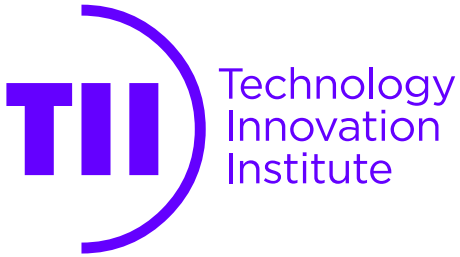


The Distracting Effect

Understanding Irrelevant Passages in RAG

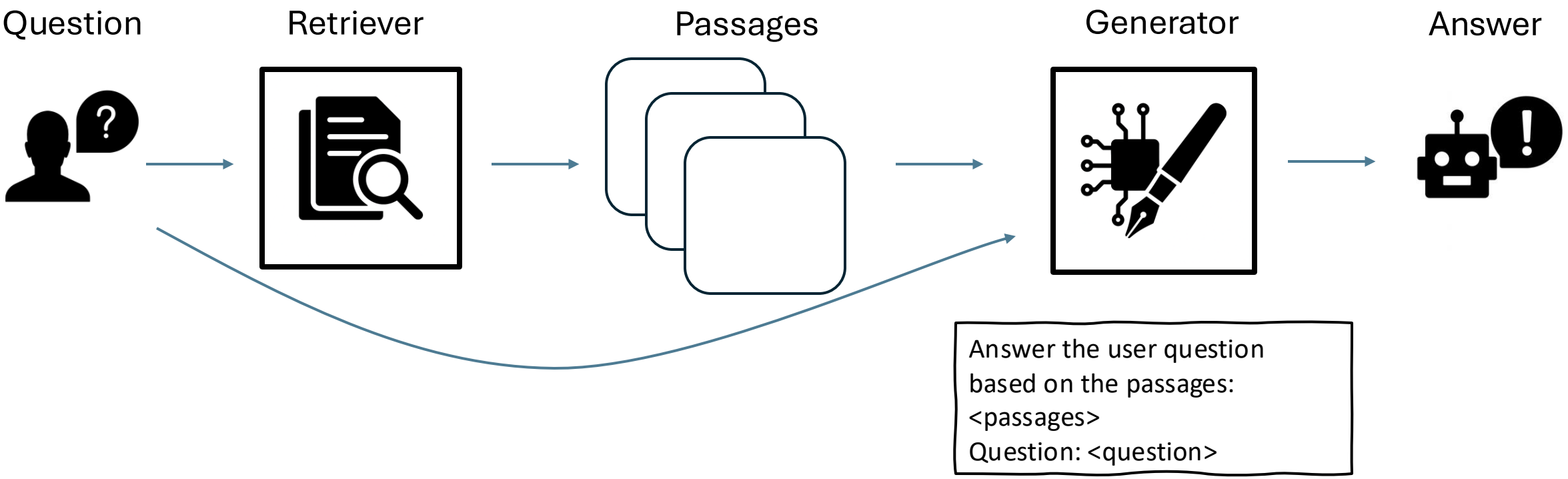
Chen Amiraz[1], Florin Cuconasu[1,2], Simone Filice[1], and Zohar Karnin[1]

Technology Innovation Institute, Haifa, Israel[1]
La Sapienza University, Rome, Italy[2]

