

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 (T)	Haiku 3.5
Gemini 2.5 Pro	89 (192)	73 (156)	88 (175)	71 (150)	63 (135)	54 (104)	71 (168)	69 (138)	42 (86)	54 (125)	96 (229)	95 (210)	85 (182)	101 (203)	87 (190)	85 (233)	87 (194)	57 (158)	89 (194)	84 (170)	124 (233)	
GPT-4o v3	73 (156)	88 (175)	71 (150)	63 (135)	54 (104)	71 (168)	69 (138)	42 (86)	54 (125)	96 (229)	95 (210)	85 (182)	101 (203)	87 (190)	85 (233)	87 (194)	57 (158)	89 (194)	84 (170)	124 (233)		
DeepSeek R1	71 (150)	63 (135)	54 (104)	71 (168)	69 (138)	42 (86)	54 (125)	96 (229)	95 (210)	85 (182)	101 (203)	87 (190)	85 (233)	87 (194)	57 (158)	89 (194)	84 (170)	124 (233)	85 (233)	87 (194)	57 (158)	89 (194)
Llama 4 Maverick Exp	71 (168)	69 (138)	42 (86)	54 (125)	96 (229)	95 (210)	85 (182)	101 (203)	87 (190)	85 (233)	87 (194)	57 (158)	89 (194)	84 (170)	124 (233)	85 (233)	87 (194)	57 (158)	89 (194)	84 (170)	124 (233)	
Gemini 2.5 Flash	96 (229)	95 (210)	85 (182)	101 (203)	87 (190)	85 (233)	87 (194)	57 (158)	89 (194)	84 (170)	124 (233)	85 (233)	87 (194)	57 (158)	89 (194)	84 (170)	124 (233)	85 (233)	87 (194)	57 (158)	89 (194)	84 (170)
Qwen3 235B	85 (233)	87 (194)	57 (158)	89 (194)	84 (170)	124 (233)	85 (233)	87 (194)	57 (158)	89 (194)	84 (170)	124 (233)	85 (233)	87 (194)	57 (158)	89 (194)	84 (170)	124 (233)	85 (233)	87 (194)	57 (158)	89 (194)
DeepSeek V3	55 (153)	59 (143)	55 (131)	41 (98)	51 (88)	83 (184)	85 (177)	85 (177)	85 (177)	85 (177)	85 (177)	85 (177)	85 (177)	85 (177)	85 (177)	85 (177)	85 (177)	85 (177)	85 (177)	85 (177)	85 (177)	
o4 Mini	73 (164)	63 (162)	86 (166)	42 (134)	45 (93)	64 (154)	77 (161)	57 (116)	77 (161)	57 (116)	77 (161)	57 (116)	77 (161)	57 (116)	77 (161)	57 (116)	77 (161)	57 (116)	77 (161)	57 (116)	77 (161)	
Mistral Medium	83 (258)	88 (229)	71 (190)	76 (181)	91 (190)	128 (275)	126 (249)	76 (147)	126 (249)	76 (147)	126 (249)	76 (147)	126 (249)	76 (147)	126 (249)	76 (147)	126 (249)	76 (147)	126 (249)	76 (147)	126 (249)	
Opus 4	78 (265)	91 (220)	81 (191)	82 (191)	88 (200)	130 (309)	120 (271)	100 (211)	94 (200)	119 (256)	119 (256)	119 (256)	119 (256)	119 (256)	119 (256)	119 (256)	119 (256)	119 (256)	119 (256)	119 (256)	119 (256)	
Gemma 3	72 (198)	70 (167)	58 (163)	63 (151)	43 (98)	82 (201)	88 (166)	52 (123)	58 (134)	90 (178)	116 (221)	116 (221)	116 (221)	116 (221)	116 (221)	116 (221)	116 (221)	116 (221)	116 (221)	116 (221)	116 (221)	
Opus 4 (T)	51 (152)	49 (154)	52 (118)	54 (123)	32 (93)	68 (149)	90 (206)	45 (107)	62 (114)	76 (172)	72 (146)	54 (125)	54 (125)	54 (125)	54 (125)	54 (125)	54 (125)	54 (125)	54 (125)	54 (125)	54 (125)	
Gemini 2.0 Flash	34 (122)	70 (176)	72 (174)	26 (59)	27 (76)	49 (120)	44 (113)	43 (107)	63 (144)	52 (116)	77 (152)	68 (144)	44 (86)	44 (86)	44 (86)	44 (86)	44 (86)	44 (86)	44 (86)	44 (86)	44 (86)	
GPT-4.1 Mini	57 (154)	55 (133)	50 (130)	48 (135)	32 (82)	53 (155)	79 (153)	49 (117)	49 (108)	73 (188)	78 (189)	71 (137)	49 (98)	49 (87)	49 (87)	49 (87)	49 (87)	49 (87)	49 (87)	49 (87)	49 (87)	
Sonnet 4 (T)	49 (168)	48 (147)	40 (93)	48 (115)	24 (87)	62 (140)	68 (166)	34 (89)	51 (108)	78 (167)	46 (106)	51 (111)	55 (134)	22 (59)	46 (87)	46 (87)	46 (87)	46 (87)	46 (87)	46 (87)	46 (87)	
Sonnet 4	89 (224)	57 (183)	63 (151)	60 (154)	55 (169)	93 (259)	88 (227)	62 (167)	62 (132)	86 (226)	105 (253)	66 (158)	48 (116)	57 (109)	83 (160)	55 (99)	55 (99)	55 (99)	55 (99)	55 (99)	55 (99)	
Qwen3 30B	52 (160)	45 (117)	30 (121)	37 (101)	35 (81)	57 (160)	77 (171)	39 (88)	34 (94)	58 (132)	89 (175)	41 (99)	44 (94)	40 (89)	36 (86)	39 (93)	71 (141)	71 (141)	71 (141)	71 (141)	71 (141)	
Llama 4 Maverick	39 (137)	53 (154)	49 (151)	29 (75)	34 (102)	48 (131)	38 (112)	32 (97)	49 (118)	52 (127)	50 (120)	48 (111)	34 (85)	40 (113)	45 (109)	39 (89)	47 (106)	30 (70)	30 (70)	30 (70)	30 (70)	
Sonnet 3.5 v2	37 (159)	49 (163)	65 (194)	29 (91)	37 (109)	53 (172)	41 (136)	52 (116)	46 (133)	54 (138)	63 (181)	69 (151)	36 (99)	130 (268)	59 (112)	40 (94)	42 (126)	50 (107)	55 (103)	55 (103)	55 (103)	
Sonnet 3.7	66 (198)	59 (173)	65 (168)	30 (128)	33 (106)	76 (206)	44 (151)	35 (101)	59 (153)	64 (197)	50 (178)	65 (151)	32 (110)	66 (150)	54 (126)	42 (109)	62 (154)	50 (102)	52 (107)	71 (145)	71 (145)	
Sonnet 3.7 (T)	50 (216)	54 (170)	37 (119)	49 (159)	48 (125)	66 (230)	81 (219)	32 (135)	49 (122)	82 (211)	82 (233)	69 (145)	40 (113)	49 (94)	50 (116)	41 (102)	75 (198)	37 (112)	35 (77)	50 (99)	75 (137)	75 (137)
Haiku 3.5	37 (165)	43 (152)	40 (168)	29 (119)	32 (127)	50 (163)	55 (178)	42 (113)	45 (140)	54 (172)	63 (191)	45 (141)	41 (109)	42 (121)	31 (116)	51 (116)	56 (147)	30 (77)	42 (107)	61 (138)	49 (102)	53 (127)