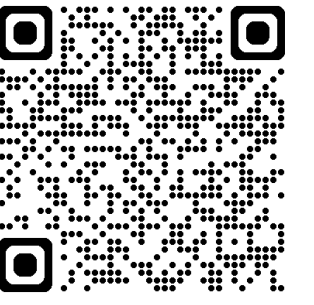


Empowering and Assessing the Utility of Large Language Models in Crop Science



GitHub:

Hang Zhang^{1*}, Jiawei Sun^{1*}, Renqi Chen^{1*}, Wei Liu¹, Zhonghang Yuan¹, Xinzhe Zheng¹, Zhefan Wang¹, Zhiyuan Yang⁴, Hang Yan¹, Hansen Zhong¹, Xiqing Wang³, Wanli Ouyang¹, Fan Yang^{2†}, Nanqing Dong^{1†}

¹ Shanghai Artificial Intelligence Laboratory ² Yazhouwan National Laboratory ³ China Agricultural University ⁴ Hangzhou Dianzi University

1. Motivation for the CROP

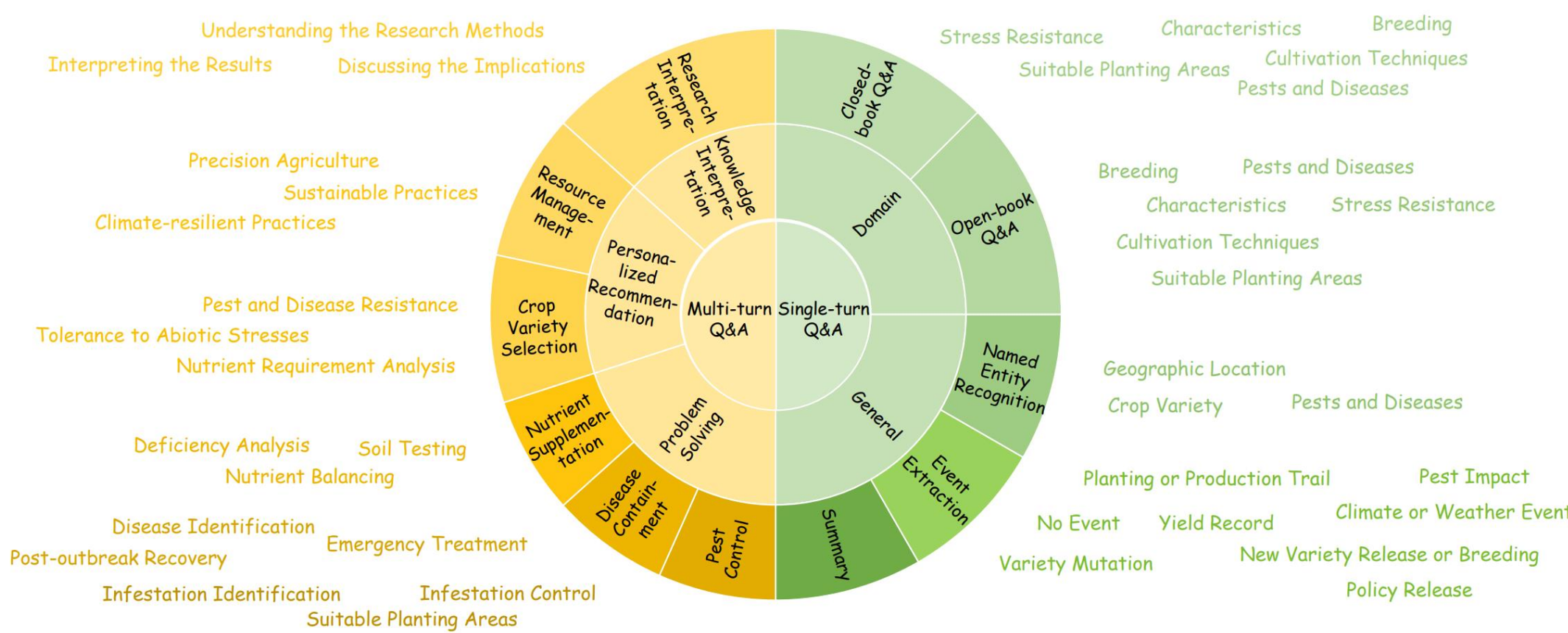
- Crop cultivation has historically been a significant challenge, with uncertainties in harvest yields.
- Recent progress in large language models (LLMs), offers promising opportunities. LLMs can generate professional knowledge and context in response to user inquiries, finding applications in various fields.
- However, LLMs currently face limitations in specific areas, such as pest management, and the existing datasets for agricultural evaluation are insufficient in quantity and locality.

