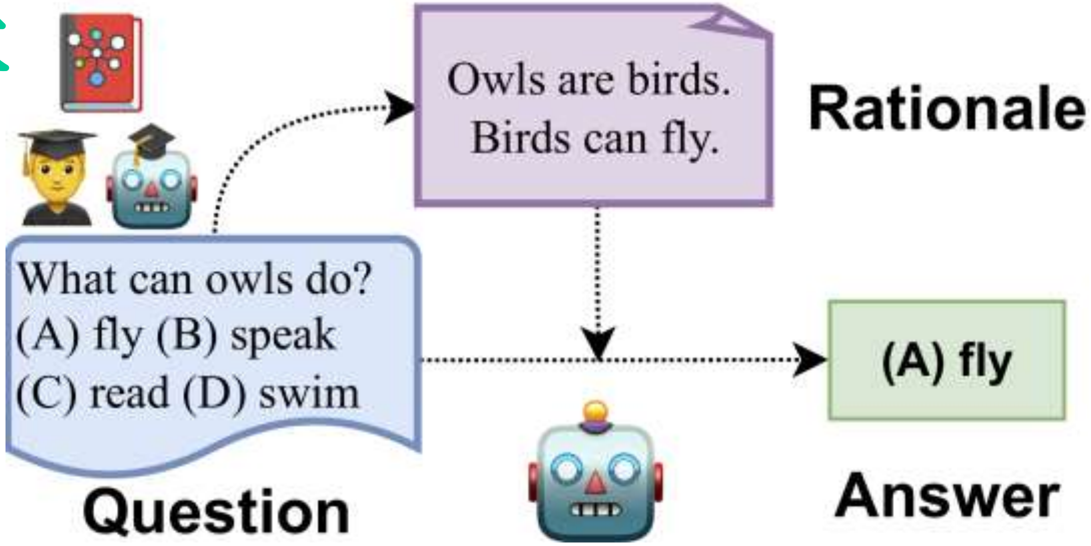
LEROS: Learning Explicit Reasoning on Synthesized Data for Commonsense Question Answering.

Chenhao Wang
Institute of Automation, Chinese Academy of Sciences

Background

- Commonsense Question Answering (CQA) requires understanding and reasoning with unstated background knowledge
- Providing explicit rationales (relevant knowledge and reasoning details) is useful
  - Performance
  - Interpretability



