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

selfrag.github.io

Akari Asai ¹
Zeqiu Wu ¹
Yizhong Wang ¹
Avirup Sil ³
Hannaneh Hajishirzi ¹²

¹ UWNLP
² AI²
³ IBM Research

May 7, 2024 | ICLR 2024
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

*Where is ICLR 2024?*
Retrieval-augmented generation (RAG)

**Step 1**: Retrieve $K$ documents

Retriever
(e.g., Google BM 25)

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.
Retrieval-augmented generation (RAG)

**Step 1: Retrieve K documents**

```
Retriever (e.g., Google BM 25)
```

```
Where is ICLR 2024?
```

Answer my question using references.

References:

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.

**Question:** Where is ICLR 2024?
Retrieval-augmented generation (RAG)

**Step 1: Retrieve K documents**

Where is ICLR 2024?

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?

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.
Retrieval-augmented generation (RAG)

**Step 1: Retrieve K documents**

Retriever (e.g., Google BM 25)

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.

**Step 2: Prompt LM with K docs and generate**

Answer my question using references.

References: 1 2

Question: Where is ICLR 2024?

ICLR 2024 will be in Vienna, Austria. 1 2

factual
up-to-date
attributable

LMs (e.g., GPT-4, Llama 3)
Retrieval-augmented generation (RAG)

**Step 1: Retrieve K documents**

Retriever (e.g., Google BM 25)

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.

**Step 2: Prompt LM with K docs and generate**

Answer my question using references.

References: ① ②

Question: Where is ICLR 2024?

ICLR 2024 will be in Vienna, Austria. ①②

- factual
- up-to-date
- attributable

Ram et al. In-Context Retrieval-Augmented Language Models. TACL 2023
RAG has been widely and successfully adopted
RAG has been widely and successfully adopted

RAG has shown effective in many benchmarks

RAG has been widely and successfully adopted.

RAG has shown effective in many benchmarks.

Widely adopted in real-world systems.

RAG has been widely and successfully adopted

RAG has shown effective in many benchmarks

Widely adopted in real-world systems

RAG-based search systems

RAG has been widely and successfully adopted

RAG has shown effective in many benchmarks

Widely adopted in real-world systems

RAG-based search systems

Libraries for customized RAG

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

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.
3. UTAH: Name taken from the Ute Indians who inhabited that region.
Standard RAG isn’t (always) reliable

**Step 1: Retrieve K documents**

How did US states get their names?

1. Of the fifty states, eleven including New York, Georgia, Washington named after an individual person.
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

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**

- How did US states get their names?
- Of the fifty states, eleven including New York, Georgia, Washington named after an individual person.
- The history of human activity in Michigan began with settlement by Paleo-Indians.
- 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?

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)

- How did US states get their names?

- Of the fifty states, eleven including New York, Georgia, Washington named after an individual person.

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

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

Easily distracted by unhelpful documents

Step 2: Prompt LM with K docs and generate

Prompt: How did US states get their names?

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)

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

Easily distracted by unhelpful documents

**Step 2: Prompt LM with K docs and generate**

Prompt How did US states get their names?

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)

How did US states get their names?

1. Of the fifty states, eleven including New York, Georgia, Washington named after an individual person.
3. UTAH: Name taken from the Ute Indians who inhabited that region.

Easily distracted by unhelpful documents

**Step 2: Prompt LM with K docs and generate**

Prompt: How did US states get their names?

- 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)

- Prompt: How did US states get their names?

1. Of the fifty states, eleven including New York, Georgia, Washington named after an individual person.
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?

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**

How did US states get their names?

1. Of the fifty states, eleven including New York, Georgia, Washington named after an individual person.
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

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**

- **Prompt**: How did US states get their names?

  1. Of the fifty states, eleven including New York, Georgia, Washington named after an individual person.
  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?

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)

- How did US states get their names?
  - Of the fifty states, eleven including New York, Georgia, Washington named after an individual person.
  - The history of human activity in Michigan began with settlement by Paleo-Indians.
  - 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?

