

*Straightforward
RAG (9-13%)*

*SOTA Industry
RAG (51%)*
*Manual Eval
*Different retrieval pool

*LLM-Only
(3-20%)*

*CRAG KDD
Cups (36%)*
Manual Eval

Truthfulness

