

p-value of Dunn test for  $\beta$  (order 3)

	GPT-3.5	GPT-4	Human	LaMA-13B	LaMA-30B	LaMA-65B	LLaMA-7B
GPT-3.5	1.00	1.00	0.00	0.00	0.00	0.00	0.00
GPT-4	1.00	1.00	0.00	0.00	0.00	0.00	0.00
Human	0.00	0.00	1.00	0.02	1.00	1.00	0.00
LaMA-13B	0.00	0.00	0.02	1.00	1.00	1.00	1.00
LaMA-30B	0.00	0.00	1.00	1.00	1.00	1.00	0.13
LaMA-65B	0.00	0.00	1.00	1.00	1.00	1.00	0.04
LLaMA-7B	0.00	0.00	0.00	1.00	0.13	0.04	1.00