

Reproducibility Statement for Multi-target Parallel Drug Discovery with Multi-agent Orchestration (#333)

To ensure the scientific validity and reproducibility of our findings, this study was conducted with a commitment to transparent and rigorous methods. The entire end-to-end workflow, from initial target identification to final molecule generation and evaluation, is encapsulated within three sequential Jupyter notebooks (`0_target_mining_stanfordAIAgentConf.ipynb`, `1_ml_training_stanfordAIAgentConf.ipynb`, and `2_molecule_evaluation_stanfordAIAgentConf.ipynb`). All data used for training and validation were sourced from publicly accessible databases, including PubMed Abstracts for literature mining, ChEMBL for target-specific bioactivity data, and the Therapeutics Data Commons (TDC) datasets for ADMET property prediction. The large language model utilized for target hypothesis generation was Google's Gemini, and the generative chemistry model was NVIDIA's MolMIM.

For all machine learning tasks involving XGBoost, model training was performed across five distinct, fixed random seeds (13, 17, 23, 29, 31) to ensure the stability and robustness of our predictive models. The use of scaffold-based data splits for model validation further guarantees that the reported performance metrics are a conservative and realistic estimate of the models' ability to generalize to novel chemical matter. Besides, in our anonymized code repository, we have disclosed all the results yielded from every step, and all software dependencies are explicitly listed in the README.md file, allowing for the straightforward recreation of the computational environment and reproduction of the results. This comprehensive approach ensures that any researcher can replicate, validate, and extend upon the results presented herein.