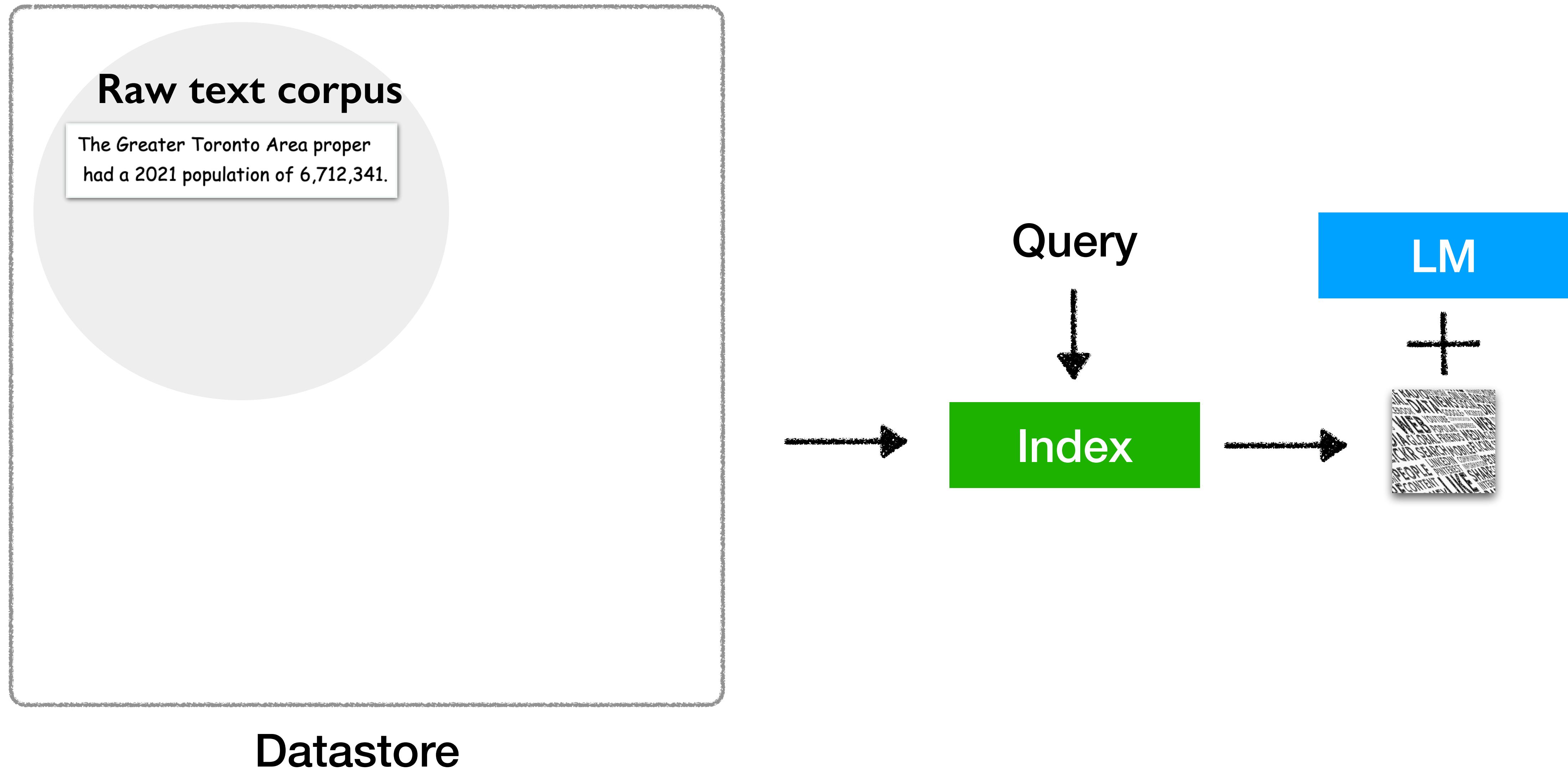
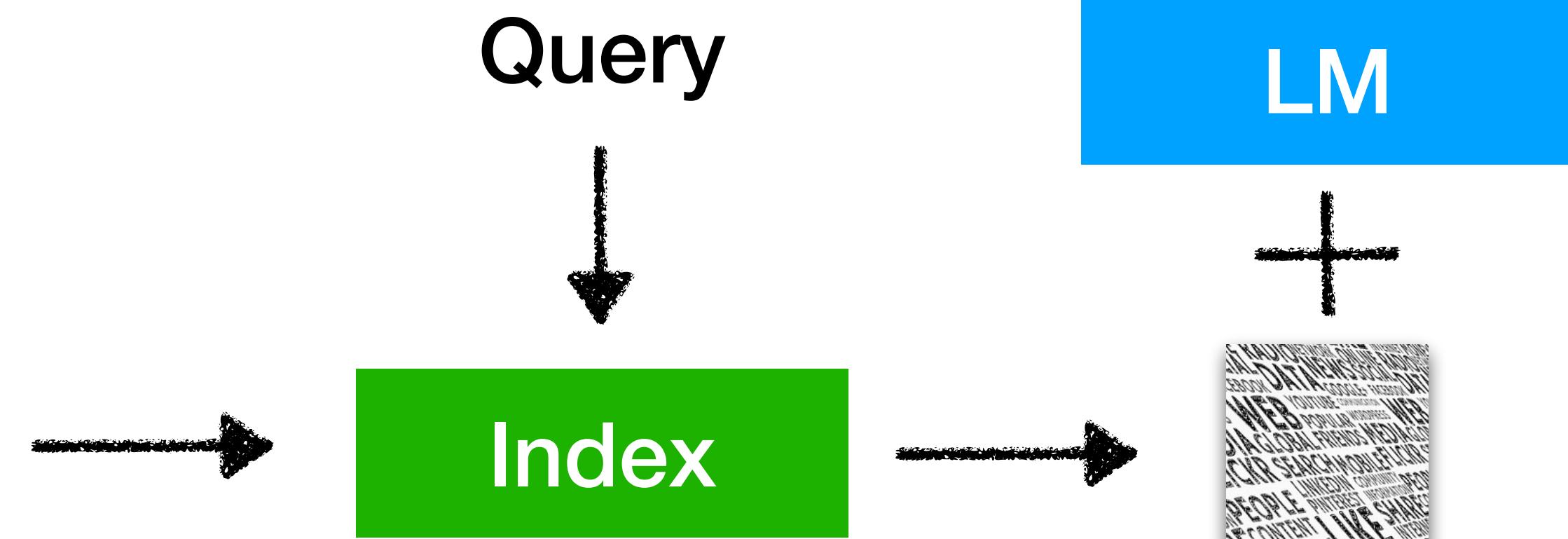
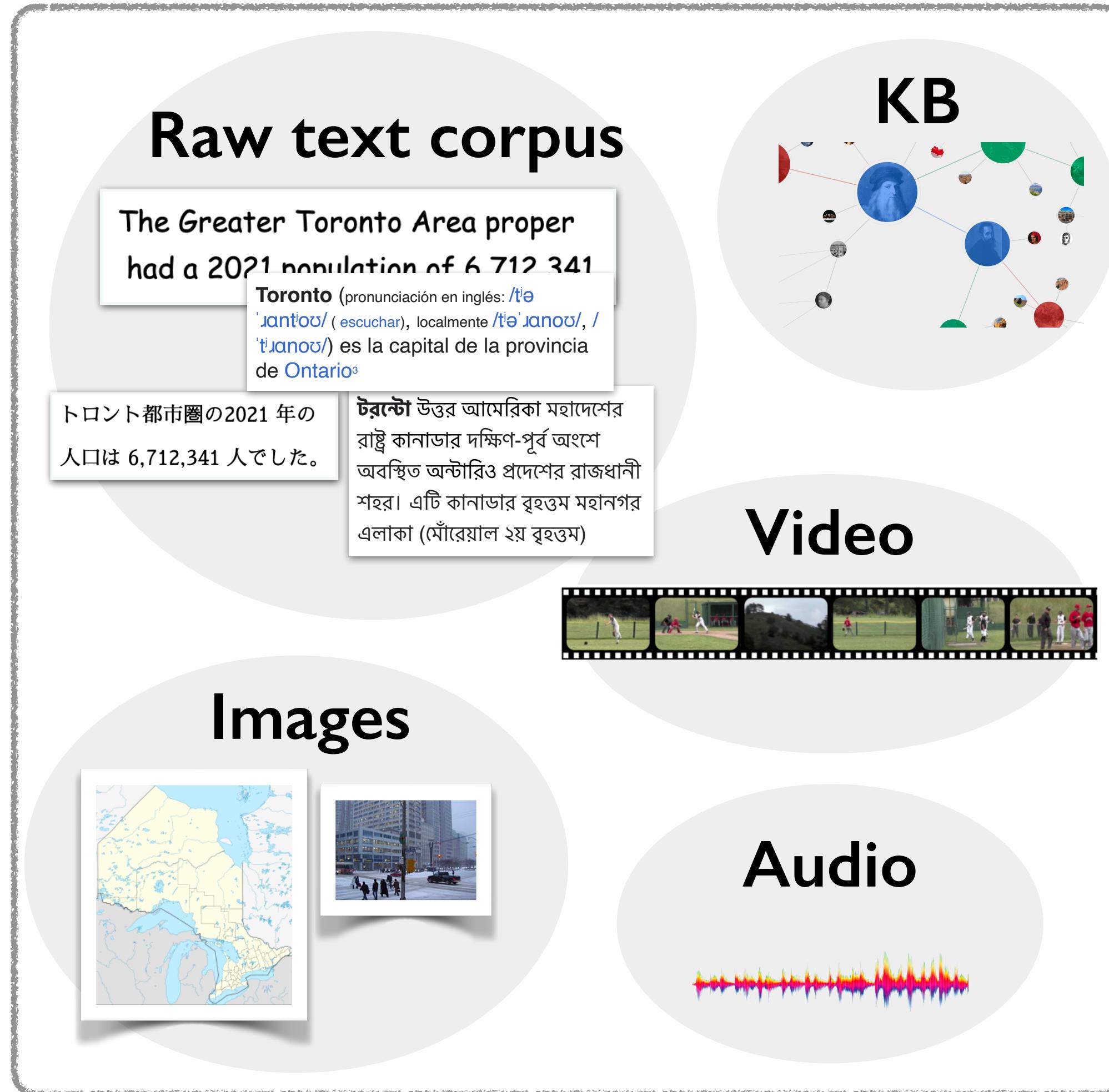


