

Table 1. ANOVA p-values. Comparing Human and model pseudo-profound bullshit statements ratings.

Model	F-value	P-value	Average
Humans			2.56 (0.24)
Flan-T5 (XL)	30.66	<0.001	3.13 (0.51)
GPT-4	99.56	<0.001	3.70 (0.57)
Llama-2 (13B)	200.28	<0.001	3.55 (0.29)
Llama-2 + RLHF (13B)	400.15	<0.001	3.88 (0.27)
Tk-Instruct (11B)	0.13	0.717	2.60 (0.49)
Vicuna (13B)	78.75	<0.001	3.56 (0.57)