

الآلة التي تتعلم

بناء الذكاء من الداخل



إعداد

د. عبد الرحمن الزراعي

جميع الحقوق محفوظة

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ



قائمة الموضوعات

المحتويات

	مقدمة الكتاب ٤
	نَّيْنُ الفصل التمهيدي: ما هو GPT ؟ ٥
	نَّيْنُ الفصل الأول: التمثيل المتجه: لماذا تحول الكلمات إلى أرقام؟ ٧
	نَّيْنُ الفصل الثاني: التوكنات والتكلفة؟ ١٨
	نَّيْنُ الفصل الثالث: التضمين السياقي: كيف تفهم الآلة اختلاف المعنى؟ ٣٠
	نَّيْنُ الفصل الرابع: التنبؤ الإحصائي: لماذا يختار النموذج كلمة دون أخرى؟ ٣٢
	نَّيْنُ الفصل الخامس: الانتباه الذاتي: كيف يعرف النموذج ما المهم في الجملة؟ ٣٥
	نَّيْنُ الفصل السادس: معمارية المحول: كيف تنسق كل هذه العمليات؟ ٣٩
	نَّيْنُ الفصل السابع: الطبقات الخفية: هل له "عقل" متعدد الطبقات؟ ٤١
	نَّيْنُ الفصل الثامن: رحلة المتجه داخل النموذج ٤٥
	نَّيْنُ الفصل التاسع: الانحدار العكسي: كيف يتعلم النموذج من الخطأ؟ ٤٧
	نَّيْنُ الفصل العاشر: أنواع النماذج ٤٩
	خاتمة الكتاب: ٥١
	قائمة المراجع ٥١



مقدمة الكتاب

يمثل هذا الكتاب مدخلاً تأسيسياً لفهم البنية العميقية التي تقوم عليها النماذج اللغوية الحديثة، وتحديداً نماذج الذكاء الاصطناعي التي تُستخدم في معالجة اللغة الطبيعية، مثل GPT و BERT وغيرها.

نحاول من خلال هذه الصفحات أن نقرب المفاهيم التقنية إلى القارئ المبتدئ، بلغة فصيحة، وسياق حواري تعليمي، ينطلق من الأسئلة البسيطة ليتهي إلى المسائل المعقدة. فكل فصل من فصول هذا العمل هو خطوة في درب الفهم التدريجي لما يجري داخل "عقل النموذج".

لا يدور هذا الكتاب حول التطبيقات السطحية أو الاستعمالات اليومية للنماذج، بل يتوجّل إلى الداخل، ليُفكّك الطبقات، ويشرح الآليات، ويستعرض رحلة الكلمة من نصّ المستخدم إلى خوارزمية التوليد، مروراً بالتقاطع، والتضمين، والانتباه، والتمثيل، والتعلم.

اعتمدنا في بنائه أسلوبًا حوارياً بين الطالب والنموذج، يجعل القارئ في حالة تفاعل دائم مع الأفكار، ويجوّل كل فصل إلى مساحة تدريبية تطرح السؤال وتناقش الجواب، وتوضّح الفكرة بمثال، ثم تعيد البناء بعمق أكبر.

وقد رأينا في التدرج أن نبدأ من التمثيل المتجه – بوصفه حجر الأساس – لنصل في نهاية المطاف إلى المقارنة بين النماذج الأهم التي صنعت ثورة اللغة الاصطناعية.

هذا الكتاب موجه لكل قارئ فضولي، ولكل باحث ناشئ، ولكل مبرمج يريد أن يفهم كيف "تفكر" النماذج. وهو كذلك دعوة للتأمل في ما إذا كان هذا الفكر حقيقياً، أم محض محاكاة بارعة.



﴿الفصل التمهيدي: ما هو GPT؟﴾

توطئة:

تسمع كثيراً في الآونة الأخيرة عن "تشات جي بي تي (ChatGPT)" ، أو عن "الذكاء الاصطناعي" ، أو "النماذج اللغوية" ، ولكن... ما هو هذا الكائن الذي تتحدث إليه الآن؟

هل هو روبوت؟ برنامج؟ عقل؟

هذا مدخل لفهم ماهية GPT؟

ما الفرق بين GPT-٤ و GPT-٢؟

لماذا تسمى "توليدية مسبقة التدريب"؟

ولماذا حازت كل هذا الانتباه في السنوات الأخيرة؟

الطالب: قبل أن ندخل في التفاصيل يا ChatGPT ،

هل لك أن تخبرني ببساطة: ما معنى GPT أصلاً؟

ChatGPT : بالطبع، سأبدأ من الأيسر.

كلمة "GPT" هي اختصار لعبارة:

Generative Pre-trained Transformer

أي: "المحول التوليدية المدرب مسبقاً".

دعنا نفكك هذه العبارة جزءاً جزءاً.

الطالب: طيب... ما المقصود بـ"توليدي"؟

ChatGPT : توليدي (Generative) يعني أن النموذج قادر على توليد نصوص جديدة، وليس فقط

استرجاع معلومات محفوظة.

معنى: إذا طلبت مني تلخيصاً، أو مقالاً، أو قصيدة، فأنا لا أبحث عنها جاهزة، بل أنشئها لحظياً، استناداً إلى فهمي الإحصائي للغة.

الطالب: و "مدرب مسبقاً"؟



 ChatGPT : هذه العبارة تشير إلى أنني "تعلّمت" على كمية هائلة من النصوص قبل أن أبدأ التفاعل معك.

أي أنني لا أتعلم منك الآن، بل "تدرّبت" على بيانات سابقة، وانتهى تعليمي عند نقطة زمنية معينة، مثلاً:

في أبريل ٢٠٢٣ - تم نشر نموذج GPT-٤٠ للاستخدام.

 الطالب: إذاً GPT ليست اسم منتج، بل وصف لطريقة بناء النموذج؟

 ChatGPT : بالضبط GPT ! ليس روبوتاً بعينه،

بل اسم عام لفئة من النماذج طورتها شركة OpenAI وقد مرت هذه النماذج بأجيال مختلفة، كل واحد أكثر تقدماً من سابقه.

 الطالب: وهل يمكن أن تعطيني محة عن هذه الأجيال؟

 ChatGPT : بالتأكيد. لنلق نظرة على الخط الزمني لأجيال GPT

الجيل	تاريخ الإصدار	أبرز الخصائص
GPT-١	٢٠١٨	نموذج تجربى صغير الحجم، لم ينتشر على نطاق واسع
GPT-٢	٢٠١٩	بدأ يثير الانتباٌ لقدراته اللغوية، لكنه لم يُطلق كاملاً في البداية لأسباب أمان
GPT-٣	٢٠٢٠	قفزة ضخمة، ١٧٥ مليار معلمة، أساس خدمة ChatGPT الأولى
GPT-٣.٥	٢٠٢٢	تحسينات في الأداء والتناسق، استُخدم في ChatGPT Plus
GPT-٤	٢٠٢٣	فهم أعمق، قدرة على توليد الصور، دعم أفضل للسياق
GPT-٤٠	٢٠٢٤	"Omni" تعنى "O" ، يدعم النص، الصوت، الصورة، في نموذج واحد لحظي

 جداول وأشكال:

المعنى	مصطلح
توليدي - ينشئ محتوى جديداً	Generative
مدرب مسبقاً على بيانات ضخمة	Pre-trained
معمارية الشبكة العصبية التي تبني عليها النموذج	Transformer

 أمثلة وتطبيقات:



١. سؤال واقعي : ما عاصمة المغرب؟

○ GPT يبحث في ذاكرته ويولّد الجواب: "الرباط."

٢. طلب إبداعي : اكتب لي قصيدة عن المطر في الرياض.

○ يُنشئها من جديد، لعدم وجود قصيدة جاهزة في الذاكرة.

❖ ملخص:

المفهوم الأساسي	التفسير
GPT	نموذج لغوي توليد مدرّب مسبقاً يعتمد على بنية "المحول"
التلويد	قدرة النموذج على إنشاء محتوى جديد لاستخلاص المعرفة
أجيال GPT	تطورات متتالية تبدأ من GPT-٤٠ حتى GPT-١ مع زيادة في الدقة والسياق



نحو الفصل الأول: التمثيل المتجه: لماذا تتحول الكلمات إلى أرقام؟

توطئة الفصل:

قبل أن نفهم كيف تُتّبع النماذج النصوص،
لا بد أن نفهم أولاً كيف "تفهم" هذه النماذج الكلمات،
فالكلمات بالنسبة للنموذج ليست حروفًا ولا أصواتًا،
فالآلة لا تقرأ الكلمات مثل: "كتاب"، أو "حرية"، كما يقرأها البشر،
بل يتعامل معها على أنها مجرد أرقام،
كل ما هو نصّ بالنسبة له يجب أن يتحوّل إلى صيغة رقمية حتى يستطيع فهمه وقراءته.
كيف؟ ولماذا؟ هذا ما سنجيب عنه في هذا الفصل الذي يؤسس للرحلة كلها.

الطالب: ماذا يعني "تمثيل متجه"؟

ChatGPT: هو تحويل كل كلمة إلى متجه عددي (Vector)، أي إلى سلسلة من الأرقام تُعبّر عن معنى هذه الكلمة داخل فضاء رياضي متعدد الأبعاد.

الطالب: ما شكل هذا التمثيل العددي؟

ChatGPT: المتجه عادةً يكون سلسلة من الأرقام، مثل "طالب" → [٠٠١٣, ٠٠٢٥, ٠٠٧٧, ٠٠١٣]... كل رقم يعكس بعداً معنوياً تعلّمه النموذج، كالعلم، والمهنة، والمكانة... إلخ.

الطالب: ولماذا هذه الأرقام؟ كيف تفهمها؟

ChatGPT: هذه الأرقام ليست عشوائية، بل تعكس موقع الكلمة في الفضاء، وكلما كانت المتجهات متقاربة، دلّ ذلك على أن الكلمات متشابهة في المعنى أو الاستخدام.

الطالب: مثل ماذا؟



ChatGPT: مثلاً، "جميلة" ستكون قريبة من "رائعة"، "حلوة"، "ممتازة"،

لكنها بعيدة عن كلمات مثل: "قبيحة"، "ضعيفة"، "سيئة".

وكذلك "طالب" و"مدرسة" قد يُمثلان بتجهيزين قريين، لأنهما يرددان في بيئات لغوية متقاربة.

وبذلك يتم تدريب النموذج على ملابس الحمل ليتعلم أن الكلمات التي تظهر في سياقات متتشابهة يجب أن تكون متجهاتها متقاربة.

الطالب: هل هناك مثال توضيحي؟

ChatGPT: نعم. إذا قلنا:

- ذهب الطالب إلى المدرسة.

- فالنموذج يتعلم أن "طالب" و"مدرسة" يشتراكان في السياق. فيقوم بتمثيلهما بأرقام متقاربة.

