

Self-RAG: Learning to Retrieve, Generate and Critique through Self-Reflection

selfrag.github.io



Akari Asai¹



Zeqiu Wu¹



Yizhong Wang¹



Avirup Sil³



Hannaneh Hajishirzi^{1,2}



IBM Research³

Retrieval-augmented generation (**RAG**)



Where is ICLR 2024?

Retrieval-augmented generation (**RAG**)

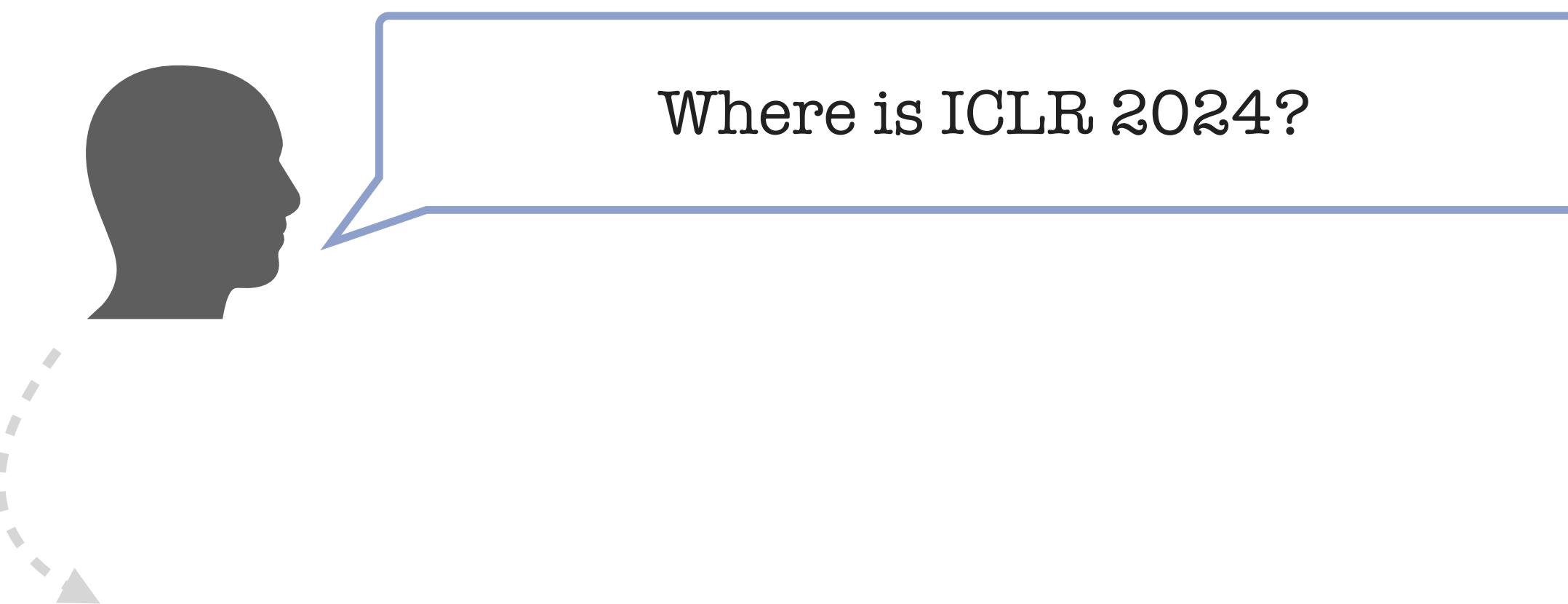
Step 1: Retrieve K documents



Where is ICLR 2024?

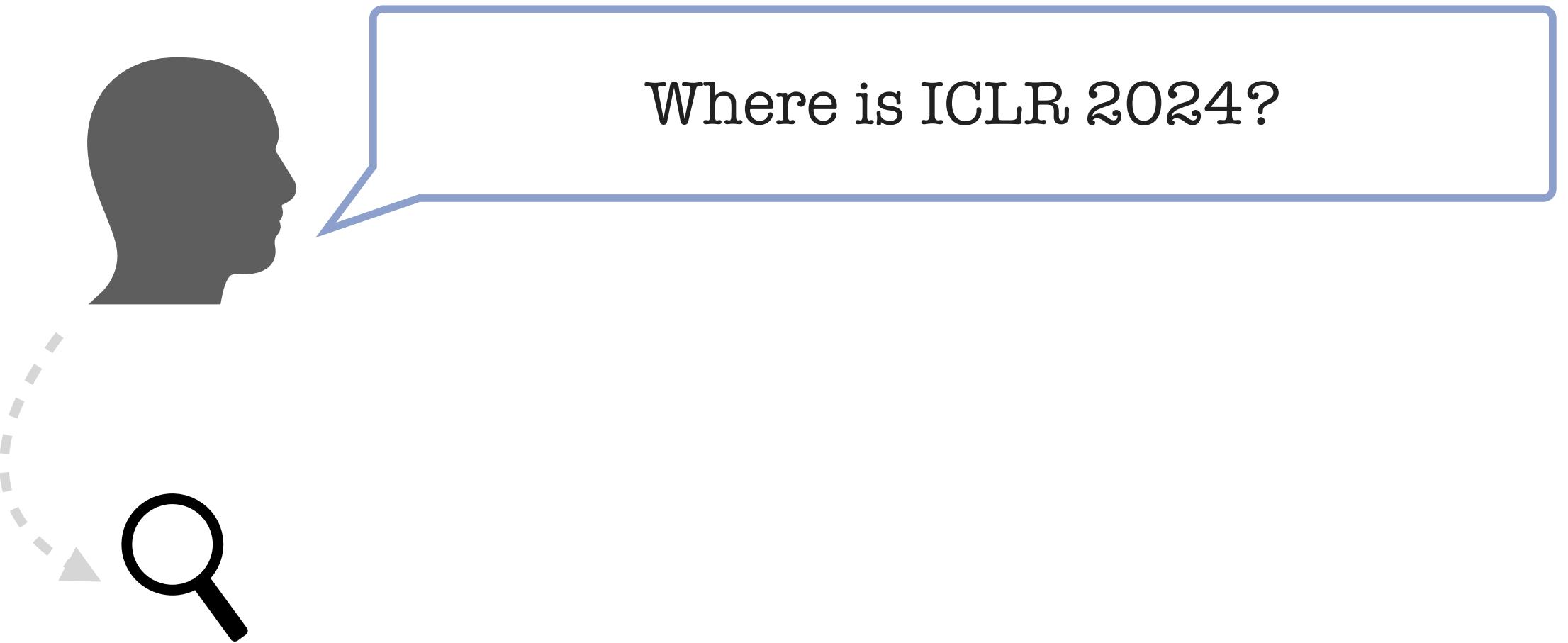
Retrieval-augmented generation (**RAG**)

Step 1: Retrieve K documents



Retrieval-augmented generation (**RAG**)

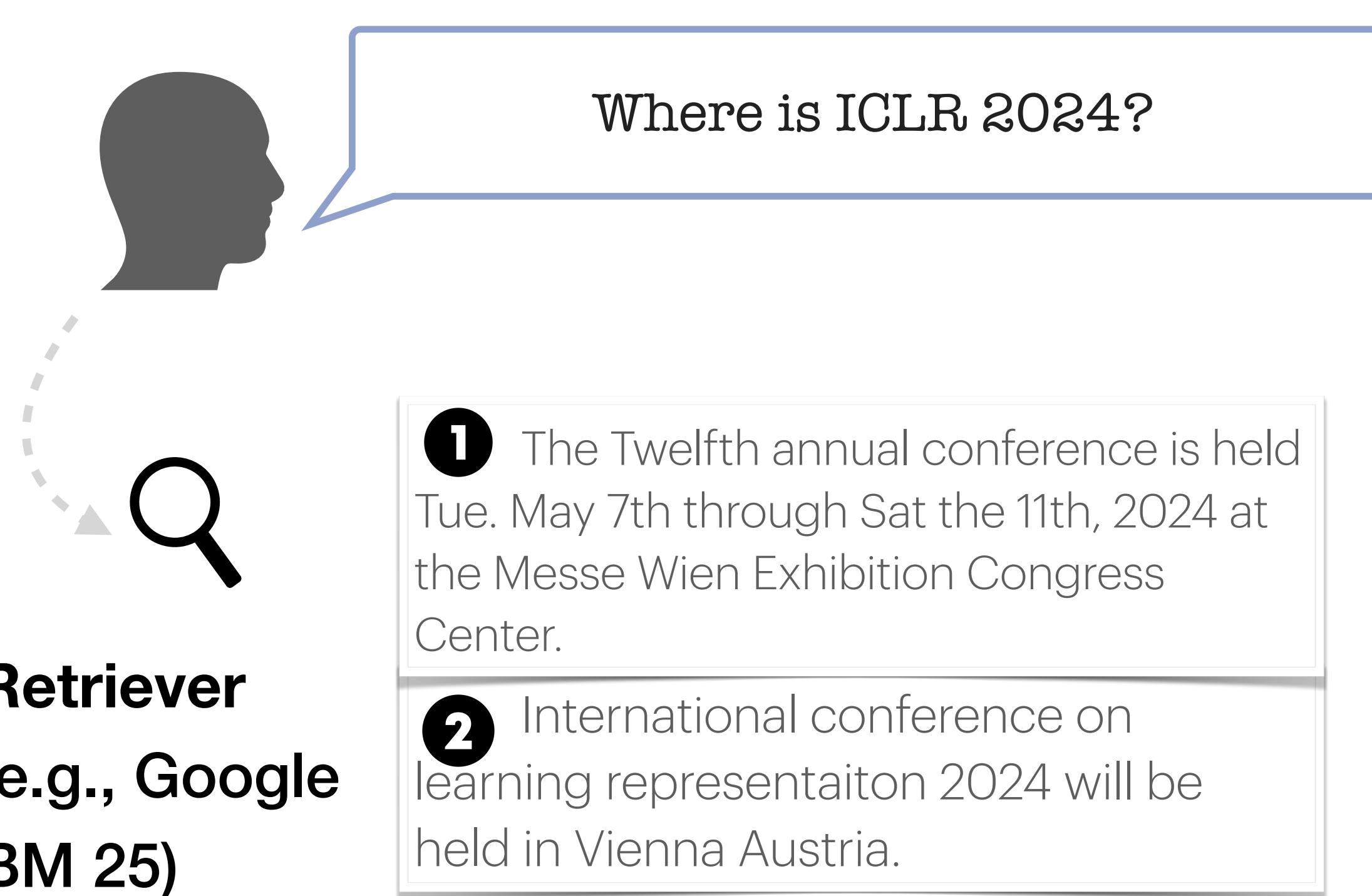
Step 1: Retrieve K documents



Retriever
(e.g., Google
BM 25)

Retrieval-augmented generation (**RAG**)

Step 1: Retrieve K documents

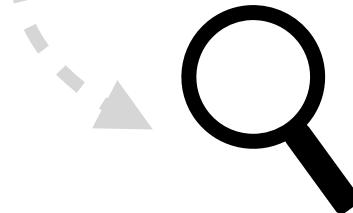


Retrieval-augmented generation (**RAG**)

Step 1: Retrieve K documents



Where is ICLR 2024?



Retriever
(e.g., Google
BM 25)

1 The Twelfth annual conference is held Tue. May 7th through Sat the 11th, 2024 at the Messe Wien Exhibition Congress Center.

2 International conference on learning representaiton 2024 will be held in Vienna Austria.

Answer my question using references.

References: **1** **2**

Question: Where is ICLR 2024?

Retrieval-augmented generation (**RAG**)

Step 1: Retrieve K documents



Where is ICLR 2024?

1 The Twelfth annual conference is held Tue. May 7th through Sat the 11th, 2024 at the Messe Wien Exhibition Congress Center.

2 International conference on learning representaiton 2024 will be held in Vienna Austria.

Retriever
(e.g., Google
BM 25)

Step 2: Prompt LM with K docs and generate

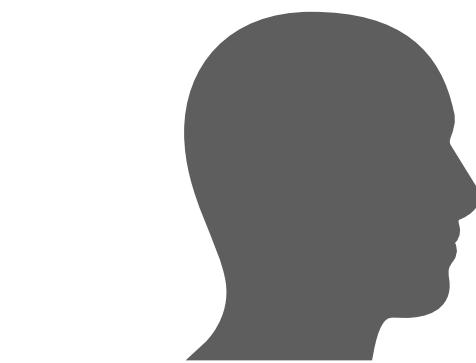
Answer my question using references.

References: **1** **2**

Question: Where is ICLR 2024?

Retrieval-augmented generation (RAG)

Step 1: Retrieve K documents



Where is ICLR 2024?

1 The Twelfth annual conference is held Tue. May 7th through Sat the 11th, 2024 at the Messe Wien Exhibition Congress Center.

2 International conference on learning representation 2024 will be held in Vienna Austria.

Retriever
(e.g., Google
BM 25)

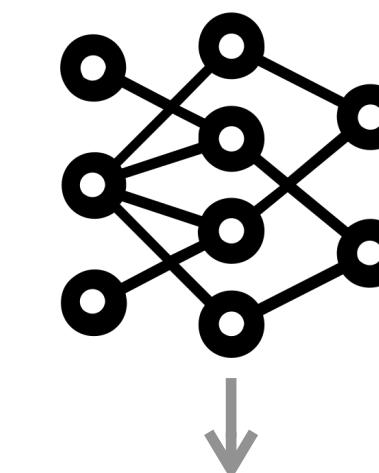
2

Step 2: Prompt LM with K docs and generate

Answer my question using references.

References: 1 2

Question: Where is ICLR 2024?



LMs
(e.g., GPT-4, Llama 3)

ICLR 2024 will be in Vienna, Austria. 1 2

factual

up-to-date

attributable

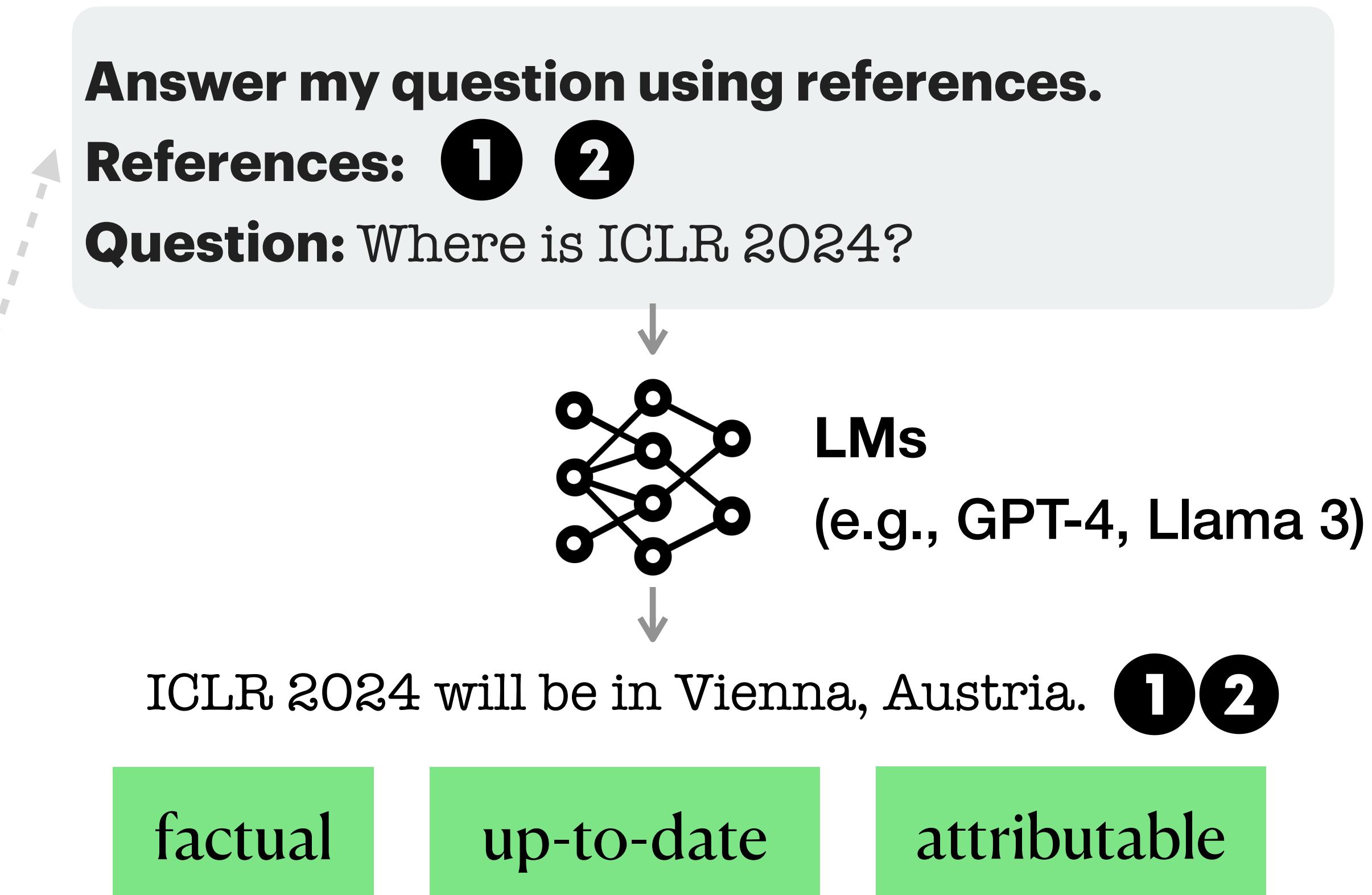
Retrieval-augmented generation (**RAG**)

Step 1: Retrieve K documents



Retriever
(e.g., Google
BM 25)

Step 2: Prompt LM with K docs and generate

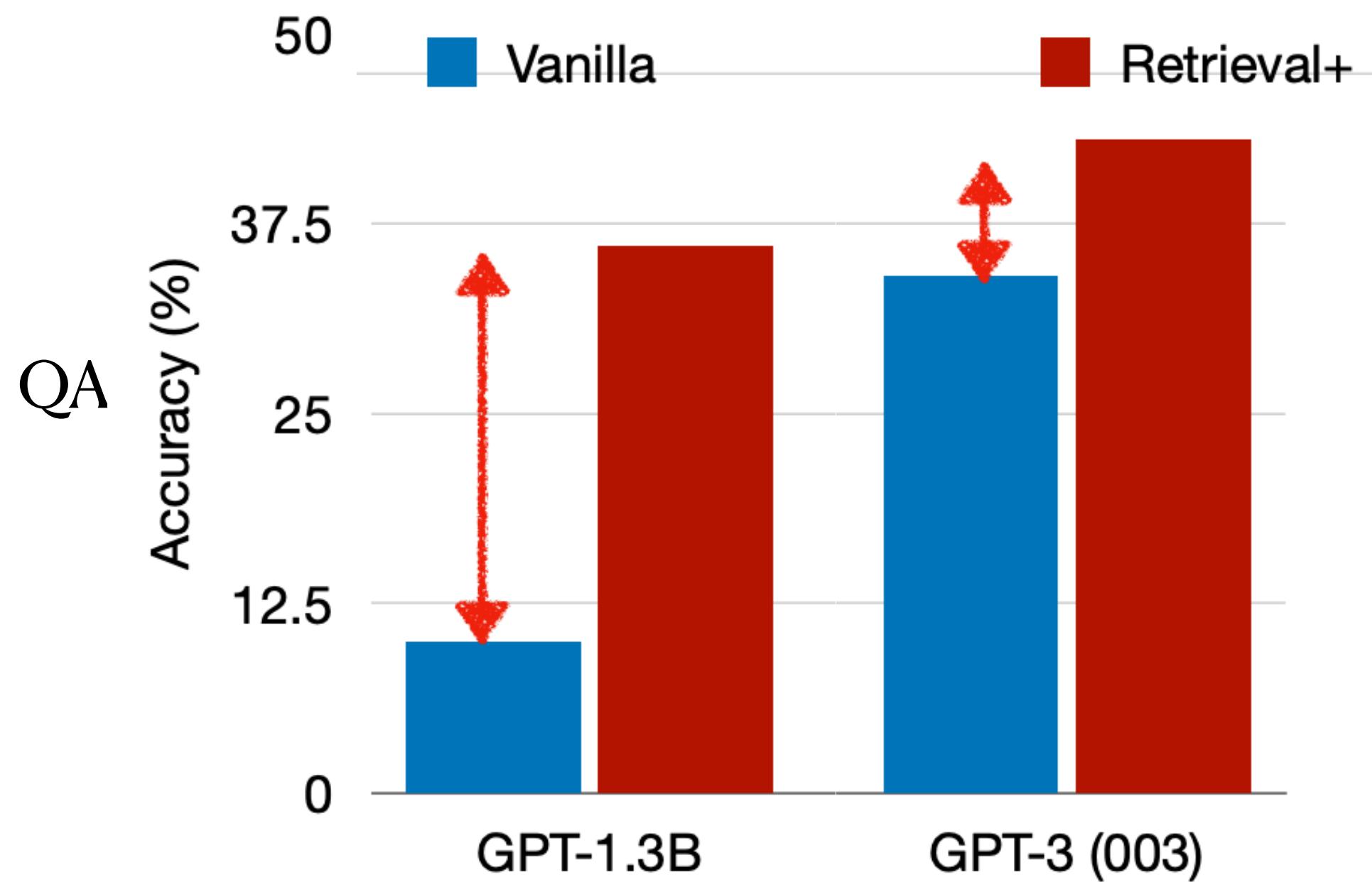


Chen et al. Reading Wikipedia to Answer Open-domain Questions. ACL 2017.
Lewis et al. Retrieval-augmented Generation for Knowledge Intensive Tasks. NeurIPS 2020.
Ram et al. In-Context Retrieval-Augmented Language Models. TACL 2023
Asai et al. Reliable, Adaptable, and Attributable Language Models with Retrieval. Arxiv 2024.

RAG has been widely and successfully adopted

RAG has been widely and successfully adopted

RAG has shown effective
in many benchmarks

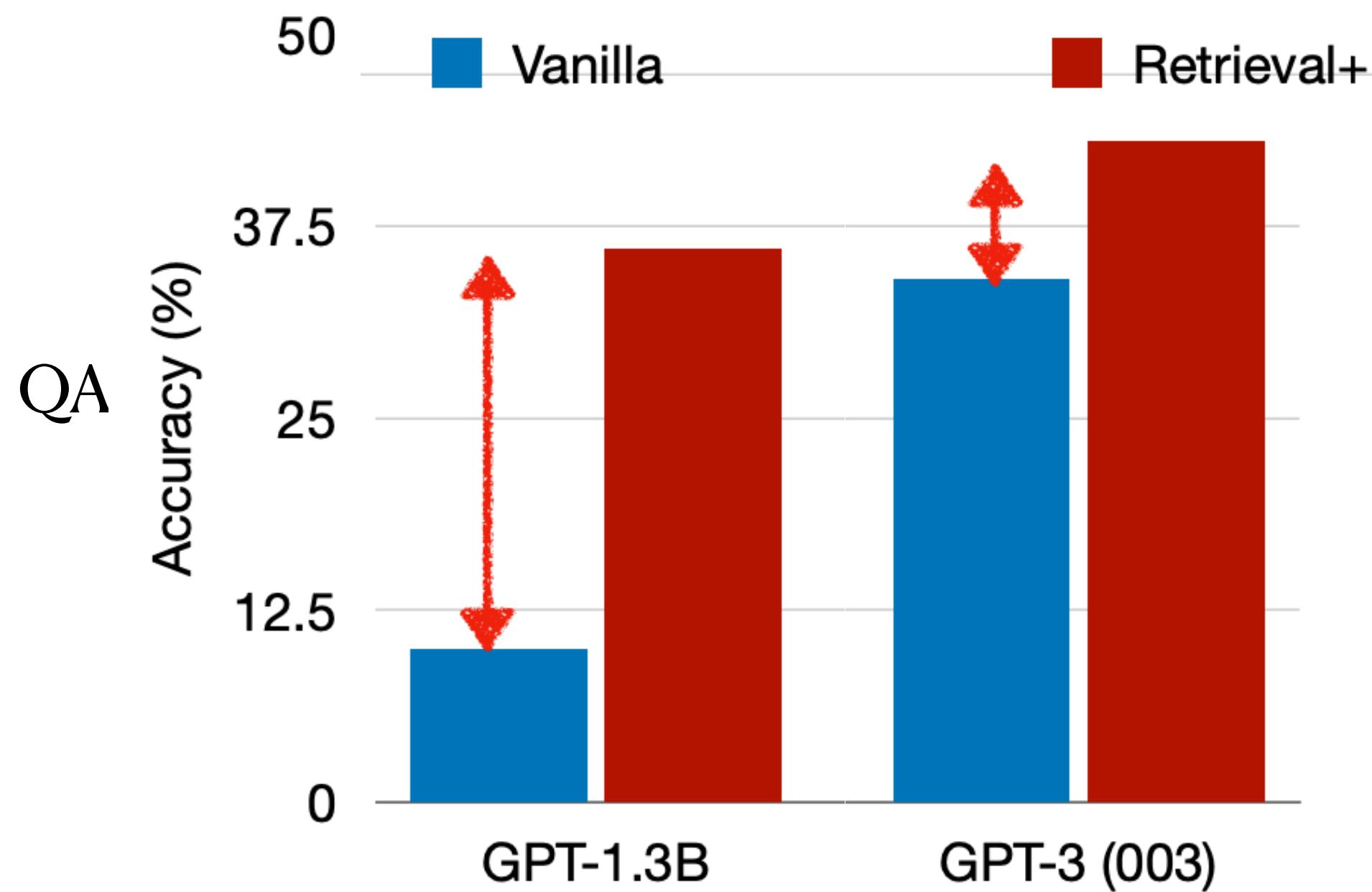


Mallen*, Asai* et al., When Not to Trust Language Models: Investigating Effectiveness of Parametric and Non-Parametric Memories. ACL 2023.

RAG has been widely and successfully adopted

RAG has shown effective
in many benchmarks

Widely adopted in real-world
systems

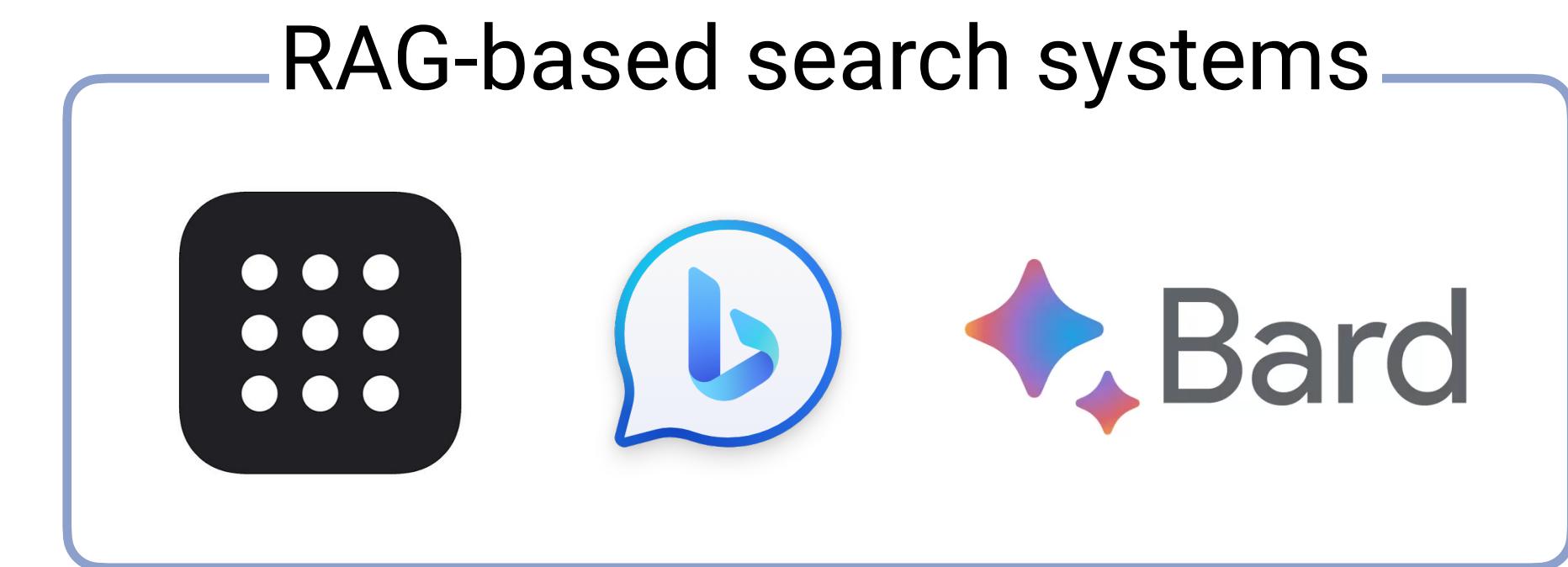
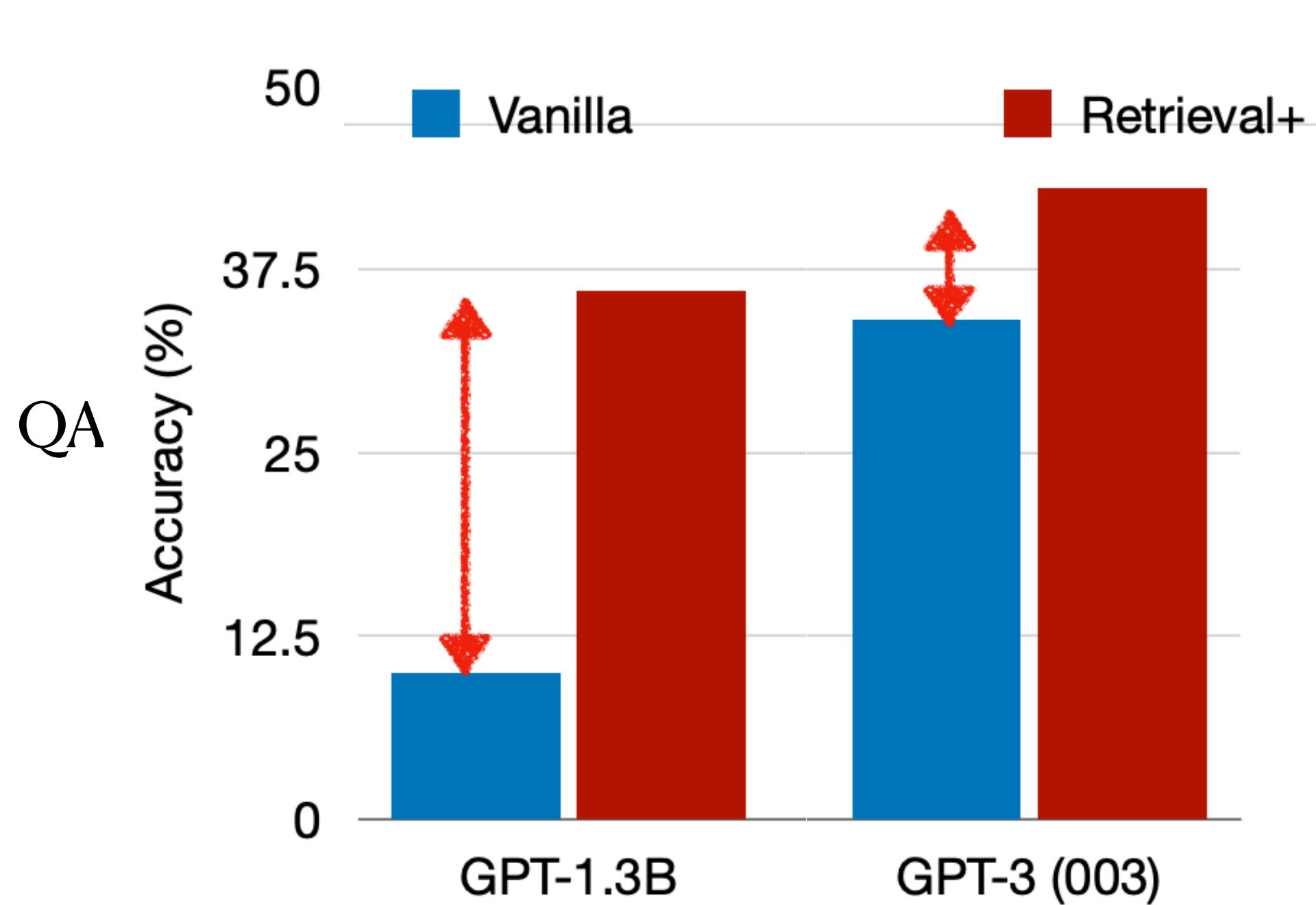


Mallen*, Asai* et al., When Not to Trust Language Models: Investigating Effectiveness of Parametric and Non-Parametric Memories. ACL 2023.

RAG has been widely and successfully adopted

RAG has shown effective
in many benchmarks

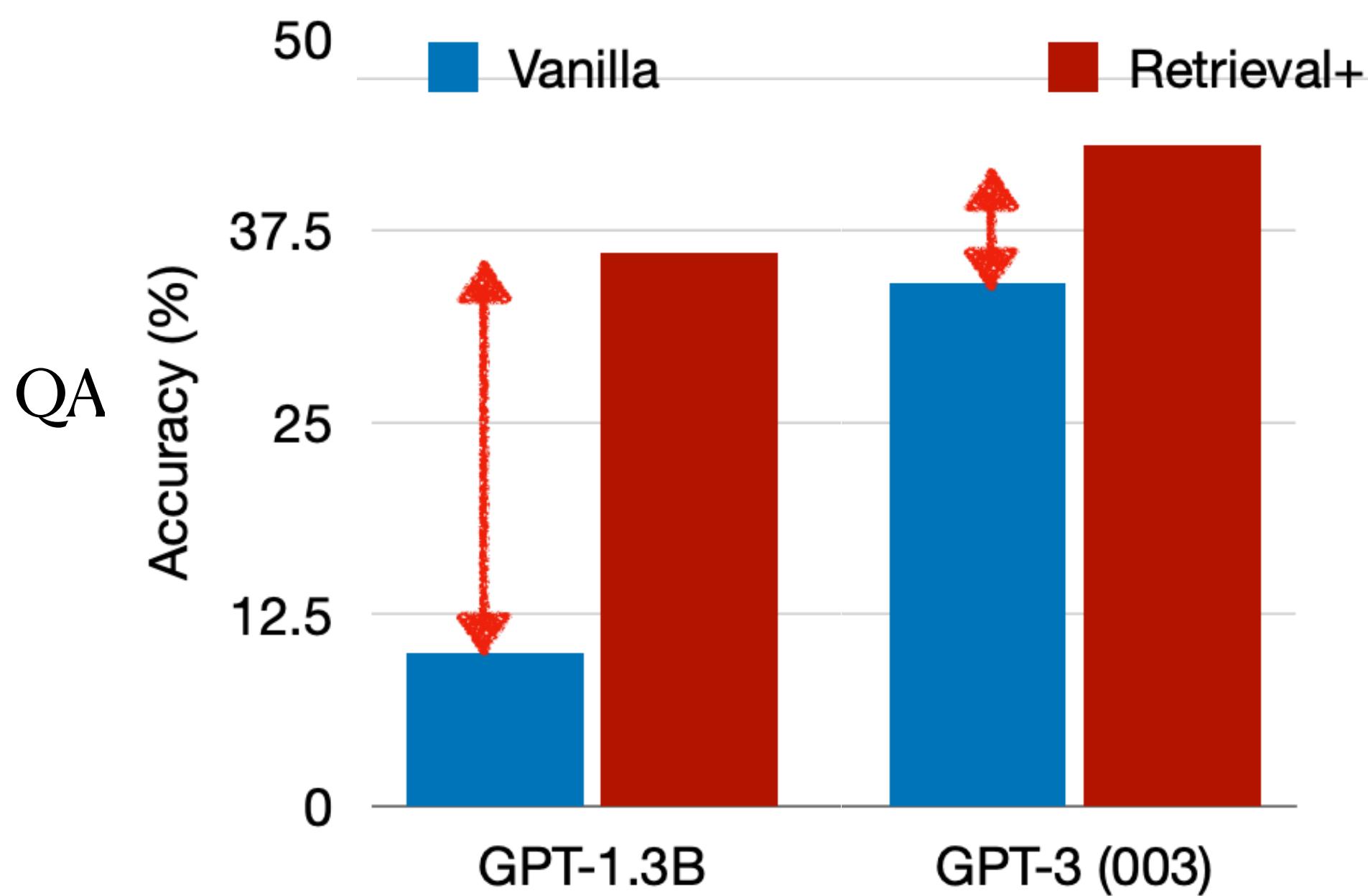
Widely adopted in real-world
systems



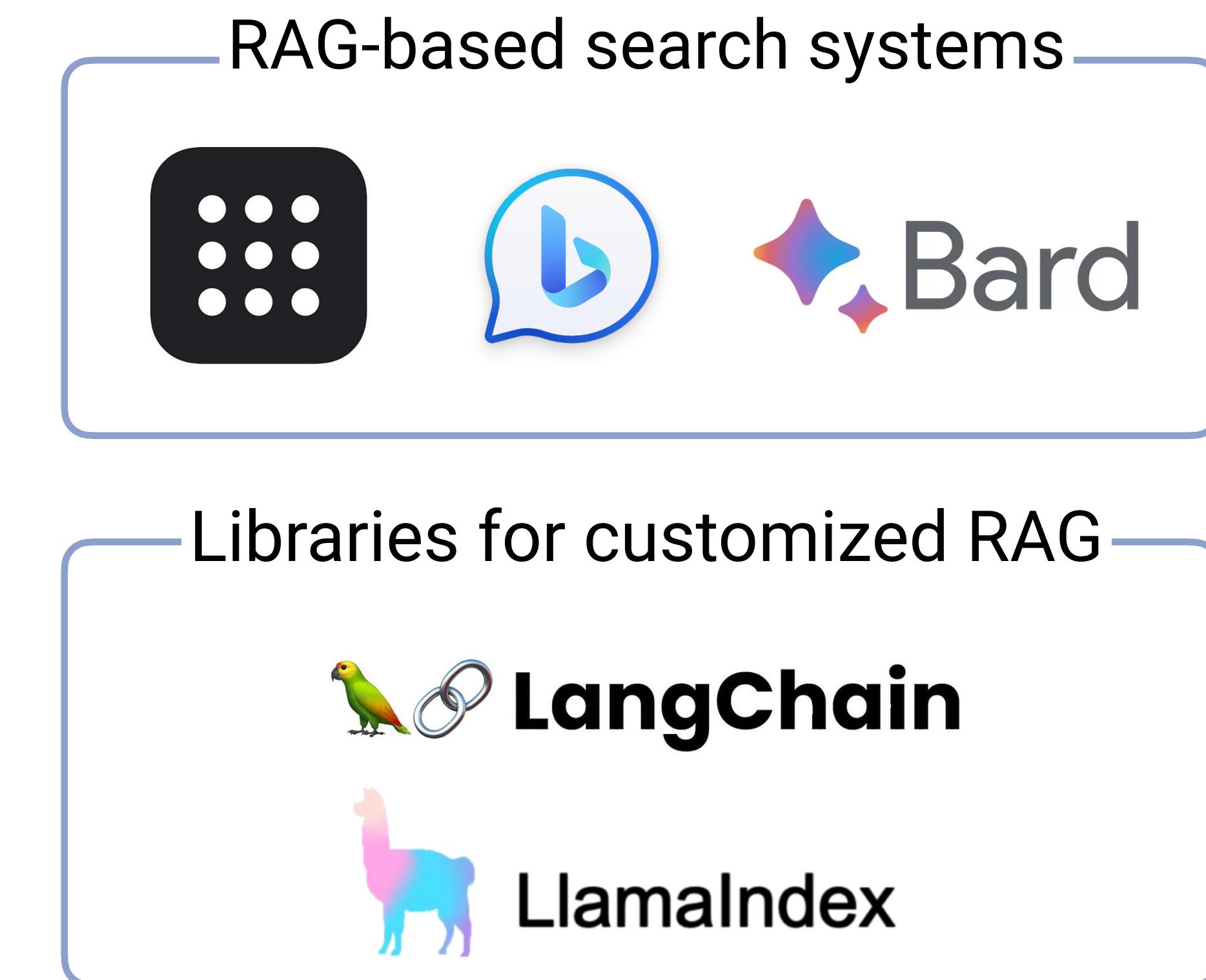
Mallen*, Asai* et al., When Not to Trust Language Models: Investigating Effectiveness of Parametric and Non-Parametric Memories. ACL 2023.

