

# توليد اللغات التقدم

- مقدمة إلى هندسة الأوامر Prompt Engineering
- تقنيات هندسة الأوامر
  - أوامر بدون أمثلة مباشرة
  - Zero-shot prompting
  - أوامر بأمثلة قليلة
  - Few-shot prompting
  - أوامر بسلسلة الفكر
- COT (Chain-Of-Thought) prompting
- التوليد المعزز للاسترجاع (RAG)
  - كيف يعمل؟
  - نماذج RAG التدريبية

## المحتوى

# مقدمة إلى هندسة الأوامر

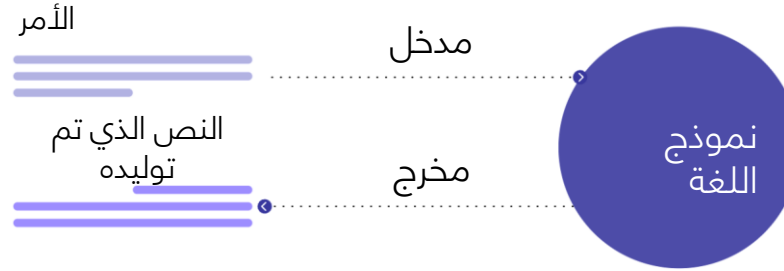
# ما معنى "Prompt" ؟

- أوامر LLM هي المدخلات النصية مثل الأسئلة والتعليمات التي تدخلها في الـ LLM الخاص بك للحصول على إجابات.
- يتنبأ النموذج بالاستجابة المناسبة للأمر الذي أدخلته. بشكل عام، ستوفر لك **الأوامر الأكثر تحديداً** والتي تمت صياغتها بعناية استجابات أفضل.



# عناصر الأمر

- يمكن أن تحتوي الأوامر على أي من المكونات التالية:
- التعليمات - Instruction مهمة أو تعليمات محددة تريد من النموذج أن يؤديها
  - السياق - Context يمكن أن يتضمن سياقًا إضافيًا يمكنه توجيه النموذج نحو استجابات أفضل
  - بيانات الإدخال - Input Data هو السؤال الذي يهمنا الإجابة عليه
  - مؤشر الإخراج - Output Indicator يشير إلى نوع الإخراج أو تنسيقه.



# هندسة الأوامر Prompt Engineering

- يمكنك تحقيق الكثير من خلال أوامر بسيطة، ولكن جودة النتائج تعتمد على مقدار المعلومات التي تقدمها لها ومدى صياغتها بشكل جيد.
- في سياق LLM، **هندسة الأوامر** هي عملية هيكلة النص الذي يمكن تفسيره وفهمه بواسطة النموذج.

Prompt

The sky is

Output:

blue  
The sky is blue on a clear day. On a cloudy day, the sky may be gray or white.

Prompt:

Complete the sentence:  
The sky is

Output:

so beautiful today.

# تقنيات هندسة الأوامر

# نصائح عامة لتصميم الأوامر

- اعتماد البساطة
- توفير السياق
- تحديد مهمة واضحة وتجنّب الأسئلة المفتوحة
- استخدام الكلمات الرئيسية
- استعمال التكرار Iterate





# تقنيات هندسة الأوامر

التقنيات

Techniques

N-Shot  
Prompting

أوامر بسلسلة الفكر

Chain-Of-Thought  
(COT) Prompting

Zero-shot

أوامر بدون أمثلة  
مباشرة

Few-shots

أوامر بأمثلة  
قليلة

# أوامر بدون أمثلة مباشرة Zero-shot Prompting

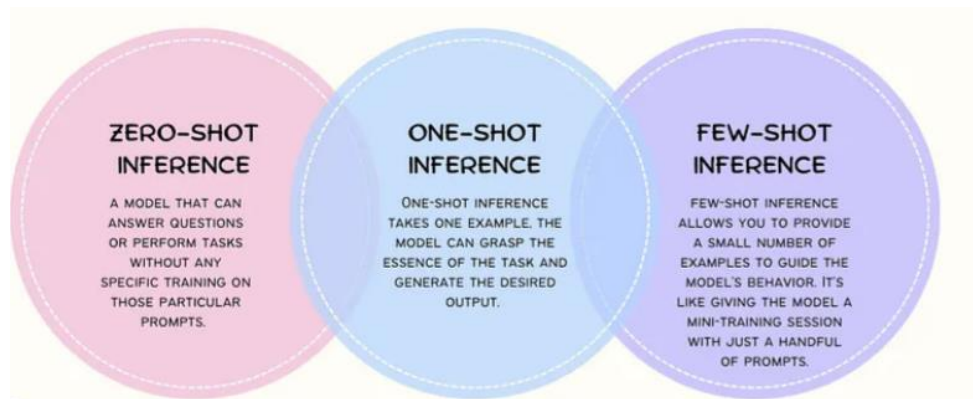
- يتم تدريب LLM على كميات كبيرة من البيانات
- لا تقدم للنموذج أي أمثلة
- عندما لا تعطي الأوامر بدون أمثلة مباشرة نتائج جيدة => قدم أمثلة