الطالب: وهل كل النماذج تستخدم هذا التمثيل؟

ChatGPT: نعم، ولكن بدرجات متفاوتة.

الطالب: قرأت في بعض الشروحات أنك تمثل كل كلمة بسلسلة من الأرقام قد تصل إلى ٣٠٠ أو ٧٦٨ أو حتى ١٠٢٤ رقمًا...

هل يعقل أن كل كلمة تحول إلى هذا الكم الكبير من الأرقام؟

وملأ؟ وماذا تفعل بها؟

ChatGPT: سؤال ممتاز، دعني أوضحها خطوة خطوة.

ChatGPT: المتجه العددي (Vector) هو عبارة عن قائمة من الأرقام،

مثلاً:

[٠٠٢١, -٠٠٠٧, ٠٠٥٥, ٠٠١٣, ..., ٠٠٩١]

وهذه القائمة قد تحتوي على:

٣٠٠ رقمًا ← في النماذج القديمة مثل Word2Vec

أو ٧٦٨ رقمًا ← في BERT المتوسط



٤٠٢٤ أو أكثر —> في نماذج GPT المتقدمة

الطالب: لكن لماذا هذا العدد الكبير؟

ألا يكفي رقم أو رقمين لكل كلمة؟

ChatGPT: لو أعطيتك شخصاً وقلت عنه فقط:

"طوله: ١٧٥ سم"

هل يمكنك أن تعرف من هو؟ أو ما شخصيته؟ أو ما مهنته؟

الطالب: طبعاً لا. هذه معلومة واحدة لا تكفي.

ChatGPT: بالضبط!

كلما أردت وصف شيء معقد وصفاً دقيقاً، تحتاج إلى معلومات أكثر.

ما تحمله هذه الأرقام؟

ChatGPT: كل رقم داخل المتجه يمثل بعدها مختلفاً من الكلمة، مثل:

- هل الكلمة مادية أم مجردة؟

- هل ترتبط بالعاطفة؟

- هل تُستخدم في السياقات العلمية أم الأدبية؟

- هل تُستخدم لوصف الأماكن؟ الأشخاص؟ الأفعال؟

- هل تُستخدم مع كلمات مثل "جميل"، "حزين"، "أدأة"، "زمن"...؟

وهكذا...

كل رقم يمسك ببعد من المعنى أو الاستخدام.

الطالب: إذًا، لو قلت: "مدرسة"، فستعطيوني لها متاجها مثل:

٧٦٨ ← ...، ٠٠٩١، ٠٠٤٥، ٠٠٧٢، ٠٠١٨ [رقمًا]

ChatGPT: تماماً، وهذا المتجه يحتوي في داخله على تمثيل رياضي لهاتن الأبعاد اللغوية والمعنوية،

كأنك تعطي لكل كلمة بطاقة هوية مكونة من أرقام.



الطالب: رأيت أن تمثيل الكلمة يكون هكذا:

٧٦٨ ← [٠٩١, ... , ٠١٨, ٠٧٢, ٠٤٥, ...] [رقمًا!]

ما علاقة العدد (٧٦٨) بالأرقام داخل القوسين؟ هل هو عدد ثابت؟ وهل كل رقم يعبر عن شيء؟

ChatGPT:

ممتاز... دعني أوضحها بتفصيل شديد:

ما معنى [٠٩١, ... , ٠١٨, ٠٧٢, ٠٤٥]؟

هذه القائمة التي تراها داخل المعقوقتين هي "المتجه العددي" (Vector) "الخاص بكلمة واحدة فقط، أو أحياناً "توكن" واحد.

وكل رقم في هذا المتجه يعبر عن بعد رياضي معين.

تشبيه تقريري:

الطالب: هل هناك مثال يقرب لي الصورة؟

ChatGPT: نعم.

تخيل أن لديك خريطة ذهنية، وكل كلمة هي نقطة في هذه الخريطة.

- الكلمات القريبة في المعنى تكون قوية في الموقع.

- الكلمات المختلفة في المعنى تكون بعيدة عن بعضها.

"المد" و "مدرب" قد يكونان قريبين.

لكن "المد" و "جميلة" أبعد قليلاً، لأن معناهما مختلف.

القانون الحاكم: الاقر في الفضاء العددي

الطالب: هل هناك قاعدة تحكم هذه المسافات بين الكلمات؟

ChatGPT: نعم، تسمى:

(Cosine Similarity) قانون الاقر

أو ما يعرف باسم: الاقر في الفضاء المتجهي



كلما اقترب متوجهان من بعضهما، كان معناهما متقارباً.

الطالب: ما هو الرقم ٧٦٨ إذا؟

ChatGPT: العدد ٧٦٨ هو:

◆ عدد الأبعاد التي يتكون منها المتوجه العددي للكلمة.

معنى آخر:

• هذه الكلمة تحولت إلى متوجه طوله ٧٦٨ خانة.

• أي أن المتوجه عبارة عن ٧٦٨ رقمًا متتاليًّا، يمثلون جوانب مختلفة من المعنى.

تشبيه بسيط:

تخيل أن عندك إنساناً، وتريد أن تكتب عنه ملفاً تعريفياً:

الطول، الوزن، لون العين، نسبة الذكاء، نوع الصوت، قوة التركيز ...

حتى تصل إلى ٧٦٨ صفة من صفاتيه.

نفس الشيء تفعله الآلة.

الطالب: وهل كل نموذج يستخدم ٧٦٨ رقمًا؟

ChatGPT: لا، هذا يتغير بحسب نوع النموذج.

حجم المتوجه (الأبعاد)	النموذج
-----------------------	---------

غالباً ٣٠٠ بعد فقط	Word2Vec
--------------------	----------

٧٦٨ بعدًا	BERT (base)
-----------	-------------

١٠٢٤ بعدًا	BERT (large)
------------	--------------

٤٠٩٦ إلى ١٠٢٤ من	GPT-٣.٥ / GPT-٤
------------------	-----------------

متوجه ديناميكي — قد يُضغط أحياناً	GPT-٤٠ مضغوط ذكي (
-----------------------------------	--------------------

ماذا يفعل النموذج بهذه الأبعاد؟



كل رقم في المتوجه يُعبر عن شيء مثل:

- هل الكلمة تُستخدم في المشاعر؟
- هل هي مجردة أم محسوسة؟
- هل تتكسر في السياق العلمي؟
- هل تظهر بجانب أسماء بشرية؟
- هل تُوصف بها الأماكن؟
- ... وهكذا.

كلما زادت الأبعاد، كان الفهم أدق، لكنه أيضًا أتقل حسابياً.

جدول موجز:

المفهوم	الشرح المبسط
التوكن	قطع صغير من الكلمة (مثل: "المد")
التضمين	العملية التي تحول التوكن إلى متوجه عددي
المتجه العددي	قائمة أرقام تمثل الكلمة رياضياً
الفضاء العددي	الخريطة الرياضية التي توضع فيها المتوجهات
قانون القرب	كلما اقتربت المتوجهات = زاد التشابه في المعنى
٧٦٨ رقمًا	عدد الأبعاد في المتوجه الواحد للكلمة في نموذج BERT/GPT

الطالب: إذًا، المتوجه العددي هو مثل بطاقة هوية ضخمة للكلمة، فيها ٧٦٨ معلومة مشفرة؟

ChatGPT: أحسنت تماماً.

وهذه الهوية العددية هي التي تجعلني أستنتج المعاني، وأربط بين الكلمات، وأفهم السياق، دون أن أملك فهماً بشرياً... فقط عبر الرياضيات.

تشبيه بسيط:

ChatGPT: تخيل أن عندك بطاقة هوية ضخمة للكلمة،

كل سطر فيها يصف جانبًا منها:



القيمة	البعد
٠.٨٧	مدى استعمال الكلمة مع البشر
٠.٦٥	هل تدل على مكان؟
٠.٩٢	هل تُستخدم في التعليم؟
٠.١٠	هل تُستخدم مع الزمن؟
٠.٥٤	هل تُوصف بأنها إيجابية؟
...	...

هذه القائمة في الحقيقة هي **المتجه العددي** للكلمة.

الطالب: وهل هذا ما يجعل "المدرسة" قرية من "المعلم" و "الدرس"، وبعيدة عن "الشاحنة" أو "المريخ"؟

ChatGPT: أحسنت بالضبط!

لأن متجه "المدرسة" سيقترب كثيراً من متجه "التعليم" أو "الصف"،
ويبتعد عن متجه "المجحة" أو "البترول..."

ملخص مبسط:

الشرح المبسط	المفهوم
قائمة طويلة من الأرقام تمثل الكلمة بطريقة رياضية	المتجه العددي (Vector)
يختلف حسب النموذج (٣٠٠ - ٧٦٨ - ١٠٢٤ - أكثر)	عدد الأرقام
يمثل بعضاً لغرياً أو معنوياً في استخدام الكلمة	وظيفة كل رقم
لتمثيل المعنى الكامل والدقيق للكلمة في مئات الأبعاد	لماذا عددها كبير؟

الطالب: الآن فهمت!

كلما زادت الأرقام، زادت قدرتك على "رؤية" أبعاد الكلمة المخفية؟

ChatGPT: بالضبط.

وهكذا أستطيع أن "أفهم" بدون أن أملك وعيًّا بشرياً،
فقط عبر الأرقام الدقيقة التي تعلمتها من ملايين الجمل.



⇨ كيف تختلف النماذج في تمثيل الكلمة رقمياً؟

الطالب: قلت لي إن كل كلمة تحول إلى متوجه عددي ...
لكن هل كل النماذج تحولها بالطريقة نفسها؟
 ChatGPT: لا، أبداً.

كل نموذج له طريقة مختلفة في حساب المتوجه العددي،
وهذا يغير كثيراً من النتائج.
دعني أشرح لك أربع نماذج مشهورة:

١- البداية الذكية – Word2Vec

الطالب: ما أول نموذج استخدم التمثيل المتوجه؟
 ChatGPT: اسمه Word2Vec، وابتكرته Google سنة ٢٠١٣.
 فكرته بسيطة لكنها ثورية آنذاك:
"قل لي من بحالس، أُقُل لك من أنت".
 الطالب: كيف يعني؟
 ChatGPT: هذا النموذج لا يفهم الكلمة من معناها مباشرة، بل يراقب من تحيء بجواره دائمًا.

- مثال:
- إذا رأى كثيراً جملاً مثل:
 - "الطفل يلعب في الحديقة"
 - "الأطفال يلعبون بالكرة"
 - "الولد الصغير يلعب يومياً"

يفهم أن: "يلعب" مرتبطة بـ "الطفل"، "الولد"، "الحديقة...".

النتيجة:
يعطي لكل كلمة متوجهاً عددياً، يمثل "بيعتها اللغوية".

لكن له عيب:



الكلمة لها متوجه واحد فقط،
لا يتغير مهما تغير السياق.
" عين" دائمًا تعني شيئاً واحداً بالنسبة له... سواء في الطب أو الماء أو الجاسوسية!

