

p-value of Dunn test for  $\beta$  (order 27)

GPT-3.5	1.00	0.00	0.22	0.18	0.62	0.02	0.51
GPT-4	0.00	1.00	0.00	0.00	0.00	0.00	0.00
Human	0.22	0.00	1.00	1.00	1.00	1.00	1.00
LaMA-13B	0.18	0.00	1.00	1.00	1.00	1.00	1.00
LaMA-30B	0.62	0.00	1.00	1.00	1.00	1.00	1.00
LaMA-65B	0.02	0.00	1.00	1.00	1.00	1.00	1.00
LLaMA-7B	0.51	0.00	1.00	1.00	1.00	1.00	1.00
	PT-3.5	GPT-4	human	A-13B	A-30B	A-65B	MA-7B