		Simple Problems					Contest-Level Problems		
LLM	Approach	HumanEval	HumanEval ET	EvalPlus	MBPP	MBPP ET	APPS	xCodeEval	CodeContest
ChatGPT	Direct	48.1%	37.2%	66.5%	49.8%	37.7%	8.0%	17.9%	5.5%
	CoT	53.9%	45.5%	-	54.5%	39.6%	7.3%	23.6%	6.1%
	Self-Planning	60.3%	46.2%	-	55.7%	41.9%	9.3%	18.9%	6.1%
	Analogical	63.4%	57.9%	58.5%	70.5%	46.1%	6.7%	15.1%	7.3%
	Reflexion	67.1%	49.4%	61.6%	73.0%	47.4%	-	-	-
	MapCoder	**80.5%** ↑ 67.3%	**70.1%** ↑ 88.5%	**70.1%** ↑ 5.4%	**78.3%** ↑ 57.3%	**54.4%** ↑ 44.3%	**11.3%** ↑ 41.3%	**27.4%** ↑ 52.6%	**12.7%** ↑ 132.8%
GPT4	Direct	80.1%	73.8%	81.7%	81.1%	54.7%	12.7%	32.1%	12.1%
	CoT	89.0%	61.6%	-	82.4%	56.2%	11.3%	36.8%	5.5%
	Self-Planning	85.4%	62.2%	-	75.8%	50.4%	14.7%	34.0%	10.9%
	Analogical	66.5%	48.8%	62.2%	58.4%	40.3%	12.0%	26.4%	10.9%
	Reflexion	91.0%	78.7%	81.7%	78.3%	51.9%	-	-	-
	MapCoder	**93.9%** ↑ 17.2%	**82.9%** ↑ 12.4%	**83.5%** ↑ 2.2%	**83.1%** ↑ 2.5%	**57.7%** ↑ 5.5%	**22.0%** ↑ 73.7%	**45.3%** ↑ 41.2%	**28.5%** ↑ 135.1%