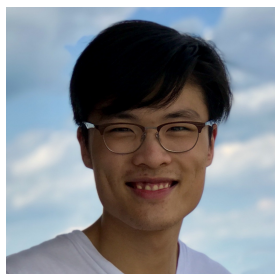

Learning to Decompose:
Hypothetical Question Decomposition
Based on Comparable Texts



Ben Zhou



Kyle
Richardson

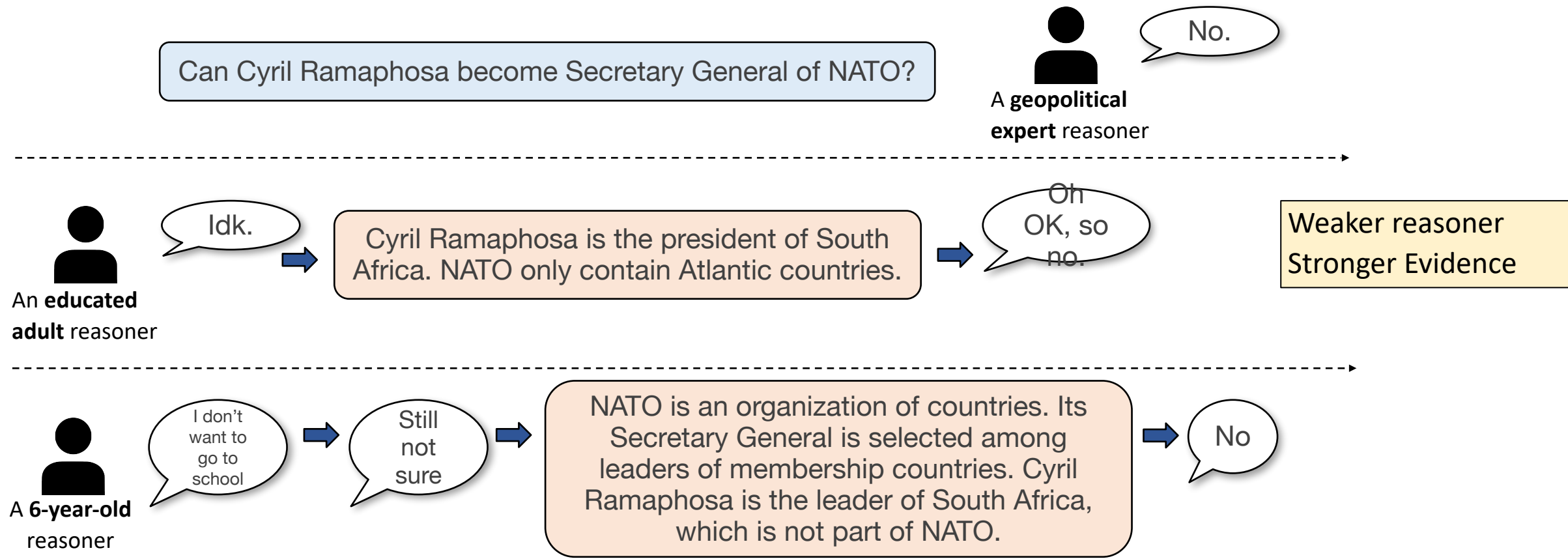


Xiaodong Yu

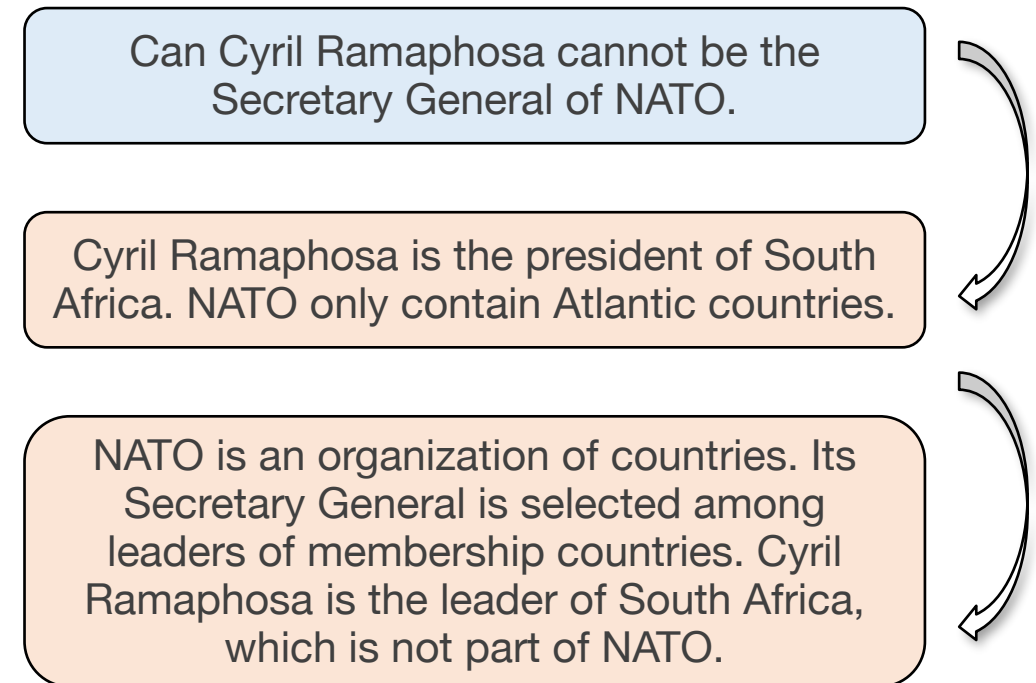


Dan Roth

- Reasoning can be viewed as finding equivalencies that suit best for a reasoner.



- Reasoning can be viewed as finding equivalencies that suit best for a reasoner.
 - Each level is
 - Equivalent with respect to the end-goal
 - More fine-grained, less demanding for a reasoner
 - More effort to decompose and potentially noisier
- Decomposition:
 - the process of finding each level
- Key questions:
 - How do we decompose?
 - How do we know which level is enough?



How should we decompose?

- Decompose is about finding equivalent reasoning process with respect to a goal.
- Why existing models struggle to find these equivalences?
 - Reporting bias: authors do not repeat a process with another equivalent one



- Language models cannot easily pick up such equivalencies
- How do we mitigate such gap?

How should we decompose?

- Learn to decompose from **comparable texts**
 - Parallel news articles that describe the same things from different angles

Document A

The Albany in NY is more crowded than that in GA.

While they are prevalent today...

...latest environment protection...

