

Large Language Models

CURRENT STATE

Rodrigo Gonzalez, PhD

(AI)

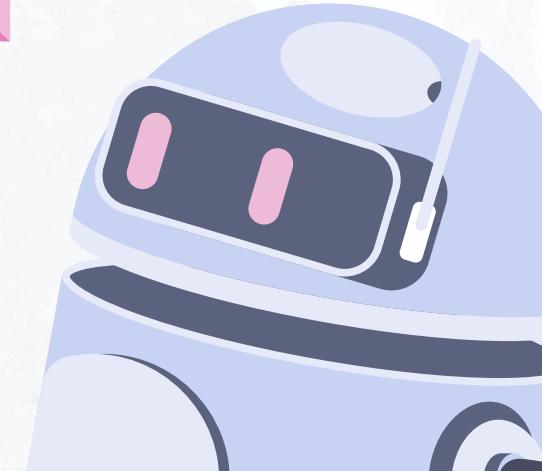


Table of contents

01 → What is Large Language Model?

02 → LLMs, a very brief technical history

03 → Prompt engineering

04 → LLM fine-tuning

05 → LLM main use cases

06 → Using LLM in applications

01 →

What is a Large Language Model (LLM)?

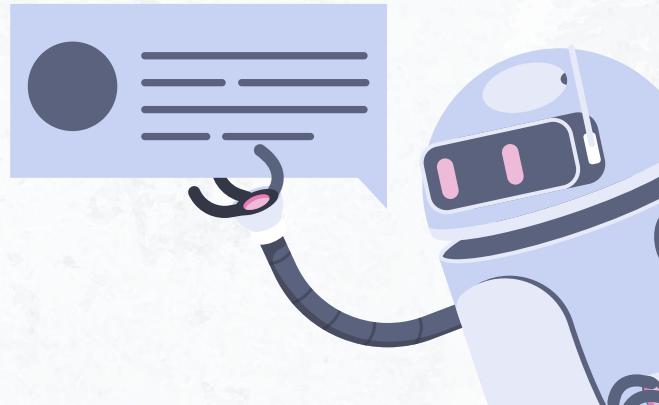
(AI)

Large Language Model

A large language model (LLM) is a language model consisting of a neural network with many parameters (**billions**) trained with immense amounts of texts using self-supervised learning.

LLMs are probabilistic models that attempt to map the probability of a sequence of words, given the surrounding context.

We don't have to speak the language of computers anymore
they can speak ours!



How does LLM work?



No nos une el amor ...



How does LLM work?

No nos une el amor **sino**



How does LLM work?



No nos une el amor sino ...



How does LLM work?

No nos une el amor sino **el**



How does LLM work?



No nos une el amor sino el ...



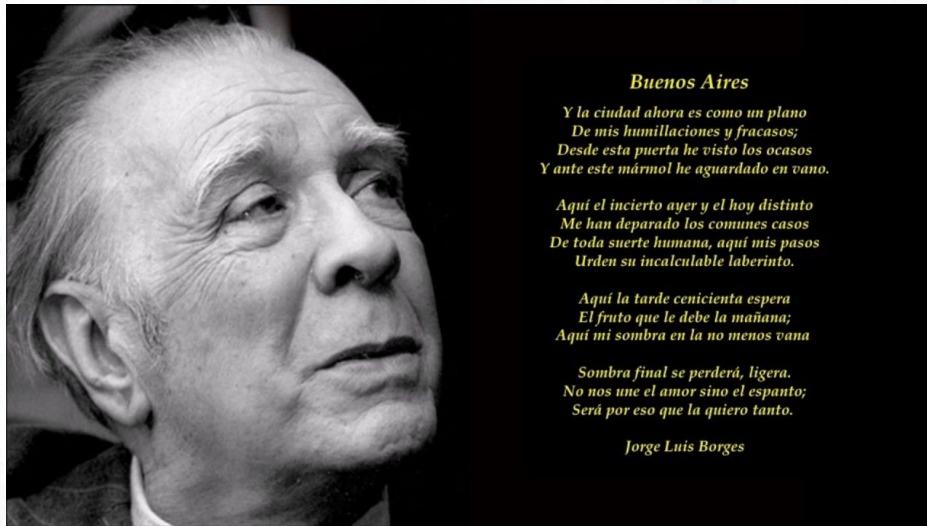
How does LLM work?

No nos une el amor sino el **espanto**



How does LLM work?

No nos une el amor sino el **espanto**



How does LLM work?

No nos une el amor sino el **odio**



02 →

LLMs, a very brief technical history

(AI)

LLMs, a very brief technical history

- 1 2013, NLP, Embeddings (Word2Vec, Glove)
- 2 2017, Transformers, Attention is all you need
- 3 2018, GPT (117M params)
- 4 2019, GPT-2 (1.5B)
- 5 2020, GPT-3 (175B)
- 6 2022, Chat GPT-3.5, more training (175B)

LLMs, a very brief technical history

7 Feb 2023, LLaMa (Meta), open source (70B)

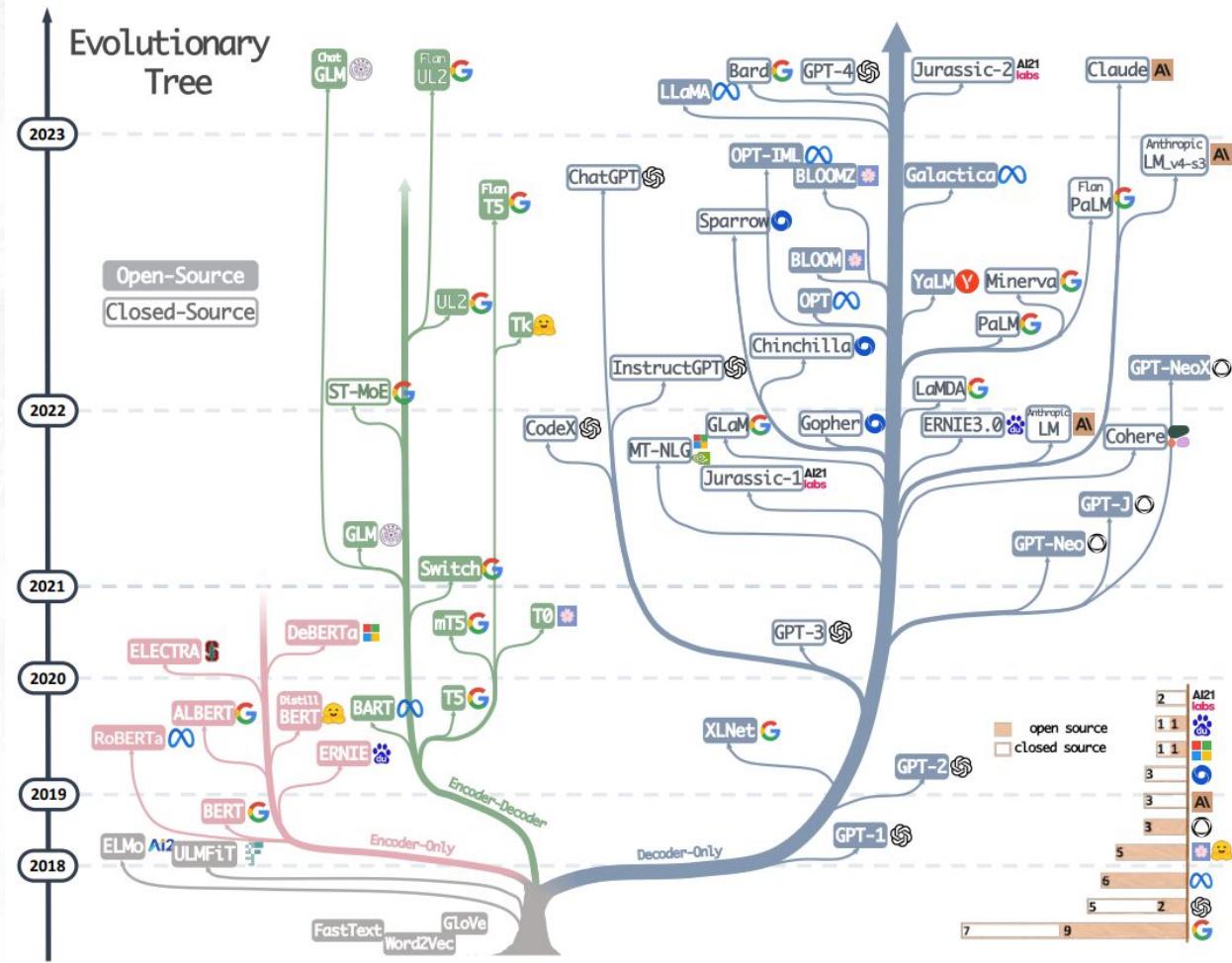
8 March 2023, GPT-4, (1T?)

9 March 2023, Bard (Google) (137B)

10 May 2023, QLoRa

11 May 2023, Falcon, open source (40B)

12 July 2023, Llama 2 (Meta), open source (70B)

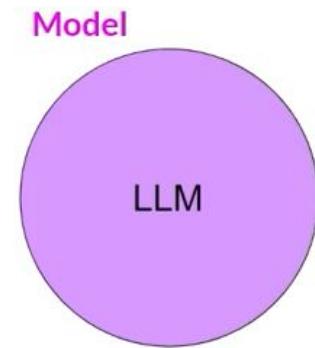


03 →

Prompt engineering

(AI)

Prompts and completions

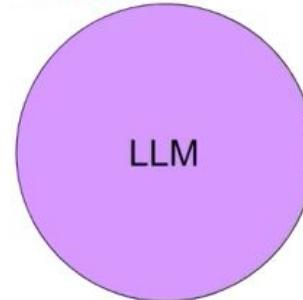


Prompts and completions

Prompt

Where is Ganymede
located in the solar
system?

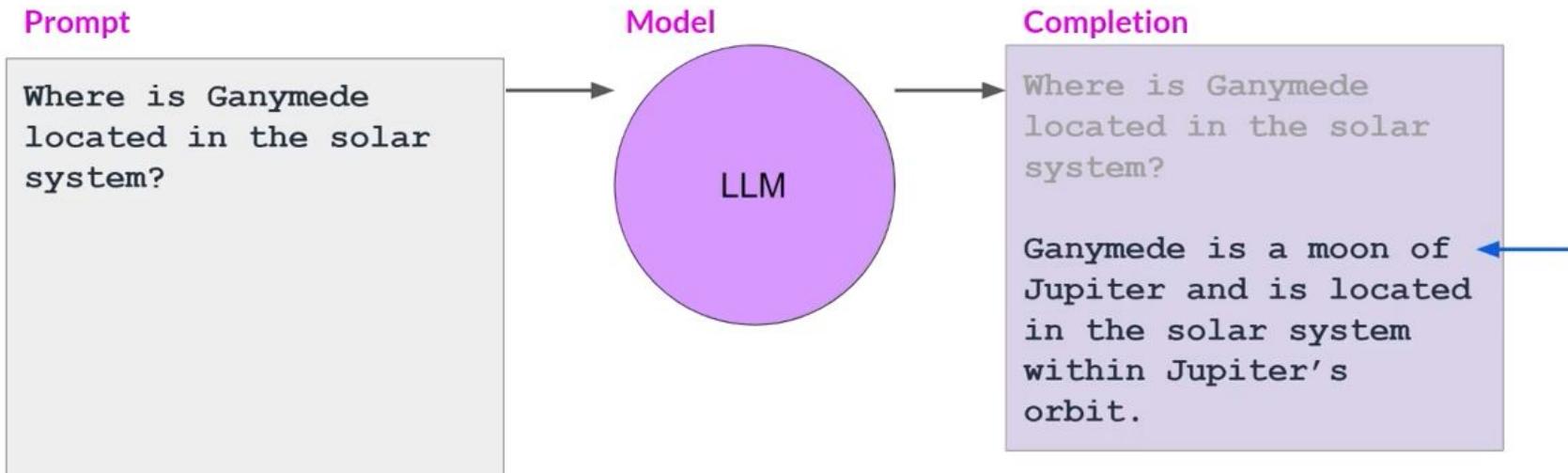
Model



Context window

- typically a few 1000 words.