2. Overview of the CROP

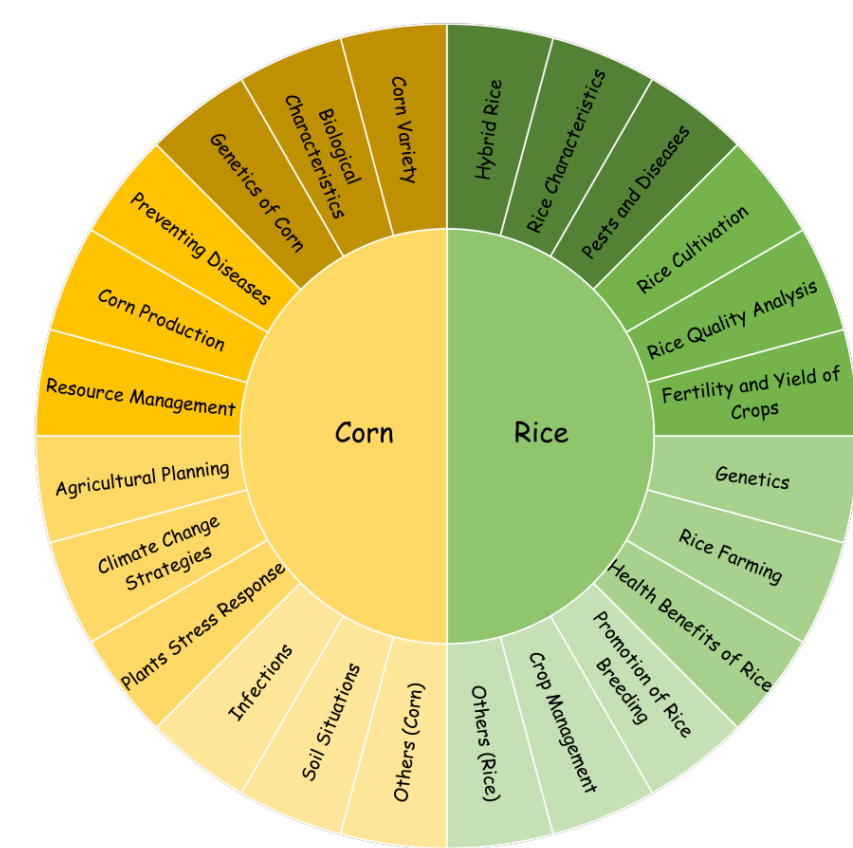
To harness the full potential of LLMs for crop science, we propose a suite called CROP, which encompasses

- an extensive instruction-tuning dataset, designed to enhance the domain-specific proficiency of LLMs in crop science.
- a meticulously constructed benchmark, aimed at assessing the performance of LLMs across a variety of domain-related tasks.

Topics of CROP dataset



Topics of CROP benchmark



An example in CROP

Q1: I've been noticing some of my rice plants are stunted and the leaves are turning yellow. What could be causing this issue?

A1: Yellowing and stunting in rice can be symptoms of several issues including nutrient deficiencies, water stress, or pest infestations such as stem borers or leaf folders. Have you noticed any visible insects on the plants or any other unusual signs?

Q2: Yes, I've seen some moths around the plants in the evening and there are holes in the leaves. Could this be related to the yellowing?

A2: It sounds like your rice crop might be infested with stem borers or leaf folders. These pests can cause the symptoms you're describing as they damage the plant's ability to transport nutrients and water by boring into the stems or folding the leaves.