Document B

The Albany in NY has more people and less space.

There are a large number of these...

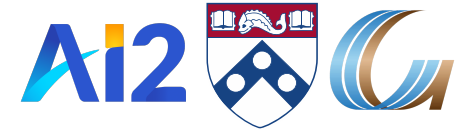
The administration... reducing methane gas...

Is cow methane safer for the environment than cars?



We need to compare the quantity of methane gas, lower the safer.

How should we decompose?



- Learn to decompose from **comparable texts**
 - Parallel news articles that describe the same things from different angles
- Provides implicative equivalencies that are richer than paraphrasing data
- In this work:
 - We collect distant supervision signals from parallel news
 - We train an equivalency-aware model DECOMPT5
 - We build a QA pipeline that follows a decomposition-entailment process

- Filter article pairs
 - We find article pairs that are roughly describing the same news from RealNews
 - Similar dates
 - Similar title (determined by a paraphrasing model)
- Filter sentence pairs
 - From selected article pairs, we find sentence pairs that provides complementary information
 - Determined by a paraphrasing model
 - Lower-bound 0.6 to ensure relevance
 - Upper-bound 0.9 to filter out paraphrasing sentences
- Topic filtering with TF-IDF
- Resulting 2.6M sentence pairs

- Using the 2.6M sentence pairs, we train a T5-large model

Input

The Albany in NY is more crowded than that in GA.

