



## BOOTCAMP

Python Aplicado à Análise de Dados e  
IA para o Setor Elétrico Brasileiro



Universidade Federal do  
Rio Grande do Norte  
IEEE Student Branch



# LLMs Powered by Retrieval-Augmented Generation (RAG)

THAÍS DE ARAÚJO DE MEDEIROS

[thais.araujo.707@ufrn.edu.br](mailto:thais.araujo.707@ufrn.edu.br)



# WHO I AM



**Thaís Medeiros**

Technician in Textiles - **IFRN**

Bachelor of Science and Technology - **UFRN**

Computer Engineer - **UFRN**

Master's Degree and PhD Candidate - **PPgEEC/UFRN**

Member of the **Conect2ai** Research Group | Projects: **CNPq** and **Rota2030**

# AGENDA



Introduction



System Building



Architecture & Key  
Components



Hands-On



# IA generativa: como estamos usando para potencializar nossa energia

*Você sabia que a inteligência artificial generativa já é usada no setor de energia? Conheça mais sobre o Projeto Cortex e como estamos fazendo isso.*

Atualizado em 18/07/2025 Postado em 18/07/2025



Energy and Buildings  
Volume 324, 1 December 2024, 114827



## Driving sustainable energy transitions with a multi-source RAG-LLM system

Muhammad Arslan <sup>a</sup>  , Lamine Mahdjoubi <sup>a</sup>, Saba Munawar <sup>b</sup>

Consumidor Moderno / CONAREC 2024 /  
Resultados reais da IA generativa no atendimento: o case Cemig

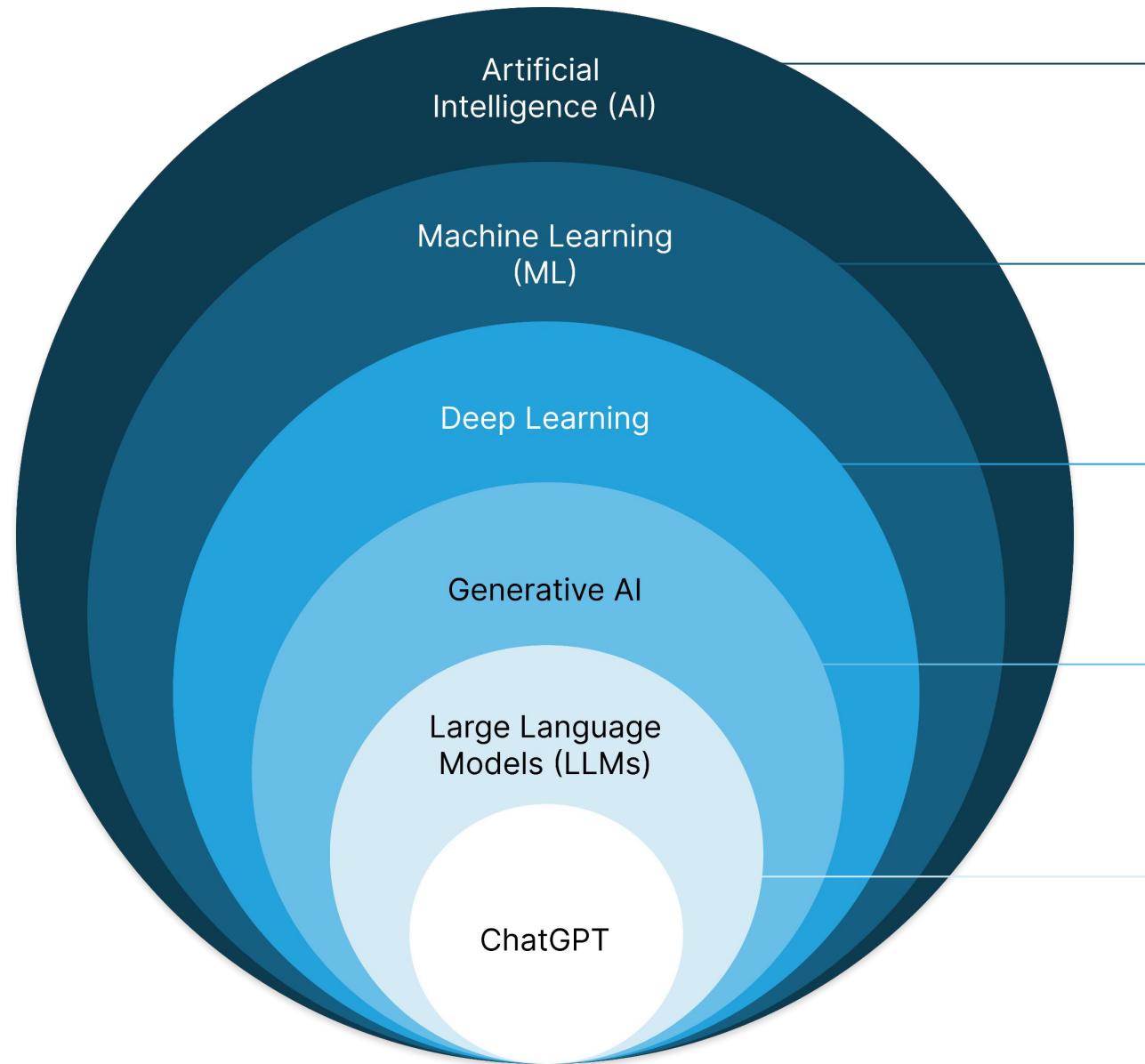
## Resultados reais da IA generativa no atendimento: o case Cemig

IBM e Cemig destacam a importância de uma implementação estratégica da GenAI para resultados efetivos e uma experiência do cliente aprimorada

Arthur Silva | Setembro 17, 2024



# OVERVIEW



Field focused on creating systems capable of performing tasks that typically require human intelligence

A subset of AI that develops algorithms able to learn patterns from data

A subfield of ML that uses neural networks with many layers to automatically learn representations

Techniques designed to create new data based on patterns learned from training data, rather than just making predictions

Generative AI models trained on massive text corpora, capable of understanding and producing human-like language across various tasks

Which Brazilian actress won the Golden Globe in 2025?



I don't know which Brazilian actress won the Golden Globe in 2025 because my training data goes only up to **June 2024**. Any events that happened after that date, including the 2025 ceremony, are outside of my knowledge. Without searching the web, I cannot provide a confirmed answer.



Which Brazilian actress won the Golden Globe in 2025?

Context:

Brazil woke up to a historic achievement: Fernanda Torres became the first Brazilian actress to win the Golden Globe in 2025. She was awarded for her performance in *I'm Still Here (Ainda Estou Aqui)*, Globoplay's first original film, in the category of Best Actress in a Motion Picture – Drama.

At the ceremony in Hollywood, Walter Salles celebrated the recognition and highlighted the film's impact on Brazilian audiences. Based on the book by Marcelo Rubens Paiva, the feature film portrays the struggle of Eunice Paiva to find her husband, Rubens Paiva, who disappeared during the military dictatorship.

Fernanda Torres won the Golden Globe in 2025, becoming the first Brazilian actress to receive the award. She earned it for her performance in *I'm Still Here (Ainda Estou Aqui)*, Globoplay's first original film, in the category of Best Actress in a Motion Picture – Drama.



---

# Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks

---

Patrick Lewis<sup>†‡</sup>, Ethan Perez<sup>\*</sup>,

Aleksandra Piktus<sup>†</sup>, Fabio Petroni<sup>†</sup>, Vladimir Karpukhin<sup>†</sup>, Naman Goyal<sup>†</sup>, Heinrich Küttler<sup>†</sup>,

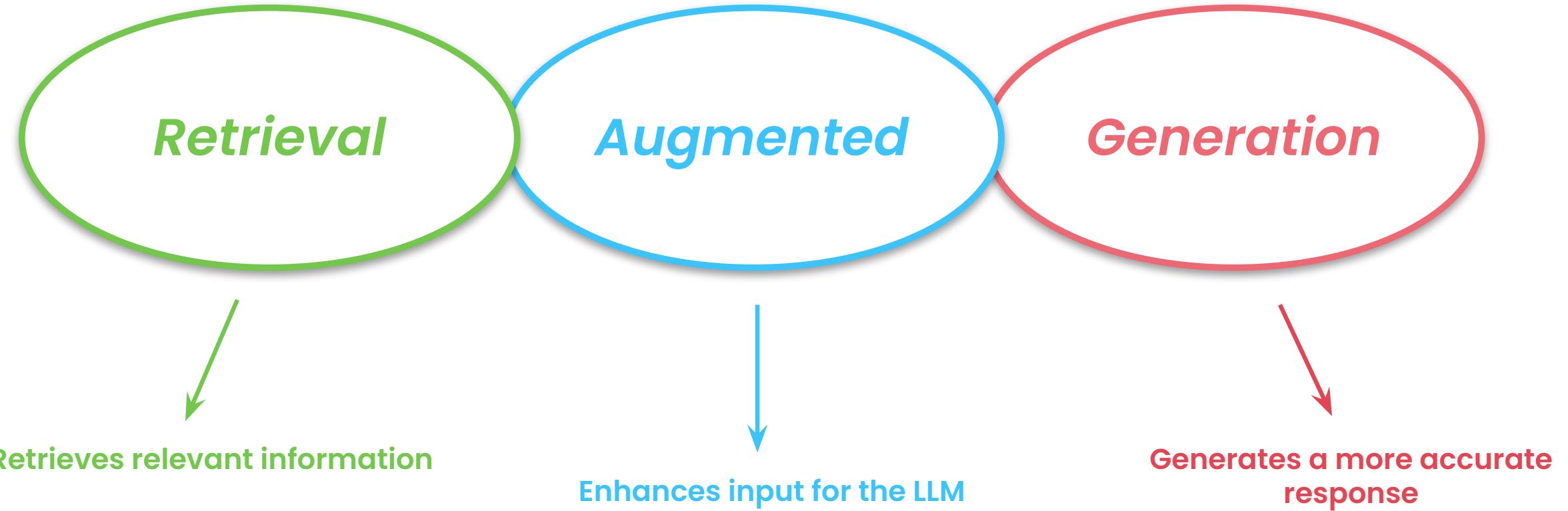
Mike Lewis<sup>†</sup>, Wen-tau Yih<sup>†</sup>, Tim Rocktäschel<sup>†‡</sup>, Sebastian Riedel<sup>†‡</sup>, Douwe Kiela<sup>†</sup>