RAG has been widely and successfully adopted

RAG has shown effective
in many benchmarks



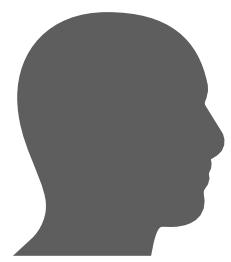
Widely adopted in real-world
systems



Mallen*, Asai* et al., When Not to Trust Language Models: Investigating Effectiveness of Parametric and Non-Parametric Memories. ACL 2023.

Standard RAG isn't (always) reliable

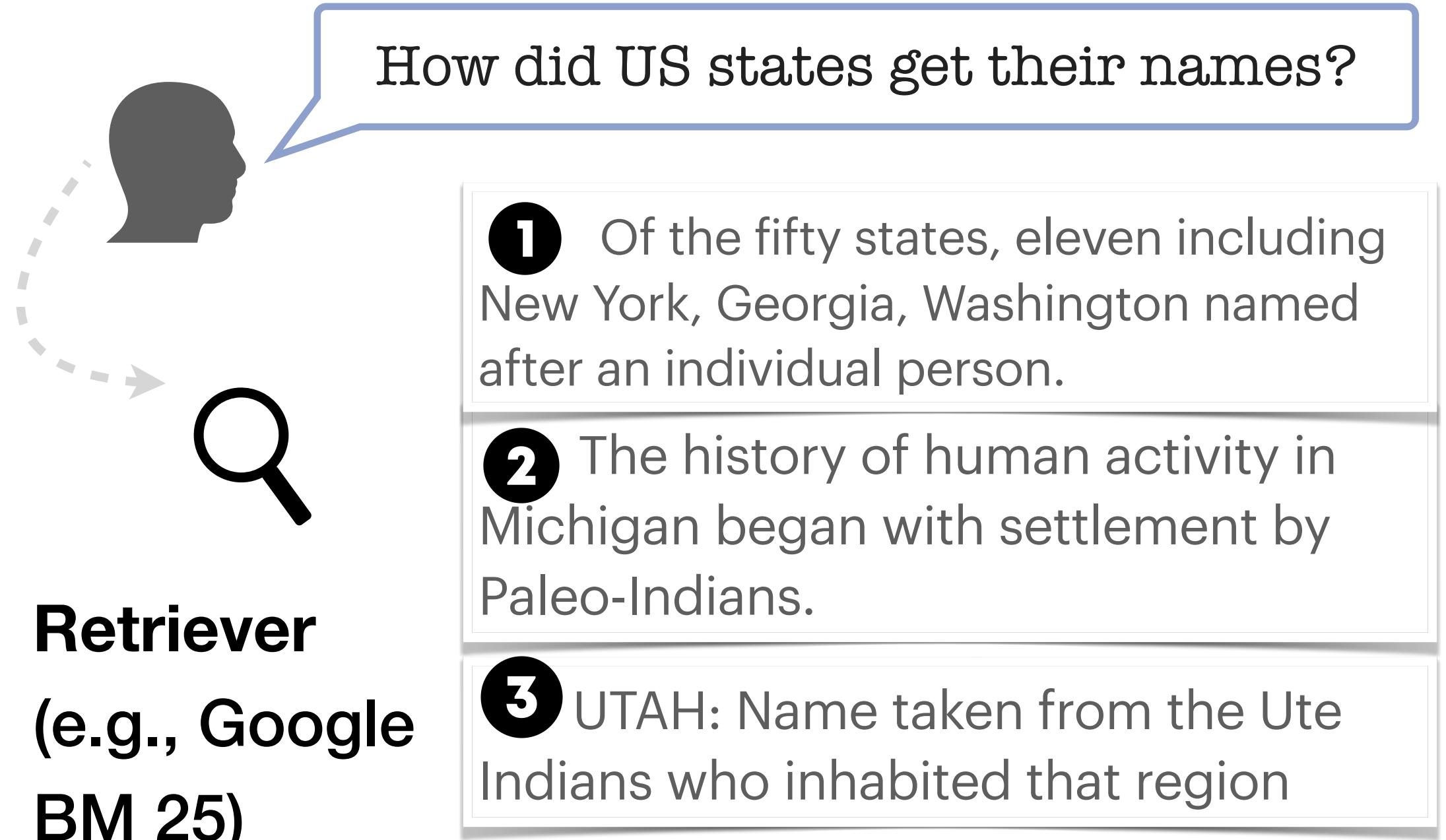
Standard RAG isn't (always) reliable



How did US states get their names?

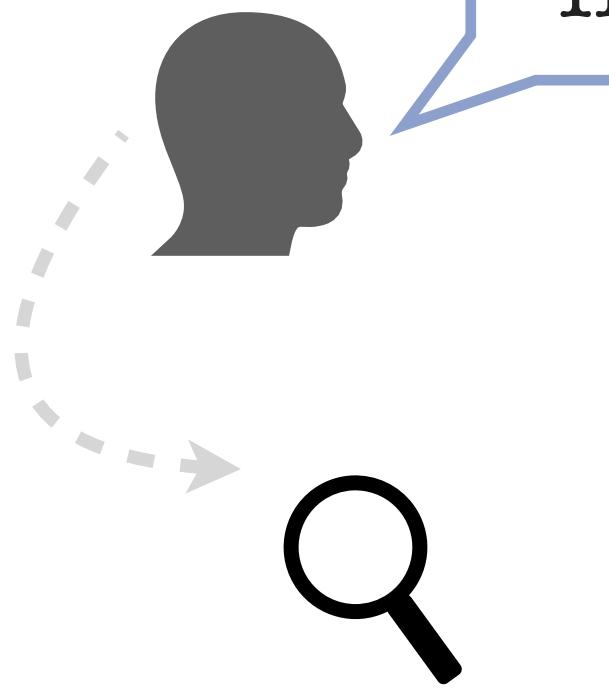
Standard RAG isn't (always) reliable

Step 1: Retrieve K documents



Standard RAG isn't (always) reliable

Step 1: Retrieve K documents



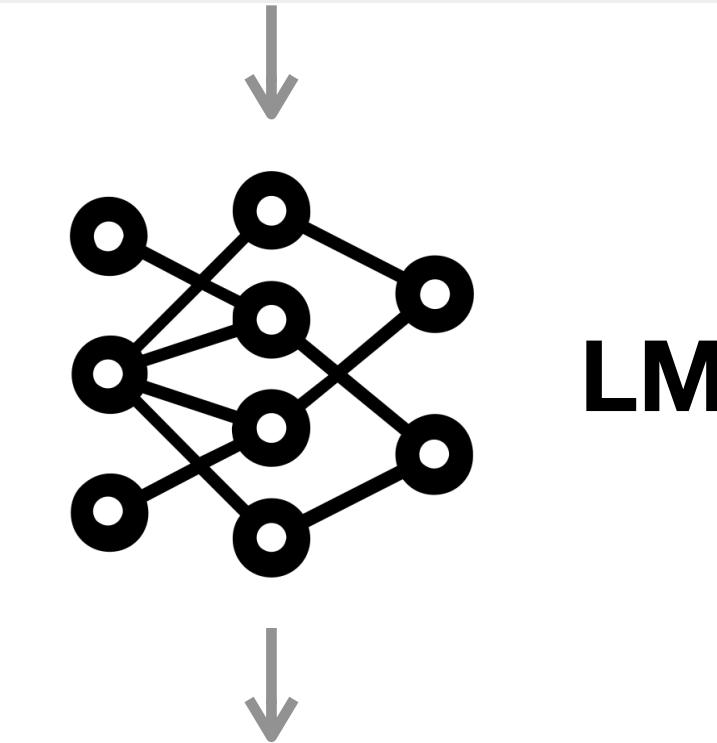
Retriever
(e.g., Google
BM 25)

How did US states get their names?

- 1 Of the fifty states, eleven including New York, Georgia, Washington named after an individual person.
- 2 The history of human activity in Michigan began with settlement by Paleo-Indians.
- 3 UTAH: Name taken from the Ute Indians who inhabited that region

Step 2: Prompt LM with K docs and generate

Prompt How did US states get their names? 1 2 3



LM

US states got their names from a variety of sources. Eleven states such as New York and Michigan are named after an individual person. Some states including Utah and Washington, are named after indigenous communities.

Standard RAG isn't (always) reliable

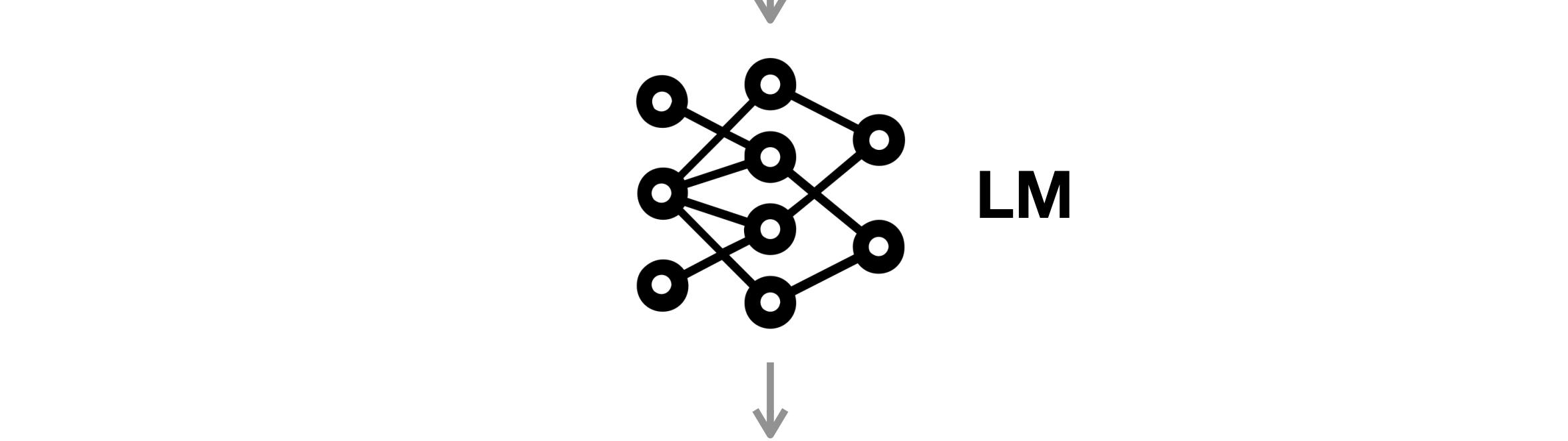
Step 1: Retrieve K documents



Retriever
(e.g., Google
BM 25)

Step 2: Prompt LM with K docs and generate

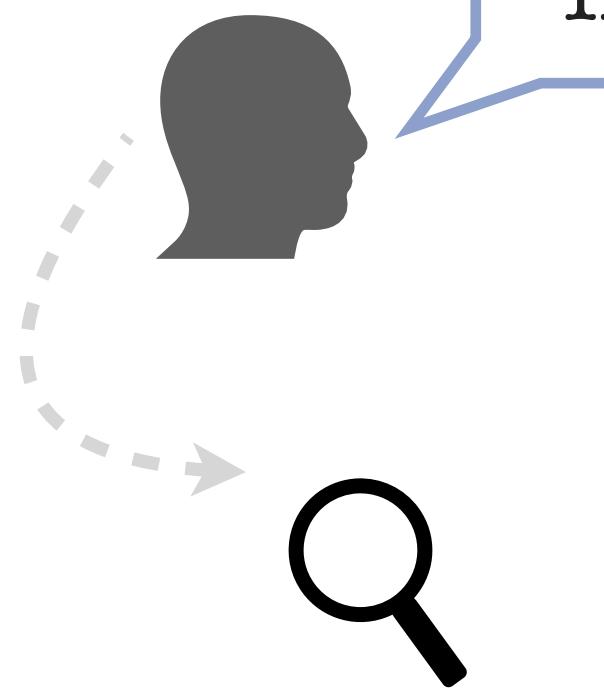
Prompt How did US states get their names? 1 2 3



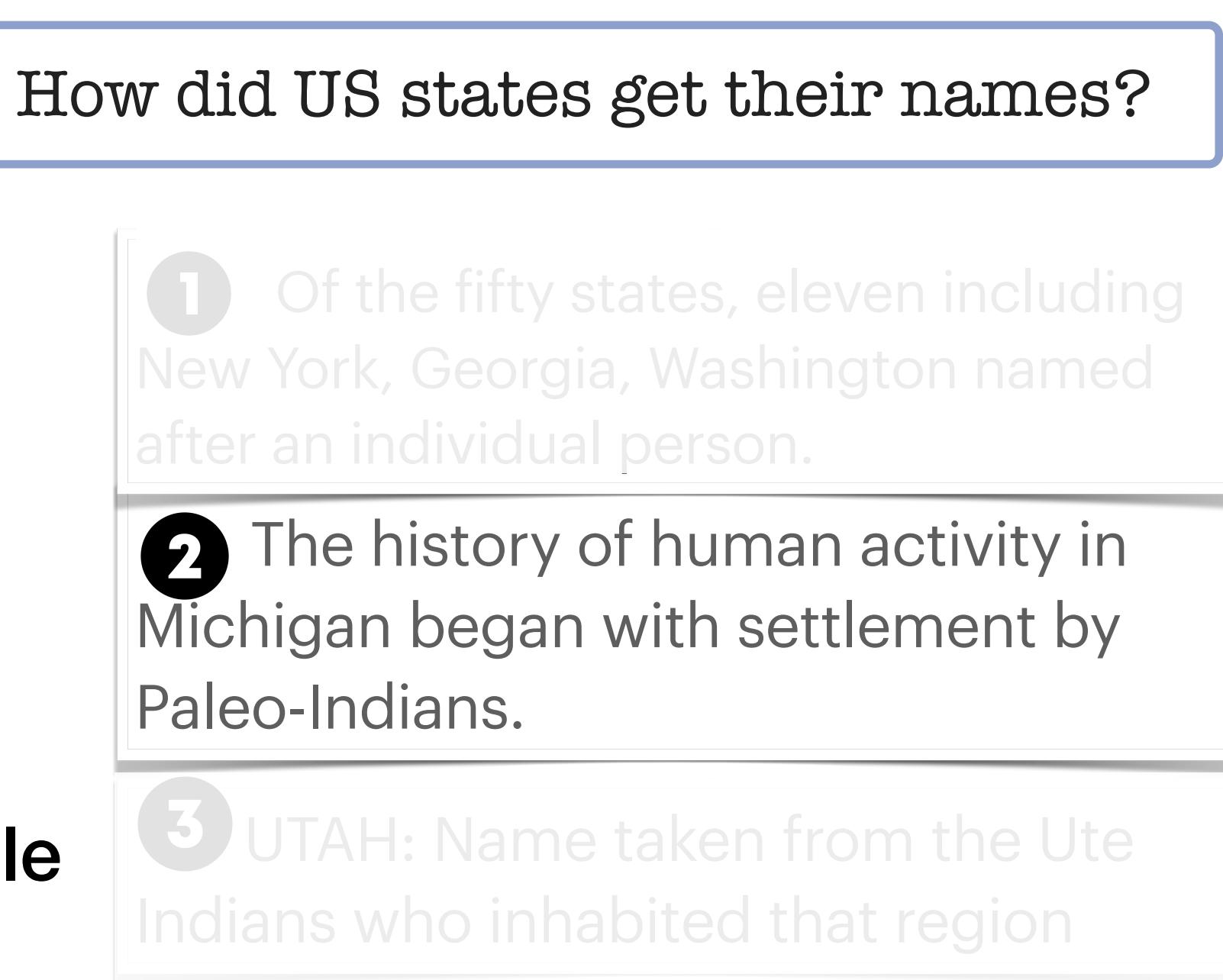
US states got their names from a variety of sources. Eleven states such as New York and Michigan are named after an individual person. Some states including Utah and Washington, are named after indigenous communities.

Standard RAG isn't (always) reliable

Step 1: Retrieve K documents



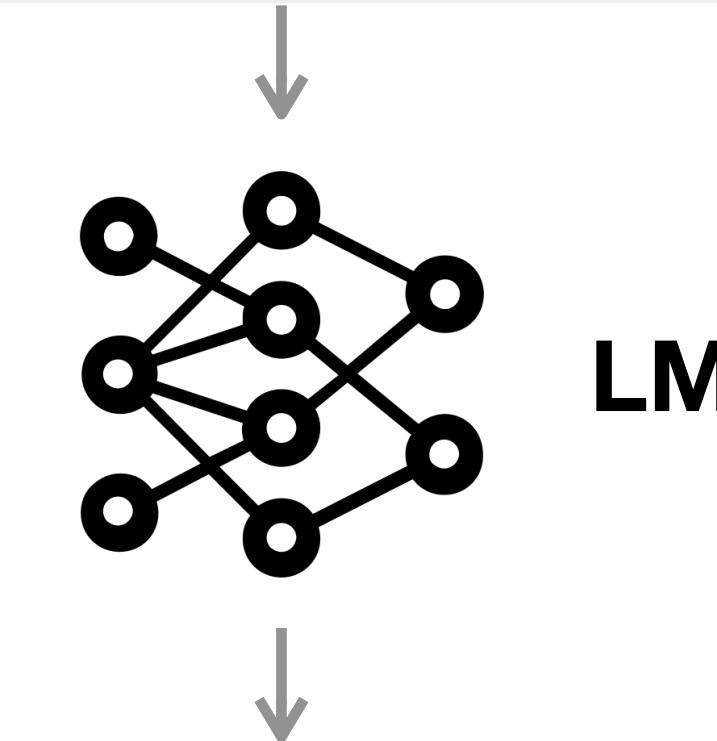
Retriever
(e.g., Google
BM 25)



Easily distracted by unhelpful documents

Step 2: Prompt LM with K docs and generate

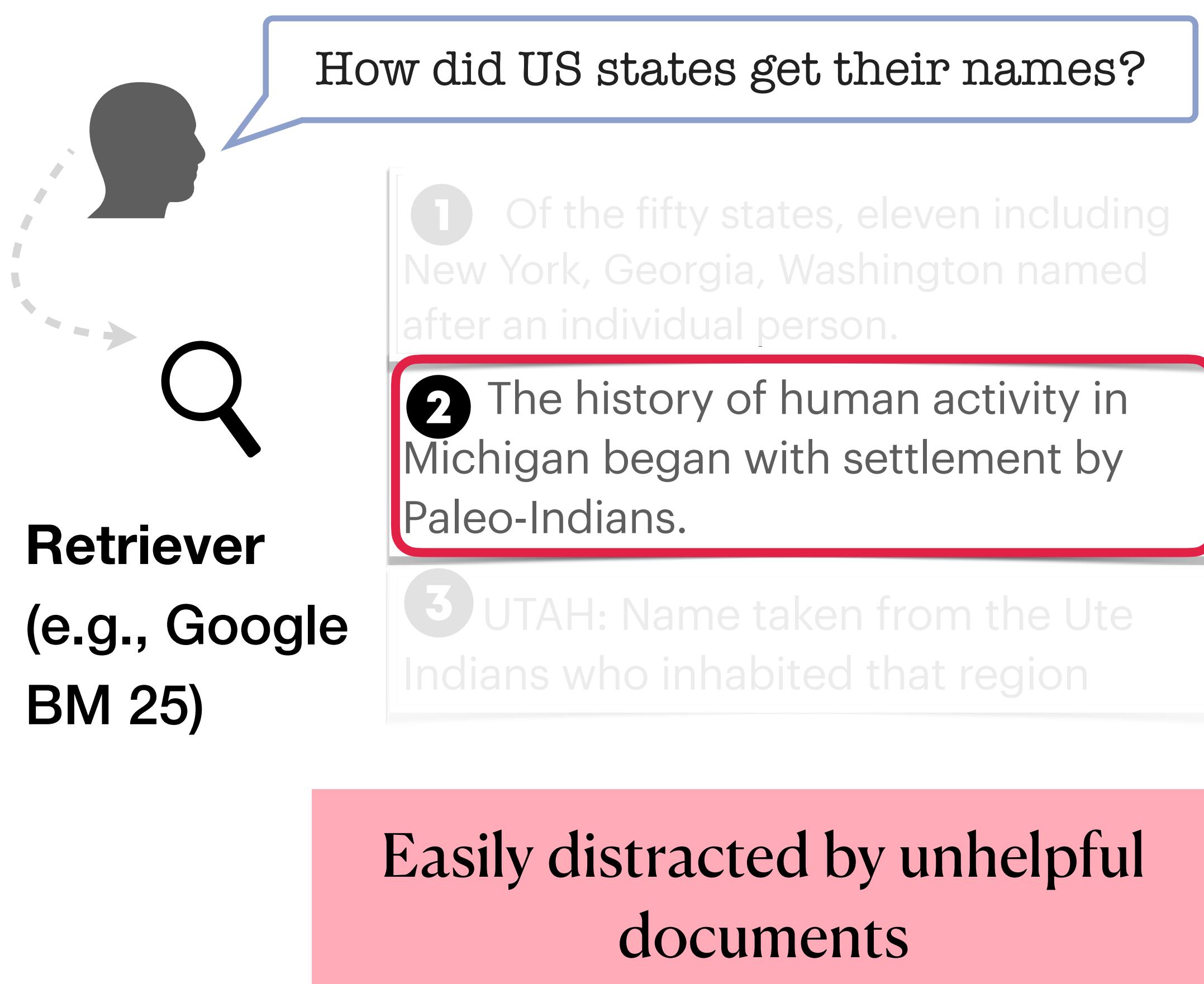
Prompt How did US states get their names? 1 2 3



US states got their names from a variety of sources. Eleven states such as New York and Michigan are named after an individual person. Some states including Utah and Washington, are named after indigenous communities.

Standard RAG isn't (always) reliable

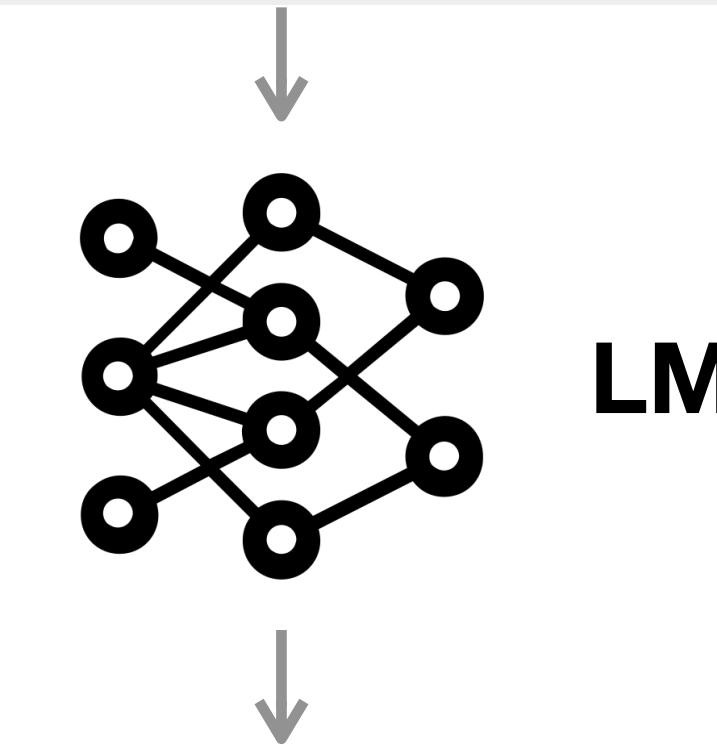
Step 1: Retrieve K documents



Retriever
(e.g., Google
BM 25)

Step 2: Prompt LM with K docs and generate

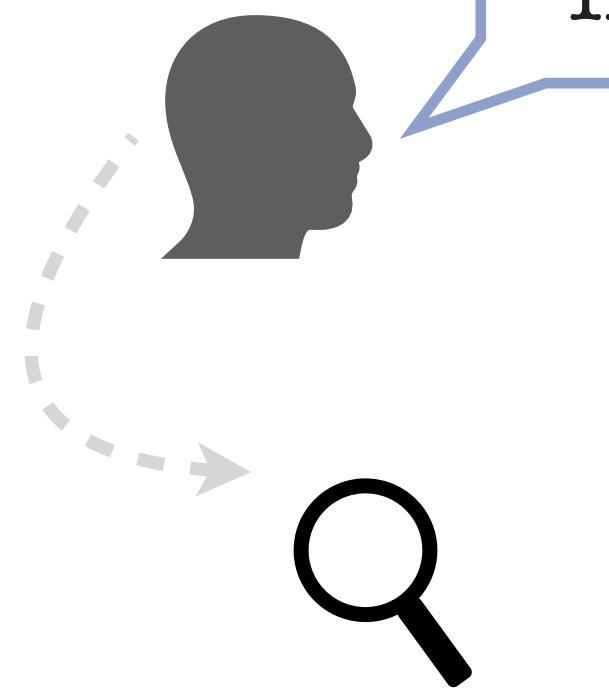
Prompt How did US states get their names? 1 2 3



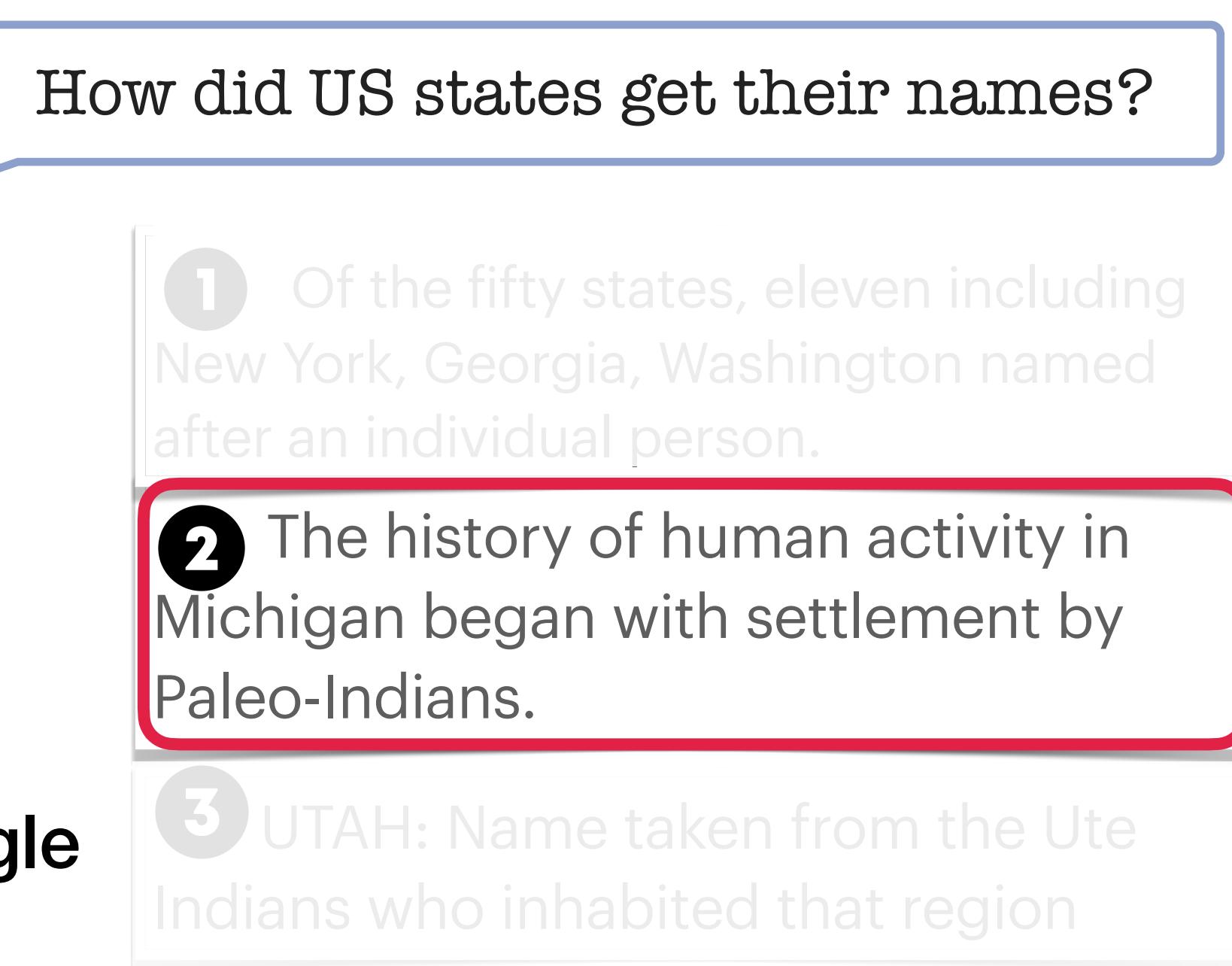
US states got their names from a variety of sources. Eleven states such as New York and Michigan are named after an individual person. Some states including Utah and Washington, are named after indigenous communities.

Standard RAG isn't (always) reliable

Step 1: Retrieve K documents



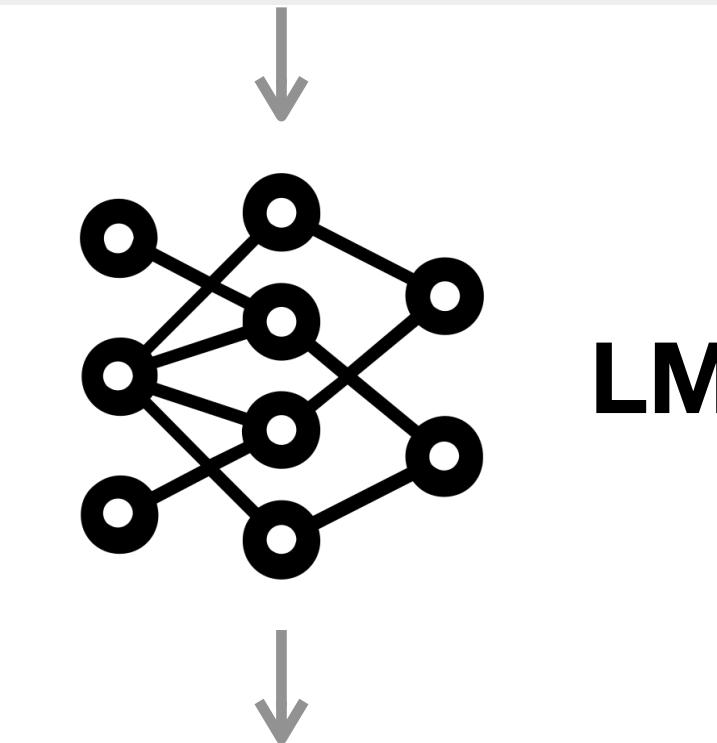
Retriever
(e.g., Google
BM 25)



Easily distracted by unhelpful documents

Step 2: Prompt LM with K docs and generate

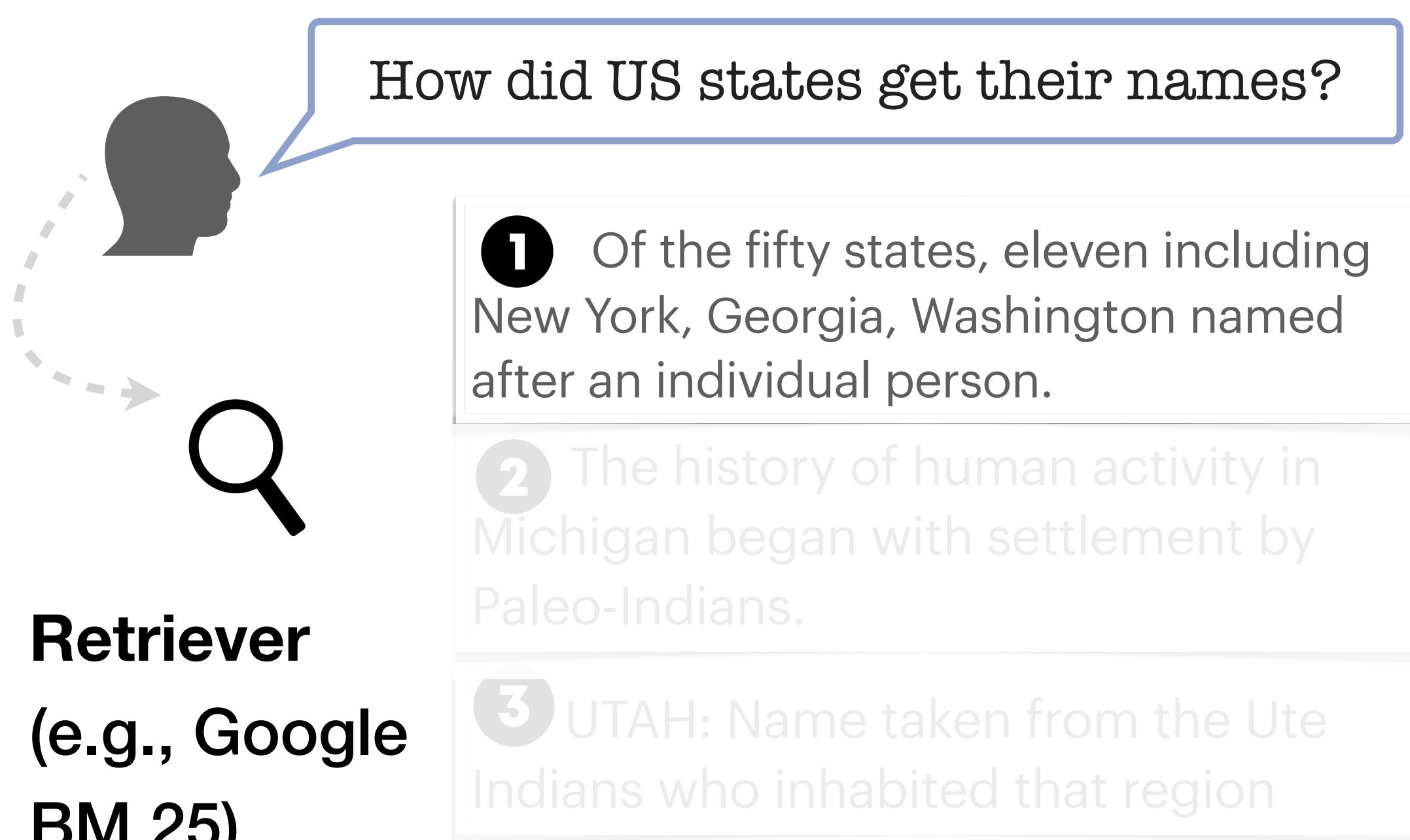
Prompt How did US states get their names? 1 2 3



US states got their names from a variety of sources. **Eleven states such as New York and Michigan are named after an individual person.** Some states including Utah and Washington, are named after indigenous communities.

Standard RAG isn't (always) reliable

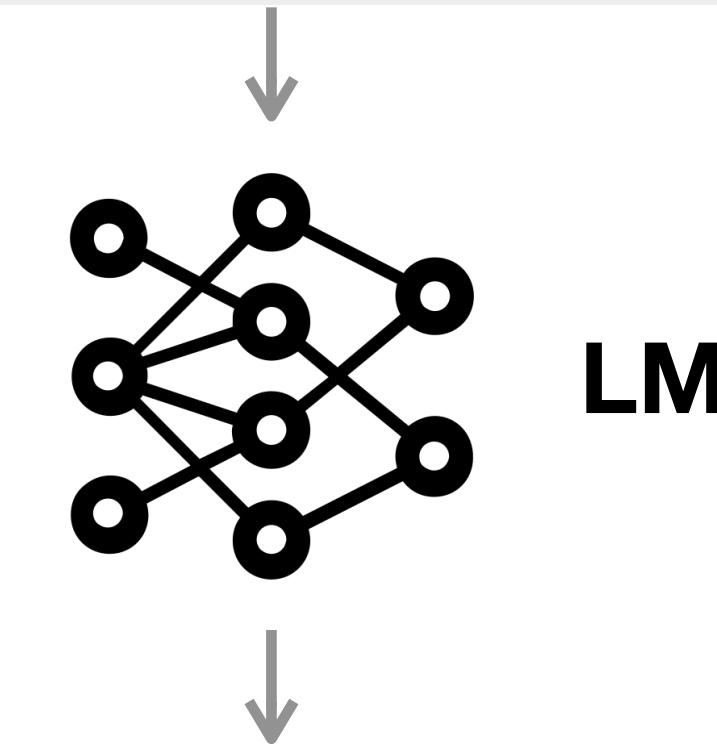
Step 1: Retrieve K documents



Retriever
(e.g., Google
BM 25)

Step 2: Prompt LM with K docs and generate

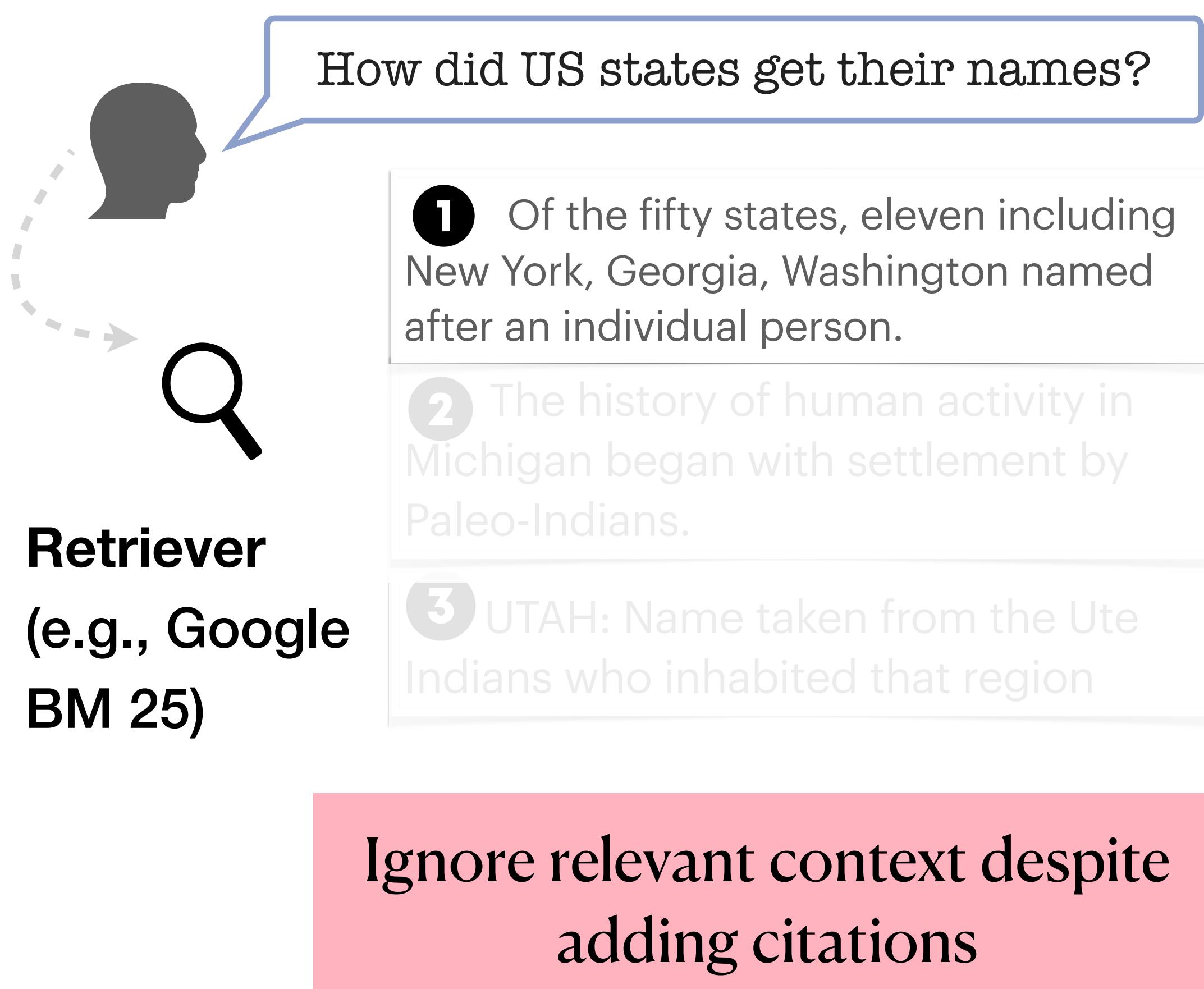
Prompt How did US states get their names? 1 2 3



US states got their names from a variety of sources. Eleven states such as New York and Michigan are named after an individual person. Some states including Utah and Washington, are named after indigenous communities.