Output

The Albany in NY has more people and less space.

- We call the resulting model DECOMPT5
 - Is it better than T5 on reasoning (equivalency-finding) benchmarks?

- On two semantic parsing datasets

- Overnight

Input (natural language query)

when is the weekly standup

apartments for rent below 1500

Output (semi-formal semantic)

start time of weekly standup

housing unit whose monthly rent is at most 1500 dollars

- TORQUE

Input (question)

What has started when X happened?

What finished before X?

Output (semi-formal semantic)

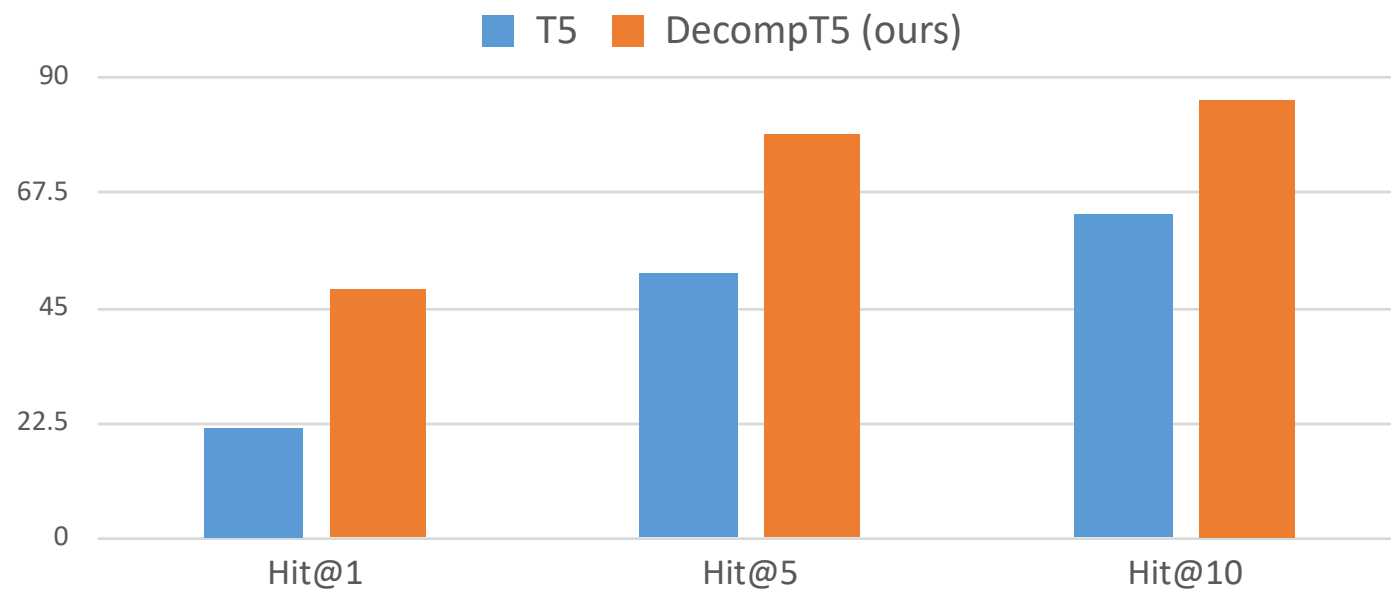
Events whose start time is smaller than the start time of X.

Events whose **end time** is smaller than the **end time** of X.

