# ReSCORE: Label-free Iterative Retriever Training for Multi-hop Question Answering with Relevance-Consistency Supervision

Dosung Lee, Wonjun Oh, Boyoung Kim, Minyoung Kim,
Joonsuk Park, Paul Hongsuck Seo

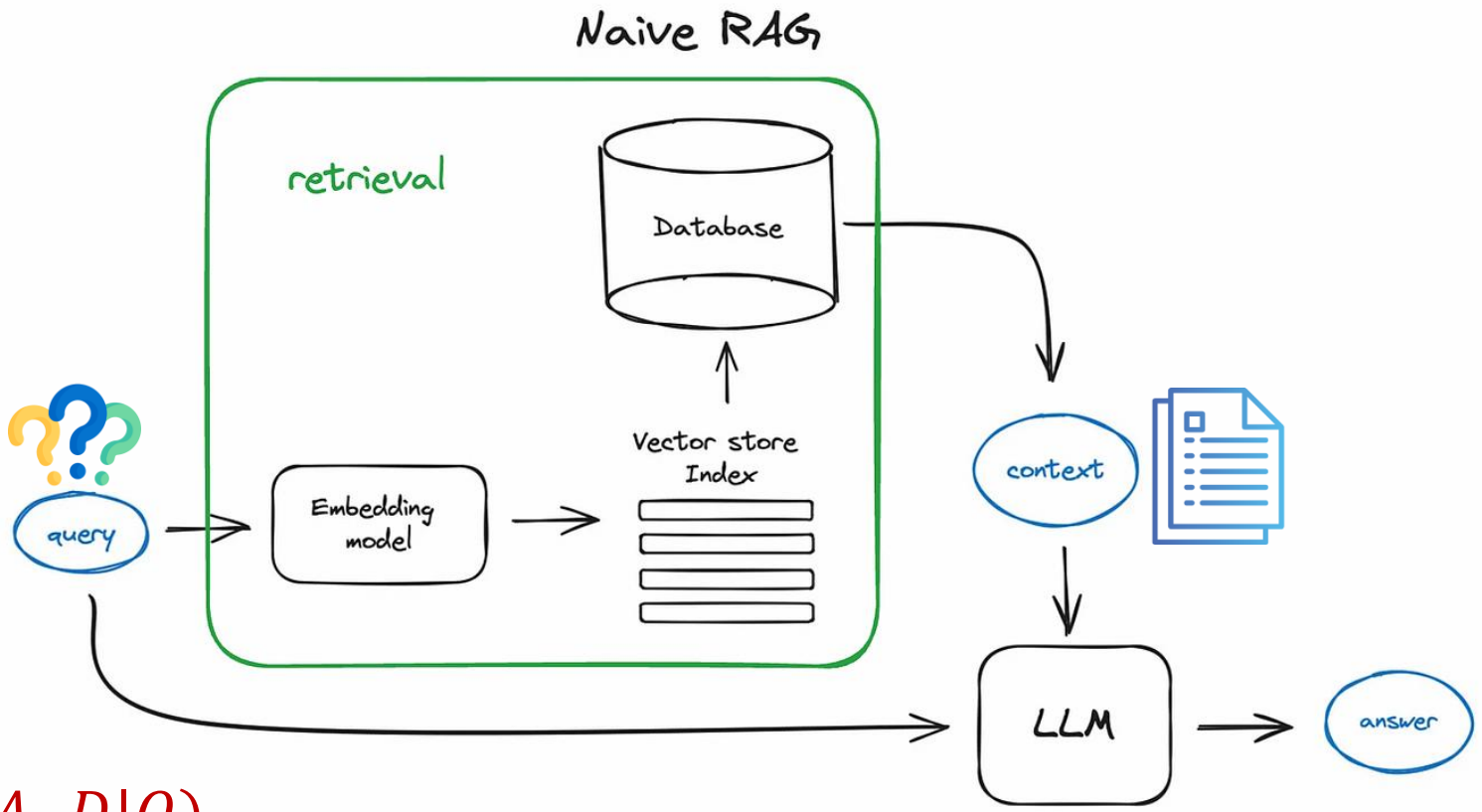
Korea University, NAVER AI Lab, NAVER Cloud, University of Richmond

{dslee1219, owj0421, bykimby, omniverse186, phseo}@korea.ac.kr, park@joonsuk.org


KOREA UNIVERSITY

MIIL Multimodal Interactive Intelligence Laboratory

# Retrieval Augmented Generation



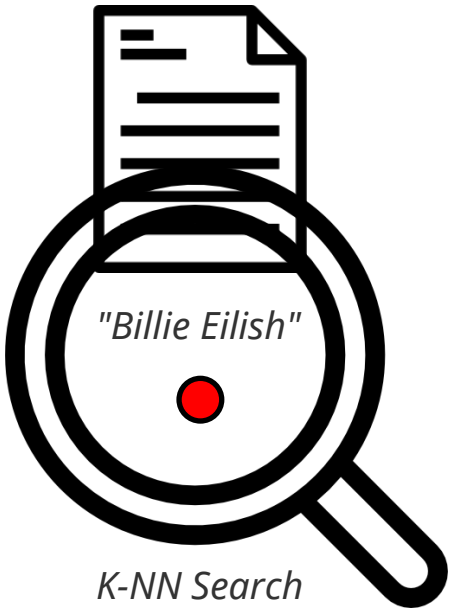

$$P_{QA}(A|Q) = \sum_{D} P(A, D|Q)$$

$$P_{RAG}(A, D|Q) = P_{generate}(A|D, Q)\, P_{retrieve}(D|Q)$$

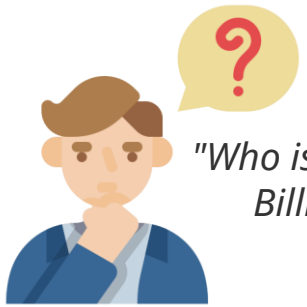
# Multi-hop Question Answering & Challenges

