Institute for Infocomm Research
I²R

EMNLP 2023

Instructive Dialogue Summarization with Query Aggregations

Bin Wang, Zhengyuan Liu, Nancy F. Chen

Aural & Language Intelligence

Institute for Infocomm Research, A*STAR

Dec. 2023

ARES PUBLIC

CREATING GROWTH, ENHANCING LIVES

What is going on for Dialogue Summarization?

Summarize the key information from the dialogue has wide applications

- Daily chat, Task-oriented dialogue
- Meeting summarization
- Social media
- Customer support
- Interviews

- & Your interaction with ChatGPT

Benjamin: Hey guys, what are we doing with the keys today?
Hilary: I've got them. Whoever wants them can meet me at lunchtime or after
Elliot: I'm ok. We're meeting for the drinks in the evening anyway and
...
Benjamin: Interesting 😃 😱 To be honest, Hilary, I almost feel like changing my mind. Wanting to take this nap might end up costing me to dear
...
Hilary: Do join us, we're going to have fun. And then you'll take the keys and take this most deserved of naps
Elliot: Sounds like a plan
Hilary: 😃 😎
Elliot: See you at 2 then xx

Output: Benjamin, Hilary and Elliot are discussing about their plans for the day after getting the apartment keys.

What is going on for Dialogue Summarization?

Previous work

- Fluency and Coherence (T5, BART, LLaMA, etc.)

- Topic awareness (**Liu et al., ASRU 2019**, Chen & Yang, EMNLP 2020)

- Co-reference (**Liu et al., SIGDIAL 2021**)

- Factuality (Tang et al., NAACL 2022) (**Wang et al., EMNLP 2022**)

- Controllability (**Liu & Chen, EMNLP 2021**) (**Wang et al., EMNLP 2023**)

Instructive Dialogue Summarization

- One model that can process diverse instructions
  - Ideally: Very diverse instructions
  - In this study:
    1. General summary
    2. Query-based summary
    3. Length control

Benjamin: Hey guys, what are we doing with the keys today?
Hilary: I've got them. Whoever wants them can meet me at lunchtime or after
Elliot: I'm ok. We're meeting for the drinks in the evening anyway and
...
Benjamin: Interesting 😃 😱 To be honest, Hilary, I almost feel like changing my mind. Wanting to take this nap might end up costing me to dear
...
Hilary: Do join us, we're going to have fun. And then you'll take the keys and take this most deserved of naps
Elliot: Sounds like a plan
Hilary: 😃 😎
Elliot: See you at 2 then xx

Instruction: Where is Hilary having lunch?
Output: La Cantina

Instruction: Highlight the key takeaways from the dialogue.
Output: Hilary has the keys to the apartment. Benjamin wants to get them and go take a nap. Hilary is having lunch with some French people at La Cantina. Hilary ...

Instruction: Summarize the dialogue with about 15 words.
Output: Benjamin, Hilary and Elliot are discussing about their plans for the day after getting the apartment keys.

Self-Generation for Data Augmentation

- Tailored as Summary-Anchored Data Augmentation

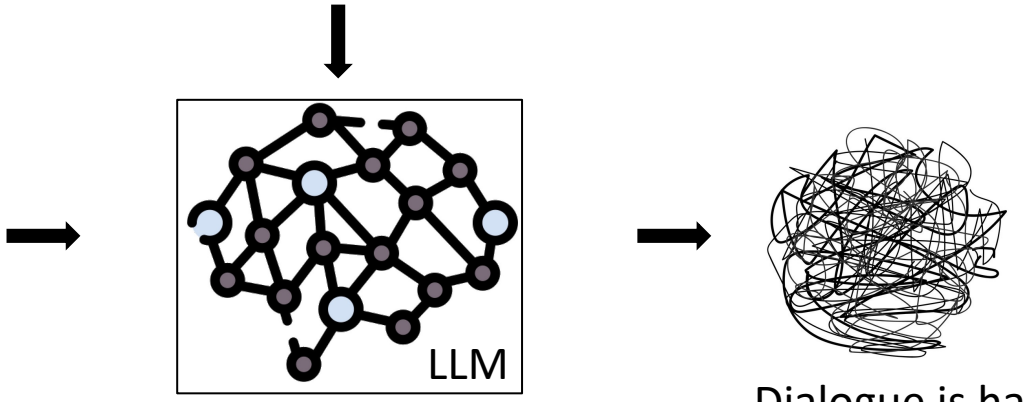
Dialogue

Benjamin: Hey guys, what are we doing with the keys today?
Hilary: I've got them. Whoever wants them can meet me at lunchtime or after
Elliot: I'm ok. We're meeting for the drinks in the evening anyway and
...
Benjamin: Interesting 😃 😱 To be honest, Hilary, I almost feel like changing my mind. Wanting to take this nap might end up costing me to dear
...
Hilary: Do join us, we're going to have fun. And then you'll take the keys and take this most deserved of naps
Elliot: Sounds like a plan
Hilary: 😃 😎
Elliot: See you at 2 then xx

