*This code packages are for reviewers’ checking only. Please do not distribute to others. We will release the codes and results after the review.*

*We test our methods with the tasks developed by previous researchers, which are all open to use for common research. Our research is consistent with their intended use.*

Here we introduce how to run our code:

Set up the environment:

conda create -n PROMST python=3.10

conda activate PROMST

conda install pytorch torchvision torchaudio pytorch-cuda=12.1 -c pytorch -c nvidia

conda install conda-forge::tiktoken

pip install openai --upgrade

pip install pygame

conda install conda-forge::transformers

conda install anaconda::scikit-learn

pip install tiktoken

Run the code:

Each environment here has very similar structures. First you need to enter into each environment directory, then change the line 11 of LLM.py to fill your own OpenAI key code.

Then to create the training and testing environments, open env{i}\_create.py, change the local path to your own paths in the last four lines,

run it with python env{i}\_create.py

After that, comment out these last four lines for environment creation.

The above environment setup is for BoxLift, BoxNet1, BoxNet2, WareHouse, GridWorld1, GridWorld2.

For Blocksworld, Logistics, we don’t need to run env{i}\_create.py, the setup procedures are as follows:

Firstly,

pip install gym==0.26.2

pip install pddlgym

pip install tarski

Then go to pddlgym package locations (the path to install your pddlgym), and substitute the whole package pddlgym with the downloaded one in env\_data\_BlocksWorld or env\_data\_Logistics

For Webarena, Alfworld, Scienceworld, the setup procedures are as follows:

**Set up the environment for tasks except WebArena**

INSTALL\_WEBARENA=false bash ./setup.sh

**Set up the environment for WebArena**

# Please check whether the dubs and Xvfb are installed before building it

# For Ubuntu or Debian

dpkg -l | grep dbus # will return the info

systemctl status dbus # will return the status(active (running))

dpkg -l | grep xvfb # will return the info

#-----------------------------------------------------------------------#

# For CentOS

yum list installed | grep Xvfb # will return the Xvfb info

systemctl status dbus # will return the status(active (running))

dnf list installed | grep dbus # will return the dbus info

**Once done, you may install the webarena environment directly.**

INSTALL\_WEBARENA=true bash ./setup.sh

**Set up the OpenAI, Claude, and Mixtral keys in lines 8-10 of AgentBoard/agentboard/LLM.py, in lines 54-56 of AgentBoard/agentboard/llm/openai\_gpt.py**

Then we start the prompt optimization:

In each environment, directly run python env{i}-box-arrange-train\_MCTS.py (note that each local path in this script should also be changed before running). The model\_name\_promptLLM and model\_name\_testLLM in the script should also be chosen by the user. 'gpt-3.5-turbo-16k-0613' or 'gpt-4-1106-preview'.

You can also terminate the running and restart from one any prompt by running python env{i}-box-arrange-train\_MCTS\_restart.py, in which the local path and the path of the initial prompt should be changed.

After the training, you can test any prompt with any LLM model using script env{i}-box-arrange-test-optimized-prompt.py

The example command for running Alfworld is:

python agentboard/env9-box-arrange-train\_MCTS.py --cfg-path eval\_configs/main\_results\_all\_tasks.yaml --tasks alfworld --log\_path ./results/gpt-3.5-turbo-16k-0613 --project\_name evaluate-gpt-4 --experiment\_trial\_num 1 --model\_name\_promptLLM gpt-4-1106-preview --model\_name\_testLLM gpt-3.5-turbo-16k-0613 --min\_level 2 --n\_children 8 --n\_selected 2 --prompt\_method PROMST --with\_score\_model 'False' --base\_path ./alfworld\_result/ --max\_num\_steps 30

The example command for running BoxLift is:

python env3-box-arrange-train\_MCTS.py -experiment\_trial\_num 1 -input\_error\_prompt\_token\_limit 15000 -model\_name\_promptLLM gpt-4-1106-preview -model\_name\_testLLM gpt-3.5-turbo-16k-0613 -min\_level 2 -n\_children 8 -n\_selected 2 -prompt\_method PROMST -with\_score\_model 'False' -Training\_path ../BoxLift/train\_set/ -Testing\_path ../BoxLift/test\_set/ -base\_path ../BoxLift/