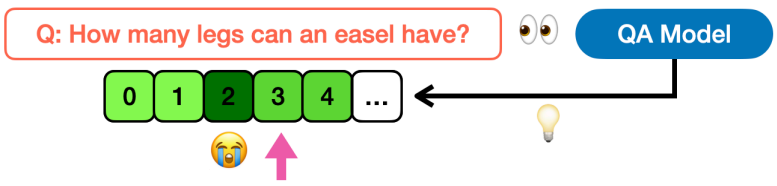
Generated Knowledge Prompting for Commonsense Reasoning

Jiacheng (Gary) Liu, Alisa Liu, Ximing Lu, Sean Welleck, Peter West,
Ronan Le Bras, Yejin Choi, Hannaneh Hajishirzi

University of Washington Allen Institute of AI

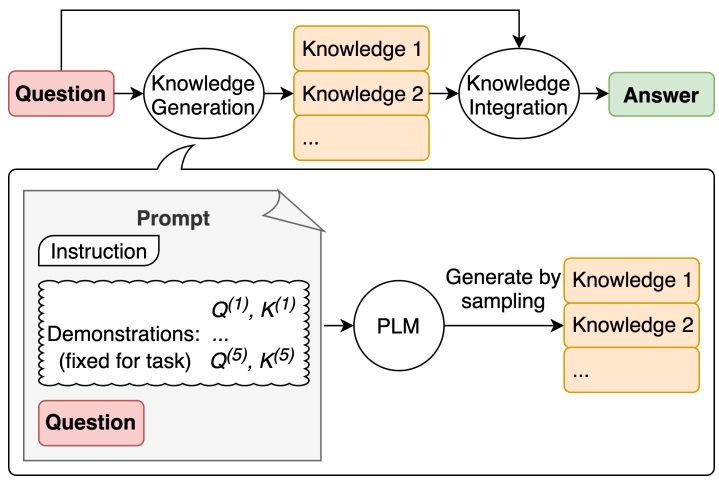
Overview

Can we let models make less mistakes?

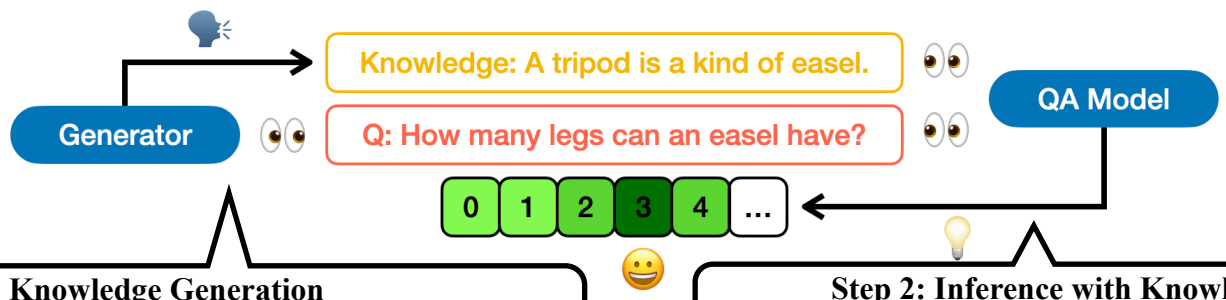


We propose a new method that substantially enhances model's commonsense reasoning performance.

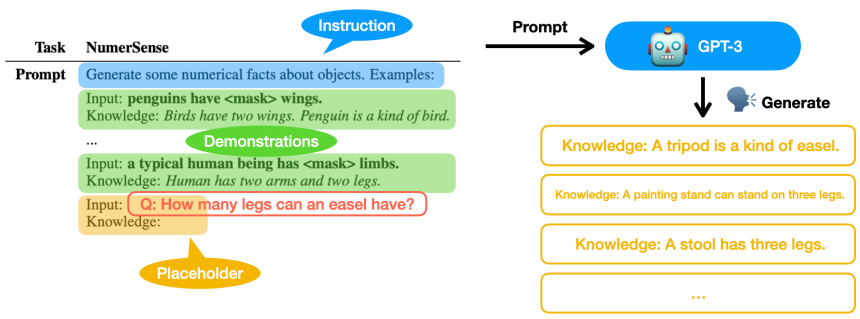
Main idea: Prompt an existing QA model with model-generated background knowledge.



