	Geometry	**Graph**			**Math**		**Game**
Model	Geometry	Maxflow	Isomorphism	Connectivity	Convexity	Parity	Winner ID
Prior LLMs without visual inputs							
Gemini-Pro	\	15.6	47.7	50.0	87.9	48.2	8.1
Claude 3 OPUS	\	56.3	50.0	82.0	93.0	77.6	74.4
Mixtral 8x7B [19]	\	8.6	50.0	62.5	69.1	41.7	7.4
LLaMA-2-70B [45]	\	18.0	50.0	50.0	74.2	33.3	12.4
Latest multimodal LLMs + Visual Sketchpad							
GPT-4 Turbo	37.5	32.8	62.5	66.0	57.0	80.5	50.4
+ Sketchpad	45.8	63.3	64.2	95.1	**93.1**	**93.1**	74.3
	+8.3	+30.5	+1.7	+29.1	+25.4	+12.6	+23.9
GPT-4o	62.5	25.0	50.8	96.1	87.2	84.4	61.1
+ Sketchpad	**66.7**	**66.3**	**65.3**	**98.1**	90.1	88.1	**81.2**
	+4.2	+41.3	+14.5	+2.0	+2.9	+3.7	+20.1

Table 1: Accuracy scores on geometry problems, graph algorithms, mathematical functions, and game. SKETCHPAD yields large performance gains across all tasks and outperform all baselines.