## *Single-hop*

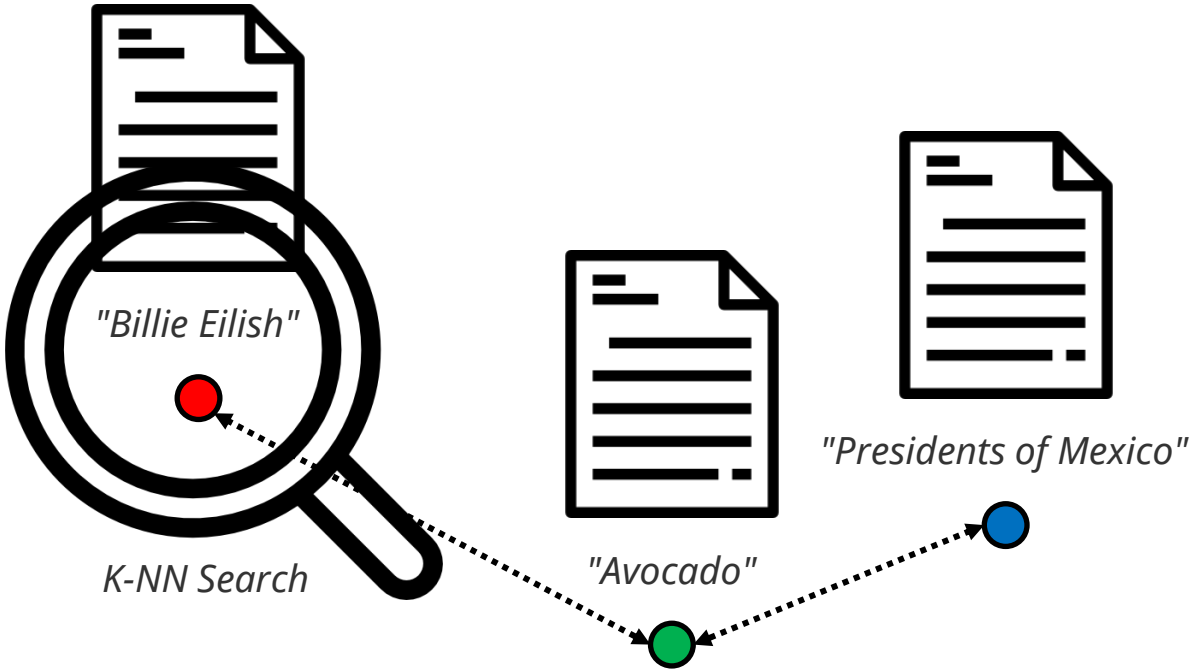
*“What is Billie Eilish’s favorite food?”*