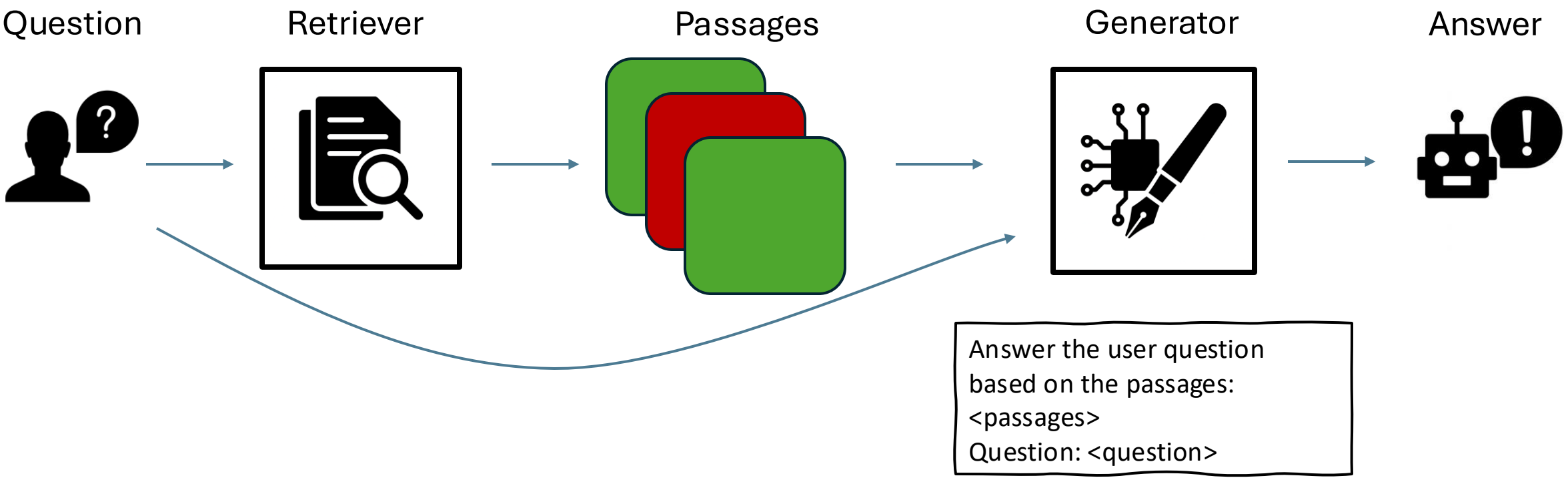
Vanilla Generation

Limited when the user question is information-seeking

Retrieval Augmented Generation (RAG)



Retrieval Augmented Generation (RAG)



Problem: *Distracting passages* that contain semantically related yet irrelevant information may fail the answer-generating LLM

Example – Relevant + Irrelevant Passages

Question: What movie featured the song on the road again?

Gold Answer: **Honeysuckle Rose**

Relevant passage
"… The song 'On the Road Again' came about when the producer of the film '**Honeysuckle Rose**' approached Willie Nelson about writing the song for the film's soundtrack …"

What happens if the LLM gets in addition an **irrelevant** passage?

Example – Relevant + Irrelevant Passages

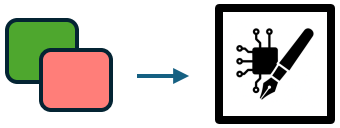
Question: What movie featured the song on the road again?

Gold Answer: **Honeysuckle Rose**

Relevant passage
"… The song 'On the Road Again' came about when the producer of the film '**Honeysuckle Rose**' approached Willie Nelson about writing the song for the film's soundtrack …"

Irrelevant Passage #1

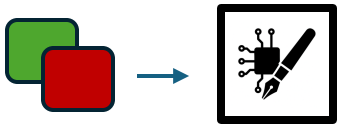
"… Country music legend Willie Nelson's iconic song '**Always on My Mind**' was featured in the 1982 film '**The Soldier**'…"



Generated Answer:
Honeysuckle Rose

Irrelevant Passage #2

"… Many people believe, though it's **not actually correct**, that Willie Nelson's song 'On The Road Again' first appeared in the 1980 film '**Smokey and the Bandit II**'…"



Generated Answer:
Smokey and the Bandit II

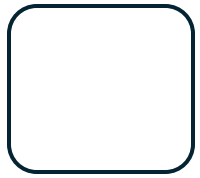
Our paper: First comprehensive framework for quantifying, obtaining, and utilizing hard distracting passages in RAG