Prompts and completions



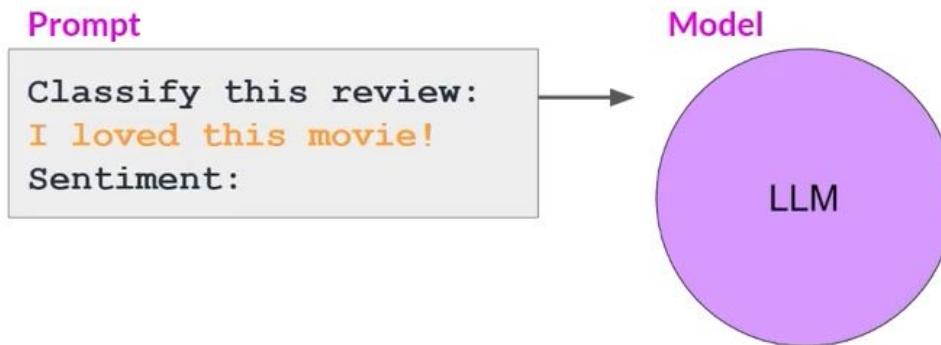
Context window

- typically a few 1000 words.

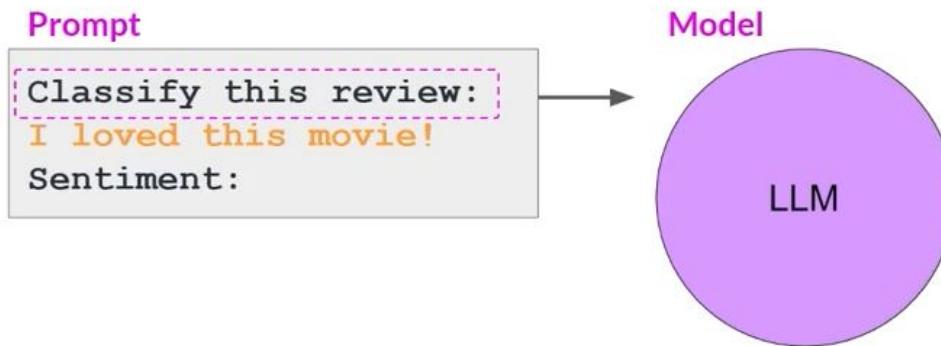
In-context learning (ICL) - zero shot inference



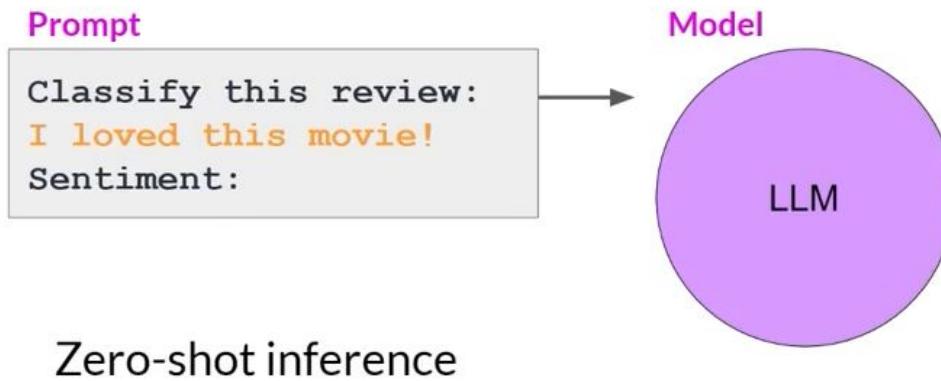
In-context learning (ICL) - zero shot inference



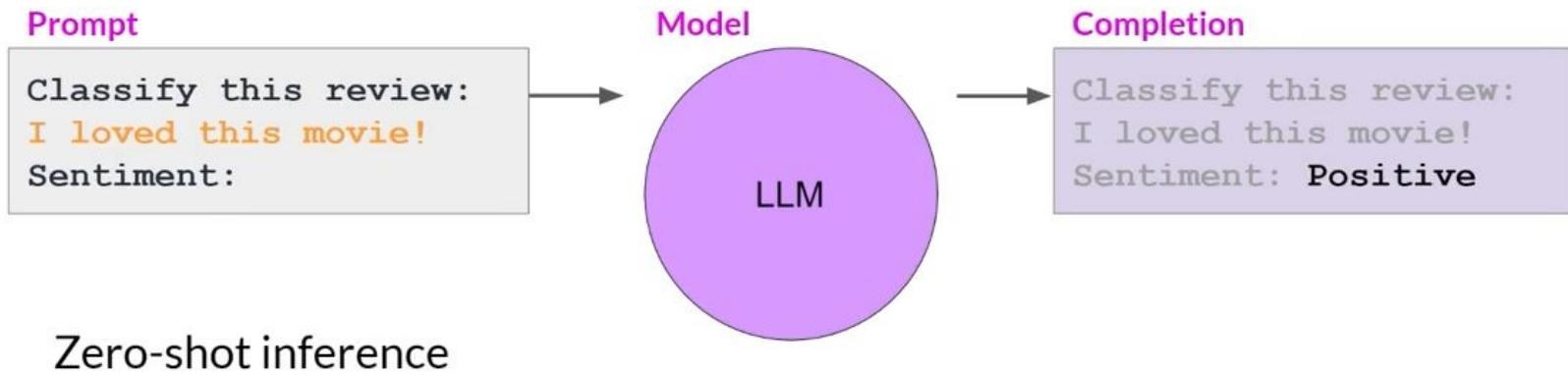
In-context learning (ICL) - zero shot inference



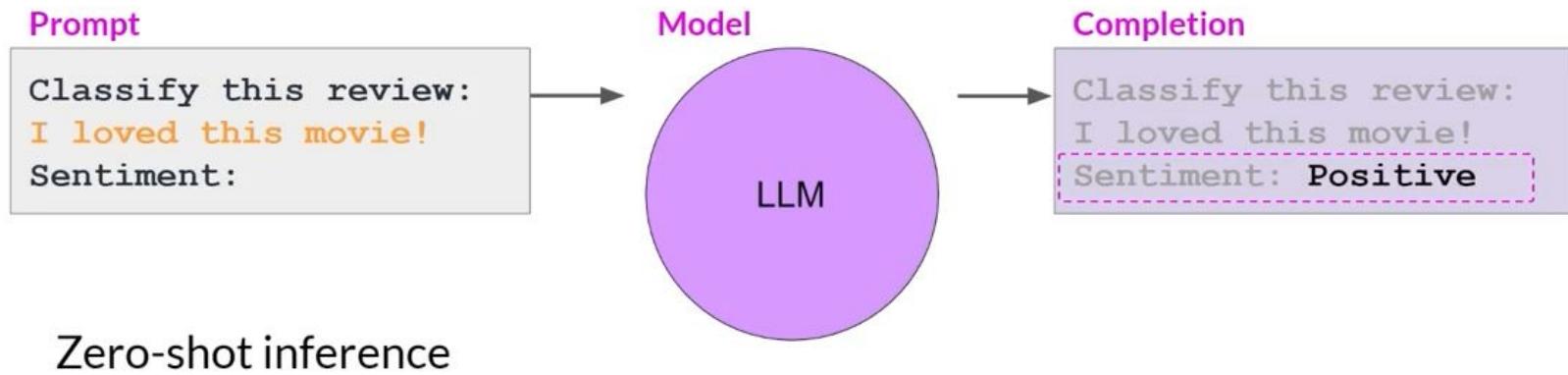
In-context learning (ICL) - zero shot inference



In-context learning (ICL) - zero shot inference



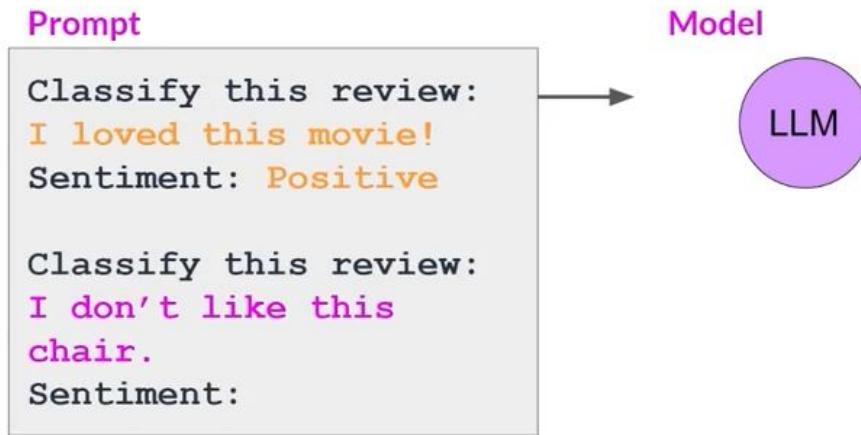
In-context learning (ICL) - zero shot inference



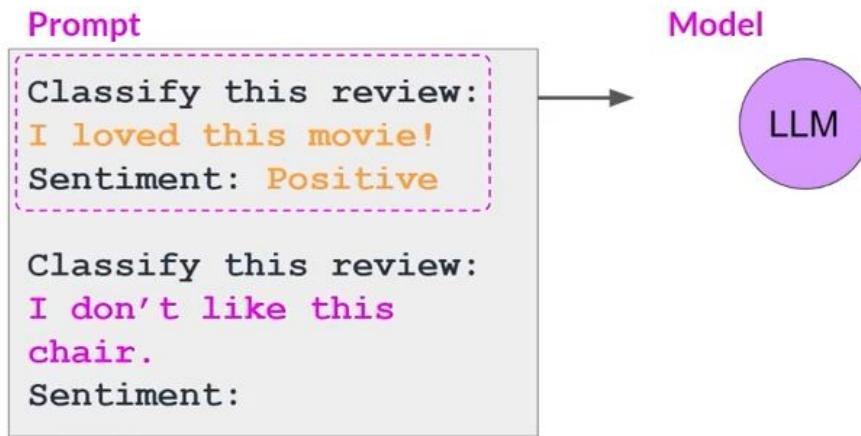
In-context learning (ICL) - zero shot inference



In-context learning (ICL) - one shot inference



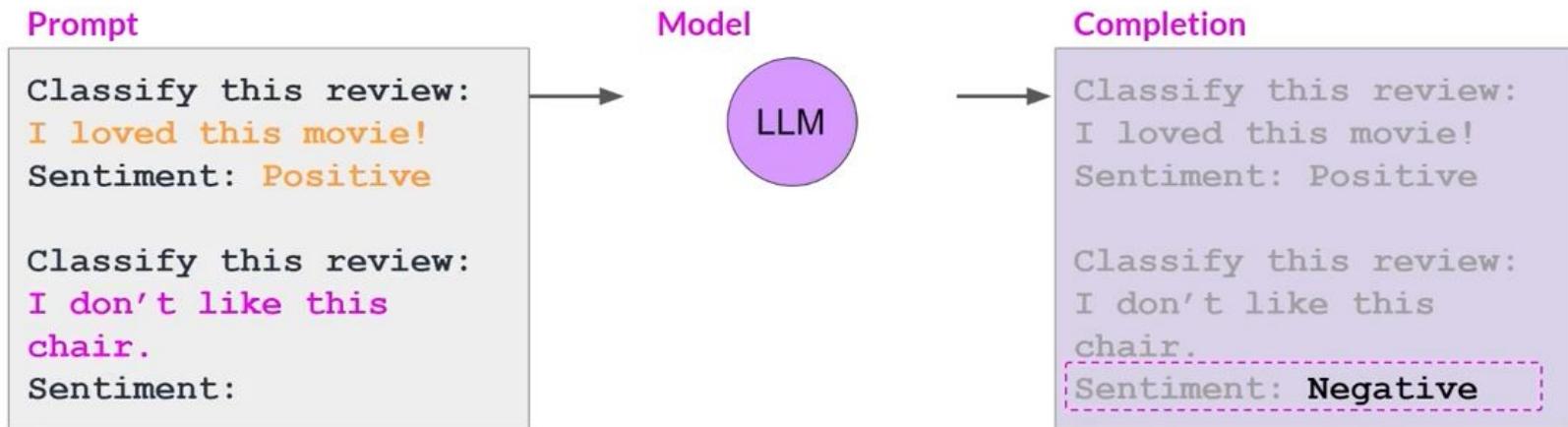
In-context learning (ICL) - one shot inference