Ways to Get Rationales

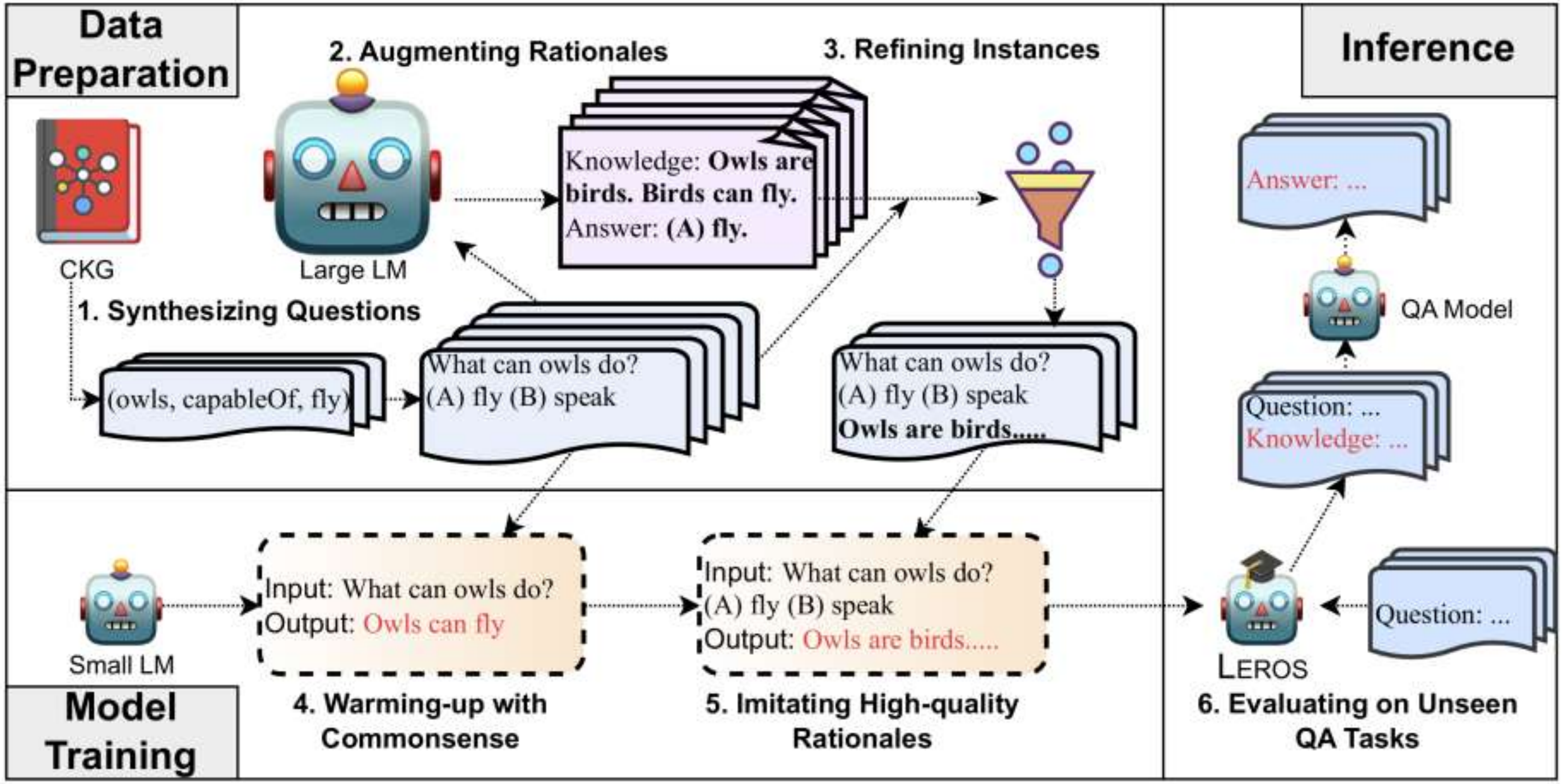
- The key problem: How to get high-quality rationales given the questions?
  - Commonsense Knowledge Graph (CKG)?
    - knowledge coverage, retrieval availability…
  - Learning to imitate human-authored rationales?
    - expensive annotation…
  - Prompting large language models (LLM) to generate rationales?
    - only for large models, expensive to deploy and further tune

Distilling the Ability to Small Models

- Distilling the rationale-generating ability of LLMs to small models? (Liu et al., 2022; Wang et al., 2022)
  - It works, but still relies on human-authored QA instances for the distillation
- **Our work**: making use of CKGs and LLMs, enabling small models to learn explicit reasoning from synthesized data
  - No need for human-authored QA instances.
  - Working in zero-shot settings.
  - Providing a strong start point for further tuning.

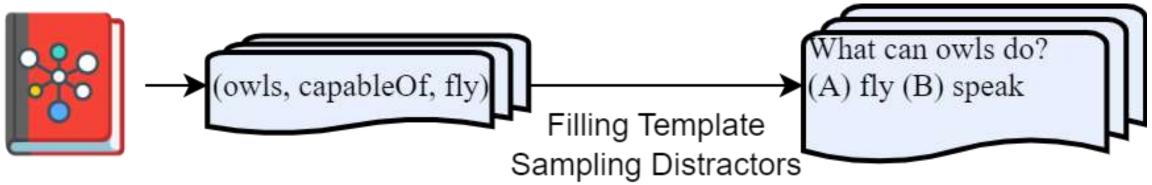
Liu et al. Rainier: Reinforced Knowledge Introspector for Commonsense Question Answering. EMNLP 2022.
Wang et al. Elaboration generating commonsense question answering at scale. EMLNP 2022.