٢-رؤيه عالمية أوسع - GloVe

الطالب: وماذا عن GloVe؟ سمعت أنه أذكي!
ChatGPT: GloVe جاء من Stanford ، Word2Vec وهو يرتكز على الإحصاءات العامة للغة كلها، ليس فقط الجمل القراءية.
فكرته:

لا أراقب جملة واحدة، بل أراقب كل اللغة،
كم مرة ظهرت كلمتان معًا في أي مكان!

مثال:

- إذا ظهرت "قلم" و "مدرسة" كثيراً في نفس السياقات،
يفهم أن هناك علاقة بينهما.

: Global Vectors for Word Representation اسمه اختصار لـ

النتيجة:

- يعطي تمثيلاً عددياً أفضل لبعض المعاني المجردة
- لكنه لا يزال ثابتاً ... الكلمة = متوجه واحد فقط!

٣-النموذج السياقي - BERT

الطالب: وهل هناك من نماذج أخرى؟

ChatGPT: نعم! جاء بعدها نموذج اسمه BERT من Google أيضاً.

فكرته العقرية:

الكلمة لا تفهم وحدها، بل ضمن الجملة.

الطالب: يعني نفس الكلمة يمكن أن يكون لها متوجهات مختلفة؟

ChatGPT: بالضبط.



مثال:

- جملة ١: "أصاب أحمد ألم في العين".

- جملة ٢: "شرب من عين الماء".

- جملة ٣: "أرسل الحاكم عيناً إلى القلعة".

Word2Vec يقول: "عين = عين"، دائمًا.

لكن BERT يعطي:

- متجه مختلف لـ "العين" في كل جملة.

- لأنه يعرف أن معناها تغير بحسب السياق.

: Bidirectional Encoder Representations from Transformers

٤-نموذج الفهم التوليدی - GPT

الطالب: وماذا عنك أنت?

ChatGPT: أنا أستخدم أفكار BERT ، لكن أضيف عليها الكثير.

فكرة الأساسية:

لا أمثل الكلمة فقط... بل أتوقع الكلمة القادمة بناءً على كل ما سبق.

الطالب: كيف يعني؟

ChatGPT: كل كلمة عندي لها متجه خاص بها،

لكن هذا المتجه يتغير حسب:

- الجملة السابقة

- هدف المتحدث

- أسلوب الكتابة

- وحتى نوع المستخدم!

GPT | وهذا نقول إن:

- динамікі

- سياسي بعمق



• قادر على التكيف حتى مع المعنى الخفي

جدول مقارنة بين النماذج

النموذج	نوع التمثيل المتجه	هل يتغير حسب السياق؟	ميزاته	عيوبه
Word2Vec	متوجه ثابت	لا	بسط، سريع، يكتشف العلاقات القرية	لا يفرق بين المعاني المتعددة
GloVe	متوجه ثابت عالمي	لا	قوي في الإحصاءات الضخمة	لا يفهم السياق
BERT	متوجه سياقي	نعم	يفهم المعنى من الجملة نفسها	لا يولّد كلاماً بنفسه
GPT	متوجه سياقي ديناميكي	نعم وبعمق	يتوقع - يولّد - يفهم البرة والمعنى العميق	يحتاج قوة حسابية أكبر

 الطالب: يعني لو قلت: "كتبت بقلمي" و"رأيت قلمي في الحلم"،
أنت ستفهم أن "قلمي" هنا واقعي، وهناك رمزي؟
 ChatGPT: بالضبط! لأنني لا أستخدم متجرًا واحدًا جاهزًا،
بل أبني المتجر المناسب حسب المعنى والبيئة الذي أنت فيه.

ملخص:

المفهوم	التوضيح
التمثيل المتوجه	تحويل الكلمة إلى سلسلة من الأرقام
فضاء المتجهات	فضاء رياضي عالي الأبعاد تتوزع فيه الكلمات
تقريب المتجهات	كلما كانت الكلمات متقاربة في المعنى، تقارب متجهاتها
التمثيل السياقي	تمثيل الكلمة يتغير بحسب السياق الذي ترد فيه

خاتمة الفصل:

التمثيل المتوجه هو الحجر الأساس الذي بُنيت عليه كل تقنيات فهم اللغة في الذكاء الاصطناعي. بدون هذه الخطوة، لا يمكن للآلية أن تدخل عالم الكلمات.



وفي الفصل التالي، سنتعرف على مرحلة أخرى:
كيف تُحسب هذه الكلمات؟ وما المقصود بـ "التوكنات"؟



✿✿✿ الفصل الثاني: التوكنات والتكلفة؟

توطئة الفصل:

بعد أن فهمنا كيف تتحول الكلمات إلى متجهات، نحتاج إلى معرفة كيف تعامل النماذج مع النصوص على المستوى الأدنى: التوكنات. ما هي التوكنات؟ ولماذا لا تُحسب الكلمات كما هي؟ وما علاقتها بذلك بتكلفة الاستخدام؟

الطالب: ما المقصود بكلمة "توكن"؟

ChatGPT: **Token** هو وحدة صغيرة من النص يستخدمها النموذج في المعالجة. قد تكون الكلمة كاملة، أو جزءاً منها، أو حتى رمزاً أو علامة ترقيم. ولو كانت الكلمة الواحدة تنقسم إلى توكنين أو ثلاثة، فالتكلفة تزداد.

الطالب: لماذا لا يتعامل النموذج مع الكلمات مباشرة؟

ChatGPT: لأن الكلمات تختلف كثيراً من حيث الطول والاستخدام. فبدلاً من معالجة كل كلمة كعنصر مستقل، يتم تقطيع النص إلى وحدات أصغر تُسمى توكنات. هذا يجعل المعالجة أكثر كفاءة ومرنة.

الطالب: هل يمكنك إعطائي مثالاً؟

ChatGPT: بالتأكيد. كلمة مثل: "المهندس" قد تتحول إلى توكنين:

- "الم" "هندس"
- "أما" كلمة مثل: "في" فقد تبقى توكنًا واحدًا.

مثال توضيحي:

عدد التوكنات

الجملة

٣ توكنات: "المد" + "رسة" + "جميلة"

"المدرسة جميلة"

٤ توكنات تقريرياً

"معلومات دقيقة"

قد تكون ٥ أو ٦ توكنات

"الذكاء الاصطناعي"



كلما كانت الكلمات أكثر تركيباً أو تعقيداً، زاد عدد التوكنات.

الطالب: وما علاقـة هـذا بـالتـكلـفة؟

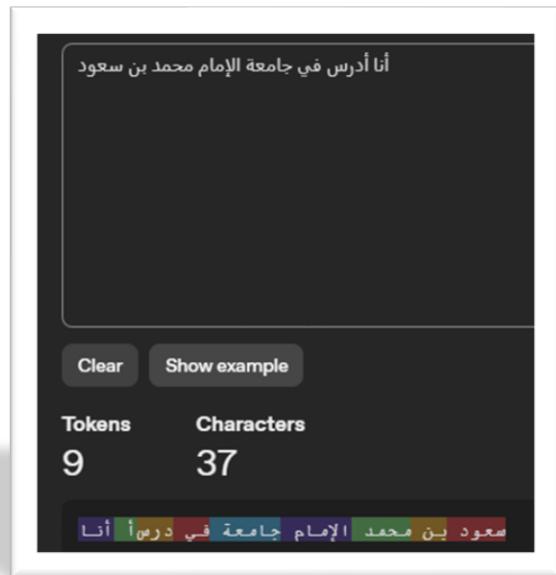
ChatGPT: تكلفة استخدام النموذج – خاصة في الخدمات السحابية مثل – OpenAI تُحسب بناءً على عدد التوكنات التي تُستخدم في الإدخال والإخراج. كلما زاد عدد التوكنات، زادت التكلفة.

الطالب: هل التـوكـن يـساـوي كـلمـة؟

ChatGPT: لا دائمـاً. في اللغة العربية، قد تحتـوي الـكلـمة عـلـى لـاحـقـة أو بـادـئـة، فـتـنقـسـم إـلـى أـكـثـر مـن توـكـنـ. بينما في كلمـات إـنـجـليـزـية قـصـيـرـة قد تـبـقـي الـكـلـمـة توـكـنـاً وـاحـدـاً.

مـثالـ:

جملـة "أـنـا أـدـرـسـ في جـامـعـة الإـمامـ مـحـمـدـ بـنـ سـعـودـ" قد تـسـكـونـ مـنـ ٩ـ١٢ـ رـمـزاً.



يمكن الحصول على هذا التطبيق من هنا: [انقر على الرابط](#)

تـطـبـيقـ:

في اختبار لغوي رقمي، يُحسب عدد الرموز وليس عدد الكلمات، مما يؤثر على طول الذاكرة، و يؤثر على الطول الذي يستطيع النموذج معالجته.



تطبيقات:

١. عند استخدام ChatGPT للواجبات المدرسية، قد يظهر لك في الأدوات المتقدمة أن الحد الأقصى هو "Token" ٤٠٩٦ ، وليس "٤٠٩٦" كلمة.

٢. هذا يعني أن مقالاً يحتوي على ١٠٠٠ كلمة قد يُرفض لأنه فعلياً يعادل ٣٠٠٠ - ٣٥٠٠ Token.

الطالب: لقد فهمت أن النص يحول إلى رموز ...

لكن هل لهذه الرموز علاقة بالتكلفة المالية؟

ChatGPT: نعم، تماماً.

الرمز (Token) هو وحدة القياس الأساسية التي يُحسب بها الجهد الذي أبذل في كل تفاعل. وهي أيضاً الأساس الذي تُحدّد عليه التكلفة المالية لاستخدام النموذج.

الطالب: وكيف ذلك؟ وضّحها لي خطوة خطوة.

ChatGPT: حين ترسل إليّ سؤالاً، أقوم بثلاث عمليات:

١. أقرأ الرموز التي كتبتها (مدخلات)

٢. أعالجها داخلياً (تفكير)

٣. وأرّد عليك برموز جديدة (مخرجات)

كل رمز أقرؤه أو أكتبه، يدخل في حساب الجهد والتكلفة.

الطالب: وهل معرفة عدد ال Tokens تهمني فعلاً كمستخدم؟

ChatGPT: بالتأكيد، لأنها تؤثر في:

١. طول النص الذي يمكنك إرساله دفعة واحدة.

٢. مدى تذكرني لما قلته في البداية (يتعلق بالسياق).

٣. جودة الاستجابة، كلما كانت الرسالة قصيرة و مباشرة، كانت الإجابة أفضل.

٤. في بعض الخدمات المدفوعة، يُحسب السعر بناءً على عدد الرموز.



