WizardLM 2

Microsoft AI
Apr 15, 2024

🤗 Models   Code   𝕏 arXiv [Coming]

We introduce and opensource WizardLM-2, our next generation state-of-the-art large language models, which have improved performance on complex chat, multilingual, reasoning and agent. New family includes three cutting-edge models: WizardLM-2 8x22B, WizardLM-2 70B, and WizardLM-2 7B.

WizardLM-2 is the latest milestone in our effort in scaling up LLM post-training. One year ago, we have been iterating on training of Wizard series since our first work on Empowering Large Language Models to Follow Complex Instructions, then we accelerate the evolution to code and math reasoning scenarios. Since then, Evol-Instruct and Instruction&Process Supervised Reinforcement Learning (RLEIF) have become fundamental technologies for GenAI community. Recently, we have further optimized our methods and data quality, resulting in significant performance improvements, the outcome is WizardLM-2.

WizardLM-2 8x22B is our most advanced model, and the best opensource LLM in our internal evaluation on highly complex tasks. WizardLM-2 70B reaches top-tier reasoning capabilities and is the first choice in the same size. WizardLM-2 7B is the fastest and achieves comparable performance with existing 10x larger opensource leading models.

Following, we will introduce the overall methods and main experimental results, and the associated details and rethinking will be presented in our upcoming paper.
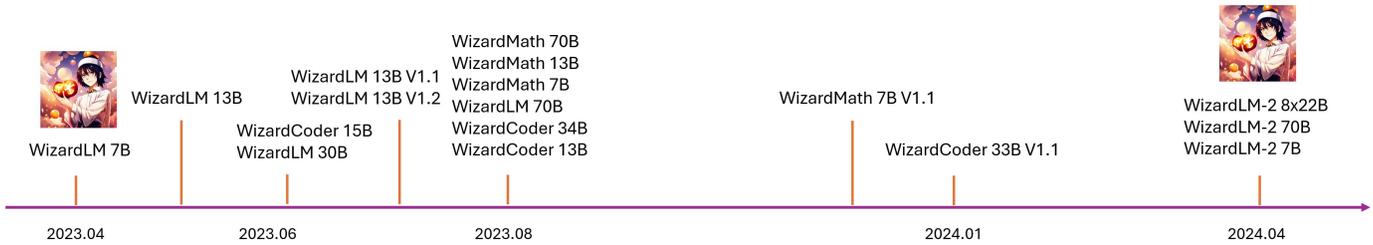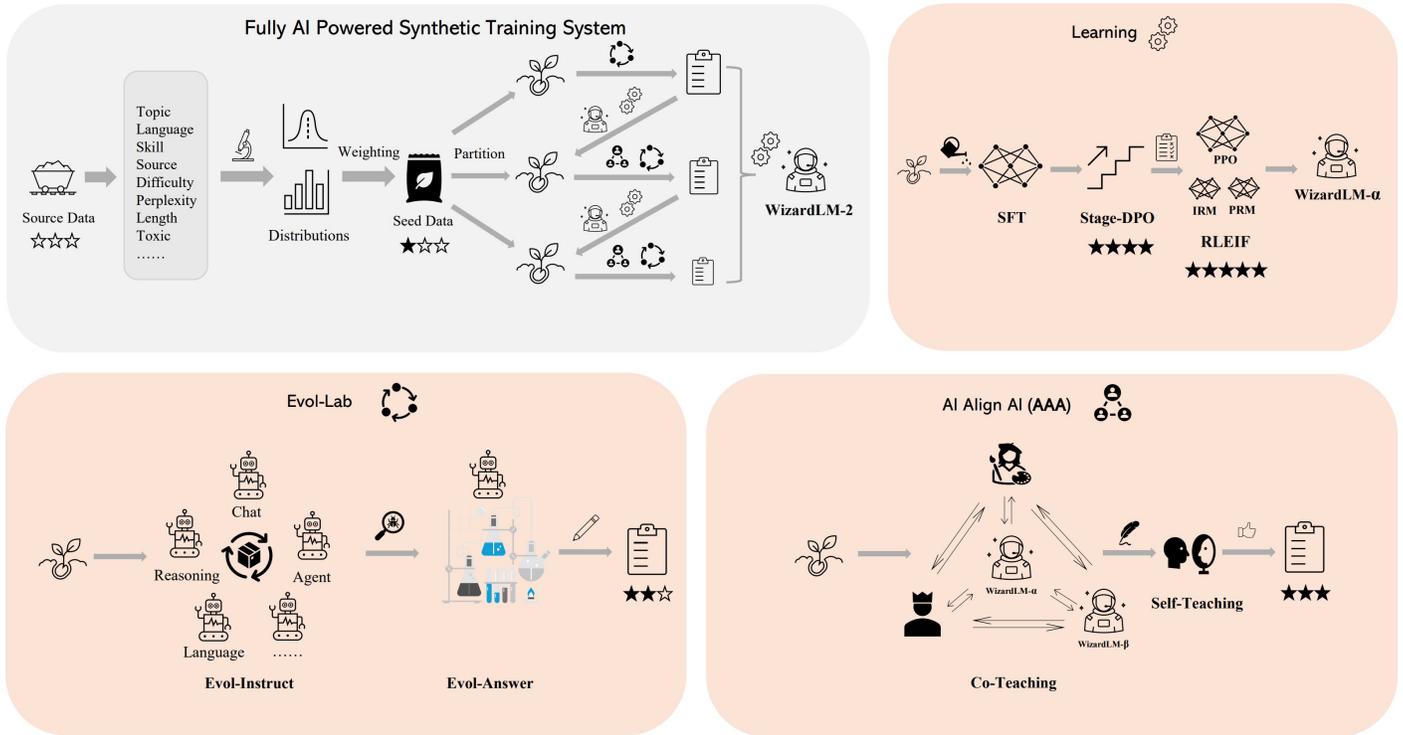

