENT already satisfy the filters except LogP and the number of hydrogen bond donors
- DRD2 affinity of generated molecules for all models is obtained by the Oracle
- The scoring model is most capable of generating molecules with DRD2 affinity, followed by the comparing model and lastly by the ranking model => major advantage of the scoring model

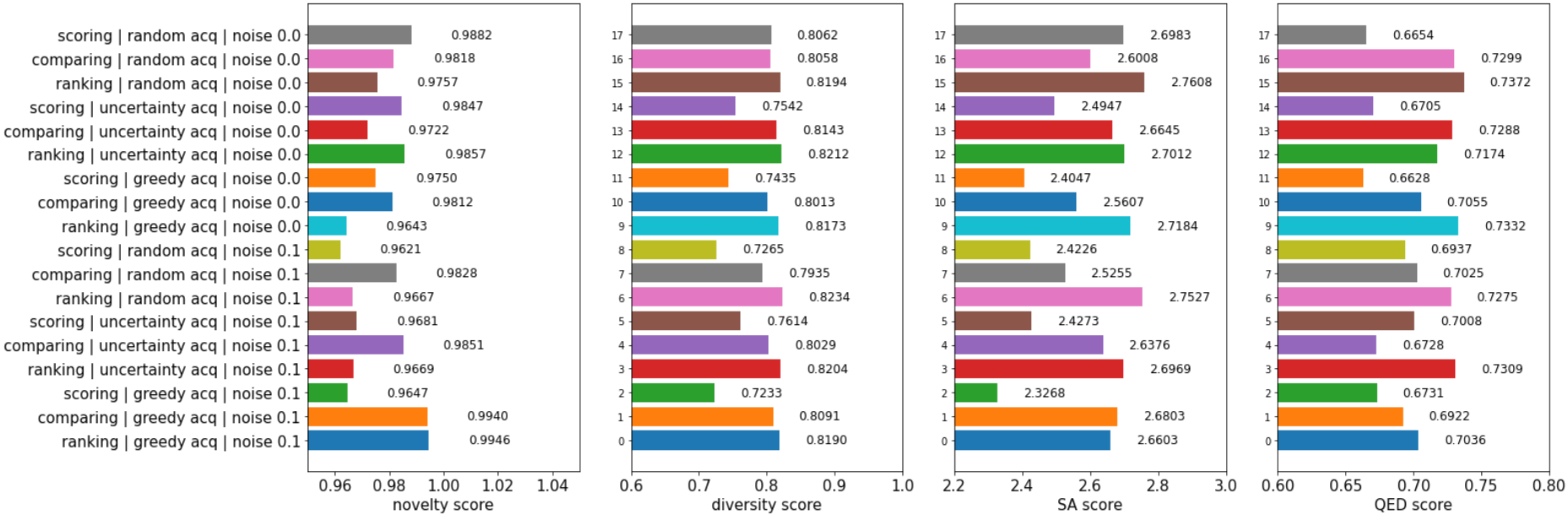
DRD2 affinity of REINVENT's generated molecules for all rounds

For generating molecules with highest DRD2 affinity, the scoring model is the best, followed by the comparing model and lastly the ranking model. This is true for both filtered and unfiltered generated SMILES



Metric scores comparison between all running cases for filtered SMILES

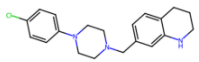
- Novelty score: it is often nearly equal for 3 models, probably comparing model frequently has highest novelty score.
- Diversity score: ranking model consistently outperforms comparing model, which in turns is better than the scoring model.
- SA score seems inversely proportional to the diversity score => scoring model has best SA score, followed by comparing model and lastly ranking model.
- QED score: ranking model consistently outperforms comparing model, which in turns better than the scoring model.



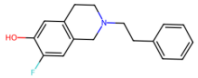
Example of best molecules for each running case

- ❖ molecules produced by the scoring model have simple structures with few branches and fewer functional groups
- ❖ comparing and ranking model shows greater structural complexity and diversity with different ring systems and functional groups. They maintain a balance between drug-likeness, diversity, and novelty

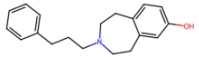
Clc1ccc(N2CCN(Cc3ccc4c(c3)NCCC4)CC2)cc1
scoring | random | noise 0.0
DRD2 proba: 0.9948 | SA: 2.0788 | QED: 0.9092



Oc1cc2c(cc1F)CN(CCc1ccccc1)CC2
comparing | random | noise 0.0
DRD2 proba: 0.9618 | SA: 2.0272 | QED: 0.9268

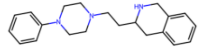


Oc1ccc2c(c1)CCN(CCCc1ccccc1)CC2
ranking | random | noise 0.0
DRD2 proba: 0.9899 | SA: 1.831 | QED: 0.9279

