

Winner	Llama 4 Maverick Exp																						
	Loser																						
	GPT-4o v3	1 (6)																					
	Gemini 2.5 Pro	7 (11)	6 (10)																				
	o3	3 (11)	7 (14)	7 (11)																			
	DeepSeek R1	1 (6)	5 (8)	5 (8)	3 (10)																		
	Gemini 2.5 Flash	8 (13)	8 (16)	4 (13)	3 (6)	7 (20)																	
	Qwen3 235B	1 (10)	4 (9)	3 (12)	5 (8)	7 (11)	4 (12)																
	Opus 4	4 (12)	8 (17)	4 (22)	7 (18)	8 (14)	13 (30)	10 (19)															
	DeepSeek V3	2 (8)	2 (5)	3 (10)	1 (7)	5 (10)	8 (12)	3 (10)	6 (12)														
	Opus 4 (T)	3 (6)	3 (7)	8 (17)	0 (8)	5 (9)	6 (12)	10 (14)	4 (12)	2 (4)													
	Sonnet 4	4 (16)	4 (10)	10 (15)	4 (9)	4 (8)	6 (21)	8 (16)	7 (18)	5 (15)	4 (6)												
	Mistral Medium	8 (19)	3 (13)	8 (20)	4 (12)	6 (14)	8 (23)	6 (16)	9 (18)	2 (6)	7 (11)	8 (16)											
	Qwen3 30B	3 (6)	2 (4)	7 (15)	4 (6)	2 (11)	4 (11)	5 (13)	11 (18)	1 (3)	5 (9)	7 (20)	6 (7)										
	Gemini 2.0 Flash	1 (4)	3 (13)	3 (7)	5 (9)	0 (5)	3 (5)	2 (5)	4 (14)	1 (5)	3 (5)	3 (13)	3 (7)	6 (6)									
	o4 Mini	1 (4)	4 (12)	2 (10)	1 (10)	4 (8)	5 (13)	1 (7)	4 (11)	4 (8)	2 (8)	4 (5)	4 (8)	2 (5)	6 (10)								
	GPT-4.1 Mini	3 (6)	0 (7)	3 (8)	1 (10)	4 (11)	4 (8)	0 (7)	4 (10)	4 (11)	3 (7)	6 (15)	3 (11)	2 (4)	5 (7)	4 (6)							
	Sonnet 4 (T)	0 (5)	2 (13)	3 (11)	2 (6)	2 (4)	4 (8)	5 (13)	2 (6)	1 (7)	2 (6)	3 (6)	5 (10)	2 (5)	1 (3)	3 (7)	3 (4)						
	Gemma 3	1 (8)	4 (18)	9 (19)	6 (16)	0 (7)	5 (20)	2 (9)	6 (14)	6 (11)	5 (6)	6 (15)	3 (10)	3 (6)	6 (9)	1 (6)	3 (11)	4 (6)					
	Llama 4 Maverick	3 (6)	4 (11)	3 (9)	3 (11)	0 (7)	2 (6)	2 (6)	3 (10)	2 (6)	3 (7)	2 (5)	1 (8)	1 (3)	1 (5)	5 (9)	2 (3)	2 (4)	2 (7)				
	Sonnet 3.7 (T)	2 (10)	1 (10)	4 (17)	5 (8)	1 (10)	3 (17)	2 (15)	6 (14)	3 (5)	3 (4)	9 (17)	3 (15)	5 (13)	3 (7)	2 (7)	4 (9)	4 (7)	2 (9)	5 (7)			
	Sonnet 3.7	0 (12)	4 (15)	4 (15)	4 (13)	2 (11)	6 (16)	5 (13)	3 (14)	3 (12)	2 (4)	3 (8)	4 (18)	3 (10)	2 (12)	2 (8)	4 (11)	5 (12)	6 (12)	3 (7)	9 (12)		
	Sonnet 3.5 v2	1 (8)	3 (14)	1 (7)	4 (12)	3 (5)	5 (13)	2 (6)	4 (19)	2 (4)	0 (6)	2 (9)	5 (10)	3 (5)	2 (14)	4 (11)	5 (8)	2 (8)	4 (10)	3 (7)	0 (7)	6 (14)	
	Haiku 3.5	1 (9)	3 (10)	3 (15)	1 (9)	1 (11)	4 (17)	3 (10)	2 (14)	3 (11)	0 (10)	4 (13)	5 (15)	1 (4)	4 (8)	3 (10)	2 (8)	2 (10)	0 (4)	3 (6)	4 (7)	3 (5)	4 (8)