

Table 3 Routing overhead for CAGR-R in Test Cohort I and CAGR-F in Test Cohort II; N=500 questions per row. U95 is a one-sided 95% upper bound.

Cohort	Dataset	Component mean (%)	Component U95 (%)	On/off mean (%)	On/off U95 (%)
I	HotpotQA	0.14551	0.15107	0.16032	0.29495
I	TriviaQA	0.17365	0.17899	0.31343	0.48198
II	HotpotQA	0.12514	0.12847	0.28488	0.42126
II	TriviaQA	0.13119	0.13506	0.07234	0.20152