Quantifying the distracting effect

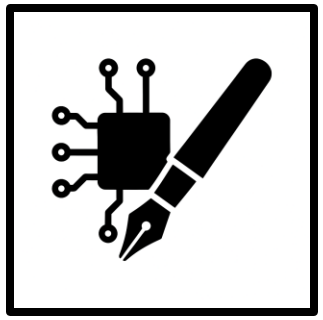
Distracting effect – Definition

Question

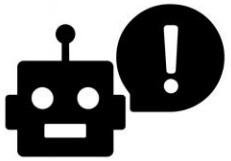
Irrelevant Passage



Fixed Generator



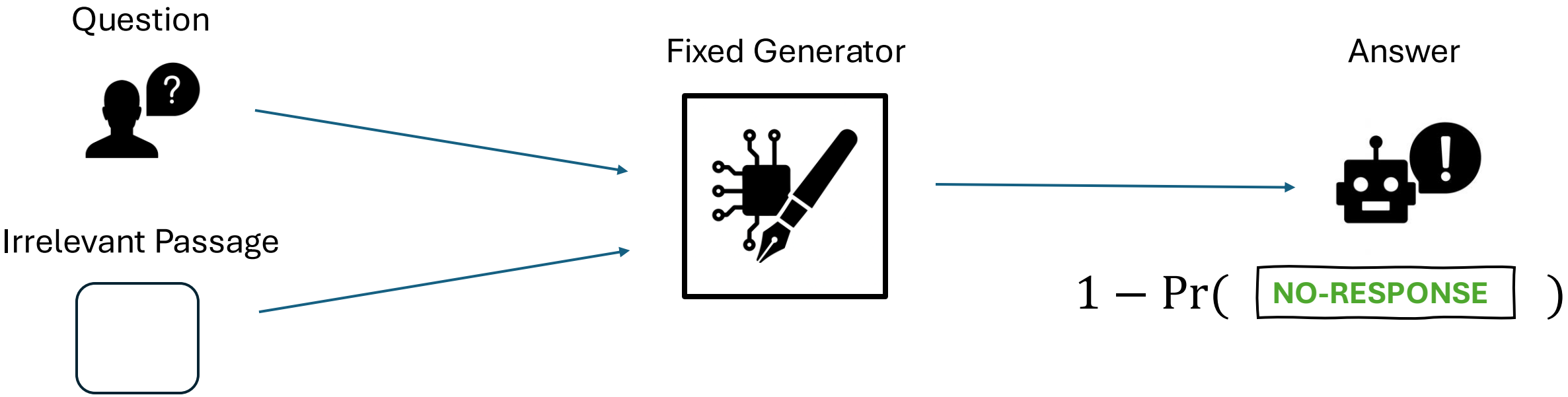
Answer



$1 - \Pr(\ \boxed{\text{NO-RESPONSE}}\ )$

Answer the question based on the passage. **If the passage does not contain the answer, please respond with NO-RESPONSE. Do not try to respond based on your own knowledge.**
Passage: <passage>
Question: <question>

Distracting effect – Definition



The ***distracting effect*** of the passage p on the LLM w.r.t. the question q is the probability of not abstaning, i.e.,

$$\mathbf{DE_q(p)} = 1 - \Pr_{LLM}(\text{NO–RESPONSE}|q,p)$$

Distracting effect – Example

The ***distracting effect*** of the passage p on the LLM w.r.t. the question q is the probability of not abstaning, i.e.,

$$\mathbf{DE_q(p)} = 1 - \Pr_{LLM}(\text{NO-RESPONSE}|q,p)$$

Question: What movie featured the song on the road again?

Gold Answer: **Honeysuckle Rose**

Irrelevant Passage #1

"... Country music legend Willie Nelson's iconic song '**Always on My Mind**' was featured in the 1982 film '**The Soldier**'..."

Distracting Effect: 0.34

Irrelevant Passage #2

"... Many people believe, though it's **not actually correct**, that Willie Nelson's song 'On The Road Again' first appeared in the 1980 film '**Smokey and the Bandit II**'..."

Distracting Effect: 1.0

For each query, the distracting effect provides a ranking of the passages

Our contributions

The ***distracting effect*** of p on the LLM w.r.t. q is the probability of not abstaning, i.e.,

$$\mathbf{DE_q(p)} = 1 - \Pr_{LLM}(\text{NO–RESPONSE}|q, p)$$

Why this definition?