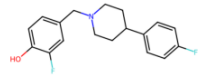


Oc1ccc(N2CCN(CCc3ccc4c(c3)CCNC4)CC2)cc1
scoring | random | noise 0.1
DRD2 proba: 0.9937 | SA: 2.2016 | QED: 0.8986

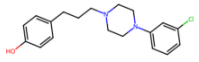
c1ccc(N2CCN(CCC3Cc4ccccc4CN3)CC2)cc1
comparing | random | noise 0.1
DRD2 proba: 1.0 | SA: 2.5813 | QED: 0.934



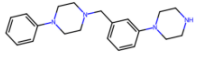
Oc1ccc(CN2CCC(c3ccc(F)cc3)CC2)cc1F
ranking | random | noise 0.1
DRD2 proba: 0.9915 | SA: 1.9134 | QED: 0.9244



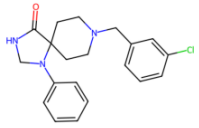
Oc1ccc(CCCN2CCN(c3cccc(Cl)c3)CC2)cc1
scoring | uncertainty | noise 0.0
DRD2 proba: 0.9901 | SA: 1.8178 | QED: 0.9029



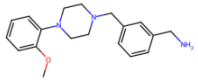
c1ccc(N2CCN(Cc3cccc(N4CCNCC4)c3)CC2)cc1
comparing | uncertainty | noise 0.0
DRD2 proba: 0.9975 | SA: 1.9259 | QED: 0.9255



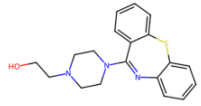
O=C1NCN(c2ccccc2)C12CCN(Cc1cccc(Cl)c1)CC2
ranking | uncertainty | noise 0.0
DRD2 proba: 0.9795 | SA: 2.6948 | QED: 0.9175



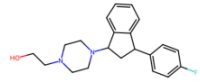
COc1ccccc1N1CCN(Cc2cccc(CN)c2)CC1
scoring | uncertainty | noise 0.1
DRD2 proba: 0.9584 | SA: 1.7881 | QED: 0.921



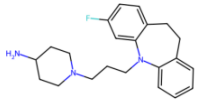
OCCN1CCN(C2=Nc3ccccc3Sc3ccccc32)CC1
comparing | uncertainty | noise 0.1
DRD2 proba: 0.9956 | SA: 2.3382 | QED: 0.9132



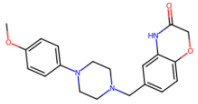
OCCN1CCN(C2CC(c3ccc(F)cc3)c3ccccc32)CC1
ranking | uncertainty | noise 0.1
DRD2 proba: 1.0 | SA: 2.9236 | QED: 0.9268



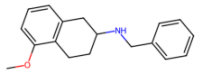
NC1CCN(CCCN2c3ccccc3CCc3ccc(F)cc32)CC1
scoring | greedy | noise 0.0
DRD2 proba: 0.9919 | SA: 2.2738 | QED: 0.9054



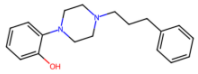
COc1ccc(N2CCN(Cc3ccc4c(c3)NC(=O)CO4)CC2)cc1
comparing | greedy | noise 0.0
DRD2 proba: 1.0 | SA: 2.0771 | QED: 0.9144



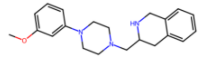
COc1cccc2c1CCC(NCc1ccccc1)C2
ranking | greedy | noise 0.0
DRD2 proba: 1.0 | SA: 2.3095 | QED: 0.917



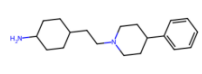
Oc1ccccc1N1CCN(CCCc2ccccc2)CC1
scoring | greedy | noise 0.1
DRD2 proba: 0.9967 | SA: 1.722 | QED: 0.918



COc1cccc(N2CCN(CC3Cc4ccccc4CN3)CC2)c1
comparing | greedy | noise 0.1
DRD2 proba: 0.9896 | SA: 2.6613 | QED: 0.9278



NC1CCC(CCN2CCC(c3ccccc3)CC2)CC1
ranking | greedy | noise 0.1
DRD2 proba: 0.991 | SA: 2.0648 | QED: 0.9126



7. Discussions

Comparison of feedback models

	Scoring model	Comparing model	Ranking model
ML model with highest MCC on DRD2 classification	3	2	1
Generate molecules that are likely to pass molecular descriptors	3	2	1
Generate molecules that are likely to pass DRD2 affinity threshold	1	2	3
Generate molecules that maximize DRD2 affinity	1	2	3
Generate molecules that maximize novelty score	2	1	3
Generate molecules that maximize diversity score	3	2	1
Generate molecules that minimize SA score	1	2	3
Generate molecules that maximize QED score	3	2	1

Rank 1 is best and Rank 3 is worst

- The scoring model (baseline) performs worst in learning chemist preferences but outperforms in generating molecules that maximize the average probability of DRD2 affinity and SA score.
- The ranking model outperforms in learning chemist preferences (with high MCC score for classification), passing most molecular descriptors, maximizing molecular diversity and achieving high QED scores.
- The comparing model (Bradley-Terry) keeps a balance between the scoring and ranking models, performing moderately well across all criteria, particularly in generating molecules with high novelty score.

