Code training guidance for R1-Code-Interpreter

Our code is based on Llama-factory (<https://github.com/hiyouga/LLaMA-Factory>) and VeRL (<https://github.com/volcengine/verl>) for the SFT and RL training and inference on multiple GPUs.

**Environment setup:**

conda create -n R1\_code\_inter python=3.11

conda activate R1\_code\_inter

pip install reasoning-gym

pip install -r requirements.txt

pip3 install torch==2.4.0 --index-url https://download.pytorch.org/whl/cu124

pip3 install flash-attn --no-build-isolation

cd verl

pip3 install -e .

cd ../LLaMA-Factory

pip install -e ".[torch,metrics]"

pip install deepspeed==0.15.2

pip install --upgrade huggingface\_hub

huggingface-cli login

**SFT training:**

cd LLaMA-Factory

sh finetune\_qwen\_7b\_1M.sh

**GRPO training:**

In train\_grpo\_3B.sh, fill your wandb key and python local path in line 1 and line 2.

sh train\_grpo\_3B.sh