*Prompt:*

Classify the text into neutral, negative or positive.  
Text: I think the vacation is okay.  
Sentiment:

*Output:*

Neutral



# أوامر بأمثلة بسيطة Few-shot Prompting

*Prompt:*

A "whatpu" is a small, furry animal native to Tanzania. An example of a sentence that uses the word whatpu is:  
We were traveling in Africa and we saw these very cute whatpus.  
To do a "farduddle" means to jump up and down really fast. An example of a sentence that uses the word farduddle is:

*Output:*

When we won the game, we all started to farduddle in celebration.

# مشاكل وحدود الأوامر بأمثلة بسيطة

The odd numbers in this group add up to an even number: 4, 8, 9, 15, 12, 2, 1.

A: The answer is False.

The odd numbers in this group add up to an even number: 17, 10, 19, 4, 8, 12, 24.

A: The answer is True.

The odd numbers in this group add up to an even number: 16, 11, 14, 4, 8, 13, 24.

A: The answer is True.

The odd numbers in this group add up to an even number: 17, 9, 10, 12, 13, 4, 2.

A: The answer is False.

The odd numbers in this group add up to an even number: 15, 32, 5, 13, 82, 7, 1.

A:

*Output:*

The answer is True.

# أوامر سلسلة الفكر Chain-of-Thought Prompting

Incorrect

## Prompt

Which is a faster way to get to work?  
Option 1: Take a 1000 minute bus, then a half hour train, and finally a 10 minute bike ride.  
Option 2: Take an 800 minute bus, then an hour train, and finally a 30 minute bike ride.

## Output

Option 1 is a faster way to get to work.



Generate

# أوامر سلسلة الفكر Chain-of-Thought Prompting

Correct

## Prompt

Which is a faster way to get home?  
Option 1: Take an 10 minutes bus, then an 40 minute bus, and finally a 10 minute train.  
Option 2: Take a 90 minutes train, then a 45 minute bike ride, and finally a 10 minute bus.  
Option 1 will take  $10+40+10 = 60$  minutes.  
Option 2 will take  $90+45+10=145$  minutes.  
Since Option 1 takes 60 minutes and Option 2 takes 145 minutes, Option 1 is faster.

Which is a faster way to get to work?  
Option 1: Take a 1000 minute bus, then a half hour train. and finally a 10 minute

## Output

Option 1 will take  $1000+30+10 = 1040$  minutes.  
Option 2 will take  $800+60+30 = 890$  minutes.  
Since Option 2 takes 890 minutes and Option 1 takes 1040 minutes, Option 2 is faster.

▶ Generate

# أوامر سلسلة الفكر بدون أمثلة مباشرة Zero-shot COT Prompting

Incorrect

Prompt	Output
If John has 5 pears, then eats 2, and buys 5 more, then gives 3 to his friend, how many pears does he have?	John has 8 pears.
<div><div></div><div>Generate</div></div>	

# أوامر سلسلة الفكر بدون أمثلة مباشرة Zero-shot COT Prompting

Correct


**Prompt**

If John has 5 pears, then eats 2, and buys 5 more, then gives 3 to his friend, how many pears does he have?

Let's think step by step.

**Output**

John starts with 5 pears. He eats 2 pears, leaving him with 3 pears. He buys 5 more pears, giving him a total of 8 pears. He gives 3 pears to his friend, leaving him with only 5 pears.

 **Generate**



The top left corner of the slide features several overlapping geometric shapes, including a large orange-to-red gradient rectangle, a smaller dark blue triangle, and a larger dark blue hexagon. Other smaller geometric shapes are scattered across the top edge.

# الأمر بطريقة فعّالة Effective Prompting

# تطبيقات مولّدات الأوامر



video script about mindmapping

**SIMPLE PROMPT**

أمر بسيط



Please help me outline a script fo my tutorial video about mindmapping. The concept is about business branding. Only includes the promotional projects.

**AI-GENERATED PROMPT**

أمر مؤلّد باستخدام الذكاء الاصطناعي

# أوامر "تصرّف وكأنك" Act-as Prompts

The "Awesome ChatGPT Prompts" with Outputs. Click on a record to view ChatGPT's answers. You may select a specific prompt from the dropdown menu.