Comparison of acquisition startegies

	random	uncertainty	greedy
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 affinity 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

Rank 1 is best and Rank 3 is worst

- Random feedback acquisition can lead to highest molecular novelty and diversity but produce molecules that do not pass most molecular descriptor filters and DRD2 affinity threshold.
- Greedy feedback acquisition is the best at generating molecules that maximize DRD2 affinity and minimize SA scores, but worst in terms of novelty and diversity.
- Uncertainty acquisition balances performance by placing second in most categories.

Comparison of labelling noise

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

Rank 1 is better than Rank 2

- Non-noisy feedback models (human = oracle) outperform in terms of producing molecules that maximize DRD2 affinity while improving synthetic accessibility
- Noisy feedback models (added noise of 0.1) outperform in terms of producing molecules that pass molecular descriptor filters while maximizing novelty, diversity, and QED scores
- Unclear how the noise added to user feedback models impacts learning performance of chemist preferences

8. Conclusions

Key takeaway: feedback mechanism, acquisition function and the noise in labelling offer robust customization in de-novo molecular design

=> Users can change them to fit specific molecules generation objectives

💡 To focus on generating molecules that directly targets chemical properties like DRD2 affinity synthesizability, the scoring model + greedy acquisition strategy + no labelling noise is a good option.

💡 To focus on generating novel molecules while keeping a balance across all metrics, the pairwise comparing Bradley-Terry model + uncertainty acquisition strategy is a good option.

💡 To focus on generating diverse and drug-like molecules as much as possible, the ranking ListNet model + random acquisition strategy + labelling noise is the best option.

🤔 Future research could further refine 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 modified molecular structure expected to be more promising than the original one)

🤔 We can validate our findings by performing real human experiments, where medicinal chemists interact with the feedback system through a Graphical User Interface.

References

[1] 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

[2] V. Gillet, De Novo Molecular Design, vol. 4. 2000

[3] 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.

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

[5] Robin Winter, Joren Retel, Frank Noé, Djork-Arné Clevert, Andreas Steffen, grünifai: interactive multiparameter optimization of molecules in a continuous vector space, Bioinformatics, Volume 36, Issue 13, July 2020, Pages 4093–4094, https://doi.org/10.1093/bioinformatics/btaa271

[6] Sundin, I., Voronov, A., Xiao, H. et al. Human-in-the-loop assisted de novo molecular design. J Cheminform 14, 86 (2022). https://doi.org/10.1186/s13321-022-00667-8

[7] Choung, OH., Vianello, R., Segler, M. et al. Extracting medicinal chemistry intuition via preference machine learning. Nat Commun 14, 6651 (2023). https://doi.org/10.1038/s41467-023-42242-1

[8] 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.

[9] 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

[10] T. Blaschke, J. Arús-Pous, H. Chen, C. Margreitter, C. Tyrchan, O. Engkvist, K. Pa- padopoulos, 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

[11] 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

[13] 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.

[14] Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. Journal of Medicinal Chemistry, 45(12), 2615-2623. doi:10.1021/jm020017n

[15] Muegge, I., Heald, S. L., & Brittelli, D. (2001). Simple selection criteria for drug-like chemical matter. Journal of Medicinal Chemistry, 44(12), 1841-1846. doi:10.1021/jm015507e

[16] 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.

[17] M. Benhenda, “Chemgan challenge for drug discovery: can ai reproduce natural chemical diversity?,” arXiv preprint arXiv:1708.08227, 2017.26

[18] Ertl, P., & Schuffenhauer, A. (2009). Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. Journal of Cheminformatics, 1, 8. doi:10.1186/1758-2946-1-8

[19] Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S., & Hopkins, A. L. (2012). Quantifying the chemical beauty of drugs. Nature Chemistry, 4, 90–98. doi:10.1038/nchem.1243

Thank you for for your attention

Questions and Answers

- Contact info: binh.nguyen@aalto.fi
  If you have inquiries about this research project

- Project code hosted at: https://github.com/SpringNuance/Human-In-The-Loop-De-Novo-Molecular-Design

Or you can scan with this QR code to access it