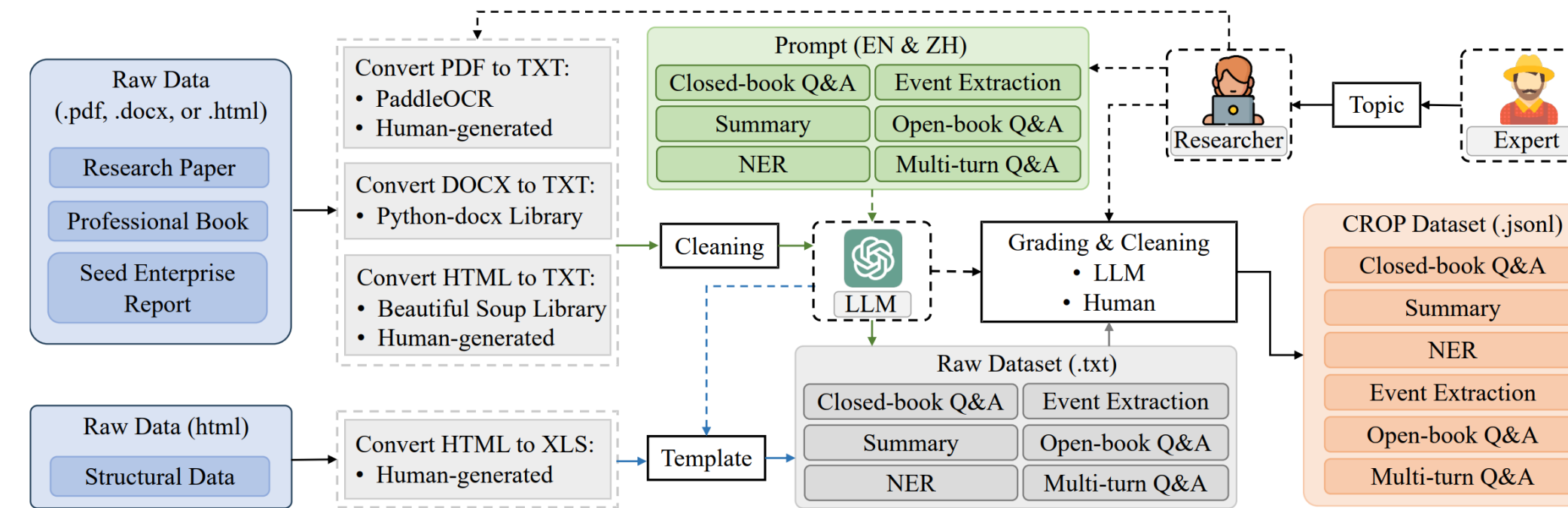
Q3: What can I do to manage these pests and protect my crop?

A3: Considering the infestation, it would be wise to apply integrated pest management practices. This includes using pheromone traps to monitor and reduce moth populations and possibly applying biological control agents like *Trichogramma* wasps, which are natural enemies of stem borers. Also, ensure that your crop is healthy with proper nutrition and water management to resist pest damage more effectively.

Q4: Thank you for the advice. I'll look into these options. Is there a specific time or method for applying these biological controls?