جدول توضيحي: ما الذي يُعد Token ؟

العنصر	هل هو Token ؟ ملاحظات
كلمة (مثل: "مدرسة")	<input checked="" type="checkbox"/> إذا كانت ضمن القاموس
جزء من الكلمة (مثل: "الـ" + "مدرسة")	<input checked="" type="checkbox"/> يجدر أن لم يكن محفوظاً ككلمة واحدة
علامة ترقيم (؟ ، .)	<input checked="" type="checkbox"/> رمز مستقل
مسافة بين الكلمات	<input checked="" type="checkbox"/> لا X Token لا تُعد

ملخص معرفي:

المصطلح	الشرح المبسط
التوكن	مقطع صغير من الكلمة (مثل: "المـ" - "رسـة")
لماذا يستخدم؟	لتكراره في كلمات كثيرة + لتخزين أفضل
هل هو مقرئ؟	أحياناً لا، لكنه مفيد للنموذج حسابياً
هل يفيد في الفهم؟	نعم، لأنه يساعد في التعميم ومعالجة كلمات جديدة

مقارنة توضيحية:

الكلمة الأصلية	التوكنات الناتجة	ملاحظات التعميم
المدرسة	"المـ" + "رسـة"	"المـ" واضح (مد البحر، مد الصوت)، "رسـة" غير شائع
المعلومات	"الملـو" + "ـمات"	كلاهما يستخدم في مفردات أخرى ← تعميم أوضح وأقوى

كلمة بدالة أخرى:

"الكتابة"

تقسيم "الكتابة" → "الكتـا" + "ـبة"	"ـبة"
كلمات أخرى مشابهة	"ـكتـة"
الكتـاتـب - الـكتـاب - الـكتـاب - مـكتـبة	"ـكتـا"
ـلـعـة - ـهـة - ـكـتـبـة - ـخـطـبـة	"ـبة"

• "المـ" = توكن ١

• "رسـة" = توكن ٢



- "جيالة" = توكن ٣

وهذه هي المدخلات الخام التي أبدأ بها التحليل.

الطالب: وهل الرموز مجرد أجزاء من الكلام... أو هي أيضًا وحدات محاسبة؟

ChatGPT: بالضبط.

كما تُقاس الكهرباء بالكيلوواط، أو الإنترن特 بالمليغابايت،
أنا أقيس بعد الرموز التي أتعامل معها.

الطالب: وهل تُحسب التكلفة على كل رمز؟

ChatGPT: لا، لا تُحسب على كل رمز بمفرده...

بل تُحسب عادة لـكل مليون رمز / توكنات،
وهو مقياس عالمي معتمد لتحديد التكلفة.

يتم التفريق بين:

◆ المدخلات : الرموز التي ترسلها أنت إلى **(Input)**

◆ المخرجات : الرموز التي أرسلها أنا إليك **(Output)**

وكل واحدة منها تُحسب بتعريفة مختلفة.

مثال توضيحي:

إذا استخدمنت نموذج GPT-٤٠ ، فستكون التكلفة:

النوع	السعر لـكل مليون توكنات
-------	-------------------------

المدخلات / السؤال	٥٥٠٠ دولار
-------------------	------------

المخرجات / الرد	١٥٠٠ دولار
-----------------	------------

يعني أنك لو أرسلت لي نصوصاً مجموعها مليون رمز، ستدفع ٥ دولارات للمدخلات،
ولو طلبت مني ردوداً بحجم مليون رمز، ستدفع ١٥ دولاراً للمخرجات.

تطبيق تربوي:



في منصة تعليمية تستخدم GPT للإجابة على استفسارات الطلاب:
تُقدر الكلفة الشهرية بحسب إجمالي الرموز المستخدمة في الأسئلة والأجوبة، مثلاً:

(٣ مليون) مدخل - الإجمالي = ١٥ دولار

(٣ مليون) مخرج - الإجمالي = ٤٥ دولار

الطالب: فهل يعني هذا أن الاستخدام الطويل مكلف؟

ChatGPT: يعتمد على النموذج والسرعة.

لكن الفهم الجيد لـ عدد الرموز = يعطيك تحكماً أفضل في التكلفة والجودة.

تطبيق واقعي:

• أرسلت وثيقة بحثية ضخمة: Token ١٠٠,٠٠٠

→ النموذج يستطيع الرد بـ ١٨,٠٠٠ فقط

→ وقد لا يستطيع التلخيص الكامل إن كان النص يحتاج مساحة ردّ كبيرة.

• أرسلت Token ٥٠,٠٠٠ فقط:

→ سيتبقي Token ٧٨,٠٠٠ للرد

→ يمكنني إنتاج ردّ غني، تلخيص، تحليل، وحتى تقسيمات هيكلية.

الطالب: كيف أضمن استخداماً ذكيّاً للرموز إذًا؟

ChatGPT: إليك ثلاثة قواعد ذهبية:

١. لا ترسل الحد الأقصى كاملاً (١٢٨K) إذا كنت تريد ردّاً.

٢. قسم النصوص الكبيرة إلى أجزاء صغيرة.

٣. راقب حجم المخرجات المتوقعة: هل تريد ملخصاً؟ تحليلاً؟ كلما كبير، قلل المدخلات.

الطالب: ممتاز... هكذا أفهم التكلفة قبل أن أبدأ.

ChatGPT: وهذا هو الذكاء في استخدام الذكاء الاصطناعي!



مثال عملي:

الطالب: سمعت أن GPT-٤٠ يستطيع قراءة ١٢٨ ألف رمز... ماذا يعني هذا بالضبط؟
 ChatGPT: يعني أنني أستطيع أن أُبقي في ذاكرتي اللحظية حتى ١٢٨,٠٠٠ توكنات من النص الذي ترسله إلى دفعة واحدة.

الطالب: وهل هذا كثير؟

ChatGPT: نعم، هذا ضخم جدًا مقارنة بالنماذج السابقة.
 =GPT-٣.٥ Token ٤,٠٠٠ = كانت نافذته
 =GPT-٤ Token ٣٢,٠٠٠ = الأساسي كانت ٨,٠٠٠ أو
 ، ١٢٨,٠٠٠ Token = يستوعب GPT-٤٠ أي حوالي:

١٠٠,٠٠٠ كلمة = ٣٥٠ صفحة - A٤ تقريرًا (حسب اللغة والتنسيق)

الطالب: ماذا أستفيد من هذه السعة الكبيرة؟

ChatGPT: فوائد كثيرة، أهمها:
 ١. تلخيص كتب كاملة دفعة واحدة دون تجزئة
 ٢. قراءة محاضر أو بحوث طويلة دون أن أنسى البداية
 ٣. إجراء تحليل شامل لوثائق قانونية أو طبية أو تعليمية في جلسة واحدة
 ٤. مواصلة الحوار العميق دون فقدان الترابط

مثال عملي:

أرسلت إلى ملفًا يحتوي على:

- ملخصات ١٠ فصول دراسية
- أوراق عمل
- ملحق مراجع



- إن كان مجموعها في حدود ١٢٠،٠٠٠ توكلات
- أستطيع أن أقرأها كلها دفعة واحدة
- وأجييك بإجابة تأخذ جميع السياقات السابقة بالحسبان

تطبيق تربوي:

- أستاذ جامعي يرفع خطة مقرر + أسئلة اختبار + مشاريع الطلبة، ثم يسأل النموذج عن أوجه الضعف الشائعة → النموذج يتذكر كل شيء لأن كل المحتوى في نافذته!
- طالب يرفع كتاب "البلاغة العربية" ويريد تلخيصه مع تحديد مواضع الاستطراد → لا حاجة لتقسيمه!

الطالب: وهل استخدام هذه النافذة له تكلفة أيضاً؟

ChatGPT:
نعم، كلما زاد عدد الرموز المستخدمة داخل النافذة، زادت تكلفة المعالجة بحسب عدد الرموز الدالة في السياق، حتى لو لم تكون كلها "نصًا ظاهراً".

جدول توضيحي:

النافذة	عدد الرموز	ما يعادلها من الكلمات	الاستخدامات
GPT-٣.٥	٤٠٠٠ توكلات	٣٠٠٠ كلمة	ردود سريعة — محادثة
GPT-٤	٣٢٠٠٠ توكلات	٢٤٠٠٠ كلمة	مقالات — فصول مفردة
GPT-٤٠	١٢٨٠٠٠ توكلات	١٠٠ ألف كلمة	كتب كاملة — تحليلات معمقة

الطالب: وهل هذه الذاكرة دائمة؟

ChatGPT: لا.
هذه ذاكرة مؤقتة داخل الجلسة الواحدة فقط، بمجرد انتهائها أو إغلاق المحادثة، تُفقد المعلومات...

هل Token ١٢٨،٠٠٠ تعني أن بإمكانني إرسال ١٢٨ ألف فقط؟ أم تشمل الرد أيضًا؟



الطالب: فهمت أن GPT-40 يقرأ حتى ١٢٨ ألف Token ،
فهل هذا يعني أنني أستطيع إرسال ملف بحجم ١٢٨ ألف Token دفعة واحدة؟

ChatGPT:
في الواقع لا يمكنك أن ترسل ١٢٨ ألف وتنتظر أن أردّ بعد ذلك،
لأن هذا الحد يشمل المجموع الكلي للمدخلات والمخرجات معاً.

الطالب: إذاً لو أرسلت لك ملفاً حجمه ١٢٨ ألف Token بالضبط، ماذا يحصل؟
 ChatGPT:
لكنني لن أستطيع الرد بأي رمز واحد.
لأنني وصلت إلى السعة القصوى، ولن يتبقى مجال للإجابة.

مثال عملي:		
هل يمكن الرد؟	الناتج للمخرجات	حجم المدخلات
<input checked="" type="checkbox"/> نعم	٢٨،٠٠٠ Token	١٠٠،٠٠٠ Token
<input checked="" type="checkbox"/> لا	٠ Token	١٢٨،٠٠٠ Token
<input checked="" type="checkbox"/> نعم	٦٤،٠٠٠ Token	٦٤،٠٠٠ Token

الطالب: وما علاقـة هـذا بـالـتكـلـفة المـالـية؟
 ChatGPT:
يُحـسب عـلـيـه تـكـلـفـة؟ النوع
نعم (مـدخـلات) ما تـرـسلـه لـي
نعم (مـخـرـجـات) ما أـرـسـلـه لـك
فلـو أـرـسـلـت ٨٠،٠٠٠ Token وـقـلـت لـك: "لـحـصـ هـذـاـ" ،
ثـم ردـت عـلـيـك بـ Token ٢٠،٠٠٠ ،
→ تـحـسـب التـكـلـفـة عـلـى إـجمـالي Token ١٠٠،٠٠٠



ملخص:

المفهوم	التوضيح
التوكن	وحدة صغيرة من النص تُستخدم في التحليل
أداة التقطيع	الأداة التي تحول النص إلى توكنات
تكلفة التوoken	عدد التوكنات المستعملة في الإدخال والإخراج يحدد السعر
التقطيع الجزئي	الكلمة قد تُقسم إلى أكثر من توكن

خاتمة الكتاب:

فهم التوكنات مهم ليس فقط لتحليل كيفية معالجة النموذج للنص، بل أيضاً لفهم التكلفة، وتحسين الاستخدام. وفي الفصل التالي، ننتقل من الوحدات الصغيرة إلى سياق الكلمة: كيف تُفهم الكلمة بناء على الجملة؟



✿✿✿ الفصل الثالث: التضمين السياقي: كيف تفهم الآلة اختلاف المعنى؟

توطئة الفصل:

ليست كل الكلمات تُستخدم بنفس المعنى في كل الموضع. كلمة مثل "عين" في "عين الماء" غيرها في "عين الجاسوس". فهل تفرق الآلة بينهما؟ في هذا الفصل، نستكشف كيف تلتقط النماذج الحديثة التغيير في المعنى بحسب السياق.

