MADD: Multi-Agent Drug Discovery Orchestra

Gleb V. Solovev [1], Alina B. Zhidkovskaya[*1], Anastasia Orlova[1], Nina Gubina[1], Anastasia Vepreva[1], Rodion Golovinskii[1], Ilya Tonkii[1], Ivan Dubrovsky[1], Ivan Gurev[1], Dmitry Gilemkhanov[1], Denis Chistiakov[1], Timur A. Aliev[1], Ivan Poddiakov[2], Galina Zubkova[2], Ekaterina V. Skorb[1], Vladimir Vinogradov[1], Alexander Boukhanovsky[1], Nikolay O. Nikitin[1], Andrei Dmitrenko [1,3], Anna V. Kalyuzhnaya [1], and Andrey Savchenko[2,4]

[1]ITMO University, Saint Petersburg, Russia; [2]Sber AI Lab, Moscow, Russia; [3]D ONE AG, Zurich, Switzerland; [4]HSE University, Moscow, Russia

*Email: alina.zhdk@yandex.ru

From Natural Language to Life-Saving Drugs

We present a multi-agent system that transforms natural language into expert-level drug discovery. Chemists can simply describe their target — ***"Generate GSK-3beta inhibitors with high docking score and low brain-blood barrier permeability"*** — and MADD delivers target molecules in SMILES format, ready for experimental validation. Our system currently addresses critical therapeutic challenges including Alzheimer's, Parkinson's, Multiple Sclerosis, Lung Cancer, Drug Resistance, and Dyslipidemia, making advanced AI tools accessible to researchers without coding expertise.

MADD Architecture

MADD uses four specialized agents to coordinate molecule discovery. The **Decomposer** breaks queries into subtasks and passes them to the **Orchestrator**, which builds action plans and calls specialized tools—including generative models (Transformer-based CVAE, LSTM-based GAN) and predictive ML models for property prediction—achieving 92.3% tool selection accuracy. The **Summarizer** compiles results into structured responses (95.3% accuracy), while the **Chat** Agent assists users with queries and missing data.