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
Write an essay about your summer vacation

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 is a 2023 Spanish comedy film directed by Santiago Segura which stars Segura, Leo Harlem, Cristina Gallego, and Patricia Conde.
Standard RAG isn’t efficient

Write an essay about your summer vacation

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

User input

No retrieval needed
Self-RAG: *Self-reflective Retrieval-Augmented Generation*

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

User input → No retrieval needed
Self-RAG: Self-reflective Retrieval-Augmented Generation

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

User input → No retrieval needed → User input
Self-RAG: *Self-reflective Retrieval-Augmented Generation*

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

- User input
- **No retrieval needed**
- User input
- Avoid unnecessary retrieval
Self-RAG: Self-reflective Retrieval-Augmented Generation

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

User input

No retrieval needed

User input

Retrieval needed

Avoid unnecessary retrieval
Self-RAG: Self-reflective Retrieval-Augmented Generation

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

- User input
- No retrieval needed
- Retrieval needed
- User input
- Avoid unnecessary retrieval
Self-RAG: *Self-reflective Retrieval-Augmented Generation*

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

- **User input**
  - No retrieval needed
  - Retrieval needed
  - Avoid unnecessary retrieval

- **Retrieve**
  - User input

![Diagram showing retrieval and inference process](image)
Self-RAG: Self-reflective Retrieval-Augmented Generation

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

User input

No retrieval needed

Retrieval needed

1

2

3

Avoid unnecessary retrieval
Self-RAG: *Self*-reflective Retrieval-Augmented Generation

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

User input

No retrieval needed

Retrieval needed

1. Retrieve

2. User input

3. Avoid unnecessary retrieval

User input
Self-RAG: Self-reflective Retrieval-Augmented Generation

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

User input

No retrieval needed

User input

Avoid unnecessary retrieval

Retrieval needed: Retrieve

1

2 × Irrelevant

3
**Self-RAG: Self-reflective Retrieval-Augmented Generation**

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

- **User input**
  - No retrieval needed
  - Retrieval needed

1. **Identify unhelpful evidences**
2. **Avoid unnecessary retrieval**
3. **No retrieval needed**

User input → Retrieval needed → Retrieve → Irrelevant → User input → Avoid unnecessary retrieval
Self-RAG: Self-reflective Retrieval-Augmented Generation

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

User input

No retrieval needed

Retrieval needed

Retrieve

Identify unhelpful evidences

Avoid unnecessary retrieval

User input

1 Relevant

2 Relevant

3 Relevant

1 Irrelevant

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

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

- User input
  - No retrieval needed
  - Retrieval needed: Retrieve
  - Identify unhelpful evidences

- Generation 1
  - Relevant
  - Supported

- Avoid unnecessary retrieval
Self-RAG: Self-reflective Retrieval-Augmented Generation

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

User input

No retrieval needed

Retrieval needed

Identify unhelpful evidences

Avoid unnecessary retrieval

User input

Retrieval needed

1. Relevant
   Generation 1
   Supported

2. Irrelevant

3. Relevant
   Generation 2
   Partially
Self-RAG: Self-reflective Retrieval-Augmented Generation

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

User input

No retrieval needed

Retrieve needed

Identify unhelpful evidences

Avoid unnecessary retrieval

User input

Retrieve

Generation 1

Supported

Generation 2

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

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

- **User input**
- **No retrieval needed**
- **Retrieve**
- **Retrieval needed**
- **Identify unhelpful evidences**
- **Generate output supported by evidence**

Avoid unnecessary retrieval

1. Relevant
2. Irrelevant
3. Relevant

Generation 1: Supported
Generation 2: Partially
Self-RAG: **Self-reflective Retrieval-Augmented Generation**

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

User input → Retrieval needed → Retrieve → User input

- No retrieval needed
- Retrieval needed

1. Identify unhelpful evidences
2. Generate output supported by evidence
3. Avoid unnecessary retrieval

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

1. Relevant Generation 1 (Supported)
2. Irrelevant Generation 2 (Partially)
Self-RAG: *Self-reflective Retrieval-Augmented Generation*

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

- **User input**
  - No retrieval needed
  - Retrieval needed
  - Identify unhelpful evidences
  - Generate output supported by evidence
  - Avoid unnecessary retrieval