Section 6: Multilingual & Multimodal

Retrieval-based LM for diverse knowledge sources

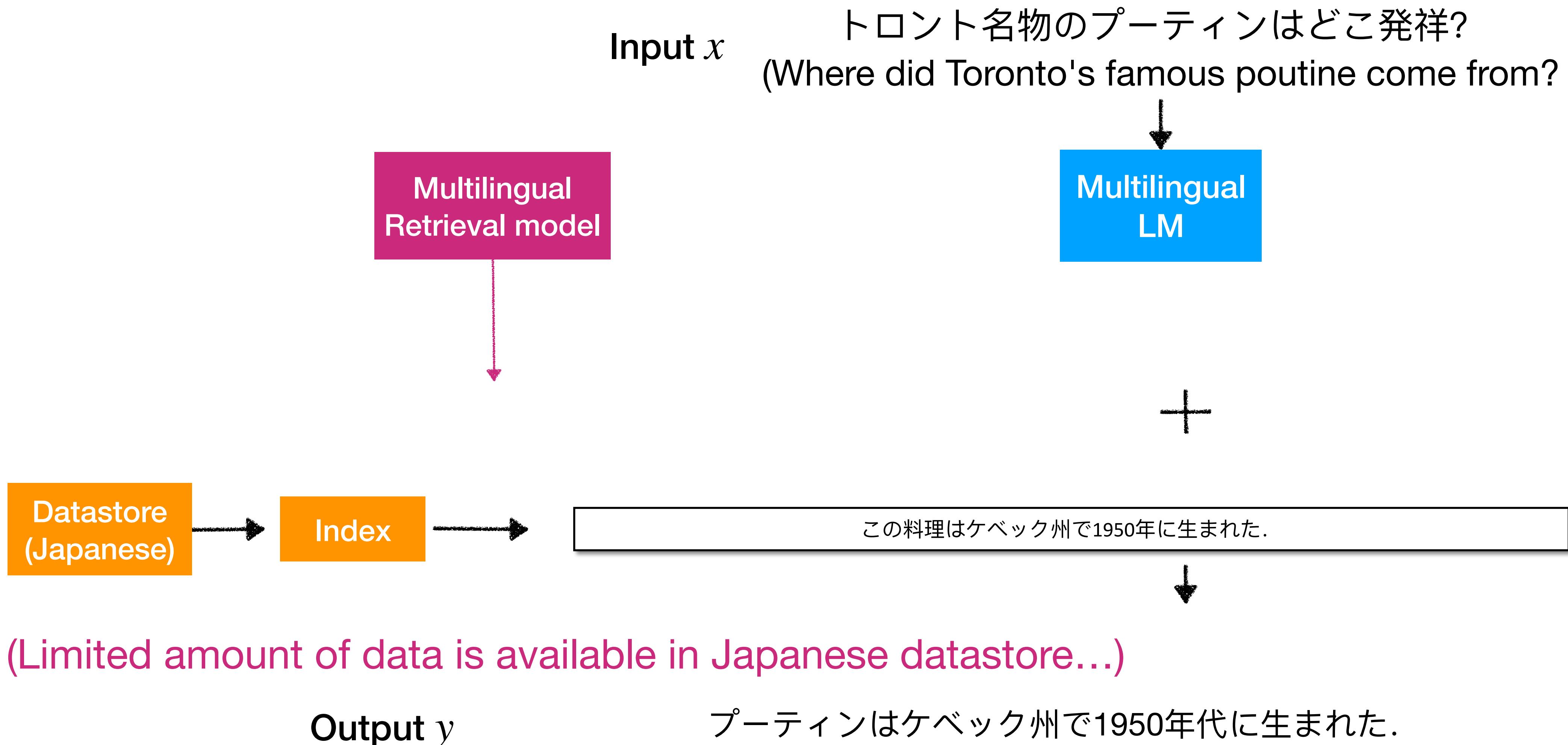


Retrieval-based LM for diverse knowledge sources