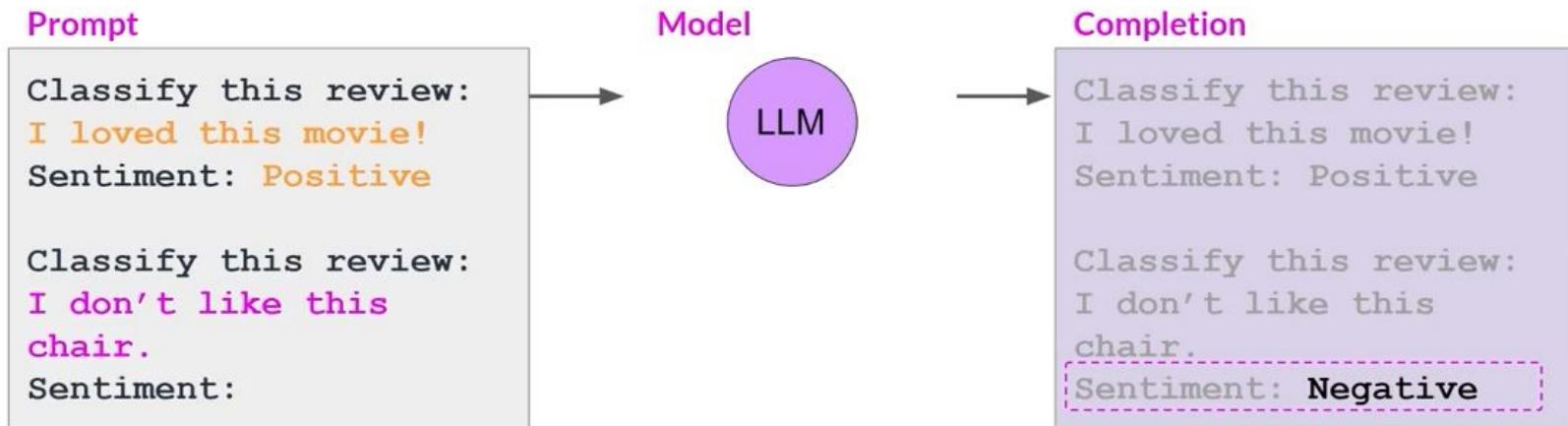
In-context learning (ICL) - one shot inference



In-context learning (ICL) - one shot inference



In-context learning (ICL) - one shot inference



One-shot inference

In-context learning (ICL) - few shot inference

Model



In-context learning (ICL) - few shot inference

Prompt

Classify this review:

I loved this DVD!

Sentiment: Positive

Classify this review:

I don't like this
chair.

Sentiment: Negative

Classify this review:

This is not great.

Sentiment:

Model



In-context learning (ICL) - few shot inference

Prompt

Classify this review:

I loved this DVD!

Sentiment: Positive

Classify this review:

I don't like this
chair.

Sentiment: Negative

Classify this review:

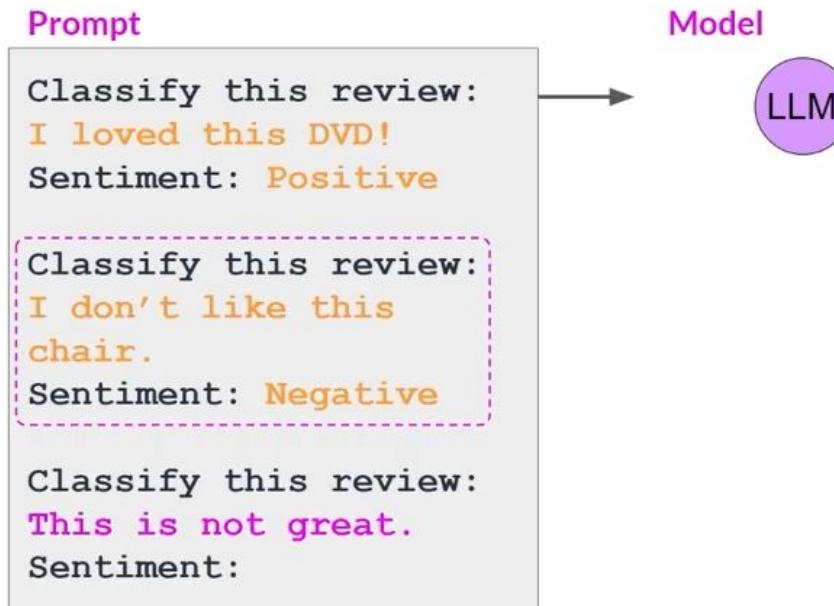
This is not great.

Sentiment:

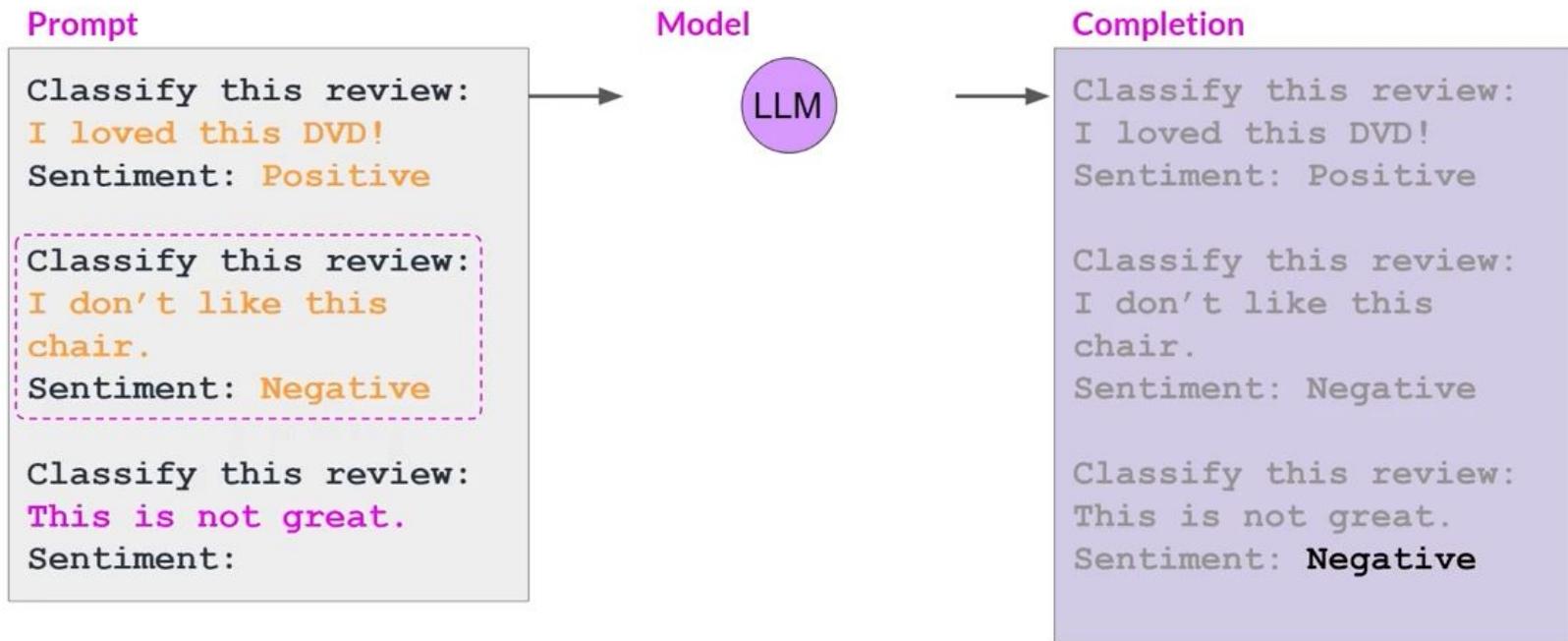
Model



In-context learning (ICL) - few shot inference



In-context learning (ICL) - few shot inference



Summary of in-context learning (ICL)

Prompt // Zero Shot

Classify this review:
I loved this movie!
Sentiment:

Prompt // One Shot

Classify this review:
I loved this movie!
Sentiment: **Positive**

Classify this review:
I don't like this chair.
Sentiment:

Prompt // Few Shot

Classify this review:
I loved this movie!
Sentiment: **Positive**

Classify this review:
I don't like this chair.
Sentiment: **Negative**

Classify this review:
Who would use this product?
Sentiment:

Summary of in-context learning (ICL)

Prompt // Zero Shot

Classify this review:
I loved this movie!
Sentiment:

Context Window
(few thousand words)

Prompt // One Shot

Classify this review:
I loved this movie!
Sentiment: Positive

Classify this review:
I don't like this chair.
Sentiment:

Prompt // Few Shot

Classify this review:
I loved this movie!
Sentiment: Positive

Classify this review:
I don't like this chair.
Sentiment: Negative

Classify this review:
Who would use this product?
Sentiment:

Summary of in-context learning (ICL)

Prompt // Zero Shot

Classify this review:
I loved this movie!
Sentiment:

Context Window
(few thousand words)

Prompt // One Shot

Classify this review:
I loved this movie!
Sentiment: **Positive**

Classify this review:
I don't like this chair.
Sentiment:

Prompt // Few Shot >5 or 6 examples

Classify this review:
I loved this movie!
Sentiment: **Positive**

Classify this review:
I don't like this chair.
Sentiment: **Negative**

Classify this review:
Who would use this product?
Sentiment:

04 →

LLM fine-tuning

(AI)

Limitations of in-context learning

Classify this review:

I loved this movie!

Sentiment: Positive

Classify this review:

I don't like this chair.

Sentiment: Negative

Classify this review:

This sofa is so ugly.

Sentiment: Negative

Classify this review:

Who would use this product?

Sentiment:

Limitations of in-context learning

Classify this review:

I loved this movie!

Sentiment: Positive

Classify this review:

I don't like this chair.

Sentiment: Negative

Classify this review:

This sofa is so ugly.

Sentiment: Negative

Classify this review:

Who would use this product?

Sentiment:

- In-context learning may not work for smaller models  LLM

Limitations of in-context learning

Classify this review:

I loved this movie!

Sentiment: Positive

Classify this review:

I don't like this chair.

Sentiment: Negative

Classify this review:

This sofa is so ugly.

Sentiment: Negative

Classify this review:

Who would use this product?

Sentiment:

Context Window

Even with
multiple
examples

- In-context learning may not work for smaller models LLM
- Examples take up space in the context window

Limitations of in-context learning

Classify this review:

I loved this movie!

Sentiment: Positive

Classify this review:

I don't like this chair.

Sentiment: Negative

Classify this review:

This sofa is so ugly.

Sentiment: Negative

Classify this review:

Who would use this product?

Sentiment:

Context Window

Even with
multiple
examples

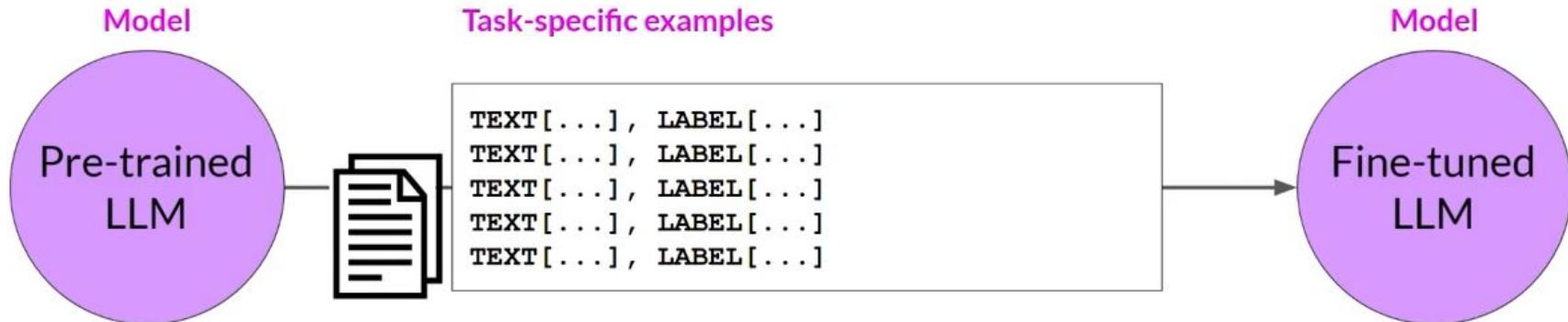
- In-context learning may not work for smaller models LLM