<sup>†</sup>Facebook AI Research; <sup>‡</sup>University College London; <sup>\*</sup>New York University;  
plewis@fb.com

## Abstract

Large pre-trained language models have been shown to store factual knowledge in their parameters, and achieve state-of-the-art results when fine-tuned on downstream NLP tasks. However, their ability to access and precisely manipulate knowledge is still limited, and hence on knowledge-intensive tasks, their performance lags behind task-specific architectures. Additionally, providing provenance for their decisions and updating their world knowledge remain open research problems. Pre-trained models with a differentiable access mechanism to explicit non-parametric memory have so far been only investigated for extractive downstream tasks. We explore a general-purpose fine-tuning recipe for retrieval-augmented generation (RAG) — models which combine pre-trained parametric and non-parametric memory for language generation. We introduce RAG models where the parametric memory is a pre-trained seq2seq model and the non-parametric memory is a dense vector index of Wikipedia, accessed with a pre-trained neural retriever. We compare two RAG formulations, one which conditions on the same retrieved passages across the whole generated sequence, and another which can use different passages per token. We fine-tune and evaluate our models on a wide range of knowledge-intensive NLP tasks and set the state of the art on three open domain QA tasks, outperforming parametric seq2seq models and task-specific retrieve-and-extract architectures. For language generation tasks, we find that RAG models generate more specific, diverse and factual language than a state-of-the-art parametric-only seq2seq baseline.