Only in
test set

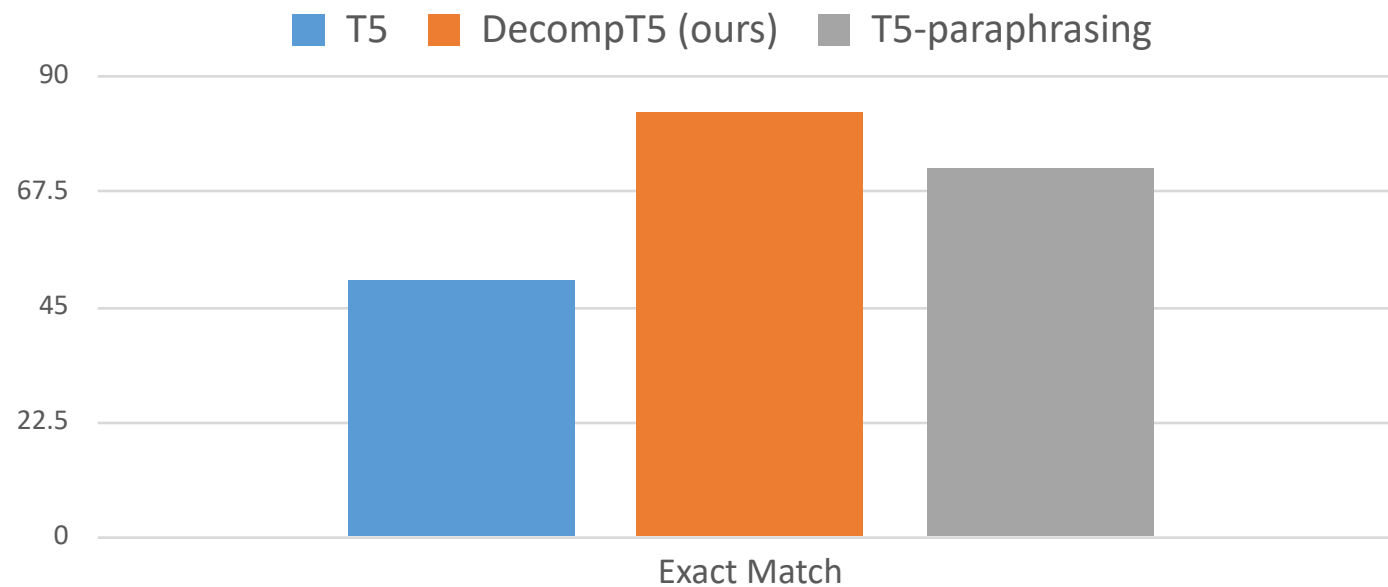
- Overnight

- Hit@K accuracy



- TORQUE

- Exact match accuracy



- DECOMPT5 works well on equivalency-finding tasks without knowledge
 - We need to extend it to more complicated tasks such as general QA
- Challenges:
 - Comparing to semantic parsing, QA requires a reasoner (execution) that finds the final answer given the decompositions (equivalencies)
- Key question:
 - What is a proper reasoner?
 - What level of decomposition is needed?

- We propose to use an entailment model as the reasoner
 - On StrategyQA, given gold decompositions, a T5-3B entailment model gets 90% accuracy (GPT-3: 65%)
- Consequently, we use StrategyQA's definition for the granularity of decompositions
- We propose DECOMPENTAIL
 - First decompose the question to equivalent forms that are easier for entailment models
 - Then run an entailment model to derive the final answer

Original Question

Can Cyril Ramaphosa become Secretary General of NATO?