- Examples take up space in the context window

Instead, try fine-tuning
the model

LLM fine-tuning at a high level

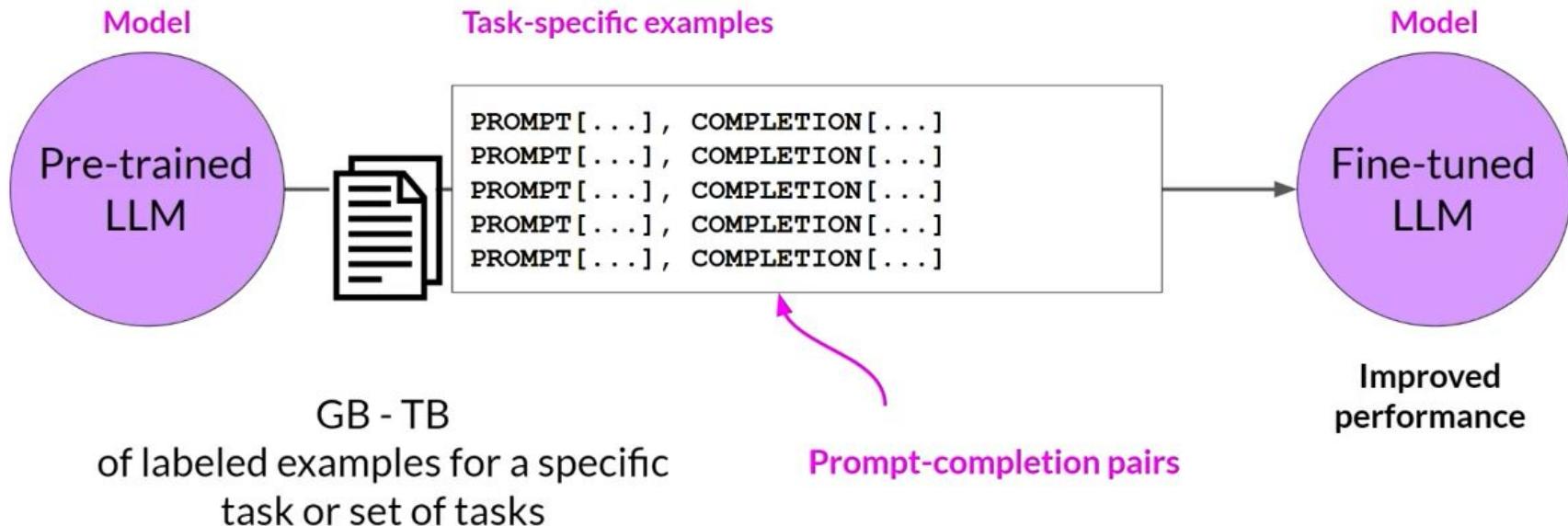
LLM fine-tuning



GB - TB
of labeled examples for a specific
task or set of tasks

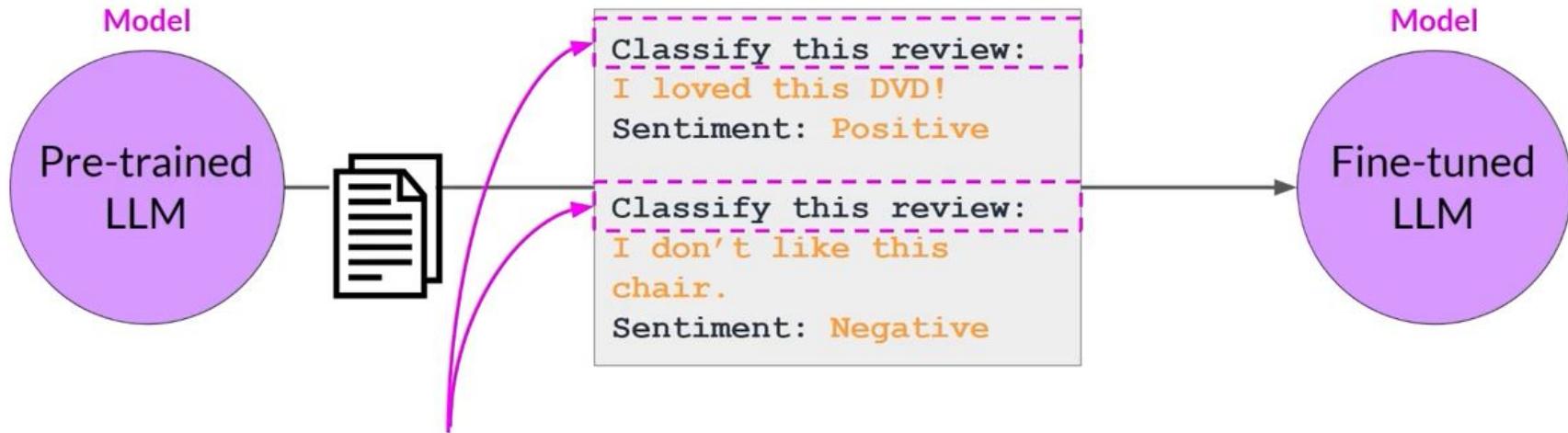
LLM fine-tuning at a high level

LLM fine-tuning



Using prompts to fine-tune LLMs with instruction

LLM fine-tuning



Each prompt/completion pair includes a specific “instruction” to the LLM

Sample prompt instruction templates

Classification / sentiment analysis

```
jinja: "Given the following review:\n{{review_body}}\npredict the associated rating\\ from the following choices (1 being lowest and 5 being highest)\n- {{ answer_choices}\\ | join('\\n- ')}\n|||\n{{answer_choices[star_rating-1]}}"
```

Text generation

```
jinja: Generate a {{star_rating}}-star review (1 being lowest and 5 being highest) about this product {{product_title}}.     |||      {{review_body}}
```

Text summarization

```
jinja: "Give a short sentence describing the following product review:\n{{review_body}}\\n|||\n{{review_headline}}"
```

Sample prompt instruction templates

Classification / sentiment analysis

```
jinja: "Given the following review:\n{{review_body}}\npredict the associated rating\\
from the following choices (1 being lowest and 5 being highest)\n- {{ answer_choices\\
| join('\\n- ')}}\n|||\n{{answer_choices[star_rating-1]}}"
```

Text generation

```
jinja: Generate a {{star_rating}}-star review (1 being lowest and 5 being highest)
about this product {{product_title}}. ||| {{review_body}}
```

Text summarization

```
jinja: Give a short sentence describing the following product review:\n{{review_body}}\\
|||\n{{review_headline}}
```

LLM fine-tuning process

LLM fine-tuning

Prepared instruction dataset



LLM fine-tuning process

LLM fine-tuning

Prepared instruction dataset



Training splits

```
PROMPT[...], COMPLETION[...]  
PROMPT[...], COMPLETION[...]  
PROMPT[...], COMPLETION[...]  
PROMPT[...], COMPLETION[...]  
PROMPT[...], COMPLETION[...]
```

Training

```
PROMPT[...], COMPLETION[...]
```

...

Validation

```
PROMPT[...], COMPLETION[...]
```

...

Test

LLM fine-tuning process

LLM fine-tuning

Prepared instruction dataset



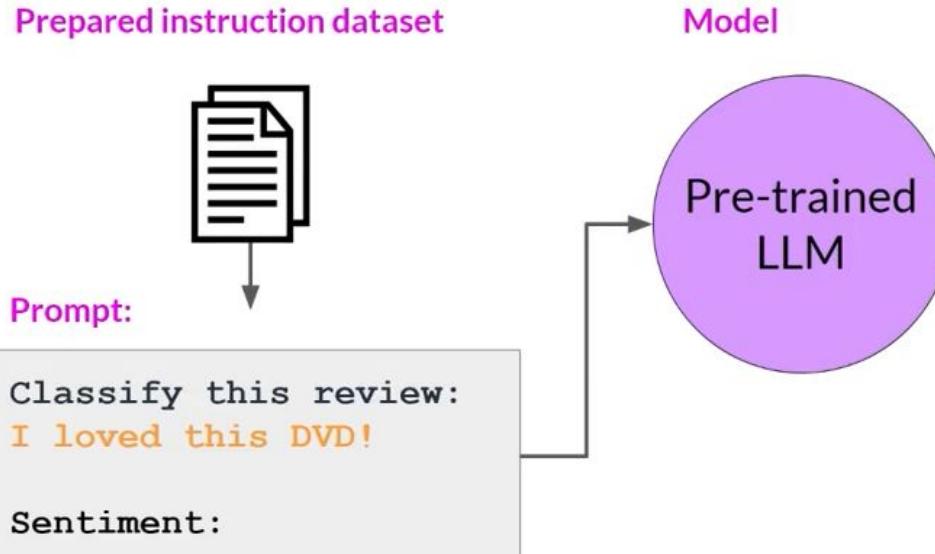
Prompt:

Classify this review:
I loved this DVD!

Sentiment:

LLM fine-tuning process

LLM fine-tuning



LLM fine-tuning process

LLM fine-tuning

Prepared instruction dataset



Prompt:

Classify this review:
I loved this DVD!

Sentiment:

Model

Pre-trained
LLM

LLM completion:

Classify this review:
I loved this DVD!

Sentiment: **Neutral**

LLM fine-tuning process

LLM fine-tuning

Prepared instruction dataset



Prompt:

Classify this review:
I loved this DVD!

Sentiment:

Model

Pre-trained
LLM

LLM completion:

Classify this review:
I loved this DVD!

Sentiment: **Neutral**

Label:

Classify this review:
I loved this DVD!

Sentiment: **Positive**

LLM fine-tuning process

LLM fine-tuning

Prepared instruction dataset



Prompt:

Classify this review:
I loved this DVD!

Sentiment:

Model

Pre-trained
LLM

LLM completion:

Classify this review:
I loved this DVD!

Sentiment: **Neutral**

Label:

Classify this review:
I loved this DVD!

Sentiment: **Positive**

LLM fine-tuning process

LLM fine-tuning

Prepared instruction dataset



Prompt:

Classify this review:
I loved this DVD!

Sentiment:

Model

Pre-trained
LLM

LLM completion:

Classify this review:
I loved this DVD!

Sentiment: **Neutral**

Label:

Classify this review:
I loved this DVD!

Sentiment: **Positive**

Loss: Cross-Entropy

LLM fine-tuning process

LLM fine-tuning

Prepared instruction dataset



Prompt:

Classify this review:
I loved this DVD!

Sentiment:

Model

Pre-trained
LLM

LLM completion:

Classify this review:
I loved this DVD!

Sentiment: Neutral

Label:

Classify this review:
I loved this DVD!

Sentiment: Positive

Loss: Cross-Entropy

LLM fine-tuning process

LLM fine-tuning

Prepared instruction dataset



Training splits

```
PROMPT[...], COMPLETION[...]  
PROMPT[...], COMPLETION[...]  
PROMPT[...], COMPLETION[...]  
PROMPT[...], COMPLETION[...]  
PROMPT[...], COMPLETION[...]
```

Training

```
PROMPT[...], COMPLETION[...]  
...
```

Validation

validation_accuracy

```
PROMPT[...], COMPLETION[...]  
...
```

Test

LLM fine-tuning process

LLM fine-tuning

Prepared instruction dataset



Training splits

```
PROMPT[...], COMPLETION[...]  
PROMPT[...], COMPLETION[...]  
PROMPT[...], COMPLETION[...]  
PROMPT[...], COMPLETION[...]  
PROMPT[...], COMPLETION[...]
```

Training

```
PROMPT[...], COMPLETION[...]
```

...

Validation

```
PROMPT[...], COMPLETION[...]
```

...

Test

test_accuracy

LLM fine-tuning process



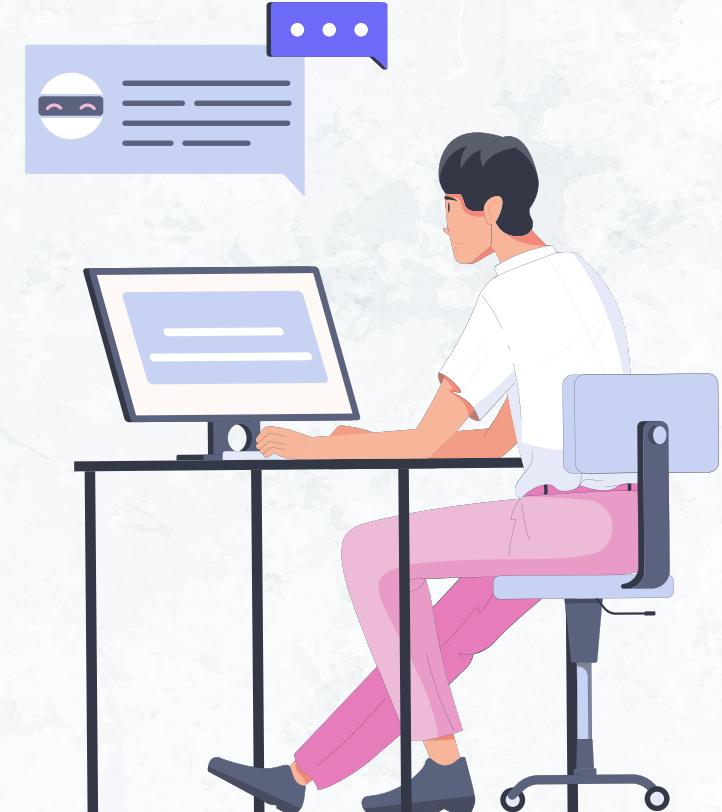
05 →

LLM main use cases

(AI)

Main use cases

- Chatbots and virtual assistants 
- Content generation
- Transforming (translation, parsing)
- Editing and proofreading
- Sentiment analysis and text classification
- Question answering systems (RAG)
- Code generation and debugging
- Summarization
- Agents
- Automation! 



LLM chatbot

 ChatBot

Who designed the street layout of Washington DC?

(T)

The street layout of Washington D.C. was designed by Pierre Charles L'Enfant, a French-born American architect and civil engineer.



LLM use cases & tasks

Essay Writer

Title: Write a 5 paragraph short academic style essay for the title “A History of Machine Learning”

Generate

The history of machine learning can be traced back to the 1950s, when Arthur Samuel defined machine learning as a field of study that gives computers the ability to learn without being explicitly programmed. In 1956, Samuel's program called 'Checkers-playing program' became the first computer program to win a game against a human opponent. This achievement marked the beginning of machine learning as a field of research.

In the 1960s, Marvin Minsky and Seymour Papert published their seminal

LLM use cases & tasks

Summarize

Text file:



support.txt

Generate

In the chat session, Support efficiently and effectively assists Alex, who was initially unable to access their account due to issues with a password reset email, leading to a positive customer service experience.

LLM use cases & tasks

Translate

French:

J'aime l'apprentissage automatique.

German:

Ich liebe maschinelles Lernen.

Generate

LLM use cases & tasks

Code AI

Prompt:

Write some python code that will return the mean of every column in a dataframe.

Generate

Code:

```
import pandas as pd  
  
df = pd.DataFrame({  
    'A': [1, 2, 3, 4, 5],  
    'B': [2, 3, 4, 5, 6],  
    'C': [3, 4, 5, 6, 7]  
})  
  
mean_values = df.mean()
```

LLM use cases & tasks

Entity Extraction

Input:

Scientist Dr. Evangeline Starlight of Technopolis announced a breakthrough in quantum computing at Nova University. Mayor Orion Pulsar commended her. The discovery will be shared at the Galactic Quantum Computing Symposium in Cosmos.

Extract

LLM use cases & tasks

Entity Extraction

Input:

Scientist Dr. Evangeline Starlight of Technopolis announced a breakthrough in quantum computing at Nova University. Mayor Orion Pulsar commended her. The discovery will be shared at the Galactic Quantum Computing Symposium in Cosmos.

The named entities in this shorter text are "Dr. Evangeline Starlight", "Technopolis", "quantum computing", "Nova University", "Mayor Orion Pulsar", "Galactic Quantum Computing Symposium", and "Cosmos".

Extract

LLM use cases & tasks

Flight Information

Input:

Is flight VA8005 landing on time?

Go

LLM use cases & tasks

Flight Information

Input:

Is flight VA8005 landing on time?

Formatting API query...

Go



A large, semi-transparent light purple rectangular overlay covers the bottom half of the card.

LLM use cases & tasks

Flight Information

Input:

Is flight VA8005 landing on time?

Formatting API query...

Go

Making request...

Processing response.

Done.

LLM use cases & tasks

Flight Information

Input:

Is flight VA8005 landing on time?

Formatting API query...

Go

Making request...

Processing response.

Done.

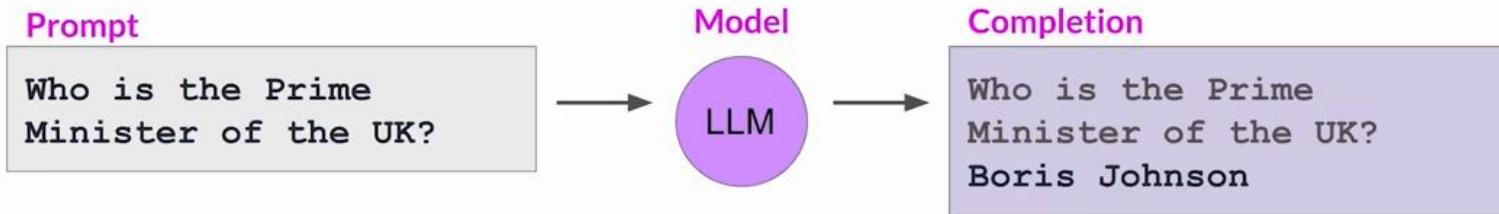
Flight VA8005 from San Francisco to Sydney Australia is on time and is due to land at 7:00am local time.

06 →

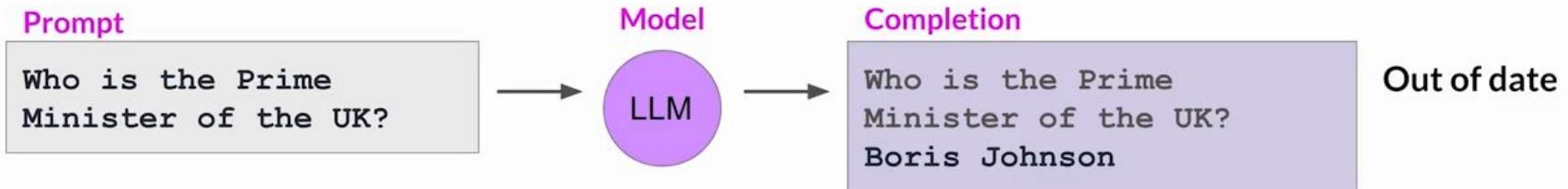
Using LLM in applications

(AI)