# CONCEPT



# HOW TO BUILD?

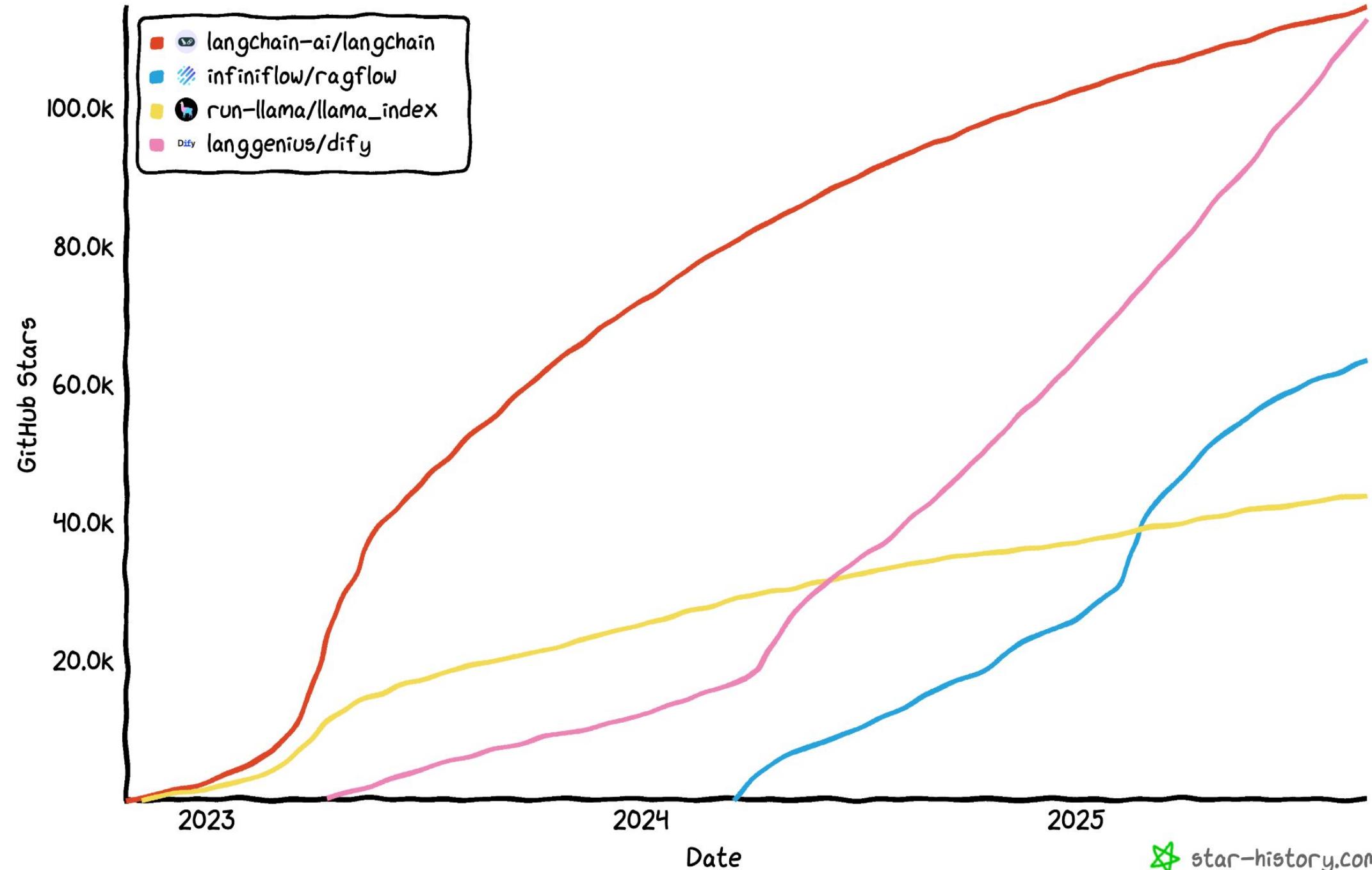


LlamaIndex

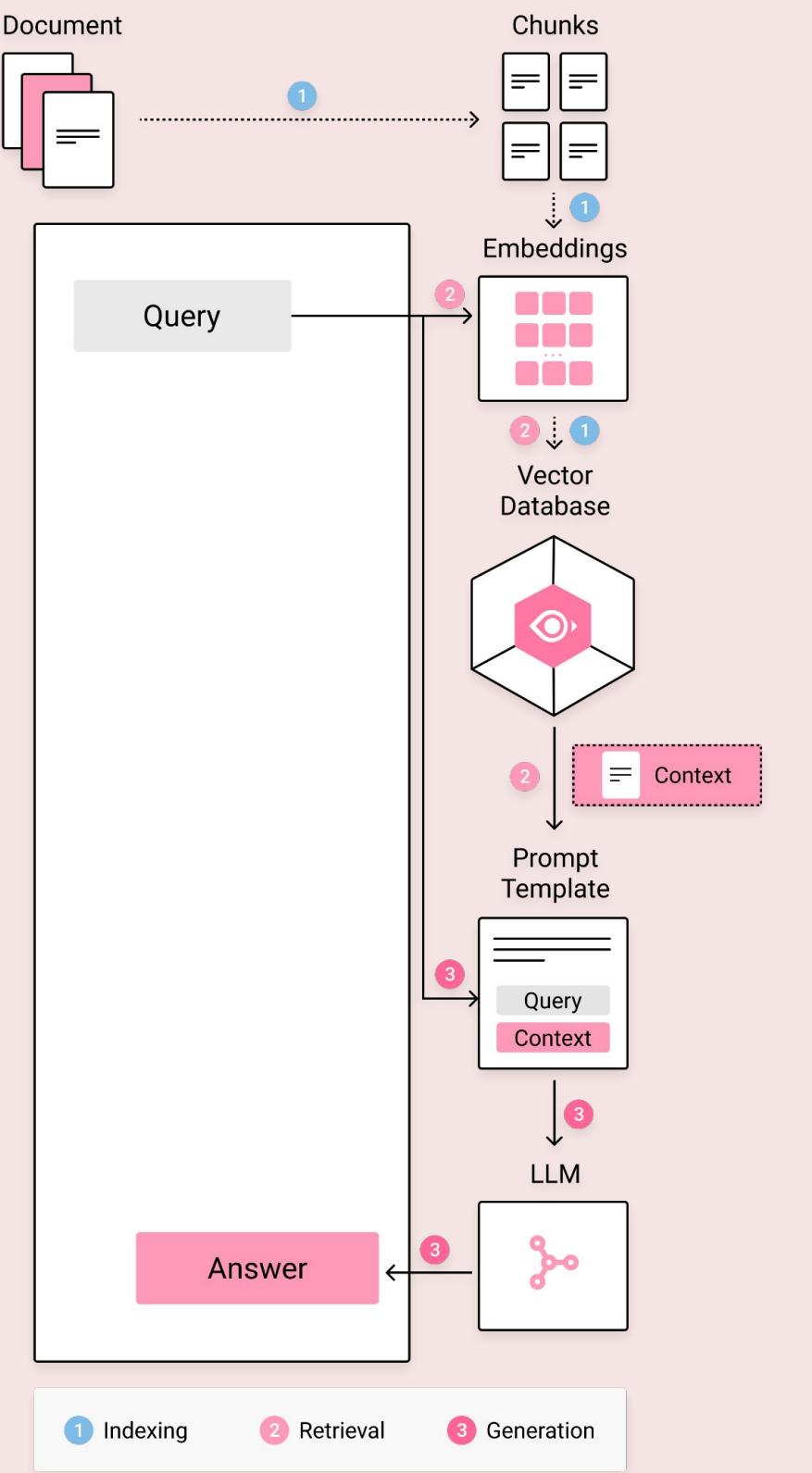


LangChain

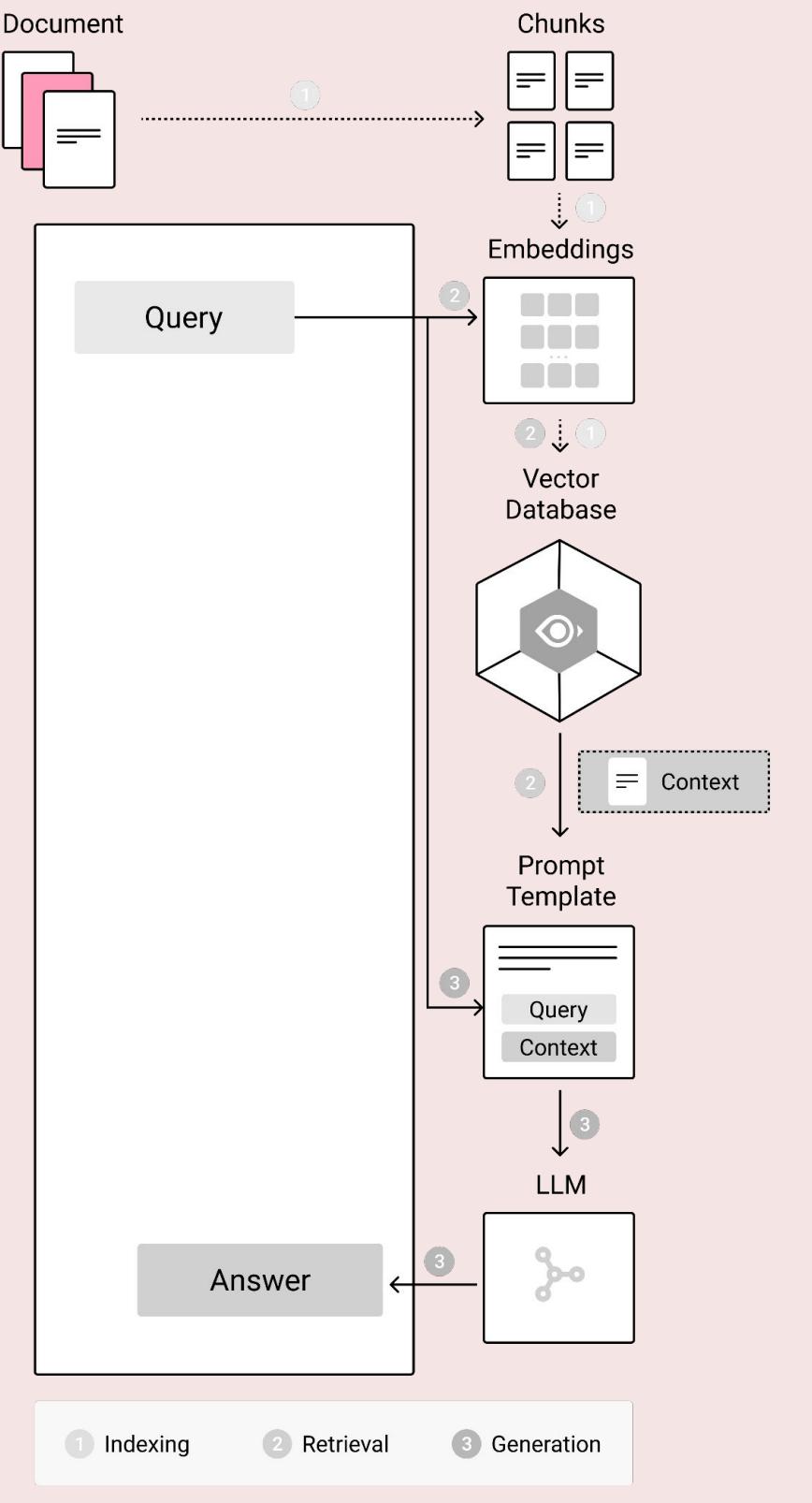
## Star History



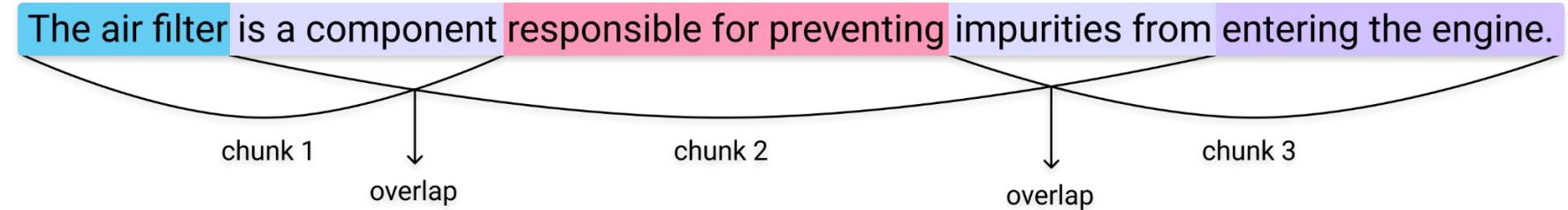
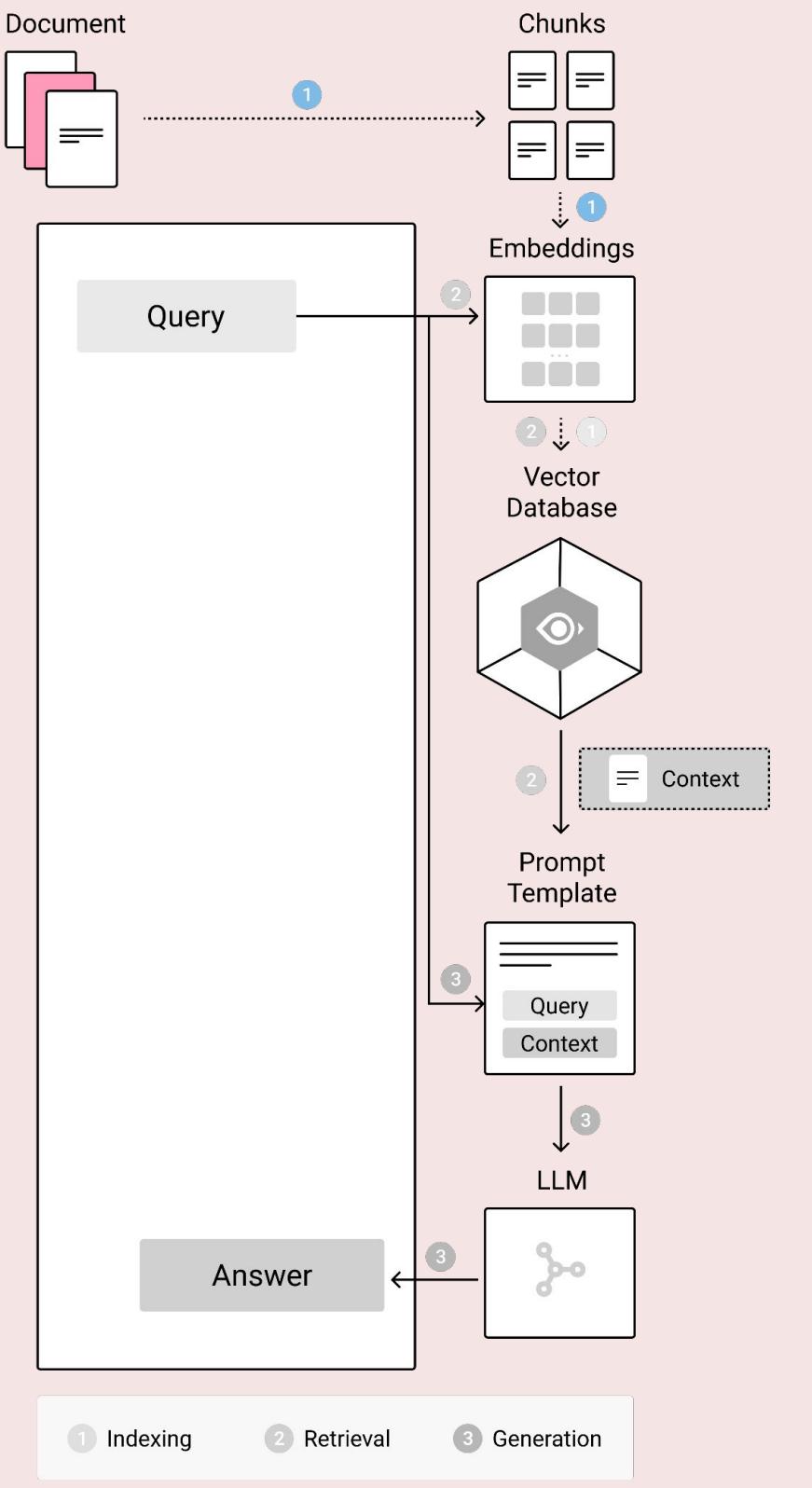
# A TYPICAL RAG SETUP



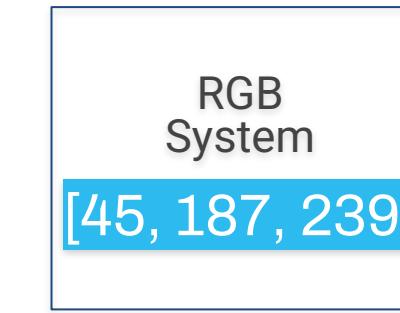
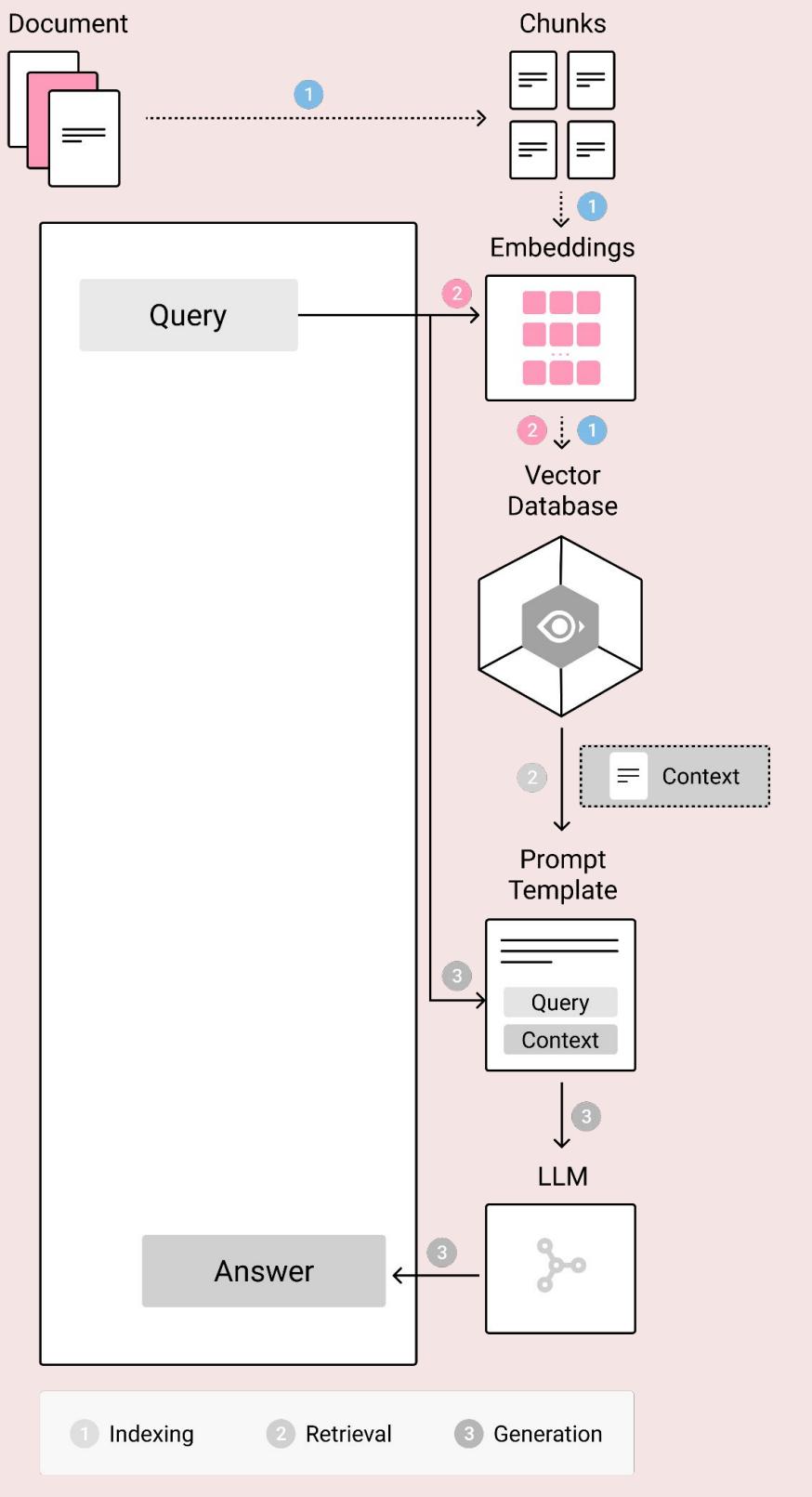
# TYPES OF DOCUMENTS



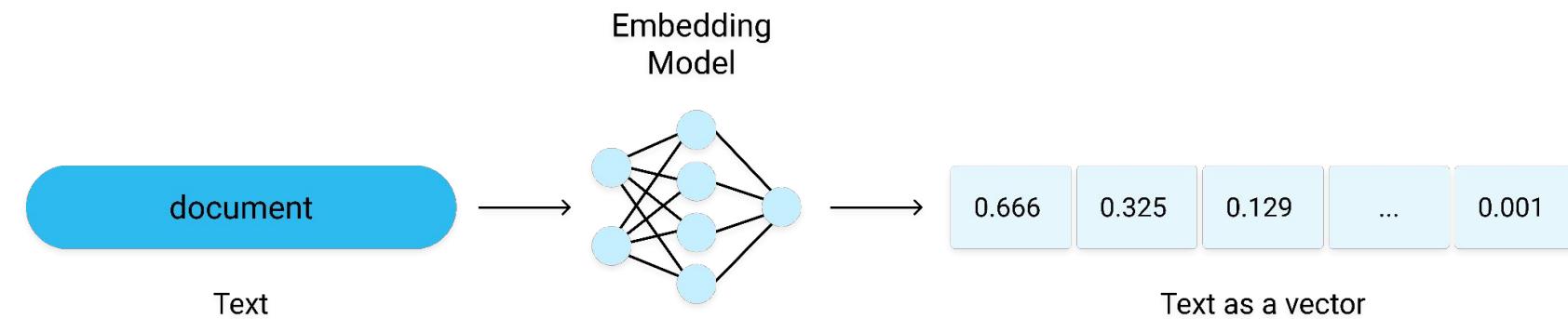
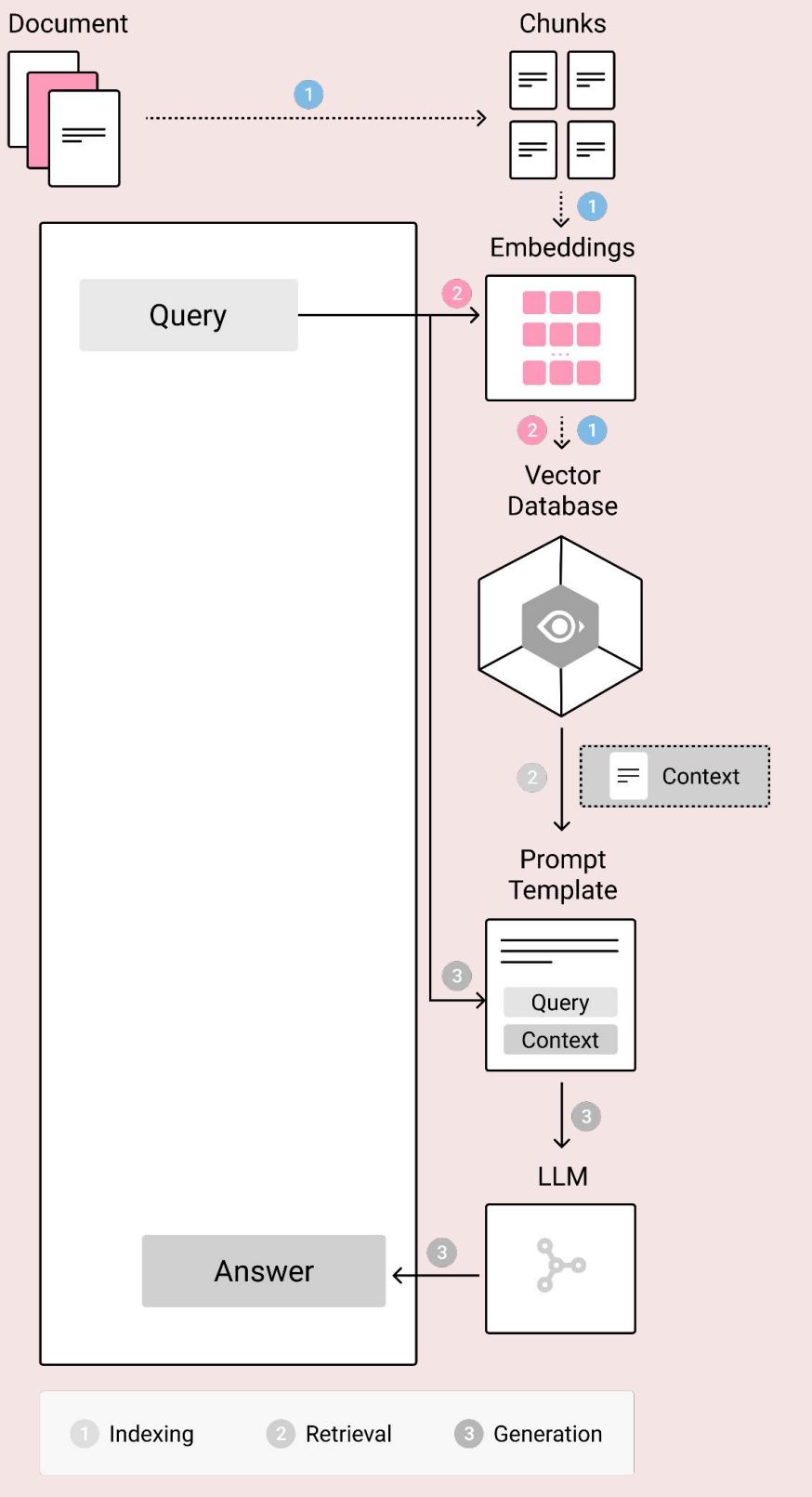
# CHUNKING