Human Written Summary

Output: Hilary has the keys to the apartment. Benjamin wants to get them and go take a nap. Hilary is having lunch with some French people at La Cantina. Hilary ...

- You're a helpful assistant.
- Please generate some questions based on the context.



Dialogue is harder to understand
1. Factual errors
2. Not answerable

Self-Generation for Data Augmentation

- Tailored as Summary-Anchored Data Augmentation

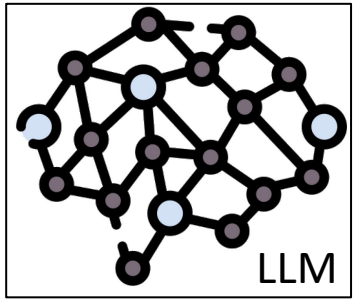
Dialogue

Benjamin: Hey guys, what are we doing with the keys today?
Hilary: I've got them. Whoever wants them can meet me at lunchtime or after
Elliot: I'm ok. We're meeting for the drinks in the evening anyway and
...
Benjamin: Interesting 😃 😱 To be honest, Hilary, I almost feel like changing my mind. Wanting to take this nap might end up costing me to dear
...
Hilary: Do join us, we're going to have fun. And then you'll take the keys and take this most deserved of naps
Elliot: Sounds like a plan
Hilary: 😃 😎
Elliot: See you at 2 then xx

Human Written Summary

Output: Hilary has the keys to the apartment. Benjamin wants to get them and go take a nap. Hilary is having lunch with some French people at La Cantina. Hilary ...

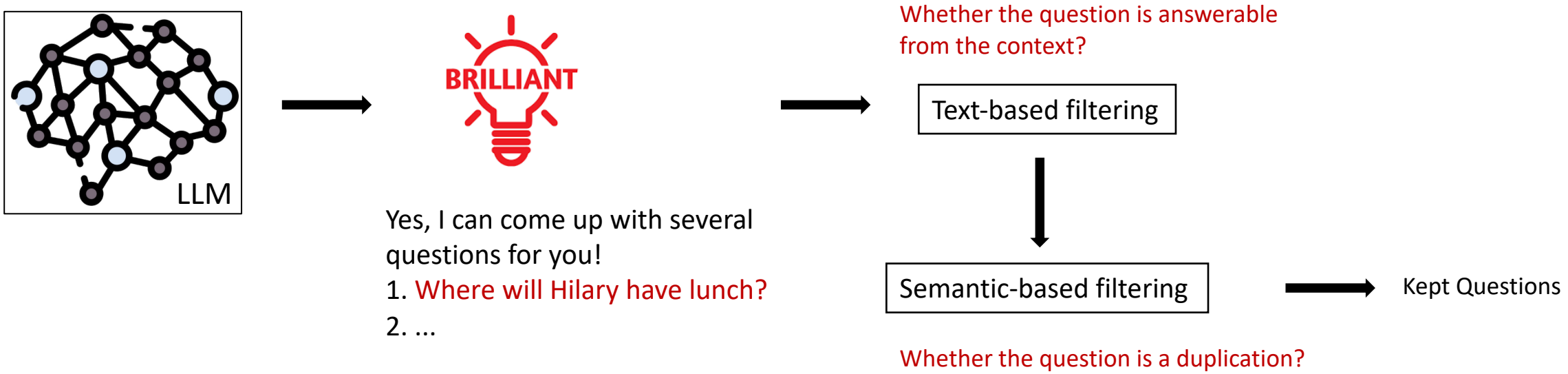
- You're a helpful assistant.
- Please generate some questions based on the context.



Yes, I can come up with several questions for you!
1. Where will Hilary have lunch?
2. ...