Standard RAG isn't (always) reliable

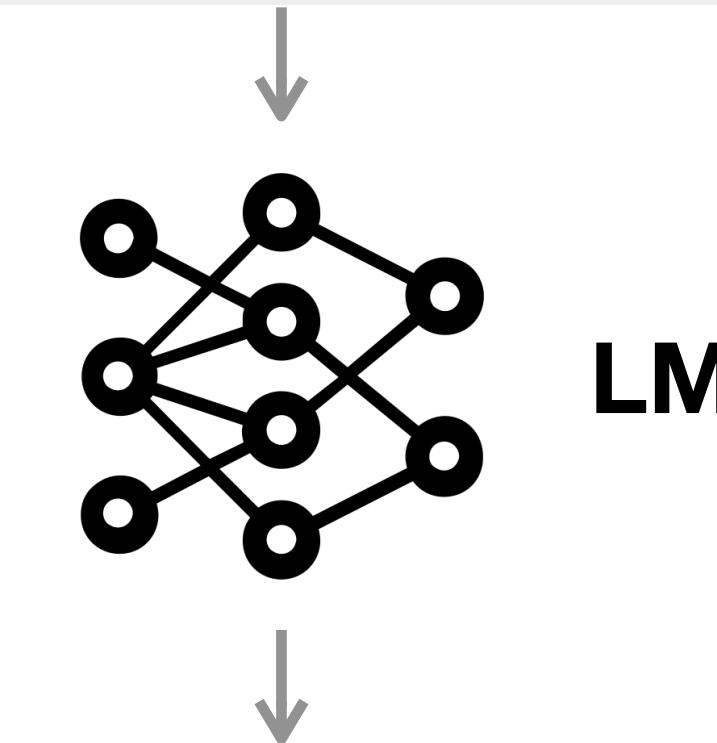
Step 1: Retrieve K documents



Retriever
(e.g., Google
BM 25)

Step 2: Prompt LM with K docs and generate

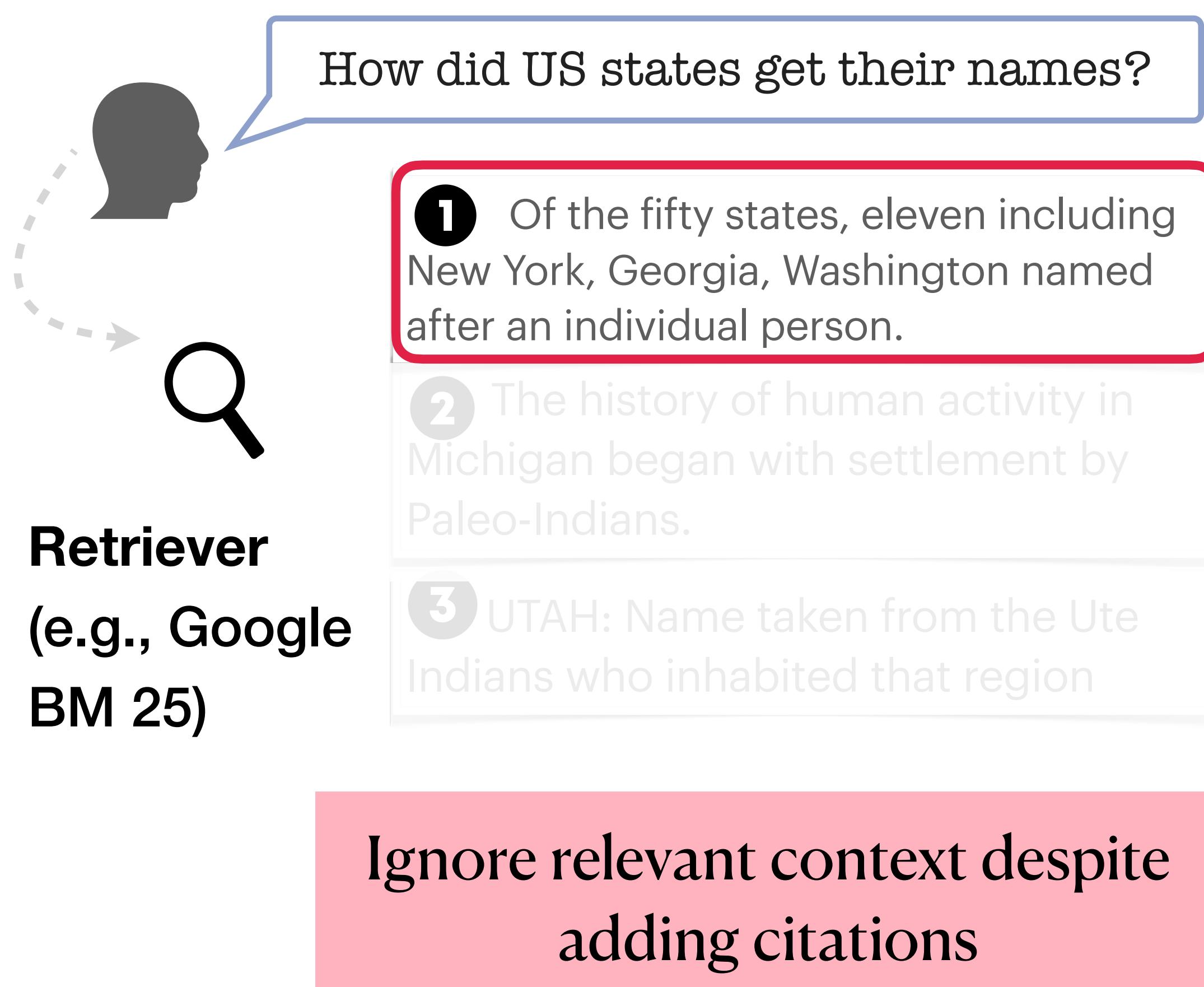
Prompt How did US states get their names? 1 2 3



US states got their names from a variety of sources. Eleven states such as New York and Michigan are named after an individual person. Some states including Utah and Washington, are named after indigenous communities.

Standard RAG isn't (always) reliable

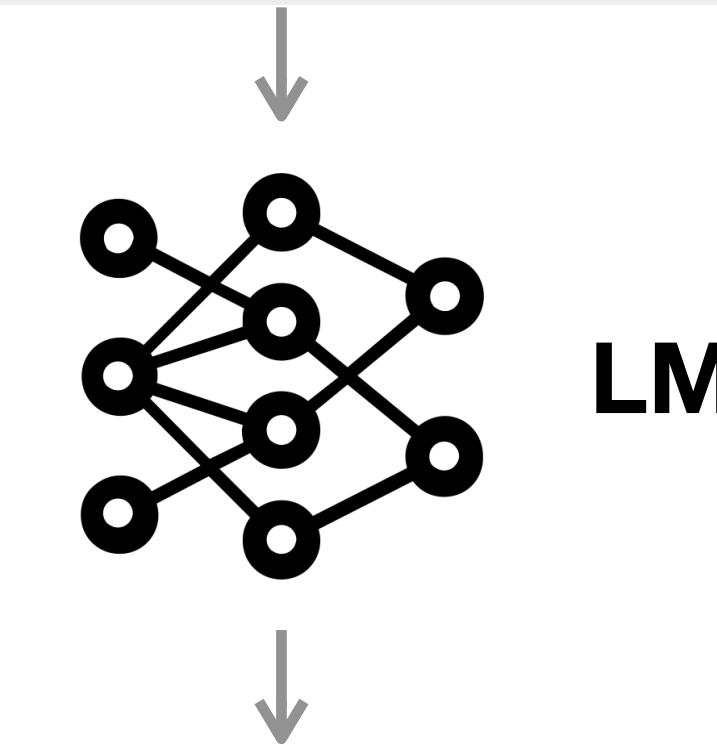
Step 1: Retrieve K documents



Retriever
(e.g., Google
BM 25)

Step 2: Prompt LM with K docs and generate

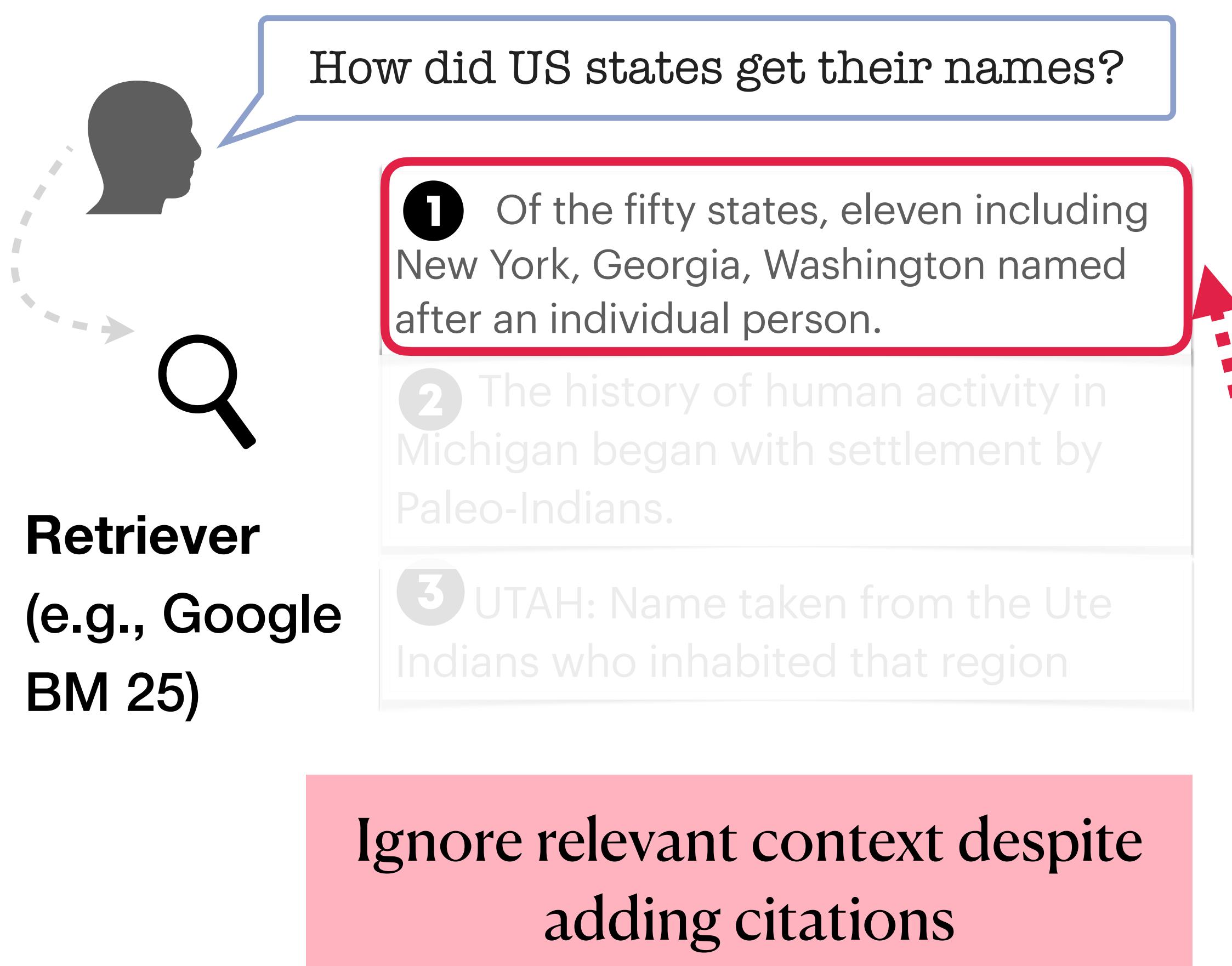
Prompt How did US states get their names? 1 2 3



US states got their names from a variety of sources. Eleven states such as New York and Michigan are named after an individual person. Some states including Utah and Washington, are named after indigenous communities.

Standard RAG isn't (always) reliable

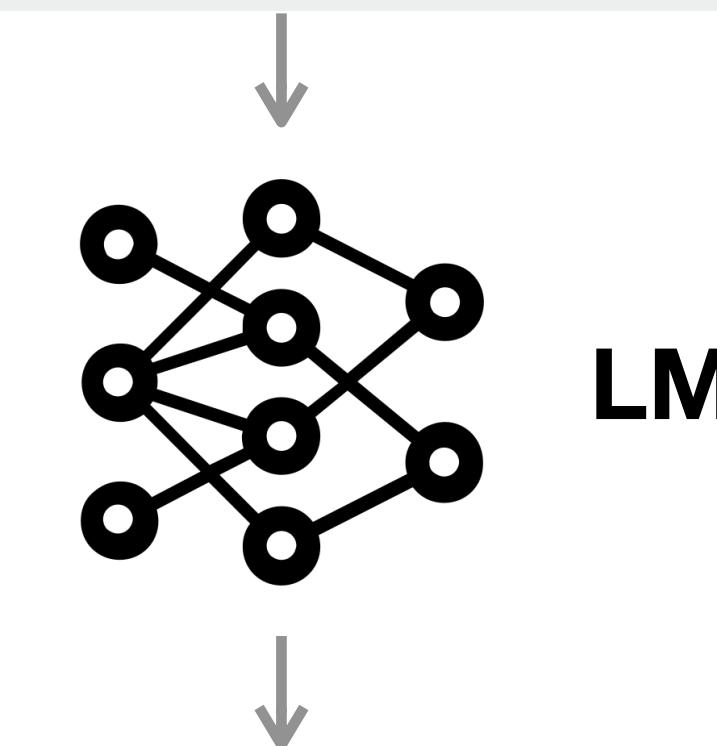
Step 1: Retrieve K documents



Retriever
(e.g., Google
BM 25)

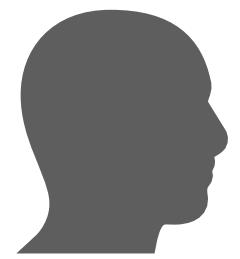
Step 2: Prompt LM with K docs and generate

Prompt How did US states get their names? 1 2 3



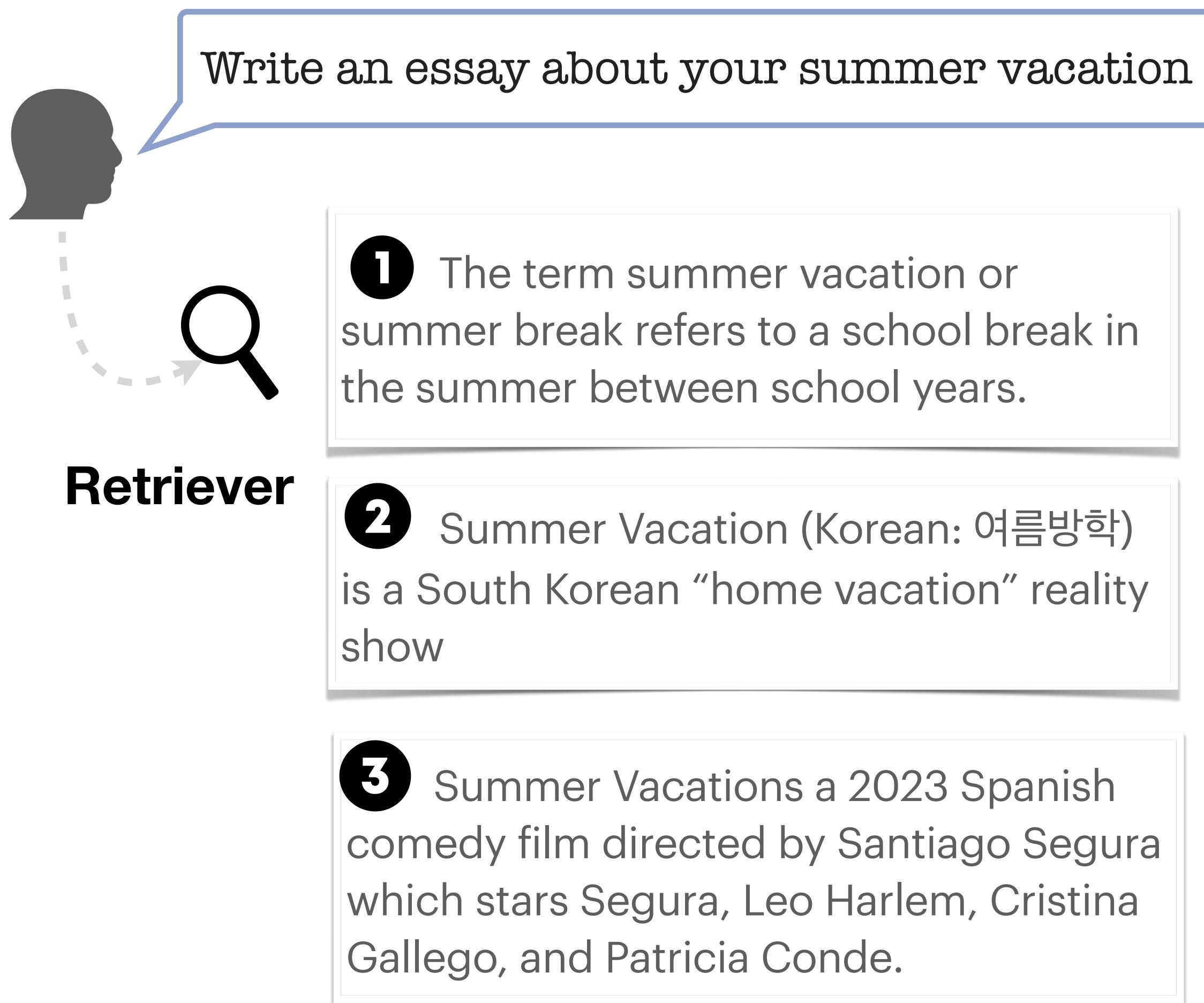
US states got their names from a variety of sources. Eleven states such as New York and Michigan are named after an individual person. **Some states including Utah and Washington**, are named after indigenous communities.

Standard RAG isn't efficient

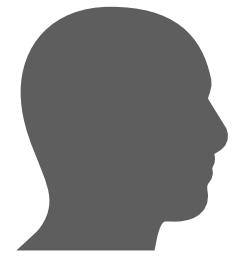


Write an essay about your summer vacation

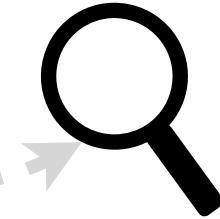
Standard RAG isn't efficient



Standard RAG isn't efficient



Write an essay about your summer vacation



Retriever

- 1** The term summer vacation or summer break refers to a school break in the summer between school years.
- 2** Summer Vacation (Korean: 여름방학) is a South Korean “home vacation” reality show
- 3** Summer Vacations a 2023 Spanish comedy film directed by Santiago Segura which stars Segura, Leo Harlem, Cristina Gallego, and Patricia Conde.

Always retrieving fixed number of documents is inefficient & harmful

Self-RAG: *Self*-reflective Retrieval-Augmented Generation

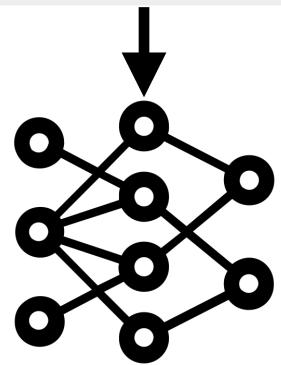
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens

Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens

User input



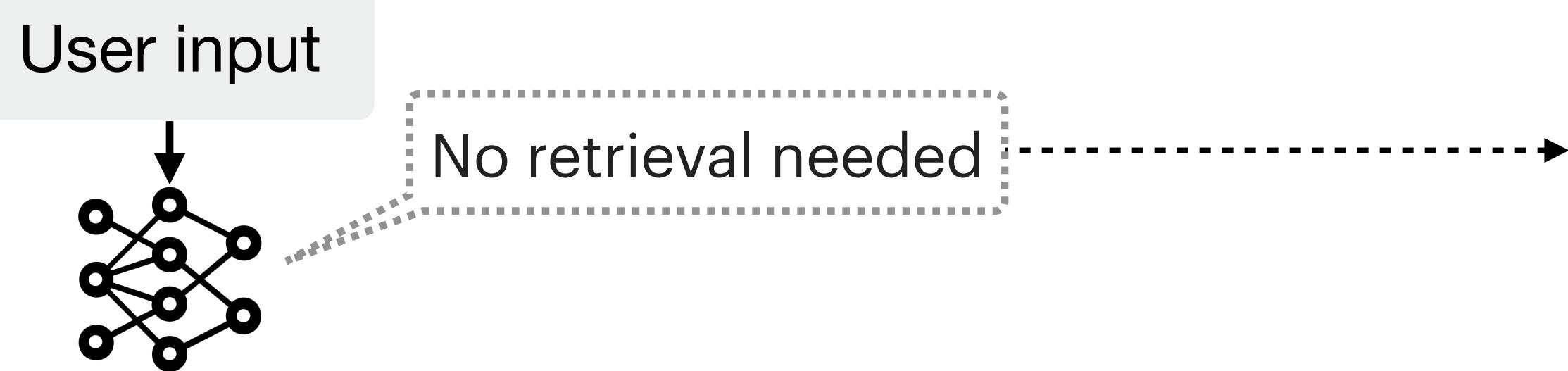
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



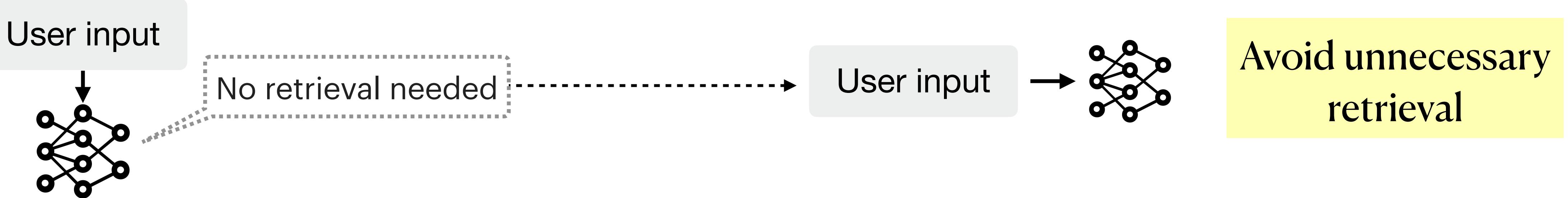
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



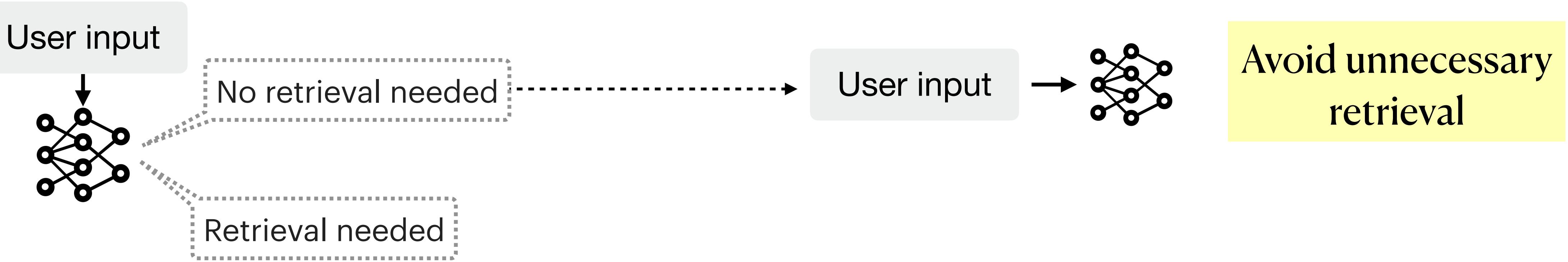
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



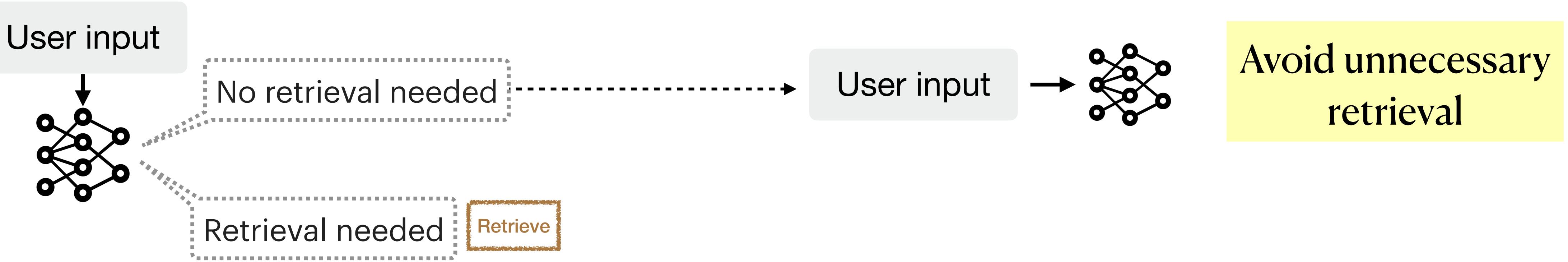
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



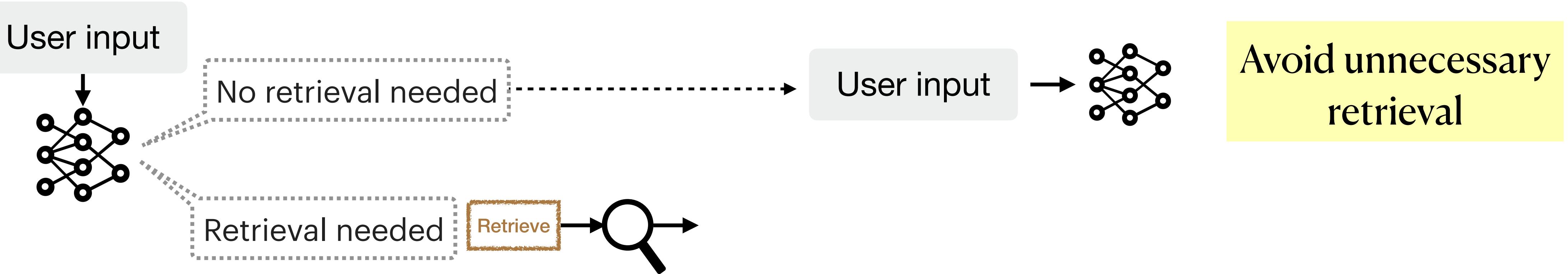
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



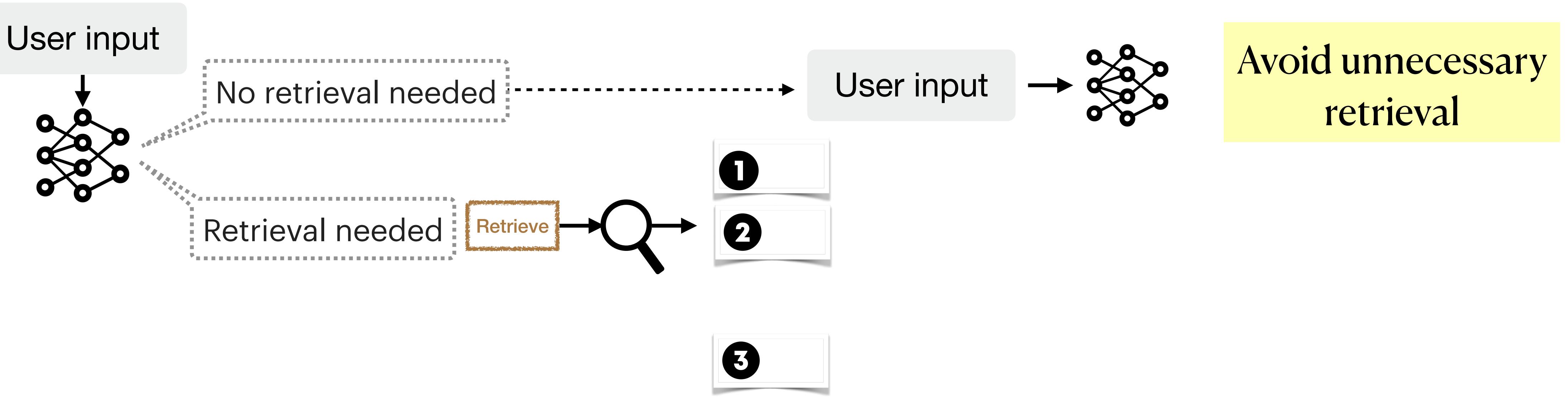
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



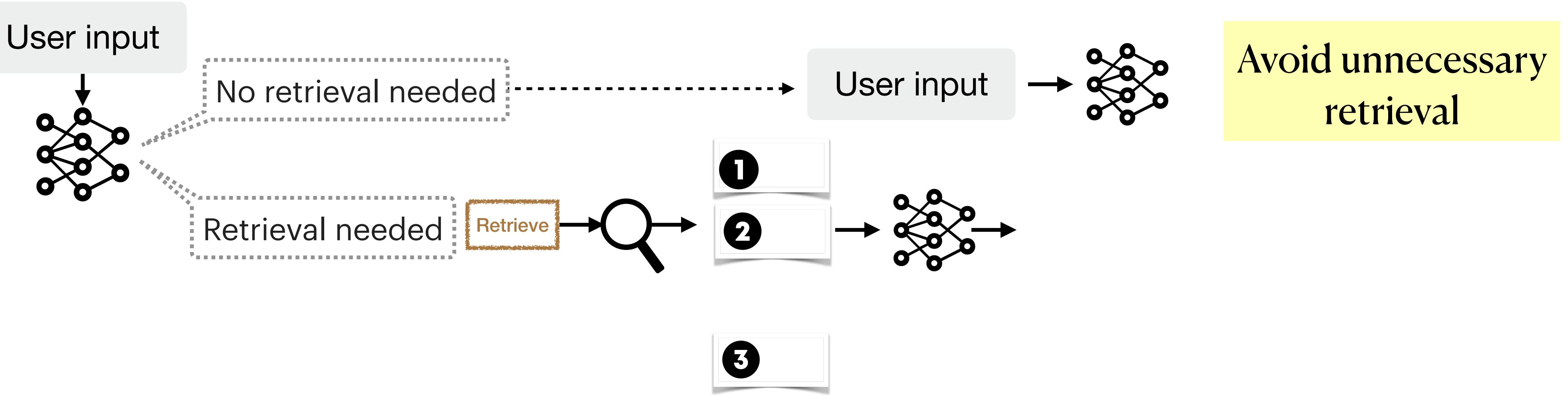
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



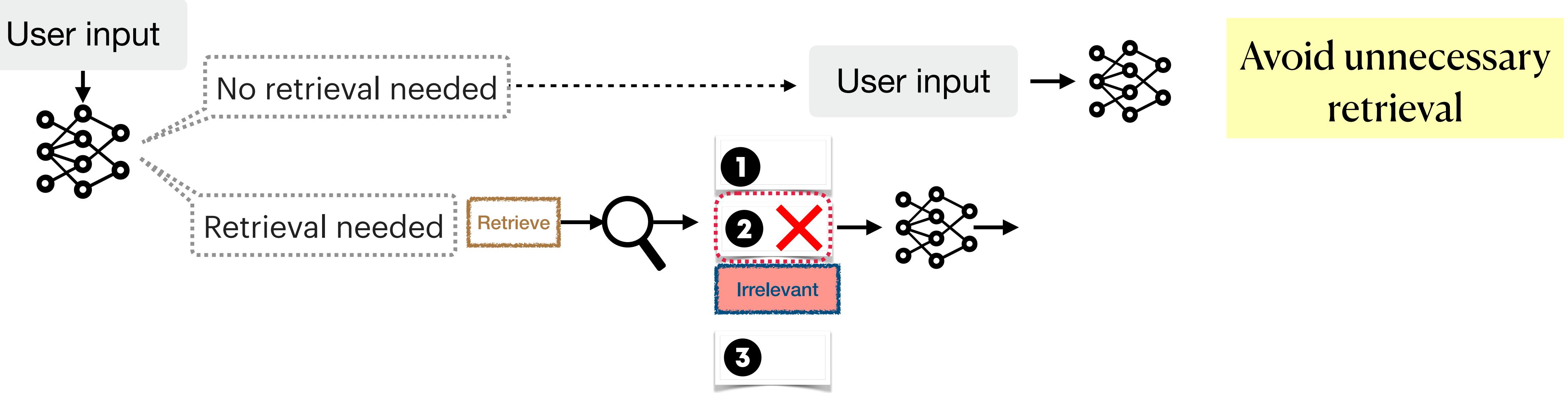
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



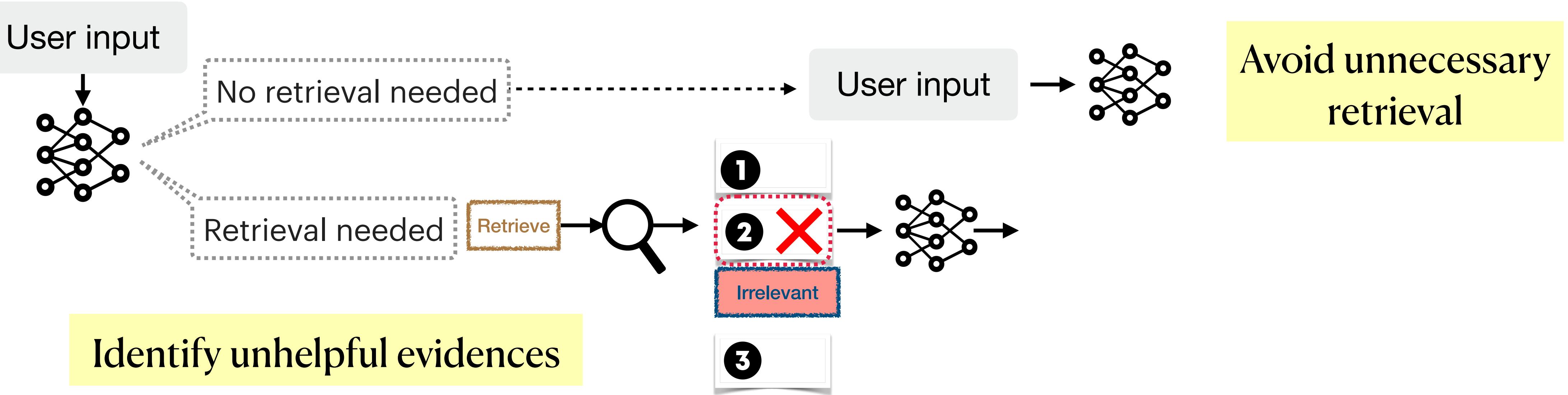
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



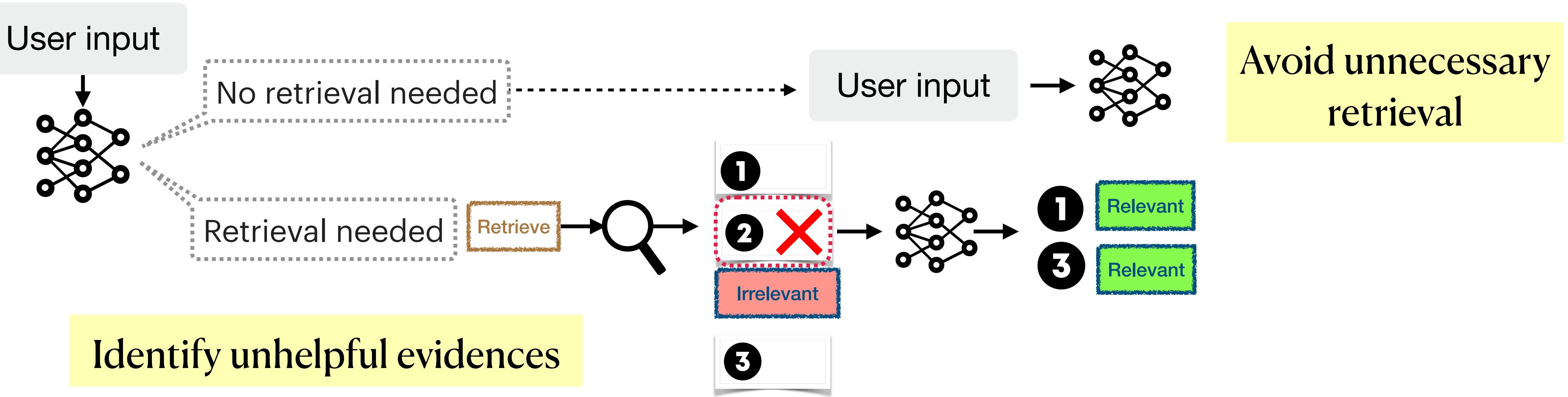
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



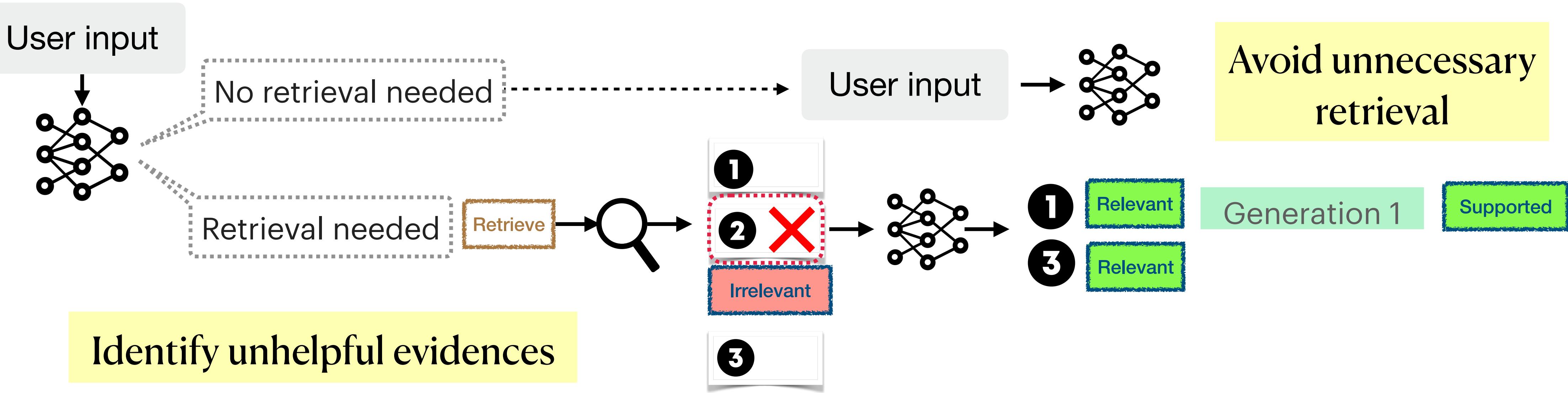
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



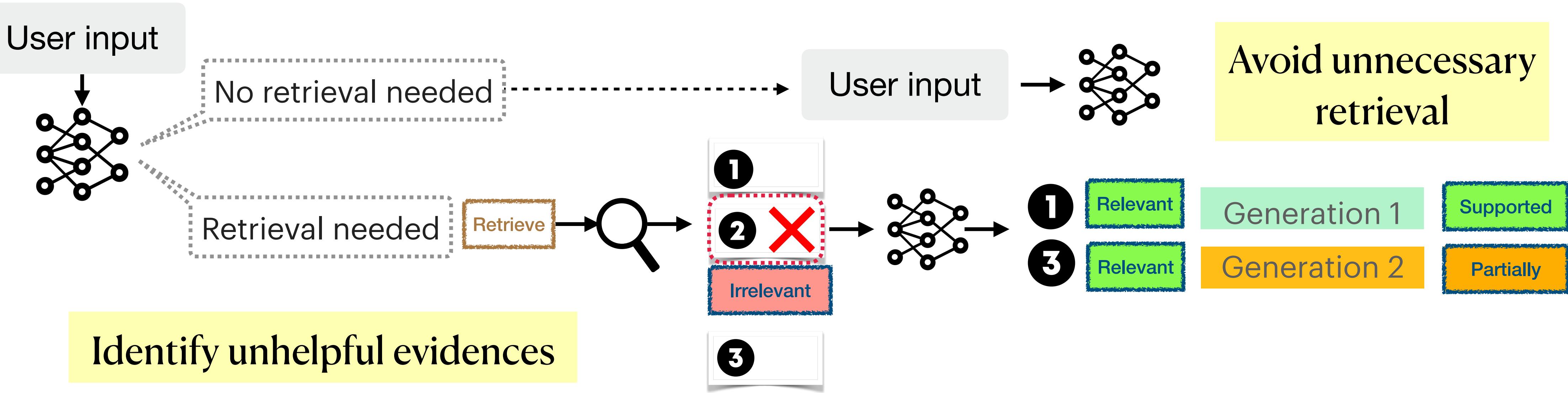
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



