Methods		ChatGPT			LLaMA2-70B		
		HotpotQA	StrategyQA	MMLU	HotpotQA	StrategyQA	MMLU
Baselines	Standard	29.8(0)	57.6(0)	69.0(0)	30.0(0)	65.9(0)	64.7(0)
	CoT	34.8(0)	55.9(0)	49.1(0)	35.6(0)	63.8(0)	39.3(0)
	Self-Consistency	39.4(0)	58.5(0)	70.0(0)	37.4(0)	67.7(0)	67.2(0)
	ReAct	32.0(2,114)	55.5(709)	55.8(1,824)	32.4(2,094)	58.1(890)	30.7(2,808)
	ReAct+Backoff	35.4(2,114)	61.6(709)	59.8(1,824)	37.0(2,094)	66.8(890)	62.8(2,808)
Ours	UALA-S	38.2(403)	65.5(134)	69.8(662)	36.4(350)	69.0(298)	56.7(1354)
	UALA-S+Backoff	39.2(403)	66.4(134)	71.4(662)	37.3(350)	71.6(298)	69.8(1354)
	UALA-S+Oracle	41.4(403)	67.5(134)	75.8(662)	42.2(350)	80.1(298)	70.5(1354)
	UALA-M	39.4(1,199)	63.8(234)	67.9(641)	36.4(1,302)	66.4(492)	56.7(1,196)
	UALA-M+Backoff	**41.2(1,199)**	**66.9(234)**	**72.2(641)**	**40.4(1,302)**	**72.2(492)**	**70.3(1,196)**