Method Overview

As the natural world's human-generated data becomes increasingly exhausted through LLM training, we believe that: the data carefully created by AI and the model step-by-step supervised by AI will be the sole path towards more powerful AI.

In the past one year, we built a fully AI powered synthetic training system:

- Data Pre-Processing:
  - Data Analysis: We use this pipline to get the distribution of different attributes for new source data. This helps us to have a preliminary understanding of the data.
  - Weighted Sampling: The distribution of the best training data is always not consistent with the natural distribution of human chat corpus, thus we need adjust the weights of various attributes in the training data based on experimental experience.
- Progressive Learning: Unlike the common practice of using all data for one-time training, we found that using different data partitions and progressively training stage-by-stage can achieve better results with less data. In each stage, we firstly feed the data slice to following Evol Lab to get more diverse and complex [instruction, response] pairs. Immediately, we leverage a new framework named "AI Align AI" (AAA) which can group multi state-of-the-art LLMs to teach and improve each other. Finally, we successively apply the Supervised Learning, Stage-DPO, and RLEIF to optimize each variant.
- Evol Lab:
  - Evol-Instruct: Recently, we have dedicated significant effort to reassessing the various issues within the original Evol-Instruct method and have initiated preliminary modifications. The new method enables various agents to automatically generate high quality instructions.
  - Evol-Answer: Guiding the model to generate and rewrite responses multiple times can improve its logic, correctness, and affinity.
- AI Align AI (AAA):
  - Co-Teaching: We collect WizardLMs, and various licensed opensource and proprietary state-of-the-art models, then let them co-teach and improve each other, the teaching contains simulated chat, quality judging, improvement suggestions and closing skill gap, etc.
  - Self-Teaching: WizardLM can generate new evolution training data for supervised learning and preference data for reinforcement learning via activate learning from itself.
- Learning:
  - Supervised Learning.
  - Stage-DPO: For more effective offline reinforcement learning, we also split the preference data to different slices, and progressively improve the model stage by stage.
  - RLEIF: We employ instruction quality reward model (IRM) combined with the process supervision reward model (PRM) to achieve more precise correctness in the online reinforcement learning

Fully AI Powered Synthetic Training System

Source Data
☆☆☆

Topic
Language
Skill
Source
Difficulty
Perplexity
Length
Toxic
......

Distributions

Weighting

Seed Data
★☆☆

Partition

WizardLM-2

Learning ⚙

SFT
★★★★

Stage-DPO

RLEIF
★★★★★

PPO

IRM   PRM

WizardLM-α

Evol-Lab ⟳

Chat
Reasoning   Agent
Language   ......

Evol-Instruct

Evol-Answer
★★☆

AI Align AI (AAA)

