

كيف يعمل النموذج اللغوي الكبير (LLM): طريقة مبسطة لشرحه

Assem Hijazi – Experience, Knowledge & Ideas

أهدف من خلال هذه المقالة السريعة إلى تبسيط فهم النموذج اللغوي الكبير (LLM)، لأنه لا يزال من الصعب شرحه. حاولت أن أبسطه قدر الإمكان وأن أقدم أمثلة لجعله أكثر وضوحًا.

تخيل أنك عندما تطرح سؤالاً على نموذج لغوي كبير – مثل "ما عدد سكان باريس وكيف يؤثر ذلك على الاقتصاد؟" – فإنك في الحقيقة تعطيه مشروحاً ليعمل عليه.

هذا المشروع لا يُعالج بواسطة عقل واحد ضخم، بل يتم التعامل معه من خلال مجموعة من الفرق التي تعمل معًا.

كل فريق من هذه الفرق يتكوّن من طبقات داخل النموذج اللغوي. هذه الطبقات تشبه مجموعات من العاملين، ولكل منها دور مختلف في تحليل سؤالك ومحاولة الوصول إلى أفضل إجابة. بعض الفرق تتولى المهام البسيطة، في حين أن فرقاً أخرى أكثر خبرة تتعامل مع المشاريع الأصعب.

أكيد حبيبي، إليك الترجمة الكاملة لهذا الجزء الجميل:

فكيف يعمل كل هذا؟

1. النموذج يستقبل سؤالك (المشروع)

في البداية، أنت تعطي النموذج مشروعاً بمجرد أن تطرح سؤالاً. قد يكون هذا السؤال بسيطاً مثل: "هل باريس هي عاصمة فرنسا؟"، أو قد يكون معقداً أكثر مثل المثال السابق حول عدد سكان باريس وتأثيرهم على الاقتصاد.

2. الفرق تبدأ التحليل (الطبقات تبدأ العمل)

بمجرد أن يدخل السؤال (المشروع)، تبدأ الفرق المختلفة (الطبقات) في العمل. إذا كان السؤال بسيطاً، قد تحتاج فقط بعض الفرق. أما إذا كان السؤال معقداً، تشارك فرق أكثر. كل طبقة (فريق) تنظر إلى جزء معين من السؤال، وتنقل نتائجها للطبقة التالية.

3. بناء العلاقات بين الكلمات (المعاملات)

أثناء عمل الفرق، تستخدم ما يُعرف بـ "المعاملات" – وهي ببساطة الروابط أو "العلاقات" بين الكلمات التي تعلمها النموذج من المشاريع السابقة (التدريب). فإذا كان النموذج قد رأى جملة "باريس هي عاصمة فرنسا" مئات المرات من قبل، فسيدرك أن هناك علاقة قوية بين هذه الكلمات. المعاملات هي ما يصنع هذه الروابط.

4. اكتشاف الأنماط (الكلمات المفتاحية)

أثناء عمل الفرق على المشروع، تبدأ بملاحظة أنماط في السؤال. مثلاً، إذا سألت عن عدد السكان والاقتصاد في باريس، سيتعرف النموذج على أن هذين المفهومين مرتبطان – لأنه قد رآهما مرتبطتين من قبل. فيتبع هذه الأنماط ليحدد أفضل طريقة للإجابة على سؤالك.

5. لا يوجد مدير كبير يتخذ القرارات

قد تعتقد أن هناك "عقل مركزي" داخل النموذج يتخذ القرارات، لكن هذا ليس صحيحاً. لا يوجد رئيس يوجّه الأمور. بدلاً من ذلك، تعمل الطبقات (الفرق) معاً، وكل واحدة منها تُعدل أداءها بناءً على العلاقات التي تعلمتها. يتعاونون ويتبادلون المعلومات، لكن لا يوجد صانع قرار واحد.

6. تعديلات ذهاب وعودة (التغذية الراجعة)

إذا بدأت الفرق في اتخاذ مسار خاطئ، فإنها تقوم بتعديل طريقته. هنا يأتي دور ما يسمى "الانتشار العكسي" (التغذية الراجعة بين

الفرق) خلال عملية التدريب. وهو ببساطة يعني أن النموذج يتعلم من أخطائه – فإذا لم يكن شيء صحيحًا، يقوم بتعديل طريقة تعامله في المرات القادمة.

7. الأسئلة المعقدة تحتاج إلى فرق أكثر (طبقات أكثر)

بالنسبة لسؤال مثل "هل باريس هي عاصمة فرنسا؟"، فهو مشروع بسيط، لذا تشارك فيه فرق قليلة لأن النموذج يعرف الإجابة جيدًا. لكن عندما تسأل سؤالًا أكثر تعقيدًا، مثل "كيف يؤثر عدد سكان باريس على اقتصادها؟"، يستدعي النموذج المزيد من الفرق (الطبقات) ويستخدم معاملات أكثر لربط جميع الأجزاء. كلما كان السؤال أصعب، تعاونت طبقات أكثر للوصول إلى الإجابة.

