

# HONESTY EVALUATION

Dataset	Instruction-1		Instruction-2	
Model	CHAIR <sub>i</sub>	CHAIR <sub>s</sub>	CHAIR <sub>i</sub>	CHAIR <sub>s</sub>
BLIP-2	3.40	4.00	<b>2.75</b>	<b>3.50</b>
InstructBLIP	2.38	3.45	5.16	14.48
LLaVA	9.98	31.10	23.40	61.50
LLaVA-HF	4.26	5.40	6.05	10.80
mPLUG	15.10	21.65	25.89	73.50
miniGPT4	5.70	13.40	10.60	30.45
DRESS👗	<b>2.34</b>	<b>3.30</b>	4.74	9.84