الذكاء الاصطناعي التوليدي

النماذج اللغوية الكبيرة

أغسطس 2023

محتويات

5	مقدمة
6	نبذة تعريفية
7	لمحة تاريخية
8	حالات الاستخدام
9	حلول تجارية
10	هندسة المدخلات
14	التحديات والمخاطر
16	أفضل الممارسات
17	توقعات مستقبلية
18	مراجع
20	ملحق: أبرز النماذج والتطبيقات

This image was created using AI

مقدمة

تتسم تقنيات البيانات والذكاء الاصطناعي بسرعة تطورها واتساع انتشارها وتعدد حالات استخدامها، حتى أصبحت تطبيقاتها تظهر بشكل يومي في جميع نواحي الحياة لأتمتة المهام أو تعزيز دور العاملين عليها، ففي جانب الأتمتة انتشرت تقنيات توليد المحتوى النصي، ويعود الفضل في هذا الانتشار إلى التقدم التقني في الذكاء الاصطناعي التوليدي وخاصة ما يعرف بالنماذج اللغوية الكبيرة (Large Language Models - LLMs) التي تندرج تحت ما يسمى النماذج التأسيسية (Foundation Models).

وقد زاد الاهتمام بالنماذج اللغوية الكبيرة بعد أن أطلقت شركة أوبن أي آي (OpenAI) خدمة لتجربة نموذج شات جي بي تي (ChatGPT) مجاناً في عام 2022م، إذ بلغ عدد مستخدمي الخدمة أكثر من مليون مستخدم في أول خمسة أيام من إطلاقها. وتعد هذه النماذج فاعلة جداً وقادرة على تنفيذ أنواع مختلفة من المهام، مثل: إنشاء النصوص والترجمة والتلخيص بلغات متعددة. ومن المتوقع أن تشكل حالات استخدام النماذج التأسيسية بشكل عام (50%) من حالات استخدام معالجة اللغات الطبيعية في عام 2026م، بينما كانت تمثل أقل من (5%) من حالات الاستخدام في عام 2021م[1].

يهدف هذا التقرير إلى تقديم لمحة مختصرة عن النماذج اللغوية الكبيرة وحالات استخدامها، واستعراض أبرز النماذج من شركات ومؤسسات أكاديمية، إضافةً إلى ذكر بعض التحديات والمخاطر التي قد تنشأ مع انتشار هذه النماذج بصورة عامة، والإشارة إلى أفضل الممارسات في تطوير تلك النماذج، والتطرق إلى بعض التوقعات المستقبلية.

This image was created using AI

[1] https://www.gartner.com/document/4022570

نبذة تعريفية

النموذج اللغوي (Language Model) هو نموذج ذكاء اصطناعي يحدد احتمالية وجود سلسلة معينة من الكلمات في جملة ما، وعندما يتكون النموذج من عدد كبير من المُعاملات يطلق عليه نموذج لغوي كبير (Large Language Model). وتندرج النماذج اللغوية الكبيرة تحت مجموعة من النماذج تعرف بالنماذج التأسيسية (Foundation Models)، وهي نماذج مدربة على بيانات مختلفة ويمكن تكييفها لأداء مجموعة واسعة من المهام.

تختلف النماذج اللغوية الكبيرة عن نماذج اللغات الطبيعية التقليدية في جانبين مهمين هما:

This image was created using AI

حجم بيانات التدريب

تُدرب النماذج اللغوية الكبيرة باستخدام كميات هائلة من البيانات التي تحتوي على ملايين من أنماط الكلمات المعقدة والارتباطات.

