

Winner

Gemini 2.5 Pro																							
Llama 4 Maverick Exp	11 (20)																						
DeepSeek R1	9 (18)	12 (18)																					
Gemini 2.5 Flash	12 (19)	8 (22)	8 (17)																				
o3	9 (20)	6 (12)	8 (13)	15 (28)																			
Opus 4 (T)	5 (7)	3 (7)	7 (12)	8 (14)	8 (11)																		
GPT-4o v3	8 (12)	2 (8)	3 (10)	6 (12)	7 (18)	8 (16)																	
Opus 4	8 (25)	2 (10)	9 (19)	8 (25)	12 (27)	5 (6)	7 (13)																
DeepSeek V3	2 (8)	6 (8)	6 (14)	7 (21)	6 (13)	5 (13)	2 (6)	11 (17)															
Qwen3 235B	11 (22)	7 (15)	12 (24)	11 (20)	6 (18)	8 (17)	5 (17)	13 (25)	11 (19)														
Gemini 2.0 Flash	4 (17)	2 (8)	3 (7)	2 (11)	4 (11)	6 (10)	7 (13)	9 (18)	3 (8)	4 (12)													
Sonnet 4	7 (28)	7 (23)	8 (21)	11 (24)	4 (19)	4 (10)	8 (13)	16 (42)	7 (17)	14 (34)	5 (10)												
Gemma 3	7 (19)	4 (13)	6 (17)	10 (23)	5 (7)	7 (13)	3 (10)	7 (23)	5 (13)	7 (14)	5 (13)	13 (26)											
Sonnet 4 (T)	7 (13)	2 (7)	2 (9)	6 (14)	4 (15)	1 (7)	3 (7)	6 (13)	2 (4)	5 (10)	2 (6)	2 (3)	3 (10)										
Mistral Medium	9 (28)	5 (14)	9 (22)	8 (23)	11 (23)	7 (12)	8 (22)	8 (27)	7 (13)	14 (30)	5 (14)	8 (21)	8 (16)	7 (13)									
GPT-4.1 Mini	6 (13)	2 (10)	5 (10)	5 (16)	8 (15)	4 (11)	5 (13)	5 (21)	4 (7)	13 (22)	8 (11)	12 (29)	6 (13)	7 (14)	4 (12)								
o4 Mini	1 (11)	3 (7)	7 (19)	3 (6)	8 (13)	5 (12)	5 (20)	11 (26)	7 (14)	9 (14)	5 (10)	5 (11)	2 (8)	3 (8)	10 (18)	6 (9)							
Qwen3 30B	4 (15)	3 (8)	6 (13)	6 (21)	4 (12)	2 (6)	3 (12)	13 (22)	3 (8)	5 (10)	4 (9)	6 (12)	7 (15)	0 (5)	4 (8)	3 (7)	4 (8)						
Sonnet 3.7	5 (13)	7 (14)	4 (20)	6 (18)	5 (14)	2 (10)	5 (17)	4 (17)	2 (12)	4 (11)	3 (8)	7 (16)	4 (7)	4 (8)	5 (20)	7 (13)	7 (14)	7 (8)					
Sonnet 3.5 v2	1 (11)	2 (4)	2 (12)	5 (10)	4 (14)	2 (9)	6 (17)	6 (16)	2 (13)	2 (9)	7 (19)	3 (17)	5 (11)	1 (5)	4 (14)	7 (12)	10 (12)	2 (6)	9 (12)				
Llama 4 Maverick	6 (16)	1 (5)	2 (8)	5 (17)	5 (15)	2 (8)	5 (14)	9 (19)	3 (10)	0 (4)	2 (8)	4 (12)	1 (4)	3 (5)	6 (13)	4 (9)	4 (6)	0 (4)	3 (9)	2 (3)			
Sonnet 3.7 (T)	5 (23)	3 (11)	6 (23)	5 (22)	2 (19)	2 (10)	5 (10)	6 (32)	3 (10)	6 (23)	6 (13)	7 (25)	8 (13)	3 (10)	13 (24)	3 (7)	6 (18)	2 (10)	7 (15)	10 (15)	5 (8)		
Haiku 3.5	2 (12)	3 (12)	4 (10)	3 (14)	4 (15)	2 (6)	2 (12)	7 (26)	2 (9)	4 (16)	1 (8)	5 (17)	3 (8)	4 (9)	5 (11)	2 (11)	5 (13)	3 (13)	6 (9)	3 (8)	5 (10)	4 (15)	
<div><div>Gemini 2.5 Pro</div><div>Llama 4 Maverick Exp</div><div>DeepSeek R1</div><div>Gemini 2.5 Flash</div><div>o3</div><div>Opus 4 (T)</div><div>GPT-4o v3</div><div>Opus 4</div><div>DeepSeek V3</div><div>Qwen3 235B</div><div>Gemini 2.0 Flash</div><div>Sonnet 4</div><div>Gemma 3</div><div>Sonnet 4 (T)</div><div>Mistral Medium</div><div>GPT-4.1 Mini</div><div>o4 Mini</div><div>Qwen3 30B</div><div>Sonnet 3.7</div><div>Sonnet 3.5 v2</div><div>Llama 4 Maverick</div><div>Sonnet 3.7 (T)</div><div>Haiku 3.5</div></div>																							

Loser