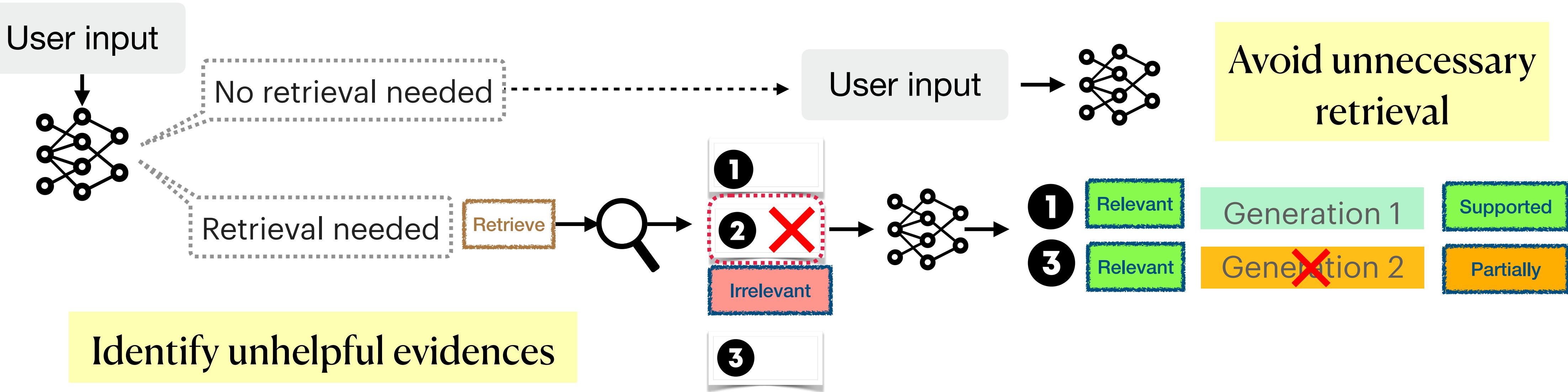
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



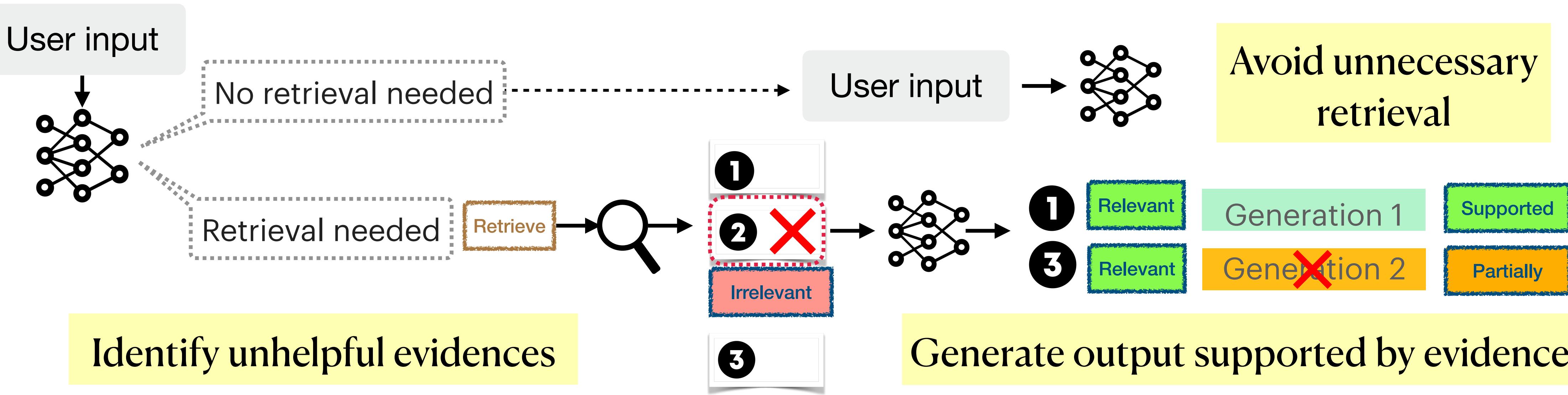
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



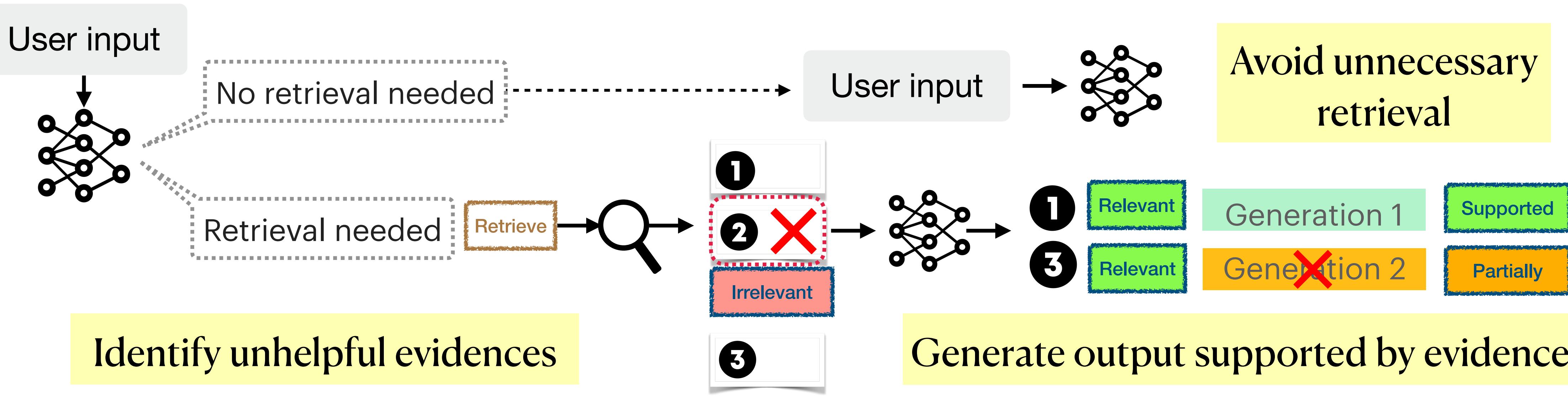
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



Self-RAG: *Self*-reflective Retrieval-Augmented Generation

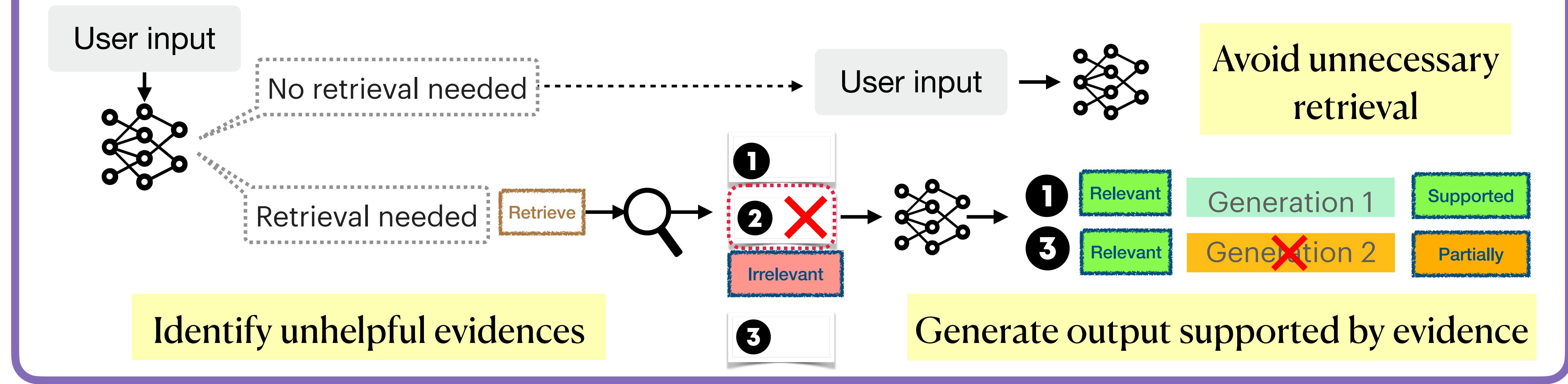
Self-RAG Inference: Learn to retrieve and reflect via special tokens



Self-RAG training: Instruction-tuning an arbitrary LM with reflection tokens

Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens

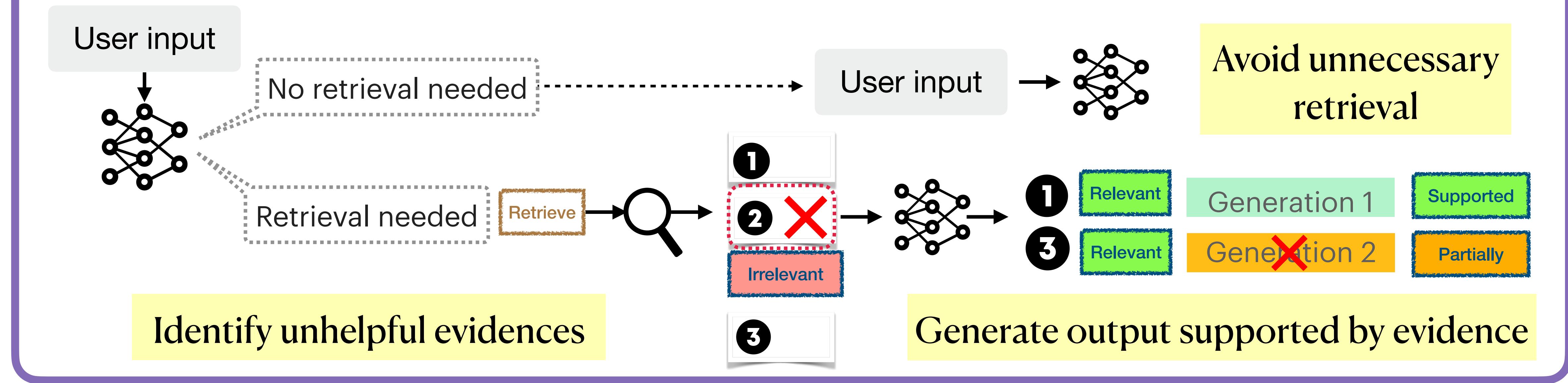


Self-RAG training: Instruction-tuning an arbitrary LM with reflection tokens

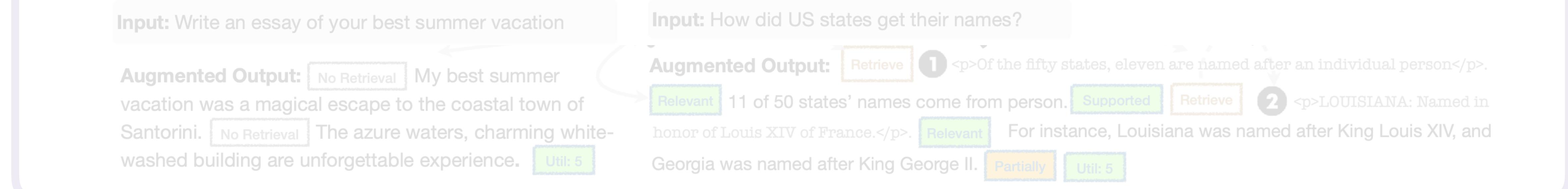


Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and reflect via special tokens



Self-RAG training: Instruction-tuning an arbitrary LM with reflection tokens



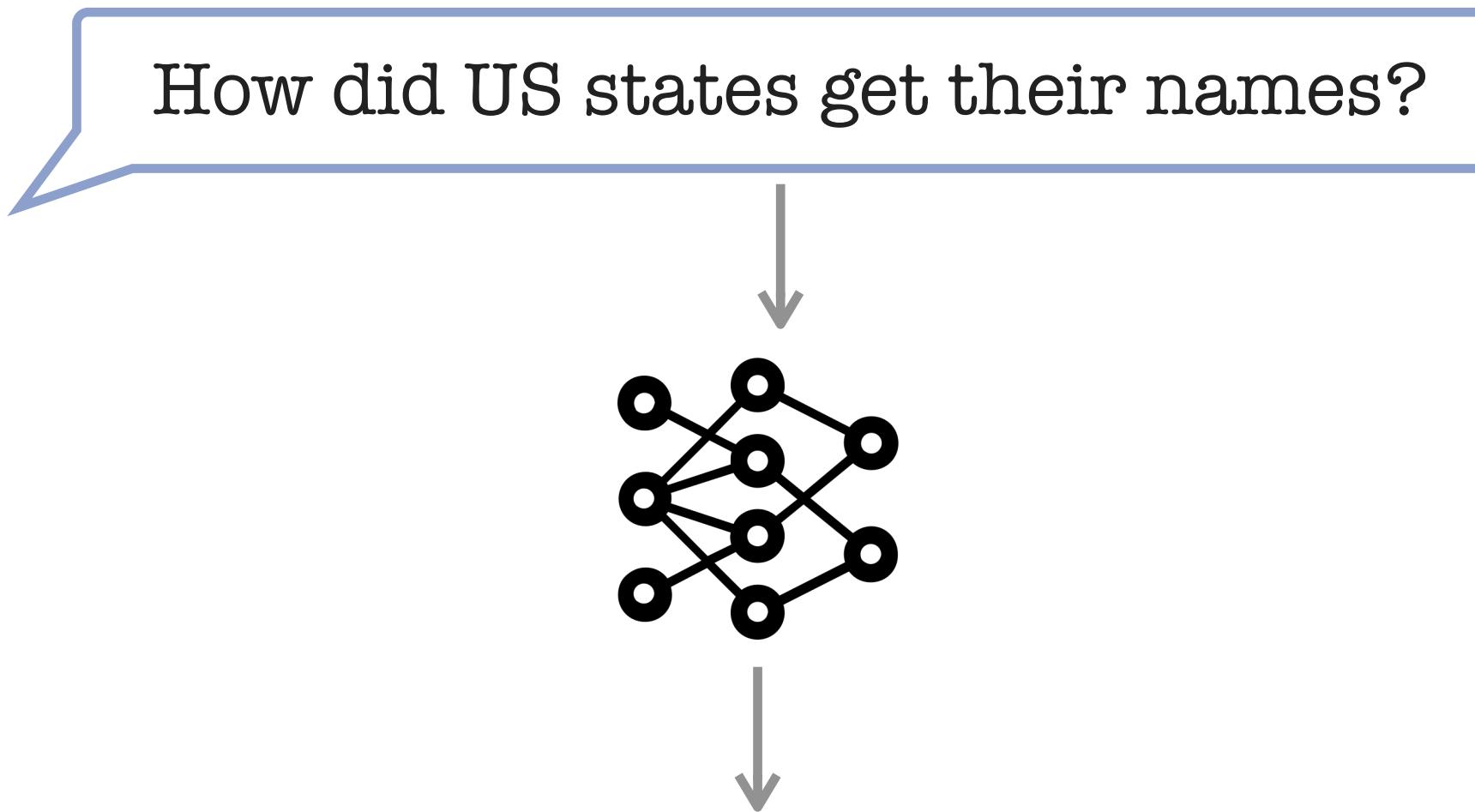
Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 1: Retrieve documents on demand

How did US states get their names?

Self-RAG – Self-Reflective Retrieval-Augmented Generation

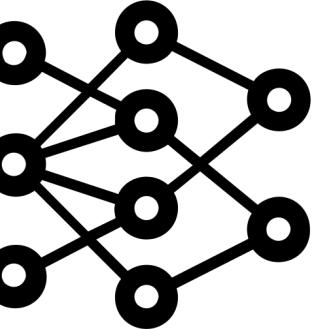
Step 1: Retrieve documents on demand



Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 1: Retrieve documents on demand

How did US states get their names?

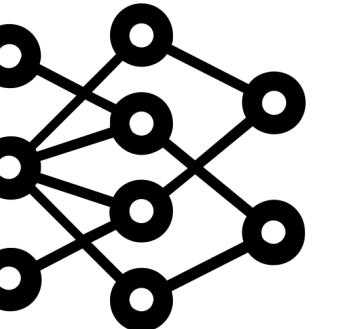


US states got their names
from a variety of sources.

Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 1: Retrieve documents on demand

How did US states get their names?

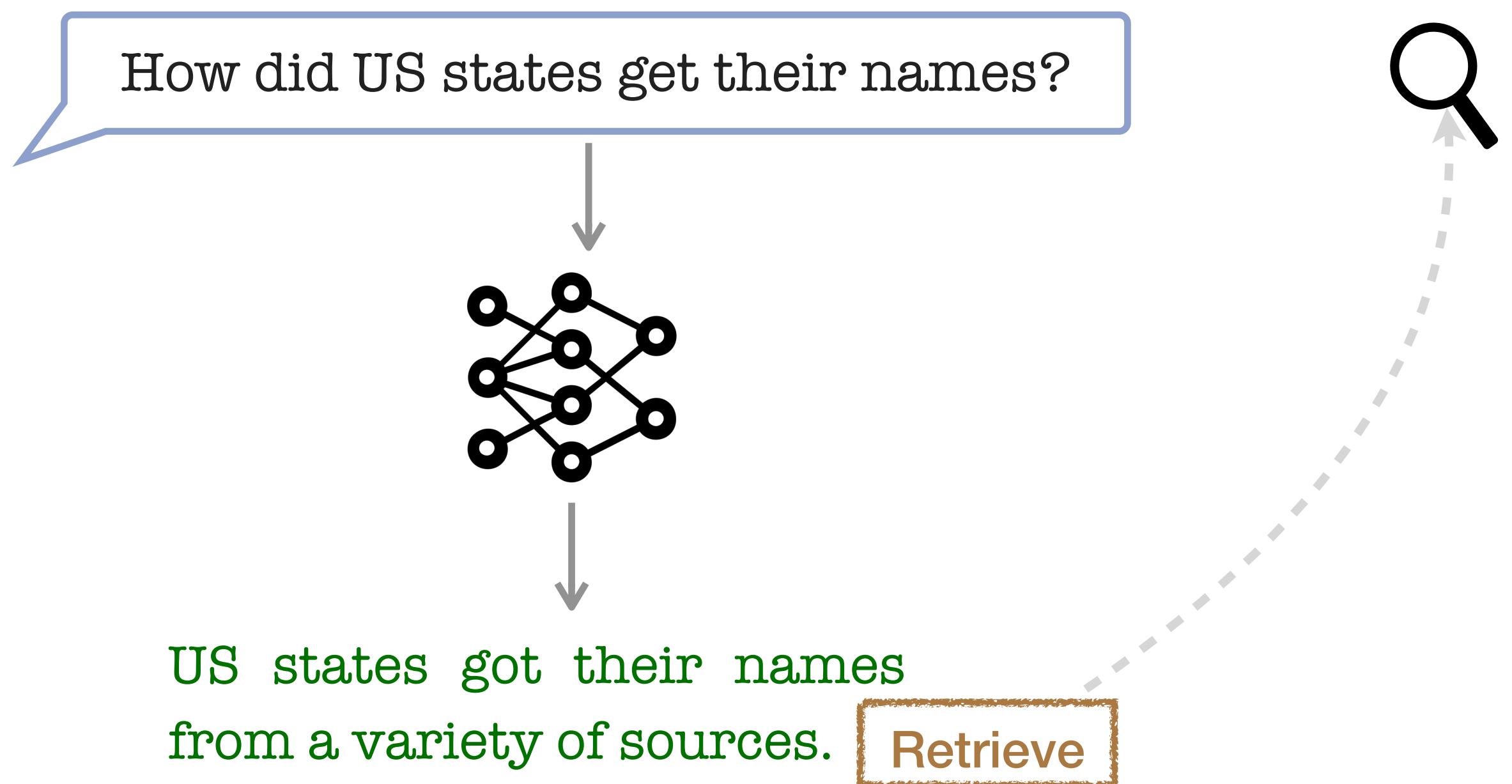


US states got their names
from a variety of sources.

Retrieve

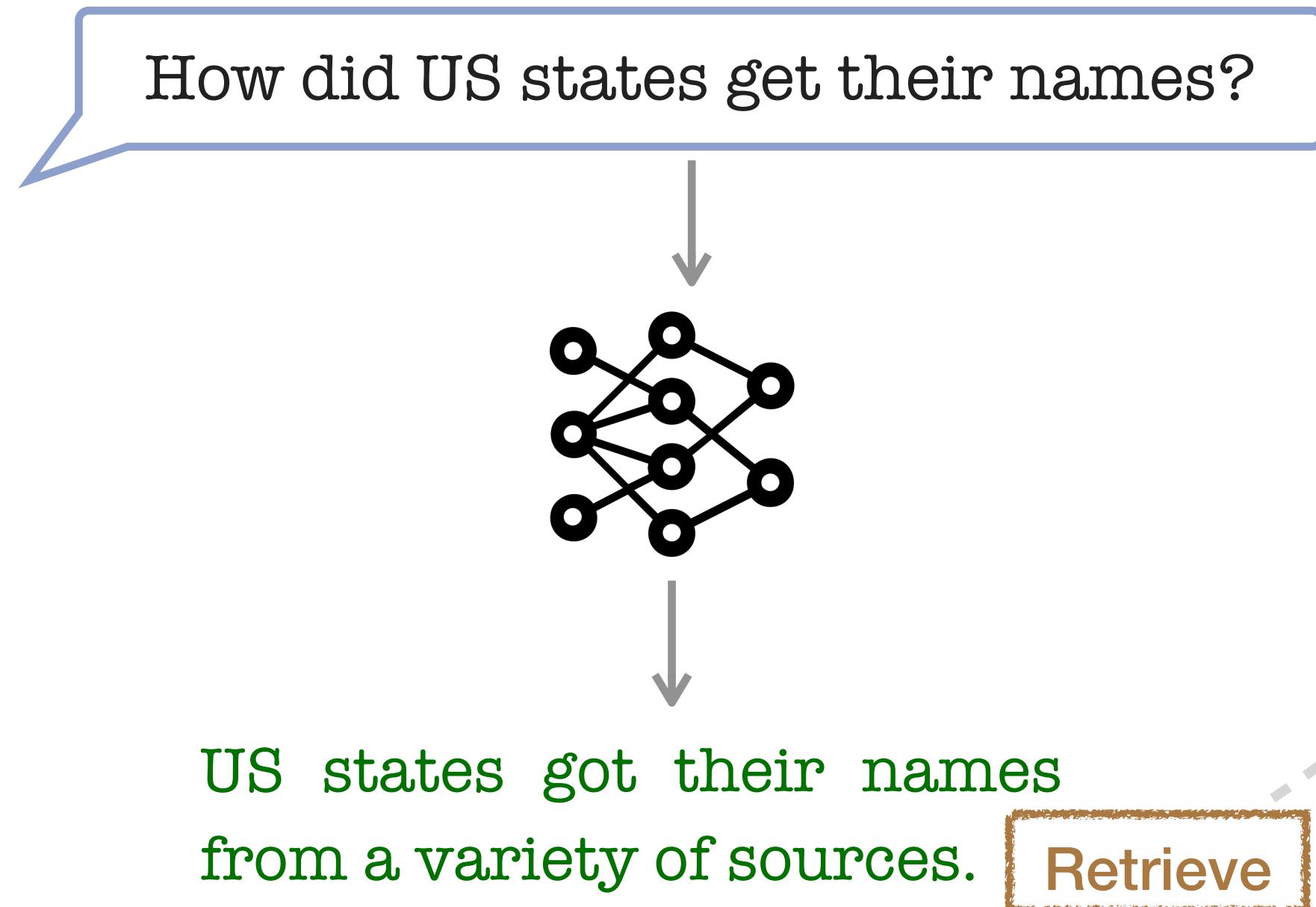
Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 1: Retrieve documents on demand

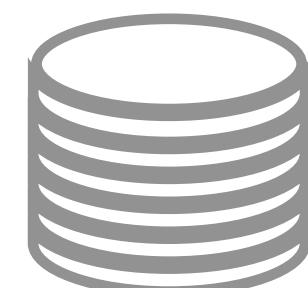


Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 1: Retrieve documents on demand

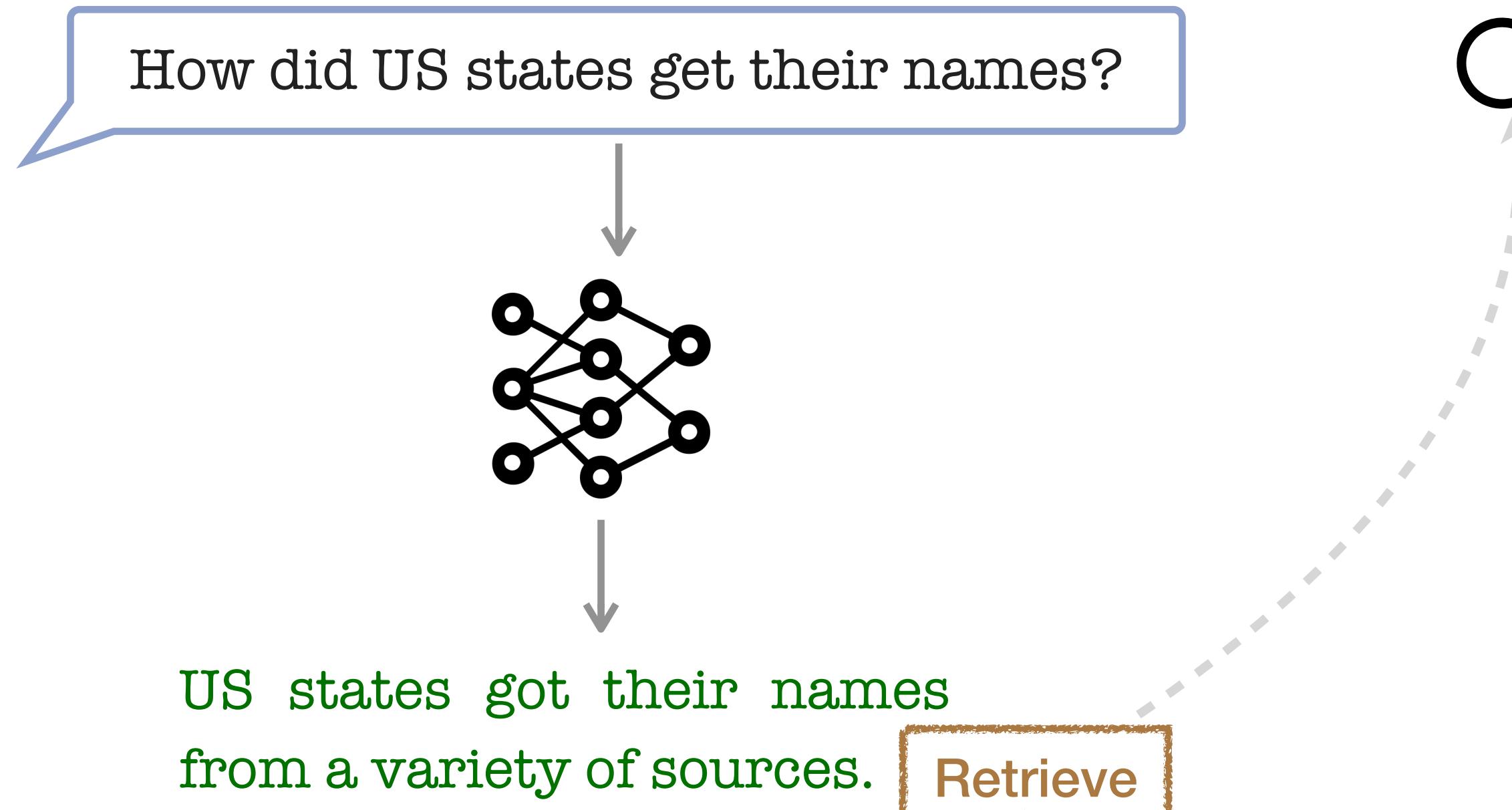


- 1** Of the fifty states, eleven including New York, Georgia, Washington named after an individual person.
- 2** The history of human activity in Michigan began with settlement by Paleo-Indians.
- 3** UTAH: Name taken from the Ute Indians who inhabited that region

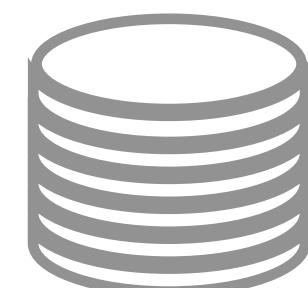


Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 1: Retrieve documents on demand



- 1 Of the fifty states, eleven including New York, Georgia, Washington named after an individual person.
- 2 The history of human activity in Michigan began with settlement by Paleo-Indians.
- 3 UTAH: Name taken from the Ute Indians who inhabited that region



Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 2: Generate segments in *parallel*

Prompt **1**

1 Of the fifty states, eleven states named after an individual person.

Prompt **2**

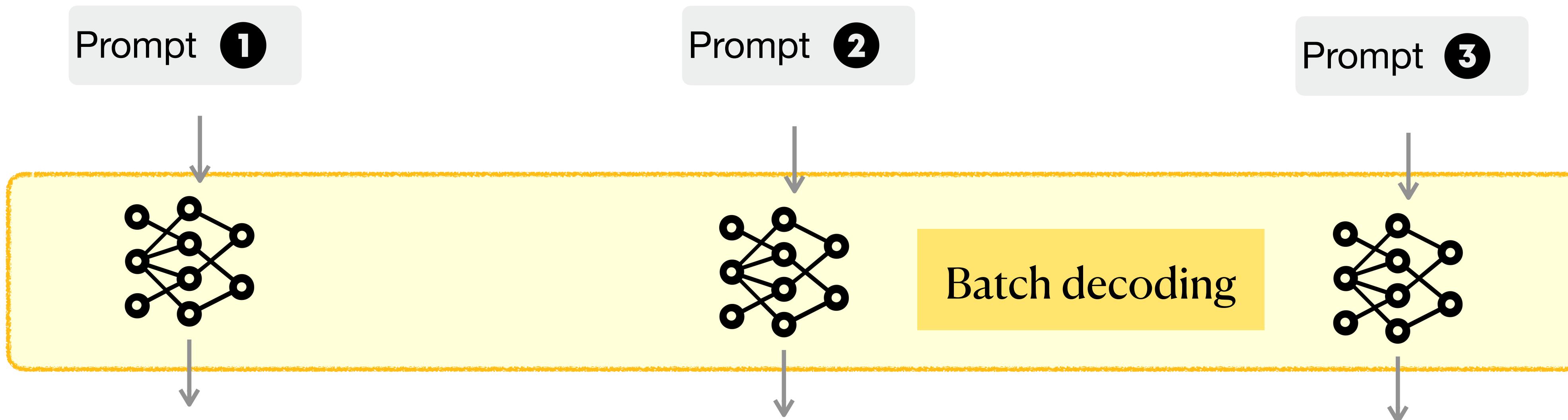
2 The history of human activity in Michigan began with settlement by Paleo-Indians.

Prompt **3**

3 UTAH: Name taken from the Ute Indians who inhabited that region

Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 2: Generate segments in *parallel*



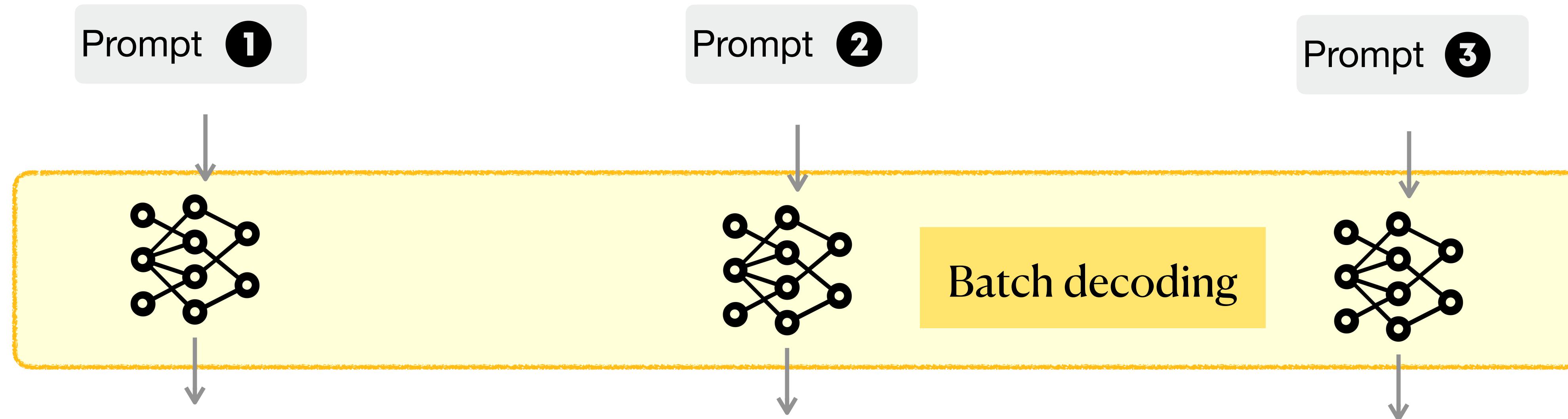
1 Of the fifty states, eleven states named after an individual person.

2 The history of human activity in Michigan began with settlement by Paleo-Indians.

3 UTAH: Name taken from the Ute Indians who inhabited that region

Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 2: Generate segments in *parallel*



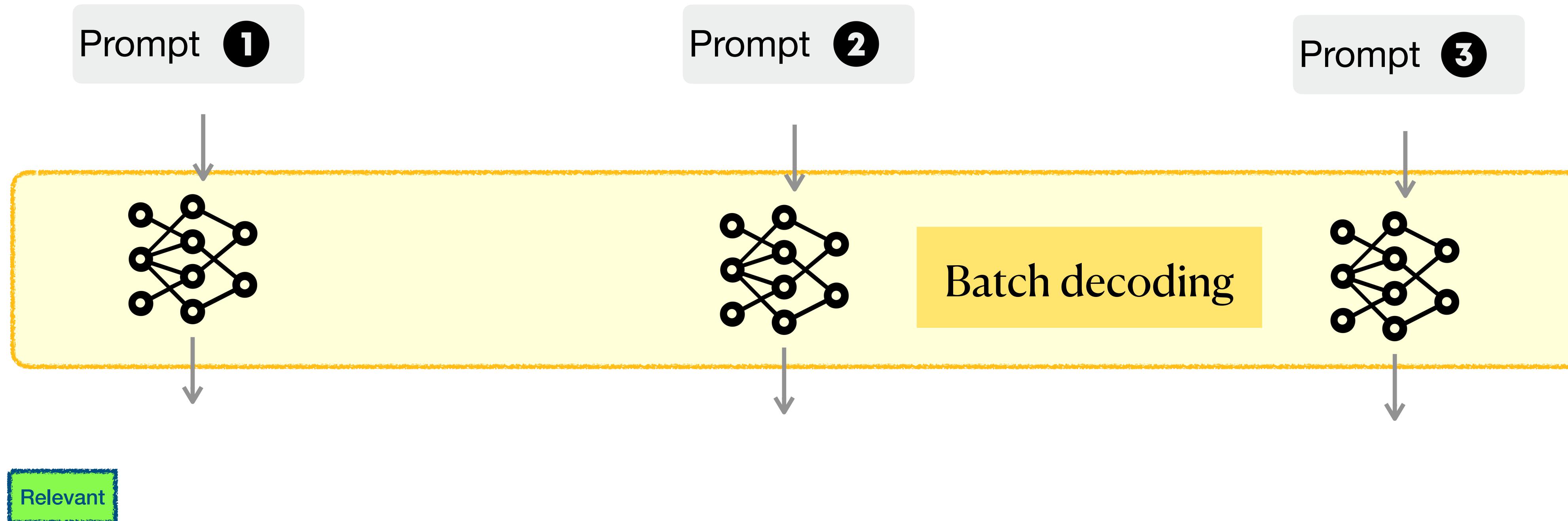
1 Of the fifty states, eleven states named after an individual person.

2 The history of human activity in Michigan began with settlement by Paleo-Indians.

3 UTAH: Name taken from the Ute Indians who inhabited that region

Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 2: Generate segments in *parallel*



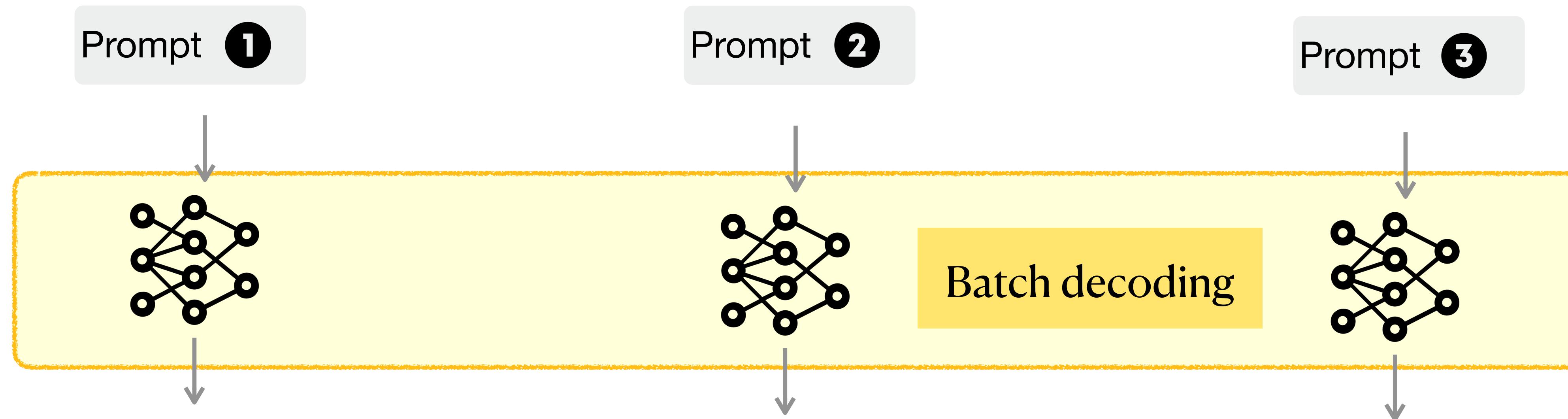
1 Of the fifty states, eleven states named after an individual person.

2 The history of human activity in Michigan began with settlement by Paleo-Indians.

3 UTAH: Name taken from the Ute Indians who inhabited that region

Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 2: Generate segments in *parallel*



Relevant

11 of 50 state names

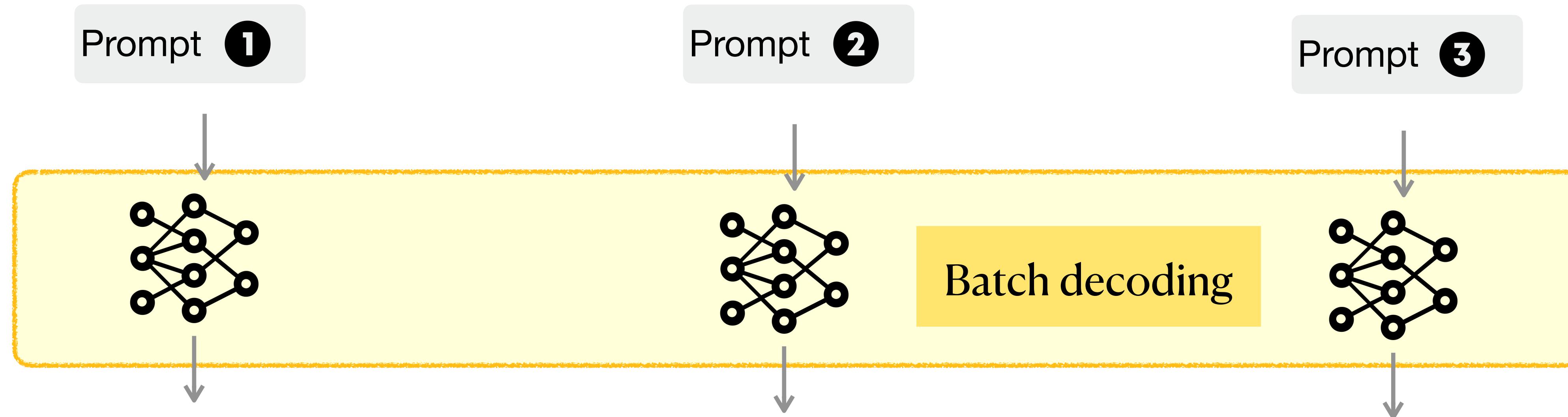
1 Of the fifty states, eleven states named after an individual person.

2 The history of human activity in Michigan began with settlement by Paleo-Indians.

3 UTAH: Name taken from the Ute Indians who inhabited that region

Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 2: Generate segments in *parallel*



Relevant

11 of 50 state names

come from persons.