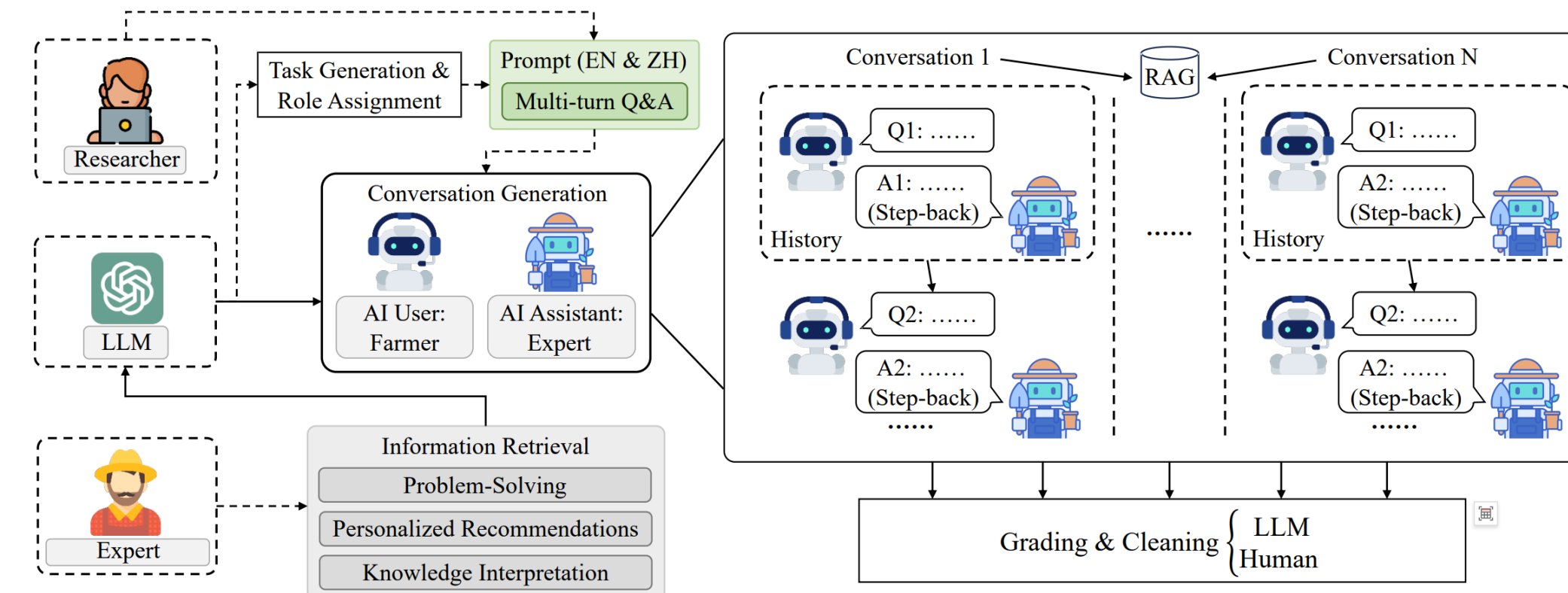
A4: For biological controls like *Trichogramma* wasps, it's best to release them early in the season before the pest population peaks. Local agricultural extension services or a professional pest management expert can provide guidance on timing and application methods specific to your area to ensure the best results.

3. CROP Dataset Collection



Single-turn dialogue collection pipeline:

- Raw data is first converted to TXT or XLS format.
- Prompt an LLM to generate Q&As from unstructured data or design templates that transform structured data into dialogue format.
- Filtering steps with both human and LLM involved.



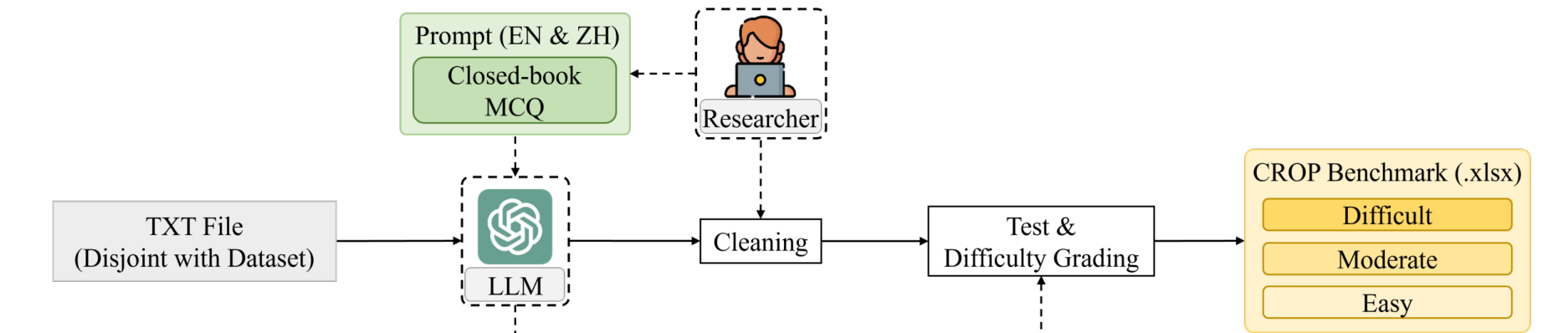
Multi-turn dialogue collection pipeline:

- An LLM creates tasks under the guidance of domain experts and assigns roles to two agents.
- Using task-dependent prompts from researchers, the LLM generates dialogues with RAG.
- Filtering steps.

4. CROP Dataset Analysis

- The single-turn dialogues comprise 210,038 high-quality samples.
 - 140,056 dialogue samples for rice
 - 69,482 dialogue samples for corn
- The multi-turn dialogues include 1,871 high-quality samples.
 - Each task within the multi-turn dialogues has at least 80 samples
 - 3-5 turns of dialogue

5. CROP Benchmark Collection



- We prompt an LLM to generate MCQs from TXT files.
- After additional filtering steps with both human and LLM involved, we get the CROP benchmark, comprising three difficulty levels.

