

Discipline	LLaVA-OneVision-72B			GPT4o		
	CoT Acc	Direct Acc	Difference	CoT Acc	DIRECT Acc	Difference
Art and Design	20.42%	37.53%	-17.12%	63.14%	61.55%	1.58%
Science	23.89%	22.61%	1.28%	46.67%	38.46%	8.22%
Business	29.26%	24.50%	4.76%	57.45%	42.79%	14.66%
Humanities and Social Science	32.14%	36.60%	-4.46%	60.08%	57.87%	2.21%
Health and Medicine	19.22%	20.78%	-1.56%	49.68%	44.34%	5.34%
Tech and Engineering	22.98%	20.65%	2.33%	37.72%	23.23%	14.49%

Table 2: Comparison of CoT and direct accuracy of two representative models across disciplines in the Vision Input setting. Difference = CoT Acc - Direct Acc.