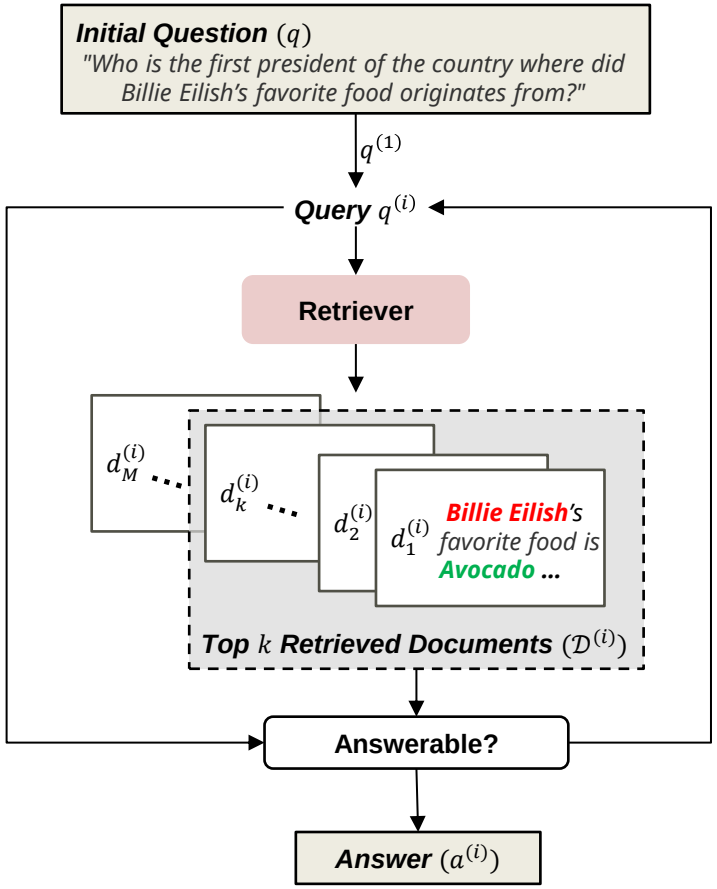

## *Multi-hop*

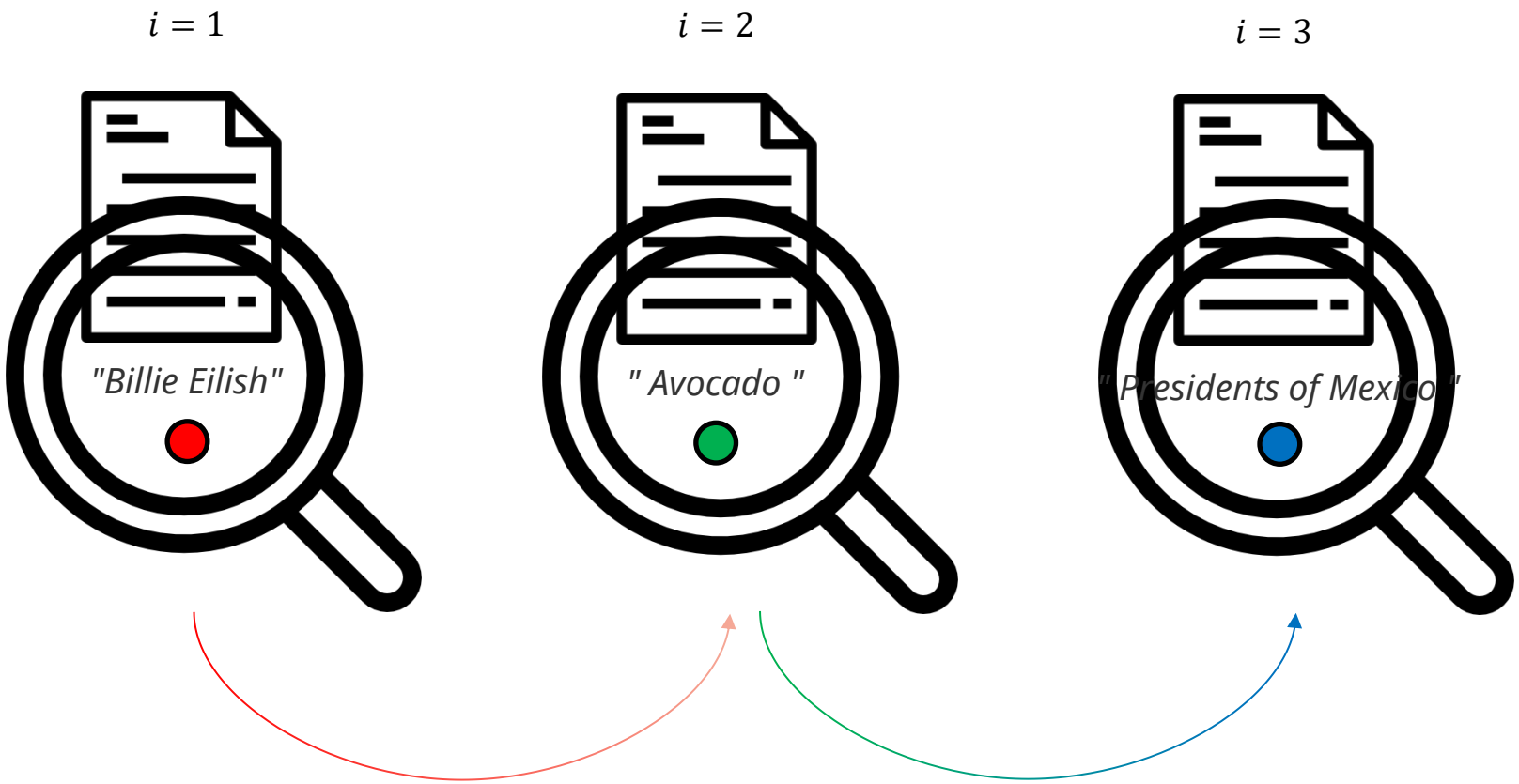


*"Who is the first president of the country where did Billie Eilish’s favorite food originates from?"*



# Iterative RAG Inference Framework



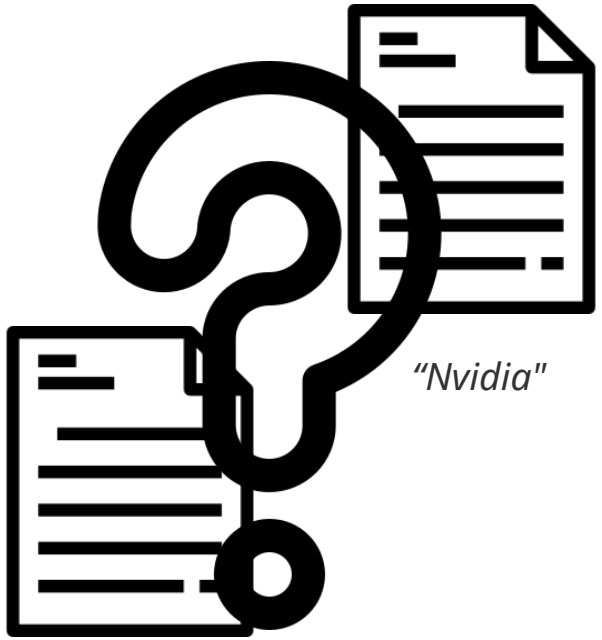



# Challenges of Labeling MHQA & Label-free Training

*"Who is the first president of the country where did Billie Eilish's favorite food originates from?"*

*"Billie Eilish"*

*"Elon Musk"*



*"Nvidia"*

*"South Korea"*

*" Presidents of Mexico "*

*" Avocado"*

*"Paris"*

# Relevance-Consistency Supervision

$$P_{QA}(A|Q) = \sum_{D} P(A, D|Q)$$

$$P_{RAG}(A, D|Q) = P_{generate}(A|D, Q)\, P_{retrieve}(D|Q)$$

$$P_{Retriever}(D|Q) \propto P_{LLM}(A, Q|D)$$

$$P_{LLM}(A, Q|D) = P_{Consistency}(A|D, Q)\, P_{Relevance}(Q|D)$$

*"Who is the first president of the country where did Billie Eilish's favorite food originates from?"*