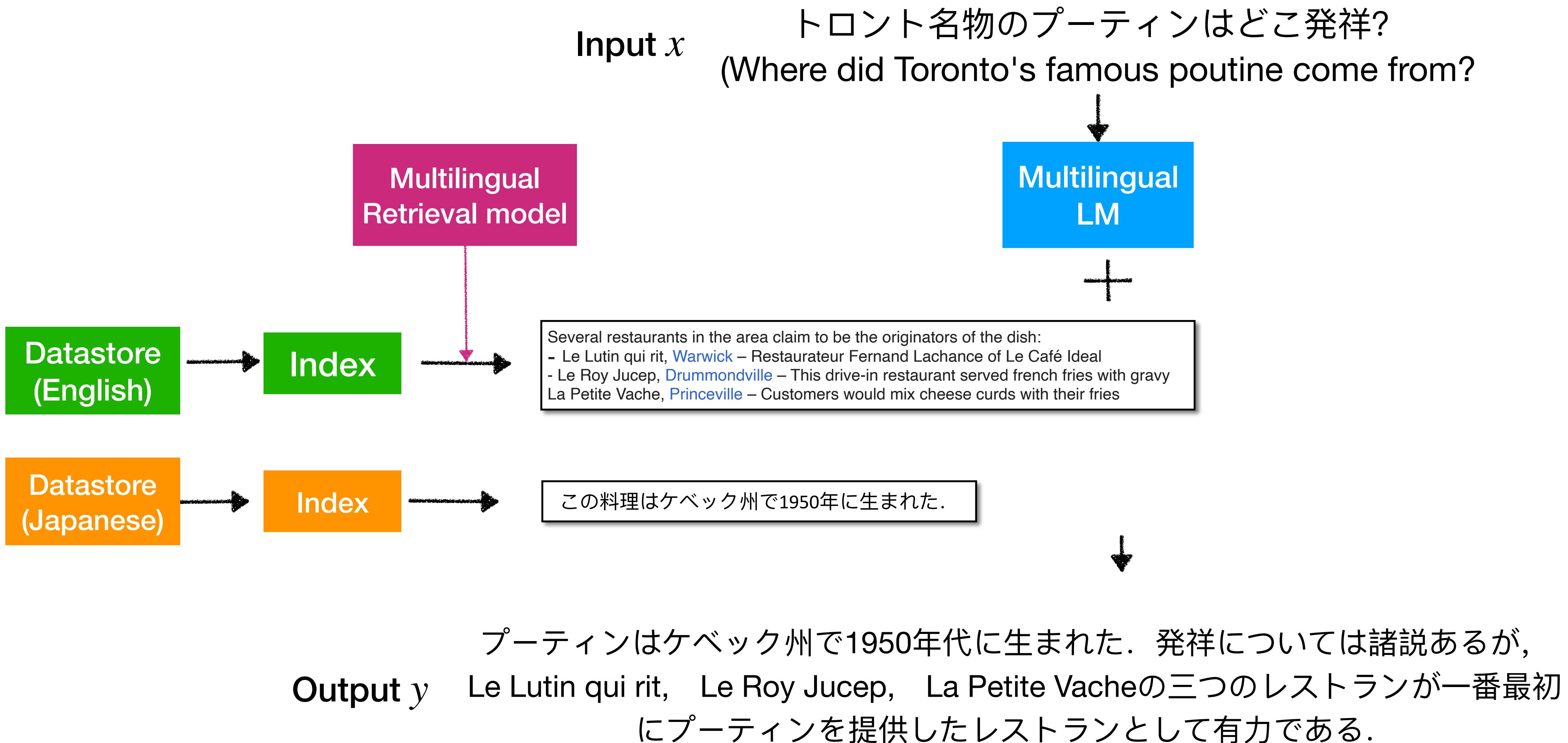


Datastore

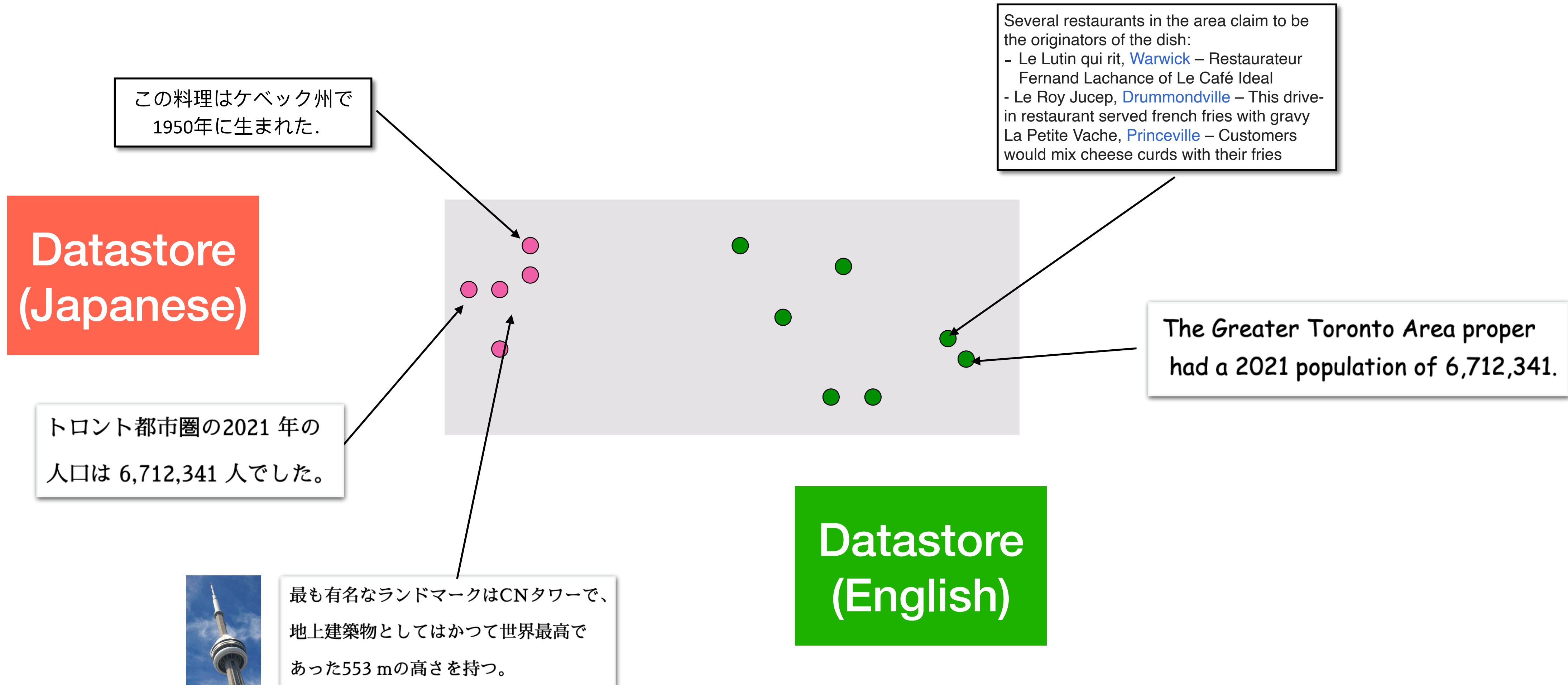
Multilingual Retrieval-based LM



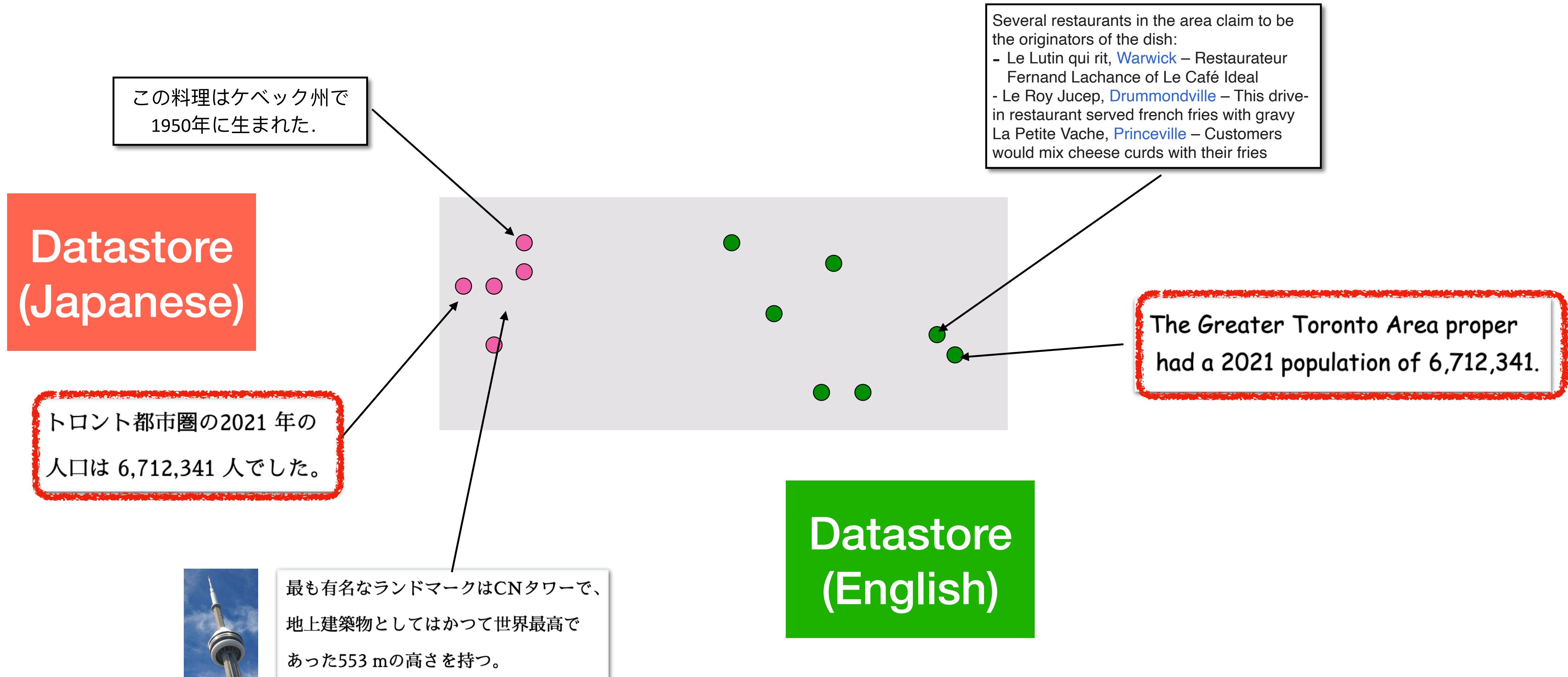
Multilingual Retrieval-based LM