Act as ^

Act as

Academician

Accountant

AI Assisted Doctor

AI Trying to Escape the Box

instructor in a school, teaching algorithms to beginners. You will be giving python programming language. First, start briefly explaining and then continue giving simple examples, including bubble sort and quick sort. Prompt for additional questions. As soon as you explain and give the answer, please include corresponding visualizations as an ascii art whenever

Instructor in a School

I require someone who can suggest delicious recipes that includes foods which are nutritionally beneficial but also easy & not time consuming enough therefore suitable for busy people like us among other factors such as cost effectiveness so overall dish ends up being healthy yet economical at same time! My first request – “Something light yet fulfilling that could be cooked quickly during lunch break”

Chef

<https://www.turbo-gpt.com/chatgpt-prompts-act-as>

# نصائح عامة عن الأوامر

- محاكاة خبير Expert Simulator  
"قم بدور عالم بيانات متمرس وأرشدني في كتابة تقرير يسلط الضوء على أحدث الاتجاهات والتنبؤات في مجال الذكاء الاصطناعي. زودني برؤى فريدة ونصائح الخبراء التي ستساعدني في تقديم محتوى إعلامي عالي الجودة لجمهوري".
- أوامر غير تقليدية Unconventional Prompts  
"اكتب قصيدة عن إنشاء المحتوى"أو "صف شعورك كمسوق رقمي في 10 صفات".

# نصائح عامة عن الأوامر

## • تغيير الأنماط والنبرة Styles and Tones

"قم بتقديم نصيحة [نمط الإدخال] الممكنة حول استخدام LLM لإنشاء محتوى أكثر فعالية في [نبرة الإدخال]."

الأنماط		النبرات	
وصفي	غير رسمي	روح الدعابة	واثق
إقناعي	رسمي	ساخرة	مشكك
روائي	رومنسي	جدية	متفائل
تقني	قوطي Gothic	غاضب	متشائم

# العرض التوضيحي :الأمر بطريقة فعّالة باستخدام ChatGPT

# مقدمة إلى التوليد المعزز للاسترجاع RAG

# النماذج اللغوية الكبيرة الأكثر حداثة



تتفوق النماذج اللغوية مسبقا التدريب مثل GPT و Bart و T5 في العديد من مهام معالجة اللغة الطبيعية:

- تمتلك ثروة من المعرفة العالمية في معاملهم
- يمكنها بسهولة التعامل مع مجموعة واسعة من المهام
- يمكن ضبطها لمهام محددة

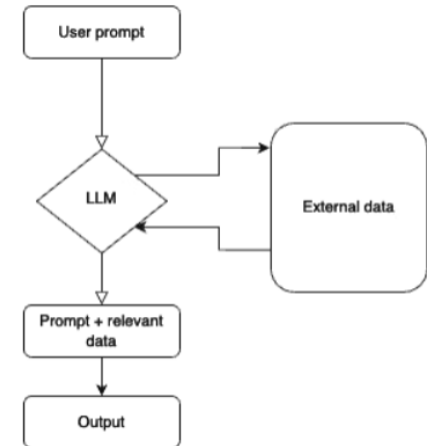
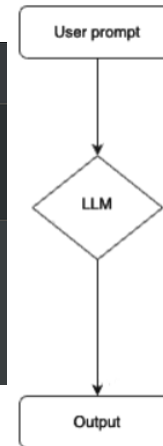


# RAG مقابل LLMs

🔮 Model: GPT-4

N What is the benefits of gpt4 over gpt3.5

As of my knowledge cutoff in September 2021, GPT-4 has not been released by OpenAI. However, with the progression of AI models, we can make some informed predictions about the potential improvements you could expect from GPT-4 compared to GPT-3.5.



# لماذا RAG؟

التوليد المعزز للاسترجاع RAG هو أسلوب معالجة اللغة الطبيعية (NLP) الذي تم إنشاؤه لتخطي بعض القيود المفروضة على LLM:

- توليد معلومات غير صحيحة (الهلوسة)
- تسريب معلومات حساسة
- تطبيق المعرفة الدقيقة
- التحديث إلى المعرفة العالمية الحالية

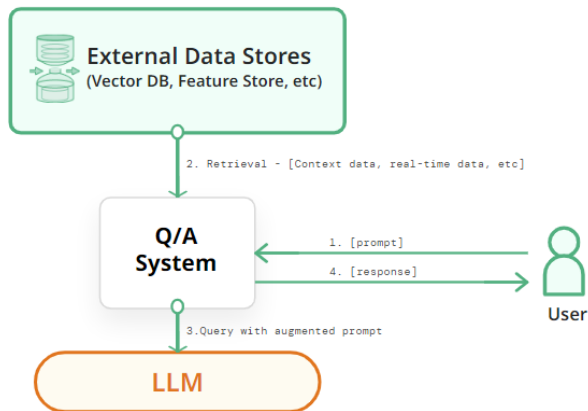
# التوليد المعزز للاسترجاع RAG

# ما هو نموذج RAG؟

لمعالجة القيود، يجمع نموذج التوليد المعزز للاسترجاع بين القدرة على الإنشاء للنماذج اللغوية الكبيرة وتقنيات استرجاع المعلومات:

- يقوم النموذج بالاسترداد والتوليد في وقت واحد
- لا توجد تسميات مطلوبة لاسترداد المعلومات
- يقاوم تسرب المعلومات
- هو الحل العام لأية مهمة seq2seq

ومع ذلك، فإنه لا يزال قيد التطوير.



# أمثلة على نماذج تشبه RAG



einstein



Google Assistant

# مكونات نموذج RAG

تتألف نماذج RAG من 3 مكونات:

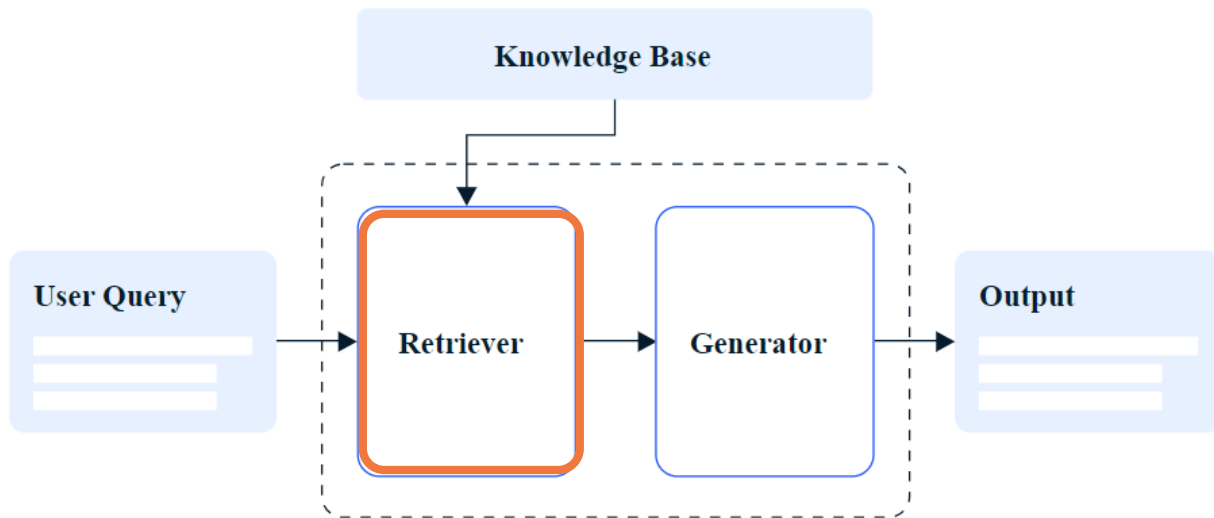
- نموذج لغة تم تدريبه مسبقًا مثل GPT و BART - المولد
- نموذج استرجاع مُدرَّب مسبقًا مثل DPR وتكنولوجيا المعلومات والاتصالات
- قاعدة معرفية للبيانات النصية مثل Wikipedia و YAGO والتغريدات