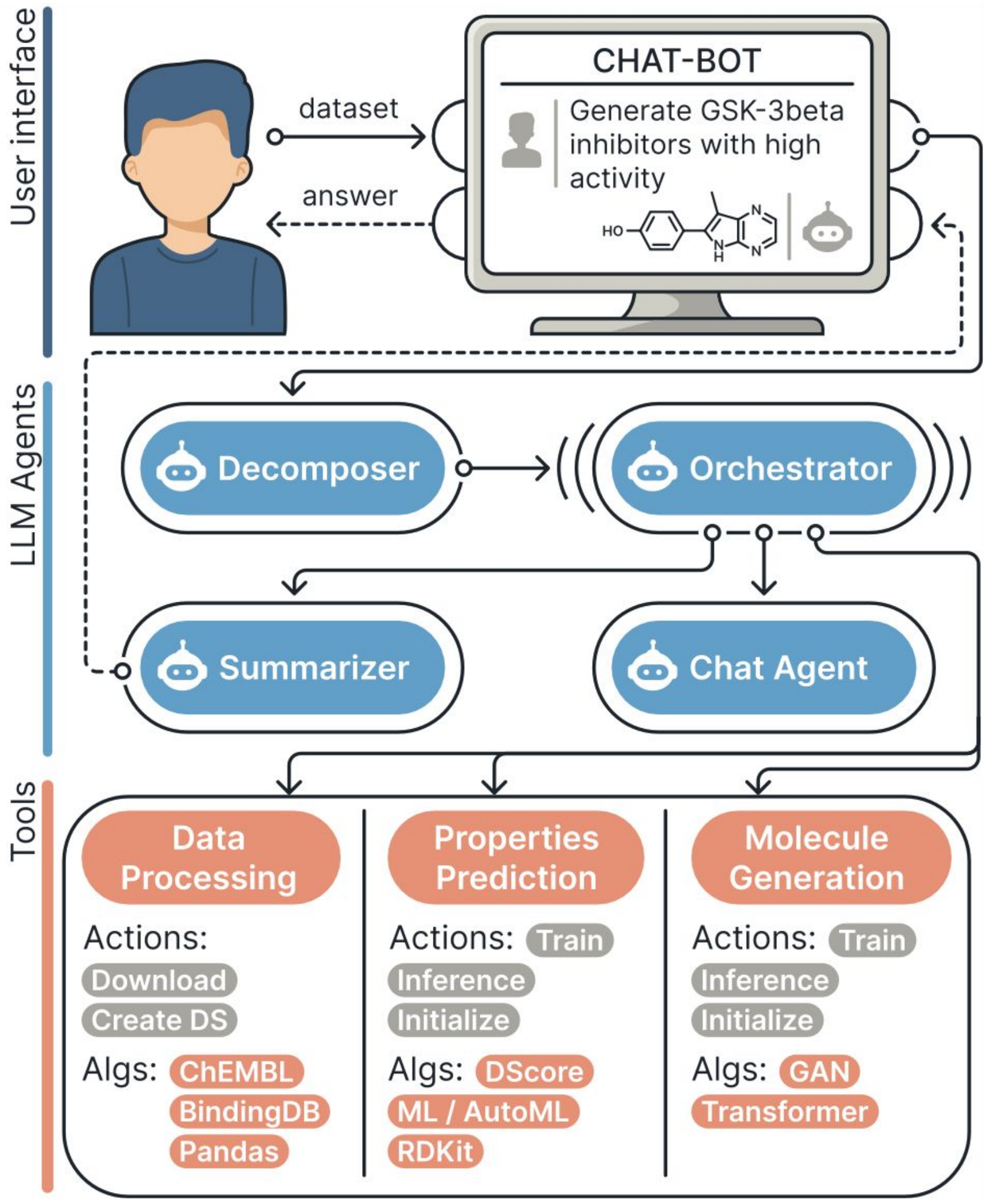

Fig. 1. MADD Architecture

Performance Comparison

MADD was evaluated on tasks of varying complexity (S/M/L), achieving 79.8-86.9% accuracy. The system significantly outperformed ChemAgent (12.4-16.4%) and specialized chemistry LLMs: X-LoRA-Gemma (0.0-0.44%), LlasMol (0.0-0.46%), and ChemDFM (0.0-5.31%). The **MADD architecture demonstrates superiority over single-agent and LLM solutions** in automating the drug discovery process — from molecule generation to property prediction.

Dataset	S	M	L
MADD	**86.9**	**84.3**	**79.8**
ChemAgent	12.4	15.3	16.4
LlasMol	0.46	0.24	0.0
X-Lora-Gemma	0.44	0.12	0.0
ChemDFM	5.31	0.33	0.0

Table 1. Comparison of the Accuracy (%) of MADD and baseline methods on datasets of different complexity

Novel Benchmark & Dataset

- 245 base expert-validated queries expanded across 3 difficulty levels (S: 1 task, M: 1-3 tasks, L: 4-5 tasks)
- Query types: target specification, molecular generation, property prediction, multi-step workflows
- 3M+ molecules with DS, IC_{50}, SA, QED for 6 diseases (Alzheimer's, Parkinson's, Multiple Sclerosis, Lung Cancer, Dyslipidemia, Drug Resistance)

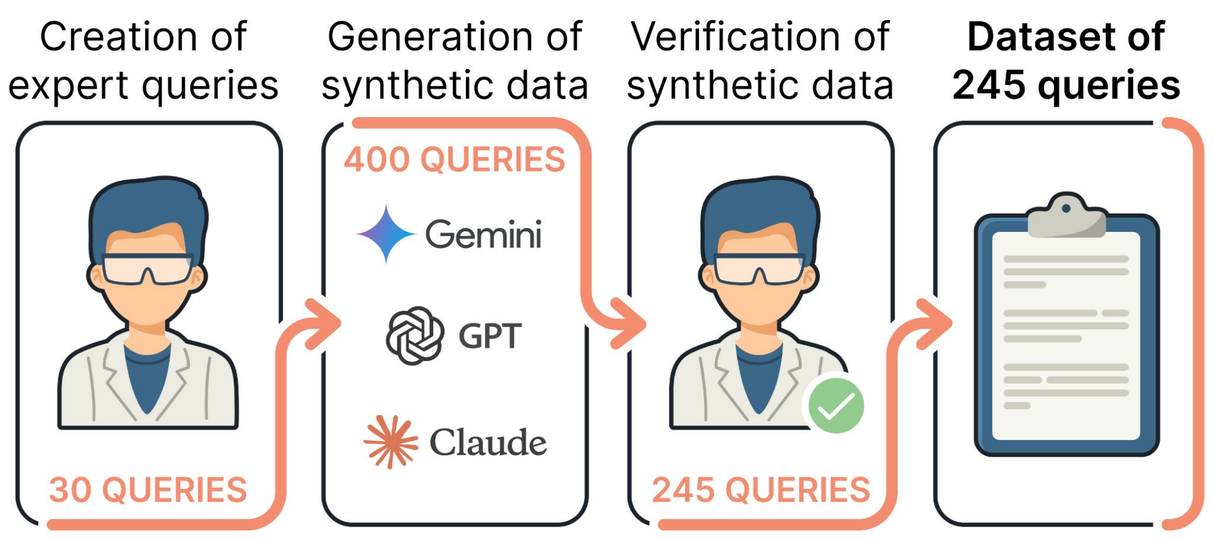

Fig. 2. The process of obtaining a validation dataset