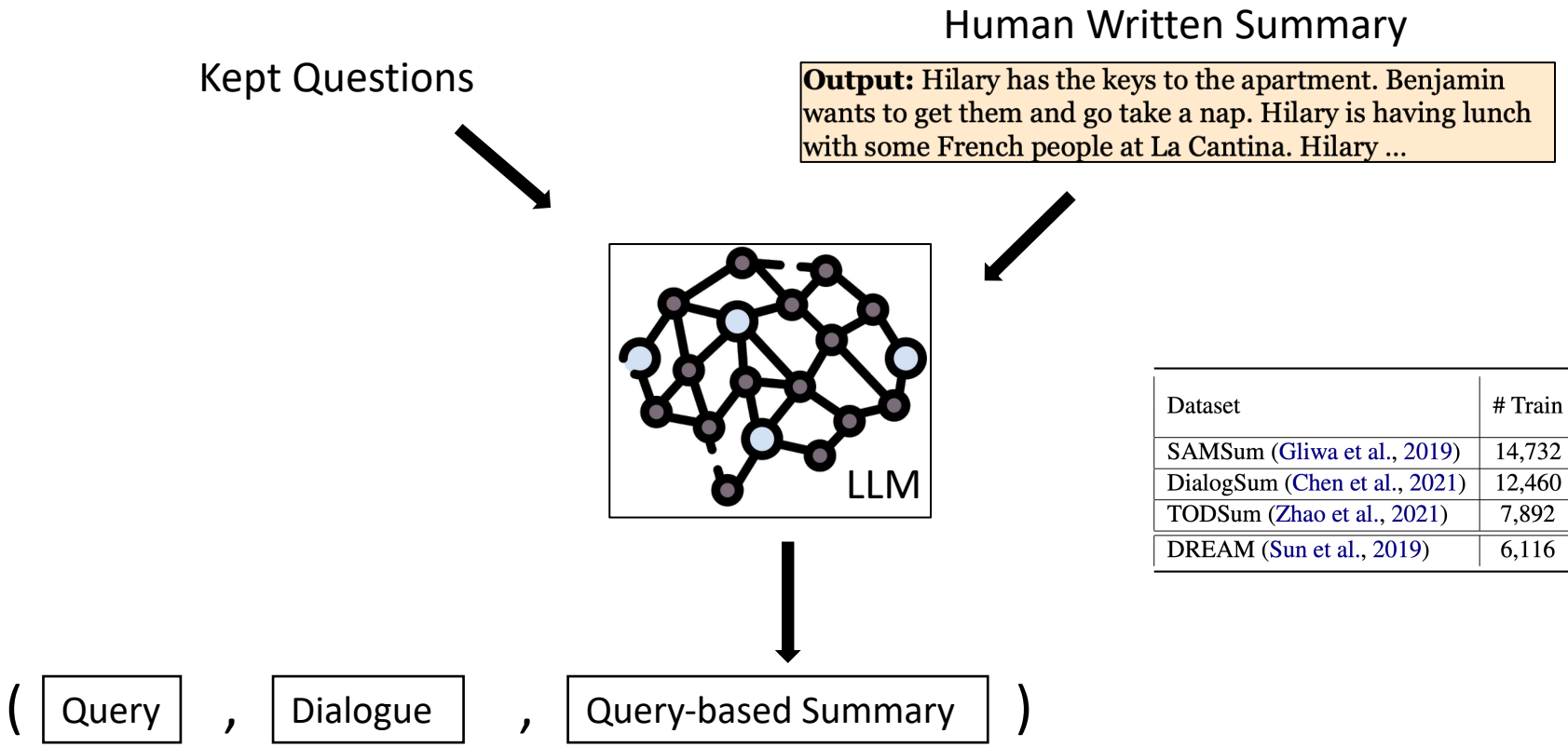
Self-Generation for Data Augmentation

- Filtering and Postprocessing



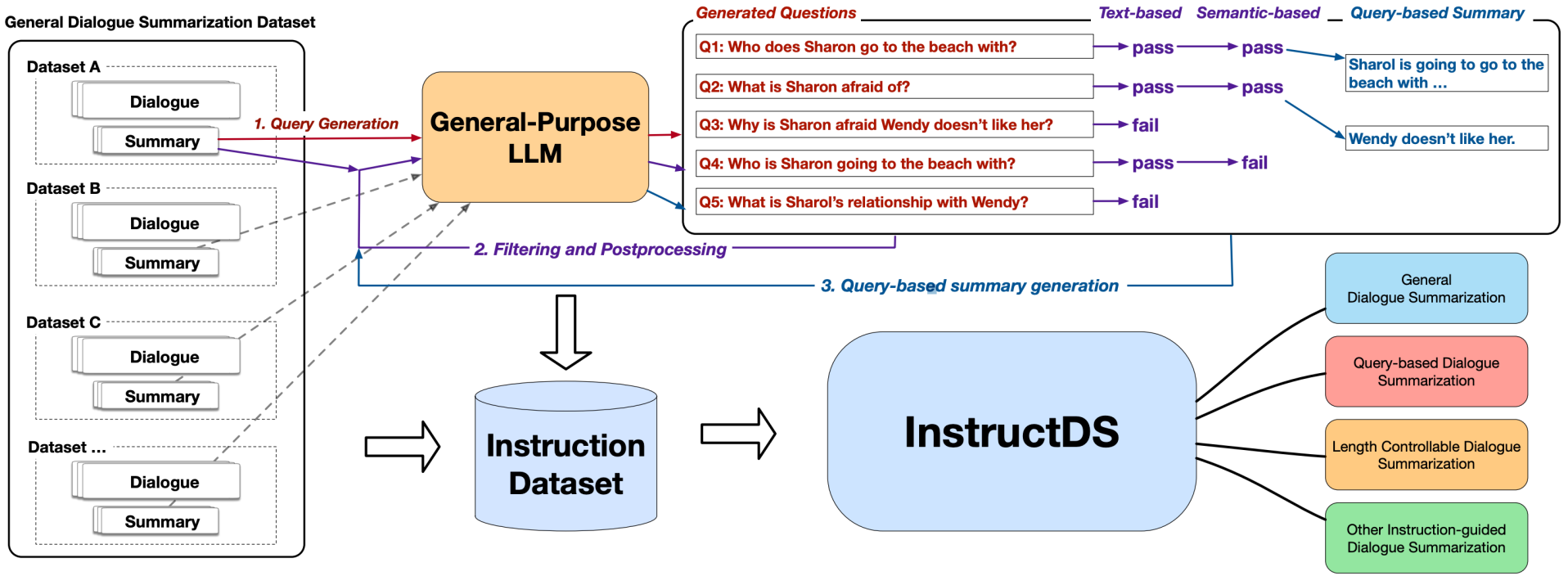
Self-Generation for Data Augmentation

- (Query, Dialogue, Summary) triples



Dataset	# Train	# Validation	# Test	# QDS Triples
SAMSum (Gliwa et al., 2019)	14,732	818	819	18,245
DialogSum (Chen et al., 2021)	12,460	500	1,500	18,600
TODSum (Zhao et al., 2021)	7,892	999	999	8,705
DREAM (Sun et al., 2019)	6,116	2,040	2,041	-

