

Table 1 Test Cohort I: eight methods on 500 HotpotQA and 500 TriviaQA questions. EM and F1 are percentages.

Method	H EM	H F1	T EM	T F1	All EM	All F1
Direct	14.40	21.13	30.40	38.63	22.40	29.88
Always-RAG	24.40	33.59	35.20	41.38	29.80	37.48
Always-Multihop	26.80	35.92	35.20	41.73	31.00	38.83
Matched Binary	21.40	30.65	35.60	43.06	28.50	36.85
TARG-style	22.40	32.02	37.20	44.40	29.80	38.21
SUGAR-L-style	23.60	33.49	39.40	44.84	31.50	39.16
CAGR-R	22.00	31.60	35.40	42.80	28.70	37.20
CAGR-R (answer expansion)	21.40	30.72	35.80	43.13	28.60	36.93

Method	Latency (ms)	Retr./q
Direct	209.43	0.000
Always-RAG	618.33	1.000
Always-Multihop	1916.15	2.000
Matched Binary	473.82	0.410
TARG-style	549.46	0.555
SUGAR-L-style	2460.22	1.240
CAGR-R	592.31	0.501
CAGR-R (answer expansion)	516.69	0.501