Table 13: **Results on MMLU**. All are tested with five-shot accuracy. We provide the reported results of the other models for comparison.

Model	**Params**	**Average**	**STEM**	**Social Sciences**	**Humanities**	**Others**
MPT	7B	26.8	25.3	27.1	26.7	28.2
	30B	46.9	39.0	52.8	44.5	52.9
Falcon	7B	26.2	26.2	24.7	26.4	27.4
	40B	55.4	45.5	65.4	49.3	65.0
ChatGLM2	6B	47.9	41.2	54.4	43.7	54.5
	12B	56.2	48.2	65.1	52.6	60.9
InternLM	7B	51.0	-	-	-	-
Baichuan2	7B	54.2	-	-	-	-
	13B	59.2	-	-	-	-
XVERSE	13B	55.1	44.5	64.4	50.5	62.9
LLaMA	7B	35.1	30.5	38.3	34.0	38.1
	13B	46.9	35.8	53.8	45.0	53.3
	33B	57.8	46.0	66.7	55.8	63.4
	65B	63.4	51.7	72.9	61.8	67.4
Llama 2	7B	45.3	36.4	51.2	42.9	52.2
	13B	54.8	44.1	62.6	52.8	61.1
	34B	62.6	52.1	71.8	59.4	69.2
	70B	68.9	58.0	80.3	65.0	74.6
Qwen	1.8B	44.6	39.6	50.0	40.4	51.0
	7B	58.2	50.2	68.6	52.5	64.9
	14B	**66.3**	**59.4**	**76.2**	**60.9**	**71.8**

Table 14: **Leaderboard results of C-Eval**. We include the results of both proprietary models and open-source models. Note that there are a number of models on the leaderboard with very few details, in terms of proprietary models, we only report the results of GPT-3.5, GPT-4, InternLM and ChatGLM2.

Model	Params	Avg.	Avg. (Hard)	STEM	Social Sciences	Humanities	Others
Proprietary models							
GPT-3.5	-	54.4	41.4	52.9	61.8	50.9	53.6
GPT-4	-	68.7	**54.9**	**67.1**	77.6	64.5	67.8
InternLM	123B	68.8	50.0	63.5	81.4	72.7	63.0
ChatGLM2	-	**71.1**	50.0	64.4	**81.6**	**73.7**	**71.3**
Open-source models							
ChatGLM2	6B	51.7	37.1	48.6	60.5	51.3	49.8
InternLM	7B	52.8	37.1	48.0	67.4	55.4	45.8
Baichuan2	7B	54.0	-	-	-	-	-
	13B	58.1	-	-	-	-	-
XVERSE	13B	54.7	33.5	45.6	66.2	58.3	56.9
Qwen	1.8B	54.7	41.8	50.8	69.9	56.3	46.2
	7B	63.5	46.4	57.7	78.1	66.6	57.8
	14B	**72.1**	**53.7**	**65.7**	**85.4**	**75.3**	**68.4**

In terms of MMLU, we report the detailed results in Table 13. In terms of C-Eval, we report the results in Table 14. For the rest of the datasets, we report the results in Table 15. Note that AGIEval includes

[5] https://github.com/openai/openai-python/blob/main/chatml.md