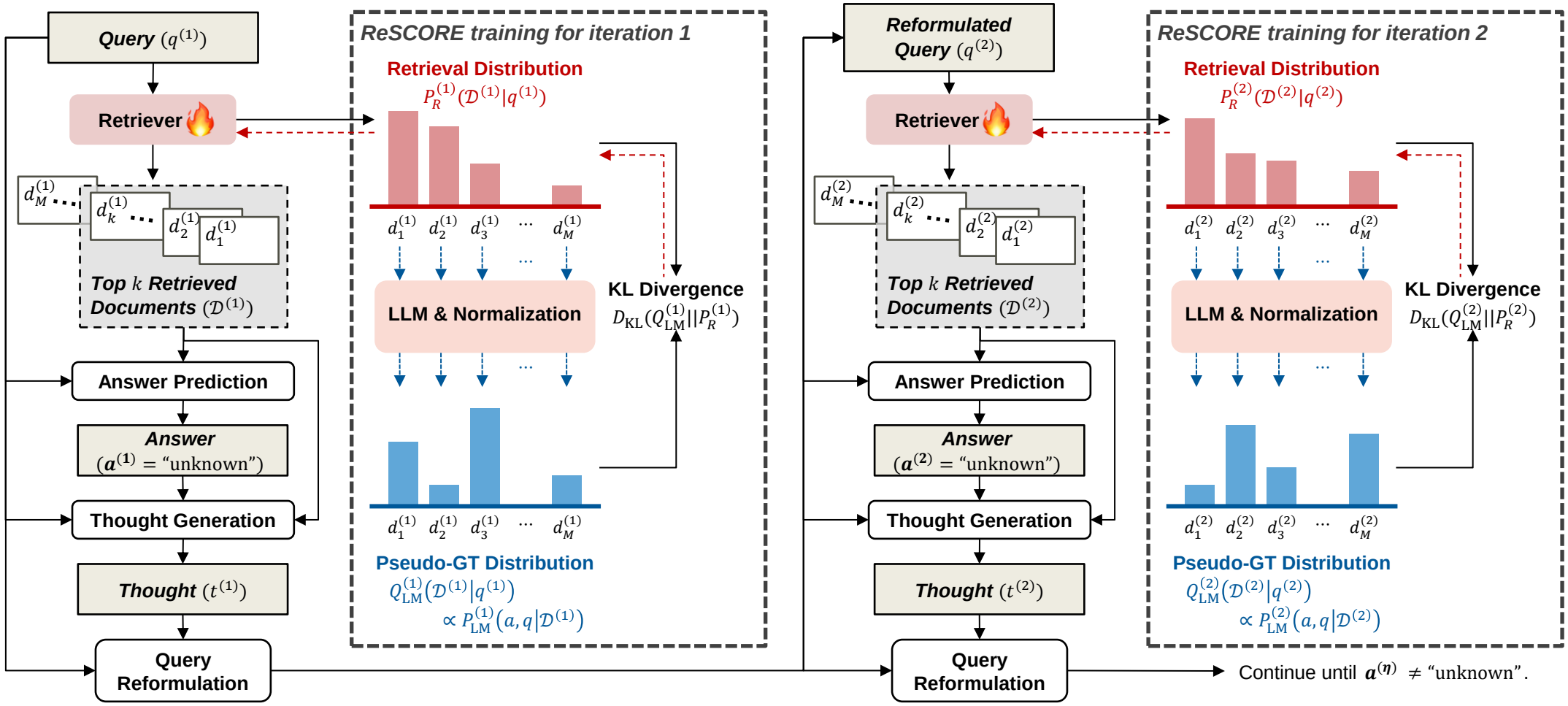
*"Presidents of Mexico"*

$P_{Consistency}(A|D, Q)$

*"Billie Eilish"*

$P_{Relevance}(Q|D)$

# ReSCORE Training



# Comparison to SOTA MHQA Baselines

| Model | MuSiQue EM | MuSiQue F1 | HotpotQA EM | HotpotQA F1 | 2WikiMHQA EM | 2WikiMHQA F1 |
|---|---|---|---|---|---|---|
| ReAcT (GPT-3.5+BM25)† | 10.2 | 19.7 | 36.0 | 46.9 | 28.0 | 37.3 |
| FLARE (GPT-3.5+BM25)† | 11.2 | 18.7 | 36.4 | 47.8 | 31.8 | 42.8 |
| Self-RAG (GPT-3.5+BM25)† | 10.6 | 19.2 | 33.8 | 44.4 | 24.4 | 30.8 |
| Adaptive-Note (GPT-3.5+BM25)† | 13.2 | 24.2 | 45.6 | 58.4 | 43.2 | 54.2 |
| IRCoT (Flan-T5-XL+BM25)‡ | 22.0 | 31.8 | 44.42 | 56.2 | 49.7 | 54.9 |
| Adaptive-RAG (Flan-T5-XL+BM25)‡‡ | **23.6** | 31.8 | 42.0 | 53.8 | 40.6 | 49.8 |
| Our Baseline (Llama-3.1-8B+BM25) | 15.2 | 23.6 | 42.2 | 55.7 | 44.6 | 52.2 |
| Our Baseline (Llama-3.1-8B+Contriever) | 15.2 | 23.8 | 39.4 | 52.3 | 32.8 | 41.6 |
| IQATR (Llama-3.1-8B+Contriever trained w/ ReSCORE) | 23.4 | **32.7** | **47.2** | **59.3** | **50.0** | **59.7** |