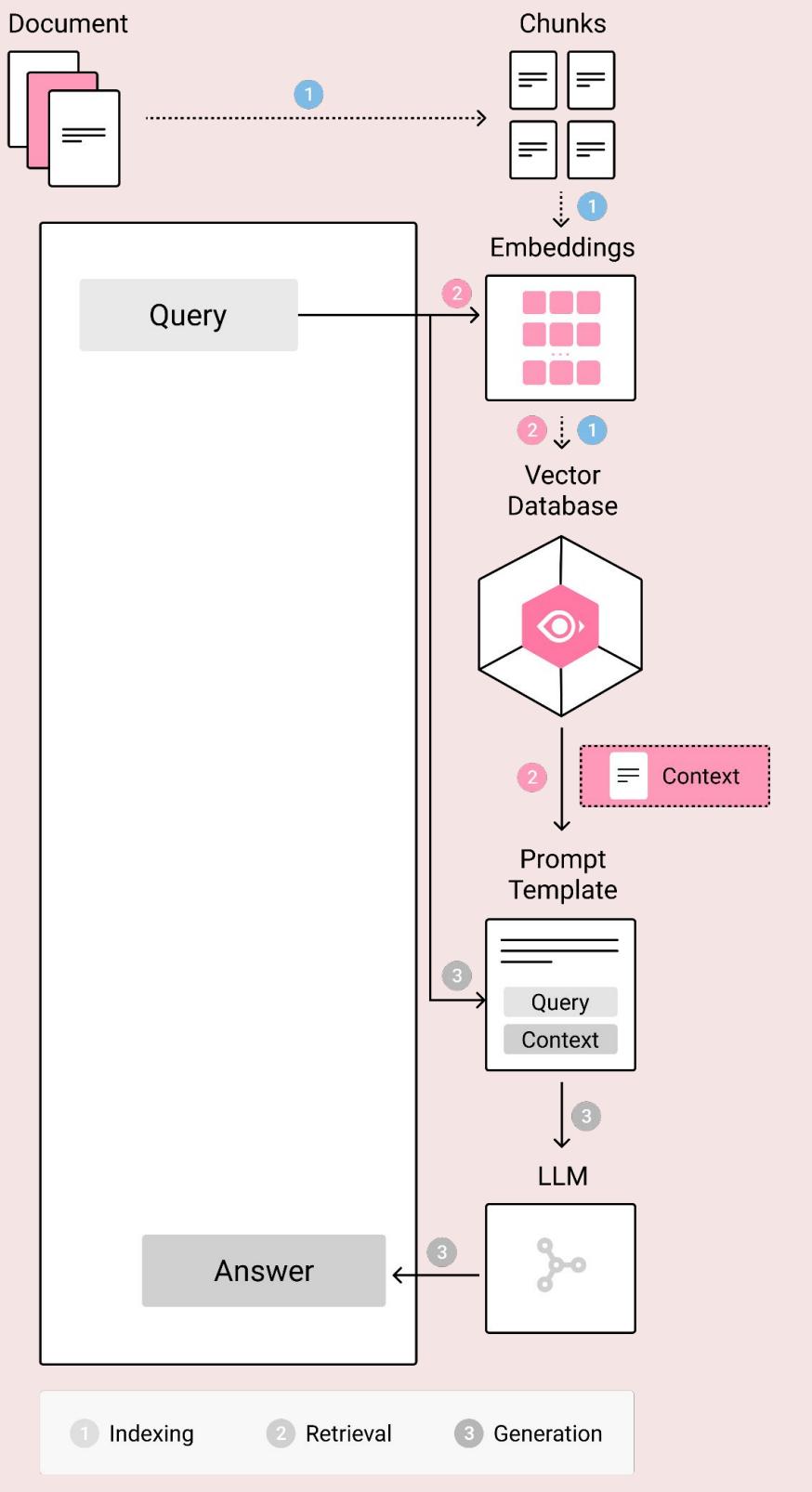
# EMBEDDING



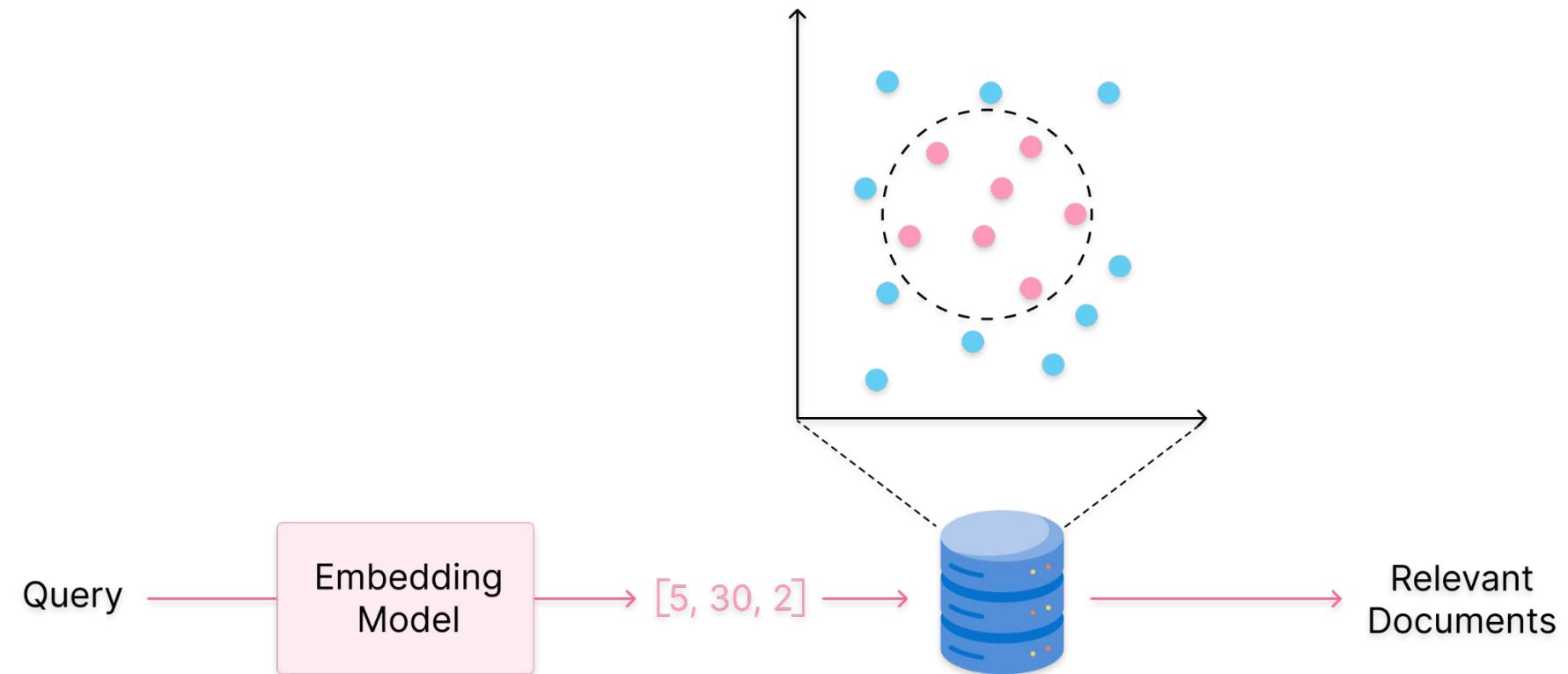
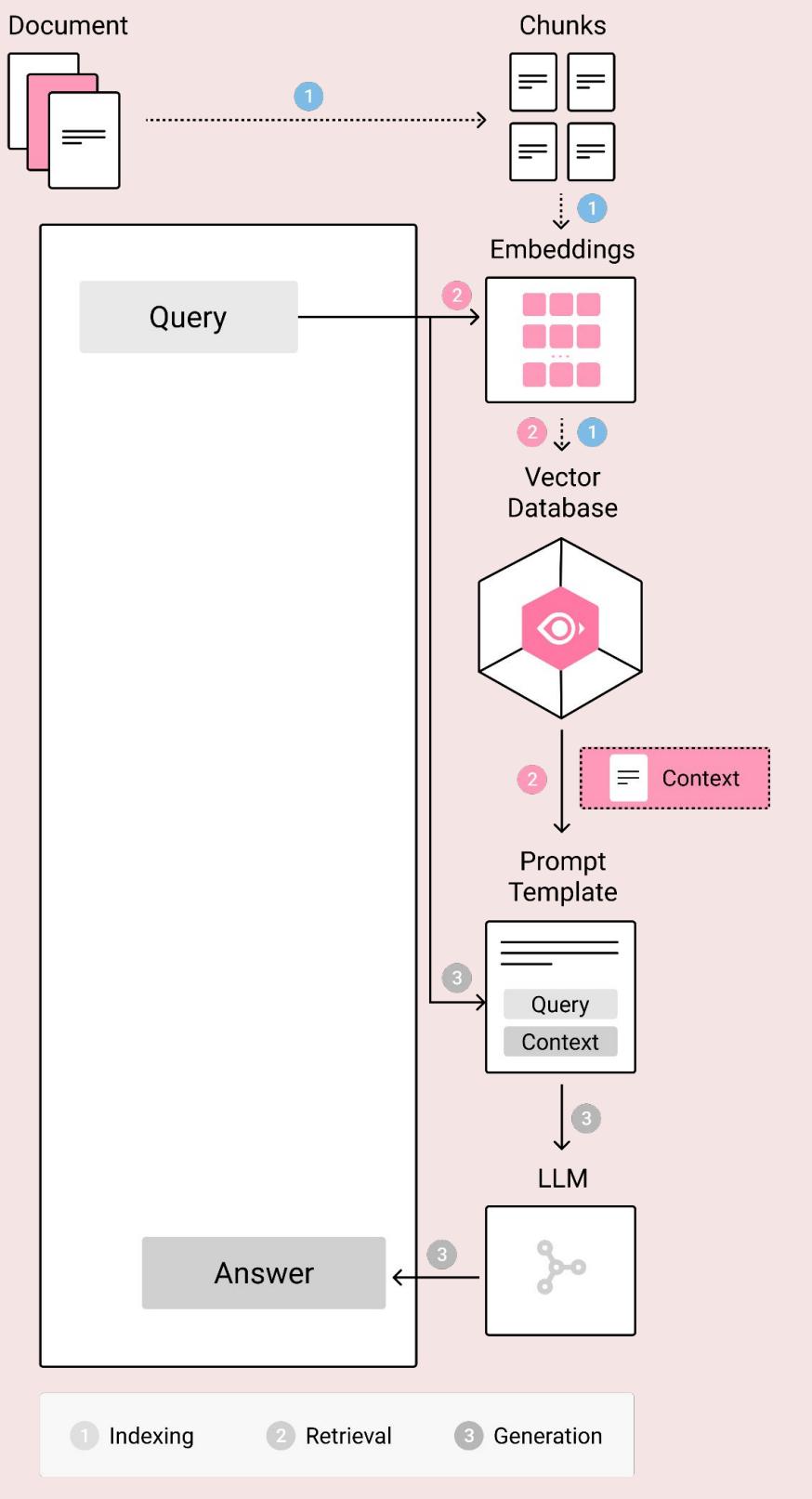
# EMBEDDING



