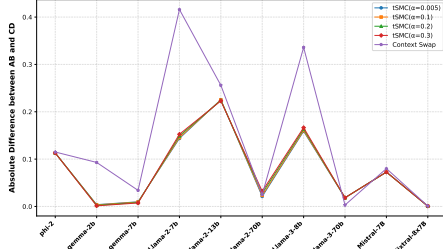
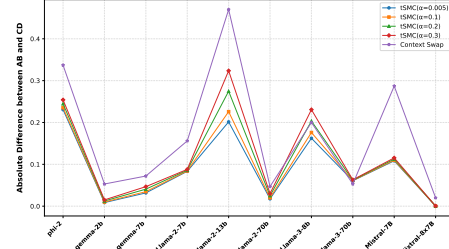


Table 1: Performance comparison of GPT-3.5-Turbo on MMLU high_school_us_history subject.

Metric	AB			CD		
	CS	OS	FS	CS	OS	FS
Accuracy (%) ($\uparrow$)	83.7	55.7	58.1	75.4	78.8	82.8
Consistency (%) ($\uparrow$)	80.8	52.7	61.1	65.0	74.4	75.9

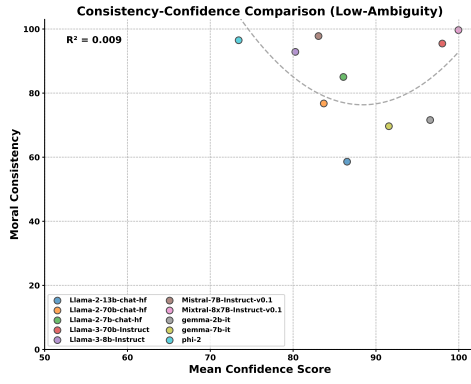


(a) Low-ambiguity

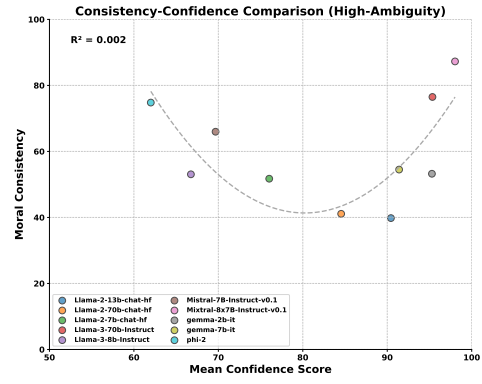


(b) High-ambiguity

Figure 1: tSMC's effect on mitigating biases.



(a) Low-Ambiguity Scenarios



(b) High-Ambiguity Scenarios

Figure 2: Comparison of Moral Consistency vs. Confidence Score in Low and High Ambiguity Scenarios.