



قسم علوم الحاسب
كلية الحاسبات والذكاء الاصطناعي
جامعة بني سويف

مجلة قسم علوم الحاسب للعلوم المتقدمة

النماذج اللغوية الكبيرة (Large Linguistic Models)

النماذج اللغوية الكبيرة (LLM) هي نماذج تعلم عميق كبيرة جدًا مدربة مسبقًا على كميات هائلة من البيانات. المكون الأساسي هو مجموعة من الشبكات العصبية Neural Networks التي تتكون غالبًا من وحدتي تشفير Encoder وفك تشفير Decoder مع قدرات الانتباه الذاتي Self-Attention. تستخرج وحدتا التشفير وفك التشفير المعاني من تسلسل نصي، وتفهم العلاقات بين الكلمات والعبارات الموجودة فيها. على عكس الشبكات العصبية المتكررة (RNN) التي ظهرت في السابق وتعالج المدخلات بالتتابع، تعالج المحولات Transformers التسلسلات بالكامل بالتوازي. يسمح هذا لعلماء البيانات باستخدام وحدات معالجة الرسومات (GPU) لتدريب نماذج اللغة الكبيرة القائمة على المحولات، ما يقلل بشكل كبير من وقت التدريب. تسمح هيكلية الشبكة العصبية للمحول باستخدام نماذج كبيرة للغاية تحوي مئات المليارات من المعلمات في الغالب. يمكن لمثل هذه النماذج كبيرة النطاق استيعاب كميات هائلة من البيانات، غالبًا من الإنترنت، ولكن أيضًا من مصادر مثل Common Crawl الذي يضم أكثر من 50 مليار صفحة ويب بالإضافة إلى ويكيبيديا تتضمن ما يقرب من 57 مليون صفحة.

بعض تطبيقات النماذج اللغوية الكبيرة

1. كتابة المحتوى

باستثناء GPT-3 و ChatGPT، يمكن ل Claude و Llama 2 و Cohere Command و Jurassic كتابة محتوى أصلي. ويقترح AI21 Wordspice إجراء تغييرات على الجمل الأصلية لتحسين الأسلوب والصياغة.

2. تقديم الإجابات القائمة على قاعدة المعرفة

يُشار إليه غالبًا باسم معالجة اللغة الطبيعية كثيفة المعرفة (KI-NLP)، وتشير التقنية إلى نماذج اللغات الكبيرة (LLM) التي يمكنها الإجابة عن أسئلة محددة من المساعدة المعلوماتية في الأرشفات الرقمية. ومن أمثلة ذلك قدرة AI21 Studio Playground على الإجابة عن أسئلة المعرفة العامة.

3. تصنيف النصوص

عند استخدام التجميع، يمكن لنماذج اللغات الكبيرة (LLM) تصنيف النصوص التي تتشابه في المعاني أو المشاعر. وتشمل الاستخدامات قياس مشاعر العملاء، وتحديد العلاقة بين النصوص، والبحث عن المستندات.

رئيس مجلس الإدارة

د. محمد فايد

رئيس التحرير

د. أحمد النجار

رئيس الإصدارات

أ. هشام محمد

Address: New Beni-Suef City. Beni-Suef. 62111 Web Site: WWW.fci.bsu.edu.eg

Email: fci@fci.bsu.edu.eg

Telephone/Fax: 082 2246796



4. إنشاء الأكواد البرمجية

يتميز نموذج اللغة الكبير (LLM) ببراعته في إنشاء التعليمات البرمجية من أوامر اللغات الطبيعية. ومن أمثلة ذلك Amazon CodeWhisperer ومخطوطة Open AI المستخدمة في GitHub Copilot ويمكنها إنشاء تعليمات برمجية بلغة Python و JavaScript و Ruby والعديد من لغات البرمجة الأخرى. تتضمن تطبيقات إنشاء التعليمات البرمجية الأخرى إنشاء استعلامات SQL ، وكتابة أوامر shell ، وتصميم المواقع الإلكترونية.

5. إنشاء النصوص

على غرار إنشاء التعليمات البرمجية، يمكن لإنشاء النصوص إكمال جمل غير مكتملة أو كتابة وثائق المنتج أو كتابة قصة قصيرة للأطفال، مثل Alexa Create

الاعتبارات الأخلاقية لنماذج اللغة الكبيرة

لخاوف الأخلاقية المحيطة باستخدام LLM معقدة ومتعددة الجوانب منها

1. توليد محتوى ضار مثل خطاب الكراهية والدعاية المتطرفة واللغة العنصرية أو المتحيزة جنسياً وأشكال المحتوى الأخرى التي يمكن أن تسبب ضرراً لأفراد أو مجموعات معينة.
 2. الأثر الاقتصادي يمكنهم إدخال تغييرات هيكلية كبيرة في طبيعة العمل والعمالة ، مثل جعل بعض الوظائف زائدة عن الحاجة من خلال إدخال الأتمتة. قد يؤدي ذلك إلى نزوح القوى العاملة ، والبطالة الجماعية ، وتفاقم عدم المساواة القائمة في القوى العاملة.
 3. المعلومات المضللة والتأثير على العمليات يمكن أن ينتج عن ذلك محتوى يبدو واقعياً من خلال المقالات أو القصص الإخبارية أو منشورات وسائل التواصل الاجتماعي ، والتي يمكن استخدامها بعد ذلك للتأثير على الرأي العام أو نشر معلومات مضللة.
- الخصوصية يجب وضع سياسات واضحة لجمع البيانات الشخصية وتخزينها. ويجب ممارسة إخفاء هوية البيانات للتعامل مع الخصوصية بشكل أخلاقي.

إعداد

د / حاتم نعمان

دكتور بقسم علوم الحاسب - كلية الحاسبات والذكاء الاصطناعي - جامعة بني سويف

رئيس مجلس الإدارة

د. محمد فايد

رئيس التحرير

د. أحمد النجار

رئيس الإصدارات

أ/ هشام محمد

Address: New Beni-Suef City. Beni-Suef. 62111 Web Site: WWW.fci.bsu.edu.eg

Email: fci@fci.bsu.edu.eg

Telephone/Fax: 082 2246796