WizardLM-α

WizardLM-β

Self-Teaching
★★★
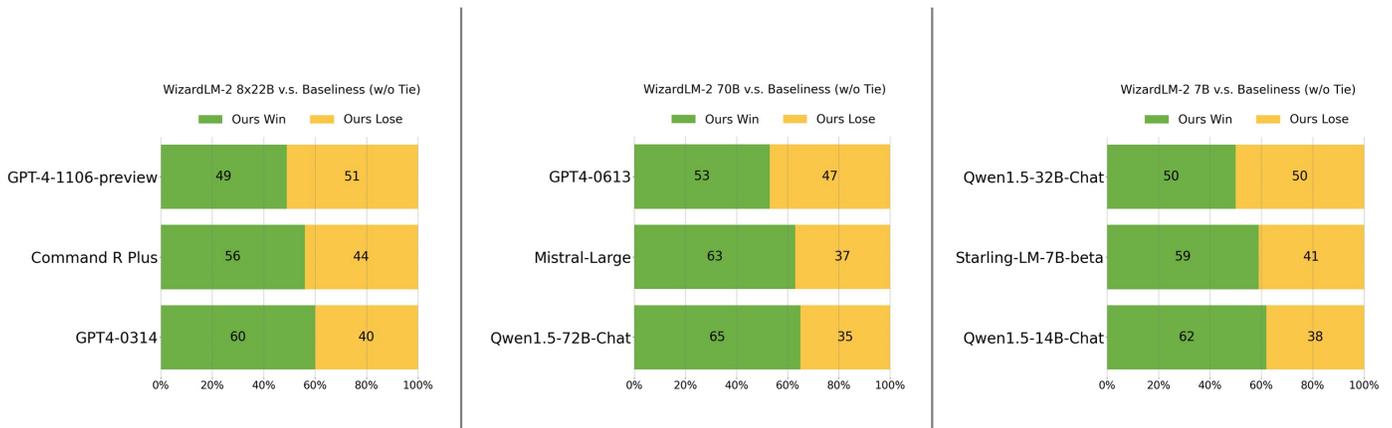
Co-Teaching

WizardLM 2 Capacities

To present a comprehensive overview of the performance of WizardLM-2, we conduct both human and automatic evaluations between our models and diverse baselines. As indicated in the following main experimental results, WizardLM-2 demonstrates highly competitive performance compared to those leading proprietary works and consistently outperforms all the existing state-of-the-art opensource models. More associated details and thinking will be presented in our upcoming paper.

Human Preferences Evaluation

We carefully collected a complex and challenging set consisting of real-world instructions, which includes main requirements of humanity, such as writing, coding, math, reasoning, agent, and multilingual. We perform a blind pairwise comparison between WizardLM-2 and baselines. To each annotator, responses from all models are presented, which are randomly shuffled to hide their sources. We report the win:loss rate without tie:

- WizardLM-2 8x22B is just slightly falling behind GPT-4-1106-preview, and significantly stronger than Command R Plus and GPT4-0314.
- WizardLM-2 70B is better than GPT-0613, Mistral-Large, and Qwen1.5-72B-Chat.
- WizardLM-2 7B is comparable with Qwen1.5-32B-Chat, and surpasses Qwen1.5-14B-Chat and Starling-LM-7B-beta.

Through this human preferences evaluation, WizardLM-2's capabilities are very close to the cutting-edge proprietary models such as GPT-4-1106-preview, and significantly ahead of all the other open source models.

WizardLM-2 8x22B v.s. Baseliness (w/o Tie)

| Ours Win | Ours Lose |

Baseline	Ours Win	Ours Lose
GPT-4-1106-preview	49	51
Command R Plus	56	44
GPT4-0314	60	40

WizardLM-2 70B v.s. Baseliness (w/o Tie)

| Ours Win | Ours Lose |

Baseline	Ours Win	Ours Lose
GPT4-0613	53	47
Mistral-Large	63	37
Qwen1.5-72B-Chat	65	35

WizardLM-2 7B v.s. Baseliness (w/o Tie)

| Ours Win | Ours Lose |

Baseline	Ours Win	Ours Lose
Qwen1.5-32B-Chat	50	50
Starling-LM-7B-beta	59	41
Qwen1.5-14B-Chat	62	38

MT-Bench

We also adopt the automatic MT-Bench evaluation framework based on GPT-4 proposed by lmsys to assess the performance of models. The WizardLM-2 8x22B even demonstrates highly competitive performance compared to the most advanced proprietary works such as GPT-4-Trubo and Glaude-3. Meanwhile, WizardLM-2 7B and WizardLM-2 70B are all the top-performing models among the other leading baselines at 7B to 70B model scales.

Model	MT-Bench	Model	MT-Bench	Model	MT-Bench
Claude 3 Opus	9.43	WizardLM-2 70B	8.92	WizardLM-2 7B	8.28
GPT-4-1106-Preview	9.32	Mistral Large	8.66	StarlingLM 7B Beta	8.12
Claude 3 Sonnet	9.18	Qwen1.5 72B Chat	8.61	Claude-2	8.06
WizardLM-2 8x22B	9.12	Mistral Medium	8.61	Yi 34B Chat	7.88
GPT-4-0314	8.96	Claude-2.1	8.18	OpenChat-3.5-0106	7.80

Usage

The model weights of WizardLM-2 8x22B and WizardLM-2 7B are shared on Huggingface, and WizardLM-2 70B and the demo of all the models will be available in the coming days. Please use the same system prompts strictly with us to guarantee the generation quality.

❗ **Note for model system prompts usage:**
WizardLM-2 adopts the prompt format from **Vicuna** and supports **multi-turn** conversation.
The prompt should be as following:

```
A chat between a curious user and an artificial intelligence assistant. The assistant gives helpful, detailed, and polite answers to the user's questions. USER: Hi ASSISTANT: Hello.</s>USER: Who
```

Inference WizardLM-2 Demo Script
We provide a WizardLM-2 inference demo code on our github.

License

The License of WizardLM-2 8x22B and WizardLM-2 7B is Apache2.0. The License of WizardLM-2 70B is Llama-2-Community.