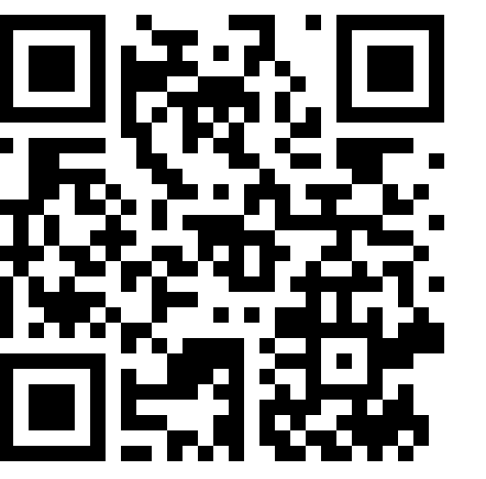


Disentangling Questions from Query Generation for Task-Adaptive Retrieval



↑↑ [Link to paper](#)



Yoonsang Lee, Minsoo Kim, Seung-won Hwang
Seoul National University

RQ: How to generate effective synthetic queries for zero-shot retrieval?

- ✗ Simple question form
- ✗ Few-shot learning



We incorporate **task-specific search intents!**

Efficient Generalized Generator

Dataset	Task	Intent	Query description
Fever	Fact Checking	Transactional	Claim
Arguana	Argument Retrieval	Transactional	Argument
Scidocs	Citation Prediction	Navigational	Title
DBPedia	Entity Retrieval	Mixed	Entity

1. EGG-FLAN (3B)

Meta-prompt Write a {query description} related to the topic of the passage.
Diversity Do not directly use wordings from the passage.

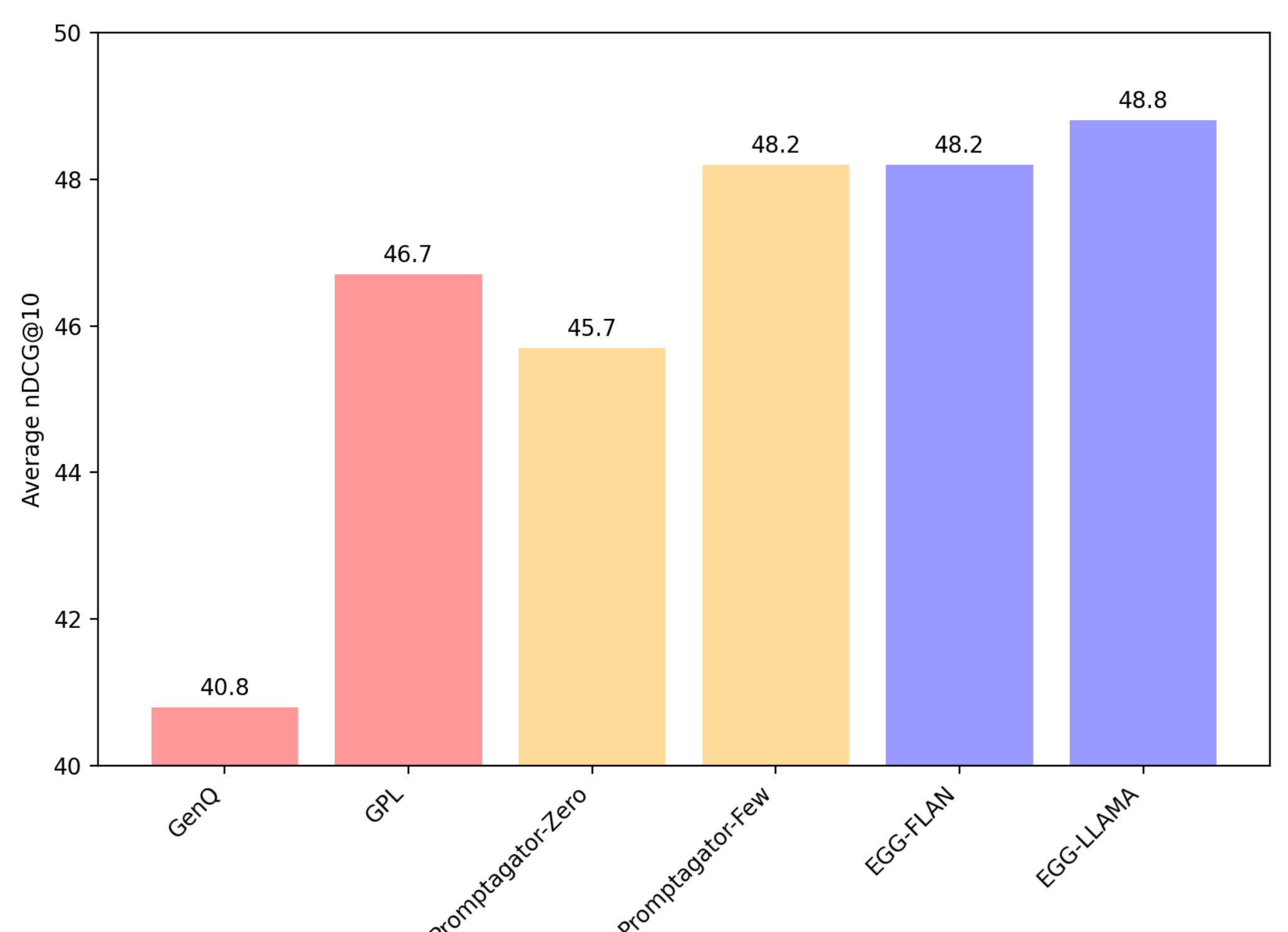
2. EGG-LLAMA (7B)

Prototype query Read the passage and generate a {query description}.
In-context learning Passage: {passage} Query: {prototype query}

EGG outperforms baselines

	Fev.	Arg.	Sci.	Dbp.	Avg.	
DPR	FLAN (zero)	54.4	53.9	14.8	30.4	38.4
	FLAN (few)	38.1	22.9	16.7	31.5	27.3
	EGG-FLAN	69.5	60.1	18.6	33.6	45.5
	Llama2 (zero)	60.9	59.0	16.0	34.2	42.5
	Llama2 (few)	67.2	62.0	17.5	30.2	44.2
	EGG-LLAMA	67.6	61.2	18.2	32.5	44.9
GPL	FLAN (zero)	73.2	55.8	14.9	39.6	45.9
	FLAN (few)	77.0	3.5	16.3	37.7	33.6
	EGG-FLAN	79.4	58.7	16.9	40.0	48.8
	Llama2 (zero)	73.0	56.4	15.0	40.2	46.2
	Llama2 (few)	79.8	55.8	17.0	39.1	47.9
	EGG-LLAMA	78.3	57.1	17.0	40.2	48.2

EGG outperforms prior works



Check out more analysis and ablations in the paper!