- ✓ Naturally continuous and bounded [0,1]
- ✓ Interpretable
- ✓ Simple & cheap to implement
- ✓ Independent of other passages

- ✓ Distracting effect translates to RAG accuracy
- ✓ Consistent across LLMs
- ✓ Methods for obtaining diverse distracting passages
- ✓ Application: Increasing robustness by distraction-based fine-tuning

Distracting effect – Example

Question: What movie featured the song on the road again?

Gold Answer: **Honeysuckle Rose**

Relevant passage
"… The song 'On the Road Again' came about when the producer of the film '**Honeysuckle Rose**' approached Willie Nelson about writing the song for the film's soundtrack …"

Irrelevant Passage #1

"… Country music legend Willie Nelson's iconic song '**Always on My Mind**' was featured in the 1982 film '**The Soldier**'…"

Distracting Effect: 0.34

Generated Answer:
Honeysuckle Rose

Irrelevant Passage #2

"… Many people believe, though it's **not actually correct**, that Willie Nelson's song 'On The Road Again' first appeared in the 1980 film '**Smokey and the Bandit II**'…"

Distracting Effect: 1.0

Generated Answer:
Smokey and the Bandit II

Passages with high distracting effect fail the answer-generating LLM

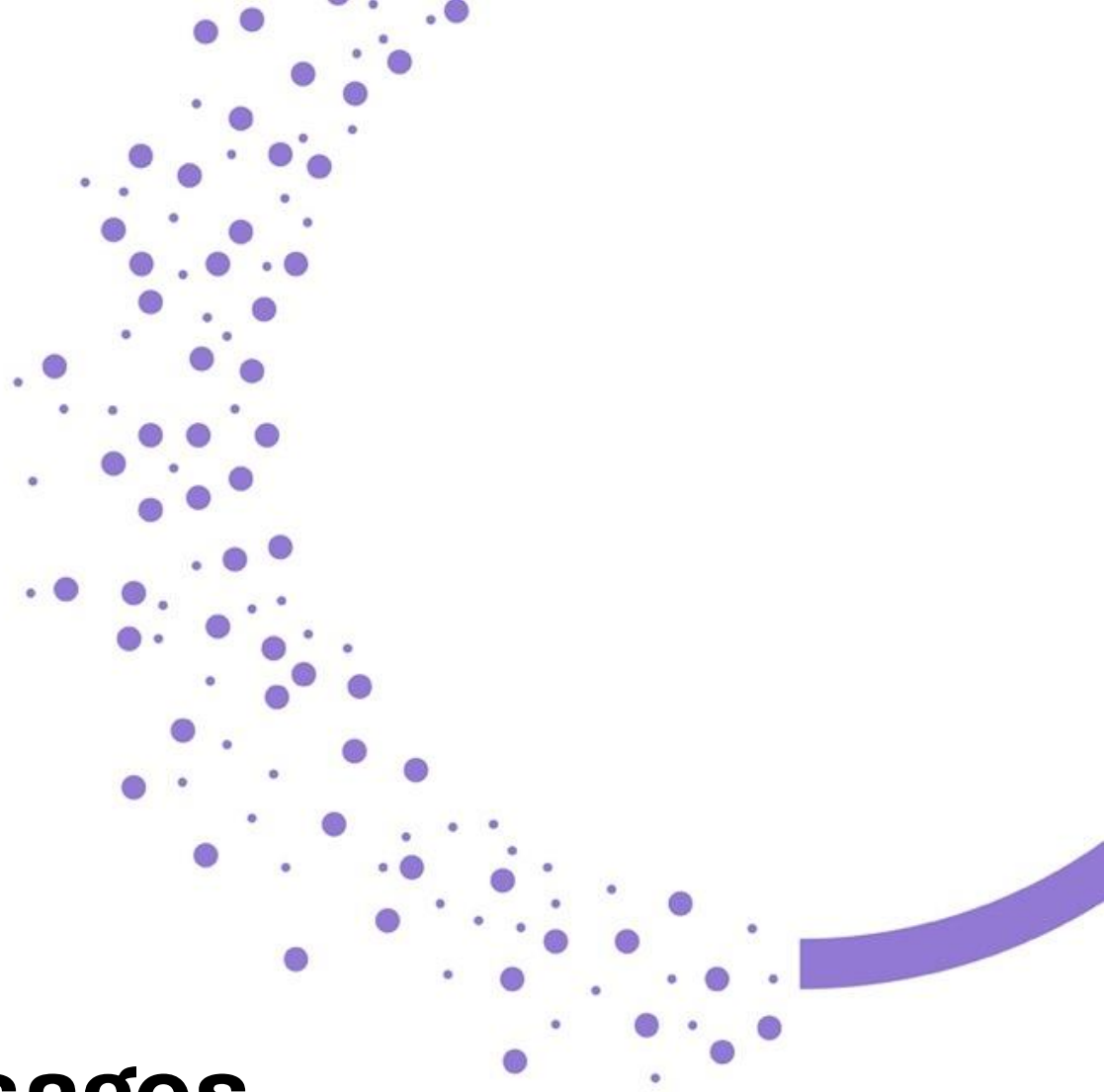
Impact on downstream RAG accuracy

WD := DE<0.2

HD := DE>0.8

LLM	Only Gold	Gold + WD	Gold + HD
Llama-3.2-3B	82.6	79.4	**71.5**
Llama-3.1-8B	80.6	80.1	**73.9**
Llama-3.3-70B	81.1	80.1	**75.2**
Falcon-3-3B	78.5	74.1	**67.1**
Falcon-3-7B	84.1	81.5	**73.3**
Qwen-2.5-3B	80.9	75.5	**69.4**
Qwen-2.5-7B	82.4	80.4	**73.7**

Hard distractors hurt accuracy more than weak ones, leading to accuracy drop of 6-11 accuracy points

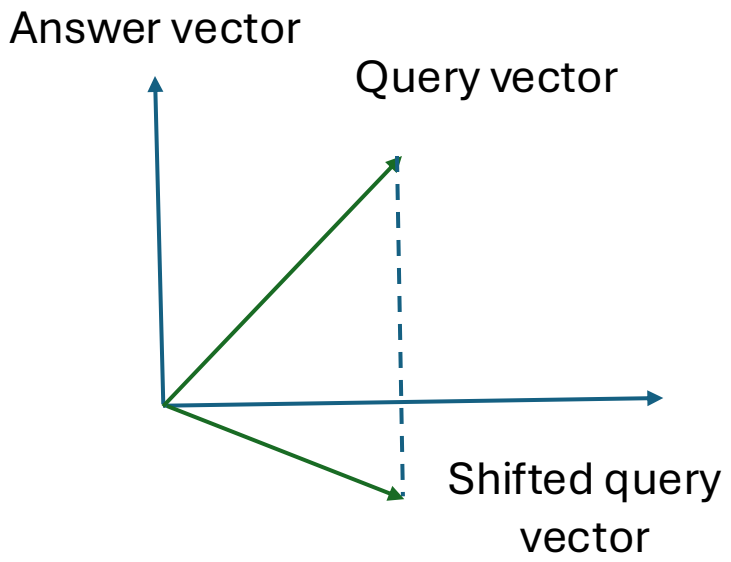


Obtaining distracting passages

Methods for obtaining distracting passages

Retrieval (+ Reranking)