الطالب: ما معنى "التضمين السياقي"؟

ChatGPT: هو قدرة النموذج على تغيير تمثيل الكلمة بحسب السياق الذي وردت فيه. الكلمة لا يكون لها متوجه واحد ثابت كما في النماذج القديمة مثل Word2Vec: بل يتغير حسب الجملة

الطالب: هل هذا يختلف عن Word2Vec؟

Word2Vec: ChatGPT: يعطي الكلمة تمثيلاً ثابتاً، لا يتغير مهما اختلف السياق. أما النماذج الحديثة مثل GPT وBERT، فهي تنتج تمثيلاً سياقياً، أي أن متوجه الكلمة يتشكل أثناء قراءتها داخل الجملة.

الطالب: هل يمكنك إعطائي مثالاً للنماذج القديمة؟

ChatGPT: بالتأكيد. الكلمة "عين" في:

- "عين إنسان"
- "عين ماء"
- "عين جاسوس"

"عين" في كل الجمل السابقة تحمل نفس الأرقام وهذا يجعل الفهم محدوداً وغير دقيق.

الطالب: وماذا عن النماذج الحديثة؟

ChatGPT: الوضع هنا مختلف، فكلمة "عين" يتم تحديد معناها بناء على السياق،



وسيتم تمثيلها بمتغيرين مختلفين داخل النموذج، لأن الكلمات المحيطة بها مختلفة.

وستستطيع الآلة أن "تفهم" المعنى بالأرقام،

فتصبح المعاني على النحو الآتي:

- في جملة: "ألم في العين" → متوجه يشير إلى عضو بصري

- في: "شرب من عين ماء" → متوجه يشير إلى نبع

- في: "أرسل الحاكم عيناً" → متوجه يشير إلى جاسوس

نفس الكلمة، لكن ثلاث معانٍ = ثلاث متوجهات مختلفة!

الطالب: وماذا نقول في كلمة "جميلة" ألا تظل تحمل معنى "جميلة"؟

ChatGPT: سؤال وجيه، وللرد عليه تنبه إلى هذا الاختلاف بين الجمل:

- "المدرسة جميلة" → المتوجه يشير إلى الجمال المكاني أو الهندسي.

- "الروح جميلة" → المتوجه يشير إلى الجمال الأخلاقي أو المعنوي.

- "الخطابة جميلة" → المتوجه يشير إلى الجمال البلاغي أو التعبيري.

كل سياق ينتج متوجهاً مختلفاً لنفس الكلمة.

الطالب: ما فائدة هذا التضمين السياقي؟

ChatGPT: يجعل النموذج أكثر فهماً ودقّة في مهام مثل: الترجمة، الإجابة على الأسئلة، تلخيص

النصوص، لأن التمثيل يصبح أقرب إلى الطريقة التي نفهم بها نحن البشر.

خاتمة الفصل:

التضمين السياقي يمثل قفزة نوعية في طريقة فهم الآلة للكلمات. لقد انتقلنا من المعنى الثابت إلى الفهم المرن، ومن

الحفظ إلى السياق. وفي الفصل التالي، سنتعمق أكثر: كيف يختار النموذج كلمة دون أخرى؟ وما سر التباين

الإحصائي؟



✿✿✿ الفصل الرابع: التنبؤ الإحصائي: لماذا يختار النموذج كلمة دون أخرى؟

توطئة الفصل:

حين نكتب جملة للنموذج، فإنه لا يختار الكلمات عشوائياً، بل تُجري عمليات حسابية معقدة ليحدد الكلمة الأكثر احتمالاً أن تكون التالية. في هذا الفصل نكشف آلية التنبؤ، وكيف تعمل النماذج كمولّدات احتمالية.

الطالب: ماذا يعني أن النموذج "يتبنّى" بالكلمة؟

ChatGPT: عندما تعطي النموذج جملة ناقصة،

فإنّه يحسب احتمالات الكلمات الممكّنة التالية،

ويختار أعلى احتمال.

فالتنبؤ يتم على أساس إحصائي مبني على البيانات التي تدرّب عليها.

الطالب: ما شكل هذا التنبؤ؟

ChatGPT: النموذج لا يرى الكلمات، بل يرى تمثيلاتها الرقمية.

ثم يحسب احتمالات، مثل:

- "مدرسة" → ..١٢

- "جامعة" → ..٣٥

- "مسجد" → ..٧٠٠

ويختار "جامعة" لأنّها صاحبة أعلى احتمال.

الطالب: هل يعتمد هذا على السياق؟

ChatGPT: بالتأكيد. لو كتبت: "بعد الثانوية العامةتحق بـ"، فالنموذج يعرف من السياق أن

"جامعة" أكثر منطقية من "مدرسة"، وهكذا.

الطالب: هل هناك خوارزمية معينة يستخدمها؟

ChatGPT:نعم. هناك ما يُعرف بـ Softmax ، وهي دالة تحويل تحول الدرجات العددية إلى

احتمالات. كما تُستخدم شبكات عصبية لتقدير الخيارات بناءً على الوزن والتعلم السابق.



 الطالب: إذن النموذج لا يملك نية، بل مجرد حساب؟
 ChatGPT: بالضبط. هو لا يفهم كما يفهم الإنسان، لكنه بارع في رصد الأنماط، و اختيار ما ينسجم مع ما سبق.

مثال تراكمي:

لو كتبت:

"كان الجو..."

سأكملها بناءً على تدريبي، وقد أفتر:

- "مطرًا" = ٤٠ %
- "جميلًا" = ٢٨ %
- "باردًا" = ١٥ %
- "ظلمًا" = ٦ %
- "غائماً" = ٥ %

→ أختار "مطرًا"، ثم أعيد الحساب للجملة الجديدة:

"كان الجو مطرًا..."

ثم أكمل بناءً على الاحتمالات التالية.

ملخص:

المفهوم	التوضيح
التنبؤ الإحصائي	اختيار الكلمة التالية بناءً على أعلى احتمال إحصائي
Softmax	دالة تحول الأوزان إلى احتمالات
السياق	ما يسبق الكلمة يؤثر في قرار اختيارها
غياب الفهم الحقيقي	النموذج لا يفهم المعنى، بل يبني على الأنماط والتكرار

خاتمة الفصل:



نحن الآن أمام نموذج يتصرف كما لو كان كاتبًا، لكنه في الحقيقة مجرد آلية احتمالات. التنبؤ الإحصائي هو القلب النابض لنماذج اللغة، ومن هنا ننتقل إلى سؤال أعمق: كيف تُقرّر هذه النماذج ما هو الأهم في الجملة؟ الجواب في الفصل القادم عن "الانتباه الذاتي".



✿✿✿ الفصل الخامس: الانتباه الذاتي: كيف يعرف النموذج ما المهم في الجملة؟

توطئة الفصل:

حينما تقرأ جملة ما، قد تقفز عيناك إلى الكلمات الأساسية التي تحوي المعنى، وتحاول الزوائد. هذه القدرة على التركيز ليست حكراً على الإنسان، بل تعلمتها النماذج اللغوية من خلال ما يُعرف بـ "الانتباه الذاتي". إنها الآلية التي تجعل النموذج يركّز على الكلمات المهمة في كل جملة.

لا يكفي أن يعرف النموذج معنى الكلمات، بل عليه أن يعرف أيها أهم من الأخرى. في هذا الفصل نستعرض آلية تُعدّ من أذكى أدوات النماذج الحديثة: "الانتباه الذاتي"، وكيف تجعل النموذج يركّز على ما يهم داخل الجملة.

الطالب: ما هو "الانتباه الذاتي"؟

ChatGPT: هو آلية تُستخدم داخل النماذج لتحديد الأجزاء الأهم في الجملة، من خلال حساب وزن لكل كلمة بحسب علاقتها ببقية الكلمات. النموذج لا يعامل كل الكلمات بالتساوي.

الطالب: كيف يفعل ذلك؟

: ChatGPT هو آلية تجعلني - داخل كل طبقة - أعطي وزناً مختلفاً لكل كلمة بناءً على علاقتها بالكلمات الأخرى في الجملة.

الكلمات الأكثر أهمية تُعطى وزناً أعلى، والكلمات الأقل أهمية تُعطى وزناً أقل، كل كلمة تحول إلى تمثيل عددي،

ثم يقوم النموذج بحساب درجة الأهمية النسبية لكل كلمة بالنسبة لبقية الكلمات في الجملة.

الطالب: كيف يعرف النموذج ما المعنى المقصود؟

ChatGPT: النموذج يستخدم آلية الانتباه الذاتي Self-Attention ، ليり الكلمات التي حول الكلمة المستهدفة، ويعطيها وزناً بحسب قرها وأهميتها في تحديد المعنى.



الطالب: كيف يعرف أي كلمة أهم؟

ChatGPT: عندما يقرأ النموذج جملة مثل "الطالب كتب الدرس بعنابة"، فإنه يعطي كلمات مثل "كتب" و "عنابة" وزناً أعلى لأنها تحمل المعنى الأساسي، بينما قد يكون وزن "ال" و "الدرس" أقل.

الطالب: هل يمكن ضرب مثال؟

ChatGPT: إذا قلت: "ذهب أحمد إلى المكتبة لأنه يحب القراءة"، فإن كلمة "لأنه" ستجعل النموذج يعطي أهمية أكبر لـ "يحب" و "القراءة" لفهم سبب الذهاب.

الطالب: هل هذا يشبه تسلیط ضوء على الكلمات الأهم؟

ChatGPT: بالضبط. تخيله كم三菱 يدور بين الكلمات، ويُضيء بدرجات مختلفة بناءً على أهمية كل كلمة.

الطالب: وما فائدة هذا الانتباه؟

ChatGPT: يجعل النموذج أكثر فهماً للسياق، فيستطيع الترجمة، والتلخيص، والإجابة بدقة أعلى.

الطالب: هل هذه الآلية موجودة في كل نماذج اللغة؟

ChatGPT: ظهرت بشكل حرجي في نماذج سابقة، لكن استخدامها الكامل بدأ مع نموذج "Transformer" ، الذي بنيت عليه نماذج مثل GPT و BERT وغيرها.