Method



Step 1: Knowledge Generation

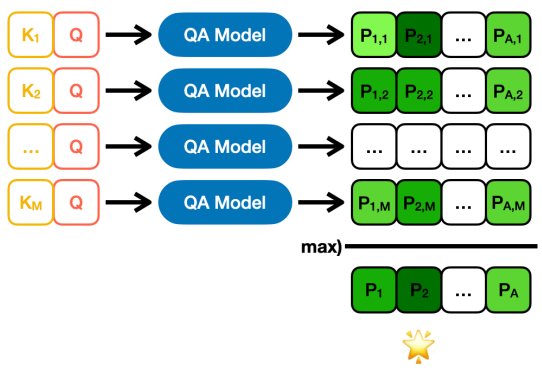


- We develop a task-specific prompt.
- For a new question, we plug it into the prompt and sample multiple *knowledge statements* from GPT-3's text continuation to the prompt:

$$K(q) = \{k_m : k_m \sim p_G(k \mid \text{prompt}, q)\}$$

- where q is the question, $K(q)$ is the set of knowledge statements, and $p_G(\cdot)$ is the generative distribution given by GPT-3 and decoding with nucleus sampling.

Step 2: Inference with Knowledge Prompting



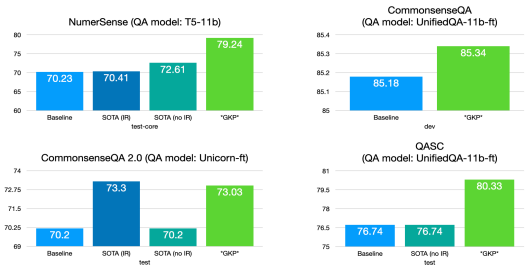
- Each knowledge statement individually *prompts* the QA model.
- We collect the confidence scores of choices in all model passes.
- Aggregated score for choice a: $P_a = \max_{1 \le m \le M} P_{a,m}$
- Final prediction: $\hat{a} = \max_{1 \le a \le A} P_a = \max_{1 \le a \le A} \max_{1 \le m \le M} P_{a,m}$

Experimental Results

Tasks: NumerSense, CommonsenseQA, CommonsenseQA 2.0, QASC

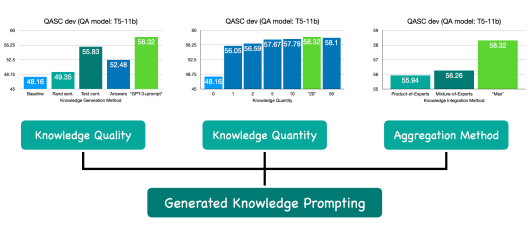
Main Results

- New SOTA among non-retrieval methods.
- Comparable with or outperforms retrieval methods.



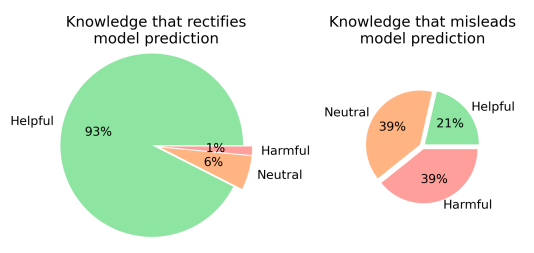
Ablations

Three key factors in the success of our method: knowledge quantity, knowledge quality, aggregation method.



Human Evaluation

Most knowledge are helpful to the QA model in a human-interpretable way.



Qualitative examples

Our generated knowledge support various reasoning procedures: induction, deduction, abduction, analogy, etc.

Dataset	Question / *Knowledge*	Prediction	Reasoning
NumerSense	**an easel can have [M] or four legs.** *A tripod is a kind of easel.*	two three	Commonsense Induction
CSQA	**Where does a heifer's master live?** *The master of a heifer is a farmer.*	slaughter house farm house	Commonsense Deduction
CSQA	**I did not need a servant. I was not a what?** *People who have servants are rich.*	in charge rich person	Commonsense Abduction
QASC	**[M] is used for transportation.** *Bicycles are used for transportation.*	plastic boats	Commonsense Analogy

Large pre-trained language models as source of flexible, high-quality knowledge

Code/Data
https://github.com/liujch1998/GKP