| | MuSiQue | | | HotpotQA | | | 2WikiMHQA | | |
|---|---|---|---|---|---|---|---|---|---|
| | cEM | EM | F1 | cEM | EM | F1 | cEM | EM | F1 |
| Self-Ask | - | 13.8 | 27.0 | - | - | - | - | 30.0 | 36.1 |
| Self-Ask + SE | - | 15.2 | 27.2 | - | - | - | - | 40.1 | 52.6 |
| SearChain + ColBERT | 17.1 | - | - | 56.9 | - | - | 46.3 | - | - |
| **IQATR (Ours)** | 30.4 | 23.4 | 32.7 | 59.6 | 47.2 | 59.3 | 57.0 | 50.0 | 59.7 |

| Model | QA EM | QA F1 | MHR$_i$@8 $i=1$ | MHR$_i$@8 $i=2$ | MHR$_i$@8 $i=\eta_n$ |
|---|---|---|---|---|---|
| **MuSiQue** | | | | | |
| Self-RAG* | 1.2 | 8.2 | 25.8 | 25.8 | 25.8 |
| +ReSCORE | 2.8 | 10.8 | 24.9 | 31.6 | 31.6 |
| FLARE | 7.3 | 13.3 | 31.0 | 37.1 | 37.1 |
| +ReSCORE | 8.2 | 15.3 | 30.9 | 40.1 | 43.3 |
| Adaptive-Note | 9.6 | 17.7 | 44.9 | 50.2 | 50.2 |
| +ReSCORE | 11.2 | 20.5 | 45.1 | 49.8 | 55.3 |
| Our Baseline | 15.2 | 23.8 | 44.9 | 51.6 | 51.6 |
| +ReSCORE | 23.4 | 32.7 | 46.8 | 63.0 | 65.2 |
| **HotpotQA** | | | | | |
| Self-RAG* | 5.6 | 17.9 | 36.1 | 36.5 | 36.5 |
| +ReSCORE | 8.7 | 19.2 | 33.8 | 37.2 | 37.2 |
| FLARE | 27.5 | 38.9 | 37.2 | 48.4 | 48.4 |
| +ReSCORE | 31.4 | 42.5 | 39.2 | 48.5 | 51.7 |
| Adaptive-Note | 42.0 | 55.3 | 44.8 | 49.8 | 50.1 |
| +ReSCORE | 43.8 | 58.0 | 47.3 | 63.3 | 77.2 |
| Our Baseline | 39.4 | 52.3 | 44.8 | 47.5 | 47.5 |
| +ReSCORE | 47.2 | 59.3 | 46.6 | 69.3 | 72.4 |
| **2WikiMHQA** | | | | | |
| Self-RAG* | 3.0 | 19.1 | 26.3 | 27.1 | 27.1 |
| +ReSCORE | 5.6 | 21.2 | 25.9 | 28.4 | 32.8 |
| FLARE | 23.2 | 35.0 | 32.5 | 42.9 | 42.9 |
| +ReSCORE | 26.5 | 38.0 | 33.2 | 45.6 | 45.6 |
| Adaptive-Note | 35.7 | 46.1 | 45.7 | 59.2 | 59.2 |
| +ReSCORE | 37.4 | 49.3 | 49.8 | 63.2 | 67.5 |
| Our Baseline | 32.8 | 41.6 | 45.7 | 56.9 | 56.9 |
| +ReSCORE | 50.0 | 59.7 | 51.2 | 81.2 | 88.0 |

# Pseudo-GT Supervision Ablation

| Method | MuSiQue | | | HotpotQA | | | 2WikiMHQA | | |
|---|---|---|---|---|---|---|---|---|---|
| | R@8 | EM | F1 | R@8 | EM | F1 | R@8 | EM | F1 |
| None | 47.06 | 15.2 | 23.8 | 61.65 | 39.4 | 52.3 | 58.85 | 32.8 | 41.6 |
| $P_{LM}(a\|d,q)$ | 41.41 | 5.8 | 12.3 | 42.75 | 19.2 | 26.4 | 41.85 | 18.8 | 26.5 |
| $P_{LM}(q\|d)$ | 47.93 | 15.9 | 25.9 | 65.9 | 42.0 | 53.9 | 63.23 | 39.2 | 47.9 |
| $P_{LM}(q,a\|d)$ | 55.66 | 16.4 | 26.3 | 68.3 | 43.6 | 56.4 | 67.10 | 41.4 | 51.7 |