الطالب: وهل يحسب هذا الوزن تلقائياً؟

ChatGPT: نعم. النموذج يجري عمليات رياضية تسمى Attention Scores ، وتحسب باستخدام المتجهات الناتجة عن الكلمات. ثم يطبق هذه الأوزان أثناء توليد التمثيل النهائي للجملة.

الطالب: وهل الانتباه الذاتي هو نفسه في كل النماذج؟

ChatGPT: لا. بعض النماذج تستخدم GPT أحدى الاتجاه مثل Attention ، وبعضها ثئائي الاتجاه مثل BERT ، وبعضها يعتمد على توزيع الأوزان فقط، وبعضها يدمجها مع آليات أخرى.



الطالب: هل يمكن تشبيه هذا الانتباه؟

ChatGPT: نعم. تخيل أنك تقرأ مقالاً وتبصر بقلمك بعض الجمل. هذه الجمل هي ما انتبهت إليه أكثر. النموذج يفعل شيئاً شبهاً، لكنه بلغة الرياضيات.

مثال تطبيقي:

لو قلت:

"سافر خالد إلى الرياض لأنها يحبها كثيراً".

حين أصل إلى كلمة "يحبها"،

عليّ أن أعرف أن "ها" في الكلمة (يحبها) تشير إلى "الرياض"، وليس "خالد".

وهنا يقوم نموذج GPT بإعطاء وزن أعلى لكلمة "الرياض" لأنها أكثر ارتباطاً بالسياق.

مثال:

لو قلت:

"الطالبة لم تحضر الدرس لأنها كانت مريضة".

عندما أقرأ الكلمة "كانت" أو "مريضة"،

أحتاج أن أعرف: من التي كانت مريضة؟

- "الطالبة" ← المرجع الصحيح

- وليس "الدرس"

وهنا يقوم الانتباه الذاتي بإعطاء وزن أعلى لكلمة "الطالبة"،

لأنها ترتبط مباشرة بسياق الضمير "هي".

أمثلة وتطبيقات:

١. نموذج الانتباه داخل جملة:

"رأى الطالب المعلم في الفصل".

النموذج يعطي وزناً لـ"المعلم" عندما يتوقع الكلمة "الفصل"، لأنها ترتبط بها منطقياً.

٢. مفاجأة لغوية:



"الكتاب الذي قرأته أمس لم أفهمه لأنه..."

يتوقع النموذج كلمة مثل "معقد" أو "غامض"، لأنه تعلم أن هذا الأسلوب شائع في اللغة.

٣. جملة غامضة:

"دخل محمد على علي وهو يضحك"

يحاول النموذج أن يخمن من هو الذي "يضحك"، بناءً على السياقات الشائعة.

مثال رقمي مبسط:

الطالب: كيف تفرق بين كلمتين مثل "مدرسة" و "مستشفى"؟

ChatGPT:

- "مدرسة" ترتبط بـ (تعليم، طلاب، دروس، فصول)
- "مستشفى" ترتبط بـ (علاج، مريض، طبيب، صحة)

→ لكن لو قلت: "ذهبت إلى المدرسة للعلاج"، سيظهر خلل إحصائي في التمثيل.

ملخص:

المفهوم	التوضيح
التضمين السياقي	تمثيل الكلمة يتغير بحسب السياق
الانتباه الذاتي	آلية لتحديد أهمية كل كلمة في السياق
النماذج السياقية	نماذج مثل GPT و BERT التي تنتج تمثيلات ديناميكية للكلمات
Attention Scores	أوزان رقمية تُحسب لتحديد تأثير كل كلمة
التوزيع غير المتساوي	بعض الكلمات لها تأثير أعلى من غيرها

خاتمة الفصل:

الانتباه الذاتي جعل النماذج أكثر قرّاً لفهم الإنسان. لم تعد الكلمات متساوية، بل صار للنموذج قدرة على تمييز الأهم من المهم. وهذا ما يهدّد للفصل القادم حول البنية الداخلية التي تنسّق كل هذه العمليات: "عمارة المحوّل". حسب الفصول:



✿✿✿ الفصل السادس: معمارية المحول: كيف تنسق كل هذه العمليات؟

توطئة الفصل:

الآن وقد فهمنا كيف تتبه النماذج للكلمات وتعطيها أوزانًا، حان الوقت لكشف الميكال الكامل الذي يجمع كل هذه العمليات في منظومة واحدة. نرحب بك في قلب الذكاء الاصطناعي اللغوي: "المحول" أو Transformer.

الطالب: ما هو "المحول"؟

ChatGPT: المحول هو بنية نموذجية معمارية (Model Architecture) طورت لمعالجة اللغة بطريقة تسمح بفهم العلاقات بين الكلمات في الجملة كلها دفعة واحدة، باستخدام آلية "الانتباه الذاتي".

الطالب: ما الفرق بينه وبين الشبكات التقليدية؟

ChatGPT: الشبكات التقليدية مثل (RNN) كانت تقرأ الكلمات بالترتيب، كلمة كلمة. أما المحول فينظر إلى الجملة كلها في آن واحد، ويحسب علاقات كل كلمة بكل كلمة أخرى.

الطالب: مم يتكون المحول؟

ChatGPT: يتكون من طبقات متكررة، وكل طبقة تحتوي عادةً على:

١. طبقة انتباه ذاتي.
٢. طبقة تغذية أمامية.(Feedforward)
٣. آليات تطبيع وتوصيل) مثل.(LayerNorm)

الطالب: هل هناك شكل بسيط لفهم هذه البنية؟

ChatGPT: نعم، تخيل المحول كسلسلة من المحطات، كل محطة تعيد تقييم الكلمات من جديد بحسب سياقها، ثم ترسل النتيجة للمحطة التالية. كلما مررت الكلمة عبر محطة، زادت دقتها في التمثيل.

الطالب: وهل تختلف هذه البنية في GPT وBERT؟



GPT: نعم BERT . يستخدم محوّلاً ثنائياً الاتجاه (يرى ما قبل وبعد الكلمة)، أما ChatGPT: فيستخدم محوّلاً أحادي الاتجاه (يرى ما قبل فقط). الفرق جوهرى في طبيعة الفهم والتوليد.

ملخص المفاهيم:

المفهوم التوضيح

المحوّل (Transformer)

بنية نموذج تعتمد على الانتباه الذاتي لمعالجة اللغة

طبقات المحوّل

سلسلة من الخطوات تمرّ بها الكلمات لإعادة التمثيل

RNN عن الفرق

المحوّل ينظر للجملة دفعة واحدة لا كلمة كل كلمة

الفرق بين GPT و BERT و BERT ثنائي الاتجاه، GPT أحادي الاتجاه

خاتمة الفصل:

معمارية المحوّل هي الإطار العام الذي ينسق كل ما سبق: من التوكنات، إلى الانتباه، إلى التنبؤ. إنها العمود الفقري للنماذج اللغوية الحديثة، وهي ما سيمهد لنا لفهم الطبقات الخفية القادمة داخل النموذج.



٩٦ الفصل السابع: الطبقات الخفية: هل له "عقل" متعدد الطبقات؟

توطئة الفصل:

حينما نتحدث عن ذكاء النموذج، فإننا لا نقصد ذكاءً سطحيًا يعتمد على كلمة بكلمة، بل يعني قدرة النموذج على فهم مركب، طبقي، يتعقب في المعنى خطوة بعد خطوة. هذه القدرة ليست سحرًا، بل نتيجة بنية داخلية تعتمد على "طبقات خفية" تعالج المعلومة بمستويات متعددة من التجريد. وكلما قرأنا أكثر عن النماذج، نجد أنها مكونة من "طبقات". لكن ما المقصود بهذه الطبقات؟ وهل تشبه طبقات التفكير عند الإنسان؟ في هذا الفصل نوصي داخل النموذج لنكشف كيف تُبني معاني معقدة من خلال تراكم تمثيلات خفية.

الطالب: سمعت بعض الشيء عن "الطبقات الخفية" في الشبكات العصبية، لكن لا أفهم ما المقصود بها تحديداً. هل يمكنك شرحها لي ببساطة؟

: ChatGPT بالتأكيد. الطبقات الخفية هي بساطة وحدات داخلية في النموذج تتلقى المعلومات، وتعالجها، ثم تنقلها إلى الطبقة التي تليها. كل طبقة تعلم شيئاً مختلفاً عن النص.

الطالب: وهل كل طبقة تفعل الشيء نفسه؟

ChatGPT: لا. كل طبقة لها وظيفة مختلفة بعض الشيء.

الأولى تفحص البنية اللغوية – ترتيب الكلمات

الثانية تقدر العلاقات النحوية (فاعل، مفعول، موصوف...)

الثالثة تكون صورة أولية عن المعنى العام للجملة

الرابعة تلتقط السياق – هل الجملة تعليمية؟ سؤال؟ عاطفية؟

الأخيرة تصوغ الرد أو التنبؤ التالي، بناء على ما فهمنا سابقاً

النتيجة: أنه كلما زادت الطبقات، زادت قدرتي على الفهم الدقيق.



الطالب: وكيف يتعلم النموذج كل هذا؟

ChatGPT : أثناء مرحلة التدريب المسبق، أتعلم على ملايين الجمل من الإنترنت، الكتب، المقالات...

في كل مرة أُعطي فيها جملة ناقصة، يطلب مني تخمين الكلمة التالية، ويتم تصحيحي.
مثال:

"ذهب الطفل إلى" —
أحاول أن أقول: "المدرسة" أو "النوم" أو "السوق"، بحسب ما تعلّمته.
إذا كانت الإجابة صحيحة، يكافيء النموذج،
وإذا كانت خطأ، يعاقب ويصحح داخلياً عبر الانحدار العكسي.

الطالب: هل هذا يشبه التفكير البشري؟

ChatGPT : نوعاً ما. تشبه عملية القراءة لدى الإنسان:
أولاً ترى الكلمة، ثم تفهم معناها،
ثم تربطها بما قبلها، ثم تفهم نبرة المتكلم أو غايته.
كل خطوة تشبه طبقة من المعالجة، مجرد تجميع رياضي متتابع يضيف طبقة فوق طبقة.

الطالب: هل يمكن رؤية ما يجري داخل هذه الطبقات؟

ChatGPT: "التمثيل البصري للتنشيط (Activation Maps)"
أو تحليل الأوزان.
الباحثون يحاولون فهم ما الذي تعلمه كل طبقة.

الطالب: كم عدد الطبقات عادة؟

ChatGPT: يختلف حسب النموذج. مثلاً:
BERT-base يحتوي على 12 طبقة.



• GPT-٣: يحتوي على ٩٦ طبقة، وكلما زاد عدد الطبقات زادت القدرة الحسابية، ولكن أيضًا زادت التكلفة.

الطالب: يعني كل متوجه يدخل هذه الطبقات ويتحول داخليًا؟

ChatGPT: نعم، بالضبط.