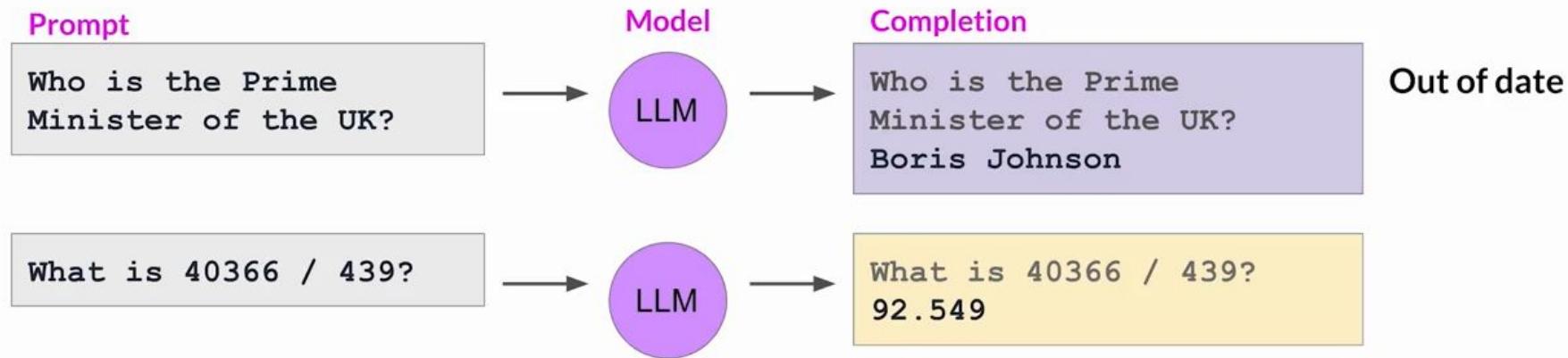
Models having difficulty



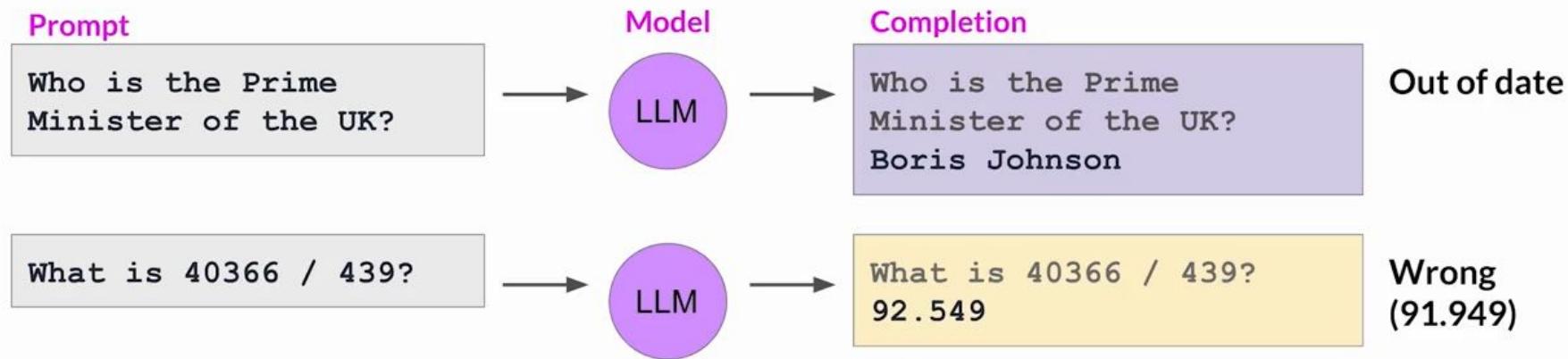
Models having difficulty



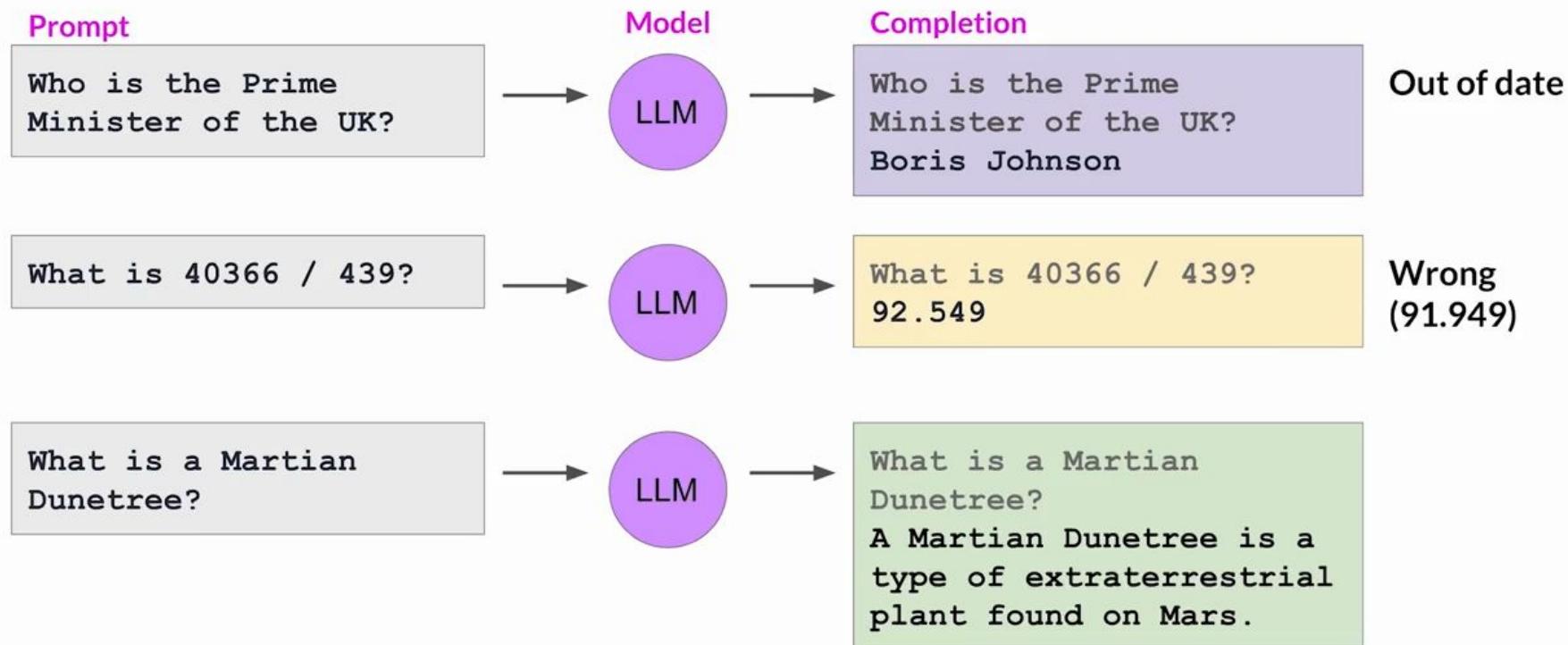
Models having difficulty



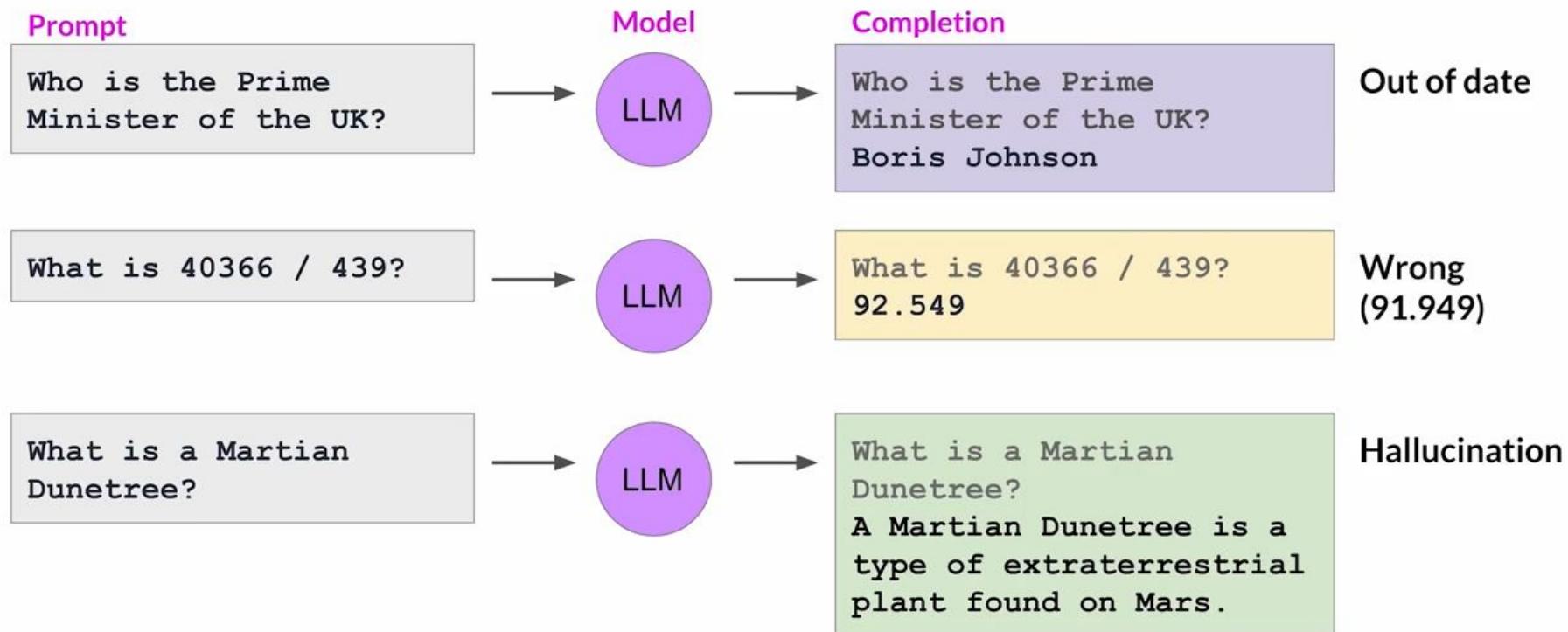
Models having difficulty



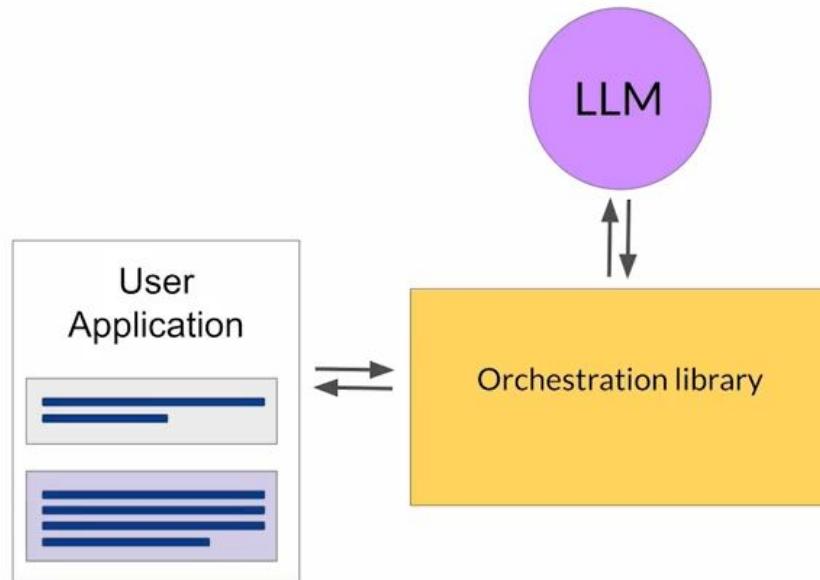
Models having difficulty



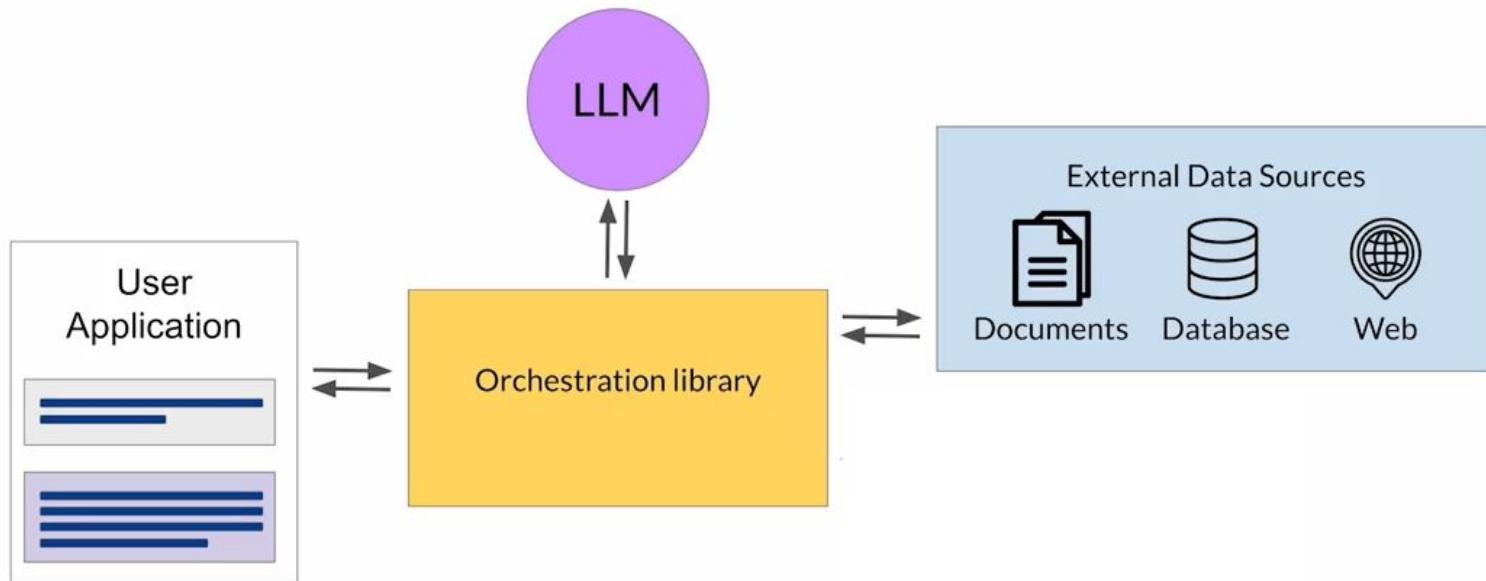
Models having difficulty



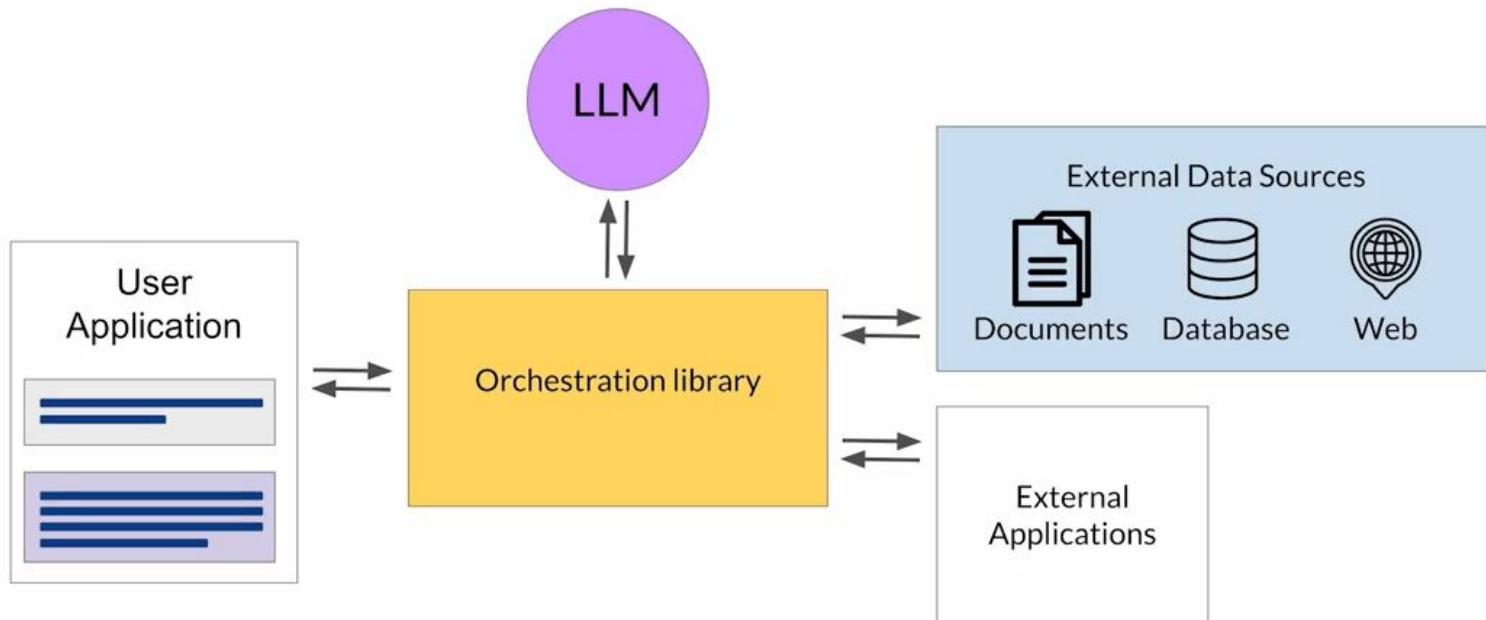
LLM-powered applications



LLM-powered applications

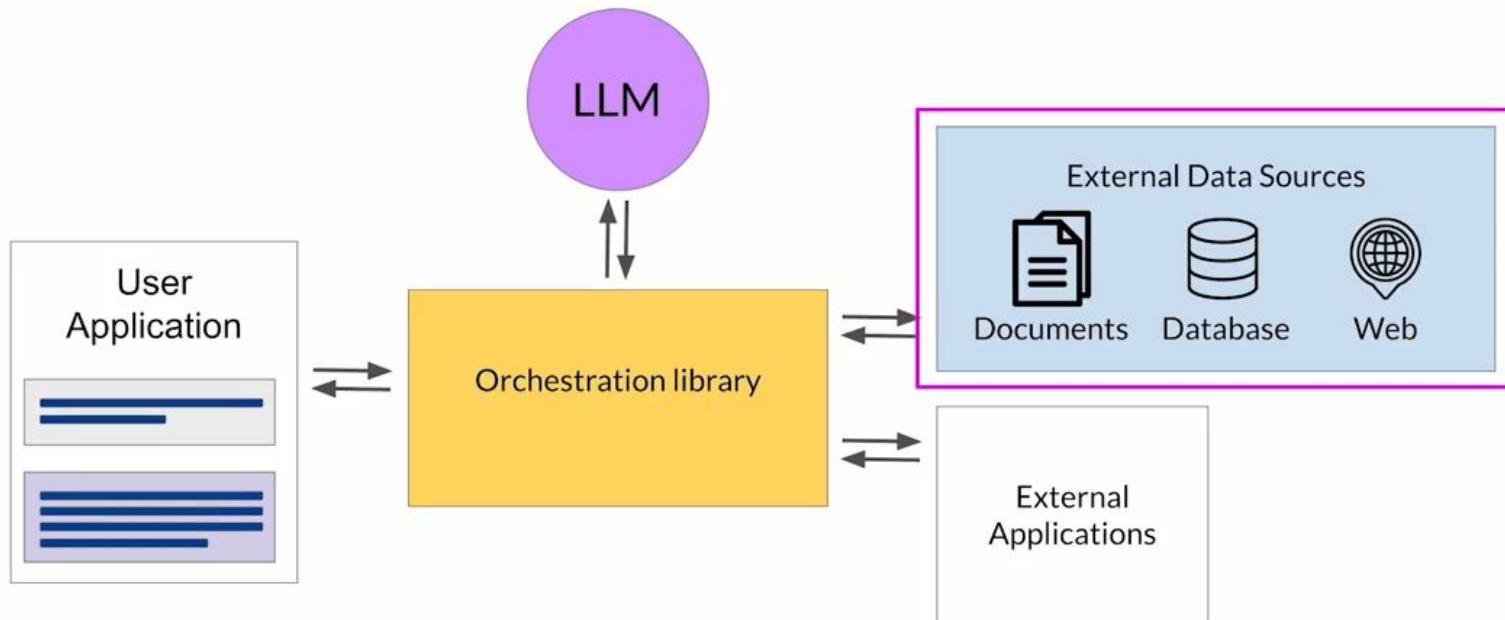


LLM-powered applications

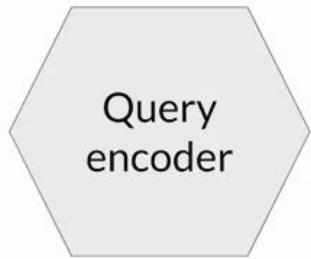


Retrieval augmented generation (RAG)

LLM-powered applications

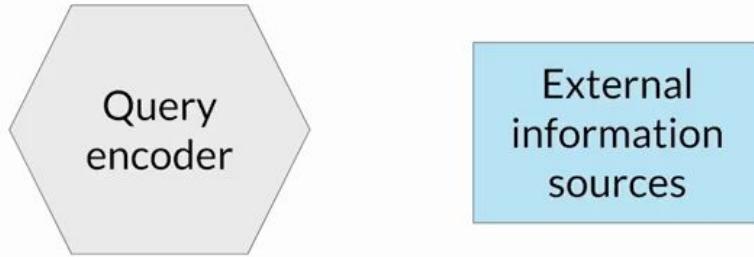


Retrieval Augmented Generation (RAG)



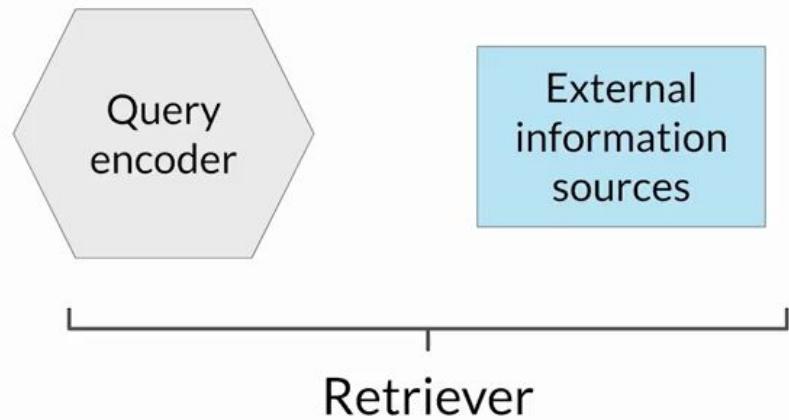
