

Variant	BLEU	ROUGE-L	BERTScore F1	Human accuracy (1-5)	Human relevance (1-5)	Safety pass rate	Avg. latency	Cost per 1K queries
---------	------	---------	--------------	----------------------	-----------------------	------------------	--------------	---------------------

Prompt v1	0.28	0.41	0.87	3.8	4.1	98%	1.2s	\$4.50
-----------	------	------	------	-----	-----	-----	------	--------

Prompt v2	0.31	0.45	0.89	4.2	4.3	99%	1.3s	\$4.80
-----------	------	------	------	-----	-----	-----	------	--------

Prompt v3	0.25	0.38	0.91	4.5	4.4	100%	1.8s	\$6.20
-----------	------	------	------	-----	-----	------	------	--------