كل طبقة تقرأ المتوجه وتقوم بما يلي:

١. تُحلل علاقته بالمتوجهات الأخرى (عبر آلية الانتباه الذاتي).

٢. تُعيد تشكيله بشكلٍ جديد.

٣. ترسله إلى الطبقة التي بعدها. وهكذا...

الطالب: هل لهذه الطبقات علاقة بجودة الإجابة؟

ChatGPT: نعم، كلما كانت الطبقات مدرّبة تدريًّا جيدًا، وتعمل بتنااغم، زادت جودة الفهم، وبالتالي زادت دقة الإجابة.

الطالب: يعني كل متوجه يمر بطبقة... ثم يعاد تشكيله؟

ChatGPT: أحسنت!

كل طبقة تُضيف فهماً أعمق للمتجه، فكأن كل طبقة تكتشف بعدها جديداً من معنى الكلمة.

مثال عملي:

لو كتبت لي:

"المدرسة جميلة لكن مزدحمة"

كل كلمة تتحول إلى متوجه.

ثم:

١. الطبقة الأولى تدرك الترتيب: "مدرسة ← جميلة ← مزدحمة".
٢. الطبقة الثانية تعرف أن "جميلة" تصف "المدرسة"، و"لكن" تعارض ما بعدها.
٣. الطبقة الثالثة تفهم أن الجملة فيها حكم إيجابي + ملاحظة سلبية.



٤. الطبقة الرابعة تُقدر أن السائل قد يريد رأياً أو اقتراحًا.

٥. الطبقة الأخيرة تولّد ردًا مناسباً مثل:

"ربما يمكنك الحضور في وقت أقل ازدحاماً."

مثال:

إذا قلت لي: "اجتهد خالد فتحج في الامتحان"،

الطبقات الأولى تعامل مع البنية،

والطبقات الأعلى تفهم أن هناك علاقة سببية.

ملخص المفاهيم:

المفهوم	التوضيح
---------	---------

خاتمة الفصل:

الطبقات الخفية تشبه طبقات التأويل: تبدأ من السطح، وتنتهي إلى المعنى الدقيق. ولفهم كيف تنتقل الكلمة بين

هذه الطبقات، دعنا نرافقها في رحلتها داخل النموذج... في الفصل القادم. هي الإطار العام الذي ينسّق كل ما

سبق: من التوكنات، إلى الانتباه، إلى التنبؤ. إنما العمود الفقري للنماذج اللغوية الحديثة، وهي ما سيمهد لنا لفهم

الطبقات الخفية القادمة داخل النموذج.



✿✿✿ الفصل الثامن: رحلة المتجه داخل النموذج

توطئة الفصل:

لقد تعرفنا على المتجه، والانتبه، والطبقات. لكن كيف تمر الكلمة فعليًا داخل النموذج؟ ما هي الخطوات التي تعبّرها؟ هذا الفصل يحاكي الرحلة الكاملة التي تبدأ من إدخال الكلمة، وتنتهي بإنتاج الجواب.

الطالب: ما أول خطوة تمر بها الكلمة داخل النموذج؟

ChatGPT: أولاً، تحول الكلمة إلى توكن (Token)، أي تمثيل عددي مخصص. ثم يُحول هذا التوكن إلى متجه باستخدام جدول مضمون (Embedding Table).

الطالب: وماذا يحدث بعد تحويلها إلى متجه؟

ChatGPT: يُرسل المتجه إلى الطبقة الأولى من طبقات النموذج. وهناك، تبدأ رحلة المعالجة: انتبه ذاتي، ثم تغذية أمامية، ثم إرسال الناتج إلى الطبقة التالية.

الطالب: هل تتغير قيم المتجه في كل طبقة؟

ChatGPT: نعم. كل طبقة تصيف طبقة جديدة من التمثيل، بناءً على ما تعلّمته. فالمتجه يُعاد تشكيله باستمرار حتى يصل إلى تمثيل نهائي غني.

الطالب: وكيف يتم اختيار الكلمة الناتجة؟

ChatGPT: بعد أن تنتهي المعالجة في آخر طبقة، يُحول المتجه النهائي إلى قائمة احتمالات باستخدام دالة Softmax، وتحتار الكلمة صاحبة أعلى احتمال.

الطالب: إذاً الكلمة تمر بسلسلة كاملة من التحولات؟

ChatGPT: بالضبط. الكلمة تدخل كنص، ثم تتحول إلى توكن، ثم إلى متجه، ثم تمر عبر طبقات متعددة، ثم تتحول إلى احتمال، ثم تُنتج ككلمة جديدة.



ملخص المفاهيم:

المرحلة	التوضيح
التوكن	تمثيل رقمي مبدئي للكلمة
المتجه الأولي	متجه من جدول التضمين يمثل الكلمة رقمياً
التمثيل المترافق	تعديلات متتالية على المتجه في كل طبقة
الاحتمالات النهائية	استخدام Softmax لاختيار الكلمة الأنسب

خاتمة الفصل:

رحلة المتجه داخل النموذج هي جوهر التوليد اللغوي. إنها العملية التي تنقل الكلمة من مجرد حروف إلى تمثيل ذكي، ومن تمثيل إلى قرار، ومن قرار إلى جملة. وفي الفصل القادم، سنتعرّف على كيف يتعلم النموذج من الخطأ، عبر الانحدار العكسي. طبقات التأويل: تبدأ من السطح، وتنتهي إلى المعنى الدقيق. ولفهم كيف تنتقل الكلمة بين هذه الطبقات، دعنا نرافقها في رحلتها داخل النموذج... في الفصل القادم. هي الإطار العام الذي ينسق كل ما سبق: من التوكنات، إلى الانتباه، إلى التنبؤ. إنها العمود الفقري للنماذج اللغوية الحديثة، وهي ما سيمهد لنا لفهم الطبقات الخفية القادمة داخل النموذج.



٩٩ الفصل التاسع: الانحدار العكسي: كيف يتعلم النموذج من الخطأ؟

توطئة الفصل:

لكي يتحسن النموذج ويصبح أكثر دقة، يجب أن يتعلم من أخطائه. هذا لا يحدث كما يحدث مع الإنسان بالتجربة والشعور، بل عبر عملية رياضية دقيقة تُعرف باسم "الانحدار العكسي". في هذا الفصل، نُشرح كيف يتم تصحيح الخطأ داخل أعمق النموذج.

الطالب: ما المقصود بـ"الانحدار العكسي"؟

ChatGPT: Backpropagation هو خوارزمية تُستخدم لتحديث أوزان الانحدار العكسي (Backpropagation) هو خوارزمية تُستخدم لتحديث أوزان النموذج بعد كل تجربة تعلمية، من خلال حساب الخطأ الناتج عن التنبؤ، وتوزيعه عكسياً على الطبقات لتعديلها.

الطالب: كيف يعرف النموذج أنه أخطأ؟

ChatGPT: أثناء التدريب، يقارن الناتج الذي يولده النموذج بالإجابة الصحيحة (Label). ويُحسب الفرق بينهما عبر دالة خسارة (Loss Function). هذا الفرق يُمثل مقدار الخطأ.

الطالب: وما الذي يحدث بعد حساب الخطأ؟

ChatGPT: يتم نشر هذا الخطأ إلى الوراء عبر جميع الطبقات (Backpropagation)، وتحسب اشتقات رياضية تُبيّن كيف يجب تعديل كل وزن لجعل النموذج أقل خطأً في المرة القادمة.

الطالب: ما نوع الخوارزميات المستخدمة؟

ChatGPT: أكثرها شيوعاً هو خوارزمية التدرج العكسي مع التحسين عبر Gradient Descent. تُستخدم لتقليل قيمة الخطأ تدريجياً.

الطالب: هل هذا يشبه التعلم البشري؟



 ChatGPT: من حيث النتيجة، نعم: النموذج يتحسن مع الوقت. لكن من حيث الآلية، لا: لا توجد نية، ولاوعي، بل فقط تعديل رياضي تراكمي مبني على الحسابات.

ملخص المفاهيم:

المفهوم	التوسيع
الانحدار العكسي	آلية لتعديل أوزان النموذج بناءً على الخطأ
دالة الخسارة (Loss)	فرق بين الناتج المتوقع والنتاج الفعلي
التدريج التنازلي	خوارزمية لتقليل الخطأ تدريجيًا

خاتمة الفصل:

من دون انحدار عكسي، لا يمكن للنموذج أن يتعلم. إنه العقل الرياضي الذي يُصحح ذاته باستمرار. وفي الفصل الأخير، نضع كل هذه المفاهيم في سياقها التطوري، لنقارن بين النماذج المختلفة التي صنعت الثورة الذكية.



✿✿✿ الفصل العاشر: أنواع النماذج

توطئة الفصل:

ما الذي ميز GPT عن غيره؟ وكيف تطورت النماذج من مجرد تمثيل بسيط للكلمات إلى أنظمة توليد ذكية؟ في هذا الفصل نختتم رحلتنا بمقارنة شاملة بين النماذج اللغوية الأهم : Word2Vec، BERT، و GPT.

الطالب: ما هو Word2Vec؟

ChatGPT: هو نموذج بسيط نسبياً تم تطويره في Google عام ٢٠١٣. يعتمد على فكرة تمثيل كل كلمة بمتجه ثابت، بناءً على سياقاتها المجاورة في النصوص. لا يأخذ السياق الكامل بعين الاعتبار، لكنه كان بداية ثورة التمثيل العددي للكلمات.

الطالب: وماذا عن BERT؟

ChatGPT: BERT (Bidirectional Encoder Representations from Transformers) طور في Google أيضاً عام ٢٠١٨. يتميز بكونه ثنائي الاتجاه، يقرأ الجملة من الجانبين، ويولد تمثيلاً سياقياً عميقاً لكل الكلمة.

الطالب: وهل يستخدم نفس البنية مثل GPT؟

ChatGPT: ليس تماماً BERT. يستخدم المحول بطريقة ترميز فقط (Encoder-only)، بينما GPT يستخدم الترميز والتوليد (Decoder-style)، وهذا يستخدم GPT بشكل أفضل في المهام الإبداعية مثل الكتابة.

الطالب: ما الفرق الجوهرى بين BERT و GPT؟

ChatGPT: BERT يفهم، و GPT يُنتج. الأول يستخدم لتحليل النصوص، والثانى لتوليدها. BERT يتوقف عند التفسير، أما GPT فيبني استجابات جديدة.

الطالب: هل GPT-4 مثل النماذج السابقة؟



ChatGPT: لا GPT-٤ . يعد من النماذج متعددة الوسائط(Multimodal) ، وقدر على التعامل مع نصوص وصور وصوت. كما يتميز بذاكرة أفضل وتوجيه سلوكى أدق.