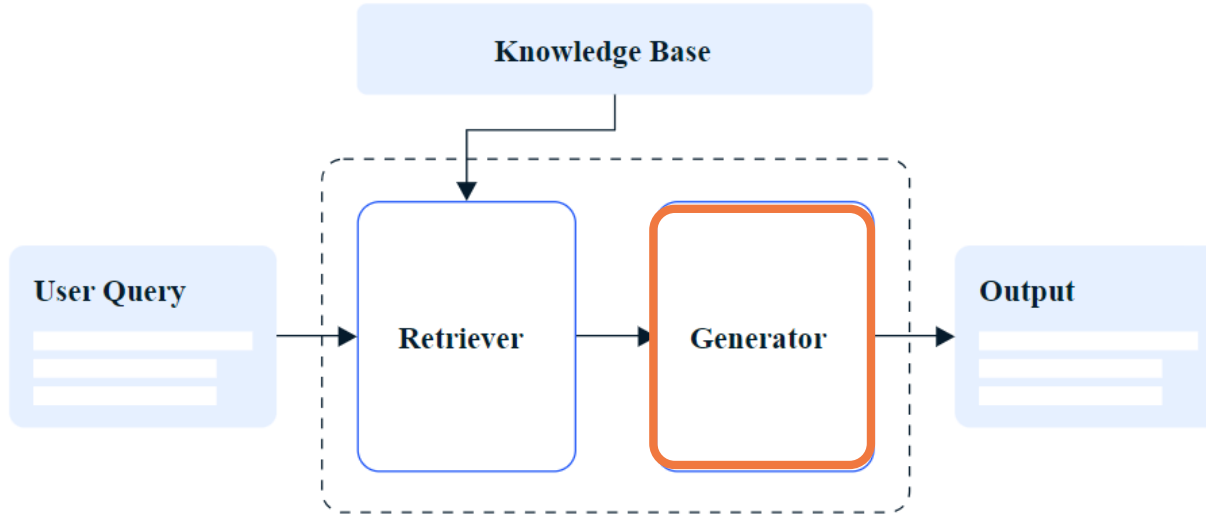
DPR



# المسترد The Retriever

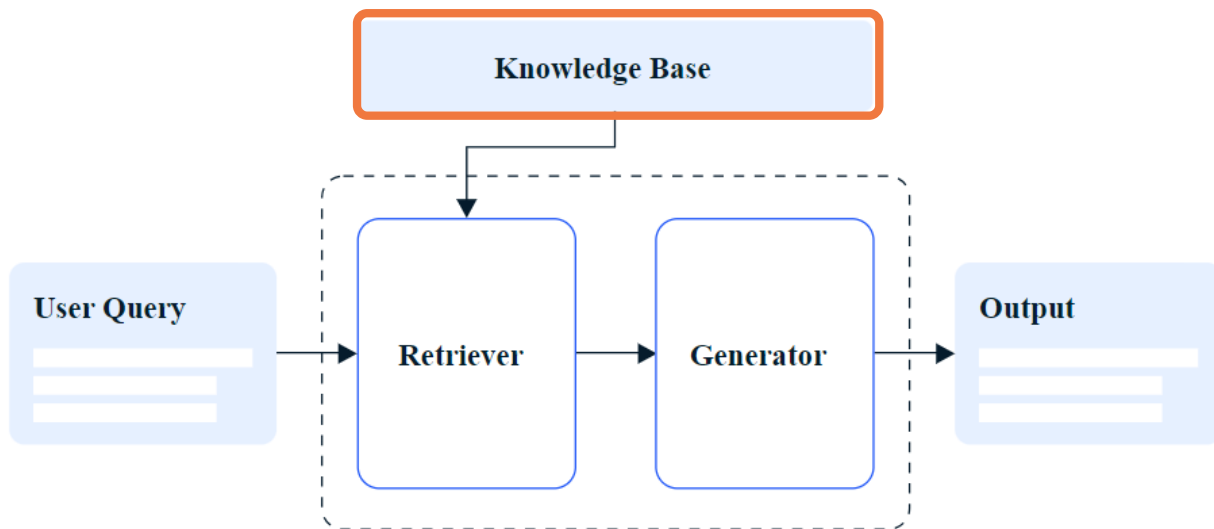


# The Generator المولد

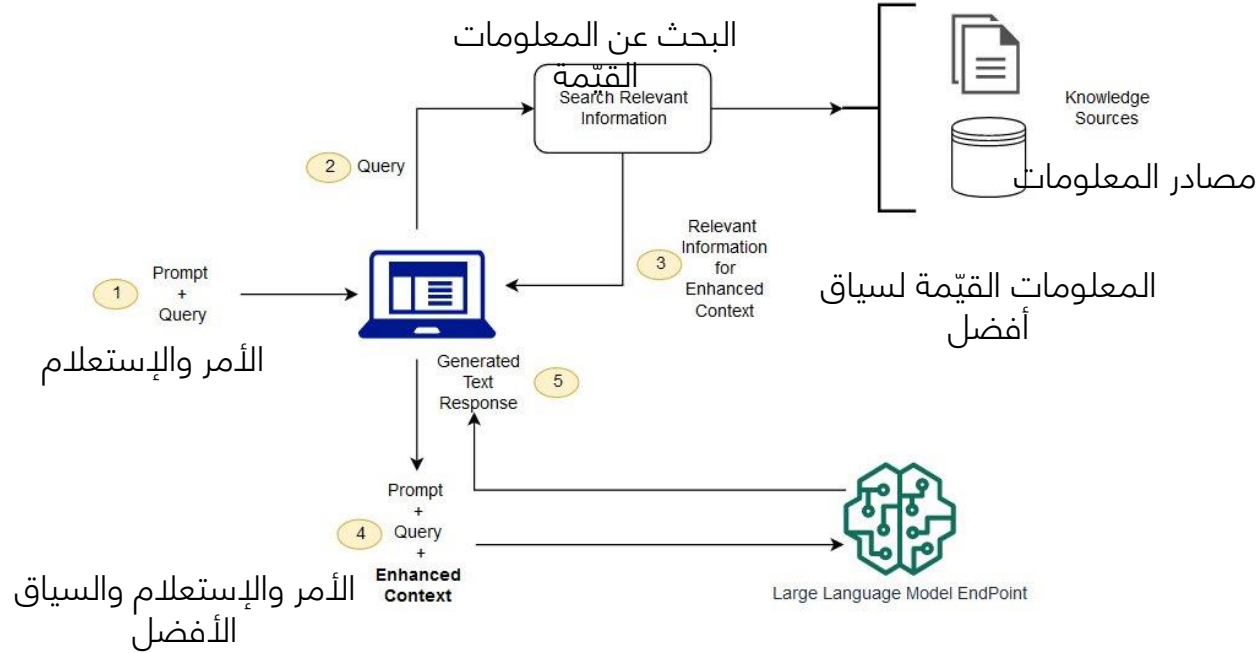




# قاعدة المعرفة



# بنية RAG



<https://docs.aws.amazon.com/sagemaker/latest/dg/jumpstart-foundation-models-customize-rag.html>

# تتفوق RAG في المهام كثيفة المعرفة

توفر RAG أحدث النتائج لضمان الجودة العامة

- الجمع بين نقاط القوة في الذاكرة المعاملية وغير المعاملية
  - الذاكرة المعاملية هي نموذج seq2seq مُدرَّب مسبقًا
  - غير المعاملية هو فهرس المنتج من قاعدة المعرفة التي يتم الوصول إليها بواسطة نموذج الاسترجاع
- معلومات أكثر واقعية ودقيقة من seq2seq فقط

**الإدخال:** أي كوكب في النظام الشمسي لديه أكبر عدد من الأقمار؟