$$P_{QA}(A|Q) = \sum_{D} P(A, D|Q)$$

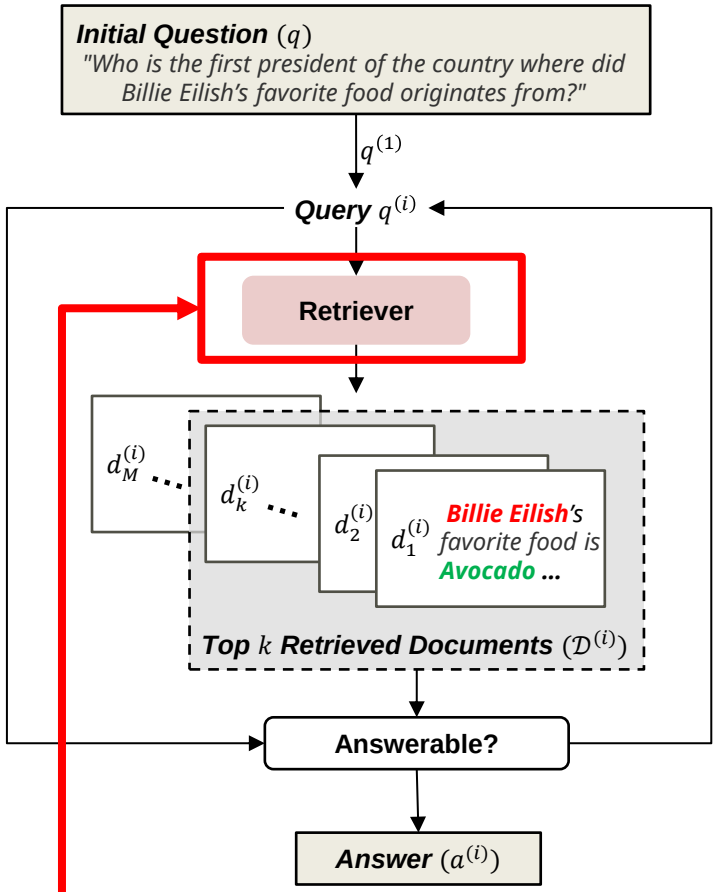
$$P_{RAG}(A, D|Q) = P_{generate}(A|D, Q)\, P_{retrieve}(D|Q)$$

$$P_{Retriever}(D|Q) \propto P_{LLM}(A, Q|D)$$

$$P_{LLM}(A, Q|D) = P_{Consistency}(A|D, Q)\; P_{Relevance}(Q|D)$$

| Pseudo-GT Label | R@2 | R@4 | R@8 | R@16 |
|---|---|---|---|---|
| **MuSiQue** | | | | |
| None | 32.71 | 40.10 | 47.06 | 53.61 |
| $P_{\text{LM}}(q \mid d)$ | 34.64 | 41.09 | 47.93 | 54.24 |
| $P_{\text{LM}}(a \mid q, d)$ | 28.94 | 35.10 | 41.41 | 47.84 |
| $P_{\text{LM}}(q, a \mid d)$ | **42.68** | **50.31** | **55.66** | **60.38** |
| **HotpotQA** | | | | |
| None | 49.40 | 56.45 | 61.65 | 66.25 |
| $P_{\text{LM}}(q \mid d)$ | 55.15 | 62.35 | 65.85 | 69.10 |
| $P_{\text{LM}}(a \mid q, d)$ | 27.50 | 34.35 | 42.75 | 52.50 |
| $P_{\text{LM}}(q, a \mid d)$ | **58.05** | **64.60** | **68.30** | **70.65** |
| **2WikiMHQA** | | | | |
| None | 46.40 | 54.30 | 58.85 | 63.35 |
| $P_{\text{LM}}(q \mid d)$ | 50.78 | 59.08 | 63.23 | 66.13 |
| $P_{\text{LM}}(a \mid q, d)$ | 26.10 | 33.26 | 41.85 | 51.20 |
| $P_{\text{LM}}(q, a \mid d)$ | **53.73** | **62.98** | **67.10** | **68.73** |

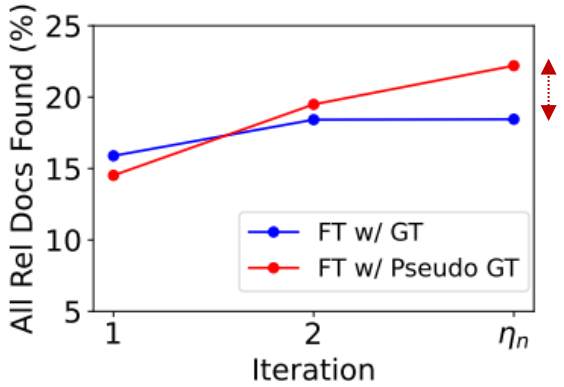
# Comparison with GT



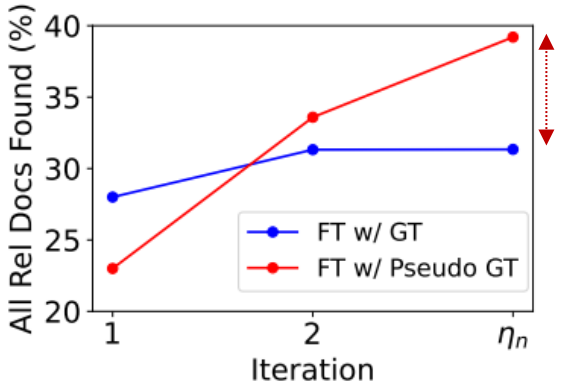

Replaced the Retriever with GT-trained

| Label | QA EM | QA F1 | MHR$_i$@8 $i=1$ | MHR$_i$@8 $i=2$ | MHR$_i$@8 $i=\eta_n$ |
|---|---|---|---|---|---|
| **MuSiQue** | | | | | |
| None | 15.2 | 23.8 | 44.9 | 51.6 | 51.6 |
| GT | 15.8 | 24.9 | 46.7 | 54.8 | 54.8 |
| Pseudo-GT | 23.4 | 32.7 | 46.8 | 63.0 | 65.2 |
| **HotpotQA** | | | | | |
| None | 39.4 | 52.3 | 44.8 | 47.5 | 47.5 |
| GT | 39.2 | 45.8 | 48.7 | 52.7 | 52.7 |
| Pseudo-GT | 47.2 | 59.3 | 46.6 | 69.3 | 72.4 |
| **2WikiMHQA** | | | | | |
| None | 32.8 | 41.6 | 45.7 | 56.9 | 56.9 |
| GT | 37.1 | 46.2 | 48.5 | 61.7 | 61.7 |
| Pseudo-GT | 50.0 | 59.7 | 51.2 | 81.2 | 88.0 |