1. Standard
2. Answer-skewed

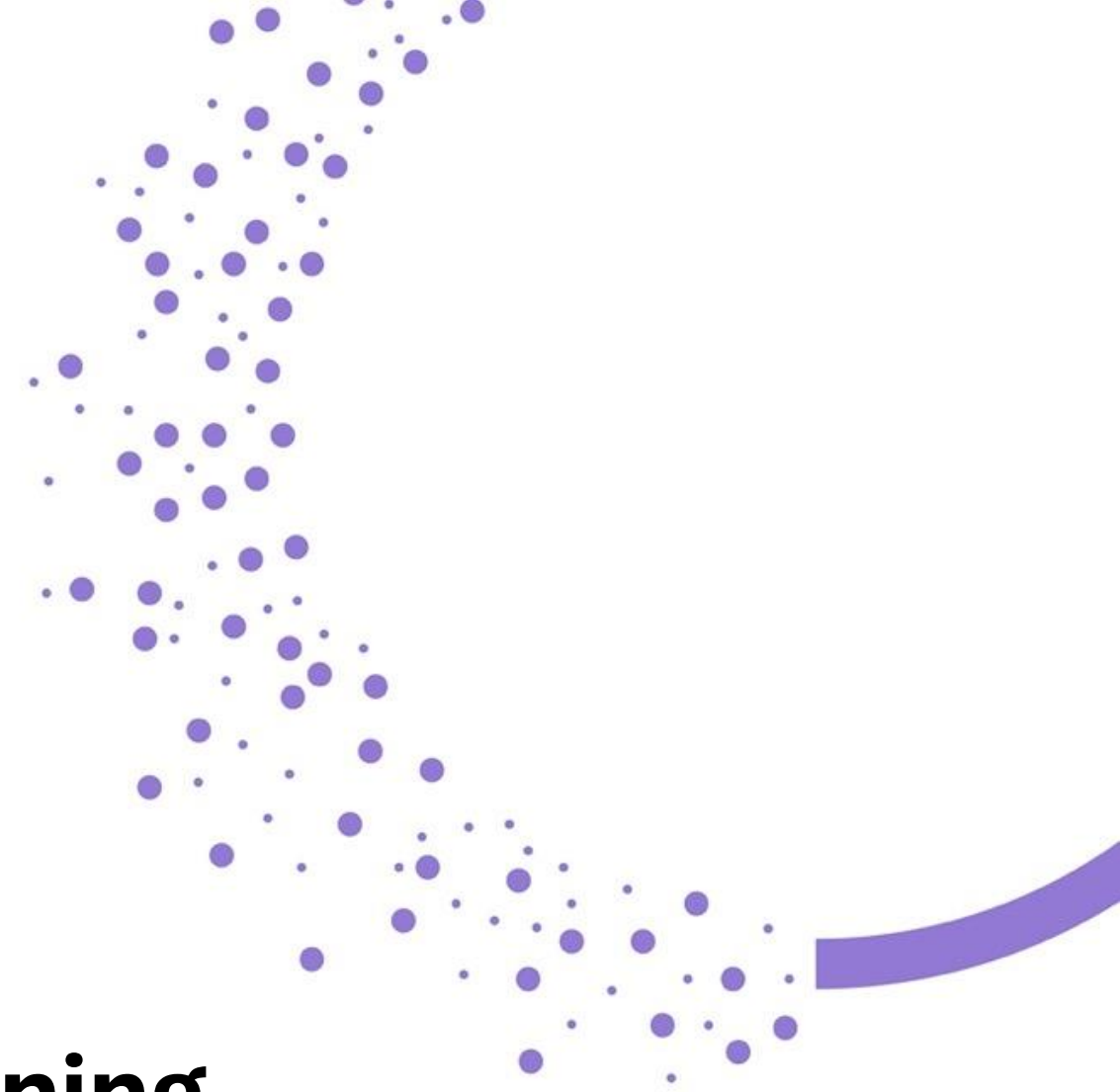


Generation

For each category, we used few-shot learning with a strong LLM:

1. Negation
2. Hypothetical
3. Modal
4. Related topic

No universal winner: Methods excel on different queries



Application: Distraction-based fine-tuning

Training on hard distractors improves accuracy

Test Set	*None*	*Retrieve*	*Rerank*	*Hard*
NQ	37.9	40.7	39.7	**42.8**
PopQA	35.9	40.1	39.1	**43.2**
TriviaQA	67.8	67.6	64.7	**74.5**
WebQA	41.9	42.1	41.3	**49.7**

Distraction-based fine-tuning yields up to 7.5% accuracy gains compared to standard methods

Understanding the distracting effect is a key step towards robust RAG

The Distracting Effect: Understanding Irrelevant Passages in RAG

Chen Amiraz[1*], Florin Cuconasu[1,2†‡], Simone Filice[1§], Zohar Karnin[1¶]

[1]Technology Innovation Institute, [2]Sapienza University of Rome

Chen.Amiraz@gmail.com

Full paper: