

# SimSage AI settings

You can set up an OpenAI or a Google Gemini Large Language models (LLMs) in SimSage. This is to aid Retrieval Augmented Generation (RAG) as supported by SimSage.

This document discusses the settings in the admin UX for enabling this feature. A *token* in the context below is anything SimSage considers to be a separate token. This includes spaces, words, punctuation items, and entities like credit cards etc. Each of such items is considered a distinct token in SimSage.

Navigate to “AI set up” in “the Mind” section for a given knowledge-base.

## None

Select “none” to disable the use of AI and large language models (default).

SimSage

Overview

User Management

demo knowledge base

Document Management

The Mind

AI set up

Synonyms

Semantics

Synsets

Large Language Model (LLM)

Amazon Bedrock

None

OpenAI

Google Gemini

Amazon Bedrock

CoPilot

runov

Amazon key \*

.....

Bedrock model

Mistral Large

Amazon region

Asia Pacific (Sydney)

search box question answering, the number of results to pass to the LLM

5

search box question answering, the number of tokens to pass for each search result

100

Save

SimSage AI Setup Guide

## Open AI settings

SimSage

Overview

User Management

demo knowledge base

Document Management

The Mind

AI set up

Synonyms

Semantics

Synsets

Large Language Model (LLM)

OpenAI

OpenAI key \*

.....

OpenAI model

gpt-3.5-turbo

gpt-3.5-turbo

gpt-4

gpt-4o

document Q&A token count

1000

document summarization token count

1000

search box question answering, the number of results to pass to the LLM

5

search box question answering, the number of tokens to pass for each search result

100

Save

SimSage AI Setup Guide

Open AI specific settings are

setting	description
---------	-------------

OpenAI key	your API key from OpenAI
OpenAI model	the OpenAI model to use for inference

## Google Gemini settings

**SimSage**

Overview | **AI set up** | Synonyms | Semantics | Synsets

User Management | demo knowledge base | Document Management | The Mind

### Large Language Model (LLM)

**Google Gemini**

Gemini key

Gemini model **gemini-pro**

document Q&A token count

document summarization token count

search box question answering, the number of results to pass to the LLM

search box question answering, the number of tokens to pass for each search result

**Save**

[SimSage.AI Setup Guide](#)

setting	description
Gemini key	your Google Gemini API key
Gemini model	the Gemini model you wish to use for inference

## Amazon Bedrock

**SimSage**

Overview | **AI set up** | Synonyms | Semantics | Synsets

User Management | demo knowledge base | Document Management | The Mind

### Large Language Model (LLM)

**Amazon Bedrock**

Amazon key id  Amazon key

Bedrock model **Select Model**

Amazon region **Asia Pacific (Sydney)**

document Q&A token count

document summarization token count

search box question answering, the number of results to pass to the LLM

search box question answering, the number of tokens to pass for each search result

**Save**

[SimSage.AI Setup Guide](#)

setting	description
Amazon key id	the Amazon key id (careful, these can be region specific)
Amazon key	the secret associated with the Amazon key id

Bedrock model	the AWS Bedrock model you wish to use for inference
Amazon region	One of the AWS regions where the model is enabled and hosted

## CoPilot

CoPilot is an Azure specific service connecting Azure to OpenAI.

The screenshot shows the SimSage AI Setup Guide interface. On the left is a sidebar with navigation links: Overview, User Management, demo knowledge-base (selected), Document Management, and The Mind. The main content area is titled 'Large Language Model (LLM)' and has tabs for AI set up, Synonyms, Semantics, and Synsets. Under the 'AI set up' tab, there is a 'CoPilot' dropdown menu. Below it are three input fields: 'CoPilot key' (with a red asterisk), 'CoPilot endpoint', and 'CoPilot model' (with a dropdown menu showing 'gpt-4'). Further down are four input fields for token counts: 'document Q&A token count' (1000), 'document summarization token count' (1000), 'search box question answering, the number of results to pass to the LLM' (5), and 'search box question answering, the number of tokens to pass for each search result' (100). A blue 'Save' button is at the bottom left of the form.

setting	description
CoPilot key	the CoPilot key to use, provided by Azure once you create a CoPilot instance.
CoPilot endpoint	the Azure endpoint for the CoPilot instance you created inside Azure (provided by Azure)
CoPilot model	the model to use for inference with CoPilot, currently only gpt-4 is available.

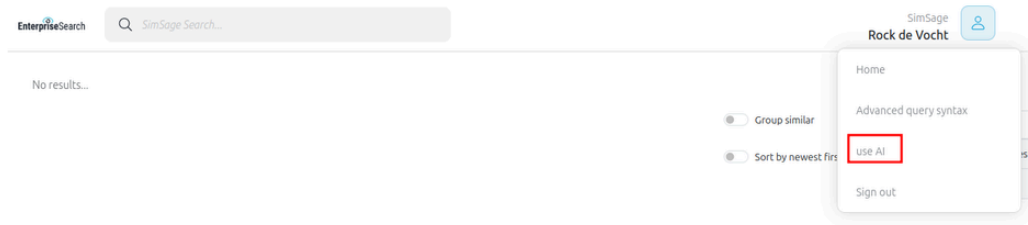
## SimSage shared settings

setting	description
document Q&A token count	the number of tokens passed, starting at the beginning of a document for a user asking questions interactively of a document.
document summarization token count	the number of tokens passed, starting at the beginning of a document for document summarization.
search box question answering, the number of results to pass to the LLM	Used for question answering while searching. This is the number of results to pass to the LLM, starting with the top result.

