	mini sh	5	1		mini	ă.	mini	Lite	nseek	Llamaric	, k	cpT i	1 ama cB	aude Ki	Claude	e ^t	ndom	
Query Type \ Model	1.5-F182	GPT-A.	GPT-A.	0,7	Genzinash 2.0-Flash	04-mini	Gemini 2.0-Flash	GPT-40	Deer 6711	Llama 4-Mayeric	03-mini	40-min2	1.12m/05B	3.5-Haz.	3.1-5011	Average	Guess	
Revisit Reasons Behind Preference Updates	0.77	0.76	0.84	0.75	0.79	0.75	0.77	0.77	0.83	0.76	0.72	0.70	0.41	0.64	0.57	0.72	0.25	1.0
Tracking Full Preference Evolution	0.65	0.68	0.67	0.67	0.70	0.73	0.68	0.68	0.68	0.66	0.54	0.60	0.38	0.55	0.45	0.62	0.25	0.8
Recall User Shared Facts	0.54	0.61	0.65	0.50	0.50	0.42	0.49	0.41	0.43	0.37	0.47	0.55	0.38	0.29	0.25	0.46	0.25	٠,
Acknowledge Latest User Preference	0.59	0.55	0.50	0.54	0.52	0.55	0.51	0.46	0.42	0.43	0.39	0.34	0.31	0.27	0.09	0.43	0.25	0.6
Provide Preference Aligned Recommendations	0.55	0.44	0.57	0.42	0.51	0.41	0.52	0.37	0.49	0.42	0.41	0.41	0.37	0.32	0.20	0.43	0.25	0.4
Generalize Reasons to New Scenarios	0.54	0.46	0.53	0.39	0.46	0.38	0.33	0.32	0.38	0.32	0.30	0.33	0.21	0.20	0.29	0.36	0.25	
Suggest New Ideas	0.15	0.27	0.19	0.25	0.15	0.17	0.16	0.24	0.16	0.20	0.11	0.10	0.20	0.06	0.28	0.18	0.25	0.2
Overall Accuracy	0.52	0.52	0.52	0.50	0.49	0.48	0.48	0.45	0.45	0.43	0.39	0.39	0.31	0.30	0.26	0.43	0.25	0.0