Method Overview

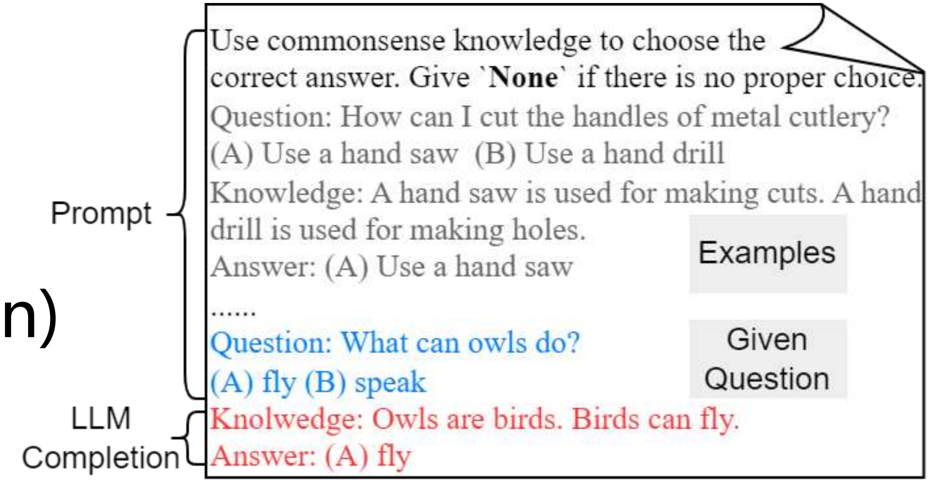


Data Preparation

- Synthesizing Questions



- Augmenting Rationales
  - Using templates
    - "owls can fly"
  - Querying LLMs (better for generalization)
    - "owls are birds; birds can fly"



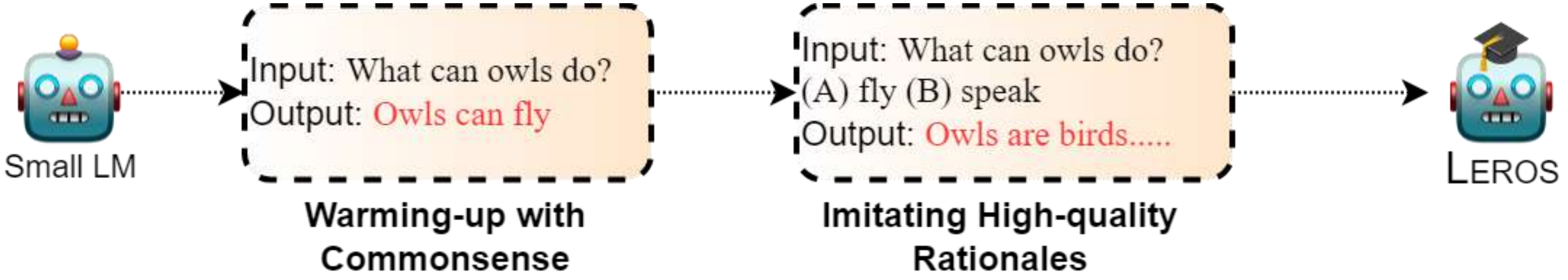
Data Preparation

- The synthesized question instances and rationales could be flawed
- Refining the results:
  - Consistency Refining
    - The LLM can predict the correct answer with its rationales
  - Helpfulness Refining
    - The rationales can improve the prediction of a general QA model (S > threshold)

$$S(k|q, a^*, A) = \frac{1}{2}$$
$$\left[\tanh\left(\log p(a^*|q \circ k) - \max_{\substack{a' \in A \\ a' \neq a^*}} \log p(a'|q \circ k)\right)\right.$$
$$\left.- \tanh\left(\log p(a^*|q) - \max_{\substack{a' \in A \\ a' \neq a^*}} \log p(a'|q)\right)\right]$$

Training

- Step1: Warming-up with commonsense knowledge
  - Input=question stem; Output=source knowledge
- Step2: Imitating high-quality rationales
  - Input=question and choices; Output=LLM generated rationales
- Optional: Reinforced learning with QA model feedback (Liu et al., 2022)
  - Using the data of the target CQA tasks to obtain the reward



Inference

- Generating multiple rationales using the LEROS model
  - Question (q) -> LEROS -> Rationale (k)
- Using the generated rationales to make the QA models
  - q,k -> QA Model -> a
  - (We can use some ensembles here)

$$\hat{a} = \arg\max_{a \in A} \max_{k \in K(q)} p(a | q \circ k)$$

Experiments

- Implementation of LEROS:
  - Commonsense Knowledge Graph: ATOMIC, ConceptNET, etc.
  - LLM (for augmenting rationales): GPT3.5-turbo
  - QA model (for refining data and evaluation): UnifiedQA-large (770M)
  - Base model for training LEROS: T5-large (770M)
- Statistics of synthesized data:

	Initial Data ($\mathcal{D}^{syn}$)	Queried Data	Consistent Data	Refined Data ($\mathcal{D}^{refine}$)
Train	823K	441K	303K	173K
Dev	67K	/	/	1K