**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. 

- **Input:** How did US states get their names?
  - **Augmented Output:** Retrieve 1 Of the fifty states, eleven are named after an individual person. Supported Retrieve 2 LOUISIANA: Named in honor of Louis XIV of France. Relevant For instance, Louisiana was named after King Louis XIV, and Georgia was named after King George II. Partially Relevant Util: 5
**Self-RAG: Self-reflective Retrieval-Augmented Generation**

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

- User input
  - No retrieval needed
- Retrieval needed
  - Retrieve
- Identify unhelpful evidences
  - Generation 1: Relevant
  - Generation 2: Partially
- Generate output supported by evidence
  - Avoid unnecessary retrieval

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

**Input:** Write an essay of your best summer vacation

**Augmented Output:** My best summer vacation was a magical escape to the coastal town of Santorini. The azure waters, charming white-washed building are unforgettable experience.

**Input:** How did US states get their names?

**Augmented Output:**
1. Retrieve: Of the fifty states, eleven are named after an individual person.
2. Relevant: 11 of 50 states’ names come from person.
3. Relevant: For instance, Louisiana was named after King Louis XIV, and Georgia was named after King George II.
**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

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.
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.
How did US states get their names?

US states got their names from a variety of sources.

1. Of the fifty states, eleven including New York, Georgia, Washington named after an individual person.
3. UTAH: Name taken from the Ute Indians who inhabited that region.
How did US states get their names?

US states got their names from a variety of sources.

1. Of the fifty states, eleven including New York, Georgia, Washington named after an individual person.


3. UTAH: Name taken from the Ute Indians who inhabited that region.
The history of human activity in Michigan began with settlement by Paleo-Indians.

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

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

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

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

UTAH: Name taken from the Ute Indians who inhabited that region
The history of human activity in Michigan began with settlement by Paleo-Indians.

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

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

UTAH: Name taken from the Ute Indians who inhabited that region.
The history of human activity in Michigan began with settlement by Paleo-Indians.

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

1. Of the fifty states, eleven states named after an individual person.
3. UTAH: Name taken from the Ute Indians who inhabited that region.

11 of 50 state names come from persons.
The history of human activity in Michigan began with settlement by Paleo-Indians.

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.

Relevant: 11 of 50 state names come from persons.
The history of human activity in Michigan began with settlement by Paleo-Indians.

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.

Relevant 11 of 50 state names come from persons.

Irrelevant Many states including Michigan
The history of human activity in Michigan began with settlement by Paleo-Indians.

Many states including Michigan are named after an individual person.

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

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

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. Relevant
   11 of 50 state names come from persons.

2. Irrelevant
   Many states including Michigan are named after an individual person.

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

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

5. Relevant
   UTAH: Name taken from the Ute Indians who inhabited that region.
The history of human activity in Michigan began with settlement by Paleo-Indians.

Many states including Michigan are named after an individual person.

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

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

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

11 of 50 state names come from persons.
The history of human activity in Michigan began with settlement by Paleo-Indians.

Many states including Michigan are named after an individual person.

Several states such as Utah and

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

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

UTAH: Name taken from the Ute Indians who inhabited that region.
The history of human activity in Michigan began with settlement by Paleo-Indians. Many states including Michigan are named after an individual person. Several states such as Utah and Alabama are named after indigenous communities. Of the fifty states, eleven states named after an individual person. The history of human activity in Michigan began with settlement by Paleo-Indians. 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. **Prompt 1**
   - Relevant: 11 of 50 state names come from persons.

2. **Prompt 2**
   - Irrelevant: Many states including Michigan are named after an individual person.
   - Relevant: Several states such as Utah and Alabama are named after indigenous communities.

3. **Prompt 3**
   - Relevant: UTAH: Name taken from the Ute Indians who inhabited that region.

Of the fifty states, eleven states named after an individual person.
**Self-RAG** — **Self-Reflective Retrieval-Augmented Generation**

**Step 2: Generate segments in parallel**

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

2. **Prompt 2**: Many states including Michigan are named after an individual person.

