

Performance and Cost Comparison

Test mit Amazon Bestellbestätigung ( <a href="https://openrouter.ai/chat?room=orc-1739464980-tu4B3mZZpIbwFIIXb8Yv">https://openrouter.ai/chat?room=orc-1739464980-tu4B3mZZpIbwFIIXb8Yv</a> )									
Model	input	output	tps	s	cost	quality	first token latency	total time	Kosten für 1000 Mails
Gemini Flash 2.0	18.225	192	118	1,627118644	0,00188	2	1,13	2,757118644	\$1,88
OpenAi o3 mini high	15.157	1.748	2.752,80	0,6349898285	0,0241	3	17,96	18,59498983	\$24,10
OpenAi gpt-4o	15.158	171	56,4	3,031914894	0,0392	3	1,33	4,361914894	\$39,20
Claude 3.5 sonnet	19.193	383	35,4	10,81920904	0,0627	4	3,6	14,41920904	\$62,70
OpenAi gpt-4o-mini	15.158	175	27,4	6,386861314	0,00235	1	0,19	6,576861314	\$2,35
DeepSeek R1	16.158	902	135,1	6,676535899	0,118	3	21,39	28,0665359	\$118,00
Test mit unserösem DM Gewinnspiel ( <a href="https://openrouter.ai/chat?room=orc-1739466855-WoAHfpFm3L2NxBSISeg">https://openrouter.ai/chat?room=orc-1739466855-WoAHfpFm3L2NxBSISeg</a> )									
Model	input	output	tps	s	cost	quality	first token latency	total time	Kosten für 1000 Mails
Gemini Flash 2.0	5.131	614	99,9	6,146146146	0,000751	3	0,68	6,826146146	\$0,75
OpenAi o3 mini high	4.729	2.994	1.481,40	2,021061158	0,0157	2	82,67	84,69106116	\$15,70
OpenAi gpt-4o	4.730	513	56,4	9,095744681	0,0112	2	0,47	9,565744681	\$11,20
Claude 3.5 sonnet	5.784	585	32,6	17,94478528	0,0259	3	2,12	20,06478528	\$25,90
OpenAi gpt-4o-mini	4.730	378	42,5	8,894117647	0,000591	1	0,56	9,454117647	\$0,59
DeepSeek R1	4.968	1.139	38,1	29,89501312	0,00664	4	36,34	66,23501312	\$6,64
Test mit unseriöser DPD Paketnachricht ( <a href="https://openrouter.ai/chat?room=orc-1739531615-giGFL67AEJ0Kv5Yg3N0C">https://openrouter.ai/chat?room=orc-1739531615-giGFL67AEJ0Kv5Yg3N0C</a> )									
Model	input	output	tps	s	cost	quality	first token latency	total time	Kosten für 1000 Mails
Gemini Flash 2.0	5483	552	109,5	5,04109589	0,000761	3	1,04	6,08109589	\$0,76
OpenAi o3 mini high	5176	3923	143,5	27,33797909	0,0227	2	23,94	51,27797909	\$22,70
OpenAi gpt-4o	5177	614	26,5	23,16981132	0,0189	2	0,6	23,76981132	\$18,90
Claude 3.5 sonnet	6264	601	45,5	13,20879121	0,0275	3	2,39	15,59879121	\$27,50
OpenAi gpt-4o-mini	5177	410	37,3	10,9919571	0,00101	1	6,12	17,1119571	\$1,01
DeepSeek R1	5412	1142	77,4	14,75452196	0,0454	3	21,33	36,08452196	\$45,40
Test mit seriösem TUM Newsletter ( <a href="https://openrouter.ai/chat?room=orc-1739533912-vwzFyPnZDvpy6HUmSN1r">https://openrouter.ai/chat?room=orc-1739533912-vwzFyPnZDvpy6HUmSN1r</a> , <a href="https://openrouter.ai/chat?room=orc-1739534460-ZubSCAv9KbivvEJRjc25">https://openrouter.ai/chat?room=orc-1739534460-ZubSCAv9KbivvEJRjc25</a> )									
Model	input	output	tps	s	cost	quality	first token latency	total time	Kosten für 1000 Mails
Gemini Flash 2.0	1959	472	113,8	4,147627417	0,000571	2	1,48	5,627627417	\$0,57
OpenAi o3 mini high	1.926	4725	125,3	37,70949721	0,0227	3	33,82	71,52949721	\$22,70
OpenAi gpt-4o	1927	398	56,7	7,019400353	0,00871	4	0,41	7,429400353	\$8,71
Claude 3.5 sonnet	2542	544	39,4	13,8071066	0,0156	4	1,74	15,5471066	\$15,60
OpenAi gpt-4o-mini	1927	341	73,8	4,620596206	0,000489	1	0,53	5,150596206	\$0,49
DeepSeek R1	2139	1074	21	51,14285714	0,00425	2	102,29	153,4328571	\$4,25
Durchschnitt aus den vorhergehenden Tests									
Model	input	output	tps	s	cost	quality	first token latency	total time	Kosten für 1000 Mails
Gemini Flash 2.0	7.700	457,5	110,3	4,240497024	0,00099075	2,5	1,0825	5,322997024	\$0,99
OpenAi o3 mini high	6.747	3.348	1.125,75	16,92588182	0,0213	2,5	39,5975	56,52338182	\$21,30
OpenAi gpt-4o	6.748	424	49	10,57921781	0,0195025	2,75	0,7025	11,28171781	\$19,50
Claude 3.5 sonnet	8.446	528,25	38,225	13,94497303	0,032925	3,5	2,4625	16,40747303	\$32,93
OpenAi gpt-4o-mini	6.748	326	45,25	7,723383068	0,00111	1	1,85	9,573383068	\$1,11
DeepSeek R1	7.169	1064,25	67,9	25,61723203	0,0435725	3	45,3375	70,95473203	\$43,57
Auswertung									
Model	quality/\$	quality/s	quality						
Gemini Flash 2.0	2523,340903	0,5895535324	2,5						
OpenAi o3 mini high	117,370892	0,1477027919	2,5						
OpenAi gpt-4o	141,0075631	0,2599436035	2,75						
Claude 3.5 sonnet	106,302202	0,2509865019	3,5						
OpenAi gpt-4o-mini	900,9009009	0,1294769392	1						
DeepSeek R1	68,85076596	0,1171086711	3						