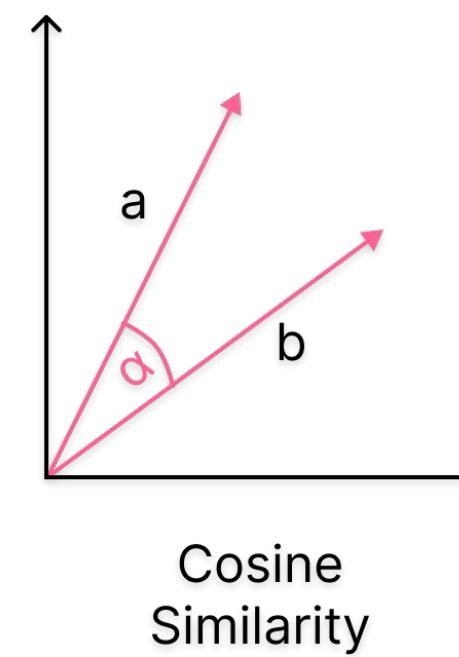
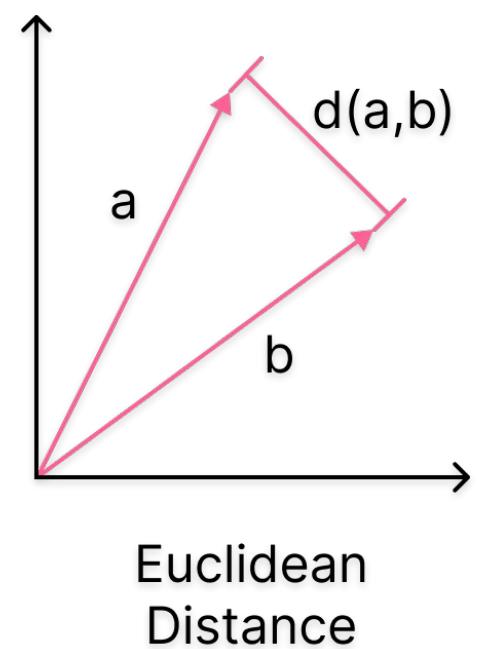
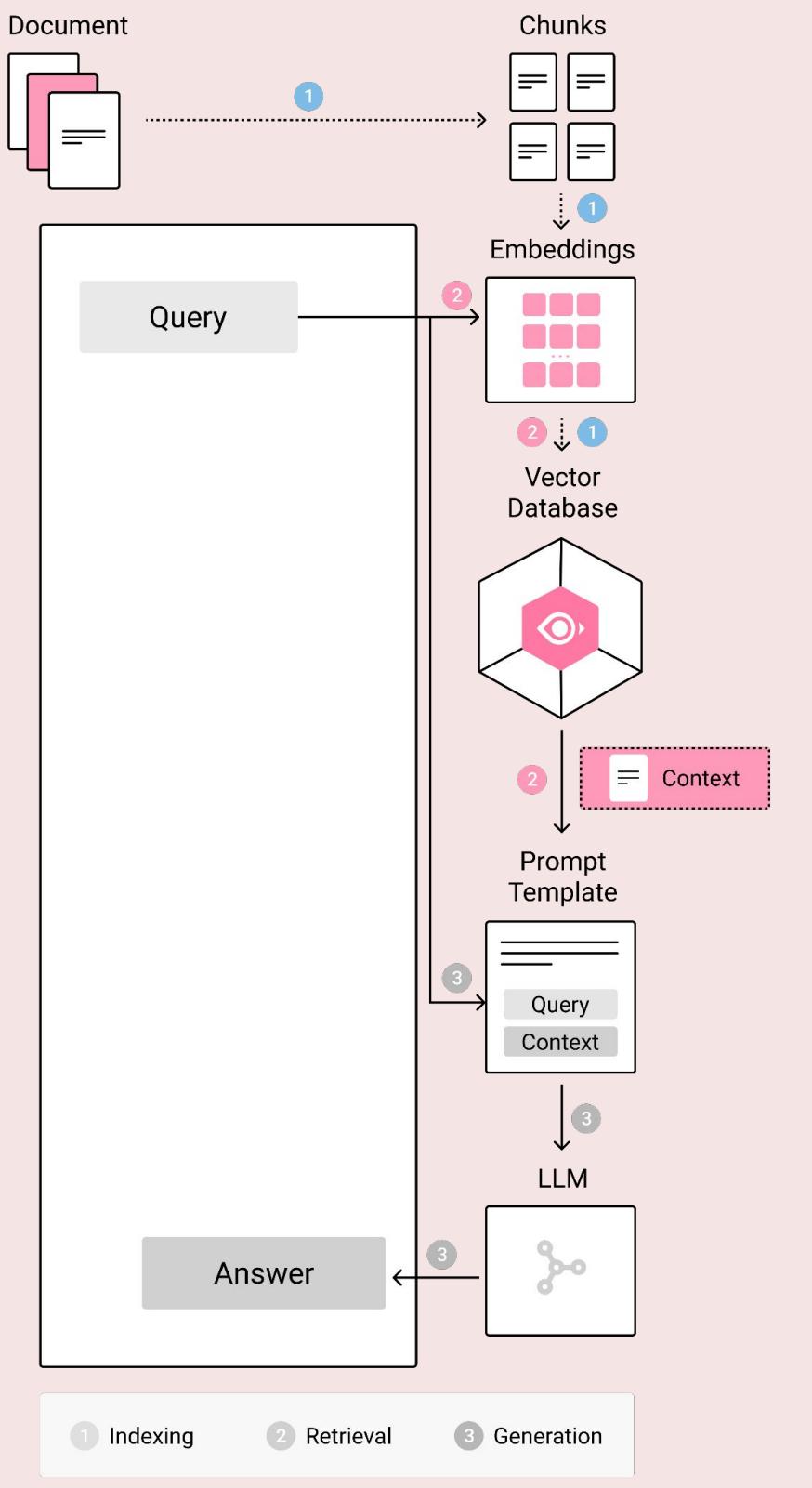
# VECTOR DATABASE



