2025/2024

سمينار شباب الباحثين

موجز السياسات Policy Brief

"تعزيز استخدام نماذج اللغات الكبيرة وتطبيقاتها "

إعداد

أ. هديل السيد أحمد الشربيني

مدرس مساعد بمركز الأساليب التخطيطية

المنسق	المنسق المشارك
د. مي عوض	**د. إسلام خليل**
مدرس إدارة الأعمال	مدرس الاقتصاد

معهد التخطيط القومي

فبراير 2025

مقدمة

نماذج اللغات الكبيرة هي نظم ذكاء اصطناعي متقدمة تم تصميمها لفهم و إنشاء نصوص تشبه الإنسان. و هي تعتمد على علم الشبكات العصبية و علم الشبكات العصبية هو فرع من الذكاء الاصطناعي والتعلم الآلي يهدف إلى محاكاة طريقة عمل الدماغ البشري من خلال بناء أنظمة حاسوبية قادرة على التعلم واتخاذ القرارات بناءً على البيانات. تعتمد هذه الأنظمة على **الشبكات العصبية الاصطناعية (Artificial Neural Networks - ANN)**، وهي نماذج رياضية مستوحاة من الخلايا العصبية في الدماغ.

تم تدريب نماذج اللغات الكبيرة على كميات ضخمة من البيانات من مصادر مختلفة : بيانات الانترنت ، كتب ،أبحاث . و تستخدم مليارات من المعلمات و هذه المعلمات تماثل نقاط الاشتباك العصبي الموجودة في مخ الإنسان . و تعمل هذه المعلمات على إنشاء العلاقات بين الكلمات و بين الجمل بعضها البعض.

السياسات المقترحة للاستفادة من نماذج اللغات الكبيرة في مصر:

1- زيادة التوعية لدى المواطنين بكيفية استخدام نماذج اللغات الكبيرة و دعم التدريب على هندسة التوجيه.

2-الاستفادة من تطبيقات عملية في القطاعات الحيوية.

3- وضع سياسات أخلاقية وقانونية.

4- تطوير بنية تحتية رقمية قوية.

النشأة

- في عام 2017، نشر باحثو Google ورقة بحثية بعنوان "**Attention is All You Need**". و سمحت هذه التقنية للنماذج بمعالجة النصوص بشكل أكثر كفاءة مقارنة بالشبكات العصبية التقليدية.

- **2018-2019: ظهور النماذج الأولى القائمة على المحولات**

  - ✓ **في عام 2018 أنتجت google نموذج Bert** و هو أول نموذج يستخدم **التعلم ثنائي الاتجاه**، مما يعني أنه يفهم السياق من الجانبين (قبل وبعد الكلمة).
  - ✓ **في عام 2019 أنتجت openAI نموذج GPT2** و هو نموذج **توليدي** قادر على إنتاج نصوص طبيعية للغاية.

- **2020-2021: ازدهار نماذج اللغة الضخمة**

  - ✓ **في عام 2020 ظهر GPT3** و كانت له القدرة على الترجمة، البرمجة، توليد النصوص الإبداعية، والمحادثات البشرية، و أدى إلى ظهور أدوات مثل **ChatGPT** المبنية على GPT-3.5و GPT-4لاحقًا.
  - ✓ **في عام 2021 أنتجت openAI نموذج codex** و هو نموذج متخصص في البرمجة.

- **2024 - حتى الآن: التوجه نحو الذكاء المتعدد الوسائط والذاكرة طويلة المدى**

  - ✓ عام **2024 أنتجت Google نموذج Gemini** و هو يتمكن من فهم الوثائق الضخمة والمحادثات الطويلة.
  - ✓ عام **2024 أنتجت OpenAI نموذج Sora** و هو نموذج قادر على **إنشاء مقاطع فيديو** بناءً على الأوصاف النصية.

الأنواع

- **حسب طريقة العمل:** النماذج التوليدية (التنبؤية) ، النماذج الفهمية، نماذج تحويل النصوص.
- **حسب الغرض:** أغراض عامة ،أغراض مخصصة لمجالات معينة، نماذج متعددة الوسائط (تتعامل مع الصوت والصور).

