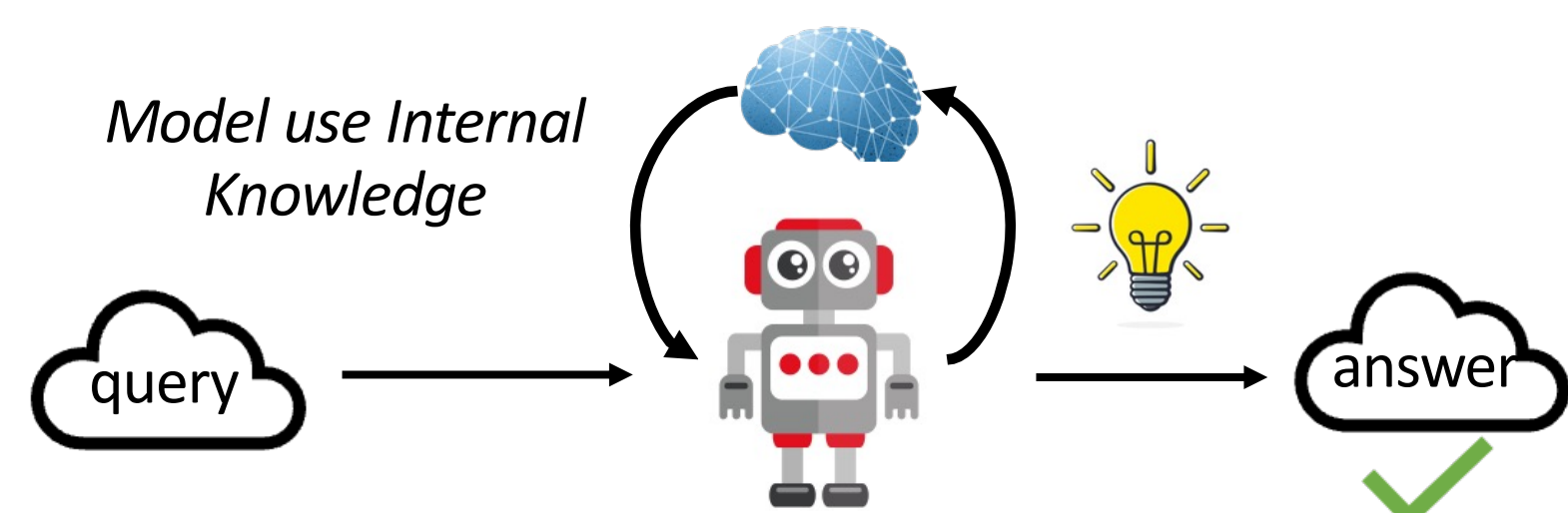


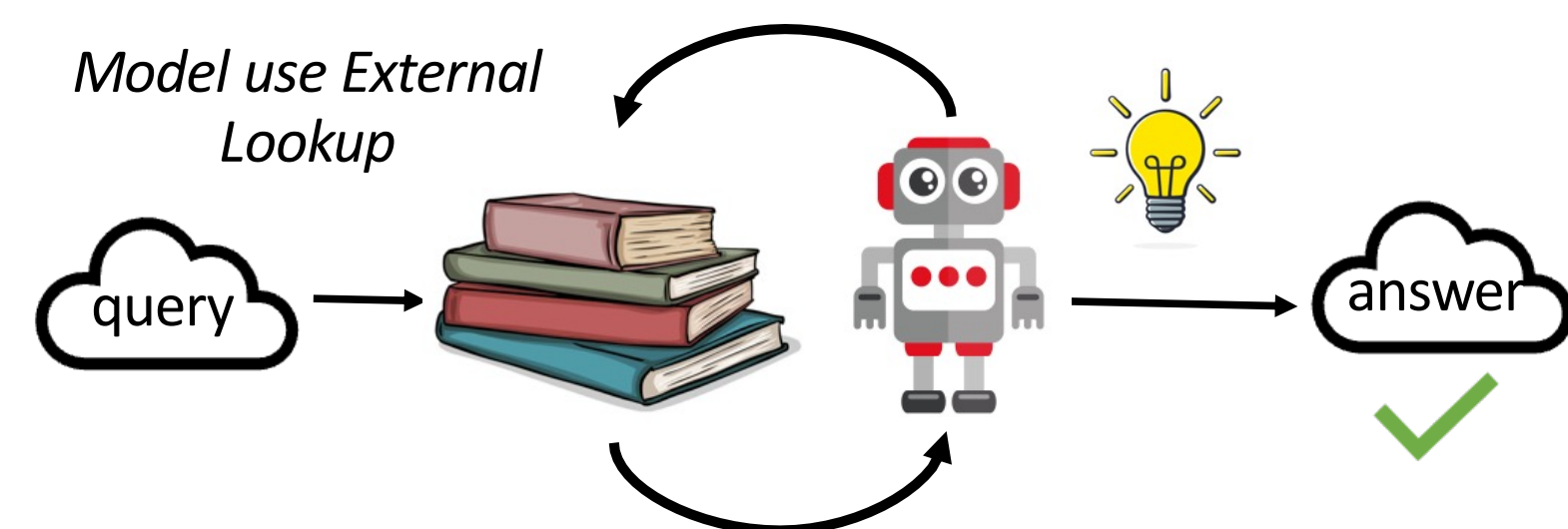
Overview

Open-book exam is an analogy referring to a student can answer the question based on :

- his/her knowledge
- search from website/books for reference

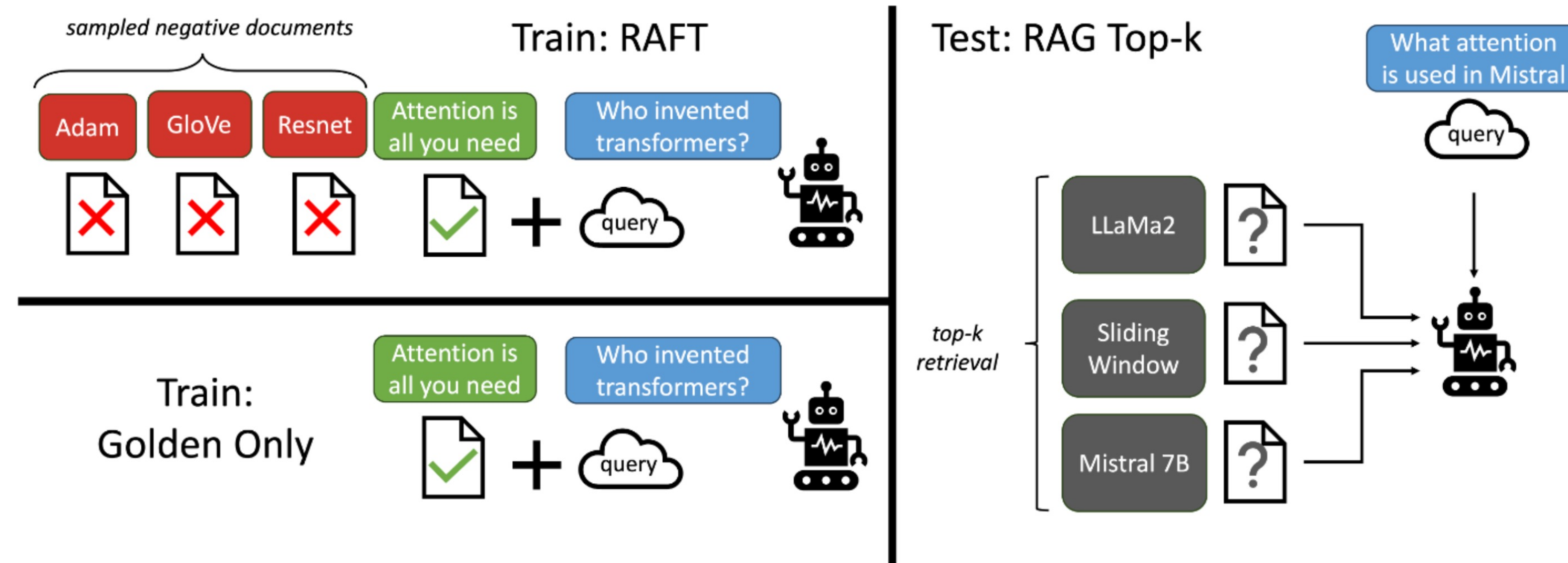


Memorize



Reading

RAFT: Retriever Aware Fine-Tuning



RAFT surpasses the base model LLaMM2-7B, Domain-Specific Finetuning and sometimes GPT-3.5

	PubMed	HotPot	HuggingFace	Torch Hub	TensorFlow Hub
GPT-3.5 + RAG	71.60	41.5	29.08	60.21	65.59
LLaMA2-7B	56.5	0.54	0.22	0	0
LLaMA2-7B + RAG	58.8	0.03	26.43	08.60	43.06
DSF	59.7	6.38	61.06	84.94	86.56
DSF + RAG	71.6	4.41	42.59	82.80	60.29
RAFT (LLaMA2-7B)	73.30	35.28	74.00	84.95	86.86

Results of RAFT on Medical (PubMed), General-knowledge (HotPotQA), and API (Gorilla) benchmarks.