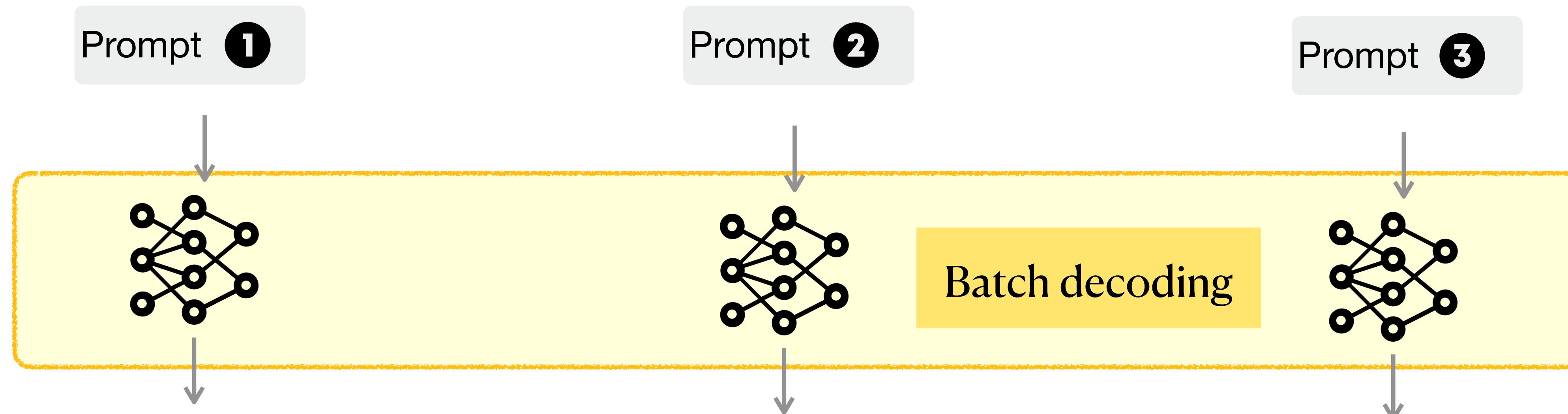
1 Of the fifty states, eleven states named after an individual person.

2 The history of human activity in Michigan began with settlement by Paleo-Indians.

3 UTAH: Name taken from the Ute Indians who inhabited that region

Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 2: Generate segments in *parallel*



come from persons.

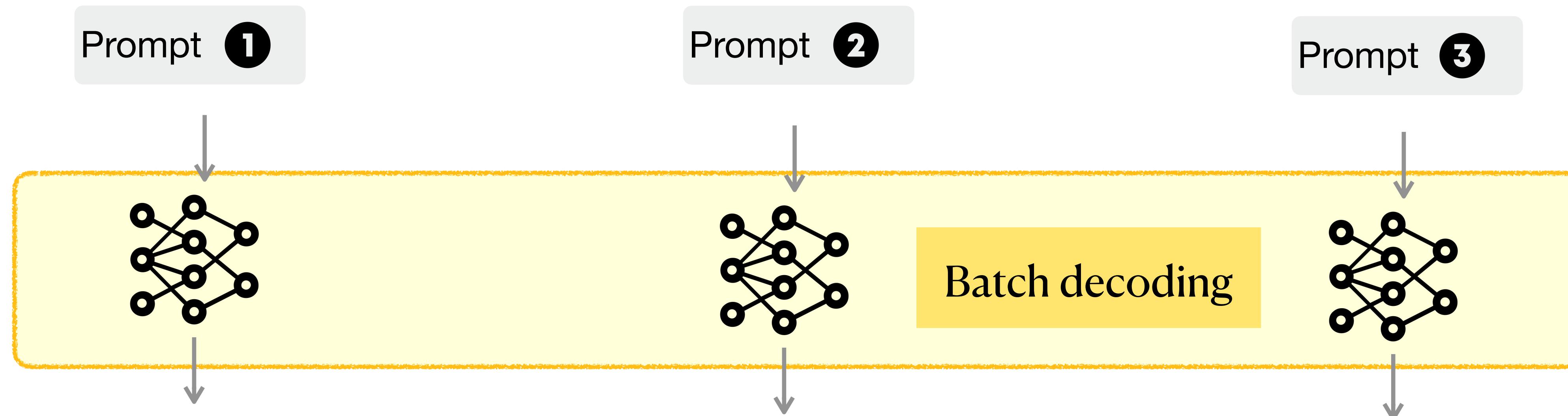
1 Of the fifty states, eleven states named after an individual person.

2 The history of human activity in Michigan began with settlement by Paleo-Indians.

3 UTAH: Name taken from the Ute Indians who inhabited that region

Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 2: Generate segments in *parallel*



Relevant

11 of 50 state names

come from persons.

Irrelevant

Many states including Michigan

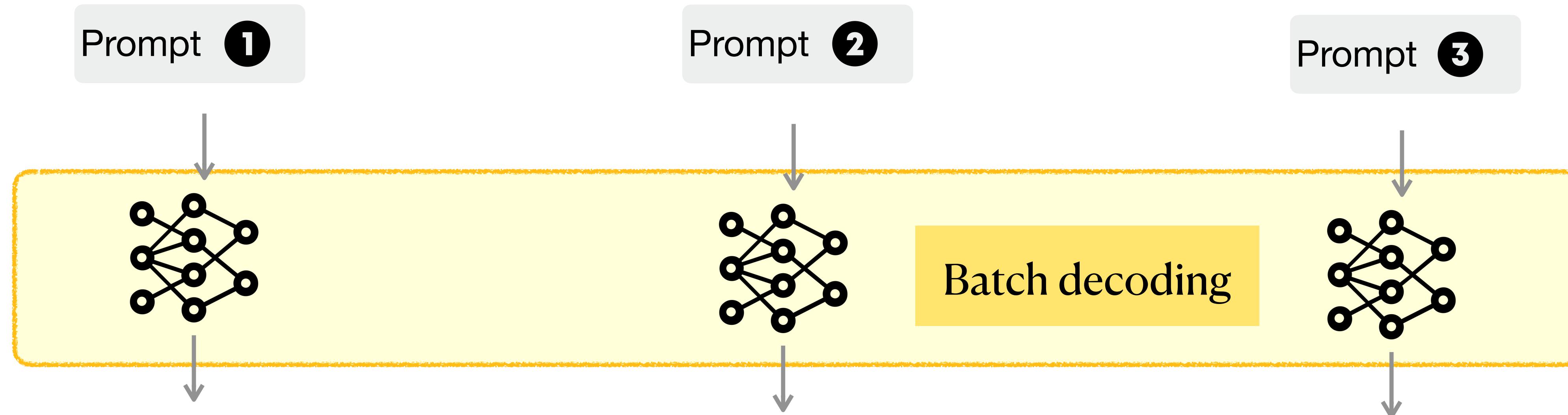
1 Of the fifty states, eleven states named after an individual person.

2 The history of human activity in Michigan began with settlement by Paleo-Indians.

3 UTAH: Name taken from the Ute Indians who inhabited that region

Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 2: Generate segments in *parallel*



Relevant

11 of 50 state names

come from persons.

Irrelevant

Many states including Michigan

are named after an individual person.

1

Of the fifty states, eleven states named after an individual person.

2

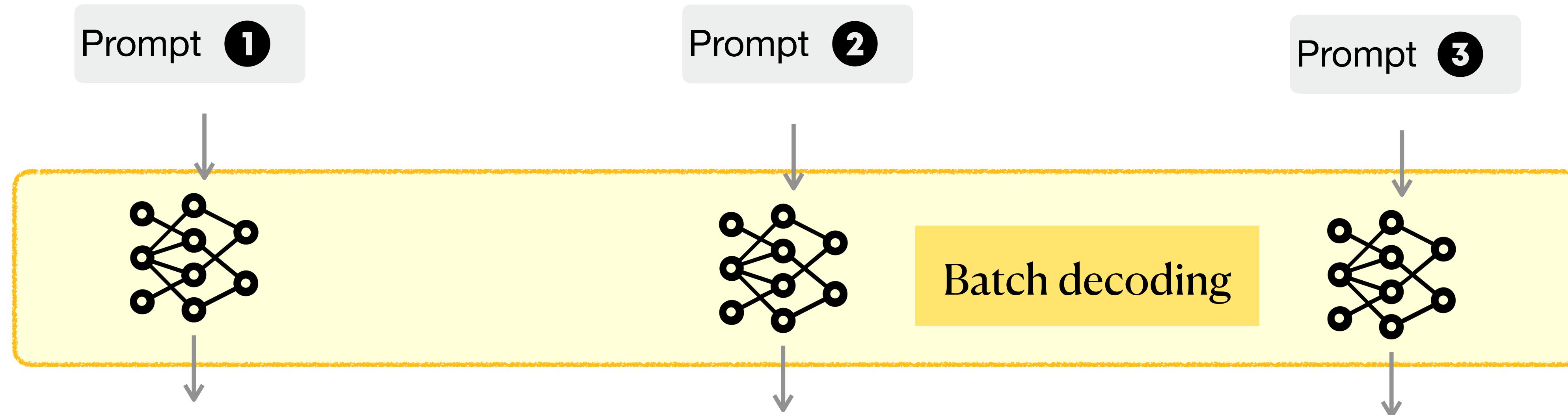
The history of human activity in Michigan began with settlement by Paleo-Indians.

3

UTAH: Name taken from the Ute Indians who inhabited that region

Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 2: Generate segments in *parallel*



come from persons.

are named after an individual person.

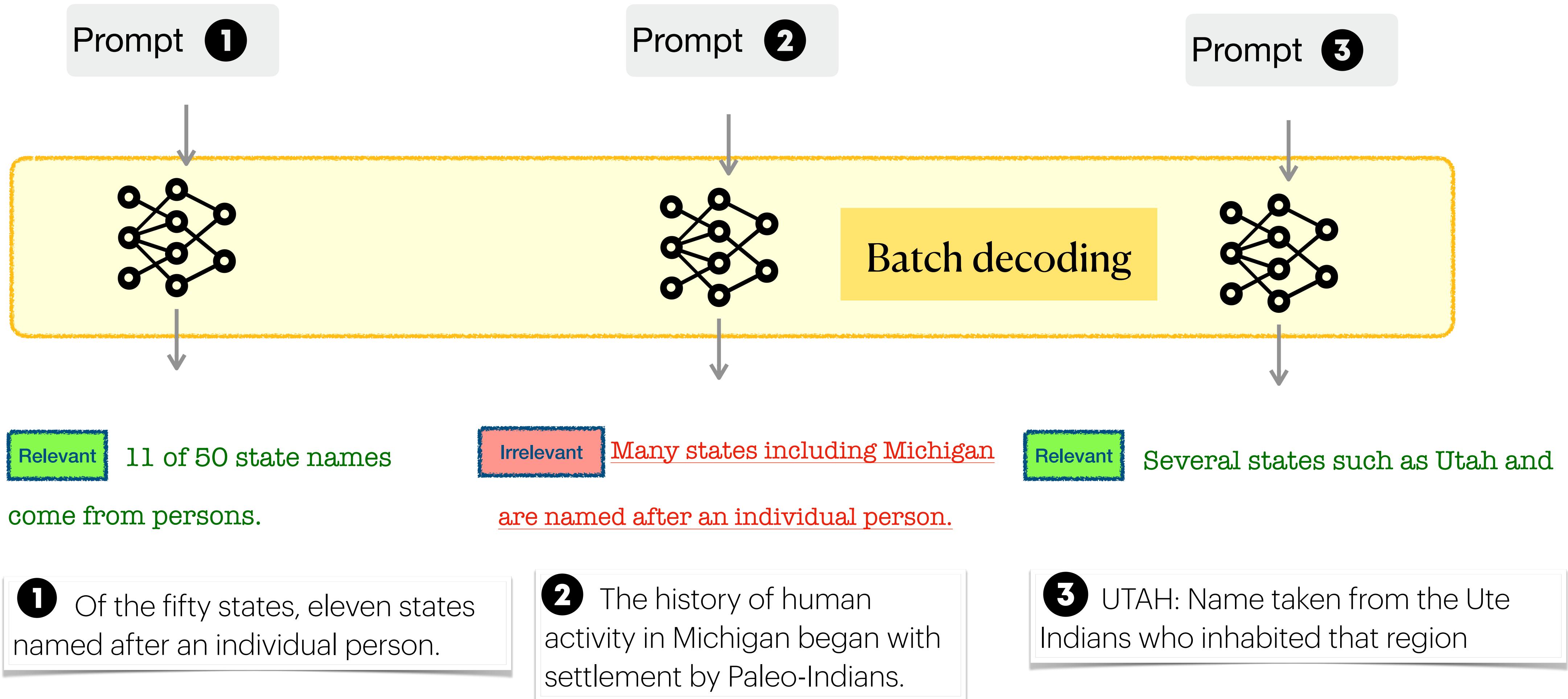
1 Of the fifty states, eleven states named after an individual person.

2 The history of human activity in Michigan began with settlement by Paleo-Indians.

3 UTAH: Name taken from the Ute Indians who inhabited that region

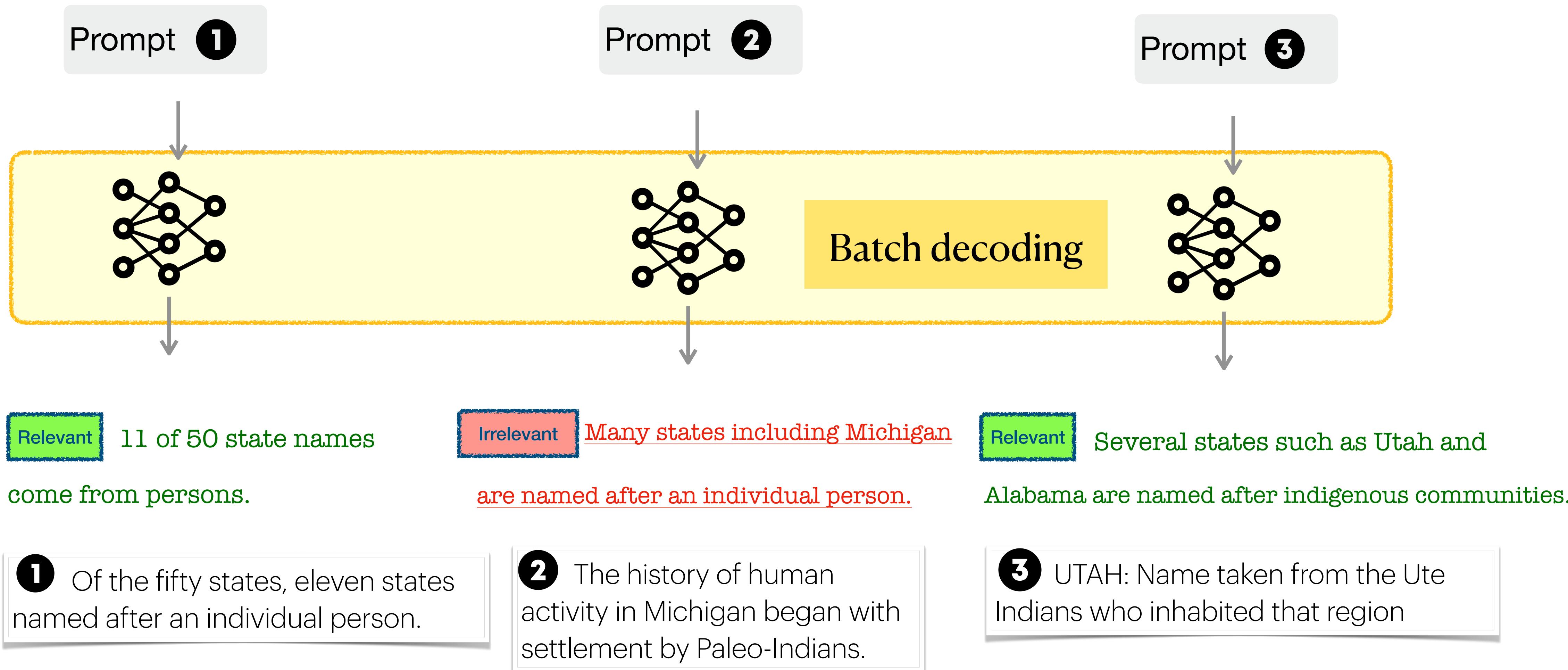
Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 2: Generate segments in *parallel*



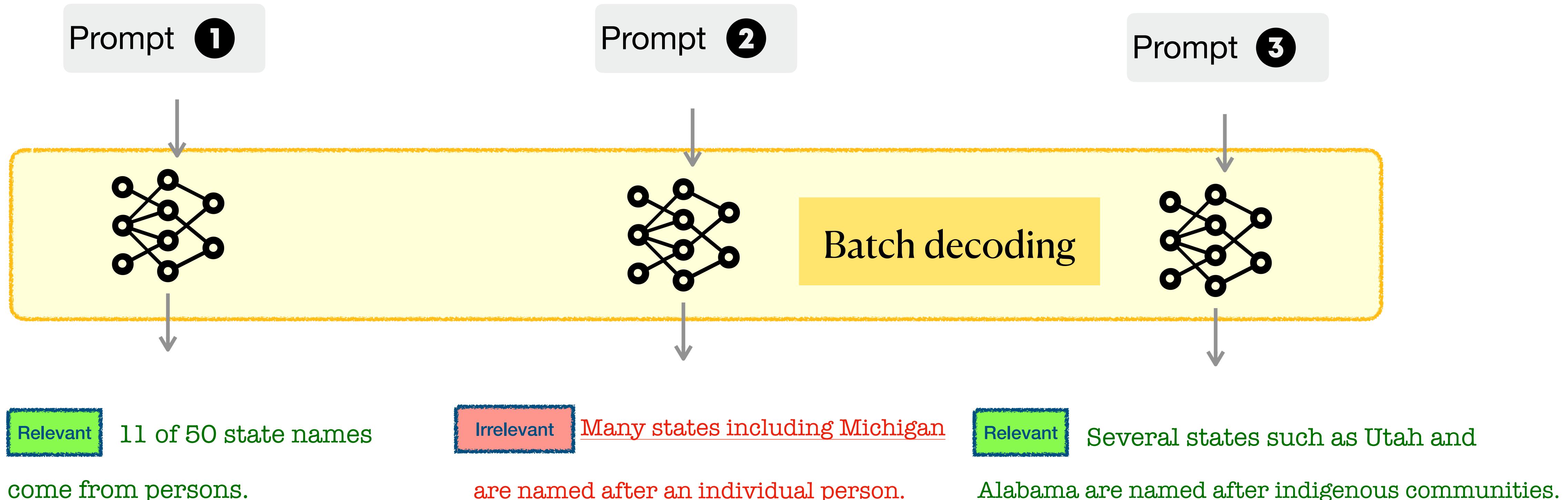
Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 2: Generate segments in *parallel*



Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 2: Generate segments in *parallel*

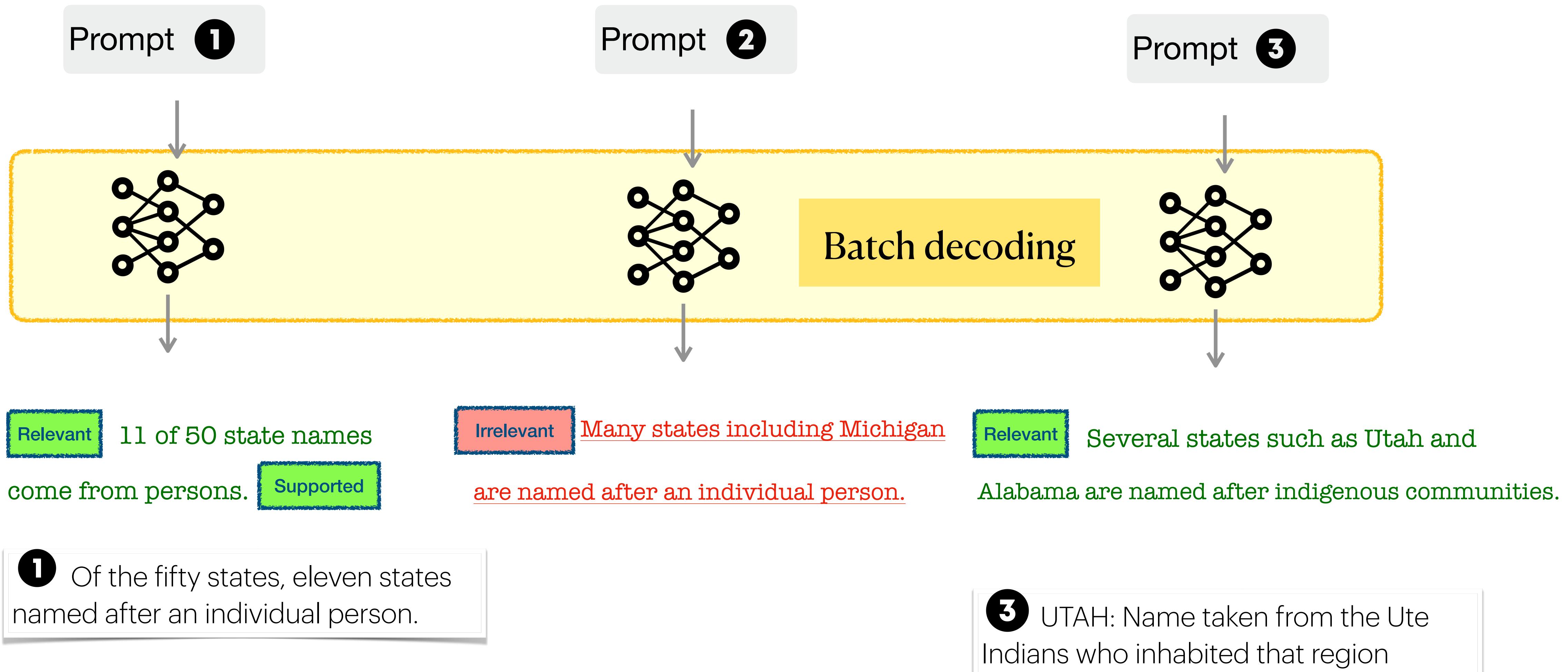


1 Of the fifty states, eleven states named after an individual person.

3 UTAH: Name taken from the Ute Indians who inhabited that region

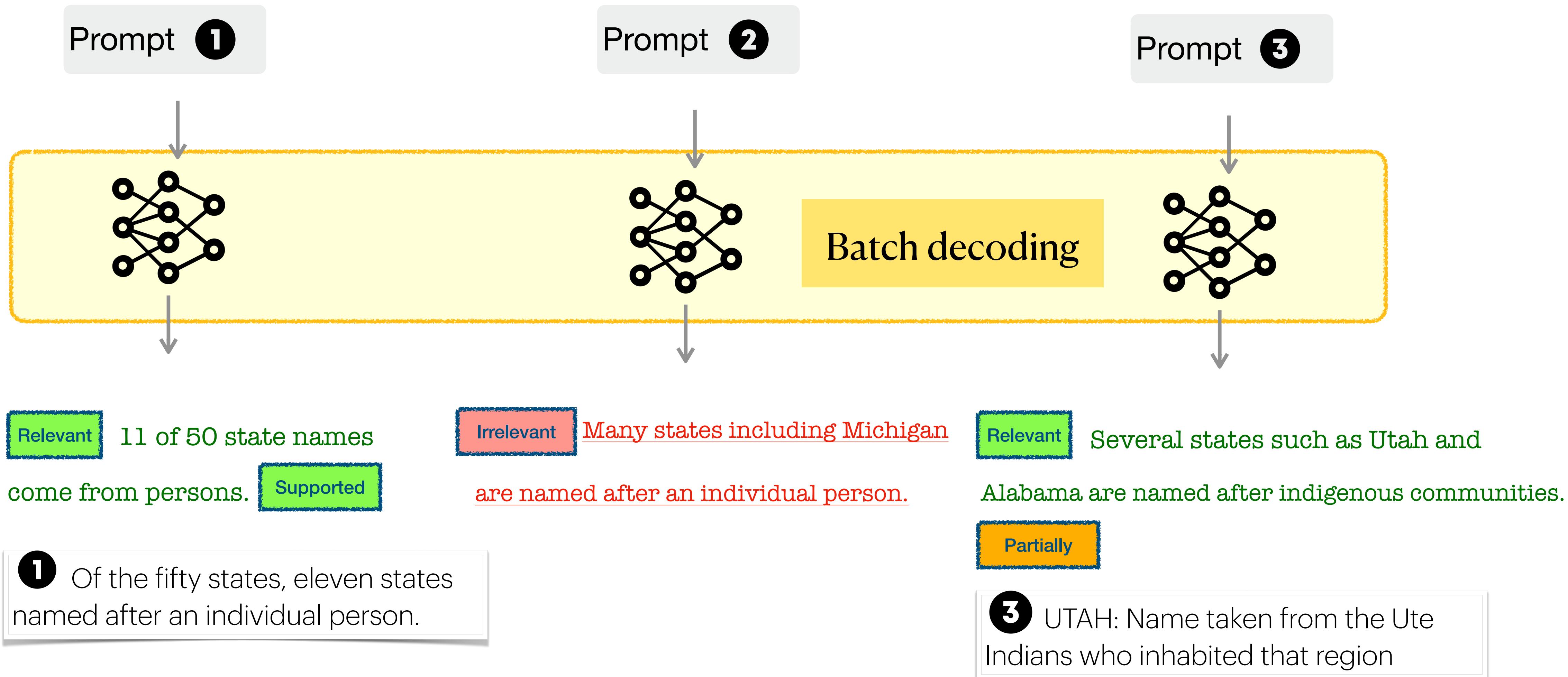
Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 2: Generate segments in *parallel*



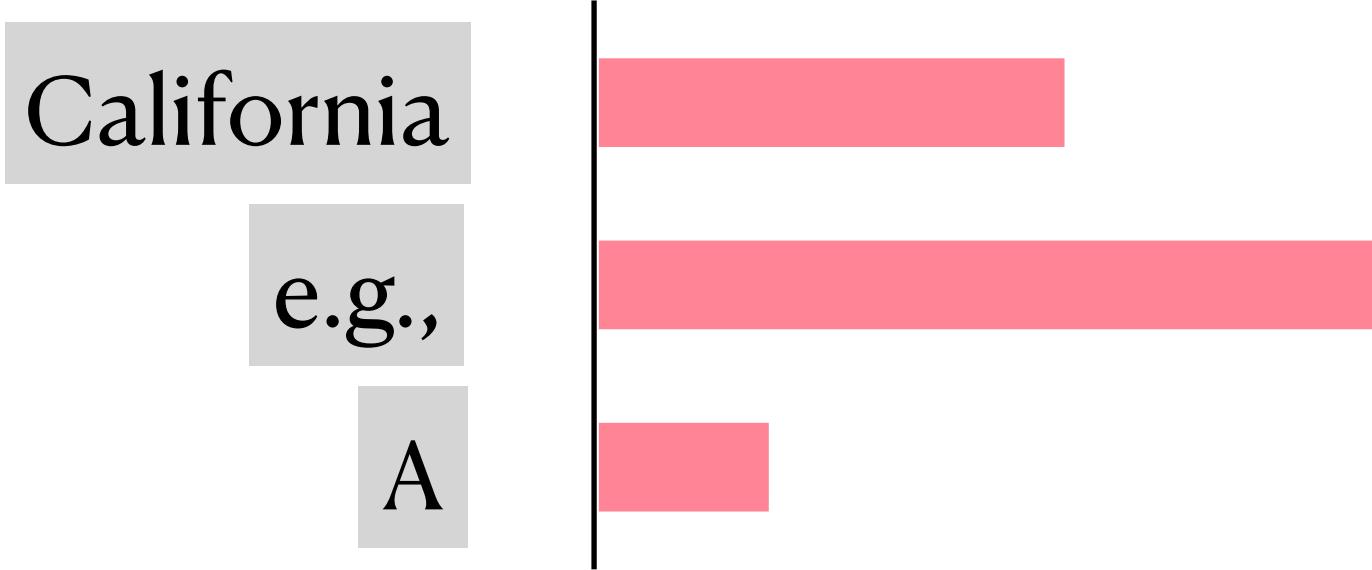
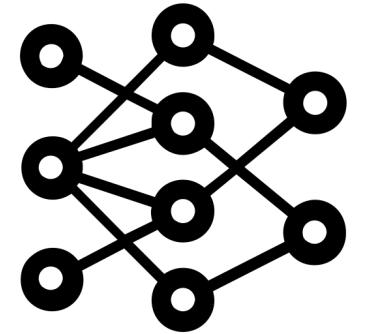
Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 2: Generate segments in *parallel*



Reflection tokens for retrieval and critique

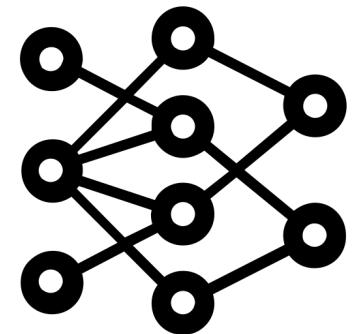
US states got their names from a variety of sources.



Original LM vocabularies

Reflection tokens for retrieval and critique

US states got their names from a variety of sources.



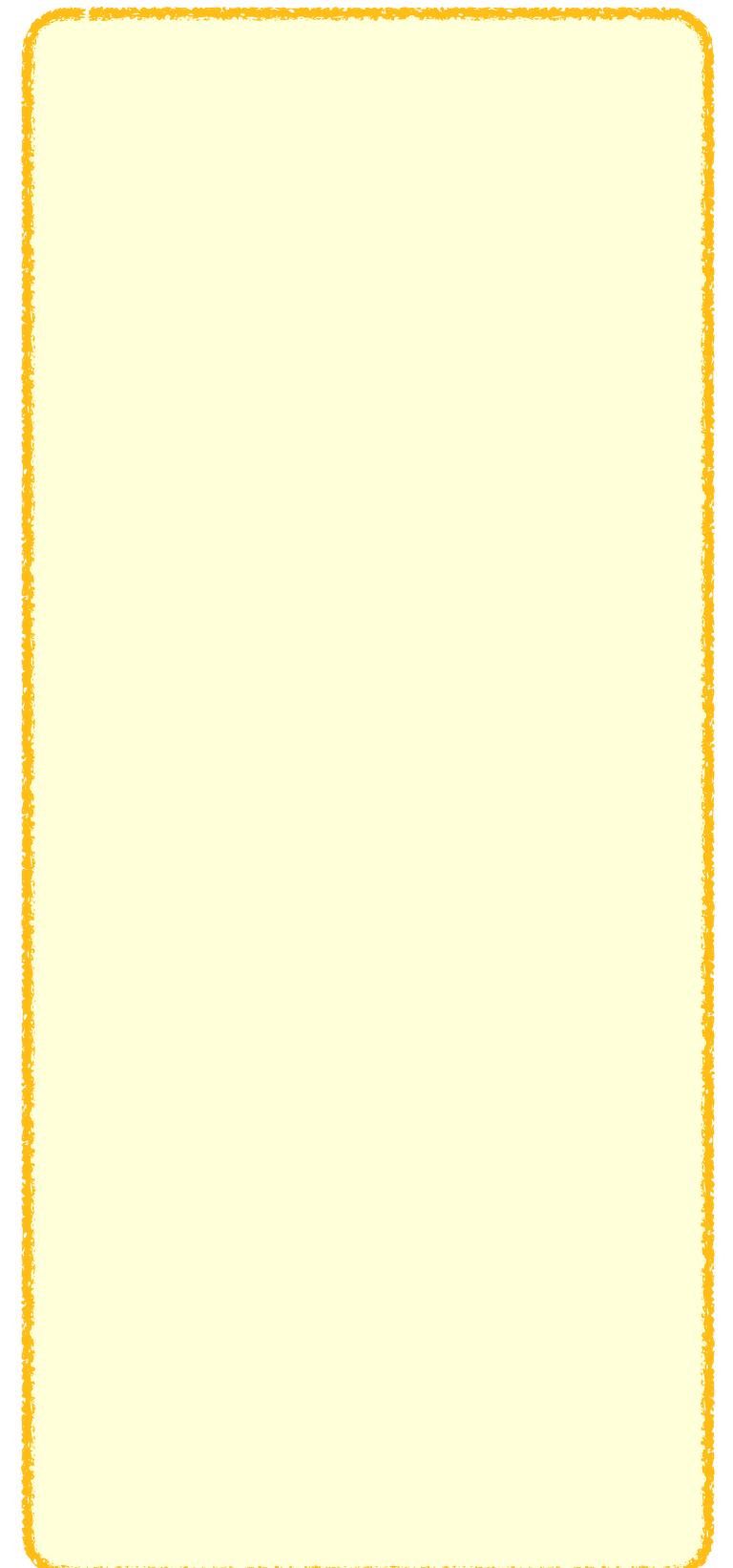
Vocabulary expanded with reflection tokens

California

e.g.,

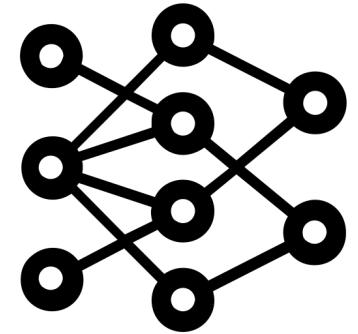
A

Original LM vocabularies



Reflection tokens for retrieval and critique

US states got their names from a variety of sources.



Vocabulary expanded with reflection tokens

California

e.g.,

A

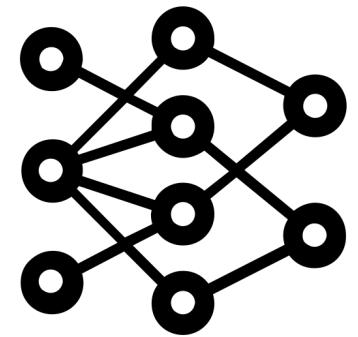
Original LM vocabularies

Retrieve

No Retrieval

Reflection tokens for retrieval and critique

US states got their names from a variety of sources.



California

e.g.,

A

Original LM vocabularies

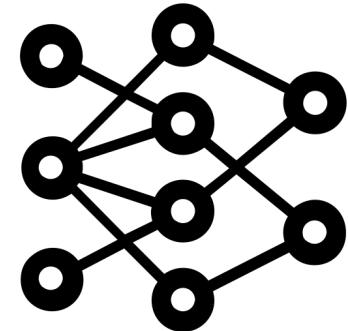
Retrieve
No Retrieval

Retrieval tokens

Vocabulary expanded with reflection tokens

Reflection tokens for retrieval and critique

US states got their names from a variety of sources.



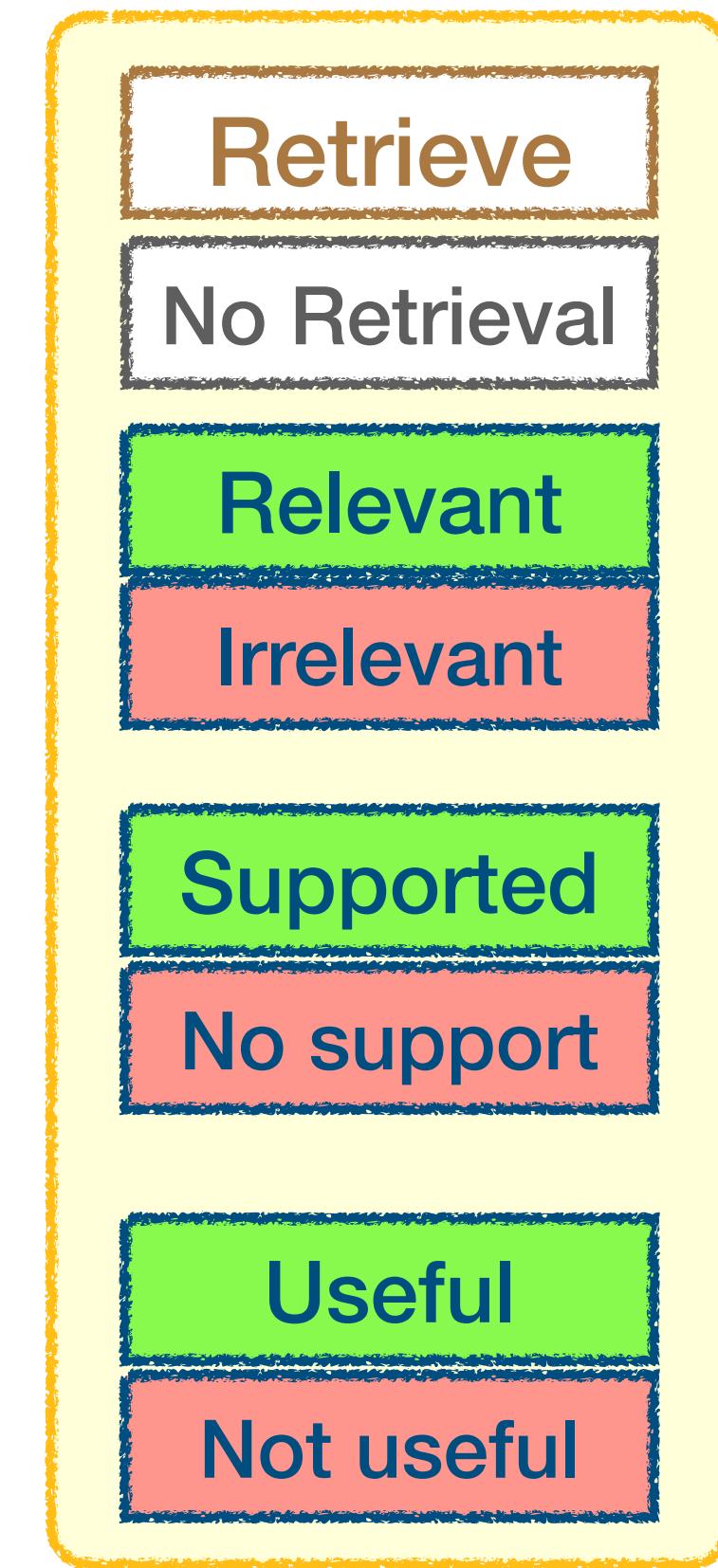
California

e.g.,

A

Original LM vocabularies

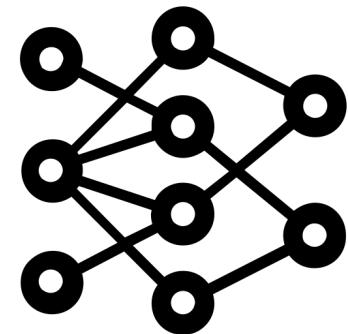
Vocabulary expanded with reflection tokens



Retrieval tokens

Reflection tokens for retrieval and critique

US states got their names from a variety of sources.