- **حسب إمكانية الوصول:** نماذج مغلقة المصدر، نماذج مفتوحة المصدر.
- **حسب الكفاءة و الحجم:** نماذج كبيرة الحجم (تحتوي على مليارات من المعلمات) ، نماذج صغيرة الحجم (تحتوي على عدد أقل من المعلمات).

طرق التحسين :

- **الضبط Fine tuning :** تدريب إضافي يتم إجراؤه على نماذج اللغات الكبيرة تم تدريبه مسبقًا، وذلك بهدف تحسين أدائه في مهمة محددة أو مجال معين، نحتاج هذه التقنية لجعل النموذج أكثر دقة ، لتقليل التحيز، تحسين الكفاءة.
- **هندسة التوجيه Prompt Engineering:** هو فن صياغة المدخلات بطريقة تجعل النموذج يقدم أفضل وأدق الإجابات.
- **استرجاع المعلومات المعزز بالإنتاجRetrieval Augmented Generation**: تهدف إلى تحسين إجابات النماذج اللغوية الكبيرة من خلال البحث عن معلومات حديثة من مصادر خارجية قبل توليد الإجابة.

التطبيقات الرئيسية:

- **توليد النصوص :** الكتابة الإبداعية، الدردشة الآلية ، ترجمة اللغة.
- **إنشاء الأكواد البرمجية :** المساعدة في توليدها أو استكمالها أو إصلاح الأخطاء.
- **فهم المحتويات المرئية:** وصف الصور ، إنشاء التسمية التوضيحية ، إنشاء الفيديوهات.

تطبيقات في بعض المجالات:

فيما يلي نستعرض تطبيقات نماذج اللغات الكبيرة في بعض المجالات.

- **مجال الصحة:** تحسين التشخيص، البحث العلمي، توصيات العلاج.
- **مجال التعليم:** سهولة التعلم الذاتي، سهولة الوصول إلى المصادر التعليمية، تصميم المحتوى التعليمي.
- **مجال الأعمال**: اتخاذ القرارت بشكل أكثر دقة، تحليل آراء العملاء و الإجابة على استفسارتهم. استخراج الرؤى أو الأفكار الهامة من التعاقدات ، الأوراق البحثية ، السياسات المختلفة. توقع احتياجات السوق.
- **مجال الزراعة:** القدرة على تحليل مراكز البيانات الكبيرة للتربة والمحاصيل وبيانات الطقس جنبا إلى جنب مع صور الأقمار الصناعية.، تقدم هذه النماذج توصيات بشأن أوقات الزراعة والري واستخدام الأسمدة وتحسين الحقول والموارد.يمكن للمزارعين الحصول على متطلبات السوق، والتنبؤ بأسعار المحاصيل ، وتوقع الكوارث الطبيعية.

التحديات

التكلفة الحسابية: يتم تدريب نماذج اللغات الكبيرة على بيانات ضخمة جدا بمليارات المعلمات و بالتالي تحتاج موارد حسابية كبيرة و بالتالي تتسبب في استهلاك كبير للطاقة مما يرفع أيضا المخاوف البيئية.

التحيز والإنصاف: بسبب تدريب نماذج اللغات الكبيرة على بيانات الإنترنت ، يمكن أن تحتوي هذه البيانات على معلومات متحيزة و من الممكن أن تكون بعض الآراء أو المجموعات غير ممثلة بشكل جيد في البيانات التي تم التدريب عليها.

الإفراط في التجهيز: هو مشكلة في تعلم الآلة تحدث عندما يكون النموذج مدربًا بشكل مفرط على بيانات التدريب، مما يجعله يتعلم الأنماط والتفاصيل الدقيقة لهذه البيانات، بدلاً من تعميم المعرفة لاستخدامها على بيانات جديدة. ونتيجة لذلك، يكون أداء النموذج جيدًا جدًا على بيانات التدريب لكنه ضعيف على بيانات الاختبار أو في سيناريوهات العالم الحقيقي.

