

LLM	Project	Total Cost (TC)	Number of Tests (T)	Failed Tests	Passed Tests	Success Rate (SR)	Coverage Branch (C)	Mutation Score (M)	Calculated Score (S)
claude	todo-api	US\$ 0,51	49	0	49	100,0%	85,7%	48,7%	78,1%
gpt	todo-api	US\$ 0,28	32	0	32	100,0%	85,7%	44,8%	76,8%
maritaca	todo-api	US\$ 0,11	45	0	45	100,0%	80,4%	47,2%	75,8%
llama	todo-api	US\$ 0,02	44	3	41	93,2%	85,7%	50,2%	76,4%
mistral	todo-api	US\$ 0,30	40	4	36	90,0%	82,1%	45,2%	72,5%
gemini	todo-api	US\$ 0,21	53	6	47	88,7%	85,7%	21,4%	65,3%
deepseek	todo-api	US\$ 0,95	34	1	33	97,1%	78,6%	43,5%	73,0%
qwen	todo-api	US\$ 0,12	47	0	47	100,0%	73,2%	47,8%	73,7%
claude	restaurants-api	US\$ 0,60	53	0	53	100,0%	75,8%	32,2%	69,3%
gpt	restaurants-api	US\$ 0,34	31	0	31	100,0%	57,8%	23,9%	60,6%
maritaca	restaurants-api	US\$ 0,12	27	4	23	85,2%	45,3%	19,3%	49,9%
llama	restaurants-api	US\$ 0,02	37	0	37	100,0%	64,1%	21,9%	62,0%
mistral	restaurants-api	US\$ 0,44	57	1	56	98,2%	64,1%	24,8%	62,4%
gemini	restaurants-api	US\$ 0,27	59	0	59	100,0%	64,1%	26,2%	63,4%
deepseek	restaurants-api	US\$ 0,97	57	0	57	100,0%	75,0%	31,4%	68,8%
qwen	restaurants-api	US\$ 0,11	44	0	44	100,0%	68,8%	25,4%	64,7%
claude	shortener-api	US\$ 0,25	13	0	13	100,0%	69,0%	43,1%	70,7%
gpt	shortener-api	US\$ 0,16	11	0	11	100,0%	65,5%	54,6%	73,3%
maritaca	shortener-api	US\$ 0,05	8	0	8	100,0%	65,5%	35,7%	67,1%
llama	shortener-api	US\$ 0,01	15	0	15	100,0%	66,7%	9,8%	58,8%
mistral	shortener-api	US\$ 0,17	13	0	13	100,0%	66,7%	9,8%	58,8%
gemini	shortener-api	US\$ 0,09	15	1	14	93,3%	66,7%	9,8%	56,6%
deepseek	shortener-api	US\$ 0,53	20	3	17	85,0%	64,3%	38,7%	62,7%
qwen	shortener-api	US\$ 0,05	15	2	13	86,7%	67,9%	40,7%	65,1%
claude	books-api	US\$ 0,30	25	0	25	100,0%	84,6%	67,0%	83,9%
gpt	books-api	US\$ 0,17	17	0	17	100,0%	84,6%	37,5%	74,0%
maritaca	books-api	US\$ 0,05	11	0	11	100,0%	84,6%	58,0%	80,9%
llama	books-api	US\$ 0,01	32	0	32	100,0%	80,8%	38,4%	73,1%
mistral	books-api	US\$ 0,18	17	0	17	100,0%	73,1%	32,1%	68,4%
gemini	books-api	US\$ 0,12	32	0	32	100,0%	84,6%	39,3%	74,6%
deepseek	books-api	US\$ 0,52	28	0	28	100,0%	84,6%	60,7%	81,8%
qwen	books-api	US\$ 0,05	24	0	24	100,0%	84,6%	39,3%	74,6%
claude	hotels-api	US\$ 0,73	58	0	58	100,0%	44,0%	17,9%	54,0%
gpt	hotels-api	US\$ 0,31	17	2	15	88,2%	17,9%	8,3%	38,1%
maritaca	hotels-api	US\$ 0,11	23	0	23	100,0%	42,9%	16,8%	53,2%
llama	hotels-api	US\$ 0,02	39	0	39	100,0%	38,1%	15,9%	51,3%
mistral	hotels-api	US\$ 0,49	89	0	89	100,0%	44,0%	12,3%	52,1%
gemini	hotels-api	US\$ 0,27	62	2	60	96,8%	44,0%	7,5%	49,4%
deepseek	hotels-api	US\$ 0,99	25	0	25	100,0%	33,3%	10,5%	47,9%
qwen	hotels-api	US\$ 0,11	31	1	30	96,8%	42,9%	10,6%	50,1%
claude	supermarket-api	US\$ 0,44	32	0	32	100,0%	71,2%	36,2%	69,1%
gpt	supermarket-api	US\$ 0,24	23	0	23	100,0%	44,2%	28,6%	57,6%
maritaca	supermarket-api	US\$ 0,07	16	0	16	100,0%	67,3%	31,1%	66,1%
llama	supermarket-api	US\$ 0,02	50	4	46	92,0%	61,5%	31,6%	61,7%
mistral	supermarket-api	US\$ 0,31	44	0	44	100,0%	48,1%	29,1%	59,1%
gemini	supermarket-api	US\$ 0,16	36	0	36	100,0%	75,0%	31,6%	68,9%
deepseek	supermarket-api	US\$ 0,72	38	0	38	100,0%	69,2%	36,7%	68,7%
qwen	supermarket-api	US\$ 0,09	47	5	42	89,4%	75,0%	35,7%	66,7%

$S = w \cdot SuccessRate + w \cdot Coverage + w \cdot MutationScore$

LLM	AVG					
	S	SR	C	M	T	TC
claude	70,85%	100,00%	71,73%	40,83%	38,3	US\$ 0,47
deepseek	67,15%	97,01%	67,50%	36,93%	33,7	US\$ 0,78
qwen	65,81%	95,47%	68,71%	33,26%	34,7	US\$ 0,09
maritaca	65,51%	97,53%	64,32%	34,69%	21,7	US\$ 0,08
llama	63,88%	97,53%	66,14%	27,97%	36,2	US\$ 0,02
gpt	63,42%	98,04%	59,28%	32,94%	21,8	US\$ 0,25
gemini	63,04%	96,46%	70,01%	22,64%	42,8	US\$ 0,19
mistral	62,20%	98,04%	63,01%	25,55%	43,3	US\$ 0,31

	S	SR	C	M
claude	claude	claude	claude	claude
deepseek	mistral	gemini	deepseek	deepseek
qwen	gpt	qwen	maritaca	
maritaca	maritaca	deepseek	qwen	
llama	llama	llama	gpt	
gpt	deepseek	maritaca	llama	
gemini	gemini	mistral	mistral	
mistral	qwen	gpt	gemini	