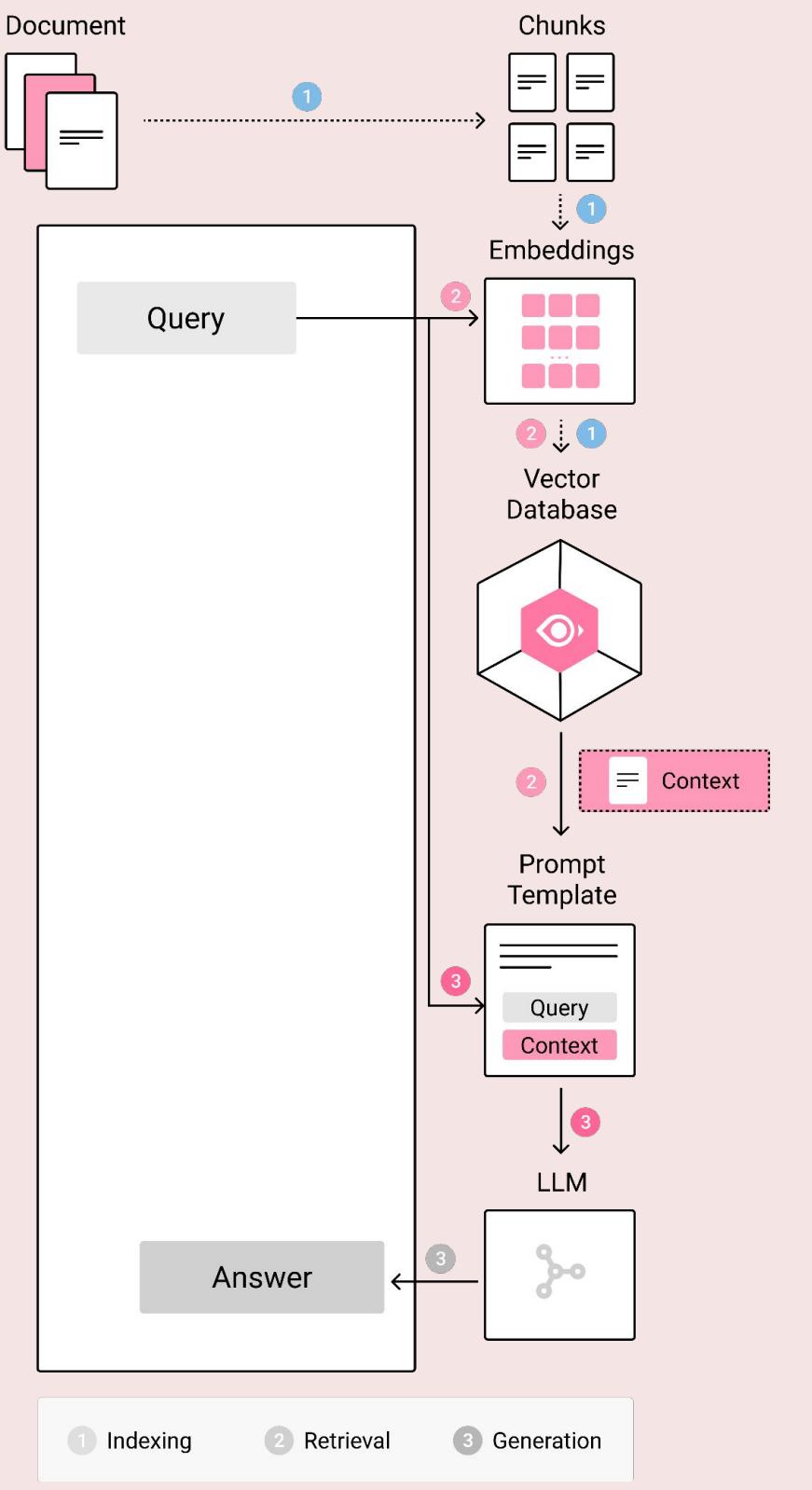
# VECTOR DATABASE

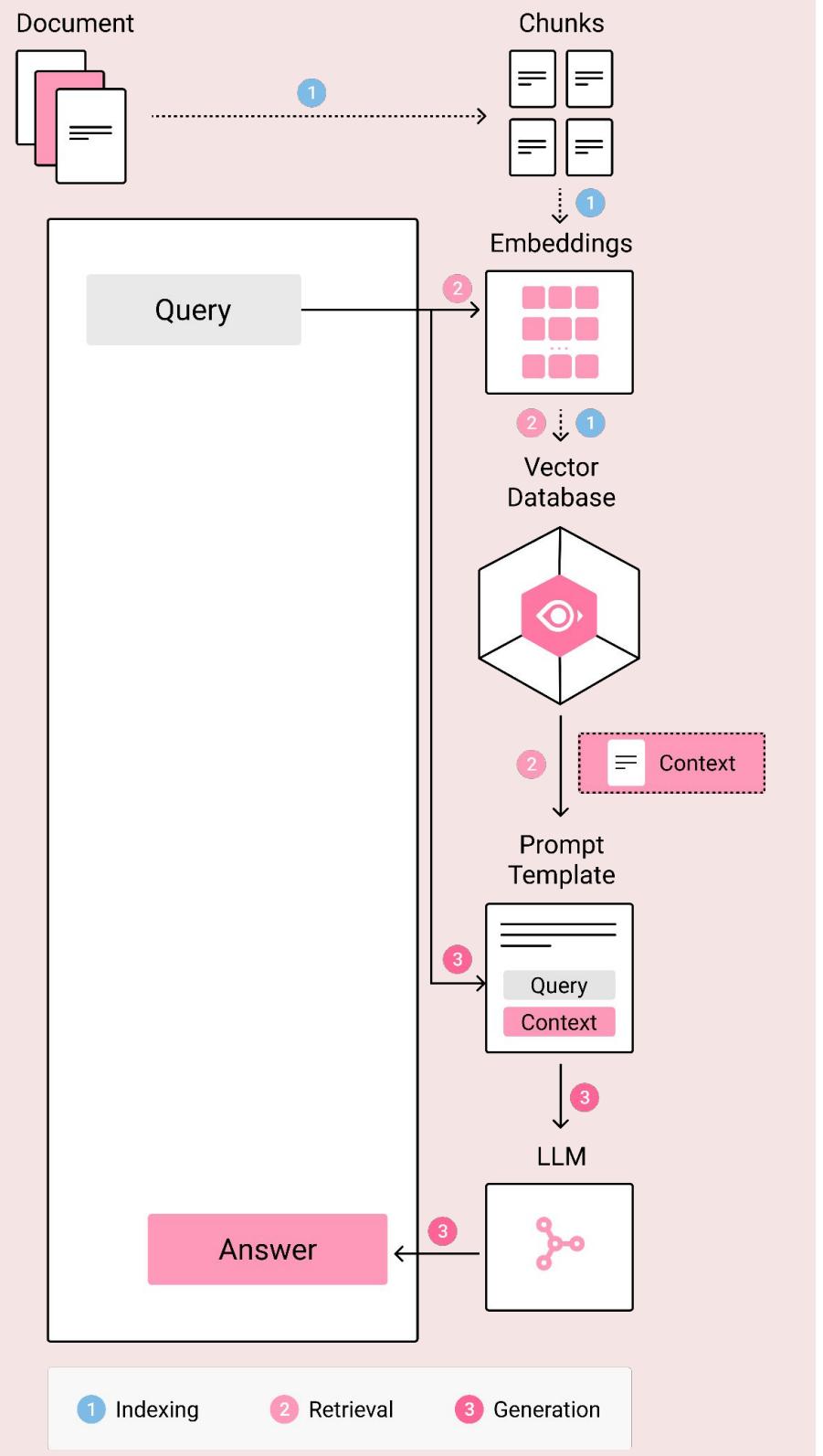


# VECTOR DATABASE

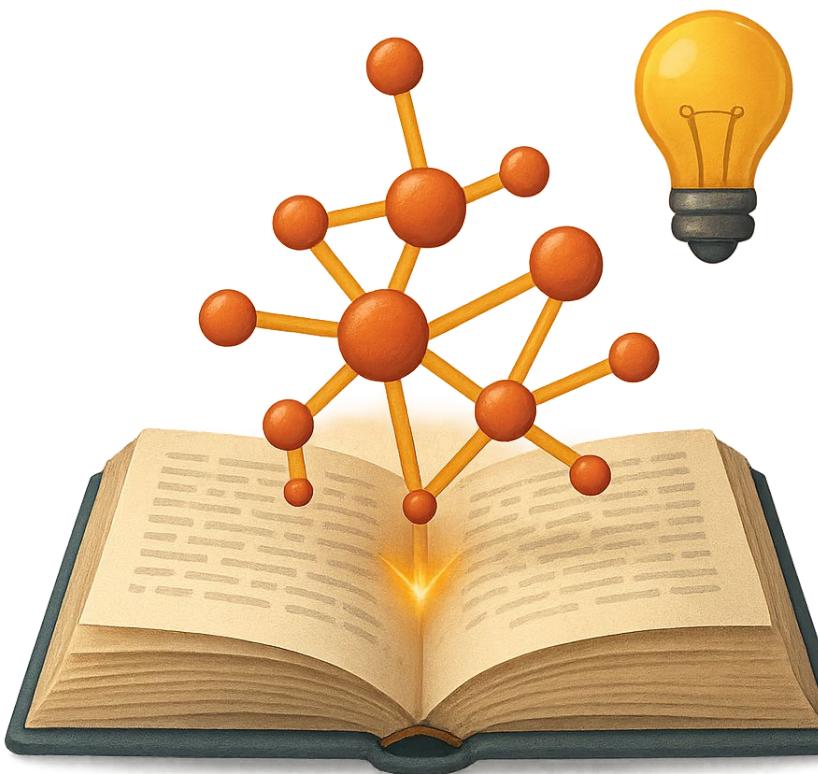


# PROMPT TEMPLATE

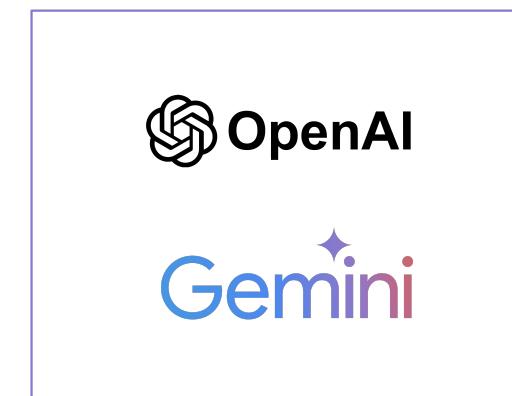
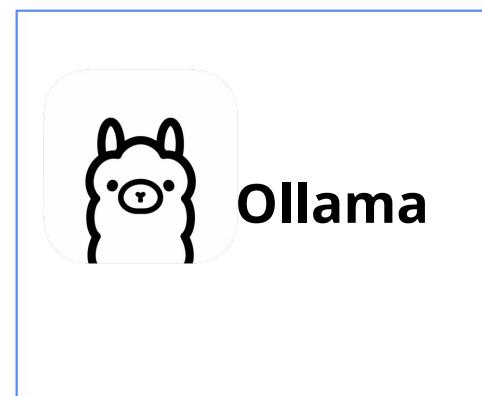
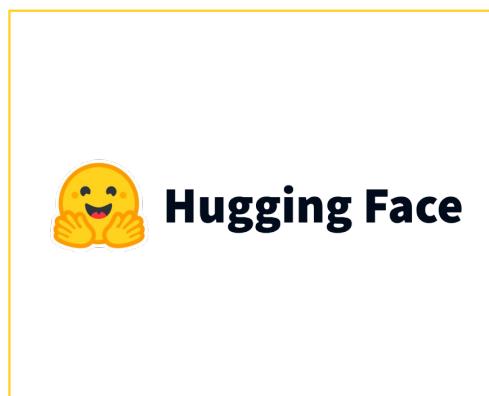
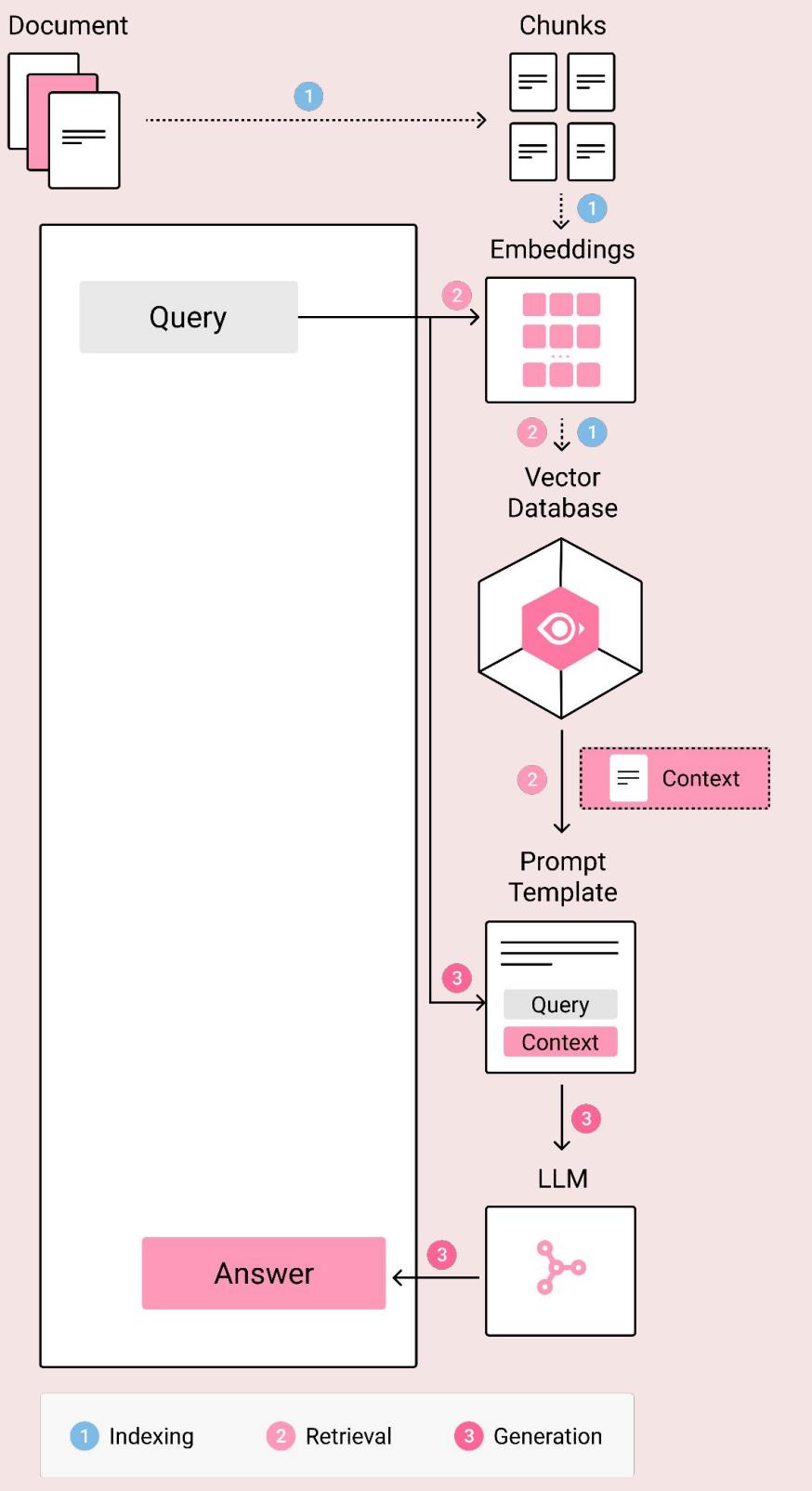