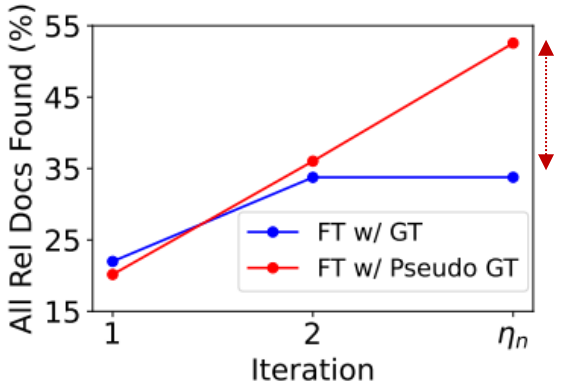
# Comparison of GT and Pseudo-GT Labels




(a) MuSiQue Dataset

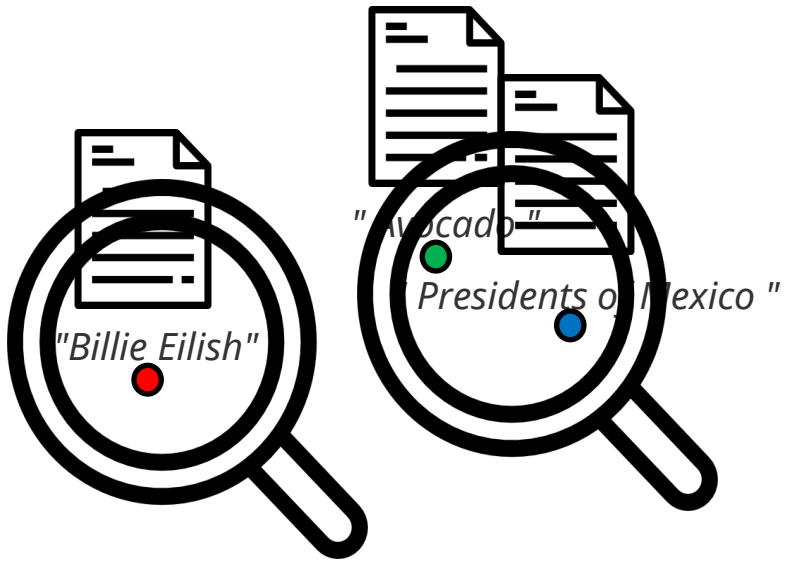



(b) HotpotQA Dataset

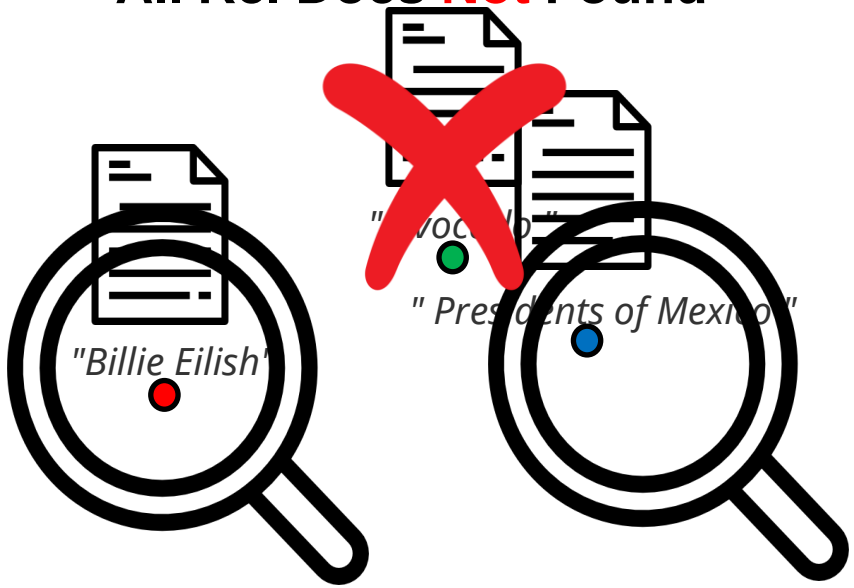



(c) 2WikiMHQA Dataset

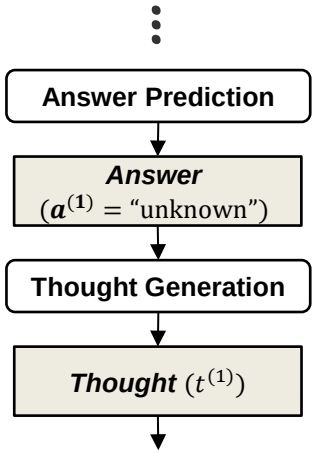
**All Rel Docs Found**




**All Rel Docs Not Found**



# Query Reformulation Ablation




*"Who is the first president of the country where did Billie Eilish's favorite food originates from?"*

*"**Billie Eilish**'s favorite food is **Avocado**"*

## Query Reformulation

**None**

*" Who is the first president of the country where did Billie Eilish's favorite food originates from?"*

*OR*

**LLM-rewrite**

*"Who is the first president of the country where did **Avocado** originates from?"*