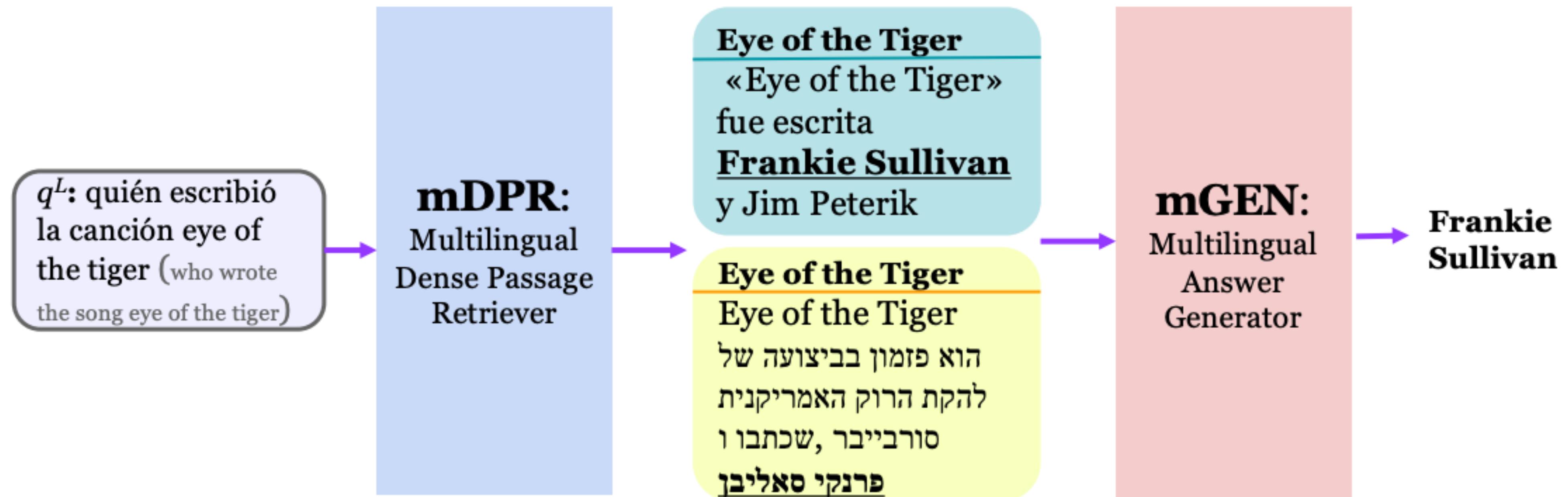
Language biases in representation spaces



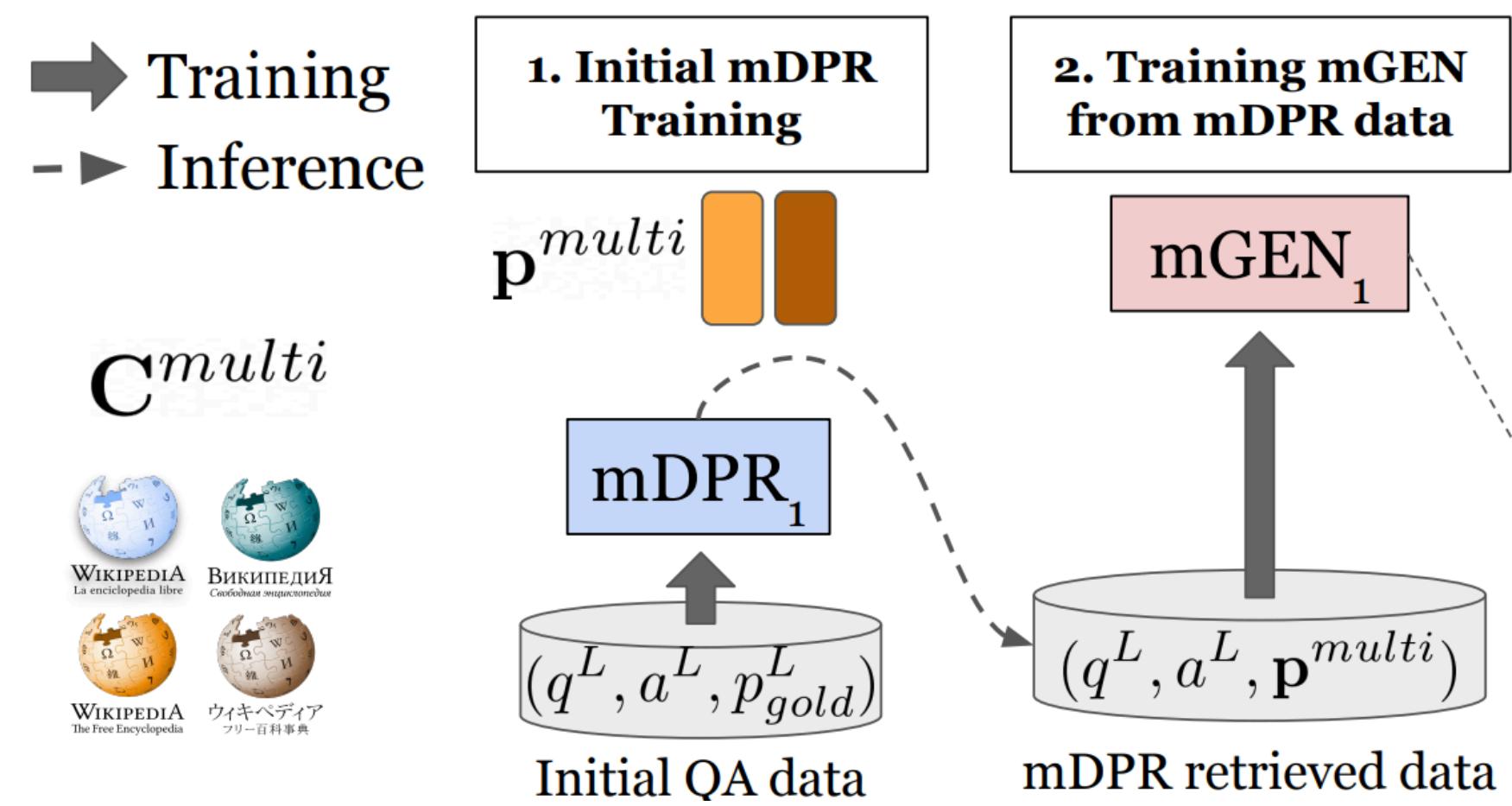
Language biases in representation spaces



CORA (Asai et al., 2021)

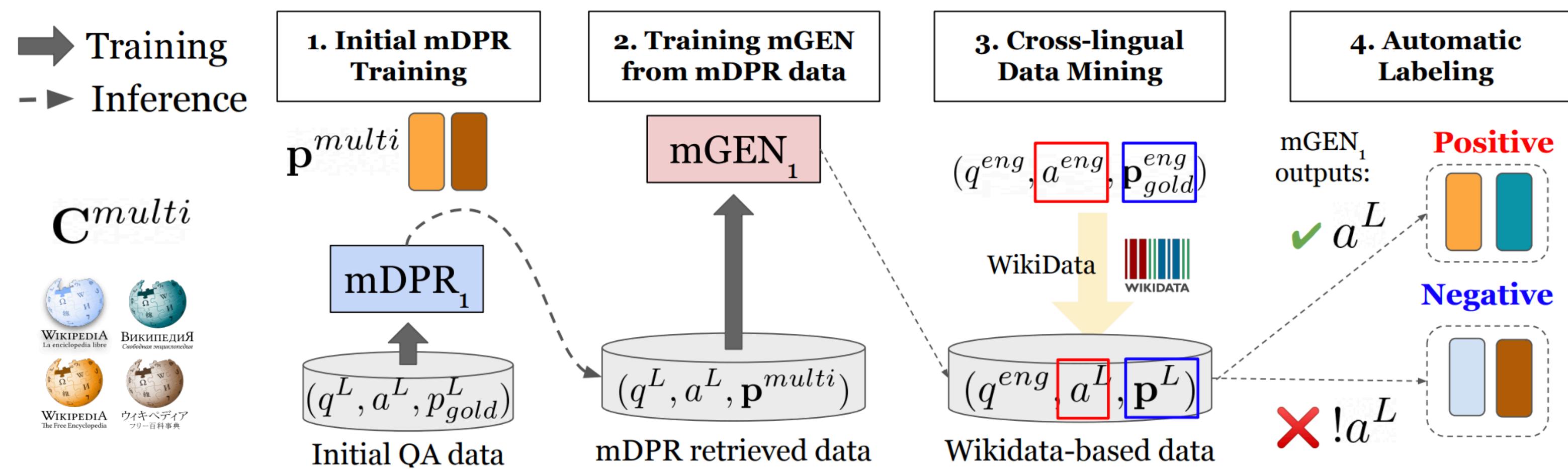


CORA: iteratively training multilingual LM & retriever



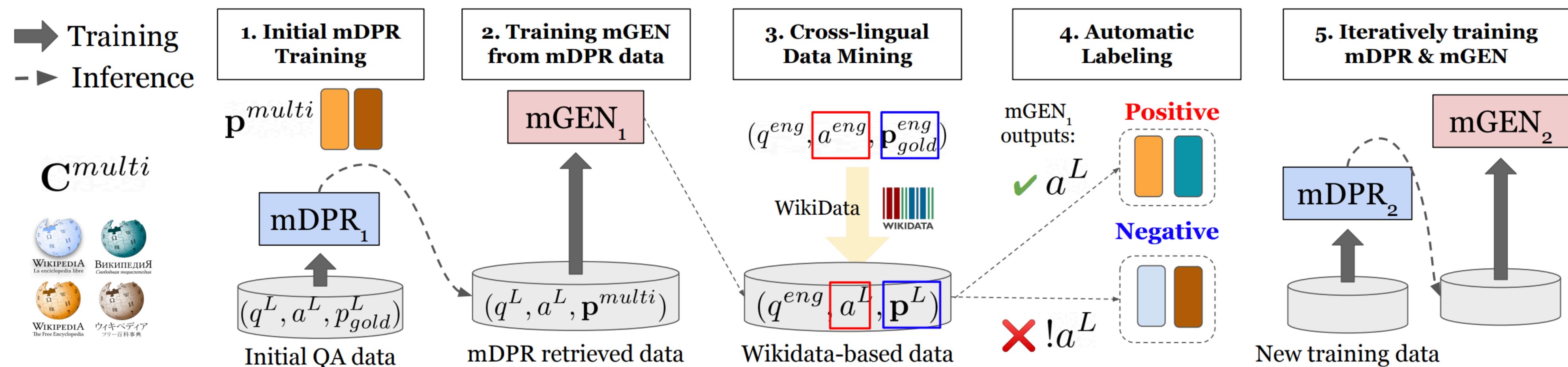
Initial training of retriever and LM

CORA: iteratively training multilingual LM & retriever

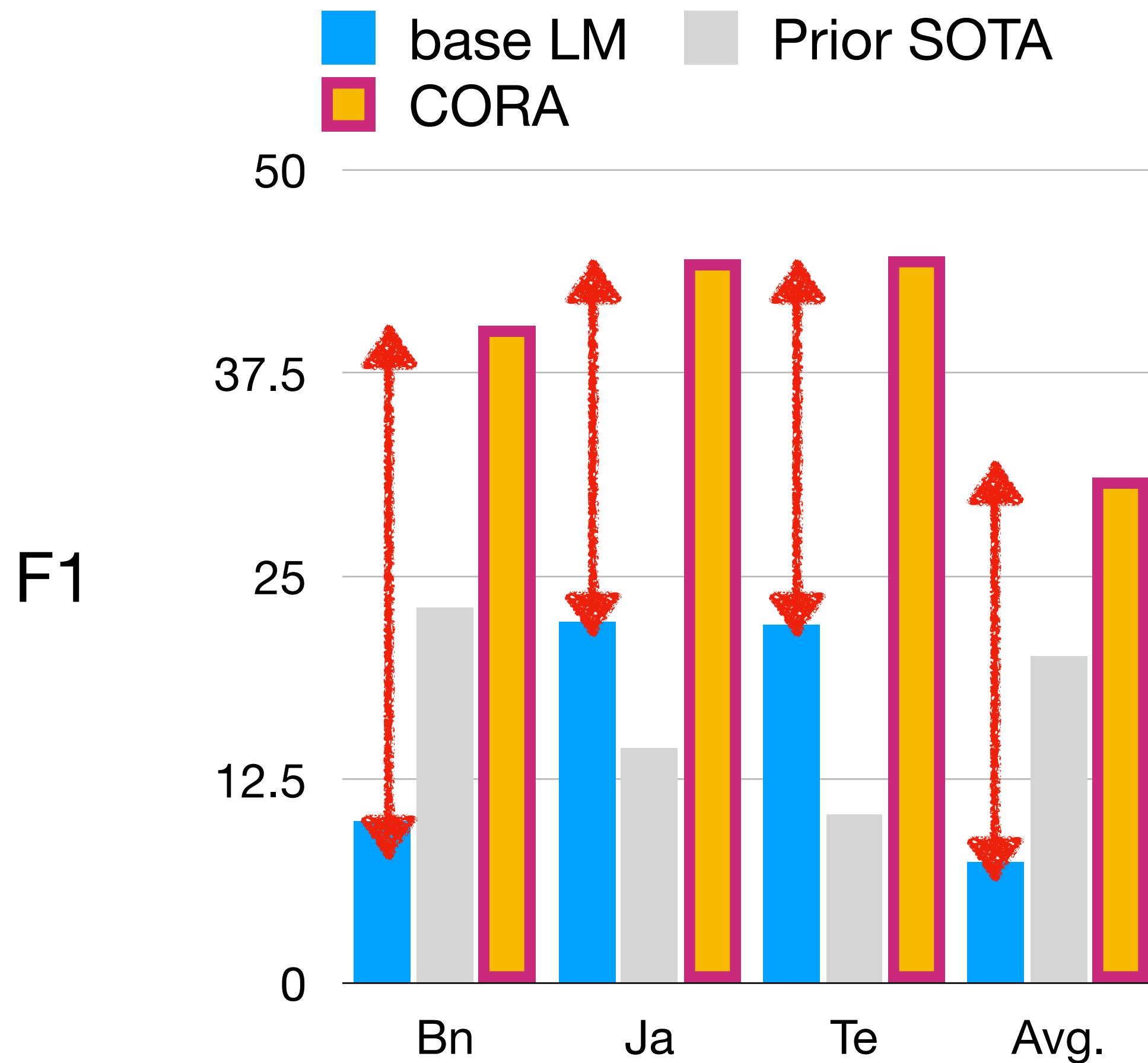


Expand training data using trained models
as well as structured cross-lingual data.

CORA: iteratively training multilingual LM & retriever

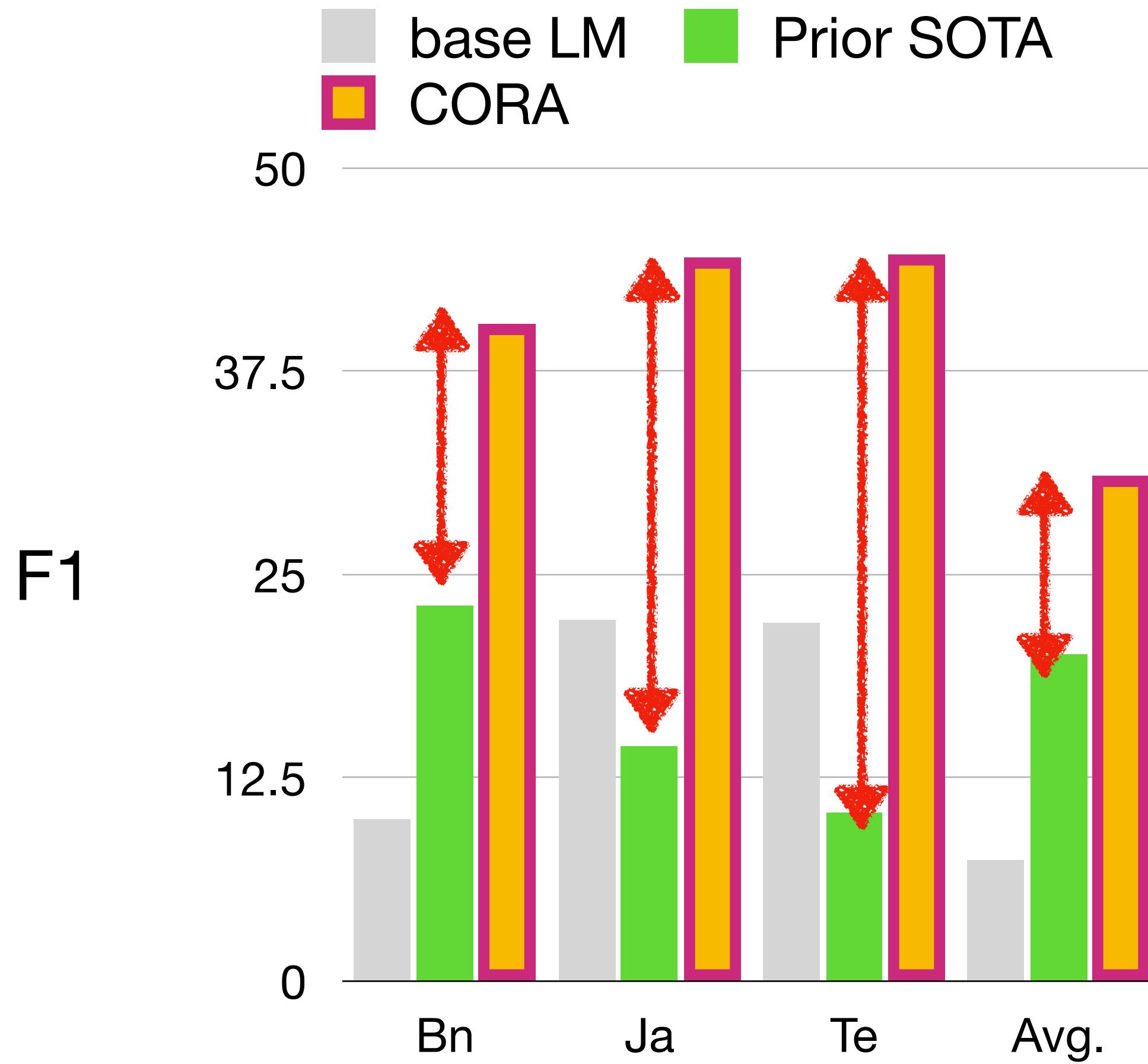


Results



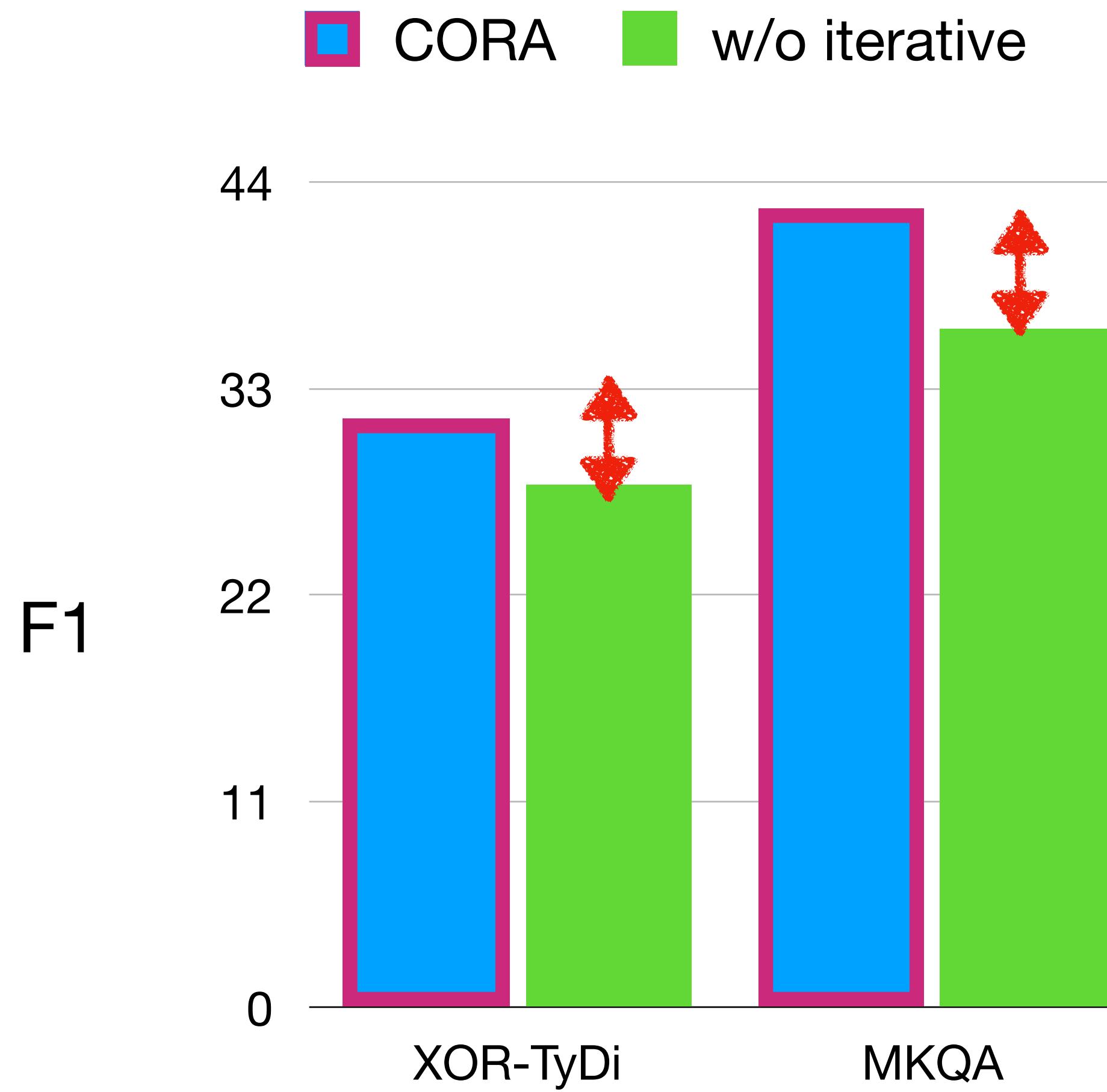
Large gains from fine-tuned LM
without retrieval

Results



Significantly Outperforms prior SOTA

Results



Iterative training of retriever and LM
gives large performance improvements

Multilingual retrieval-based LMs for diverse tasks

Question Answering

- * CL-ReLKT (Limkonchotiwat et al., 2022): knowledge transfer for better cross-lingual retrieval training
- * Gen-TyDi QA (Muller et al., 2023): generate full sentence answers for cross-lingual QA.

Fact Verification

- * CONCRETE (Hung et al., 2022): Improving cross-lingual fact-checking with cross-lingual retrieval

Dialogue

- * Cross-lingual Knowledge-grounded Dialogue (Kim et al 2021): a Korean knowledge-grounded dialogue system that learns to generate Korean response given English & Korean knowledge

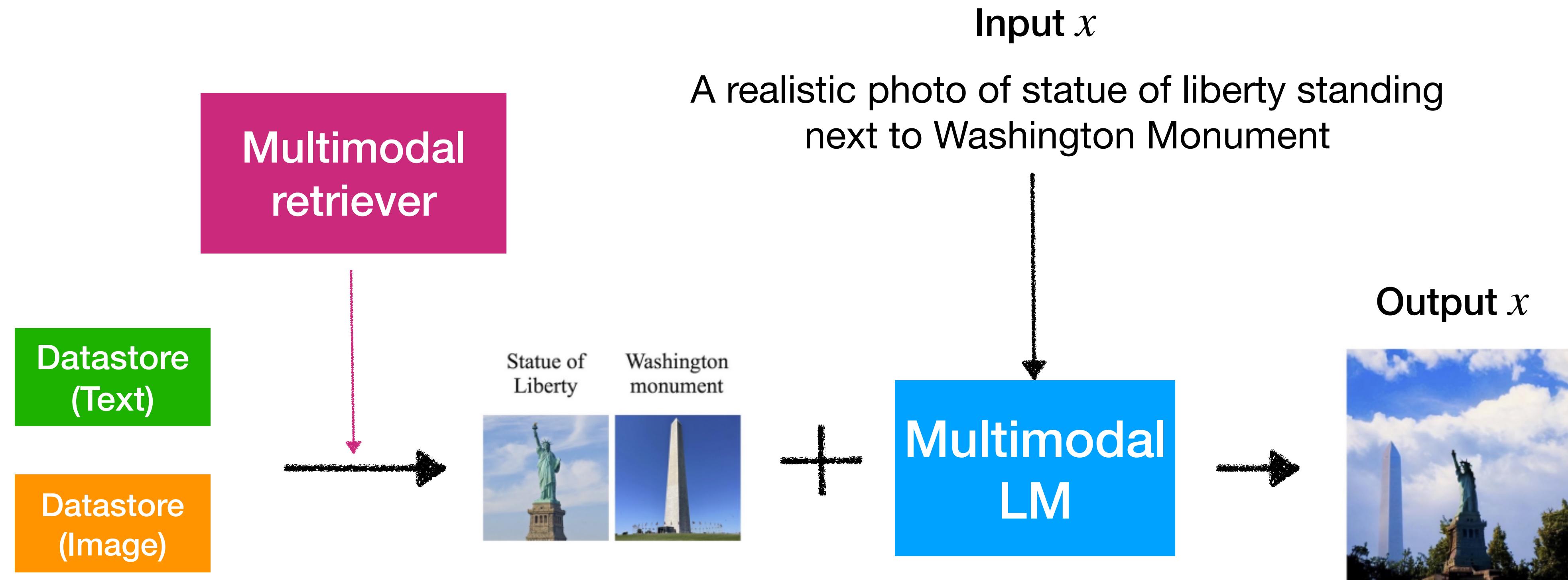
Event Extraction

- * R-GQA (Du and Ji, 2022): retrieve similar QA pairs for event argument extraction.

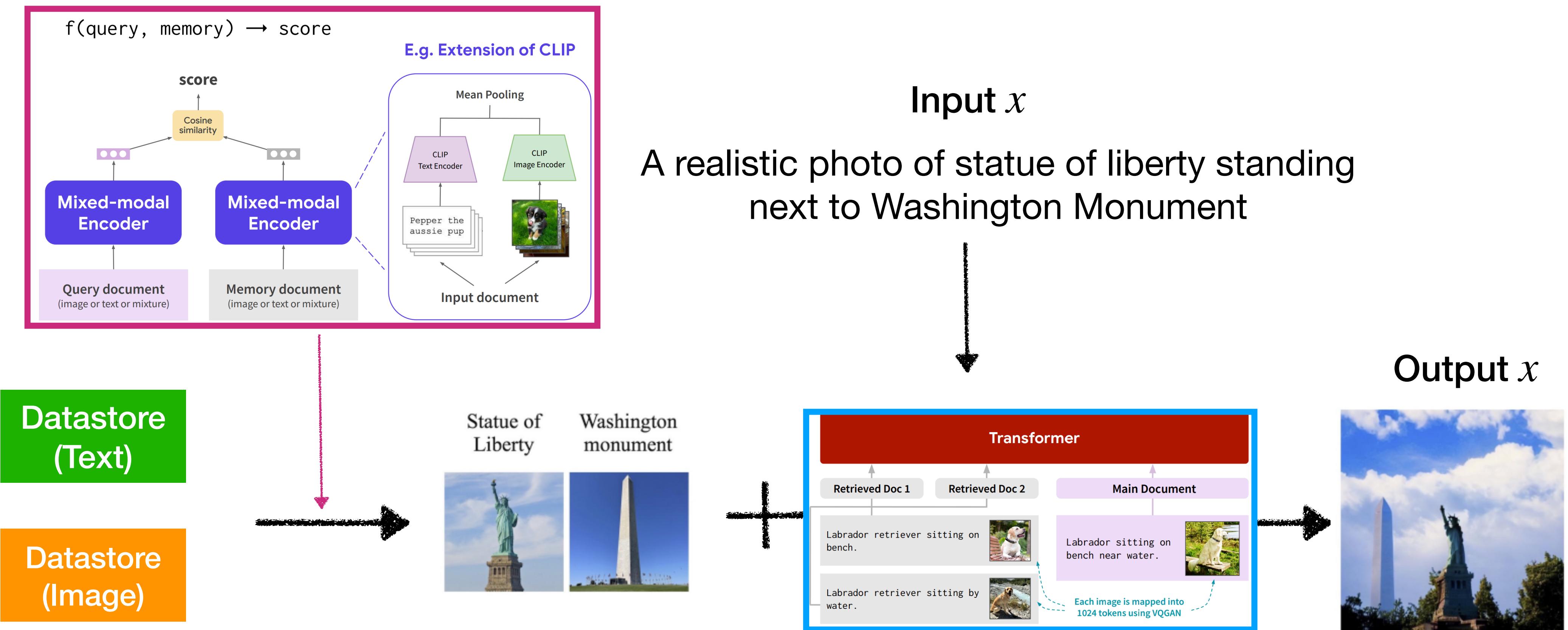
Key-phrase generations

- * Retrieval-augmented Multilingual Key phrase Generation (Gao et al 2022): Using iterative training to improve retrieval & LM for key phrase generations

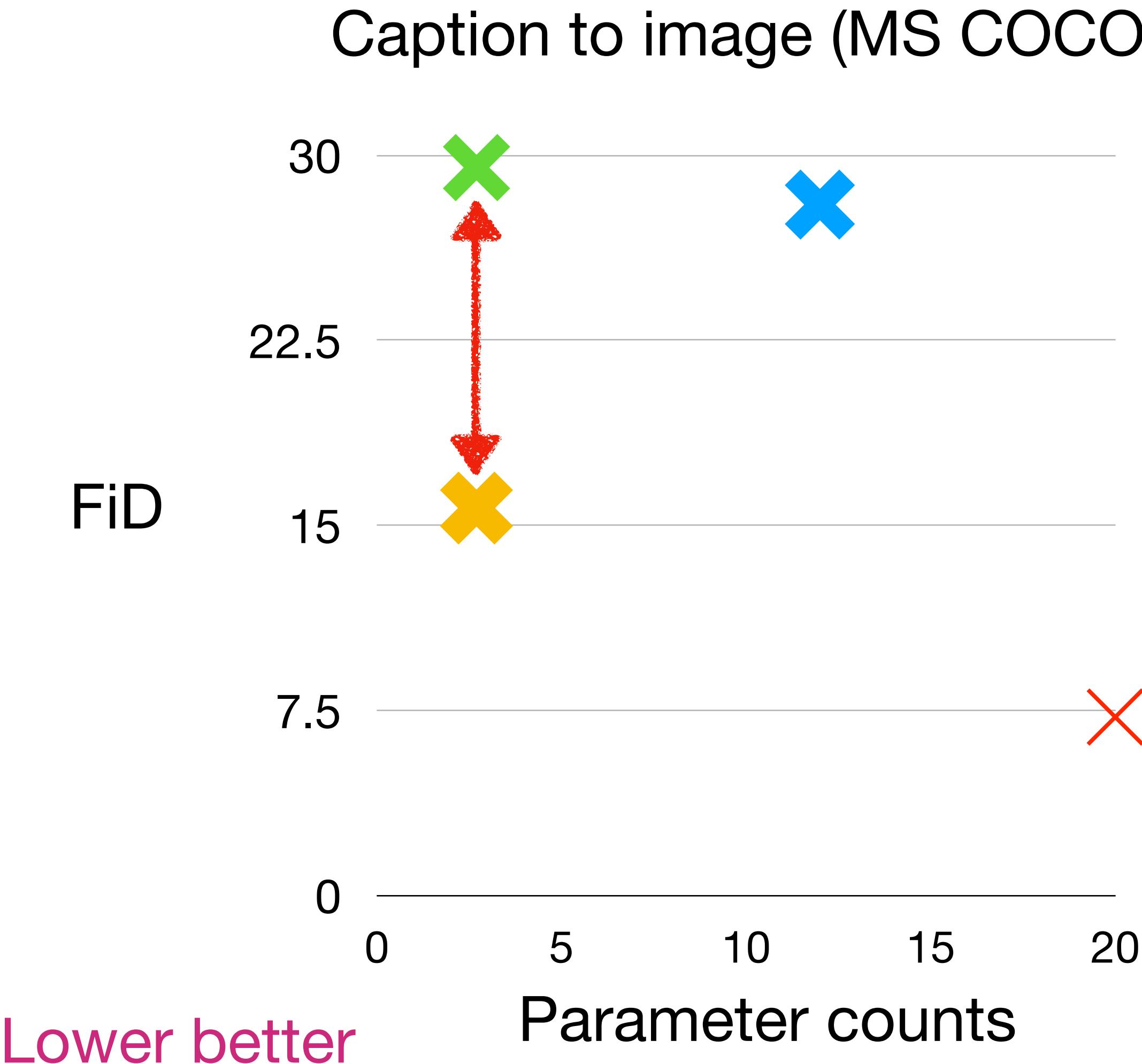
Multi-modal retrieval-augmented generations



RA-CM3 (Yasunaga et al., 2023)



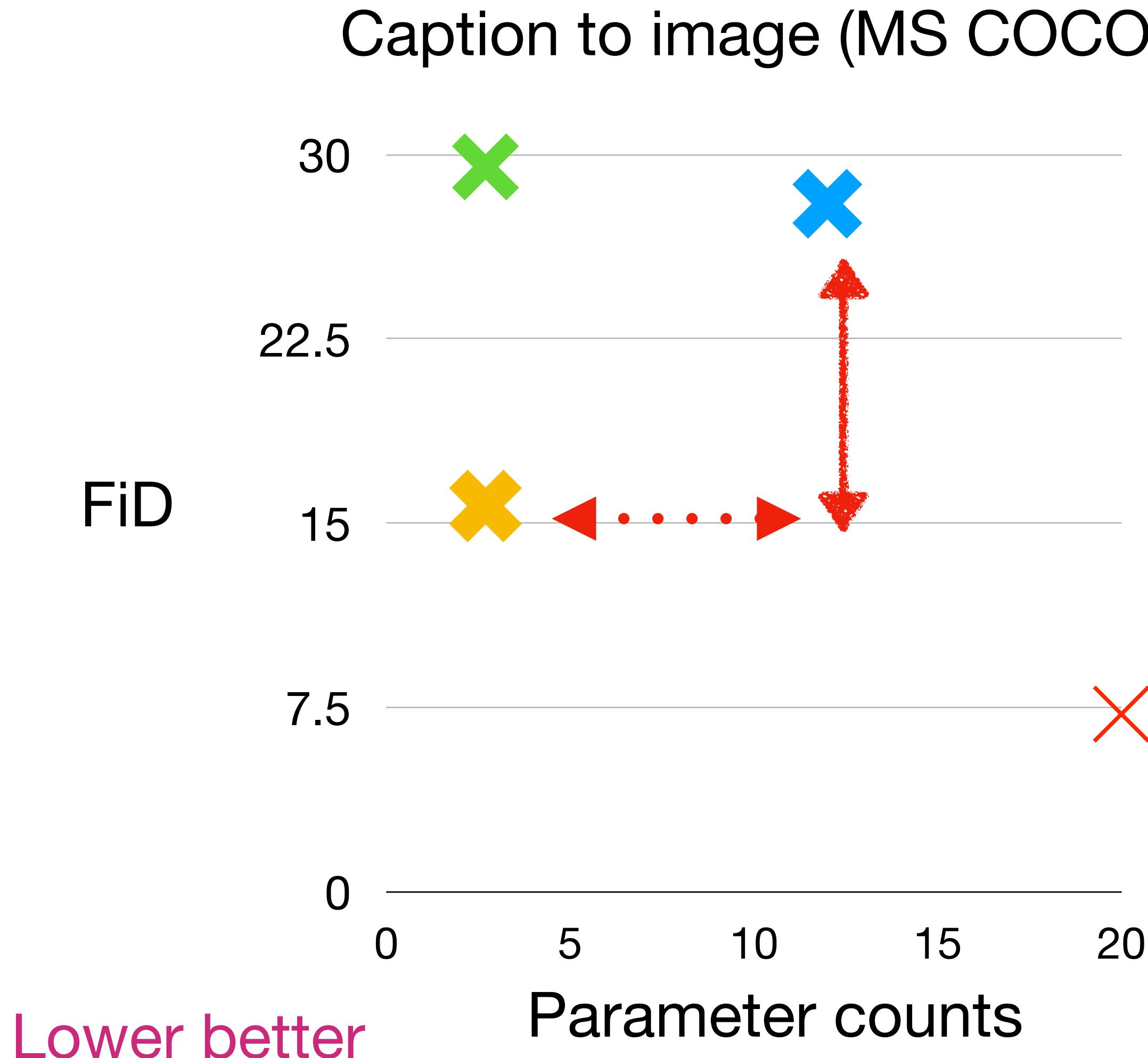
Results



Significantly outperforms base
LM (CM3)

- ✖ DALL-E
- ✖ Vanilla CM3
- ✖ RA-CM3
- ✖ Parti

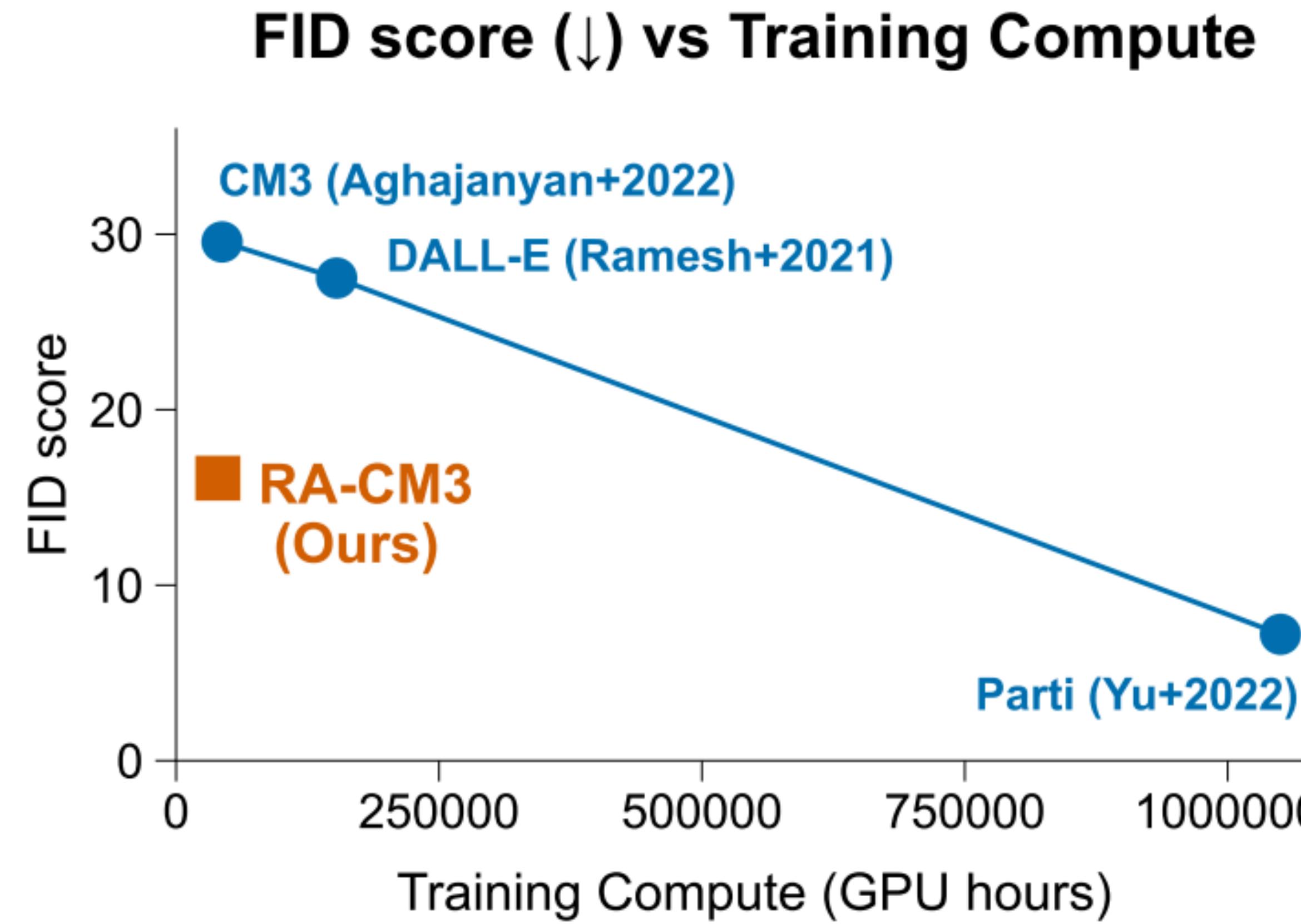
Results



Yasunaga et al. 2023. "Retrieval-Augmented Multimodal Language Modeling"

Results

Caption to image (MS COCO)



Retrieval-augmented generation achieves significantly better training efficiency

More applications beyond text

Multi-modal Retrieval-augmented Pre-training

- * RAVEAL (Hu et al 2023): Pretraining visual-language model using knowledge memory

Multi-modal Question Answering

- * MuRAG (Chen et al., 2022)

Multi-modal Classification

- * ALIGN (Gur et al., 2021)

Multimodal using image and text have been actively studied

More applications beyond text

Multi-modal Retrieval-augmented Pre-training

- * RAVEAL (Hu et al 2023): Pretraining visual-language model using knowledge memory

Multi-modal Question Answering

- * MuRAG (Chen et al., 2022)

Multi-modal Classification

- * ALIGN (Gur et al., 2021)

Retrieval-augmented training for molecules

- * Retrieval-based Molecule Generation (Wang et al., 2023)

Retrieval-augmented 3D motion generations

- * ReMoDiffus (Zhang et al., 2023)

New extensions for new input / output modality!

Wrapping up ...

The diagram illustrates the evolution of data from a raw text corpus to a knowledge base (KB) and finally to multimodal extensions. It consists of five main components arranged in a circle:

- Raw text corpus:** Shows a snippet of text about the Greater Toronto Area population in English, Spanish, and Japanese.
- KB:** A network graph where nodes represent entities and connections represent relationships, with the label "KB" at the center.
- Video:** A film strip icon with five small video frames showing people playing baseball.
- Images:** A map of the Great Lakes region and a small photo of a city street.
- Audio:** An audio waveform visualization.

Extension to multilingual

Cross-lingual retrieval and generation to overcome **datastore scarcity** in many world languages

Extension to multimodal

Key effectivenesses (Section 5; long-tail, efficiency) apply to diverse modality