

Model Index	Model Ref	Model Name	Model Name	Version	Context Length	Min	Sec	Latency (s / record) [1]	Input Token Price (\$ / 1K tokens)	Input Tokens	Output Token Price (\$ / 1K tokens)	Output Tokens	Total Price (\$ = Input + Output)	Avg Price (\$ / record)	Num Records	TP	TN	FP	FN	AP	AN	Precision	Recall	F1	Num Manual Clean
00	2.sonnet3.5v2_v213_prompt_eval.csv	anthropic.clau	Claude 3.5 Sonnet	v213 prompt	200k	15	26	0.2869	\$0.0030	4,094,894	\$0.0150	126,942	\$14.19	\$0.01	1084	117	891	44	32	149	935	0.7267	0.7852	0.7548	0
01	5.sonnet3.5v2_v217_prompt_eval.csv	anthropic.clau	Claude 3.5 Sonnet	v217 prompt	200k	14	42	0.2841	\$0.0030	3,970,234	\$0.0150	129,905	\$13.86	\$0.01	1084	113	900	35	36	149	935	0.7635	0.7584	0.7609	1
02	5.gpt4o_v217_prompt_eval.csv	gpt-4o-2024-11-20	GPT 4o	v217 prompt	128k	11	43	0.2325	\$0.0025	3,626,535	\$0.0100	112,956	\$10.20	\$0.01	1084	117	900	35	32	149	935	0.7697	0.7852	0.7774	0
1	0.sonnet3.5v2_v226_prompt_eval.csv	anthropic.clau	Claude 3.5 Sonnet	v226 prompt	200k	20	52	0.3885	\$0.0030	6,360,115	\$0.0150	168,737	\$21.67	\$0.02	1084	136	862	53	13	149	935	0.7196	0.9128	0.8047	0
2	1.sonnet3.5v2_v226_prompt_eval.csv	anthropic.clau	Claude 3.5 Sonnet	v226 v2 prompt	200k	19	7	0.3395	\$0.0030	6,642,443	\$0.0150	168,473	\$22.45	\$0.02	1084	139	895	49	11	149	935	0.7380	0.9262	0.8214	0
3	2.sonnet3.7_v226_v2_prompt_eval.csv	us.anthropic.clau	Claude 3.7 Sonnet	v226 v2 prompt	200k	22	34	0.4170	\$0.0030	6,642,443	\$0.0150	205,596	\$23.01	\$0.02	1084	136	869	66	14	149	935	0.6716	0.9060	0.7714	4
4	0.llama3.1_405b_v226_prompt_eval.csv	meta.llama3-1-405b-instruct-v1.0	Llama 3.1 405B Instruct	v226 prompt	128k	42	24	0.7583	\$0.0024	5,774,470	\$0.0024	153,201	\$14.23	\$0.01	1084	131	853	82	18	149	935	0.6150	0.8792	0.7238	1
5	1.llama3.1_405b_v226_v2_prompt_eval.csv	meta.llama3-1-405b-instruct-v1.0	Llama 3.1 405B Instruct	v226 v2 prompt	128k	43	19	0.7712	\$0.0024	6,023,790	\$0.0024	153,099	\$14.82	\$0.01	1084	131	856	79	18	149	935	0.6238	0.8792	0.7298	0
6	2.llama3.3_70B_v226_prompt_eval.csv	meta.llama3-3-70b-instruct-v1.0	Llama 3.3 70B Instruct	v226 prompt	128k	9	43	0.1974	\$0.0007	5,798,318	\$0.0007	160,331	\$4.29	\$0.00	1084	131	861	74	18	149	935	0.6390	0.8792	0.7401	4
7	3.llama3.3_70B_v226_v2_prompt_eval.csv	meta.llama3-3-70b-instruct-v1.0	Llama 3.3 70B Instruct	v226 v2 prompt	128k	9	45	0.1993	\$0.0007	6,047,638	\$0.0007	166,803	\$4.47	\$0.00	1084	133	857	78	16	149	935	0.6303	0.8926	0.7389	7
8	0.mistral_large_v226_prompt_eval.csv	mistral.mistral-large-2407-v1.0	Mistral Large 2 (24.07)	v226 prompt	131k	21	85	0.4465	\$0.0020	6,756,808	\$0.0060	171,521	\$14.54	\$0.01	1084	114	906	29	35	149	935	0.7972	0.7651	0.7808	4
9	1.mistral_large_v226_v2_prompt_eval.csv	mistral.mistral-large-2407-v1.0	Mistral Large 2 (24.07)	v226 v2 prompt	131k	21	157	0.5129	\$0.0020	7,039,732	\$0.0060	170,919	\$15.10	\$0.01	1084	117	906	29	32	149	935	0.8014	0.7852	0.7932	5
10	0.gpt4o_v226_prompt_eval.csv	gpt-4o-2024-11-20	GPT 4o	v226 prompt	128k	12	1	0.2113	\$0.0025	5,812,042	\$0.0100	144,849	\$15.98	\$0.01	1084	129	881	54	20	149	935	0.7049	0.8658	0.7771	0
11	1.gpt4o_v226_v2_prompt_eval.csv	gpt-4o-2024-11-20	GPT 4o	v226 v2 prompt	128k	14	12	0.2565	\$0.0025	6,050,522	\$0.0100	152,920	\$16.66	\$0.02	1084	130	877	58	19	149	935	0.6915	0.8725	0.7715	0

[1] ThreadPool on 8 Core Apple M2 with 5 max workers