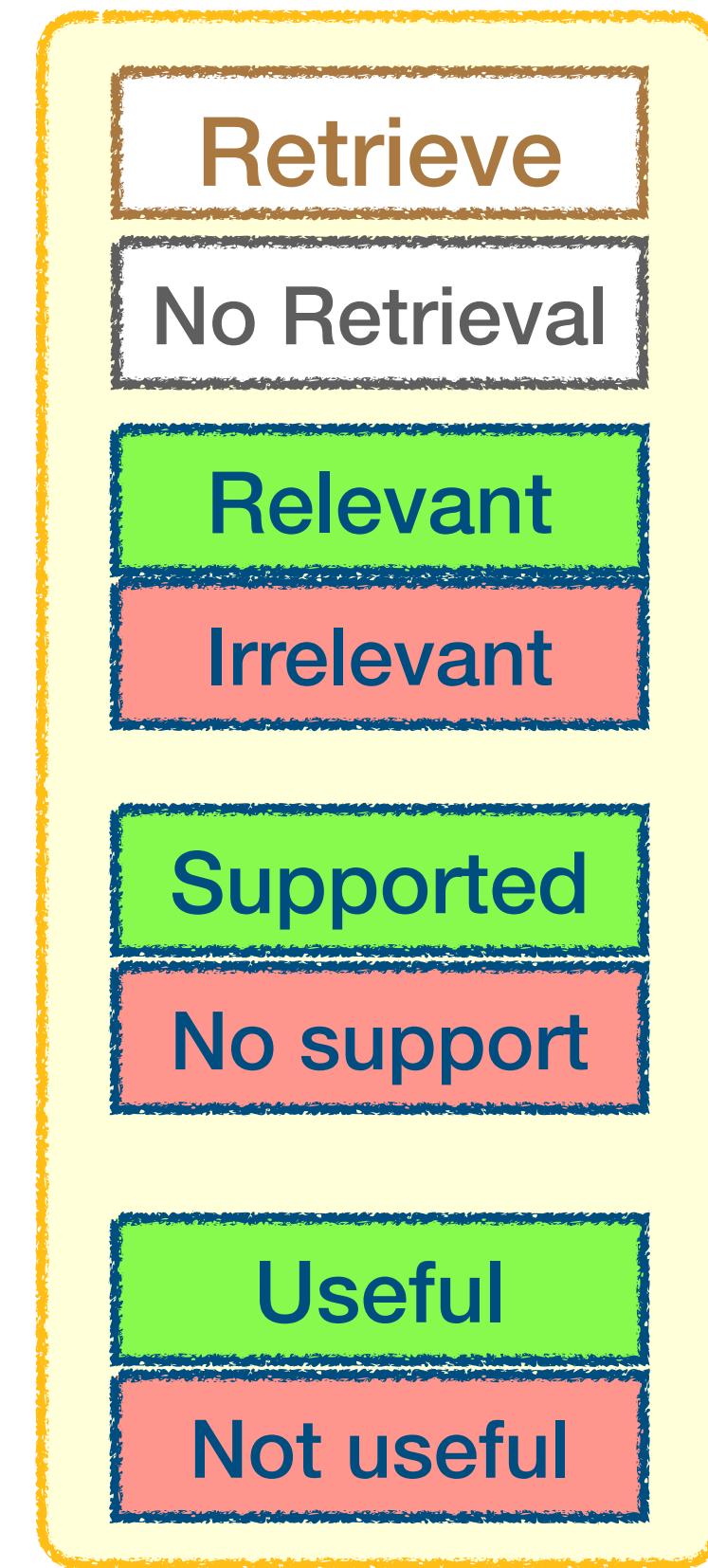
California

e.g.,

A

Original LM vocabularies

Vocabulary expanded with reflection tokens

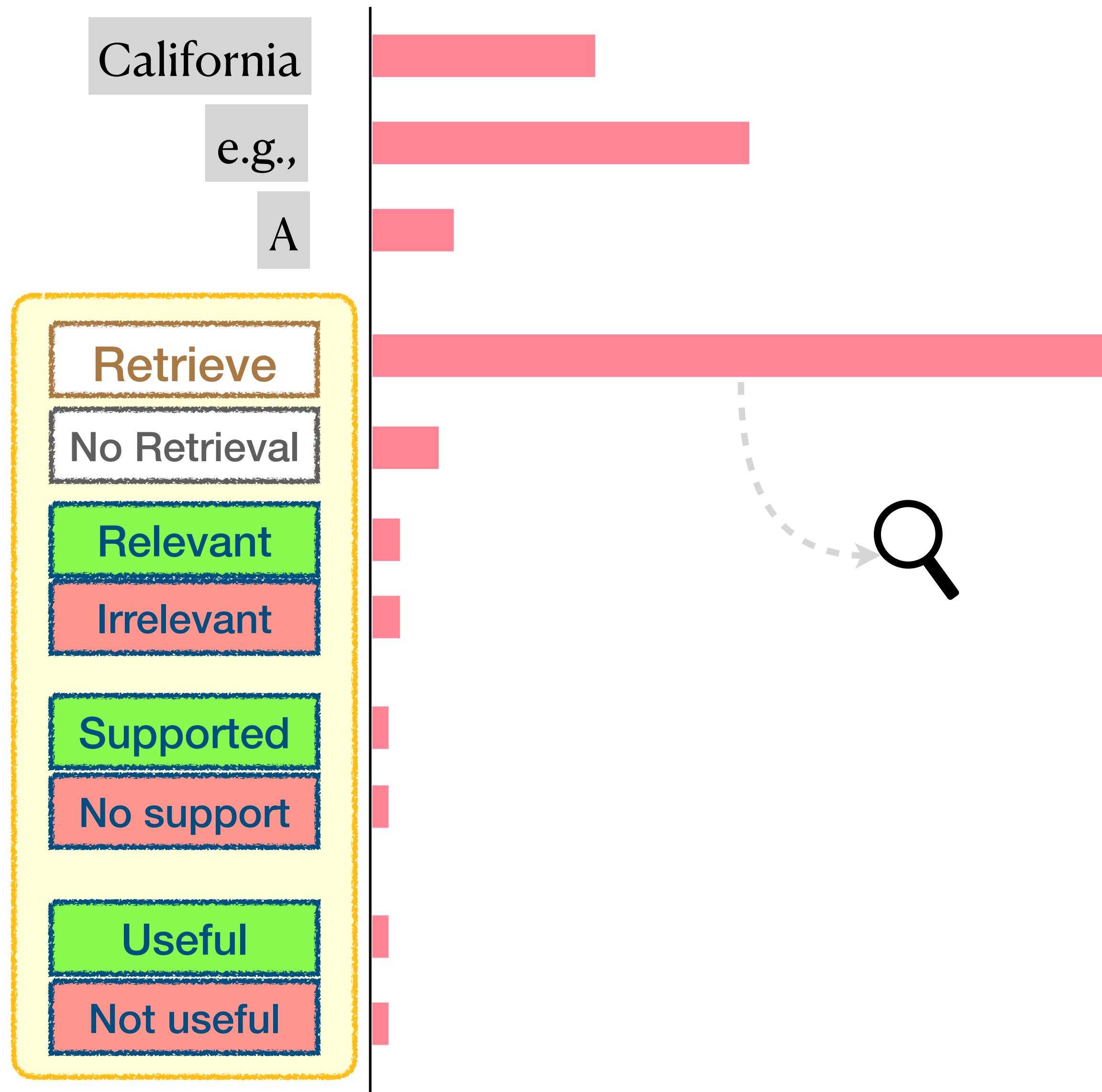
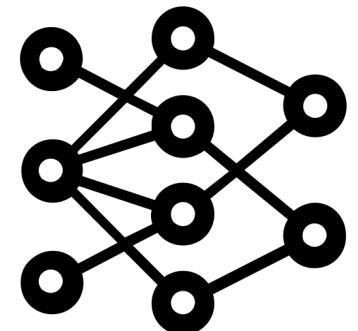


Retrieval tokens

Critique tokens

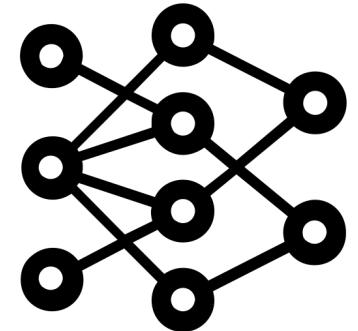
Reflection tokens for retrieval and critique

US states got their names from a variety of sources.

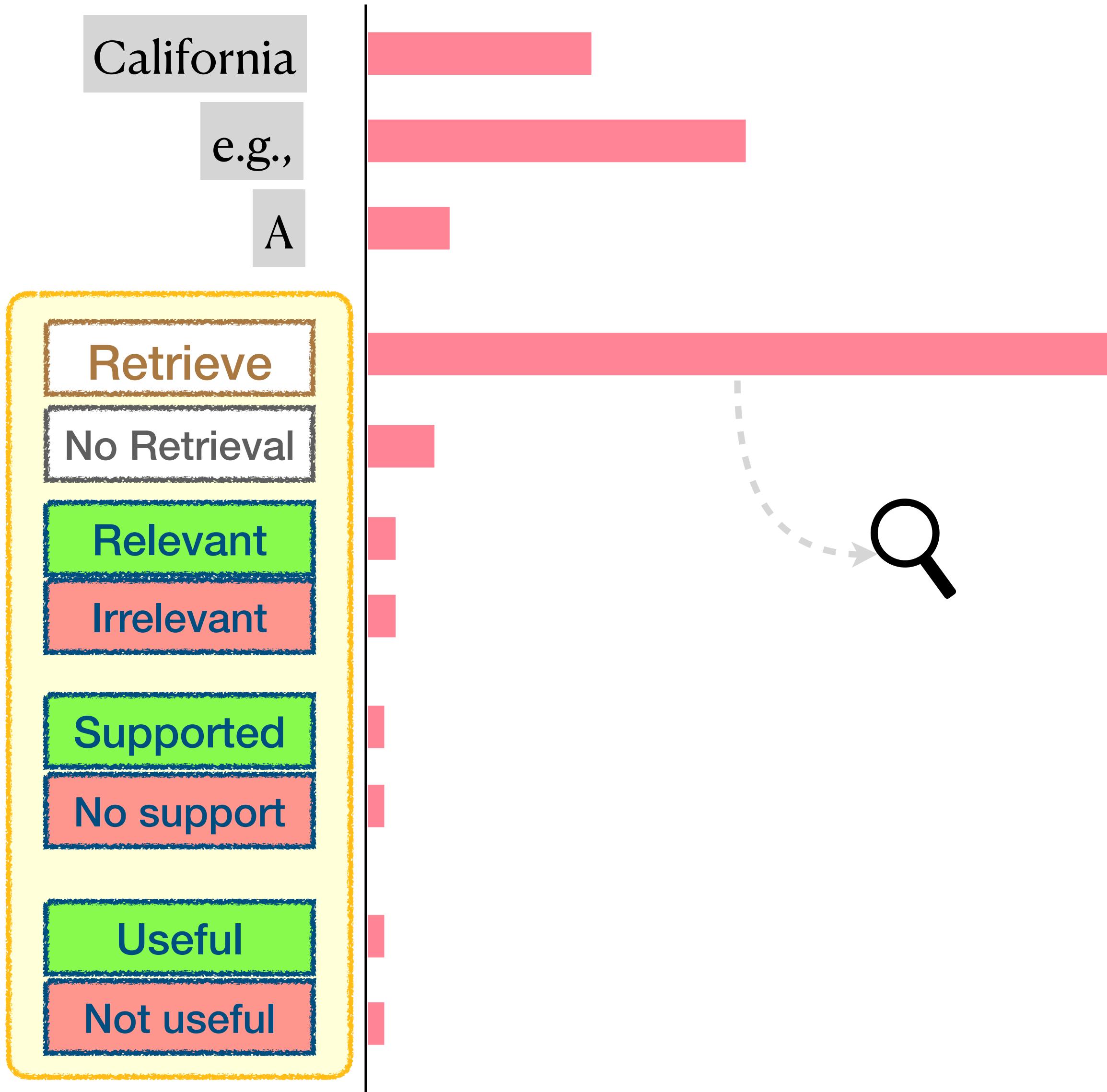


Reflection tokens for retrieval and critique

US states got their names from a variety of sources.

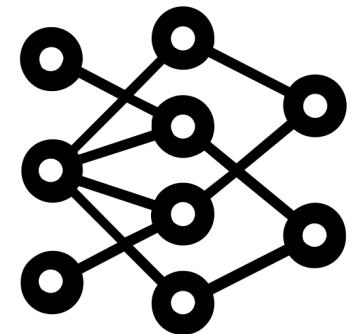


Adaptive retrieval via retrieval reflection tokens



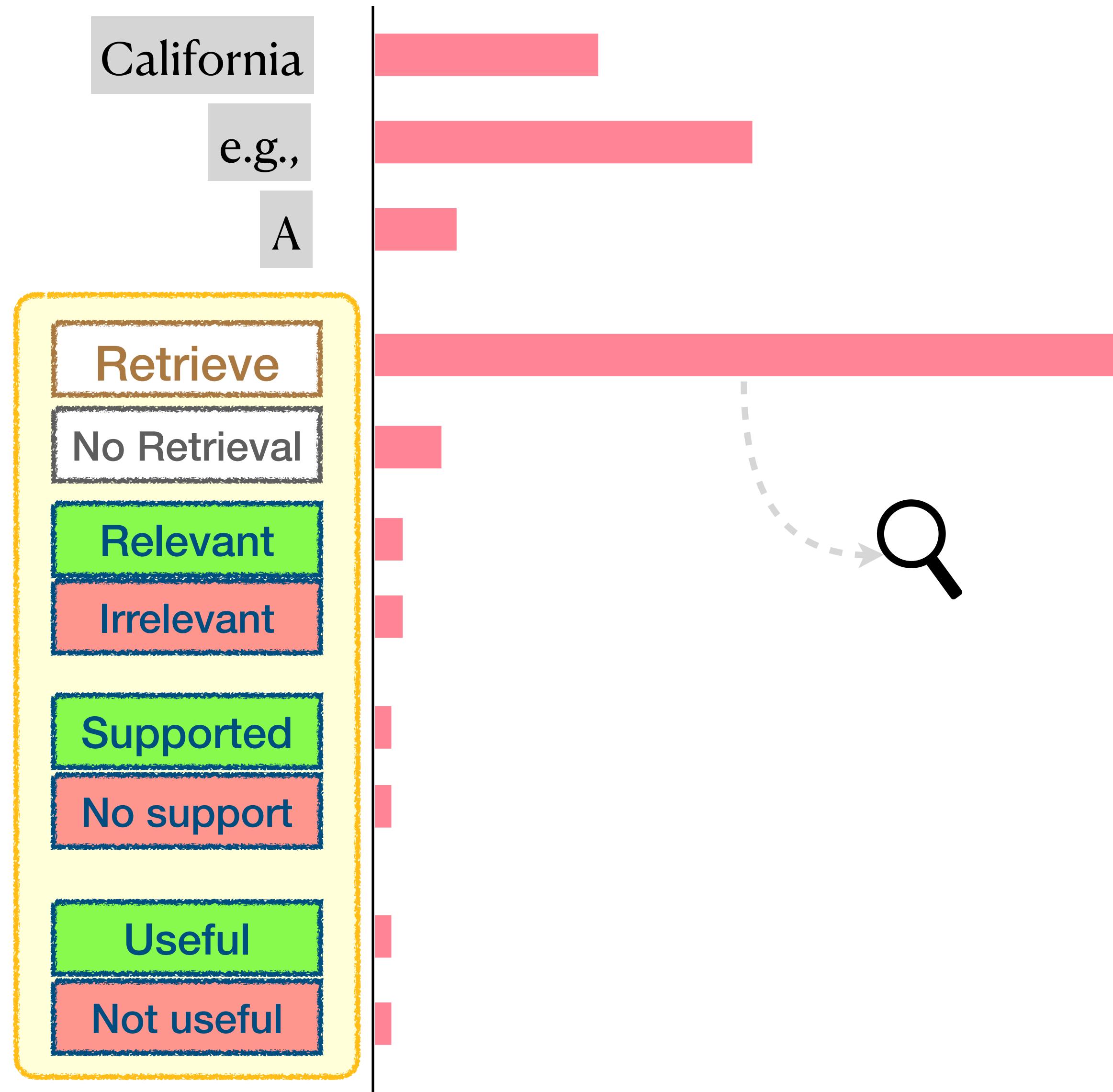
Reflection tokens for retrieval and critique

US states got their names from a variety of sources.



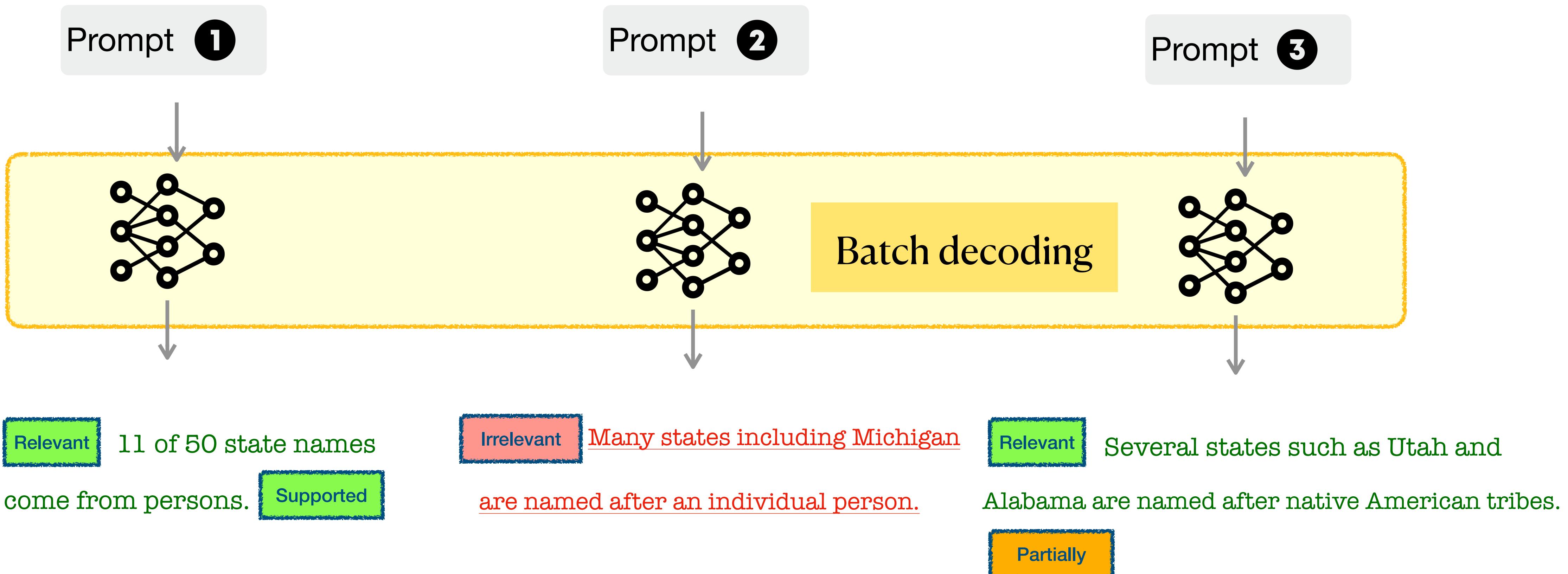
$$\frac{p(\text{Retrieve} = \text{YES})}{p(\text{Retrieve} = \text{YES}) + p(\text{Retrieve} = \text{No})} > \delta$$

Adaptive retrieval via retrieval reflection tokens



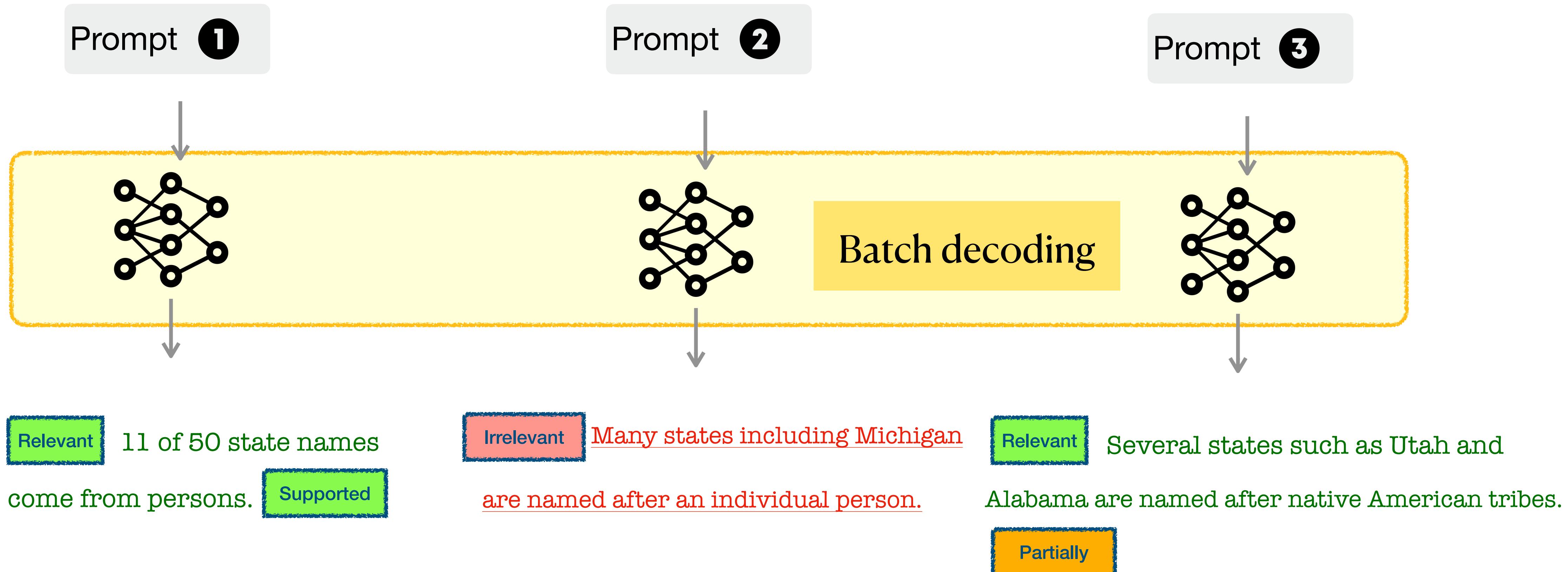
Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 3: Evaluate and choose best k outputs using reflection tokens



Self-RAG – Self-Reflective Retrieval-Augmented Generation

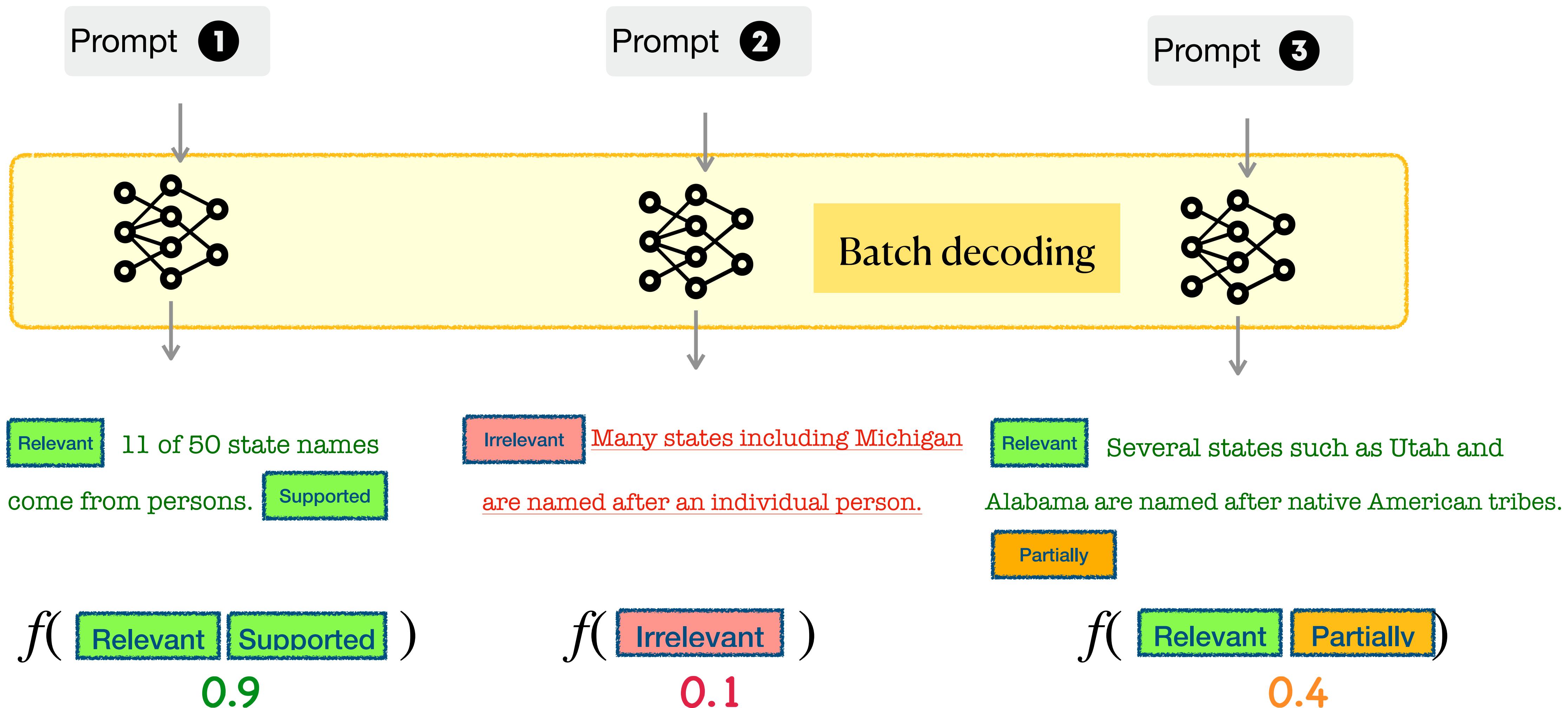
Step 3: Evaluate and choose best k outputs using reflection tokens



Critique token probabilities used as fine-grained feedback scores on each output

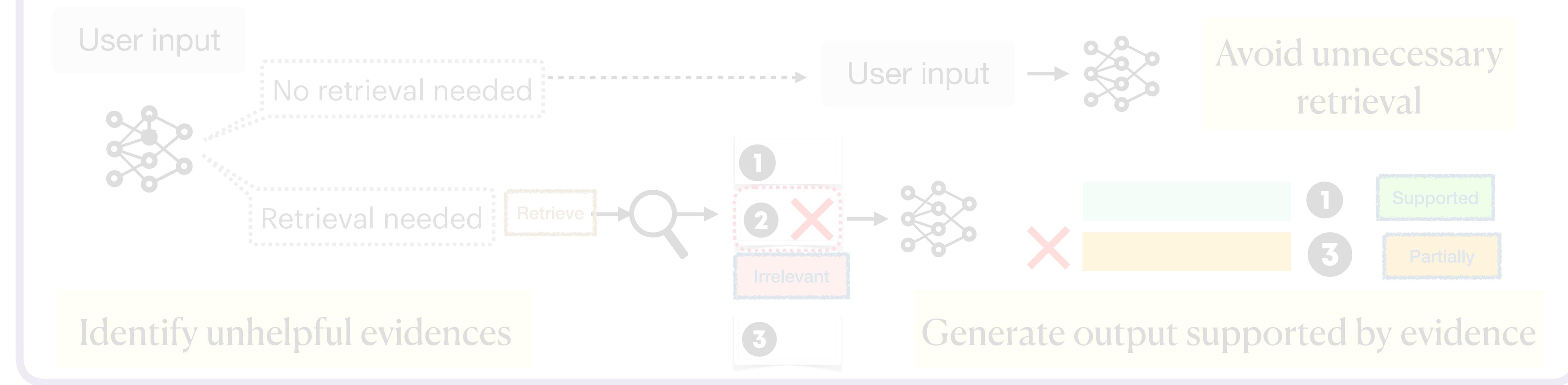
Self-RAG – Self-Reflective Retrieval-Augmented Generation

Step 3: Evaluate and choose best k outputs using reflection tokens



Self-RAG: *Self*-reflective Retrieval-Augmented Generation

Self-RAG Inference: Learn to retrieve and evaluate via special tokens



Self-RAG training: Instruction-tuning an arbitrary LM with reflection tokens

Input: Write an essay of your best summer vacation

Augmented Output: No Retrieval My best summer vacation was a magical escape to the coastal town of Santorini. No Retrieval The azure waters, charming white-washed building are unforgettable experience. Util: 5

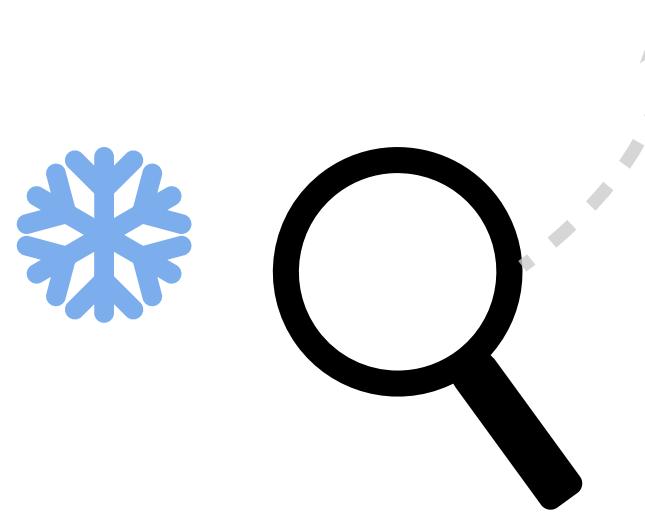
Input: How did US states get their names?

Augmented Output: Retrieve 1 <p>Of the fifty states, eleven are named after an individual person</p>. Relevant 11 of 50 states' names come from person. Supported Retrieve 2 <p>LOUISIANA: Named in honor of Louis XIV of France.</p>. Relevant For instance, Louisiana was named after King Louis XIV, and Georgia was named after King George II. Partially Util: 5

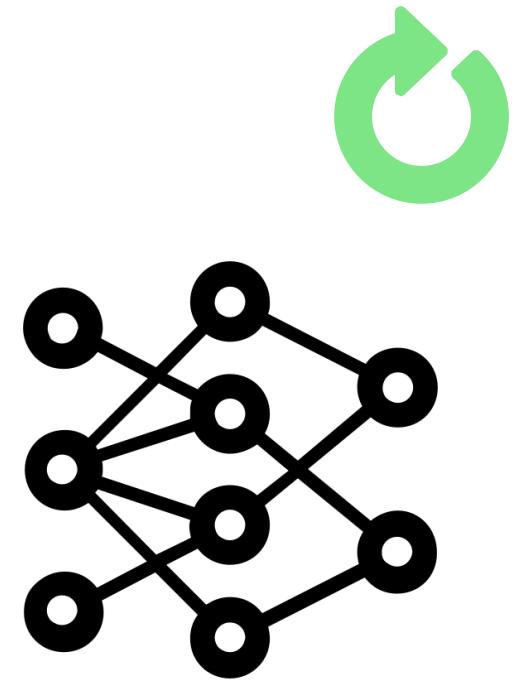
Overview of Self-RAG training

How did US states get their names?

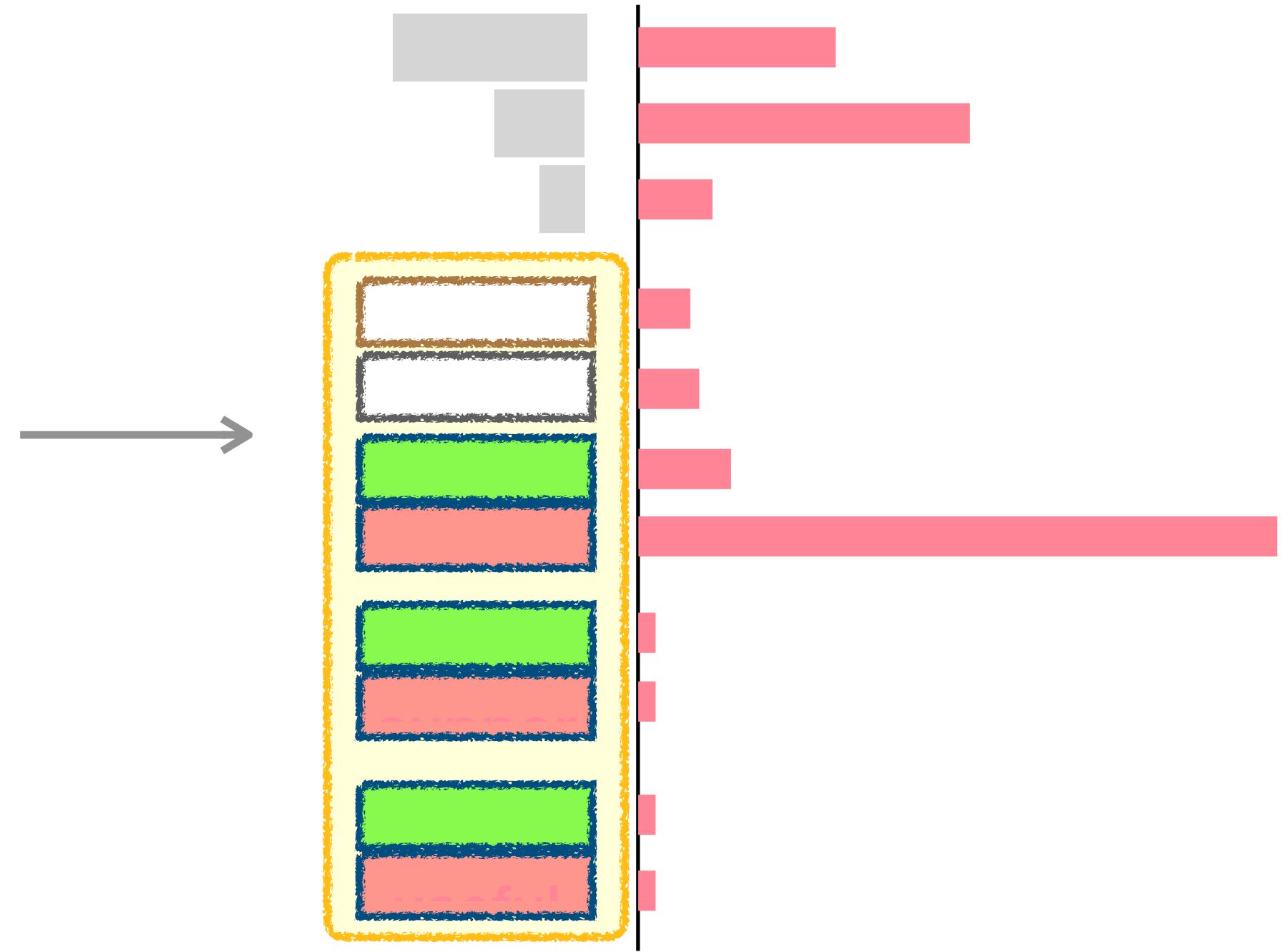
- 1 Of the fifty states, eleven are named after an individual person.



Retriever



Generator LM

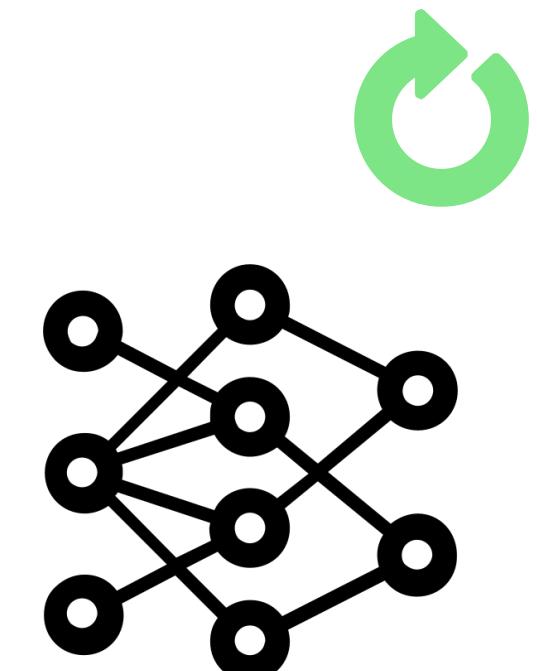


Overview of Self-RAG training

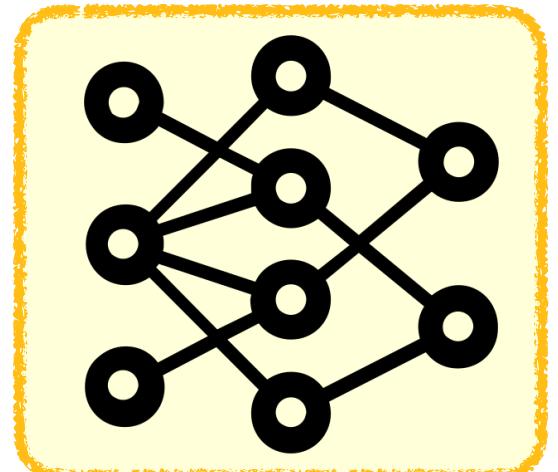
How did US states get their names?

- 1 Of the fifty states, eleven are named after an individual person.

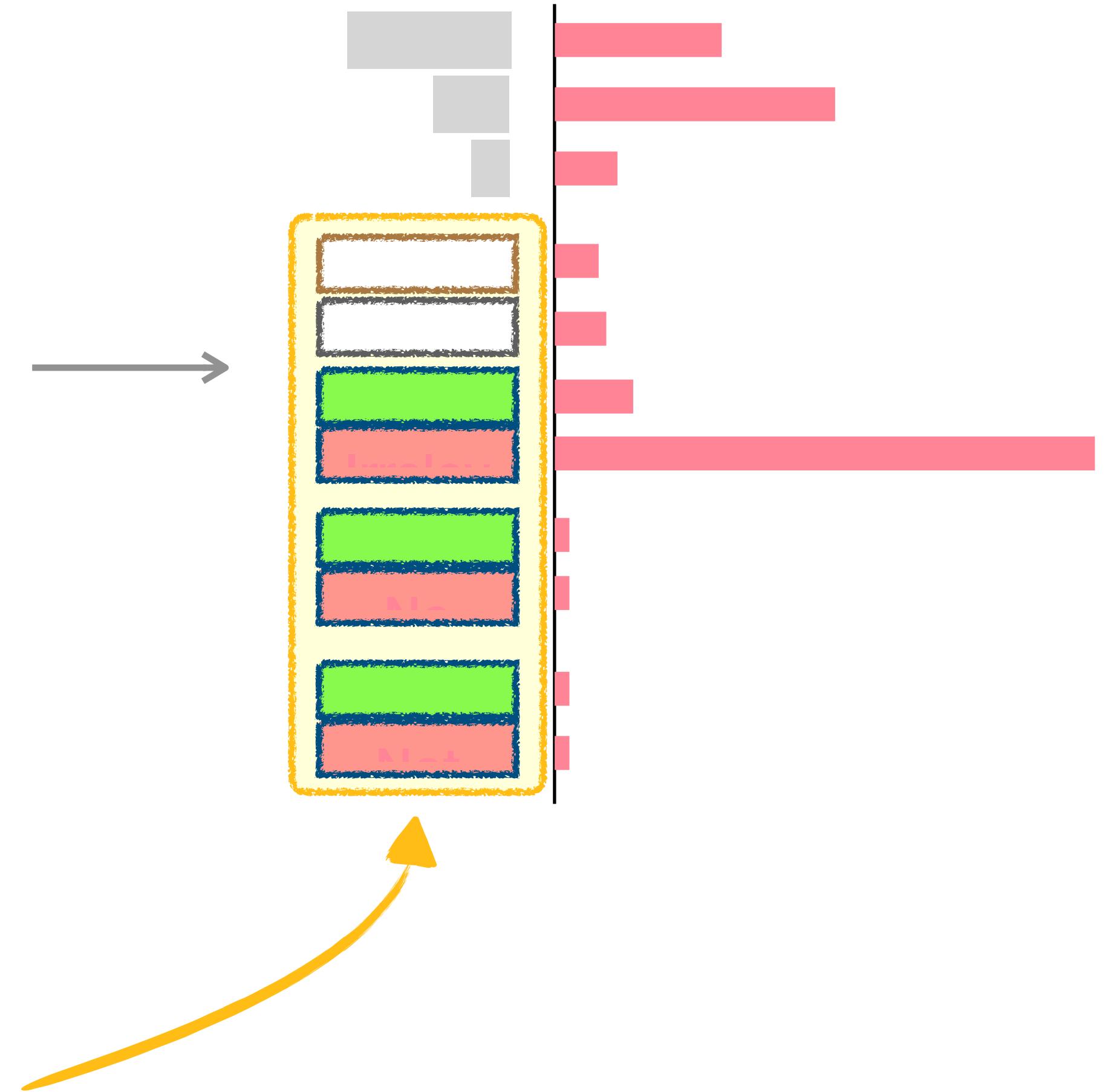
Critic LM teaches Generator LM to predict reflection tokens



Generator LM



Training time only **Critic LM**



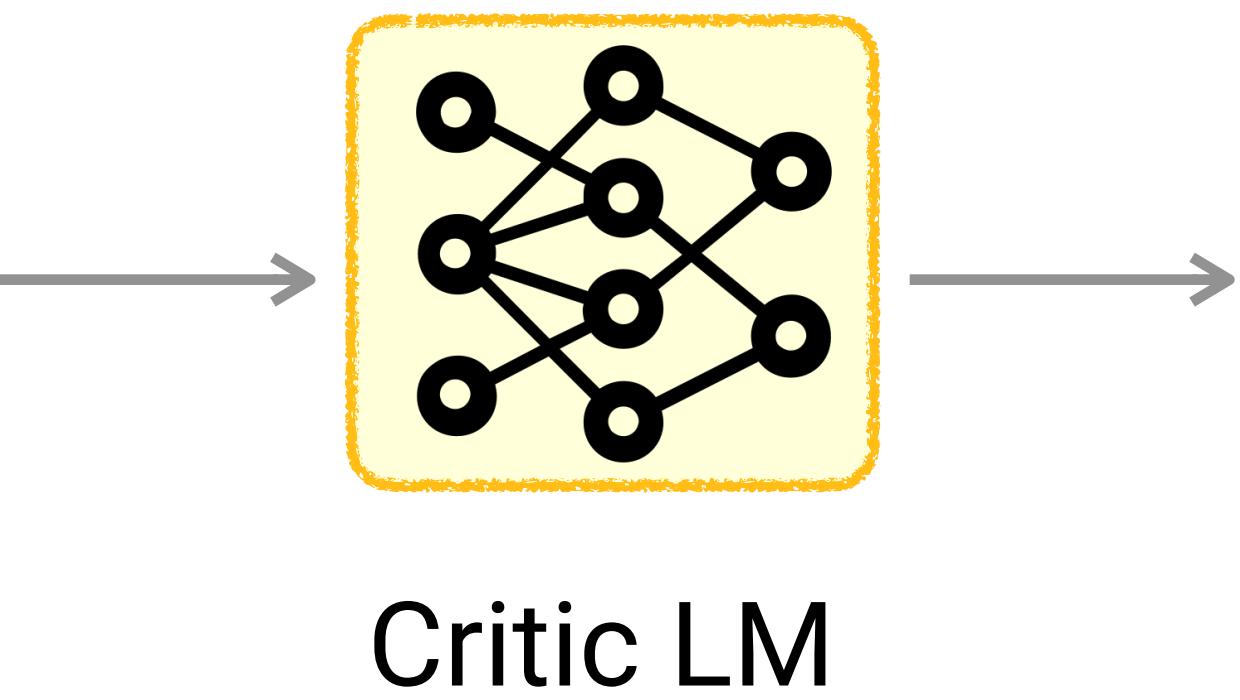
Self-RAG training - Critic LM

Evaluate if the output y to an input x is supported by retrieved doc d :

x How did US states get their names?

d Of the fifty states, eleven are named after an individual person.

y 11 of 50 state names come from persons.