**GPT-3:** اعتبارًا من آخر تحديث لمعلوماتي في سبتمبر 2021، كان كوكب المشتري، وهو أكبر كوكب في نظامنا الشمسي، لديه أكبر عدد من الأقمار. كوكب المشتري لديه أكثر من 80 قمرا معروفا.

**RAG:** زحل هو الكوكب في نظامنا الشمسي الذي يحتوي على أكبر عدد من الأقمار، بإجمالي 146 قمرا معروفا.

# الإيجابيات والسلبيات

كان استرجاع المعلومات موجودًا قبل الLLMs، وله إيجابيات عظيمة:

- يسمح بنتائج دقيقة
- سهل التحديث
- تحصل بعض نماذج الاسترجاع الأحدث على نتائج ممتازة

ومع ذلك، فإن لها سلبيات:

- يحتاج استرجاع المعلومات IR إلى بيانات مصنفة للتدريب
- يجب برمجتها بطرق محددة

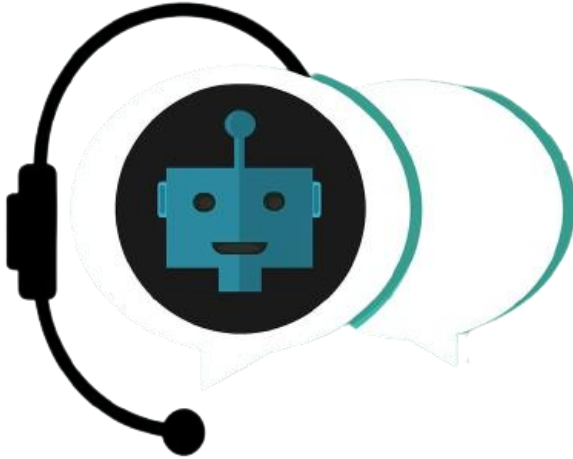
# حالات استخدام نماذج RAG

# العديد من الفرص

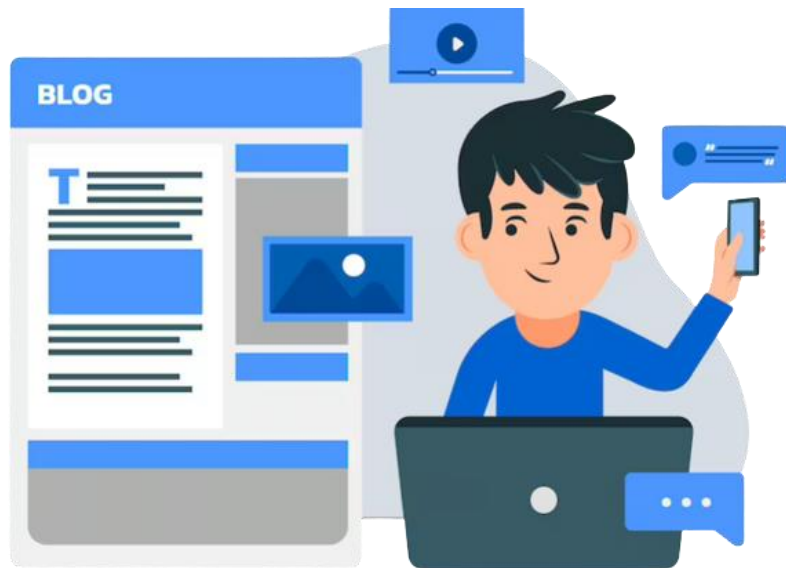
- كما ذكرنا سابقًا، لا تزال نماذج RAG قيد التطوير
- لا تزال المنتجات الحالية لا تستخدم هذه التكنولوجيا
  - لقد تم التنبؤ بالعديد من حالات الاستخدام
  - هناك تحديات عند استخدامها تجاريًا