Decompose with DECOMP5 finetuned on QA datasets

Cyril Ramaphosa is the president of Italy. **Generated Decomposition #1**

Correct the generated decomposition with GPT-3

Cyril Ramaphosa is the president of South Africa. **Corrected Decomposition #1**

Iterate until X number of decompositions are generated and corrected

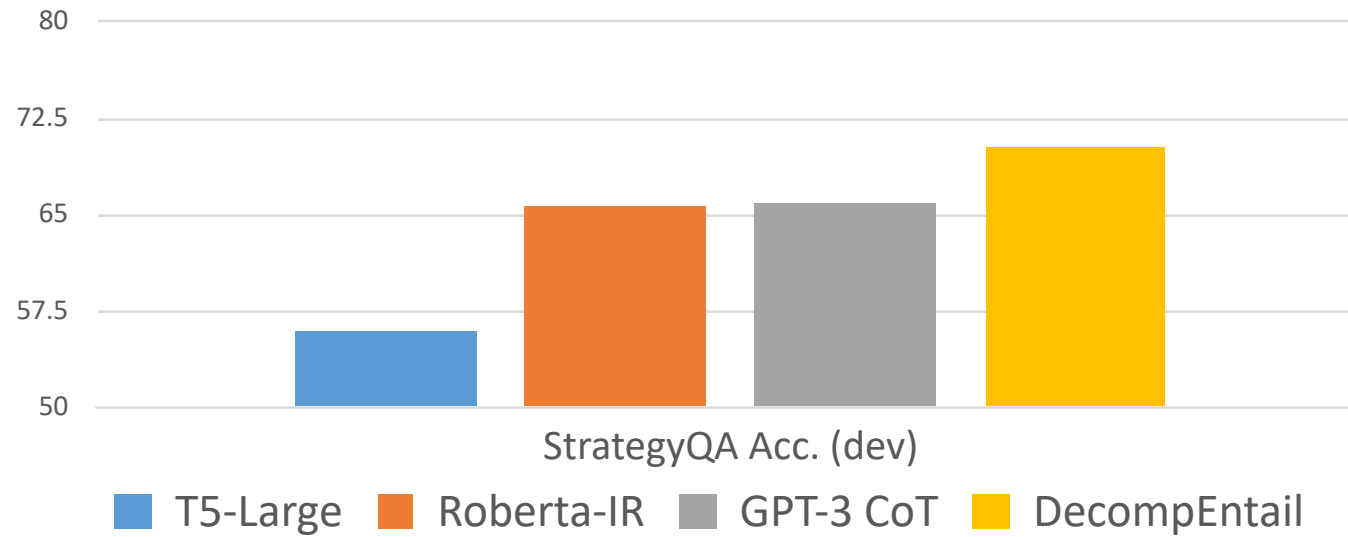
Can Cyril Ramaphosa become Secretary General of NATO?
Decompositions: Cyril Ramaphosa is the president of South Africa. South Africa is not part of NATO.

Entail → **NO Final Answer**

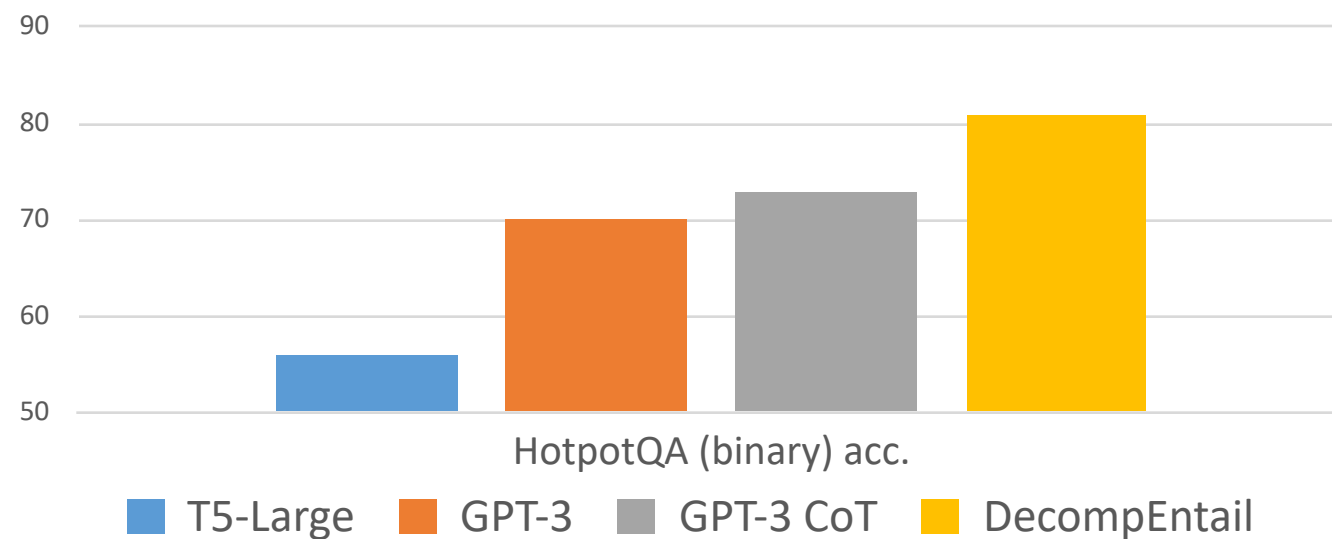
GPT-3 only sees one generated decomposition at a time, and does not involve in the reasoning process