Lewis et al. 2020 “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”

Retrieval Augmented Generation (RAG)



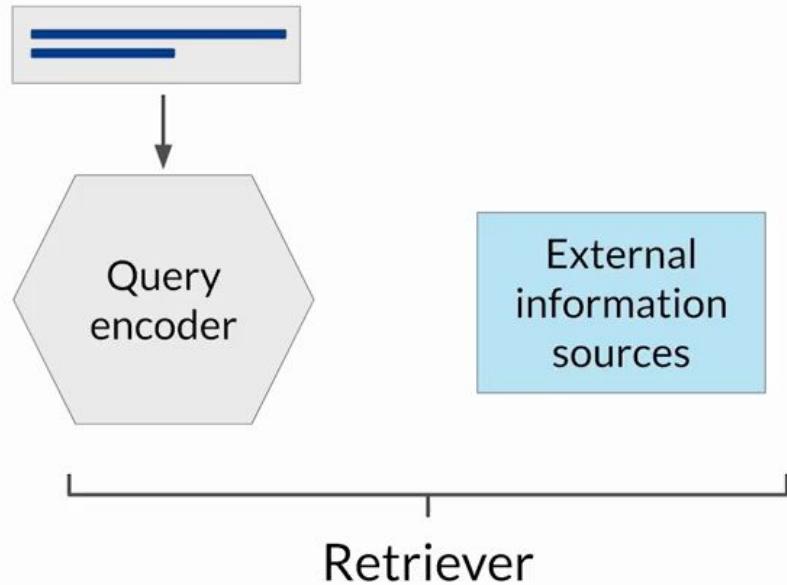
Lewis et al. 2020 “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”

Retrieval Augmented Generation (RAG)



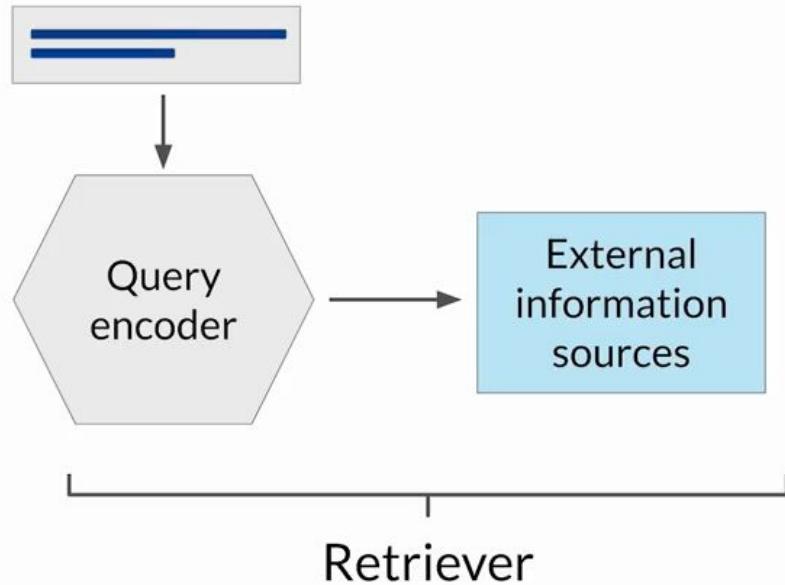
Lewis et al. 2020 “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”

Retrieval Augmented Generation (RAG)



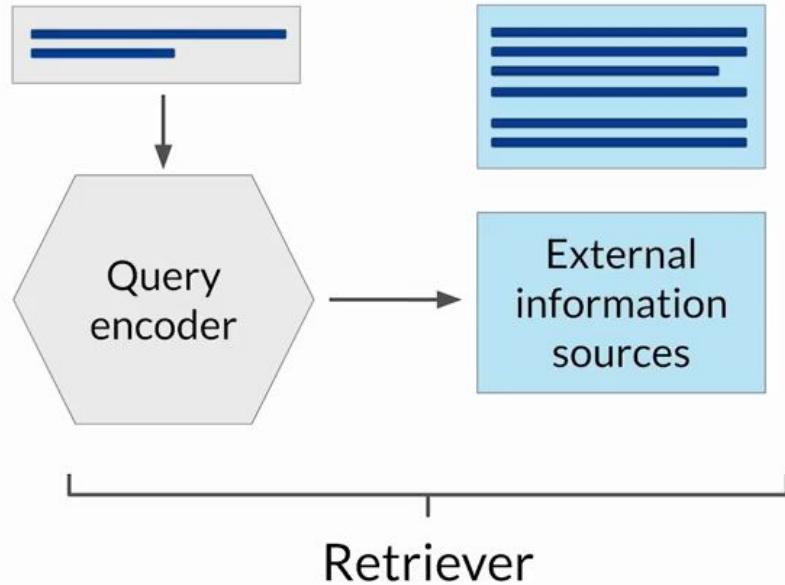
Lewis et al. 2020 “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”

Retrieval Augmented Generation (RAG)



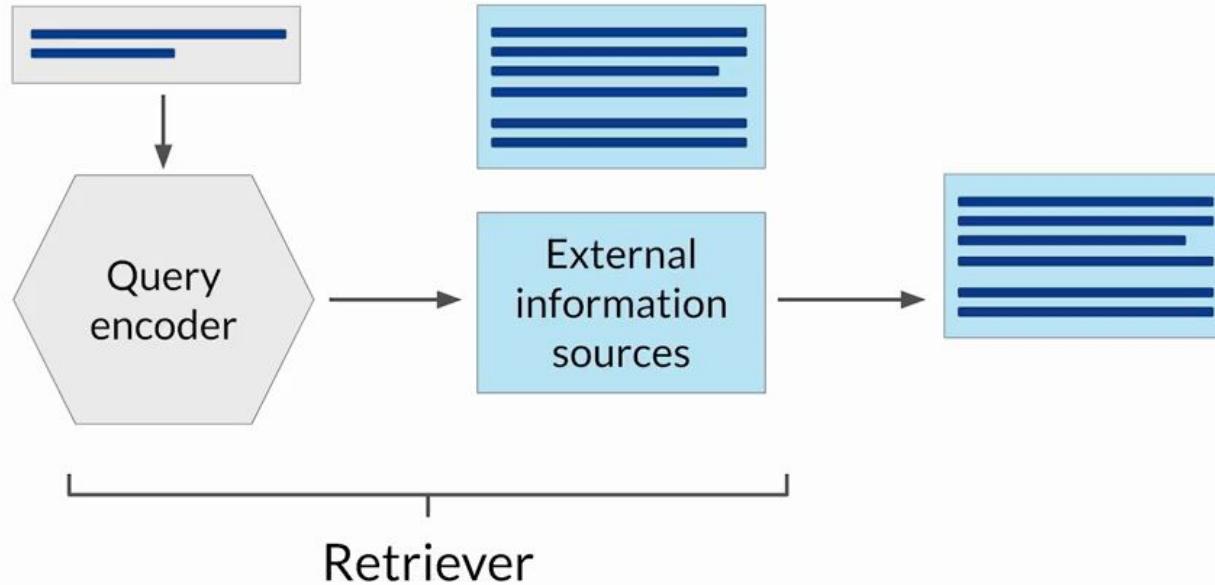
Lewis et al. 2020 “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”

Retrieval Augmented Generation (RAG)



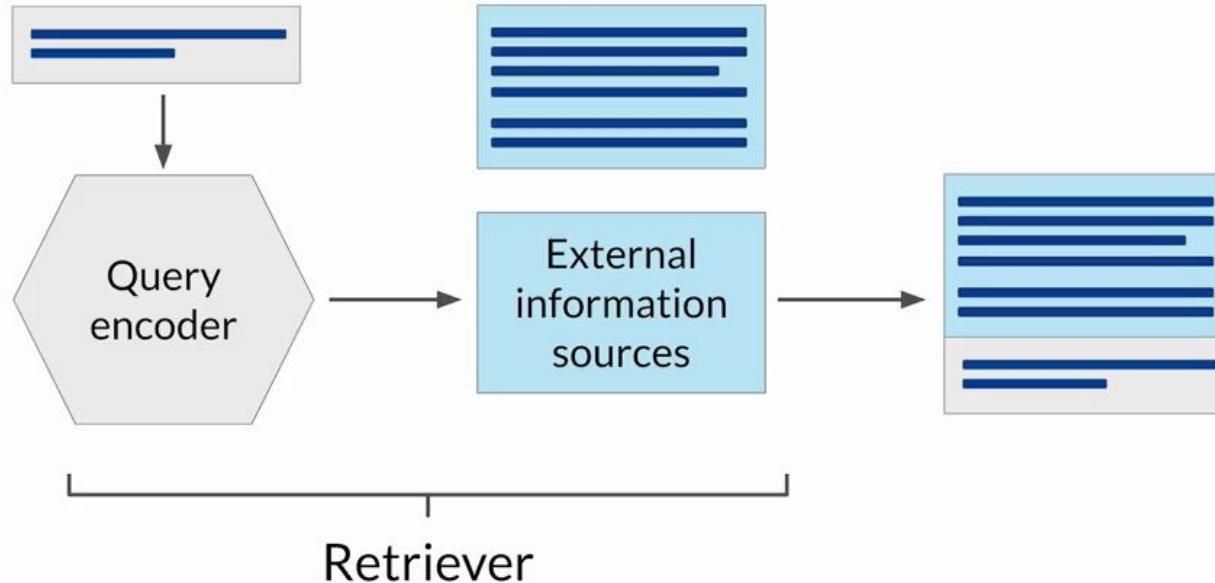
Lewis et al. 2020 “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”