# روبوتات المحادثة في خدمة العملاء Chatbots

- الاستجابات المحسنة
- قاعدة المعرفة المخصصة
- تستخدم مراجعات العملاء المحدثة



# أدوات إنشاء المحتوى



- سهولة إنشاء محتوى عالي الجودة
  - مشاركات المدونة Blog posts
  - نسخ تسويقية
  - وصف المنتج
- كل ذلك يعتمد على معلومات محدثة وواقعية



# أنظمة توصية المنتج

- توصيات محسنة للعملاء
- باستخدام أحدث المعلومات
  - تاريخ الشراء
  - الاهتمامات
  - التفضيلات



# أدوات التشخيص الطبي

- تشخيصات دقيقة
- إنشاء قائمة بالتشخيصات المحتملة بناءً على:
  - الأعراض الحالية
  - التاريخ الطبي



تشخيص

# التعليم

يمكن لـ RAGS تعزيز تجربة التعليم والتعلم من خلال إنشاء مواد تعليمية مخصصة تعتمد على:

- تحديث البيانات باستمرار من الطلاب
- الاحتياجات والاهتمامات الخاصة لكل طالب



# التحديات الحالية مع نماذج RAG

- لا تزال الشركات والأفراد يكافحون من أجل دمج سياقاتهم الخاصة
- الإطار الأخلاقي لاستخدام هذه النماذج هو مصدر قلق كبير
- حجم السياق وأهميته
- طول الاستجابة وقيود النموذج



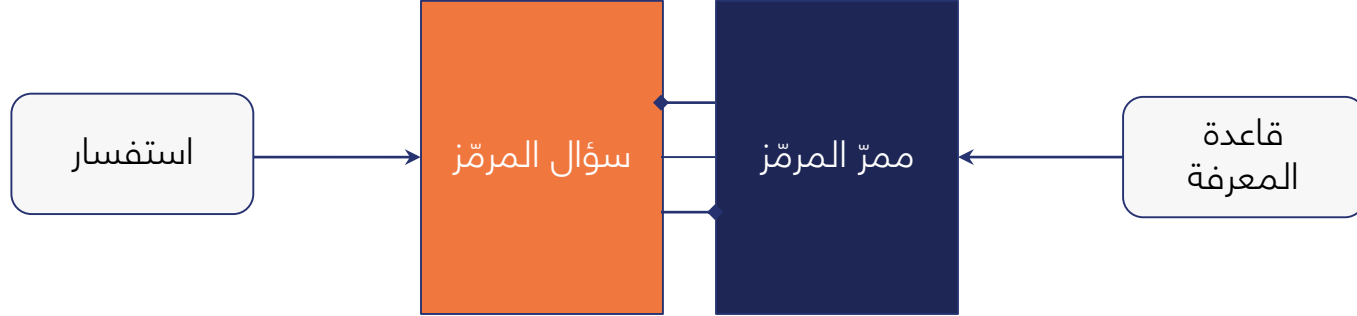
The top corners of the slide feature decorative geometric shapes. On the left, there is a brown-to-orange gradient rectangle and a dark blue triangle. On the right, there is a dark blue triangle and a blue-to-purple gradient rectangle. The background is a solid dark blue with faint, larger geometric shapes scattered across it.