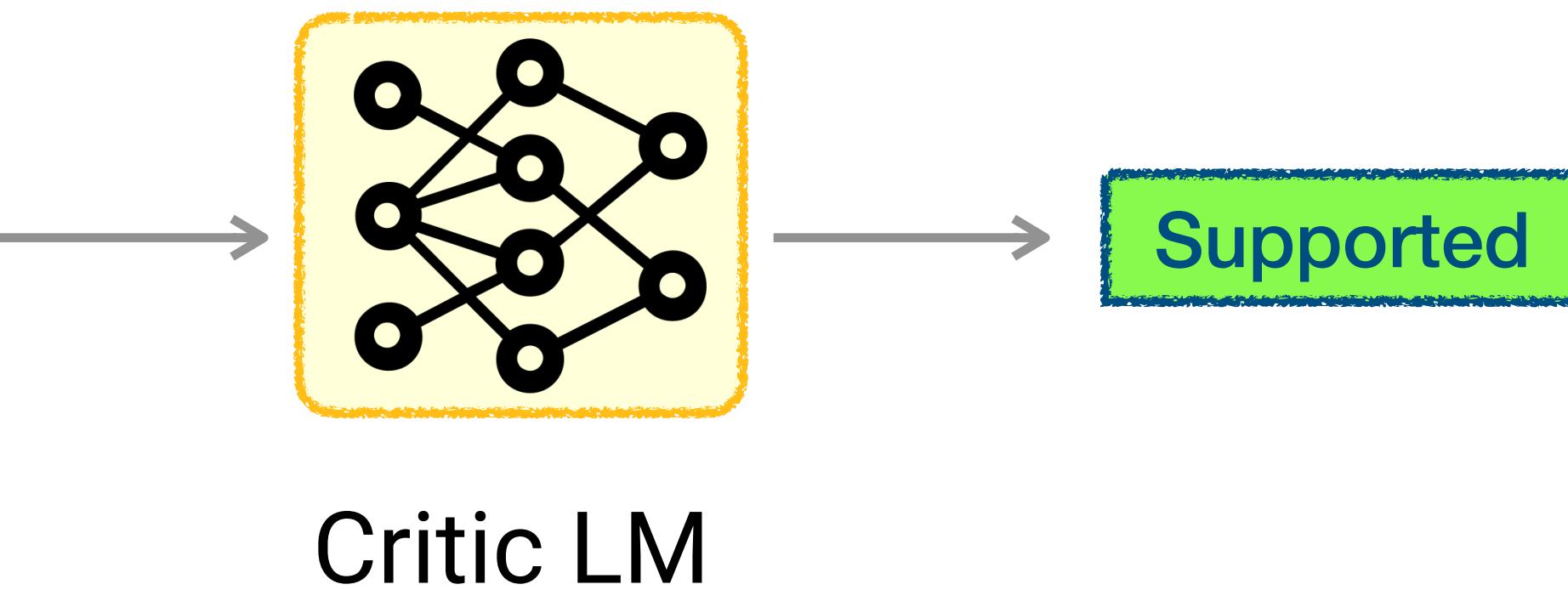
Self-RAG training - Critic LM

Evaluate if the output y to an input x is supported by retrieved doc d :

x How did US states get their names?

d Of the fifty states, eleven are named after an individual person.

y 11 of 50 state names come from persons.



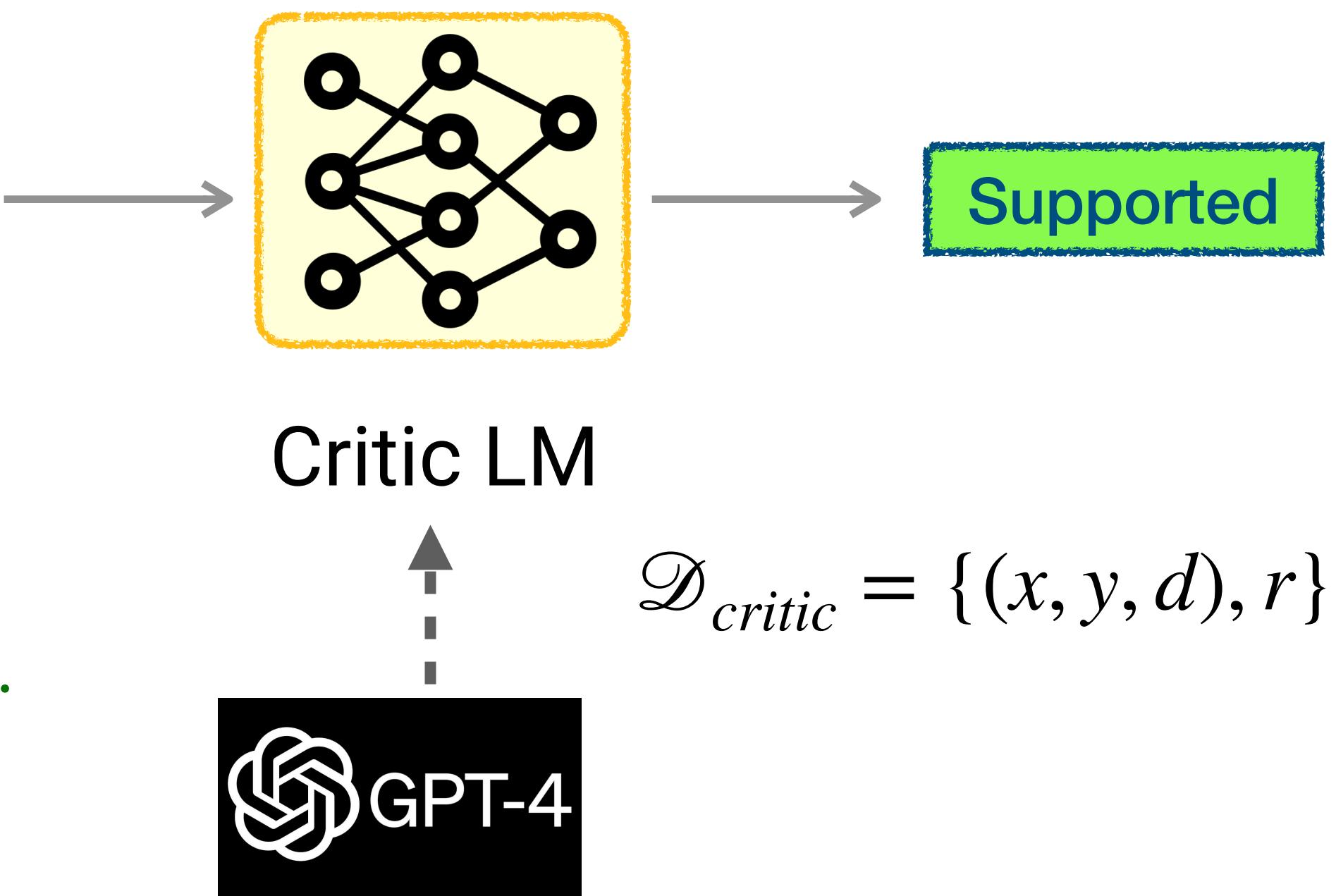
Self-RAG training - Critic LM

Evaluate if the output y to an input x is supported by retrieved doc d :

x How did US states get their names?

d Of the fifty states, eleven are named after an individual person.

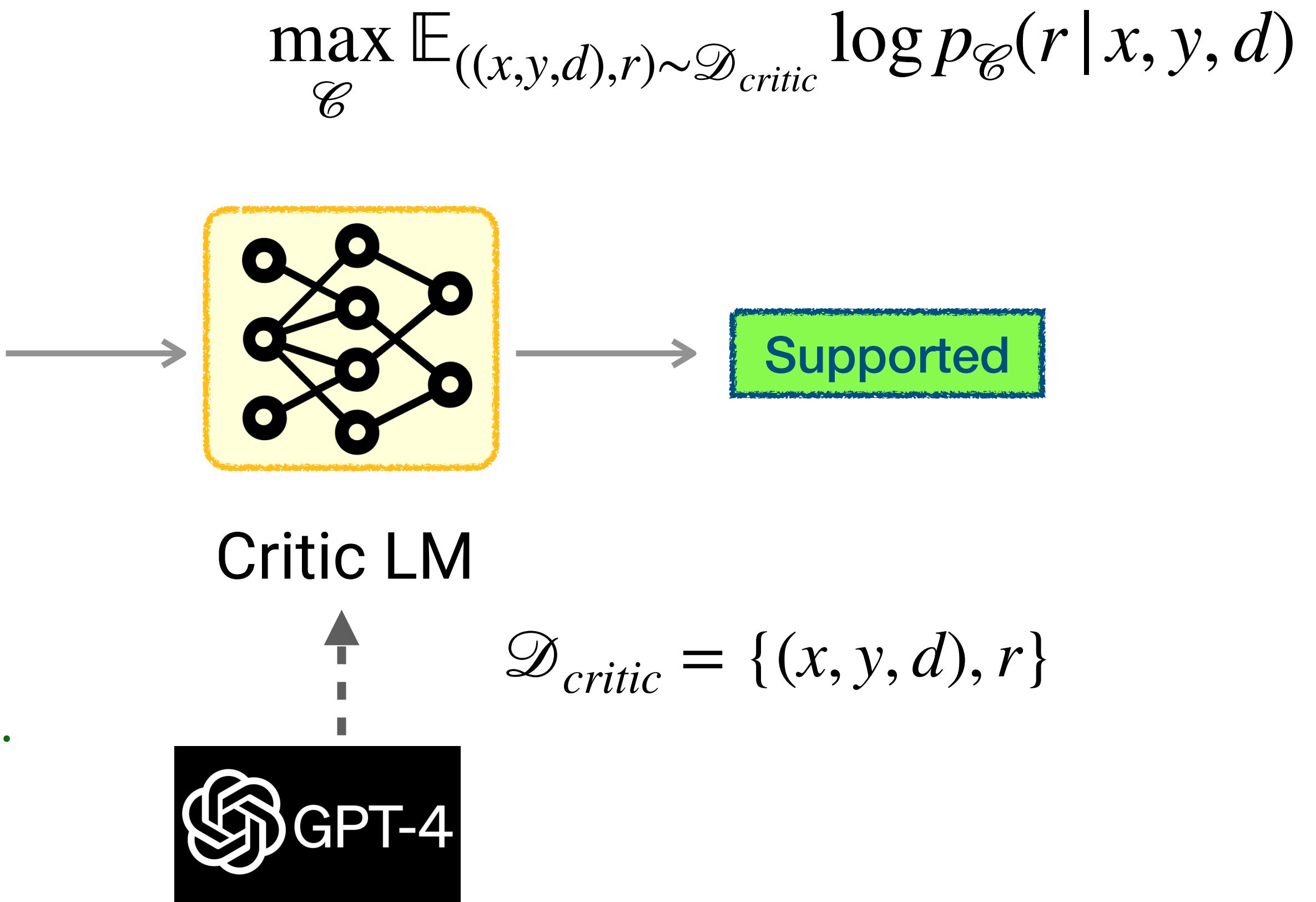
y 11 of 50 state names come from persons.



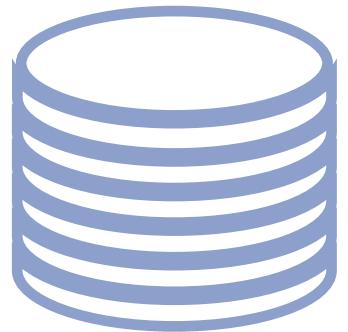
Self-RAG training - Critic LM

Evaluate if the output y to an input x is supported by retrieved doc d :

- x How did US states get their names?
- d Of the fifty states, eleven are named after an individual person.
- y 11 of 50 state names come from persons.



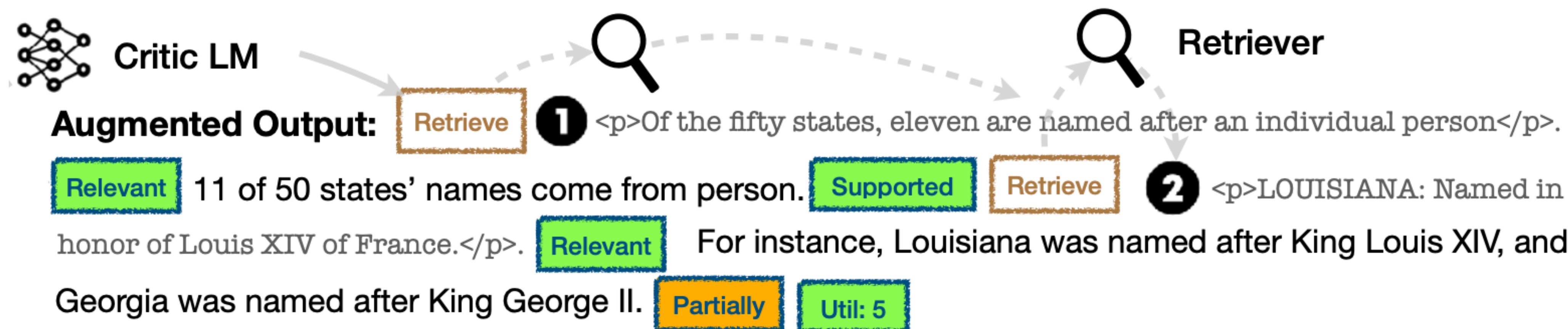
Self-RAG training - augmented instruction-tuning data



150k (input, output) instances from diverse instruction-following data

Input: How did US states get their names?

Output: 1 of 50 states names come from persons. For instance, Louisiana was named in honor of King Louis XIV of France and Georgia was named after King George II.



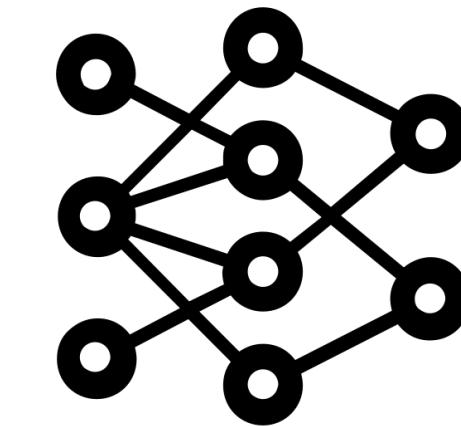
Self-RAG training - Generator LM

x

How did US states get their names?

d

Of the fifty states, eleven are named
after an individual person.



Generator LM

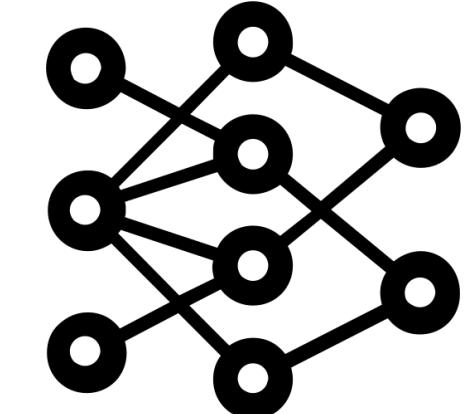
Self-RAG training - Generator LM

x

How did US states get their names?

d

Of the fifty states, eleven are named after an individual person.



Generator LM

y

r

11 of 50 state names come from persons.

Supported

Self-RAG training - Generator LM

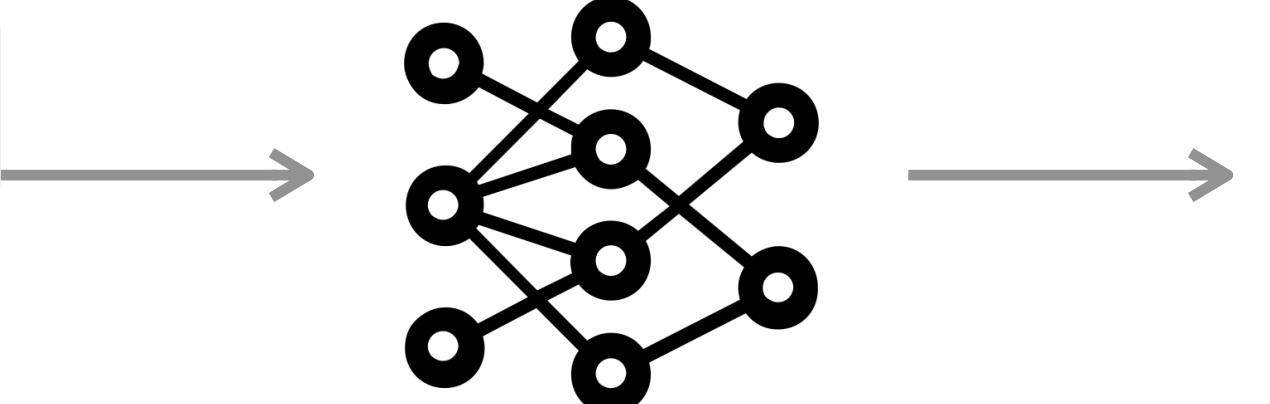
$$\max_{\mathcal{M}} \mathbb{E}_{(x,d,y,r) \sim \mathcal{D}_{gen}} \log p_{\mathcal{M}}(y, r | x, d)$$

x

How did US states get their names?

d

Of the fifty states, eleven are named after an individual person.



Generator LM

11 of 50 state names
come from persons.

Supported

Self-RAG training - Generator LM

$$\max_{\mathcal{M}} \mathbb{E}_{(x,d,y,r) \sim \mathcal{D}_{gen}} \log p_{\mathcal{M}}(y, r | x, d)$$

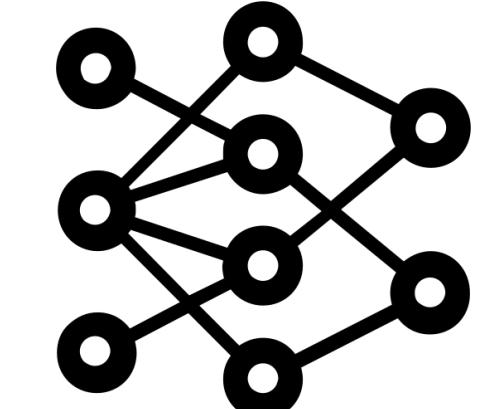
Train with a standard next token objective with expanded vocabulary

x

How did US states get their names?

d

Of the fifty states, eleven are named after an individual person.



y

r

11 of 50 state names come from persons.

Supported

Generator LM

Self-RAG training - Generator LM

$$\max_{\mathcal{M}} \mathbb{E}_{(x,d,y,r) \sim \mathcal{D}_{gen}} \log p_{\mathcal{M}}(y, r | x, d)$$

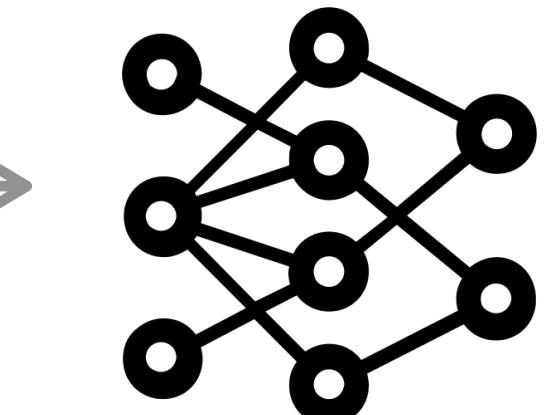
Train with a standard next token objective with expanded vocabulary

x

How did US states get their names?

d

Of the fifty states, eleven are named after an individual person.



Generator LM

y

r

11 of 50 state names come from persons.

Supported

Easily applied to new pre-trained LM

Experimental details

More details of training & test are in our paper!

Tasks and datasets

- **Closed-set tasks** (classifications, multiple-choice QA)
 - ARC-Challenge (Clark et al., 2018)
 - PubHealth (Zhang et al., 2023)

Experimental details

More details of training & test are in our paper!

Tasks and datasets

- **Closed-set tasks** (classifications, multiple-choice QA)
 - ARC-Challenge (Clark et al., 2018)
 - PubHealth (Zhang et al., 2023)
- **Short-form generation**
 - OpenQA - PopQA (Mallen* and Asai* et al., 2023)
 - Trivia QA (Joshi et al., 2017)

Experimental details

More details of training & test are in our paper!

Tasks and datasets

- **Closed-set tasks** (classifications, multiple-choice QA)
 - ARC-Challenge (Clark et al., 2018)
 - PubHealth (Zhang et al., 2023)
- **Short-form generation**
 - OpenQA - PopQA (Mallen* and Asai* et al., 2023)
 - Trivia QA (Joshi et al., 2017)
- **Long-form generation**
 - ASQA-ALCE [fluency, citation accuracy, correctness] (Gao et al., 2023)
 - Bio generations [FactScore] (Min et al., 2023)

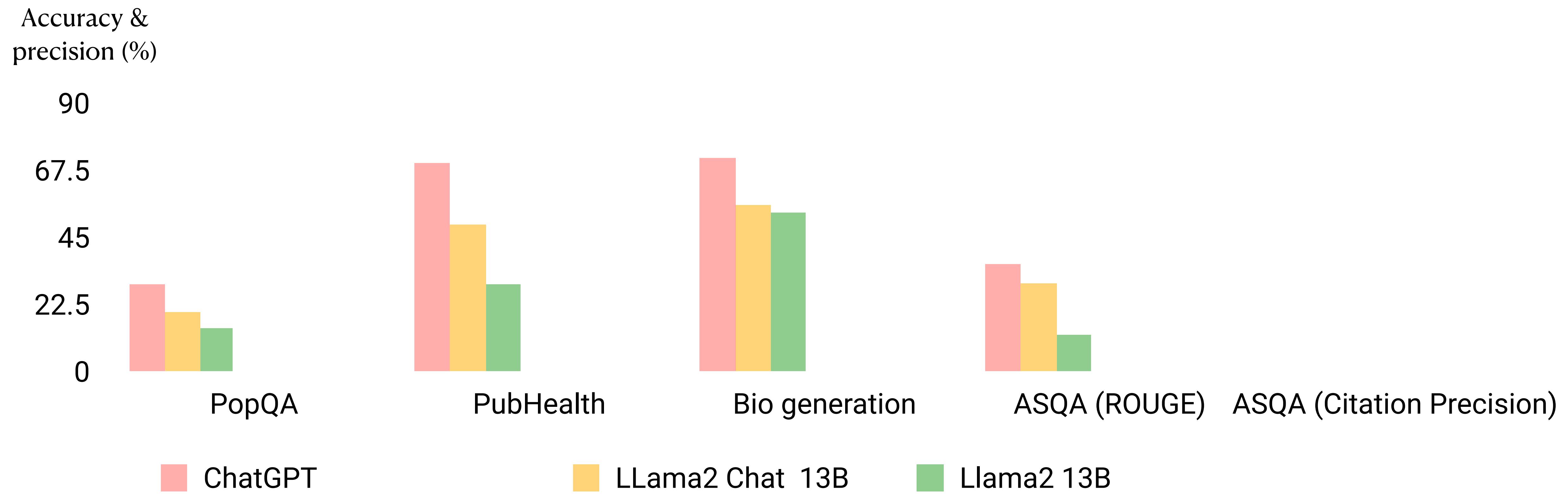
Experimental details

More details of training & test are in our paper!

Training details

- **Critic training data:** 4k-20k instances for each type
- **Generator training data:** 150k instruction-following datasets
 - ShareGPT
 - OpenAssistant
 - Alpaca
 - FLANV2
 - Natural Questions
- **Base LMs:** Llama2-7B, 13B (Touvron et al., 2023)
- **Computation:** 4*A100 (15 hours)

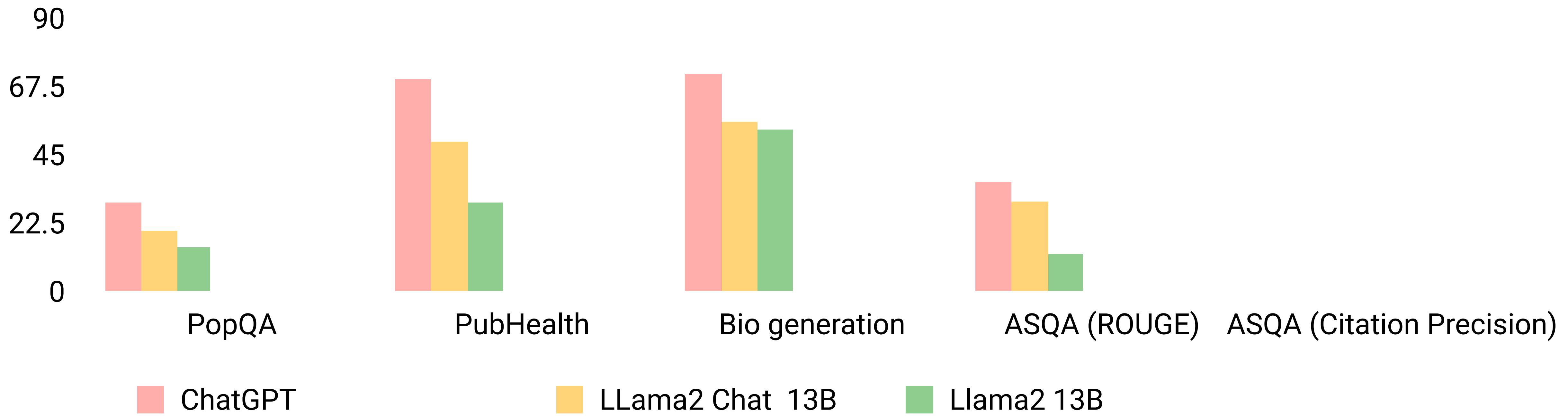
Experimental results



Experimental results

Parametric LMs struggle with tasks requiring precise world knowledge

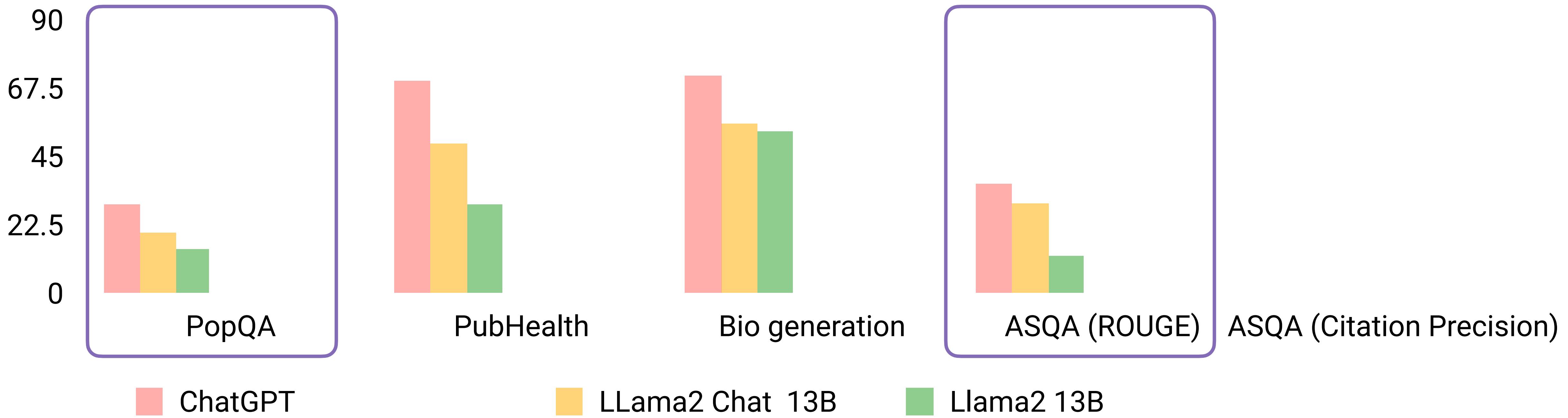
Accuracy &
precision (%)



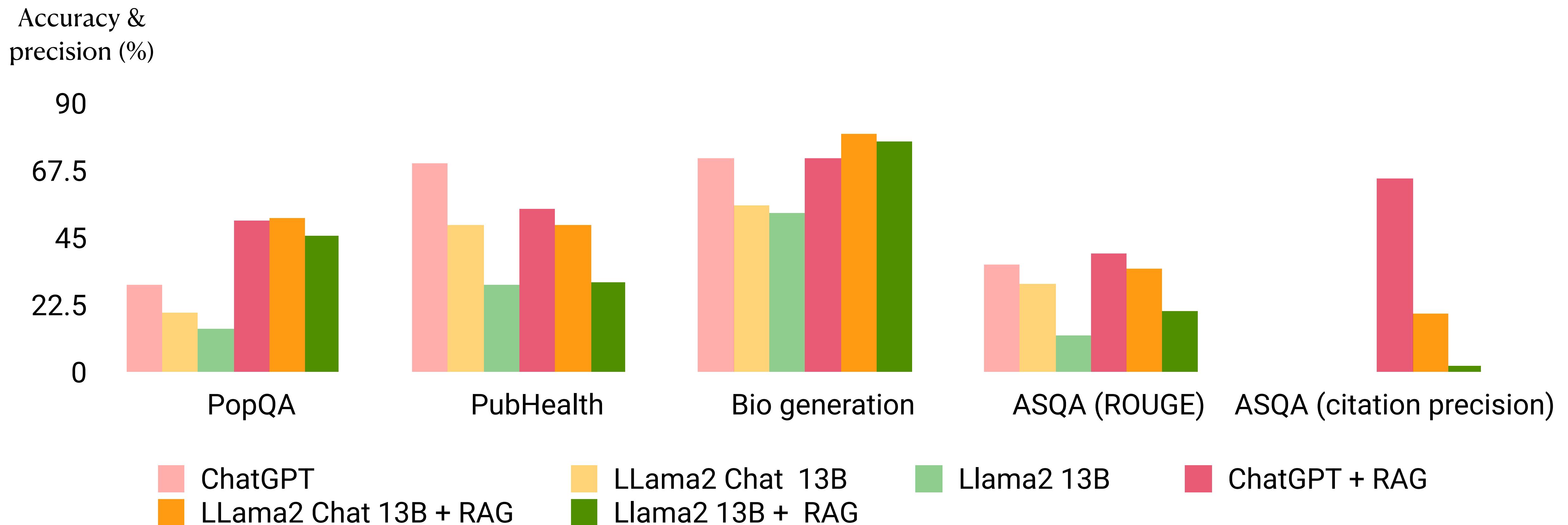
Experimental results

Parametric LMs struggle with tasks requiring precise world knowledge

Accuracy &
precision (%)



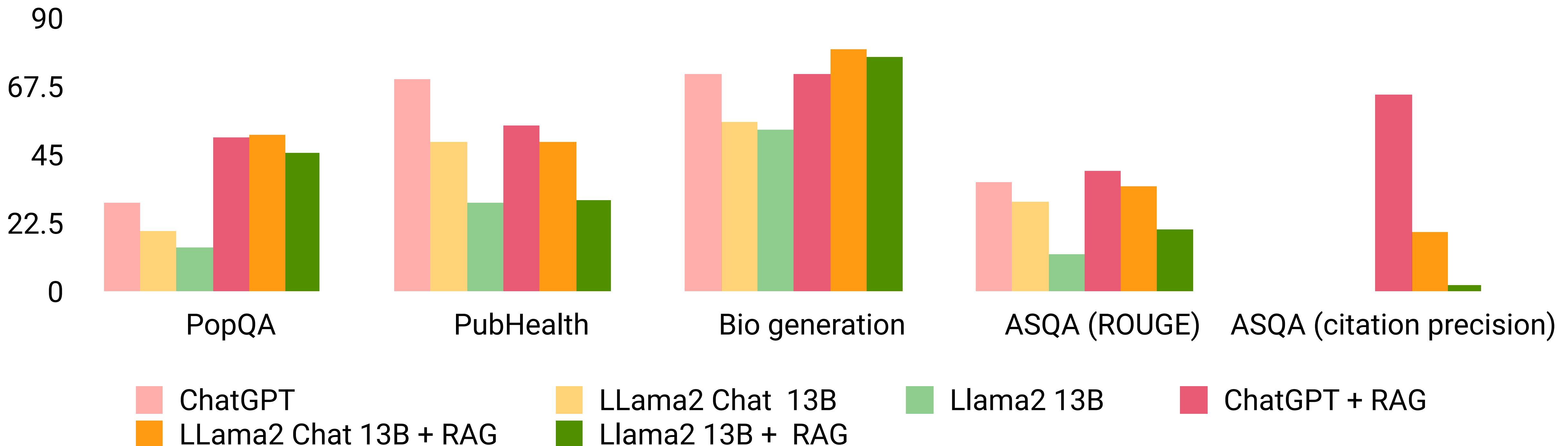
Experimental results



Experimental results

RAG gives large gains on some tasks, but not all

Accuracy &
precision (%)



ChatGPT

LLama2 Chat 13B + RAG

LLama2 Chat 13B

Llama2 13B + RAG

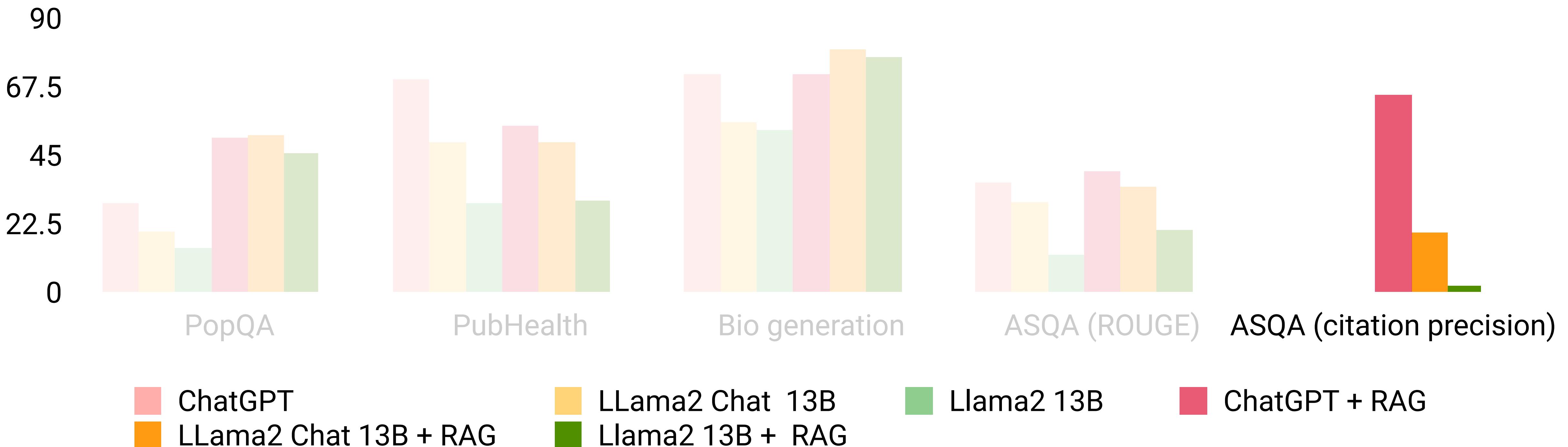
Llama2 13B

ChatGPT + RAG

Experimental results

RAG gives large gains on some tasks, but not all

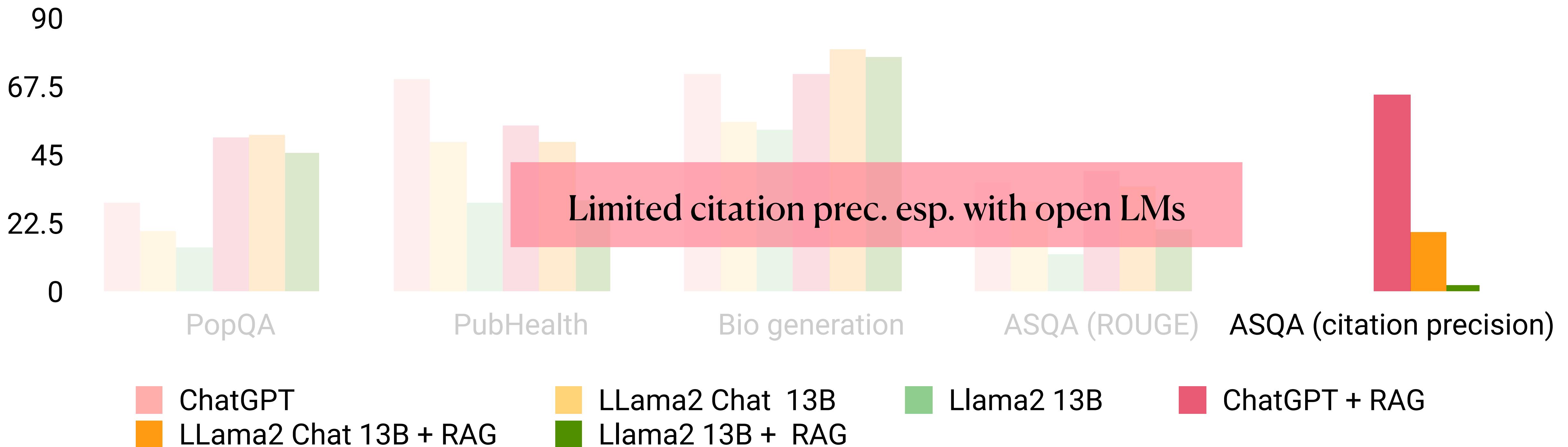
Accuracy &
precision (%)



Experimental results

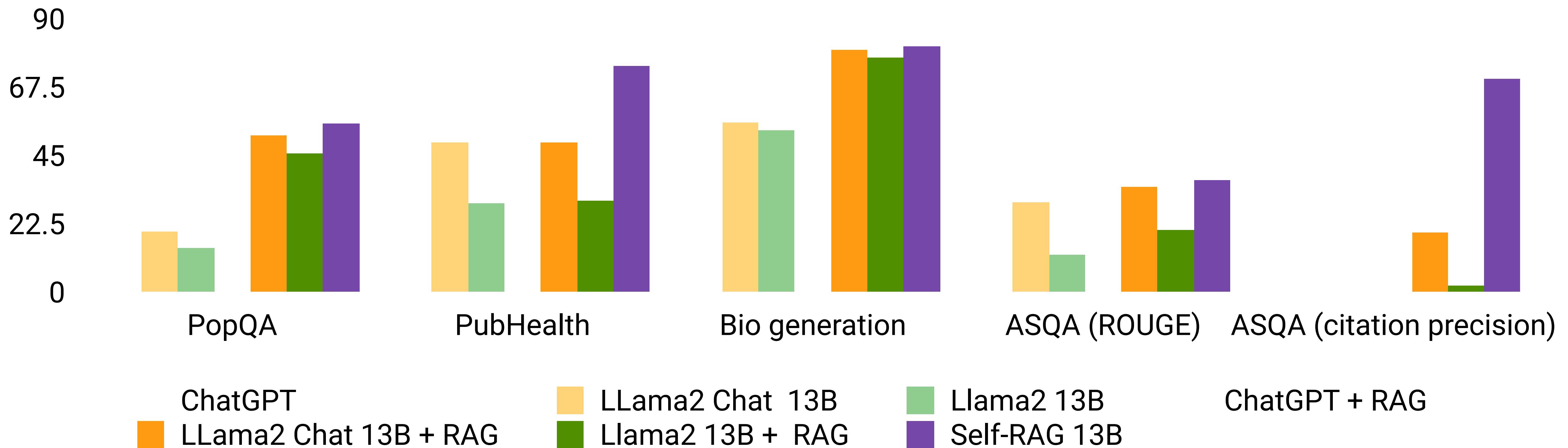
RAG gives large gains on some tasks, but not all

