

# Inter-model Spearman Correlations

$$P_{M_1}(T^{PO}|P^{PO}) \times P_{M_2}(T^{PO}|P^{PO})$$

$$P_{M_1}(T^{PO}|P^{DO}) \times P_{M_2}(T^{PO}|P^{DO})$$

$$PE_{M_1}(PO) \times P_{M_2}(PO)$$

GPT2-large	1.00	0.84	0.79	0.83	0.70	0.86	0.80	0.79
Falcon-7b	0.84	1.00	0.71	0.75	0.67	0.77	0.76	0.71
Falcon-7b-instruct	0.79	0.71	1.00	0.71	0.61	0.76	0.73	0.69
Llama-2	0.83	0.75	0.71	1.00	0.81	0.83	0.78	0.76
Llama-2-chat	0.70	0.67	0.61	0.81	1.00	0.70	0.71	0.67
Mistral-7b	0.86	0.77	0.76	0.83	0.70	1.00	0.85	0.91
Mistral-7b-instruct	0.80	0.76	0.73	0.78	0.71	0.85	1.00	0.80
Zephyr	0.79	0.71	0.69	0.76	0.67	0.91	0.80	1.00
GPT2-large								
Falcon-7b								
Falcon-7b-instruct								
Llama-2								
Llama-2-chat								
Mistral-7b								
Mistral-7b-instruct								
Zephyr								

GPT2-large	1.00	0.82	0.76	0.82	0.68	0.84	0.79	0.79
Falcon-7b	0.82	1.00	0.69	0.75	0.65	0.75	0.75	0.70
Falcon-7b-instruct	0.76	0.69	1.00	0.72	0.60	0.75	0.72	0.69
Llama-2	0.82	0.75	0.72	1.00	0.82	0.82	0.78	0.76
Llama-2-chat	0.68	0.65	0.60	0.82	1.00	0.68	0.70	0.66
Mistral-7b	0.84	0.75	0.75	0.82	0.68	1.00	0.84	0.91
Mistral-7b-instruct	0.79	0.75	0.72	0.78	0.70	0.84	1.00	0.80
Zephyr	0.79	0.70	0.69	0.76	0.66	0.91	0.80	1.00
GPT2-large								
Falcon-7b								
Falcon-7b-instruct								
Llama-2								
Llama-2-chat								
Mistral-7b								
Mistral-7b-instruct								
Zephyr								

GPT2-large	1.00	0.36	0.24	0.35	0.26	0.36	0.30	0.30
Falcon-7b	0.36	1.00	0.17	0.16	0.17	0.15	0.25	0.14
Falcon-7b-instruct	0.24	0.17	1.00	0.22	0.18	0.19	0.23	0.16
Llama-2	0.35	0.16	0.22	1.00	0.51	0.36	0.27	0.28
Llama-2-chat	0.26	0.17	0.18	0.51	1.00	0.23	0.26	0.19
Mistral-7b	0.36	0.15	0.19	0.36	0.23	1.00	0.40	0.67
Mistral-7b-instruct	0.30	0.25	0.23	0.27	0.26	0.40	1.00	0.34
Zephyr	0.30	0.14	0.16	0.28	0.19	0.67	0.34	1.00
GPT2-large								
Falcon-7b								
Falcon-7b-instruct								
Llama-2								
Llama-2-chat								
Mistral-7b								
Mistral-7b-instruct								
Zephyr								