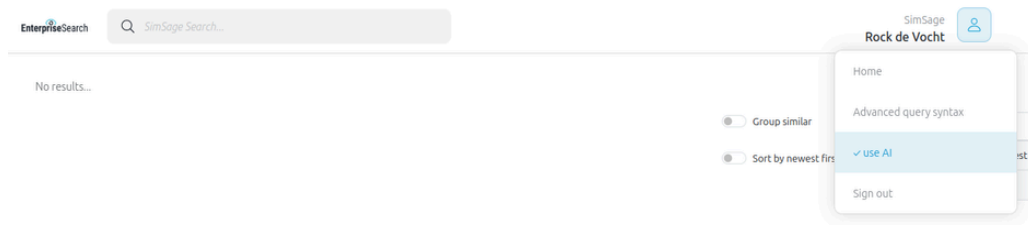
search box question answering, the number of tokens to pass for each search result	The number of tokens to pass for each search result found, starting at the sentence of each successful search.
--	--

## Search UX

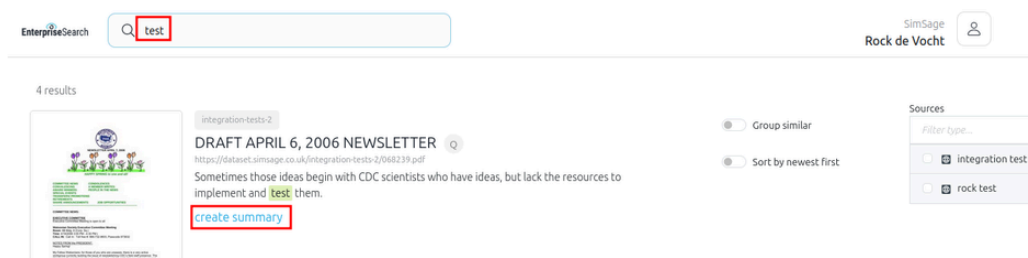
The default SimSage search interface will have extra features once a valid AI set up is set. An additional “use AI” is available (this value can be set to be checked by default by SimSage).



Make sure this value is “selected” by clicking on it.



Once selected, SimSage will automatically use the connected AI for question answering if possible (SimSage will detect if you are actually asking a question or are just using keywords).



A user searches for “test” - this is a single keyword and cannot be phrased as a question, so the search box question answer system is not used. However, once a user enters a proper query, this changes. Note the “create summary” button. This button is part of the SimSage AI system and provides a short summary for the given document.

EnterpriseSearch

SimSage  
Rock de Vocht

one result

ideas begin with CDC scientists who lack resources to implement and test them.

☐ Group similar  
☐ Sort by newest first

Sources

Filter type...

☐ integration test 2

☒ rock test

integration-tests-2

**DRAFT APRIL 6, 2006 NEWSLETTER**

<https://dataset.simsage.co.uk/integration-tests-2/068239.pdf>

Sometimes those **ideas** begin with **CDC scientists** who have **ideas**, but lack the resources to **implement** and **test** them.

[create summary](#)

The search query “what ideas begin with CDC scientists?” does look like a question. This activates SimSage’s search box query answer system and responds accordingly.

EnterpriseSearch

SimSage  
Rock de Vocht

4 results

ideas begin with CDC scientists who lack resources to implement and test them.

☐ Group similar  
☐ Sort by newest first

Sources

Filter type...

☐ integration test 2

☒ rock test

integration-tests-2

**DRAFT APRIL 6, 2006 NEWSLETTER**

<https://dataset.simsage.co.uk/integration-tests-2/068239.pdf>

Sometimes those ideas begin with CDC scientists who have ideas, but lack the resources to implement and **test** them.

[create summary](#)

integration-tests-2

**1902.09243**

<https://dataset.simsage.co.uk/integration-tests-2/1902.09243.pdf>

$L_{model} = L_{dec} + L_{refine}$  (15) At **test** time, each time step we choose the predicted word.

[create summary](#)

Select the “Q” icon after each search result to enter document Question and Answering mode for that specific document.

EnterpriseSearch

SimSage  
Rock de Vocht

4 results

ideas begin with CDC scientists who lack resources to implement and test them.

☐ Group similar  
☐ Sort by newest first

Sources

Filter type...

☐ integration test 2

☒ rock test

integration-tests-2

**DRAFT APRIL 6, 2006 NEWSLETTER**

<https://dataset.simsage.co.uk/integration-tests-2/068239.pdf>

Sometimes those ideas begin with CDC scientists who have ideas, but implement and **test** them.

[create summary](#)

integration-tests-2

**1902.09243**

<https://dataset.simsage.co.uk/integration-tests-2/1902.09243.pdf>

$L_{model} = L_{dec} + L_{refine}$  (15) At **test** time, each time step we choose

[create summary](#)

integration-tests-2

**Safety Seats**

[https://dataset.simsage.co.uk/integration-tests-2/safety\\_seats.txt](https://dataset.simsage.co.uk/integration-tests-2/safety_seats.txt)

**Test** the child seat to ensure less than 1° of movement by pulling the the front of the car.

[create summary](#)

**Document Questions**

Please ask me any question about "DRAFT APRIL 6, 2006 NEWSLETTER" (<https://dataset.simsage.co.uk/integration-tests-2/068239.pdf>)

what is this document all about?

The document is a newsletter providing updates on committee news, award winners, people in the news, special events, transfers/promotions, retirements, job opportunities, and a message from the president regarding the reestablishment of CDC's field staff presence and support for current field staff.

what people are talked about in this document?

We can now start a conversation with this document as shown above.