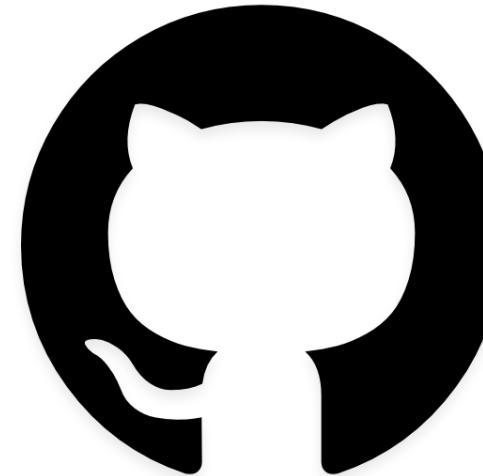
# LARGE LANGUAGE MODEL



# LARGE LANGUAGE MODEL



# REPOSITORY



<https://github.com/conect2ai/RAG-Bootcamp>

# TUTORIAL

<https://ai.google.dev/>

The screenshot shows the Google AI for Developers homepage. At the top, there's a navigation bar with links for Google AI for Developers, Models, Solutions, Code assistance, Showcase, and Community. On the right side of the header are search, language selection (Português), and other settings icons.

In the center-left, a large white text area reads "AI for every developer". Below it, a subtext says: "Unlock AI models to build innovative apps and transform development workflows with tools across platforms." A blue button at the bottom left of this area contains the text "Explore models in Google AI Studio" with a red arrow pointing to it from the left.

On the right side, there's a modal window titled "USER" and "MODELS". It shows a preview of an audio recording with a play button and the text "Split audio recording into segments and include transcriptions". The "MODELS" section lists "Gemini 2.5 Pro" (selected), "Gemini 2.5 Flash", "Gemini 2.0 Flash", "Veo 2", and "Imagen 3". At the bottom of the modal, it says "Running..." and has a "Stop" button.

At the bottom of the main page, the text "Start building" is displayed in a large white font.

<https://ai.google.dev/>

The screenshot shows the Google AI Studio interface at <https://ai.google.dev/>. A red arrow points from the left margin to the 'Get API key' button in the top right corner of the header.

**Google AI Studio**

**Chat prompt**

Chat Stream Generate media Build History Enable saving

Get API key Studio Dashboard Documentation

Run settings Gemini 2.5 Pro Token count 0 / 1.048.576 Temperature 1 Media resolution Default Thinking Thinking mode Set thinking budget Tools Structured output Edit Code execution Function calling Grounding with Google Search Source: Google Search URL context Advanced settings

Generate Python code for a simple calculator app → Run

What's new

- URL context tool Fetch information from web links
- Native speech generation Generate high quality text to speech with Gemini
- Live audio-to-audio dialog Try Gemini's natural, real-time dialog with audio and video inputs
- Native image generation Interleaved text-and-image generation with Gemini 2.0 Flash

Google AI models may make mistakes, so double-check outputs.

<https://ai.google.dev/>

Google AI Studio

Get API key Studio Dashboard Documentation ☰

API keys Usage & Billing Changelog

## Chaves de API

+ Criar chave de API

Testar rapidamente a API Gemini

Guia de início rápido da API

```
curl "https://generativelanguage.googleapis.com/v1beta/models/gemini-2.0-flash:generateContent" \
-H 'Content-Type: application/json' \
-H 'X-goog-api-key: GEMINI_API_KEY' \
-X POST \
-d '{
  "contents": [
    {
      "parts": [
        {
          "text": "Explain how AI works in a few words"
        }
      ]
    }
}'
```

Suas chaves de API estão listadas abaixo. Também é possível visualizar e gerenciar seu projeto e as chaves de API no Google Cloud.

Look up API Key for project

Número do projeto	Nome do projeto	Chave de API	Criação	Plano
Crie uma chave de API para ver seus projetos				

View status

Use as chaves de API com segurança. Não as compartilhe nem as incorpore em códigos públicos. O uso da API Gemini em um projeto com faturamento ativado está sujeito aos [preços do modelo de pagamento por uso](#).

<https://ai.google.dev/>

Google AI Studio

Get API key Studio Dashboard Documentation ☰

API keys Usage & Billing Changelog

## Chaves de API

Criar chave de API

Testar rapidamente a API Gemini

Guia de Início rápido da API

Code

```
curl "https://generativelanguage.googleapis.com/v1beta/models/gemini-2.8-flash:generateContent" \
-H "Content-Type: application/json" \
-H "X-goog-api-key: GEMINI_API_KEY" \
-X POST \
-d '{'
  "contents": [
    {
      "parts": [
        {
          "text": "Explain how AI works"
        }
      ]
    }
  ]
}'
```

**Criar chave de API**

Selecione um dos seus projetos atuais do Google Cloud

Search Google Cloud projects

🔍

↳ Criar uma chave de API em um projeto atual

Red arrow pointing to the search input field.

Suas chaves de API estão listadas abaixo. Também é possível visualizar e gerenciar seu projeto e as chaves de API no Google Cloud.

Look up API Key for project

Número do projeto	Nome do projeto	Chave de API	criado	Ação

Crie uma chave de API para ver seus projetos

View status

Use as chaves de API com segurança. Não as compartilhe nem as incorpore em códigos públicos. O uso da API Gemini em um projeto com faturamento ativado está sujeito aos [preços do modelo de pagamento por uso](#).

<https://ai.google.dev/>

Google AI Studio

Get API key Studio Dashboard Documentation

Chaves de API

Criar chave de API

Testar rapidamente a API Gemini

Guia de Início rápido da API

Code

```
curl "https://generativelanguage.googleapis.com/v1beta1/models/gemini-2.8-flash:generateContent" \
-H "Content-Type: application/json" \
-H "X-goog-api-key: GEMINI_API_KEY" \
-X POST \
-d "{\n  \"contents\": {\n    \"parts\": [\n      {\n        \"text\": \"Explain how AI works\"}\n    ]\n  }\n}"
```

Criar chave de API

Selezione um dos seus projetos atuais do Google Cloud

Search Google Cloud projects

Gemini API gen-lang-client-0581720549

Suas chaves de API estão listadas abaixo. Também é possível visualizar e gerenciar seu projeto e as chaves de API no Google Cloud.

Look up API Key for project

Número do projeto	Nome do projeto	Chave de API	criado	Plano

Crie uma chave de API para ver seus projetos

View status

Use as chaves de API com segurança. Não as compartilhe nem as incorpore em códigos públicos. O uso da API Gemini em um projeto com faturamento ativado está sujeito aos [preços do modelo de pagamento por uso](#).

<https://ai.google.dev/>

Google AI Studio

Get API key Studio Dashboard Documentation

On API keys Usage & Billing Changelog

## Chaves de API

Criar chave de API

Testar rapidamente a API Gemini

Guia de Início rápido da API

Code

```
curl "https://generativelanguage.googleapis.com/v1beta1/models/gemini-2.8-flash:generateContent" \
-H "Content-Type: application/json" \
-H "X-goog-api-key: GEMINI_API_KEY" \
-X POST \
-d '{'
  "contents": [
    {
      "parts": [
        {
          "text": "Explain how AI works"
        }
      ]
    }
  ]
}'
```

Criar chave de API

Selecionar projeto

Search Google Cloud projects

Gemini API (gen-lang-client-0581720549)

Criar uma chave de API em um projeto atual

Suas chaves de API estão listadas abaixo. Também é possível visualizar e gerenciar seu projeto e as chaves de API no Google Cloud.

Look up API Key for project

Número do projeto	Nome do projeto	Chave de API	criado	Plano

Crie uma chave de API para ver seus projetos

View status

Use as chaves de API com segurança. Não as compartilhe nem as incorpore em códigos públicos. O uso da API Gemini em um projeto com faturamento ativado está sujeito aos [preços do modelo de pagamento por uso](#).

<https://ai.google.dev/>

Google AI Studio

Get API key Studio Dashboard Documentation

On API keys Usage & Billing Changelog

## Chaves de API

+ Criar chave de API

Testar rapidamente a API Gemini

Guia de Início rápido da API

Code

```
curl "https://generativelanguage.googleapis.com/v1beta/models/gemini-2.8-flash:generateContent"  
-H "Content-Type: application/json"  
-H "X-goog-api-key: GEMINI_API_KEY"  
-X POST  
-d "{  
  \"contents\": {  
    \"parts\": [  
      {  
        \"text\": ""  
      }  
    ]  
  }  
}"
```

**Chave de API gerada**

Use suas chaves de API com segurança. Não compartilhe nem as incorpore em códigos que o público possa ver.

AlzaSyAMaONxG9FojqVGFNj9X3T9\_pv-vzErmTE  

Suas chaves de API estão listadas abaixo. Também é possível visualizar e gerenciar seu projeto e as chaves de API no Google Cloud.

Look up API Key for project

Número do projeto	Nome do projeto	Chave de API	Création	Promo
...5456	Gemini API	...mTE	25 de ago. de 2025	<a href="#">Configurar o faturamento</a> <a href="#">Ver dados de uso</a>

 View status

Use as chaves de API com segurança. Não as compartilhe nem as incorpore em códigos públicos. O uso da API Gemini em um projeto com faturamento ativado está sujeito aos [preços do modelo de pagamento por uso](#).



# HANDS-ON