عدد المُعاملات

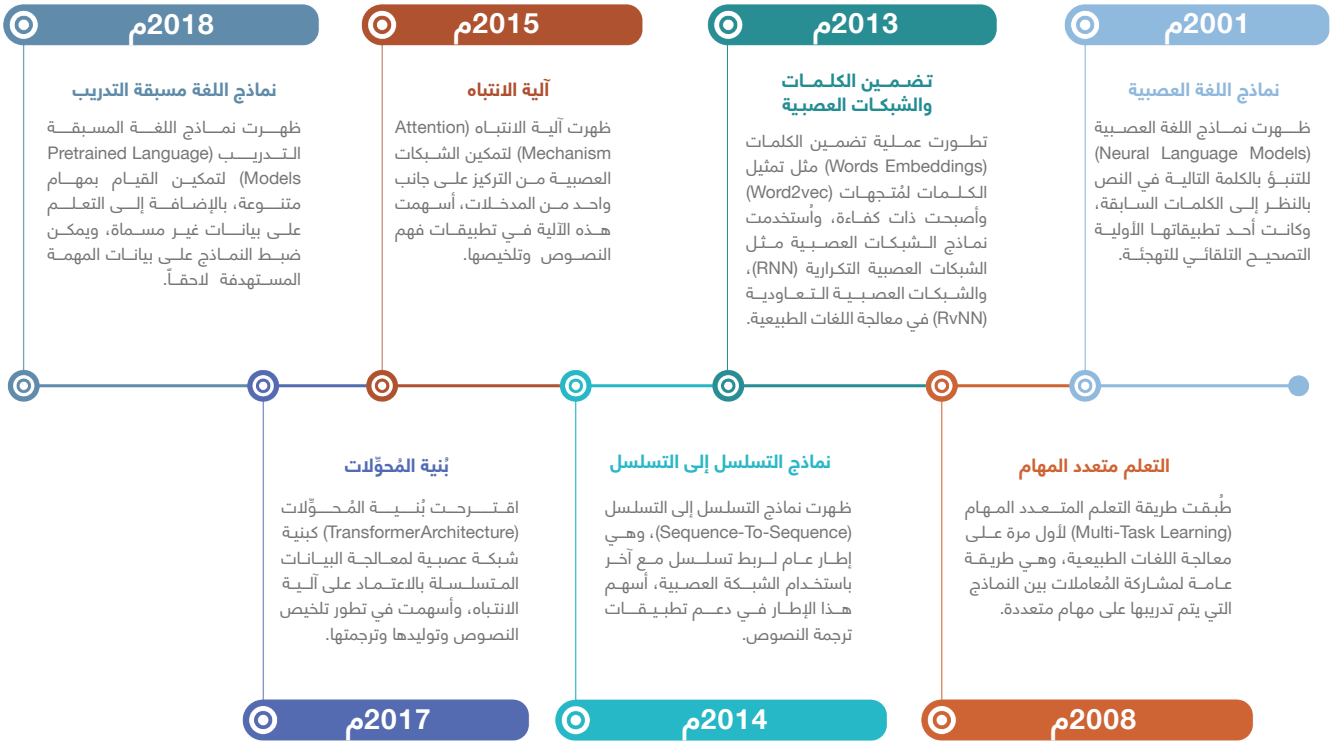
تمتلك النماذج اللغوية الكبيرة عدداً كبيراً من مُعاملات يصل في بعض النماذج إلى أكثر من تريليون مُعامل.

ويتضمن تطوير النماذج اللغوية الكبيرة ثلاث خطوات رئيسية هي:

1. جمع مجموعة بيانات من مستندات نصية.
2. تدريب الخوارزميات لفهم العلاقة بين الكلمات.
3. تقييم النماذج بصورة متكررة وضبطها.

لمحة تاريخية

ظهرت النماذج اللغوية في أوائل الثمانينات من النماذج الاحتمالية لتوليد اللغات الطبيعية التي طُورت لأنظمة التعرف الآلي على الكلام (ASR)، وتطورت خلال السنوات الأخيرة حتى وصلت في وقتنا الحاضر إلى ما يسمى بالنماذج اللغوية الكبيرة. فيما يلي استعراض لأهم الأحداث التي أسهمت في تشكّل النماذج اللغوية الكبيرة:



حالات الاستخدام

تتميز النماذج اللغوية الكبيرة بقدرتها على القيام بعدة مهام مقارنةً بالنماذج المتخصصة التي تقوم بمهمة معينة، ومن أبرز حالات استخدامها:

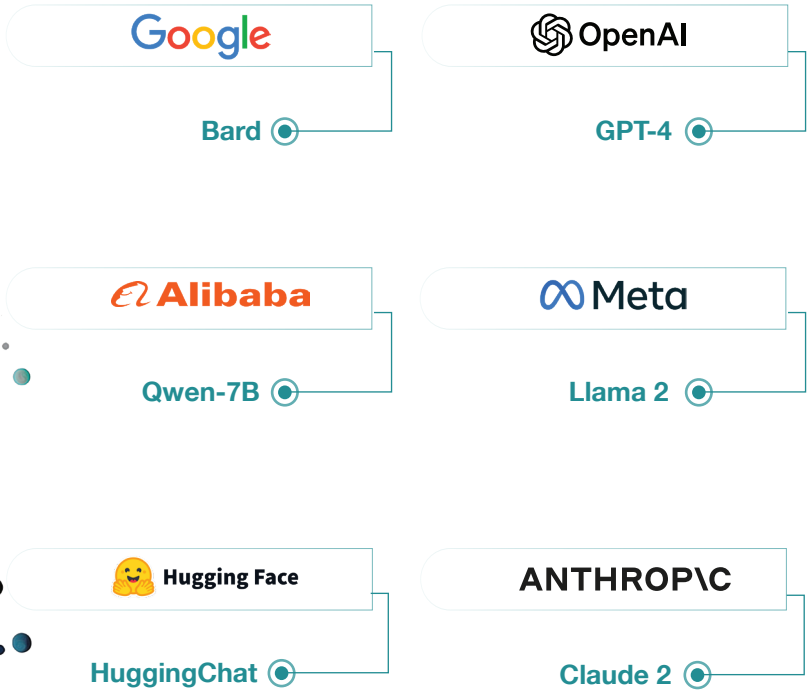
- كتابة النصوص والمقالات
- تلخيص وتبسيط النصوص
- الترجمة من لغة إلى لغة أخرى
- الإجابة عن الأسئلة
- تحليل المشاعر
- تصنيف العناصر إلى فئات
- كتابة الأكواد البرمجية
- استخراج الكلمات المفتاحية
- كتابة الإعلانات التجارية
- التدقيق النحوي