3. **Prompt 3**: Several states such as Utah and Alabama are named after indigenous communities.

**Batch decoding**

- **Relevant**: 11 of 50 state names come from persons.
- **Supported**: Relevant
- **Irrelevant**: Many states including Michigan are named after an individual person.
- **Relevant**: Several states such as Utah and Alabama are named after indigenous communities.

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

**3** UTAH: Name taken from the Ute Indians who inhabited that region.
Step 2: Generate segments in parallel

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

2. Prompt ②: Many states including Michigan are named after an individual person.

3. Prompt ③: Several states such as Utah and Alabama are named after indigenous communities.

Relevant: 11 of 50 state names come from persons. Supported
Irrelevant: Many states including Michigan are named after an individual person.
Relevant: Several states such as Utah and Alabama are named after indigenous communities.
Partially: UTAH: Name taken from the Ute Indians who inhabited that region.
Reflection tokens for retrieval and critique

US states got their names from a variety of sources.

Original LM vocabularies

California
e.g.,
A
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.,

Original LM vocabularies

Retrieve

No Retrieval
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

Retrieval tokens
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

Retrieval tokens

Retrieve
No Retrieval
Relevant
Irrelevant
Supported
No support
Useful
Not useful
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

Retrieval tokens

Critique tokens

Retrieval
No Retrieval
Relevant
Irrelevant
Supported
No support
Useful
Not useful
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.

Adaptive retrieval via retrieval reflection tokens

\[
p(\text{Retrieve} = \text{YES}) > \delta
\]

\[
p(\text{Retrieve} = \text{YES}) + p(p(\text{Retrieve} = \text{No})) > \delta
\]
**Self-RAG** — **Self-Reflective Retrieval-Augmented Generation**

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

- **Prompt 1**: Relevant 11 of 50 state names come from persons. Supported
- **Prompt 2**: Irrelevant Many states including Michigan are named after an individual person.
- **Prompt 3**: Relevant Several states such as Utah and Alabama are named after native American tribes.
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

- **Prompt 1:** 11 of 50 state names come from persons.

- **Prompt 2:** Many states including Michigan are named after an individual person.

- **Prompt 3:** Several states such as Utah and Alabama are named after native American tribes.
Self-RAG — Self-Reflective Retrieval-Augmented Generation

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

1. **Prompt 1**
   - Relevant: 11 of 50 state names come from persons.
   - Supported
   - $f(\text{Relevant, Supported}) = 0.9$

2. **Prompt 2**
   - Irrelevant: Many states including Michigan are named after an individual person.
   - $f(\text{Irrelevant}) = 0.1$

3. **Prompt 3**
   - Relevant: Several states such as Utah and Alabama are named after native American tribes.
   - Partially
   - $f(\text{Relevant, Partially}) = 0.4$
Self-RAG: **Self-reflective Retrieval-Augmented Generation**

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

- User input
- Retrieval needed
- Identify unhelpful evidences
- Generate output supported by evidence
- Avoid unnecessary retrieval

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

**Input:** How did US states get their names?

**Augmented Output:**

1. Relevant: 11 of 50 states’ names come from person.
2. Relevant: For instance, Louisiana was named after King Louis XIV, and Georgia was named after King George II.

**Input:** Write an essay of your best summer vacation

**Augmented Output:**

1. No Retrieval: My best summer vacation was a magical escape to the coastal town of Santorini.
2. No Retrieval: The azure waters, charming white-washed building are unforgettable experience.
How did US states get their names?

1. Of the fifty states, eleven are named after an individual person.
How did US states get their names?

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

Critic LM teaches Generator LM to predict reflection tokens

Training time only

Overview of Self-RAG training
How did US states get their names?

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.
How did US states get their names?

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

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.

$\mathcal{D}_{critic} = \{(x, y, d), r\}$
How did US states get their names?

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

11 of 50 state names come from persons.

Supported

\[ \max_C \mathbb{E}_{(x,y,d),r \sim \mathcal{D}_{\text{critic}}} \log p_C(r | x, y, d) \]

\[ \mathcal{D}_{\text{critic}} = \{(x, y, d), r\} \]
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.