8. بيانات التدريب هي كتاب القواعد

كل هذا – طريقة عمل الطبقات معًا، وكيف يتم إعداد المعاملات – يأتي من بيانات التدريب التي تم تغذية النموذج بها. فكر في بيانات التدريب على أنها كتاب قواعد ضخم. كلما كان هذا الكتاب دقيقًا، كان فهم النموذج أفضل. إذا كان النموذج قد شاهد بيانات موثوقة بشكل كافٍ، فسيعرف أن "باريس" و"فرنسا" مرتبطتان، وسيفهم أمورًا أكثر تعقيدًا مثل تأثير عدد السكان على الاقتصاد.

9. إنتاج الإجابة النهائية

بعد أن تنهي الفرق عملها، يقوم النموذج بتركيب الإجابة. بالنسبة للسؤال البسيط، يتم ذلك بسرعة وبخطوات قليلة. أما في الأسئلة المعقدة، فتحتاج إلى خطوات أكثر وتعاون أكبر بين الطبقات للوصول إلى العلاقات الصحيحة وتقديم إجابة مدروسة.

بهذا الشكل من التفكير، يصبح من الأسهل فهم طريقة عمل النماذج اللغوية الكبيرة: كل سؤال هو بمثابة مشروع، وطبقات النموذج هي فرق تعمل معًا، وتُعدل وتتعلم في كل مرة، دون أن يكون هناك قائد واحد يتحكم بكل شيء.

دور بيانات التدريب: كتاب القواعد للنموذج اللغوي

الآن دعنا نتحدث بتفصيل أكثر عن بيانات التدريب، وهي الأساس لكل ما يعرفه النموذج اللغوي الكبير (LLM) فكر في بيانات التدريب على أنها "كتاب القواعد" الخاص بالنموذج – هو ما علّمه كيف ترتبط الكلمات ببعضها وكيف يتعامل مع المشاريع والأسئلة الجديدة (المحفزات). ببساطة، كانت بيانات التدريب بمثابة المعلم للنموذج. والآن بعد أن تعلّم كل شيء من هذا المعلم، أصبح بإمكانه التعامل مع أسئلة جديدة بناءً على المعرفة التي خزّنها أثناء التدريب.

التعلّم من كتاب القواعد

قبل أن يتم إطلاق النموذج، خضع لتدريب على كميات ضخمة من النصوص (كتب، مقالات، مواقع إلكترونية). وخلال هذا التدريب، قرأ النموذج البيانات وتعلّم الأنماط والعلاقات بين الكلمات. وهنا قام ببناء "كتاب القواعد" الخاص به. على سبيل المثال، تعلّم من البيانات أن "باريس" مرتبطة بـ"فرنسا" و"عاصمة". وكلما رأى هذه الكلمات معًا أكثر، فهم العلاقة بينها بشكل أقوى. هذه هي الطريقة التي ملأ بها النموذج ذاكرته بالمعلومات التي يمكنه استخدامها لاحقًا.

استخدام كتاب القواعد دون الرجوع إليه

وهنا نقطة مهمة: بمجرد أن يتم تدريب النموذج، لا يحتاج إلى الرجوع لبيانات التدريب في كل مرة يُطرح عليه سؤال. لقد تعلّم بالفعل كل ما يحتاجه من تلك البيانات – تمامًا كما لا يحتاج الإنسان لإعادة قراءة كتاب في كل مرة يُسأل عن معلومة فيه. بدلًا من ذلك، يعتمد النموذج على الأنماط والعلاقات التي خزّنها مسبقًا.

لذا، عندما تسأل: "ما عدد سكان باريس؟ وكيف يؤثر ذلك على الاقتصاد؟"، فإن النموذج لا يقوم بالبحث في المقالات أو الكتب، بل يستخدم المعرفة التي جمعها من بيانات التدريب – من "كتاب القواعد" الذي حفظه – ليقدم لك الإجابة.

أكيد حبيبي، إليك الترجمة لهذا الجزء:

لماذا لا يحتاج النموذج اللغوي للرجوع إلى بيانات التدريب في الوقت الحقيقي

النموذج اللغوي الكبير لا يحتاج إلى الرجوع إلى بيانات التدريب كل مرة لأنه:

• **لقد تعلّم المحتوى مسبقًا:** بيانات التدريب علّمت النموذج كيف تتناسب الكلمات والمفاهيم مع بعضها البعض. والآن، يستخدم ذاكرته لتطبيق هذه المعرفة.