InstructDS: A Unified Model



Experiments

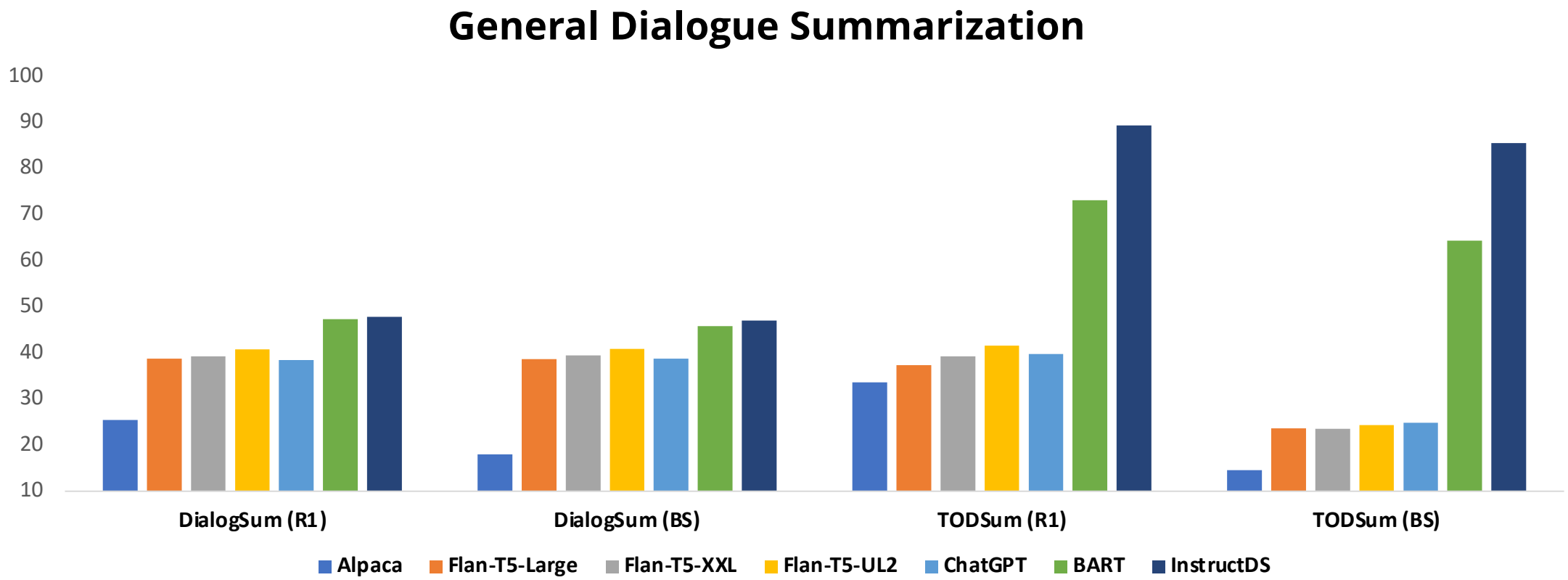
- SAMSum dataset

Models	Params	ROUGE-1			ROUGE-2			ROUGE-L			BS
		F_1	Pre	Rec	F_1	Pre	Rec	F_1	Pre	Rec	
Pointer-Generator	-	40.1	-	-	15.3	-	-	36.6	-	-	-
BART	400M	53.0	59.0	52.8	28.4	32.1	28.2	44.2	49.3	44.0	53.3
MV-BART	400M	53.9	55.7	57.4	28.4	29.3	30.6	44.4	45.7	47.5	53.6
Coref-BART	400M	53.7	56.9	56.4	28.5	30.5	29.7	44.3	46.9	46.5	53.5
ConDigSum	400M	54.3	56.0	57.6	29.3	30.4	31.2	45.2	46.6	48.0	54.0
GPT-3-finetune	175B*	53.4	-	-	29.8	-	-	45.9	-	-	-
Alpaca	7B	28.2	26.0	39.8	5.7	5.1	8.3	20.5	19.2	29.0	19.4
Flan-T5-XXL	11B	52.6	62.6	50.0	28.5	34.1	27.1	44.1	52.5	41.9	53.2
Flan-UL2	20B	53.3	60.3	52.5	28.0	32.0	27.7	44.1	50.0	43.3	53.5
ChatGPT	175B	32.7	22.4	70.2	12.3	8.4	27.1	24.7	16.9	53.6	32.5
InstructDS	3B*	**55.3**	58.8	57.5	**31.3**	33.5	32.6	**46.7**	49.7	48.6	**55.5**
w/ reference summary length											
ChatGPT	175B	40.8	39.3	43.4	13.7	13.2	14.6	31.5	30.5	33.4	40.0
InstructDS	3B*	**58.4**	58.5	58.8	**32.8**	32.9	33.0	**48.9**	49.0	49.2	**58.5**

