

p-value of Dunn test for γ (order 27)

GPT-3.5	1.00	0.00	0.09	1.00	1.00	1.00	1.00
GPT-4	0.00	1.00	0.00	0.00	0.00	0.00	0.00
Human	0.09	0.00	1.00	1.00	0.62	1.00	0.10
LaMA-13B	1.00	0.00	1.00	1.00	1.00	1.00	1.00
LaMA-30B	1.00	0.00	0.62	1.00	1.00	1.00	1.00
LaMA-65B	1.00	0.00	1.00	1.00	1.00	1.00	1.00
LLaMA-7B	1.00	0.00	0.10	1.00	1.00	1.00	1.00
	PT-3.5	GPT-4	human	A-13B	A-30B	A-65B	MA-7B