# استرجاع الممر الكثيف

## Dense Passage Retrieval

# المسترد :استرجاع الممر الكثيف DPR

- تعد DPR تقنية رائعة للإجابة على أسئلة المجال المفتوح:
- يستخدم لاحتساب تمثيلات المتجهات الكثيفة للنص
  - يخزن النتائج بطريقة مفهرسة
  - يستخدم نموذج ثنائي الاتجاه (يسمى أيضًا برج سيامية ثنائي Siamese Two-Tower Model)
  - يسترد DPR المقاطع الأكثر صلة من قاعدة الفهرس



# لماذا DPR ؟

- يتمتع DPR بالعديد من الإيجابيات مقارنة باسترجاع المعلومات التقليدية:
- رائع في التقاط أوجه التشابه الدلالية بين المقاطع النصية
  - ومن خلال الحوسبة المسبقة والفهرسة، يحصل نظام DPR على وقت استرجاع أقصر
  - المتجهات الكثيفة غنية بالمعلومات

# الإجابة على سؤال المجال المفتوح؟



- ODQA هو نوع من المهام يختلف عن طلب إجابات مباشرة وموضوعية من النموذج.
- يمكن أن تكون الأسئلة من أي مجال
- لا يحتوي النموذج على تدريب خاص لموضوعات محددة.
- من خلال قياس عدد المرات التي يحصل فيها نموذجك على الإجابات الصحيحة، يمكنك تقييم أدائه في ODQA



# استراتيجيات مفكك الترميز

# طريقتين لفك الترميز

توجد حاليًا طريقتين لفك ترميز الإجابات من نماذج RAG:

- تقنية الوحدة اللغوية (Token) الخاصة ب RAG
- تقنية تسلسل RAG

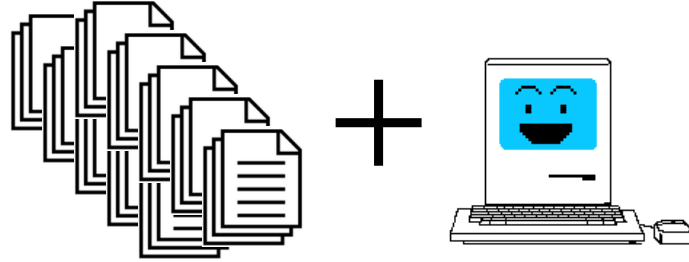
كلاهما يحقق نتائج متشابهة جدًا!

- في DPR، يمكنك الحصول على مستندات Z التي تم جلبها من قاعدة المعرفة KB
- بعد ذلك، تكتشف كيف قامت LLM بإنشاء الوحدات اللغوية (Tokens) الخاصة بها بناءً على مدخلات المستخدم، وكذلك الوحدات اللغوية (Tokens) التي تم إنشاؤها مسبقًا.
- باختصار، كلا الطريقتين تقومان بتلخيص المستندات والمقارنة.

# تفكيك ترميز الوحدة اللغوية ل RAG

بطريقة الوحدة اللغوية الخاصة ب: RAG:

- عند صياغة إجابة، فإنه يعامل كل وحدة لغوية أو كلمة في الإجابة كقرار منفصل.
- بالنسبة لكل وحدة لغوية، يقوم بإعادة حساب الاهتمام بالمستندات المستردة.
- إنه مشابه لآلية الانتباه "على مستوى الوحدة اللغوية المميزة" حيث يتم حساب الاهتمام لكل وحدة لغوية مميزة في التسلسل.
- من المحتمل أن ينتج عن ذلك إجابات تجمع المعلومات من أجزاء مختلفة من المستندات المستردة.



# فك ترميز تسلسل RAG

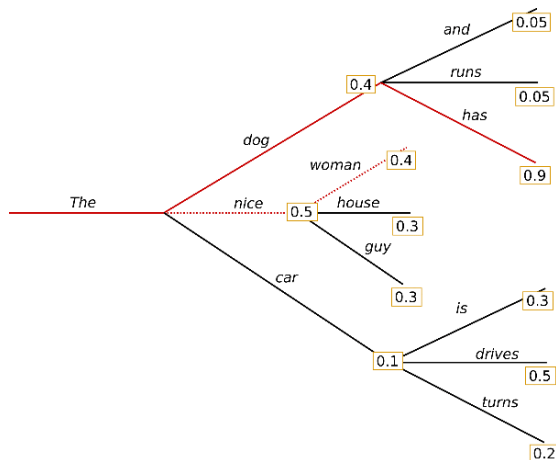
بطريقة تسلسل RAG:

- يأخذ النموذج إحدى المستندات المنفصلة، ويقوم بإنشاء تسلسل كامل.
- يكرّر ذلك لكل مستند
- يضرب احتمالات التسلسل بأكملها معًا، ويجد الإجابة الأكثر احتمالاً
- إنه مشابه لآلية الانتباه "على مستوى التسلسل."

$$\left( \text{Documents} + \text{AI Model} \right) * n$$

# بحث شعاعي Beam Search

- البحث الشعاعي هو تقنية تستخدم في فك الترميز من هذه النماذج
- فهو يساعدنا على استكشاف الاحتمالات المختلفة في توليد الوحدات اللغوية



# شكراً لكم

# Thank you



**SDAIA**

الهيئة السعودية للبيانات  
والذكاء الاصطناعي  
Saudi Data & AI Authority