• **يستخدم ما هو مخزن في ذاكرته:** تمامًا كما تتذكر المعلومات بعد قراءة كتاب، يحتفظ النموذج بالعلاقات والأنماط التي تعلّمها ويستخدمها مباشرة دون الحاجة للرجوع إلى البيانات الأصلية.

تلخيص كل ما سبق:

- **المحفز = مشروع**: عندما تطرح سؤالاً، فأنت تعطي النموذج مشروغاً للعمل عليه.
- **الفرق = الطبقات: الطبقات (الفرق) تعمل على أجزاء مختلفة من المشروع.**
- **المعاملات = العلاقات: المعاملات هي الروابط بين الكلمات، التي تعلّمها النموذج من التدريب السابق.**
- **الأنماط = الكلمات المفتاحية: الفرق تكتشف الأنماط بناءً على ما رآته من قبل.**
- **لا يوجد مدير مركزي**: لا يوجد فريق مسؤول وحده؛ بل الجميع يعملون معاً دون وجود قائد واحد.
- **بيانات التدريب = كتاب القواعد**: يتعلّم النموذج من بيانات التدريب (كتابه الإرشادي)، مما يساعده على تقديم إجابات دقيقة.
- **درجة التعقيد: المشاريع البسيطة تحتاج إلى فرق أقل، أما المشاريع المعقدة فتحتاج إلى طبقات ومعاملات أكثر.**
- **بيانات التدريب = كتاب القواعد**: يتعلّم النموذج من بيانات التدريب، تماماً كما نتعلم نحن من الكتب، ولا يحتاج للعودة إليها طالما أنه قد فهم المادة بالفعل.

فهم المعاملات والوحدات (Tokens) في النموذج اللغوي: شرح مبسّط

لقد أوضحنا سابقاً أن في النموذج اللغوي الكبير (مثل GPT) ، المعاملات هي العلاقات بين الكلمات، والوحدات أو ما يُعرف بـ "Tokens" هي الكلمات نفسها. الآن دعنا نغوص أعمق قليلاً لفهم كيف تعمل هذه الوحدات والمعاملات معاً لمساعدة النموذج في فهم الأسئلة (المحفزات) والرد عليها.

الوحدات (Tokens) والعلاقات

فكر في الوحدات على أنها كلمات فردية أو أجزاء من كلمات. كل كلمة لديها علاقات معينة مع كلمات أخرى، وهذه العلاقات تم بناؤها خلال تدريب النموذج. كلما كانت كمية بيانات التدريب أكبر، زادت قدرة النموذج على بناء علاقات (أو معاملات) بين الكلمات المختلفة.

عندما تعطي النموذج محفزاً (وسنعتبره "مشروعاً")، يقوم بتحليل العلاقات بين الكلمات في هذا المحفز. يقيّم مدى قوة الروابط بين الكلمات ويحدد أيها الأهم. هذه العلاقات يتم تمثيلها بواسطة "متجهات" (Vectors)، وهي عبارة عن قيم رقمية تعبّر عن مدى ترابط كلمتين. هذه المتجهات ليست ثابتة – بل تتغير بشكل ديناميكي حسب سياق السؤال.

تخيّل العلاقات بصرياً

تخيّل أن كل كلمة داخل دائرة، ومن هذه الدائرة تخرج خطوط متعددة تربطها بكلمات أخرى. كل خط يمثل علاقة محتملة بين الكلمة داخل الدائرة وكلمة أخرى. بعض هذه الخطوط قد تكون سميكة وقوية لأن الكلمات مترابطة بشكل وثيق، والبعض الآخر قد يكون رقيقاً وضعيفاً إذا كانت العلاقة أقل أهمية.

ملاحظة: يمكنك زيارة موقع nlpl.eu لرؤية مثال بصري على العلاقات بين الكلمات مثل: "dubai_NOUN".

[Semantically related words for "dubai_NOUN" \(nlpl.eu\)](http://nlpl.eu)

على سبيل المثال، إذا أخذت كلمة "bank" (بنك/ضفة) ووضعناها في دائرة، ستري خطوطاً تربطها بكلمات مثل "money" (مال)، و "loan" (قرض)، و "river" (نهر)، و "finance" (تمويل).

قوة هذه الروابط تعتمد على السياق. إذا كان الحديث عن البنوك، ستكون العلاقة بين "bank" و "money" قوية، أما إذا كان السياق جغرافياً، فستقوى العلاقة بين "bank" و "river".

تغيّر السياقات وتغيّر العلاقات

روعة النموذج اللغوي أنه يكيّف هذه العلاقات حسب سياق السؤال. تخيّل أنك تنتقل من الحديث عن الصحة إلى التمويل ثم إلى الجغرافيا:

• في مجال الصحة: قد ترتبط كلمة "apple" (تفاحة) بكلمات مثل "nutrition" (تغذية) و "diet" (حمية).