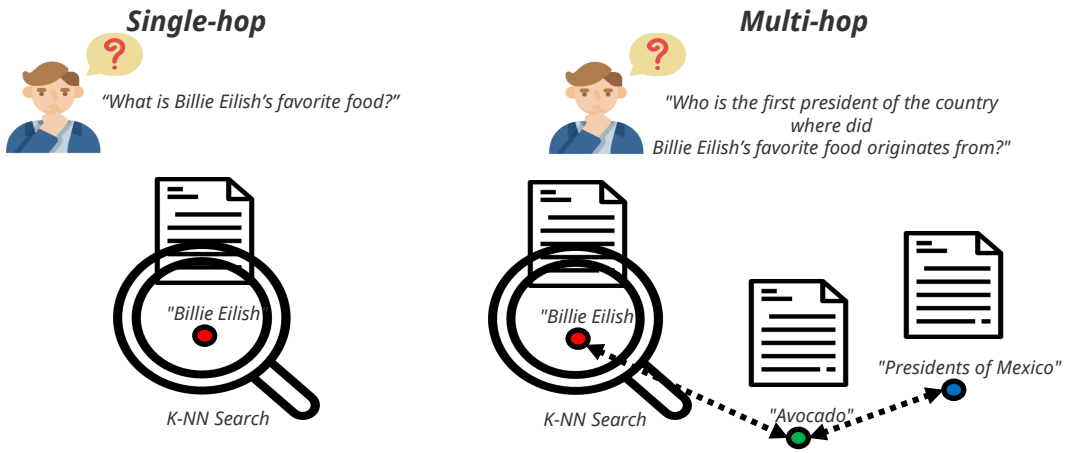
*OR*

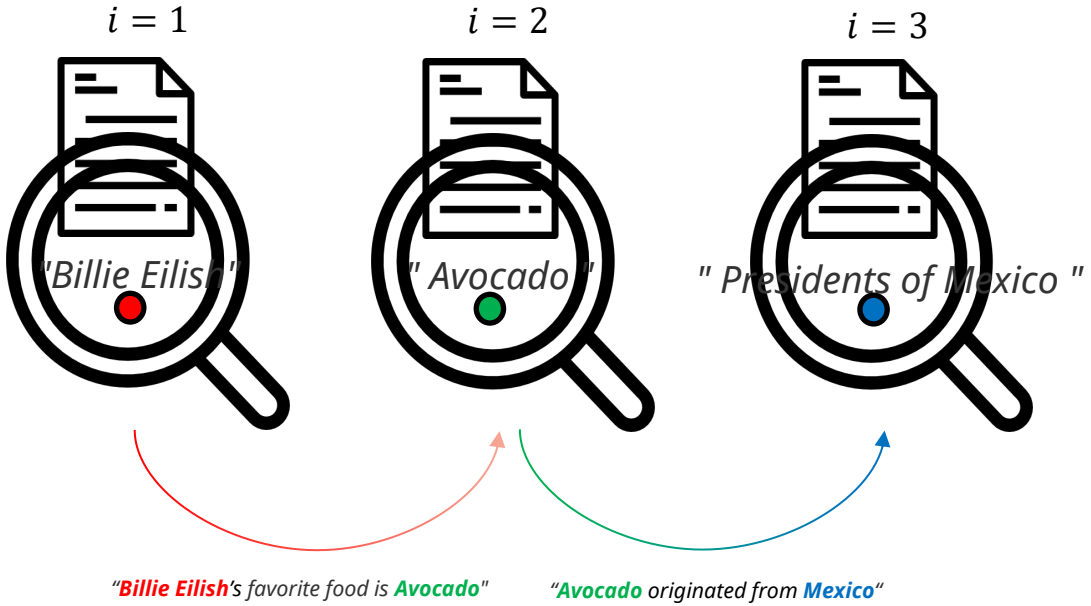
**Thought-concat**

*"Who is the first president of the country where did Billie Eilish's favorite food originates from? **Billie Eilish**'s favorite food is **Avocado.**"*

| Reformulation Method | QA EM | QA F1 | MHR$_i$@8 $i=1$ | MHR$_i$@8 $i=2$ | MHR$_i$@8 $i=\eta_n$ |
|---|---|---|---|---|---|
| **MuSiQue** | | | | | |
| None | 10.8 | 17.8 | 44.7 | 45.4 | 47.4 |
| LLM-rewrite | 21.2 | 30.5 | 45.1 | 56.7 | 63.7 |
| Thought-concat | 23.4 | 32.7 | 46.8 | 63.0 | 65.2 |
| **HotpotQA** | | | | | |
| None | 29.4 | 41.1 | 42.8 | 43.6 | 43.8 |
| LLM-rewrite | 44.2 | 57.4 | 41.9 | 54.8 | 64.7 |
| Thought-concat | 47.2 | 59.3 | 46.6 | 69.3 | 72.4 |
| **2WikiMHQA** | | | | | |
| None | 35.6 | 44.7 | 48.6 | 49.7 | 49.8 |
| LLM-rewrite | 51.7 | 60.1 | 50.0 | 86.0 | 89.5 |
| Thought-concat | 50.0 | 59.7 | 51.2 | 81.2 | 88.0 |

# Summary





$$P_{QA}(A|Q) = \sum_{D} P(A, D|Q)$$

$$P_{RAG}(A, D|Q) = P_{generate}(A|D, Q)\, P_{retrieve}(D|Q)$$

$$P_{Retriever}(D|Q) \propto P_{LLM}(A, Q|D)$$

$$P_{LLM}(A, Q|D) = P_{Consistency}(A|D, Q)\, P_{Relevance}(Q|D)$$

# E.O.D.