Experiments

- Evaluation: Question + Rationales -> UnifiedQA (UQA)
  - Varying the rationale source, comparing the performance
- Baselines:
  - UQA
  - Directly finetuning UQA with synthesized QA instances
  - Few-shot Prompting GPT3/GPT3.5 to generate rationales + UQA
  - RAINIER (a previous RLQAF rationale-generation model) + UQA
  - Golden rationales (if there are human-authored rationales) + UQA

Results

- LEROS can generate useful rationales and improve the performance in zero-shot setting
- When the data of target tasks are available, LEROS can be further tuned through reinforced learning to obtain better performance

Method	Rationale Source	QA Model	Dataset CSQA	QASC	PIQA	SIQA	WG	Average	Avg. Gain
	Gold Rationale	UQA	89.92	83.37	-	-	-	-	-
Few/Zero-shot	-	UQA	61.43	43.09	63.66	53.84	53.35	55.07	+0.00
	-	UQA_{syn}	62.24	52.27	66.05	55.42	55.25	58.25	+3.17
	Few-shot GPT-3.5-turbo	UQA	70.02	66.52	71.82	61.00	58.64	65.60	+10.53
	Self-talk GPT-3 (13B)	UQA	63.31	49.89	65.23	51.89	52.96	56.66	+1.58
	Few-shot GPT-3 (13B)	UQA	66.34	53.24	65.25	58.29	**55.56**	59.74	+4.66
	(Ours) LEROS (770M)	UQA	**67.89**	**56.59**	**67.57**	**59.77**	55.01	**61.37**	**+6.29**
Feedback Tuning	RAINIER (770M)	UQA	67.24	54.97	65.67	57.01	56.91	60.36	+5.09
	(Ours) LEROS_{RL} (770M)	UQA	**70.35**	**60.15**	**69.53**	**64.32**	**59.27**	**64.72**	**+9.65**

Results

- The trained LEROS can also work with other QA models
  - Different sizes of UQA variants
  - Zero-shot CQA models
  - Language models with QA prompts (e.g. Llama2)

QA Model→ Rationale Model↓	UQA (small)	UQA (base)	UQA (large)	UQA (3b)
-	39.07	45.51	55.07	66.51
RANIER	48.60	54.77	60.36	67.85
LEROS	**49.05**	**56.12**	**61.37**	**67.91**

Method	Average
RoBERTa-Large-CSKG (Ma et al., 2021)	64.0
LEROS + RoBERTa-Large-CSKG	**65.2**
DeBERTa-v3-Large-CAR (Wang et al., 2023a)	70.2
LEROS + DeBERTa-v3-Large-CAR	**71.1**
Llama2-7B (Few-shot)	53.4
Llama2-7B (CoT-SC)	55.8
LEROS + Llama2-7B (Few-shot)	**57.2**
Llama2-chat-7B (Few-shot)	58.6
Llama2-chat-7B (CoT-SC)	61.9
LEROS + Llama2-chat-7B (Few-shot)	**63.0**

Qualitative Results

- LEROS can generate helpful and readable rationales

Task	Question/**Rationale**	Category
CSQA	If there is a place that is hot and arid, what could it be? (A) bland (B) lifeless (C) sandy (D) neutral (E) freezing **Hot and arid can mean a place that is dry and inhospitable.**	attribute
QASC	What can measure pounds? (A) animals (B) lamphreys (C) a mouse (D) a ruler (E) humans (F) surveyor (G) a scale (H) a microscope **Measuring pounds is done using a scale.**	use
PIQA	how do you blame someone? (A) say they did it. (B) say you did it for them. **Blaming someone involves saying they did something wrong.**	subevent
SIQA	Ash always performed better at his workplace after a warm cup of coffee. What will Ash want to do next? (A) start a new task (B) take some nyquil (C) go home **After having a warm cup of coffee, people usually feel refreshed and want to continue their work.**	behavior
WG	Angela did a bunch of crunches and sit-ups but Cynthia didn't, consequentially _ had six- pack abs. (A) Angela (B) Cynthia **Doing crunches and sit-ups is a common exercise to get six-pack abs.**	taxonomy

End

The resources available at: https://github.com/wchrepo/leros

Thanks for your attention!

Contact: chenhao.wang@nlpr.ia.ac.cn