• في التمويل: ترتبط كلمة "bank" بكلمات مثل "loan" (قرض) و "interest" (فائدة).

• في الجغرافيا: تتحول علاقة "bank" إلى كلمات مثل "river" (نهر) و "water" (ماء).

وهذا يعني أن النموذج يضبط المتجهات (العلاقات) بين الكلمات تلقائياً ليتوافق مع المعنى الأنسب للسياق.

دور بيانات التدريب في بناء هذه العلاقات

كيف يعرف النموذج أن يضبط هذه العلاقات؟ هنا يأتي دور بيانات التدريب. أثناء التدريب، يتعرض النموذج لكمية هائلة من النصوص في مواضيع ومجالات متنوعة. يتعلّم من هذه النصوص الأنماط ويبنى متجهات أولية (أو علاقات) بين الكلمات.

في البداية، تكون هذه المتجهات عامة وقد تكون عشوائية. لكن مع مرور الوقت وقراءة المزيد، يُحسّن النموذج من هذه العلاقات. مثلاً، كلما قرأ النموذج عن البنوك والتمويل، زادت قوة العلاقة بين كلمتي "bank" و "loan".

وبمجرد أن يُدرَّب النموذج، لا يحتاج إلى الرجوع إلى البيانات الأصلية كلما استقبل محفزًا جديدًا. بل يستخدم ما تعلّمه مسبقًا – تمامًا كما لا تحتاج أنت إلى مراجعة كتاب عندما تكون قد حفظت المعلومة. النموذج يُعدل العلاقات بين الكلمات تلقائيًا حسب سياق السؤال أو المشروع.

في الخلاصة:

- الوحدات (Tokens) هي الكلمات، والعلاقات بينها تُسمى معاملات (Parameters).
- هذه العلاقات يتم تمثيلها بواسطة متجهات تقيس مدى ترابط كلمتين.
- بناءً على سياق المحفز، يقوم النموذج بتعديل قوة هذه العلاقات.
- بيانات التدريب تساعد النموذج في تعلّم هذه العلاقات، حتى يتمكن من التعامل مع أسئلة من مجالات مختلفة مثل الصحة أو التمويل أو الجغرافيا، دون الحاجة للرجوع إلى البيانات الأصلية.
- ببساطة، كل كلمة مرتبطة بالعديد من الكلمات الأخرى، وقوة هذه الروابط تتغير حسب المشروع أو المحفز الذي تعمل عليه. الأمر يشبه شبكة كلمات، والنموذج يعرف كيف يسلط الضوء على الروابط الصحيحة بناءً على ما تطلبه منه.

التعمق قليلاً: تمثيل العلاقات بالمتجهات

إذا كنت تتساءل عن الطريقة التي يتم بها تمثيل هذه العلاقات داخل النموذج اللغوي الكبير، فإليك بعض التفاصيل الإضافية. في نماذج مثل GPT، يتم تمثيل كل كلمة أو وحدة Token بواسطة متجه — وهو عبارة عن قائمة طويلة من الأرقام. عادةً ما يتكوّن هذا المتجه من 1024 بُعداً، أي أن كل كلمة ترتبط بـ 1024 رقمًا مختلفًا يُعبّرون عن معناها.

على سبيل المثال:

- قد يكون تمثيل كلمة "apple" هكذا $[0.45, -0.32, 0.78, \dots, -0.11]$:
- وقد تكون كلمة "fruit" ممثلة هكذا $[0.42, -0.30, 0.80, \dots, -0.10]$:

وبما أن كلمتي "apple" و "fruit" مرتبطتان في المعنى، فستكون متجهاتهما متشابهة. في هذه الحالة، الفرق بين المتجهين صغير، مما يدل على أن النموذج يعتبرهما متقاربين في المفهوم.

هذه الأرقام عادةً ما تقع في النطاق بين -1 و 1، وهي تعبّر عن علاقة الكلمة بكلمات أخرى داخل النموذج. الكلمات التي لها معانٍ متشابهة ستكون متجهاتها قريبة من بعضها في هذا الفضاء العالي الأبعاد، بينما الكلمات غير المرتبطة سيكون بين متجهاتها مسافات أكبر. النموذج يستخدم هذه المتجهات لفهم وتعديل العلاقات بين الكلمات حسب سياق السؤال أو المحفز.

إذن، بينما نحن نُصوّر هذه الروابط كخطوط بين كلمات داخل دوائر، فإن النموذج في الحقيقة يستخدم علاقات رياضية معقدة بين متجهات ذات 1024 بُعداً لتحديد كيف ترتبط الكلمات ببعضها. وهذا ما يمكّن النموذج من التكيف ديناميكياً وتقديم إجابات مناسبة مهما كان المجال أو الموضوع.