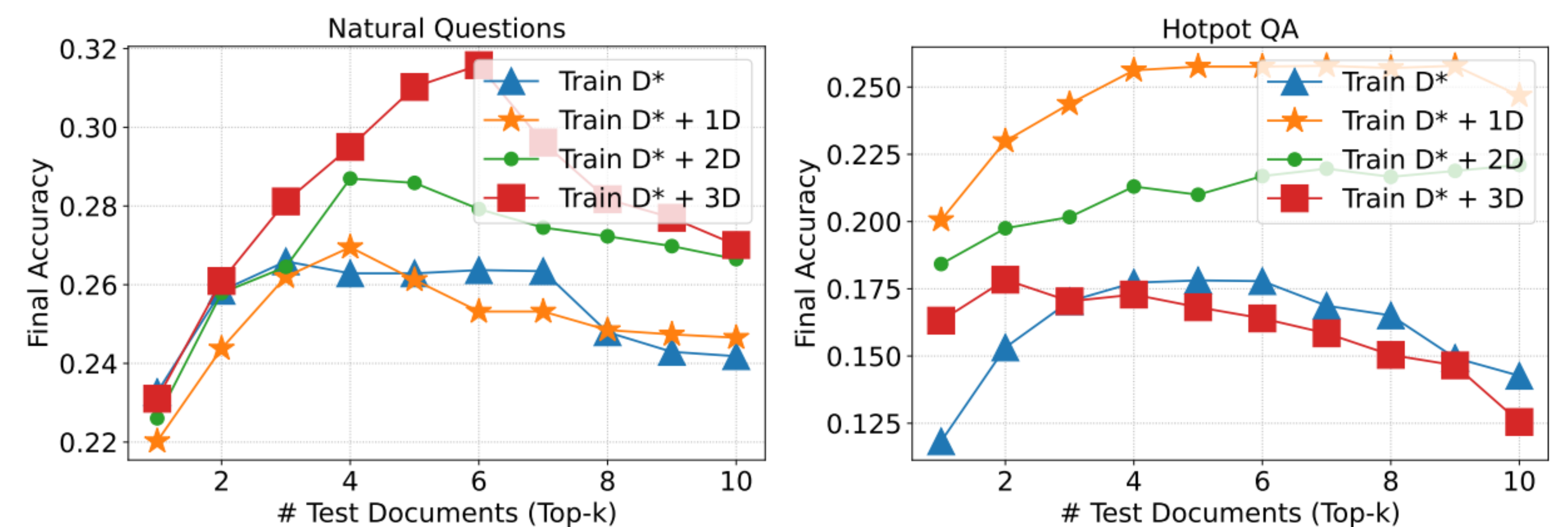
Lesson 1: Importance of CoT

	PubMed	HotpotQA	HuggingFace	Torch Hub	TensorFlow Hub
RAFT w.o CoT	68.30	25.62	59.07	86.56	83.21
RAFT	73.30	35.28	74.00	84.95	86.86

We study the importance of Chain-of-Thought (CoT) and citation from the original context. We found that adding CoT to our method gives a significant performance boost.

CoT Answer: ##Reason: The document ##begin_quote## The Oberoi family is an Indian family that is famous for its involvement in hotels, namely through The Oberoi Group. ##end_quote## establishes that the Oberoi family is involved in the Oberoi group, and the document ##begin_quote## The Oberoi Group is a hotel company with its head office in Delhi. ##end_quote## establishes the head office of The Oberoi Group. Therefore, the Oberoi family is part of a hotel company whose head office is in Delhi. ##Answer: Delhi

Lesson 2: Robust to Top-k Documents



We found that it is always beneficial to add distractor documents to our dataset. This is because at test time the model is being asked to read and extract from top-k documents, adding distractor documents will make the model learn to ignore irrelevant context.