Retrieval Augmented Generation (RAG)



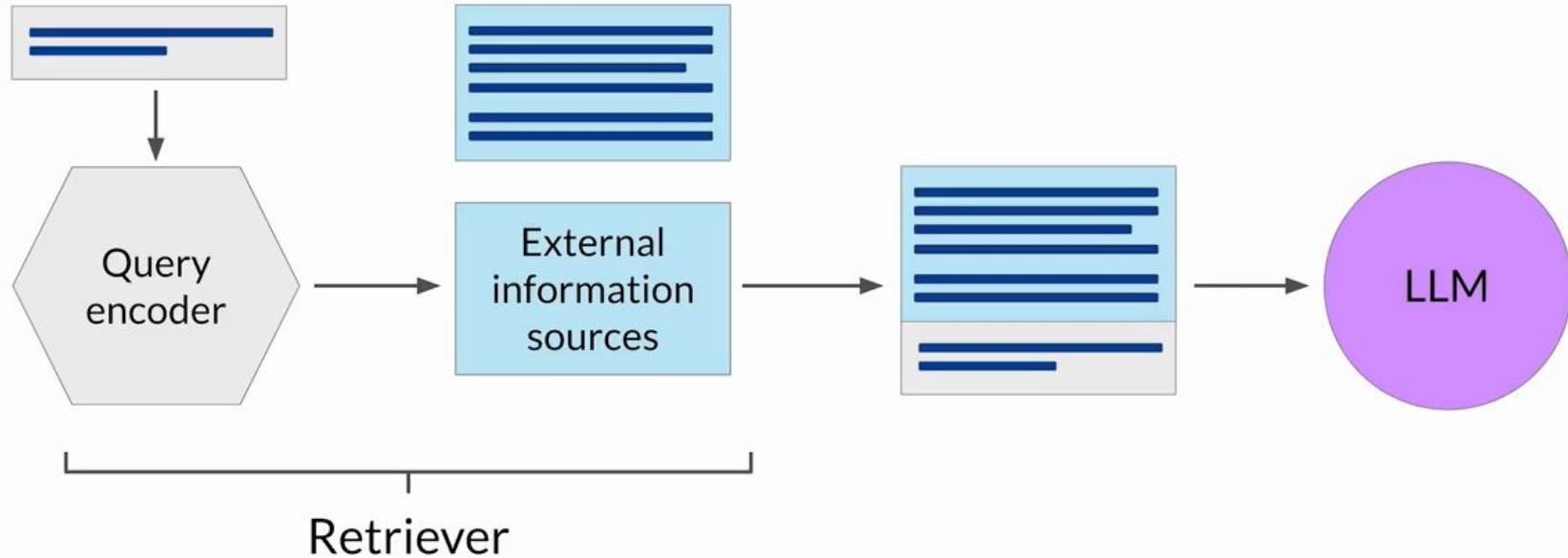
Lewis et al. 2020 “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”

Retrieval Augmented Generation (RAG)



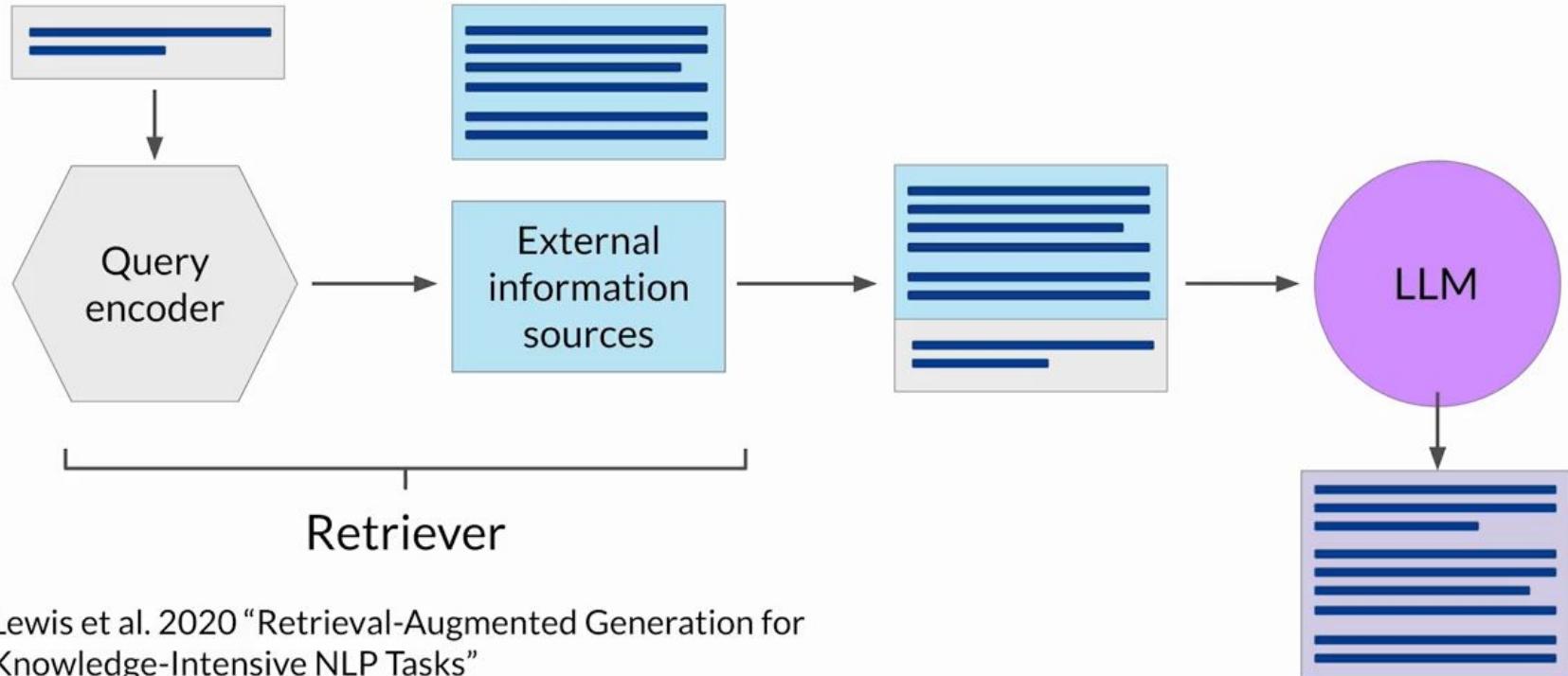
Lewis et al. 2020 “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”

Retrieval Augmented Generation (RAG)



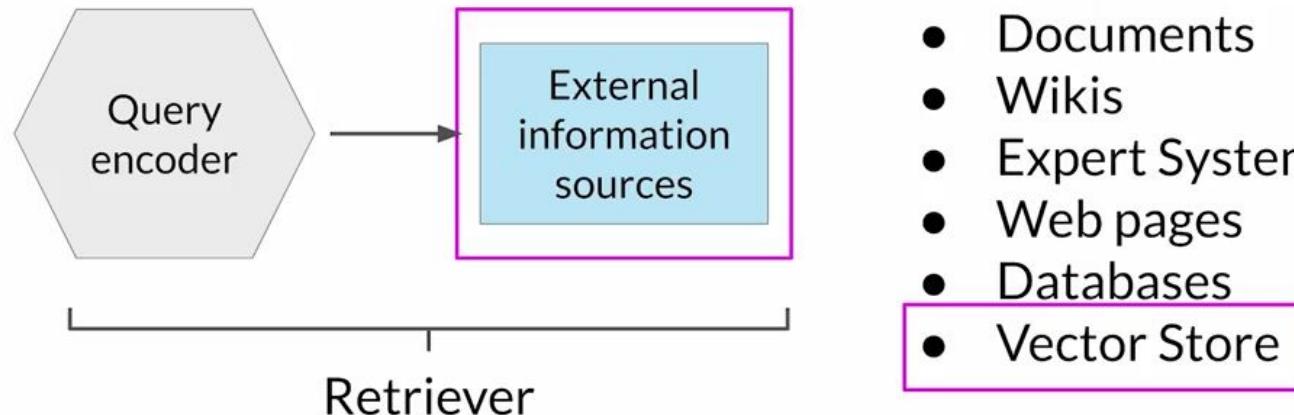
Lewis et al. 2020 “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”

Retrieval Augmented Generation (RAG)



Lewis et al. 2020 "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks"

RAG integrates with many types of data sources



Example: Searching legal documents

Input query

Example: Searching legal documents

Input query

```
Who is the  
plaintiff in case  
22-48710BI-SME?
```

Example: Searching legal documents

Input query

```
Who is the  
plaintiff in case  
22-48710BI-SME?
```



Query Encoder

Example: Searching legal documents

Input query

```
Who is the  
plaintiff in case  
22-48710BI-SME?
```



Query Encoder



documents

External Information Sources

Example: Searching legal documents

Input query

Who is the
plaintiff in case
22-48710BI-SME?

UNITED STATES DISTRICT
COURT
SOUTHERN DISTRICT OF MAINE

CASE NUMBER: 22-48710BI-SME

Busy Industries (Plaintiff)
vs.
State of Maine (Defendant)



Query Encoder



documents

External Information Sources

Example: Searching legal documents

Input query

Who is the plaintiff in case 22-48710BI-SME?



Query Encoder

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF MAINE

CASE NUMBER: 22-48710BI-SME

Busy Industries (Plaintiff)
vs.
State of Maine (Defendant)



documents

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF MAINE

CASE NUMBER: 22-48710BI-SME

Busy Industries (Plaintiff)
vs.
State of Maine (Defendant)

Who is the plaintiff in case 22-48710BI-SME?



External Information Sources

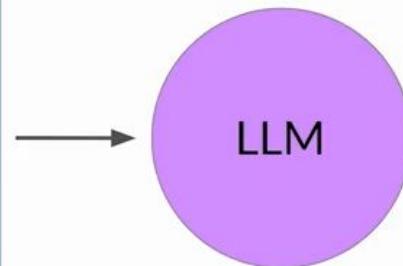
Example: Searching legal documents

UNITED STATES DISTRICT
COURT
SOUTHERN DISTRICT OF MAINE

CASE NUMBER: 22-48710BI-SME

Busy Industries (Plaintiff)
vs.
State of Maine (Defendant)

Who is the plaintiff in
case 22-48710BI-SME?



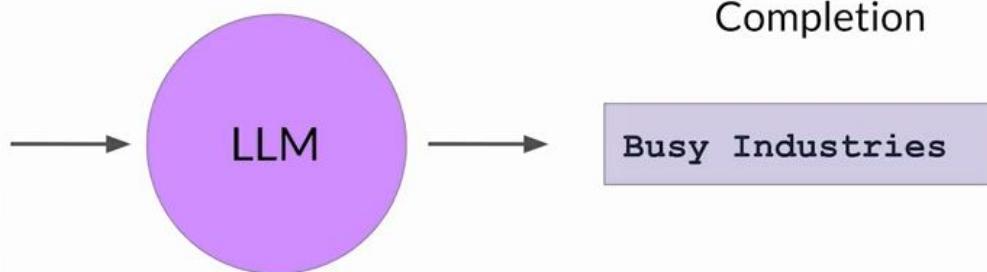
Example: Searching legal documents

UNITED STATES DISTRICT
COURT
SOUTHERN DISTRICT OF MAINE

CASE NUMBER: 22-48710BI-SME

Busy Industries (Plaintiff)
vs.
State of Maine (Defendant)

Who is the plaintiff in
case 22-48710BI-SME?



Thanks! →

Any questions?

Rodrigo Gonzalez, PhD

rodrigo.gonzalez@ingenieria.uncuyo.edu.ar