Accuracy & precision (%)

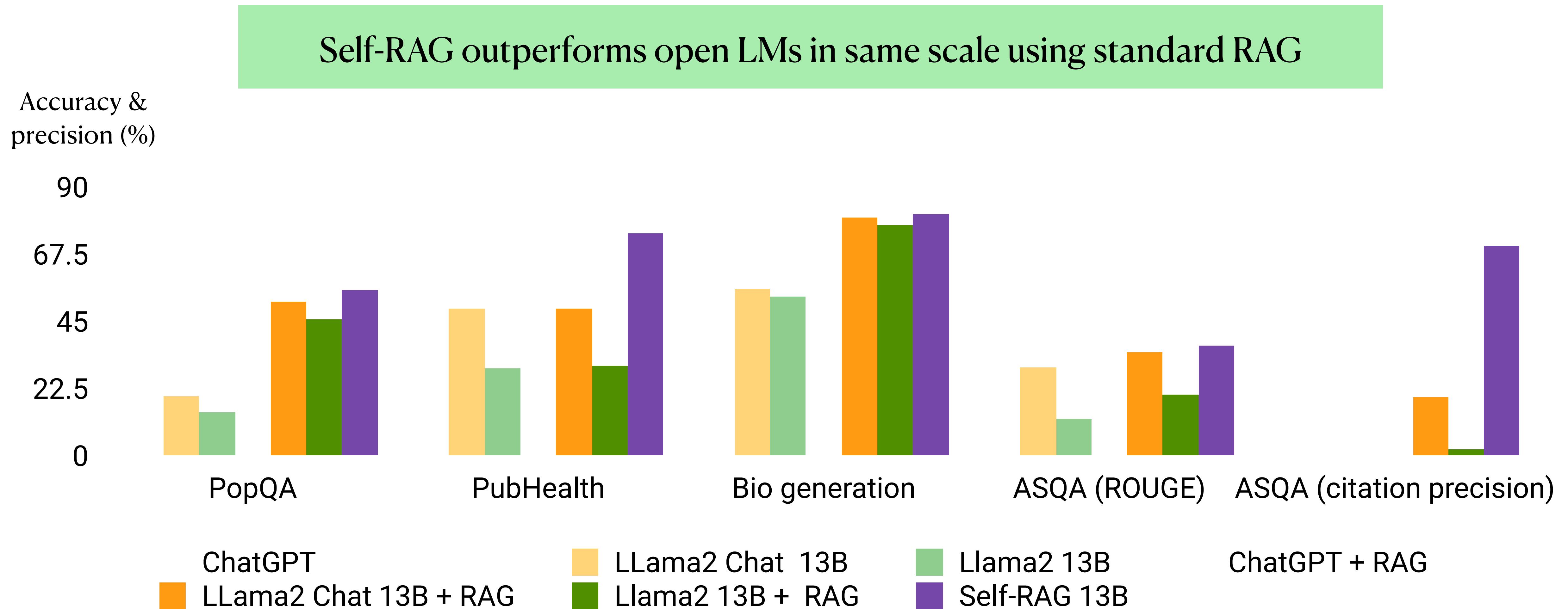


Experimental results

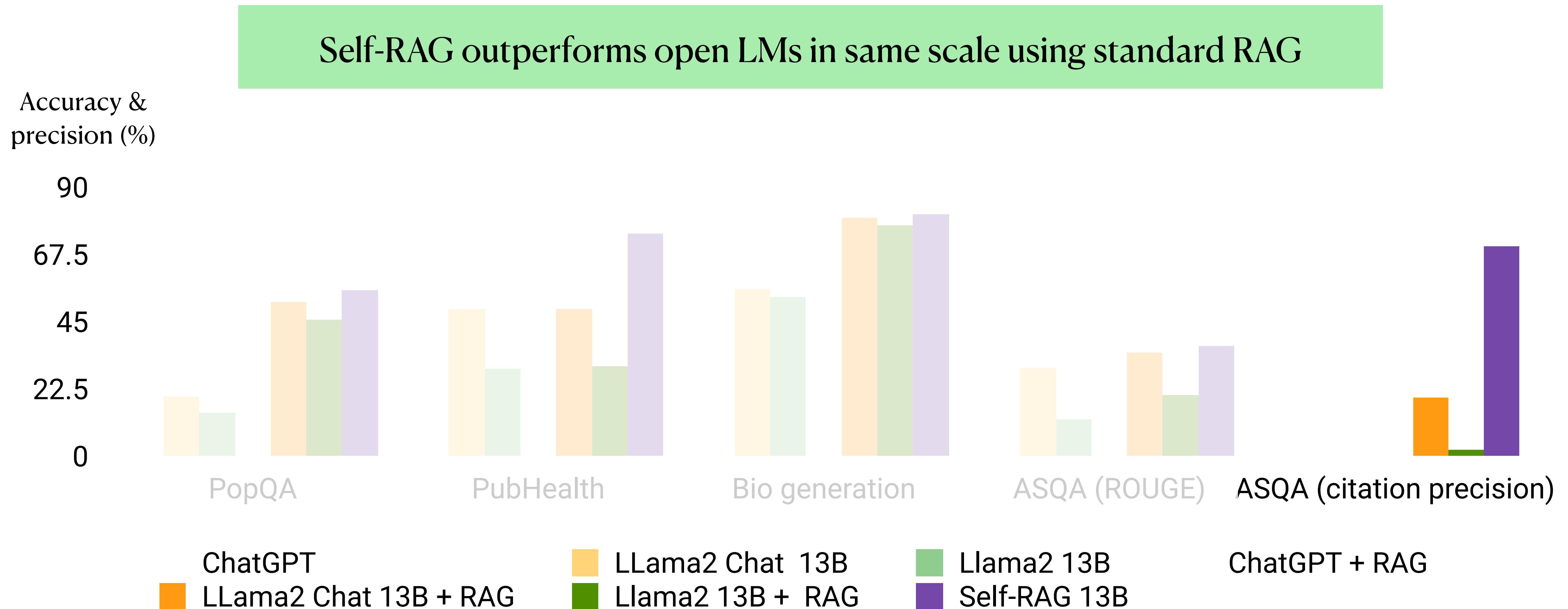
Accuracy &
precision (%)



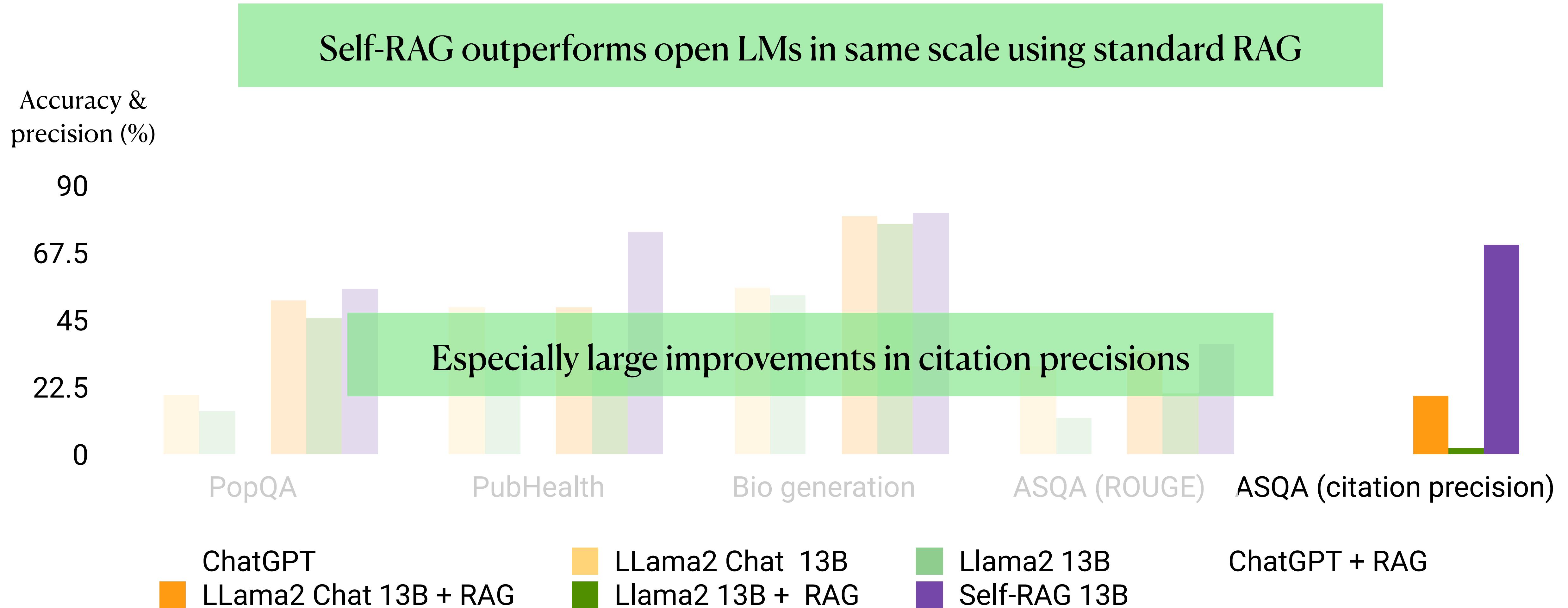
Experimental results



Experimental results

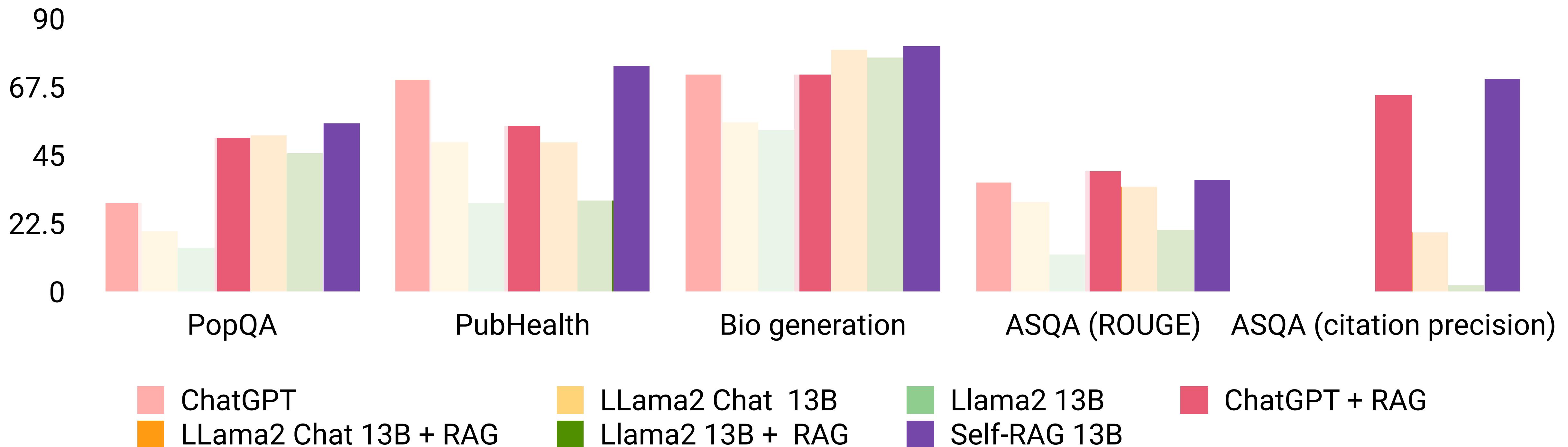


Experimental results



Experimental results

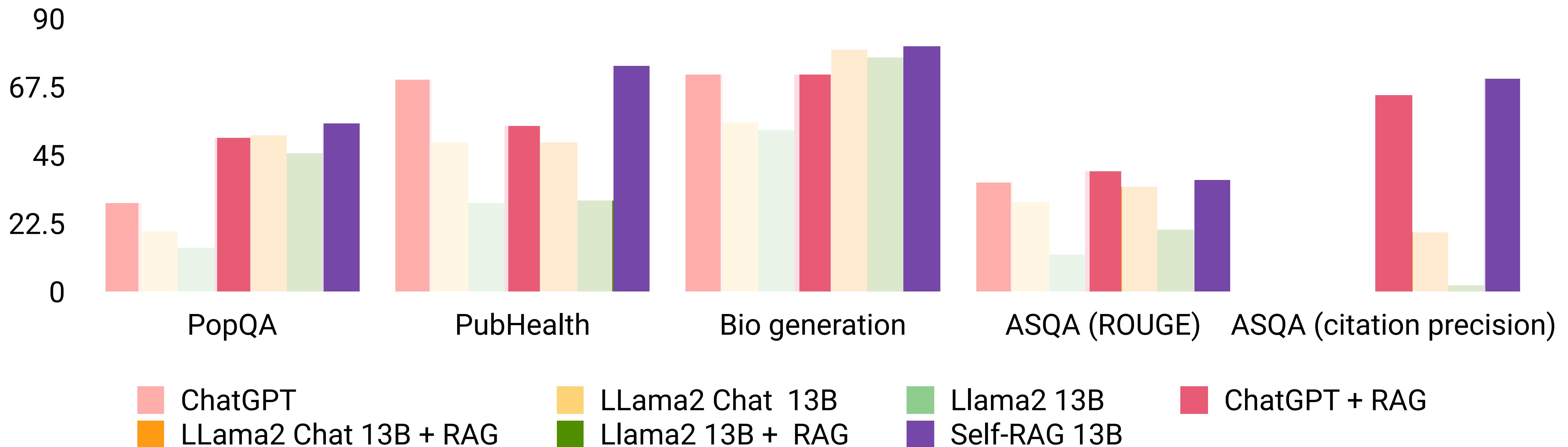
Accuracy &
precision (%)



Experimental results

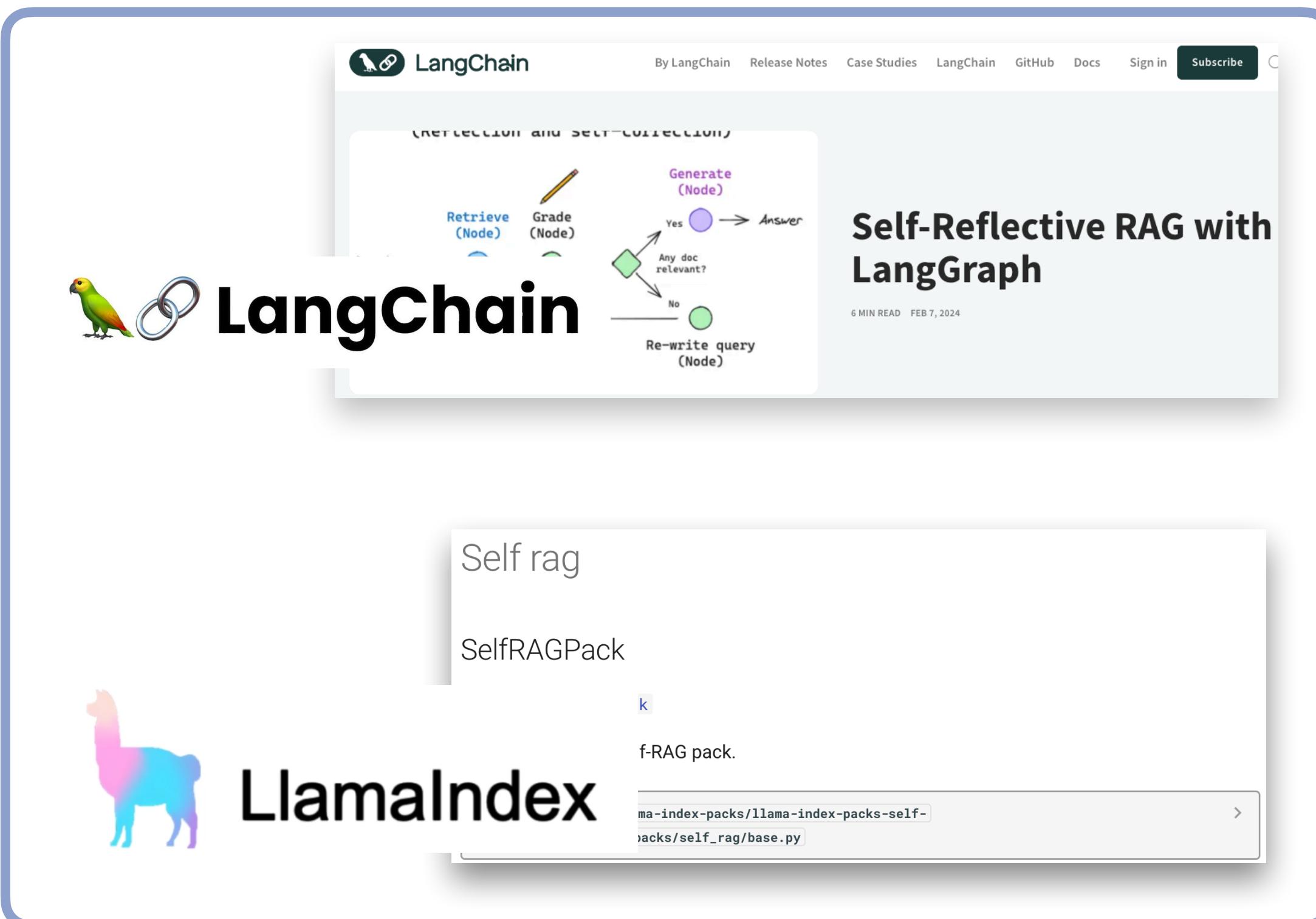
Self-RAG also outperforms ChatGPT with and without RAG

Accuracy &
precision (%)



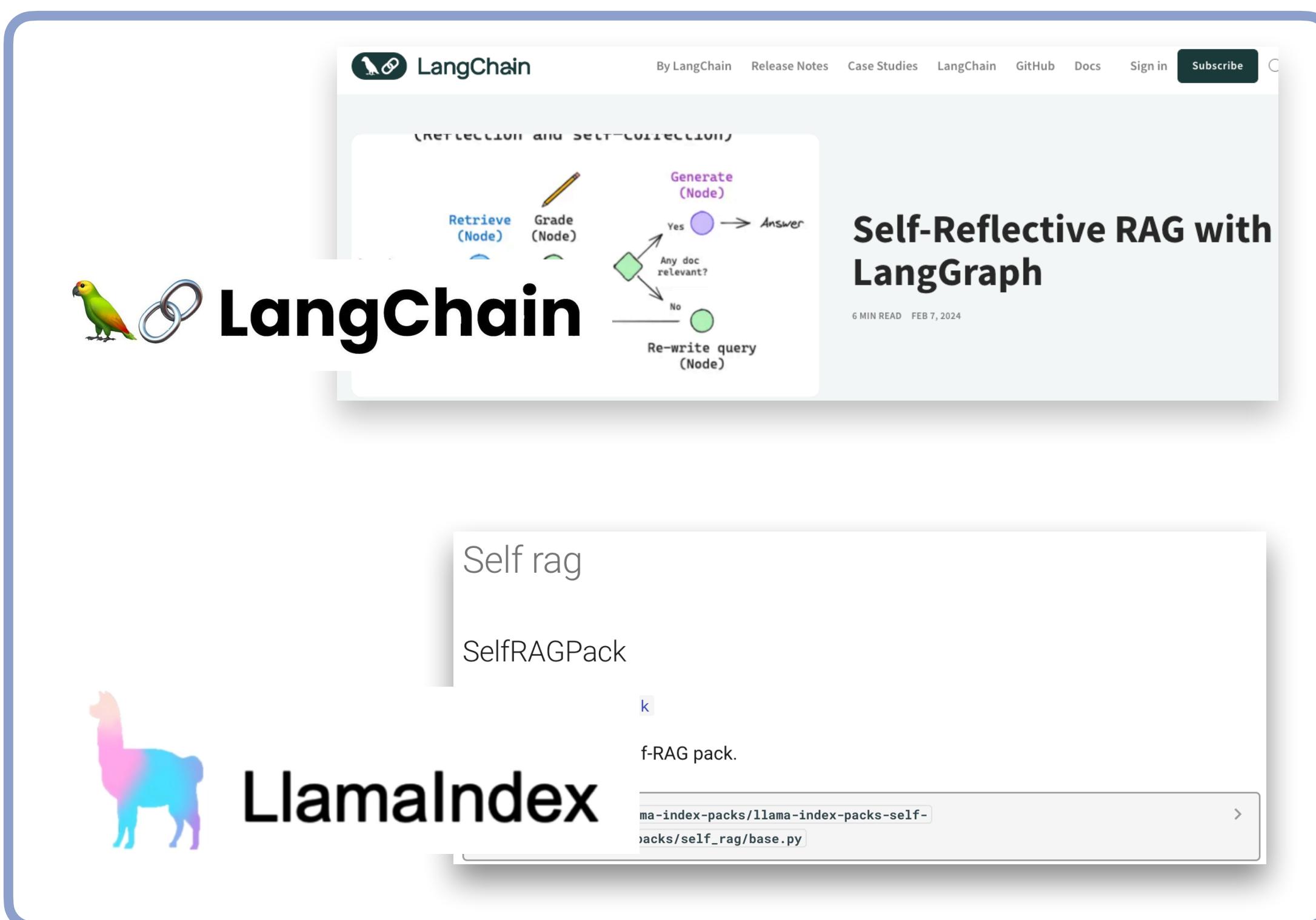
Followup of Self-RAG in academia and industry

Integrated into widely-adopted RAG libraries

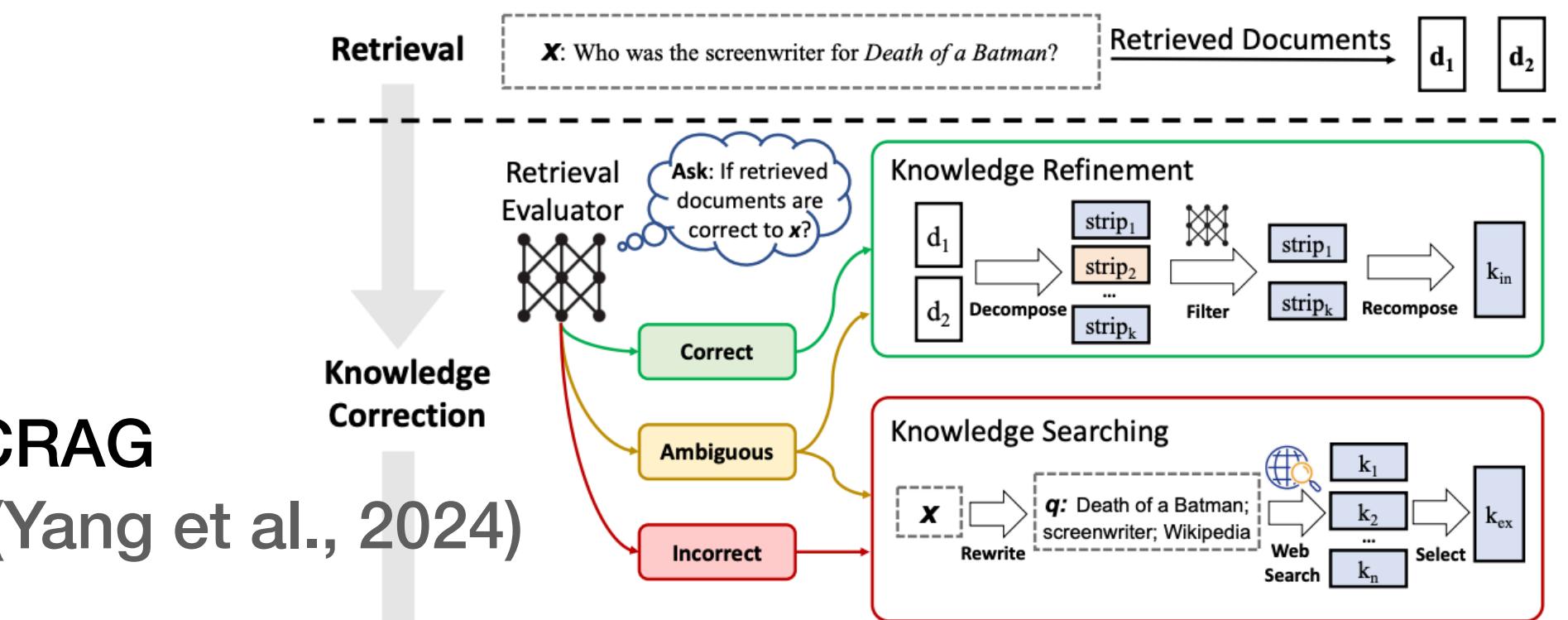


Followup of Self-RAG in academia and industry

Integrated into widely-adopted RAG libraries



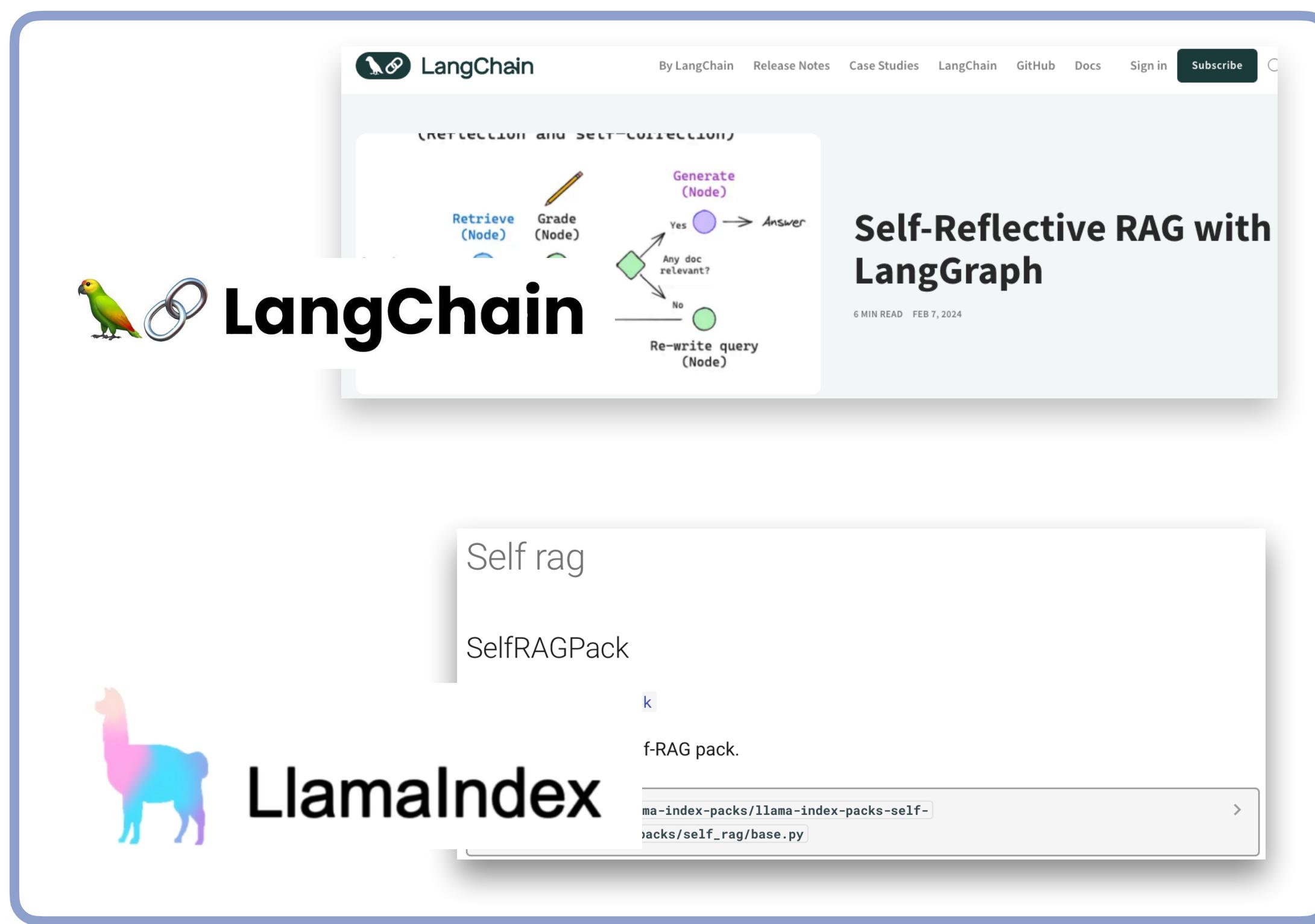
Improving Self-RAG for further reliability



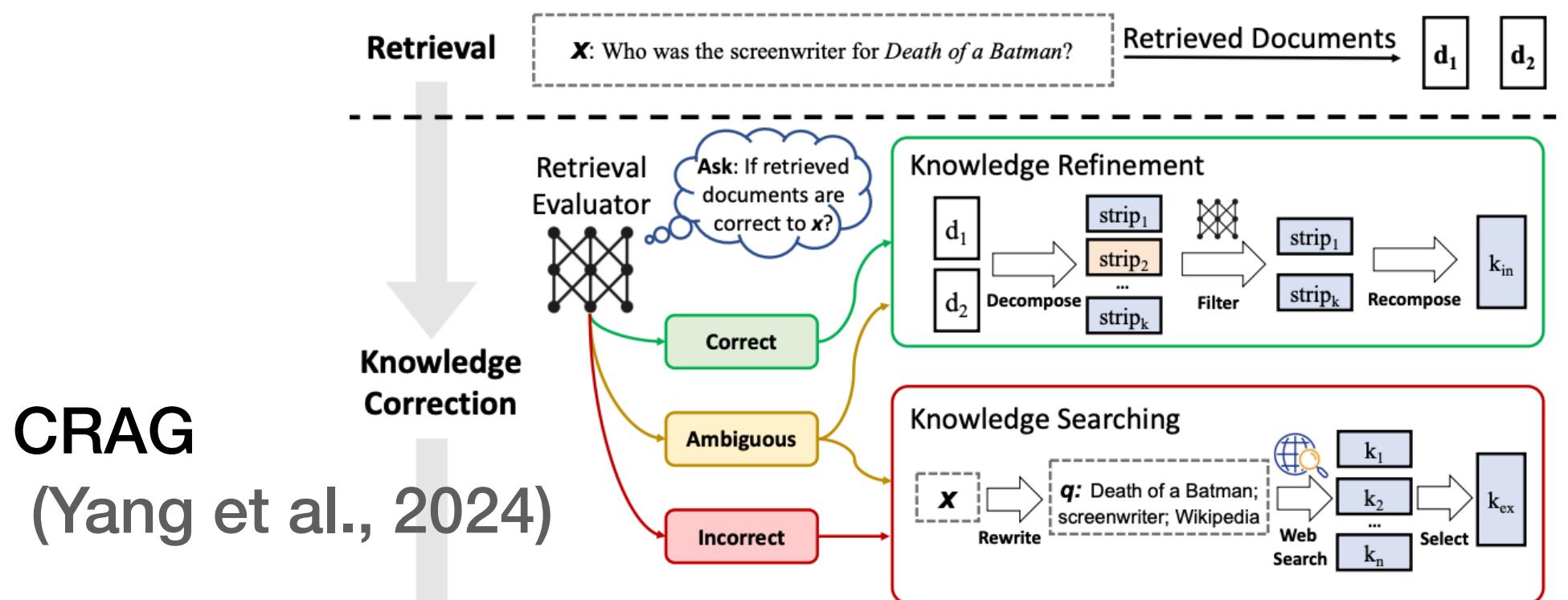
CRAG
(Yang et al., 2024)

Followup of Self-RAG in academia and industry

Integrated into widely-adopted RAG libraries



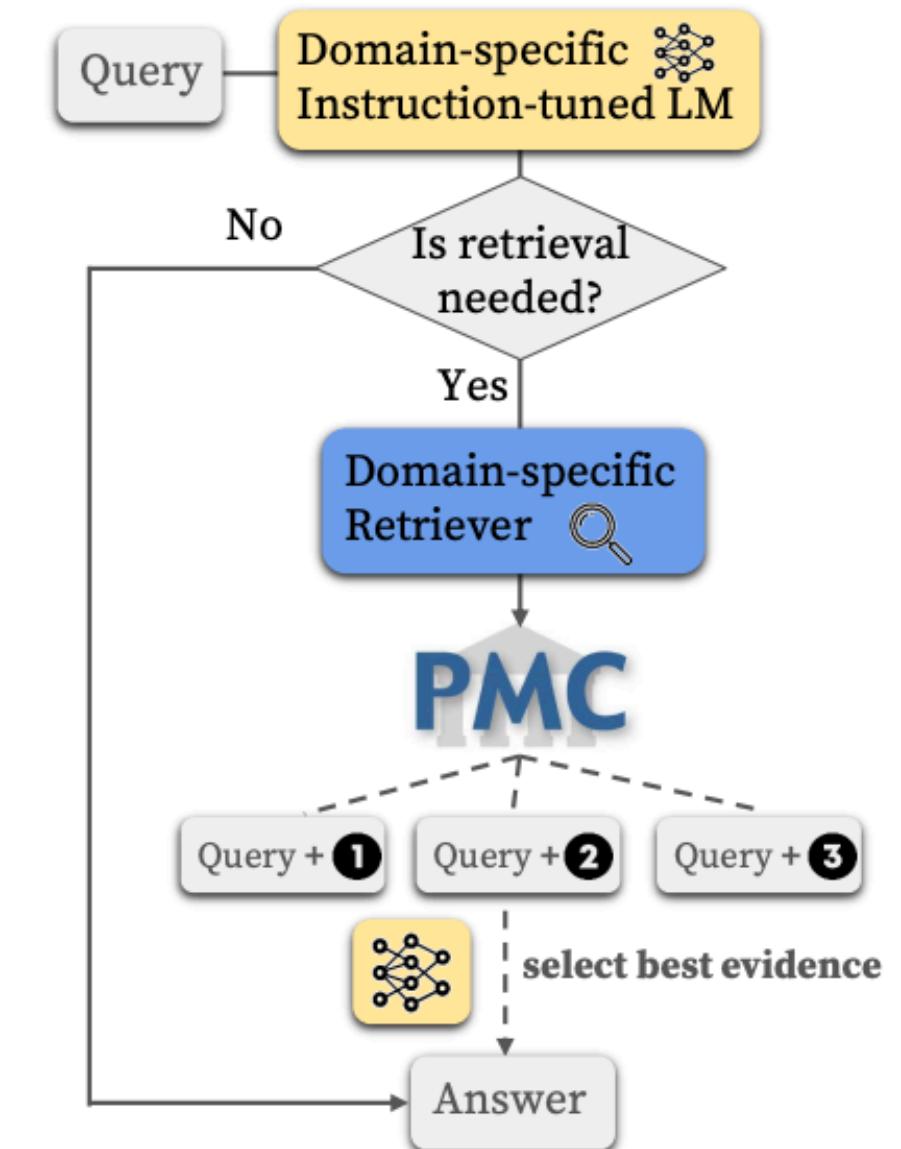
Improving Self-RAG for further reliability



CRAG
(Yang et al., 2024)

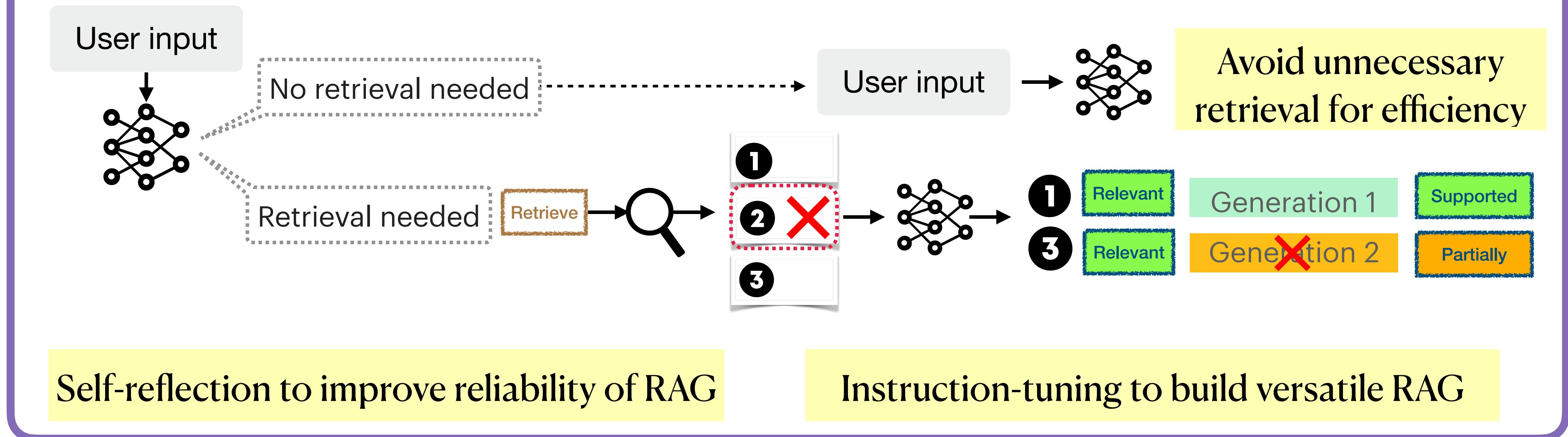
Applications to expert domains

Self-BioRAG
(Jeong et al., 2024)



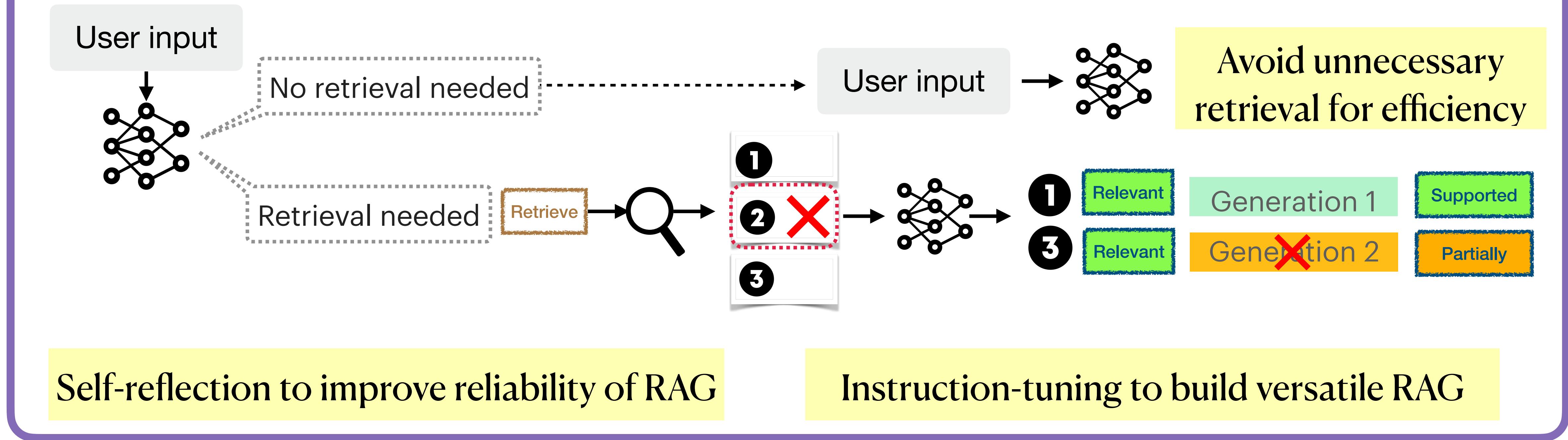
Thank you so much for listening!

Self-RAG: Learn to retrieve and reflect via special tokens



Thank you so much for listening!

Self-RAG: Learn to retrieve and reflect via special tokens



Question?

akari@cs.washington.edu

akariasai.github.io/

Poster: Hall B #63

Project page: selfrag.github.io/

Paper (Open Review): openreview.net/forum?id=hSyW5go0v8

HF checkpoints: huggingface.co/selfrag/selfrag_llama2_7b (13b)

Github: github.com/AkariAsai/self-rag (1.4k ⭐!)