Winner \ Loser	Gemini 2.5 Pro	o3	GPT-4o v3	DeepSeek R1	Llama 4 Maverick Exp	Gemini 2.5 Flash	Qwen3 235B	DeepSeek V3	o4 Mini	Mistral Medium	Opus 4	Gemma 3	Opus 4 (T)	Gemini 2.0 Flash	GPT-4.1 Mini	Sonnet 4 (T)	Sonnet 4	Qwen3 30B	Llama 4 Maverick	Sonnet 3.5 v2	Sonnet 3.7	Sonnet 3.7 (T)	Haiku 3.5
Gemini 2.5 Pro		0.39	0.53	0.36	0.48	0.65	0.50	0.29	0.84	0.41	0.28	0.28	0.91	0.40	0.46	0.56	0.83	0.68	0.35	0.35	0.25	0.35	0.20
o3			0.47	0.41	0.51	0.12	0.58	0.64	1.03	0.36	0.47	0.68	1.45	0.38	0.87	0.25	0.47	0.80	0.48	0.12	0.27	1.04	0.40
GPT-4o v3				0.48	0.81	0.34	0.30	0.26	0.67	0.42	0.48	0.49	0.29	0.62	0.86	0.76	0.70	0.54	0.48	0.28	0.87	0.80	0.31
DeepSeek R1					1.00	0.42	0.48	0.60	0.53	0.06	0.89	0.92	0.47	1.45	0.33	0.56	0.38	0.56	0.93	0.87	0.21	0.71	0.62
Llama 4 Maverick Exp						0.50	0.97	1.00	0.57	0.73	0.53	0.78	0.54	0.30	0.75	0.62	0.67	0.25	0.80	0.75	1.00	0.44	0.34
Gemini 2.5 Flash							0.46	0.67	0.23	0.30	0.23	0.50	0.48	0.84	0.38	0.18	0.35	0.63	0.23	0.44	0.21	0.28	0.24
Qwen3 235B								0.43	1.00	0.42	0.33	0.67	0.63	0.20	1.18	0.23	0.26	0.30	0.74	0.22	0.35	0.61	0.23
DeepSeek V3									0.36	0.41	0.29	0.55	0.47	0.87	0.42	0.71	0.16	0.48	0.08	0.84	0.56	0.77	0.36
o4 Mini										0.25	0.43	1.17	0.55	0.33	1.13	0.45	0.69	0.92	0.59	1.12	0.75	0.58	0.17
Mistral Medium											0.60	0.57	0.25	0.89	0.30	0.08	0.59	0.94	0.67	0.43	0.31	0.68	0.50
Opus 4												0.63	0.75	0.76	0.59	0.26	0.20	0.26	0.35	0.33	0.17	0.57	0.59
Gemma 3													0.80	0.73	0.53	0.53	0.48	0.23	0.83	0.56	0.68	1.04	0.75
Opus 4 (T)														0.27	0.32	0.65	0.59	0.44	0.40	0.95	0.60	1.10	0.93
Gemini 2.0 Flash															0.82	0.16	0.87	1.00	0.76	1.21	0.61	0.56	1.08
GPT-4.1 Mini																0.58	0.33	0.21	0.76	0.24	0.61	0.16	0.47
Sonnet 4 (T)																	0.52	1.20	0.51	0.71	0.39	0.54	0.77
Sonnet 4																		0.50	0.39	0.40	0.12	0.50	0.23
Qwen3 30B																			0.92	0.53	1.15	0.80	0.44
Llama 4 Maverick																				0.44	0.64	0.63	0.26
Sonnet 3.5 v2																					0.64	1.15	0.33
Sonnet 3.7																						0.79	0.83
Sonnet 3.7 (T)																							0.61
Haiku 3.5																							