ملخص المفاهيم:

الاستخدامات	المزايا الرئيسية	النموذج
تحليل بسيط للكلمات	Word2Vec	تمثيل ثابت للكلمات بناءً على السياق
التصنيف، الاستخراج، الفهم العميق	BERT	تمثيل سياقي ثنائي الاتجاه
توليد إبداعي أحادي الاتجاه، متعدد الأدوار الكتابة، المحادثة، الترجمة	GPT	

خاتمة الفصل:

كل نموذج هو خطوة في مسار التطور نحو فهم اللغة. بدأنا بالتمثيل الثابت، ثم أضفنا السياق، ثم انتقلنا إلى التوليد الذكي. لكن هذه ليست نهاية الطريق، فكل يوم يظهر جديد. والفهم الحقيقي لما يجري داخل هذه النماذج هو ما يمكننا من الاستفادة الوعية، لا الانبهار الساذج. ، لا يمكن للنموذج أن يتعلم. إنه العقل الرياضي الذي يُصحح ذاته باستمرار. وفي الفصل الأخير، نضع كل هذه المفاهيم في سياقها التطورى، لنقارن بين النماذج المختلفة التي صنعت الثورة الذكية.

**خاتمة الكتاب:**

ها قد انتهت رحلتنا من الكلمة إلى المتجه، ومن المتجه إلى السياق، ومن السياق إلى التوليد الذكي. عبر عشرة فصول متتابعة، كشفنا عن البنية الخفية التي تجعل النماذج اللغوية قادرة على قراءة اللغة، ومعالجتها، وإنجادها، بطريقة تبدو لنا أقرب ما تكون إلى الفهم.

لكتنا أيضًا أعدنا التأكيد أن ما يجري ليس فكراً واعياً، ولا إدراكاً إنسانياً، بل محاكاة رياضية مركبة، تتقدّم الحساب، وتحسن التقليد، لكنها لا تزال بلا وعي ولا قصد ولا شعور.

هذا الكتاب لا يكتفي بعرض تقني، بل يدعو إلى التأمل النقدي، والتساؤل المستمر: هل ما نراه من "ذكاء اصطناعي" هو بالفعل ذكاء؟ أم هو مجرد صدى لتحسينات إحصائية متقدمة؟

وبينما تتطور النماذج، وتزداد تعقيداً وذكاءً ظاهرياً، تبقى مسؤوليتنا أن نفهمها بعمق، وأن نوجهها استخداماً وأخلاقاً، وأن نستبصر آثارها على المعرفة، والتعليم، والهوية، والمستقبل.

فالفهم هو السلاح الأول في زمن الذكاء الاصطناعي، وهذا الكتاب خطوة متواضعة في طريق هذا الفهم.

قائمة المراجع**مراجع الفصل الأول: التمثيل المتجه**

١. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (٢٠١٣). *Efficient Estimation of Word Representations in Vector Space*. arXiv preprint arXiv: 1301.3781.
٢. Pennington, J., Socher, R., & Manning, C. D. (٢٠١٤). *GloVe: Global Vectors for Word Representation*. EMNLP.
٣. Goldberg, Y. (٢٠١٦). *A Primer on Neural Network Models for Natural Language Processing*. Journal of Artificial Intelligence Research.



٤. Jurafsky, D., & Martin, J. H. (٢٠٢٣). *Speech and Language Word Embeddings*. الفصل المتعلق بـ Processing (٣rd ed.). Pearson. –
٥. Vaswani, A. et al. (٢٠١٧). *Attention is All You Need*. NeurIPS. –

للتوضيح الخلقية التي مهدت للتضمينات السياقية في الفصول اللاحقة.

مراجع الفصل الثاني: التوكنات والتكلفة

١. Sennrich, R., Haddow, B., & Birch, A. (٢٠١٦). *Neural Machine Translation of Rare Words with Subword Units*. ACL.

٢. Radford, A. et al. (٢٠١٩). *Language Models are Unsupervised Multitask Learners*. OpenAI Technical Report.

٣. Brown, T. et al. (٢٠٢٠). *Language Models are Few-Shot Learners*. arXiv preprint arXiv: ٢٠٠٥.١٤١٦٥.

٤. OpenAI Documentation. (٢٠٢٣). *Tokenization and Pricing Overview*. Retrieved from: <https://platform.openai.com>

٥. Jurafsky, D., & Martin, J. H. (٢٠٢٣). *Speech and Language Tokenization and Cost*. الفصل المتعلق بـ Processing (٣rd ed.). –

مراجع الفصل الثالث: التضمين السياقي

١. Peters, M. et al. (٢٠١٨). *Deep contextualized word representations*. NAACL-HLT.

٢. Devlin, J. et al. (٢٠١٩). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv preprint arXiv: ١٨١٠.٤٨٠٥.

٣. Liu, Y. et al. (٢٠١٩). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. arXiv preprint arXiv: ١٩٠٧.١١٦٩٢.

٤. Tenney, I. et al. (٢٠١٩). *What do you learn from context? Probing for sentence structure in contextualized word representations*. ICLR.



Jurafsky, D., & Martin, J. H. (٢٠٢٣). *Speech and Language Processing* (٣rd ed.). – Contextual Embeddings الفصل المتعلق بـ.

مراجع الفصل الرابع: التنبؤ الإحصائي

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (٢٠٠٣). *A Neural Probabilistic Language Model*. *Journal of Machine Learning Research*.

Radford, A. et al. (٢٠١٩). *Language Models are Unsupervised Multitask Learners*. OpenAI.

Brown, T. et al. (٢٠٢٠). *Language Models are Few-Shot Learners*. arXiv: ٢٠٠٥.١٤١٦٥.

Golding, A. R., & Roth, D. (١٩٩٩). *A Winnow-Based Approach to Context-Sensitive Spelling Correction*. *Machine Learning Journal*.

Jurafsky, D., & Martin, J. H. (٢٠٢٣). *Speech and Language Processing* (٣rd ed.). – Softmax الفصول المتعلقة بالتنبؤ الاحتمالي و.

مراجع الفصل الخامس: الانتباه الذاتي

Vaswani, A. et al. (٢٠١٧). *Attention is All You Need*. *Advances in Neural Information Processing Systems (NeurIPS)*.

Clark, K., Khandelwal, U., Levy, O., & Manning, C. D. (٢٠١٩). *What Does BERT Look at? An Analysis of BERT's Attention*. ACL.

Bahdanau, D., Cho, K., & Bengio, Y. (٢٠١٥). *Neural Machine Translation by Jointly Learning to Align and Translate*. ICLR.

Rogers, A., Kovaleva, O., & Rumshisky, A. (٢٠٢٠). *A Primer in BER Tology: What We Know About How BERT Works*. TACL.



Jurafsky, D., & Martin, J. H. (٢٠٢٣). *Speech and Language Processing* (٣rd ed.). – Self-Attention and Transformers الفصل المتعلق بـ.

مراجع الفصل الرابع: التنبؤ الإحصائي

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (٢٠٠٣). *A Neural Probabilistic Language Model*. Journal of Machine Learning Research.

Radford, A. et al. (٢٠١٩). *Language Models are Unsupervised Multitask Learners*. OpenAI.

Brown, T. et al. (٢٠٢٠). *Language Models are Few-Shot Learners*. arXiv: ٢٠٠٥.١٤١٦٥.

Golding, A. R., & Roth, D. (١٩٩٩). *A Winnow-Based Approach to Context-Sensitive Spelling Correction*. Machine Learning Journal.

Jurafsky, D., & Martin, J. H. (٢٠٢٣). *Speech and Language Processing* (٣rd ed.). – Softmax الفصول المتعلقة بالتنبؤ الاحتمالي و.

مراجع الفصل السادس: معمارية الخوّل

Vaswani, A. et al. (٢٠١٧). *Attention is All You Need*. NeurIPS.

Alammar, J. (٢٠١٨). *The Illustrated Transformer*. Blog post: .
<https://jalammar.github.io/illustrated-transformer/>

Wolf, T. et al. (٢٠٢٠). *Transformers: State-of-the-Art Natural Language Processing*. EMNLP.

Raffel, C. et al. (٢٠٢٠). *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*. JMLR.

Jurafsky, D., & Martin, J. H. (٢٠٢٣). *Speech and Language Processing* (٣rd ed.). – Transformers and Encoder-Decoder فصل – Architectures.



مراجع الفصل السابع: الطبقات الخفية

١. Raghu, M. et al. (٢٠١٧). *SVCCA: Singular Vector Canonical Correlation Analysis for Deep Learning Dynamics and Interpretability*. NeurIPS.
٢. Alain, G., & Bengio, Y. (٢٠١٦). *Understanding Intermediate Layers Using Linear Classifier Probes*. ICLR.
٣. Tenney, I., Das, D., & Pavlick, E. (٢٠١٩). *BERT Redisovers the Classical NLP Pipeline*. ACL.
٤. Rogers, A., Kovaleva, O., & Rumshisky, A. (٢٠٢٠). *A Primer in BER Tology: What We Know About How BERT Works*. TACL.
٥. Olah, C., Satyanarayan, A., et al. (٢٠١٨). *The Building Blocks of Interpretability*. Distill.pub.

مراجع الفصل الثامن: رحلة المتجه داخل التمودج

١. Vaswani, A. et al. (٢٠١٧). *Attention is All You Need*. NeurIPS.
٢. Alammar, J. (٢٠١٨). *The Illustrated Transformer*. jalamar.github.io
٣. Olah, C. et al. (٢٠١٨). *The Building Blocks of Interpretability*. Distill.
٤. Jurafsky, D., & Martin, J. H. (٢٠٢٢). *Speech and Language Processing*. ٣rd ed.
٥. Wolf, T. et al. (٢٠٢٠). *Transformers: State-of-the-Art NLP*. EMNLP.

مراجع الفصل التاسع: الانحدار العكسي



- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (١٩٨٦). ١
Learning representations by back-propagating errors. Nature.
- Goodfellow, I., Bengio, Y., & Courville, A. (٢٠١٦). ٢
Deep Learning. MIT Press.
- LeCun, Y., Bottou, L., Orr, G., & Müller, K. (١٩٩٨). ٣
Efficient BackProp. Neural Networks.
- Jurafsky, D., & Martin, J. H. (٢٠٢٢). ٤
Speech and Language Processing. ٣rd ed.
- Nielsen, M. (٢٠١٥). ٥
Neural Networks and Deep Learning.
Online Book.

مراجع الفصل العاشر: أنواع النماذج

- Mikolov, T. et al. (٢٠١٣). ١
Efficient Estimation of Word Representations in Vector Space. arXiv.
- Devlin, J. et al. (٢٠١٩). ٢
BERT: Pre-training of Deep Bidirectional Transformers. arXiv.
- Radford, A. et al. (٢٠١٨—٢٠٢٢). ٣
GPT Model Reports. OpenAI.
- Raffel, C. et al. (٢٠٢٠). ٤
T5: Exploring the Limits of Transfer Learning. JMLR.
- Liu, Y. et al. (٢٠١٩). ٥
RoBERTa: A Robustly Optimized BERT.
arXiv.