Augmented Output:

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

2. LOUISIANA: Named in honor of Louis XIV of France.
   - Relevant
   - Partially
   - Util: 5

For instance, Louisiana was named after King Louis XIV, and Georgia was named after King George II.
How did US states get their names?

Of the fifty states, eleven are named after an individual person.
How did US states get their names?

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

11 of 50 state names come from persons.

Supported
Self-RAG training - Generator LM

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

\(x\) How did US states get their names?

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

\(y\) 11 of 50 state names come from persons.

\(r\) Supported
Self-RAG training - Generator LM

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

Train with a standard next token objective with expanded vocabulary

How did US states get their names?

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

11 of 50 state names come from persons.

Supported
**Self-RAG training - Generator LM**

Train with a standard next token objective with expanded vocabulary

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

**x**  How did US states get their names?  ↓

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

**y**  →

**r**  Supported

11 of 50 state names come from persons.

Easily applied to new pre-trained LM
Experimental details

Tasks and datasets

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

- **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)

More details of training & test are in our paper!
Experimental details

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)

More details of training & test are in our paper!
Experimental details

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)

More details of training & test are in our paper!
Experimental details

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)

More details of training & test are in our paper!
Experimental results

Accuracy & precision (%)

- PopQA
- PubHealth
- Bio generation
- ASQA (ROUGE)
- ASQA (Citation Precision)

ChatGPT
LLama2 Chat 13B
Llama2 13B
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 (%)

<table>
<thead>
<tr>
<th>Task</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>PopQA</td>
<td>ChatGPT, LLama2 Chat 13B, Llama2 13B</td>
</tr>
<tr>
<td>PubHealth</td>
<td>ChatGPT, LLama2 Chat 13B, Llama2 13B</td>
</tr>
<tr>
<td>Bio generation</td>
<td>ChatGPT, LLama2 Chat 13B, Llama2 13B</td>
</tr>
<tr>
<td>ASQA (ROUGE)</td>
<td>ChatGPT, LLama2 Chat 13B, Llama2 13B</td>
</tr>
<tr>
<td>ASQA (Citation Precision)</td>
<td>ChatGPT, LLama2 Chat 13B, Llama2 13B</td>
</tr>
</tbody>
</table>
Experimental results

Accuracy & precision (%)
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
Experimental results

RAG gives large gains on some tasks, but not all

Limited citation prec. esp. with open LMs
Experimental results

Accuracy & precision (%)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Models</th>
</tr>
</thead>
</table>
Experimental results

Self-RAG outperforms open LMs in same scale using standard RAG

Accuracy & precision (%)

- PopQA
- PubHealth
- Bio generation
- ASQA (ROUGE)
- ASQA (citation precision)

ChatGPT
LLama2 Chat 13B + RAG
Llama2 13B
Llama2 Chat 13B + RAG
Llama2 13B
Self-RAG 13B
ChatGPT + RAG
Experimental results

Self-RAG outperforms open LMs in same scale using standard RAG

Accuracy & precision (%)

0 22.5 45 67.5 90

PopQA  PubHealth  Bio generation  ASQA (ROUGE)  ASQA (citation precision)

ChatGPT  LLama2 Chat 13B + RAG  LLama2 13B + RAG  Llama2 13B  Self-RAG 13B  ChatGPT + RAG
Experimental results

Self-RAG outperforms open LMs in same scale using standard RAG

Especially large improvements in citation precisions
Experimental results

Accuracy & precision (%)

ChatGPT
LLama2 Chat 13B + RAG
Llama2 13B + RAG
Self-RAG 13B
ChatGPT + RAG
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*

- User input
  - No retrieval needed
  - Retrieval needed

- User input
  - Avoid unnecessary retrieval for efficiency

**Instruction-tuning to build versatile RAG**

**Self-reflection to improve reliability of RAG**
Self-RAG: Learn to retrieve and reflect via special tokens

User input

No retrieval needed

User input

Retrieval needed

Avoid unnecessary retrieval for efficiency

Self-reflection to improve reliability of RAG

Instruction-tuning to build versatile RAG

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/selffrag/selffrag_llama2_7b (13b)

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