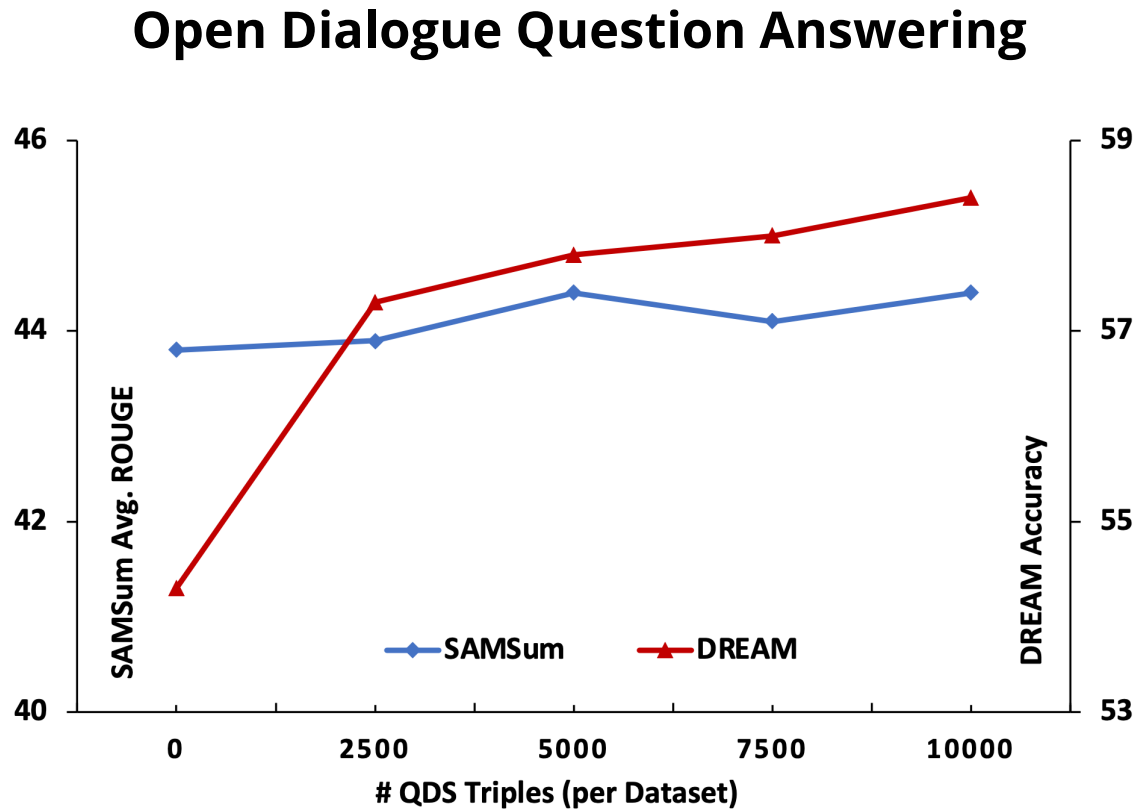
InstructDS achieves the best performance on general dialogue summarization

- Better understanding of dialogue interactions

Experiments



Experiments



- Query-Dialogue-Summary triples helps understand the dialogue
- Synthesized triples can help handle open questions

Future Work on Instructive Dialogue Summarization

- LLM performs well in general domain, inferior in dialogue processing
- Open problem
  - Handle more diverse instruction
  - Evaluation, especially on Faithfulness

Models	Human Annotator				ChatGPT			
	Faithfulness	Fluency	Informativeness	Conciseness	Faithfulness	Fluency	Informativeness	Conciseness
BART	$3.85_{(1.3)}$	$4.36_{(0.8)}$	$3.22_{(1.0)}$	$4.30_{(0.9)}$	$4.22_{(1.1)}$	$4.80_{(0.5)}$	$3.37_{(1.0)}$	$4.93_{(0.3)}$
Alpaca	$3.24_{(1.3)}$	$3.77_{(1.3)}$	$3.45_{(1.1)}$	$3.11_{(1.4)}$	$3.59_{(1.3)}$	$4.07_{(1.0)}$	$3.19_{(1.2)}$	$4.29_{(1.0)}$
Flan-UL2	$4.00_{(1.3)}$	$4.38_{(0.9)}$	$3.03_{(1.2)}$	$4.29_{(1.0)}$	$4.45_{(0.9)}$	$4.78_{(0.5)}$	$3.52_{(1.0)}$	$4.91_{(0.3)}$
ChatGPT	$4.52_{(0.9)}$	$4.38_{(0.9)}$	$4.62_{(0.6)}$	$2.77_{(1.4)}$	$4.94_{(0.3)}$	$4.94_{(0.2)}$	$4.78_{(0.4)}$	$4.89_{(0.3)}$
Human-written	$4.34_{(1.0)}$	$4.54_{(0.7)}$	$3.58_{(1.1)}$	$4.36_{(0.9)}$	$4.49_{(0.8)}$	$4.81_{(0.4)}$	$3.74_{(1.0)}$	$4.95_{(0.3)}$
InstructDS	$4.13_{(1.1)}$	$4.35_{(0.8)}$	$3.54_{(1.0)}$	$4.23_{(1.0)}$	$4.60_{(0.8)}$	$4.82_{(0.4)}$	$3.78_{(0.9)}$	$4.92_{(0.3)}$

CREATING GROWTH, ENHANCING LIVES

Institute for Infocomm Research
I²R

THANK YOU

www.a-star.edu.sg

Instructive Dialogue Summarization with Query Aggregations

Demo Code

Paper