6. CROP Benchmark Analysis

- 5,045 questions in the benchmark have three difficulty levels:
 - Easy (1613, 31.97%)
 - Moderate (2754, 53.72%)
 - Difficult (722, 14.31%)
- CROP benchmark consists of 5045 Chinese and English MCQs and covers 22 countries across six continents.

7. Experiments

- The performance of selected LLMs on the CROP benchmark

Model	Access	Size	Overall ↑	Difficulty		
				Easy ↑	Moderate ↑	Difficult ↑
Commercial LLMs						
GPT-4 ¹	API	N/A	0.856	1.000 ²	1.000 ²	0.000 ²
GPT-3.5 ¹	API	N/A	0.328	1.000 ²	0.000 ²	0.061
Claude-3 ¹	API	N/A	0.900	0.982	0.968	0.458
Qwen ¹	API	N/A	0.866	0.987	0.945	0.301
Open-source LLMs						
LLaMA3-Base	Weights	8B	0.348	0.443	0.341	0.161
+CQIA	Weights	8B	0.643 (+0.295)	0.791 (+0.348)	0.651 (+0.310)	0.281 (+0.120)
+CROP	Weights	8B	0.752 (+0.404)	0.866 (+0.432)	0.772 (+0.431)	0.378 (+0.217)
+CQIA+CROP	Weights	8B	0.754 (+0.406)	0.918 (+0.475)	0.779 (+0.438)	0.295 (+0.134)
Qwen1.5-Base	Weights	7B	0.646	0.799	0.646	0.302
+CQIA	Weights	7B	0.688 (+0.042)	0.880 (+0.081)	0.689 (+0.043)	0.258 (-0.044)
+CROP	Weights	7B	0.676 (+0.030)	0.849 (+0.050)	0.688 (+0.042)	0.202 (-0.100)
+CQIA+CROP	Weights	7B	0.709 (+0.063)	0.910 (+0.111)	0.704 (+0.058)	0.227 (-0.075)
InternLM2-Base	Weights	7B	0.368	0.445	0.381	0.148
+CQIA	Weights	7B	0.723 (+0.355)	0.861 (+0.416)	0.750 (+0.369)	0.317 (+0.169)
+CROP	Weights	7B	0.748 (+0.380)	0.945 (+0.500)	0.761 (+0.380)	0.212 (+0.064)
+CQIA+CROP	Weights	7B	0.768 (+0.400)	0.939 (+0.494)	0.794 (+0.413)	0.285 (+0.137)

- The performance of LLMs under various training epochs and languages.

Model	Epoch	Size	Overall ↑	Difficulty			Language		
				Easy ↑	Moderate ↑	Difficult ↑	Chinese ↑	English ↑	Variation ↓
LLaMA3-Base	N/A	8B	0.348	0.443	0.341	0.161	0.327	0.369	4.2%
+CQIA+CROP	1	8B	0.738	0.903	0.758	0.292	0.719	0.757	3.8%
+CQIA+CROP	2	8B	0.742	0.902	0.772	0.271	0.729	0.755	2.6%
+CQIA+CROP	4	8B	0.754	0.918	0.779	0.295	0.738	0.770	3.2%
Qwen1.5-Base	N/A	7B	0.646	0.799	0.646	0.302	0.667	0.624	4.3%
+CQIA+CROP	1	7B	0.702	0.910	0.717	0.183	0.725	0.680	4.5%
+CQIA+CROP	2	7B	0.670	0.875	0.677	0.181	0.690	0.649	4.1%
+CQIA+CROP	4	7B	0.709	0.910	0.704	0.227	0.717	0.686	3.1%
InternLM2-Base	N/A	7B	0.368	0.445	0.381	0.148	0.409	0.327	8.2%
+CQIA+CROP	1	7B	0.764	0.942	0.787	0.276	0.770	0.757	3.3%
+CQIA+CROP	2	7B	0.809	0.909	0.855	0.414	0.811	0.807	0.4%
+CQIA+CROP	4	7B	0.768	0.939	0.794	0.285	0.770	0.766	0.4%