الهلوسة: الهلوسة في نماذج اللغات الكبيرة (LLMs Hallucination) تشير إلى ظاهرة إنتاج النموذج لمعلومات غير صحيحة، غير دقيقة، أو حتى مختلقة بالكامل، رغم أنها قد تبدو منطقية أو موثوقة لغويًا. تحدث هذه المشكلة عندما يولد النموذج استجابات لا تستند إلى حقائق أو بيانات صحيحة ولكنها تعتمد على الأنماط اللغوية التي تعلمها أثناء التدريب.

عدم المساواة الاقتصادية و البحثية: بسبب التكلفة العالية لتدريب نماذج اللغات الكبيرة فهي مقتصرة على المنظمات ذات الدخل المرتفع.

السلامة و التحكم: أحيانا يمكن أن تقدم نماذج اللغات الكبيرة محتويات غير مناسبة أو غير لائقة و بالتالي يجب أن يتم التأكد بأن النموذج يستخدم بطريقة آمنة و لأهداف محددة.

الأمان و الخصوصية: نماذج اللغات الكبيرة هي عرضة لتسريب معلومات شخصية و ذلك لأنها مدربة على بيانات الإنترنت .

إلى أين تتجه نماذج اللغات الكبيرة؟

- نماذج أصغر وأسرع وأكثر كفاءة.
- دعم الذكاء متعدد الوسائط (الصوت، الصور، الفيديو).
- تحسين الدقة وتقليل الأخطاء.
- نماذج متخصصة لكل مجال.
- زيادة الأمان والخصوصية و ضمان عدم التحيز.

السياسات المقترحة للاستفادة من نماذج اللغات الكبيرة في مصر

1) **زيادة التوعية لدى المواطنين بكيفية استخدام نماذج اللغات الكبيرة و دعم التدريب على هندسة التوجيه.**
2) **تطبيقات عملية في القطاعات الحيوية:**

التعليم:
استخدام نماذج اللغات الكبيرة في تطوير أنظمة تعليمية ذكية لتحسين جودة التعليم الإلكتروني.
توفير منصات تعليمية تفاعلية تعتمد على الذكاء الاصطناعي لدعم التعلم الذاتي.

الصحة:
تطوير تطبيقات لمساعدة الأطباء في التشخيص وتوفير استشارات صحية عبر الذكاء الاصطناعي.
تحسين سجلات المرضى وتحليل البيانات الصحية لاكتشاف الأمراض مبكرًا.

الخدمات الحكومية:
استخدام روبوتات الدردشة الذكية لتحسين خدمات المواطنين والرد على الاستفسارات الحكومية.
تعزيز الشفافية وتحليل البيانات الضخمة لاتخاذ قرارات مبنية على الأدلة.

القطاع المالي:
تحسين تقنيات مكافحة الاحتيال في البنوك من خلال تحليل البيانات باستخدام نماذج الذكاء الاصطناعي.
تطوير روبوتات استشارية لمساعدة العملاء في إدارة استثماراتهم.

3) وضع سياسات أخلاقية وقانونية:

ضمان الخصوصية وحماية البيانات :وضع لوائح صارمة لحماية بيانات المستخدمين من سوء الاستخدام.

مراقبة المحتوى وتقليل التحيز :التأكد من أن نماذج اللغات الكبيرة لا تنشر معلومات مضللة أو تعزز التحيزات الاجتماعية.

توفير تشريعات تنظم استخدام الذكاء الاصطناعي :تطوير قوانين تحكم كيفية استخدام هذه النماذج في القطاعات المختلفة.

4) تطوير بنية تحتية رقمية قوية:عن طريق الاستثمار في إنشاء وتوسيع مراكز البيانات الوطنية لدعم تشغيل نماذج اللغات الكبيرة محليًا. و توفير بنية تحتية متقدمة للاتصالات والإنترنت لضمان كفاءة تشغيل هذه النماذج.