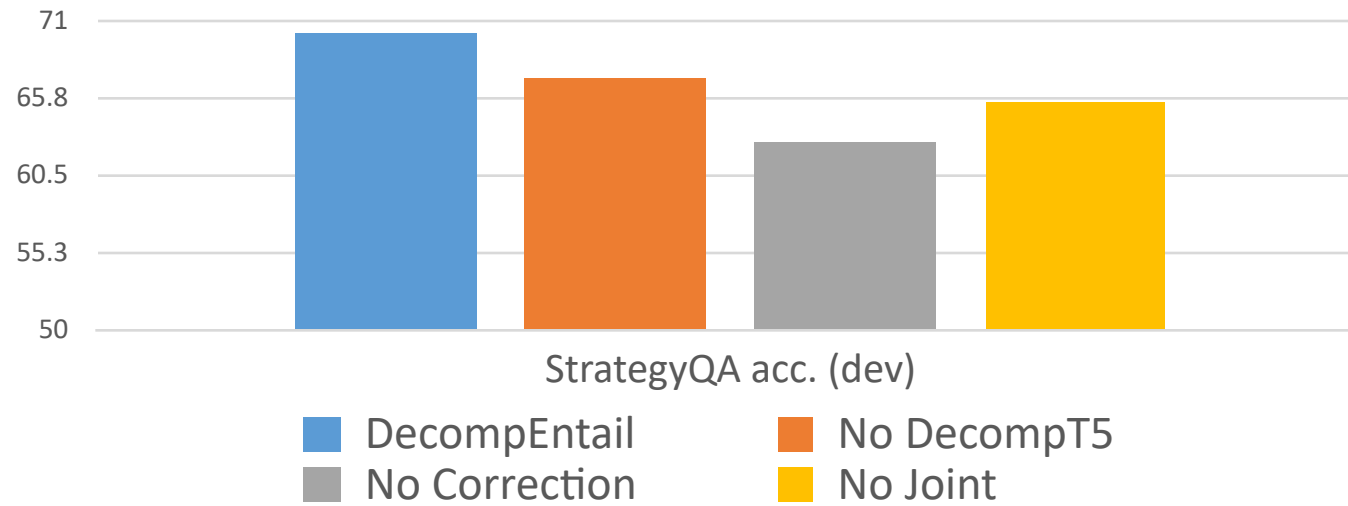
Concatenation of question and decompositions

- Joining training with StrategyQA and HotpotQA (binary questions)



- Joining training with StrategyQA and HotpotQA (binary questions)





- In this paper
 - We collect a large distant supervision dataset for equivalency-finding reasoning
 - With this dataset, we train DECOMPT5
 - improves 20% to 30% on semantic parsing datasets
 - With DECOMPT5, we propose DECOMPENTAIL
 - A general-purpose QA pipeline
 - Decompose -> Correct -> Entail
 - Improves 4% to 8% on challenging QA benchmarks
- More experiments and details in the paper



Code & Data