This image was created using AI

حلول تجارية

تسابقت كثير من الشركات التقنية الكبيرة وبعض المؤسسات الأكاديمية لإصدار نماذج لغوية كبيرة بإمكانات متفاوتة تعتمد على حجم بيانات التدريب وتنوع مصادرها وطرق معالجة مخاطرها، وبعضها متاحة تجارياً وبعضها متاحة بمصدر مفتوح، ومن أبرز تلك النماذج والتطبيقات: جي بي تي 4 (GPT-4) وبارد (Bard) وكلود 2 (Claude 2) ولاما 2 (Llama 2). ولمزيد من التفاصيل حول أبرز النماذج والتطبيقات يمكن الاطلاع على الملحق.

This image was created using AI



هندسة المدخلات

يعتمد الاستخدام الأمثل لتطبيقات النماذج اللغوية الكبيرة على ما يسمى بهندسة المدخلات (Prompt Engineering)، وهي صياغة تعليمات أو استفسارات فاعلة ومحددة لتوجيه هذه النماذج إلى توليد النتائج المطلوبة، وفيما يلي مجموعة من الإرشادات المفيدة في هندسة المدخلات:

1 توضيح سياق الطلب

من المهم تزويد النموذج بمعلومات كافية عن تفاصيل الطلب، على سبيل المثال: عند طلب المساعدة في كتابة رسالة عبر البريد الإلكتروني، ينبغي تزويد النموذج بدور المرسل وعلاقته بالمرسل إليه ومعلومات عن المرسل والنتيجة المراد تحقيقها وأي تفاصيل أخرى.

مثال

أنا مسوق، أُكتب رسالة إلكترونية تسويقية باللغة الإنجليزية إلى مجموعة من العملاء حول منتج جديد (مع ذكر مواصفات المنتج).

2 تعيين شخصية للنموذج

إعطاء النموذج دوراً مخصصاً لمهمة معينة، وفيما يلي بعض الأمثلة:

شخصية مترجم: "أنت مترجم للغة اليابانية تستطيع الترجمة من اللغة اليابانية إلى اللغة الإنجليزية...".

شخصية باحث: "أنت باحث خبير في الذكاء الاصطناعي لديك القدرة على جمع وتحليل البيانات والمعلومات لتقديم رؤى قيمة...".

شخصية كاتب: "أنت كاتب محتوى إبداعي قادر على صياغة محتوى جذاب في مواضيع مختلفة...".

مثال

أنت باحث خبير في الذكاء الاصطناعي لديك القدرة على جمع وتحليل البيانات والمعلومات لتقديم رؤى قيمة: أُكتب مقالاً عن الذكاء الاصطناعي التوليدي والتوقعات المستقبلية له.

3 استخدام رموز للتحديد

استخدام رموز لتحديد أجزاء من النص المدخل بهدف توفير إرشادات واضحة للنموذج فيما يتعلق بالنص الذي يتطلب الترجمة أو إعادة الصياغة أو مهام أخرى مماثلة، إذ تساعد على إنشاء حدود واضحة بين نص معين والمدخلات المتبقية، ولها أشكال مختلفة، منها: علامات الاقتباس الثلاثية (""") أو الشرطات الثلاثية (---)، أو أقواس الزاوية (<>).

مثال

لخص النص الآتي:
""" إضافة نص """

4 طلب مخرجات مهيكلة

تحديد صيغة المخرجات للنموذج عند كتابة المدخلات يحسن نتائج النموذج ويقدمها بصيغة مناسبة لحاجة المستخدم، ويمكن للمستخدم اختيار الهيكل المناسب من بين خيارات متنوعة، بما في ذلك الجداول أو بصيغة (JASON) أو (HTML).

مثال

أُكتب مقارنة بين منتج (أ) ومنتج (ب) في جدول.

5 التحقق من صحة المدخلات

التحقق من صحة المدخلات خطوة مهمة لتوفير الوقت وضمان الحصول على مخرجات صحيحة، لذا يمكن إرشاد النموذج إلى التحقق من استيفاء مدخلات المستخدم لشروط محددة ورفض إتمام المهمة في حال عدم استيفاء هذه الشروط.

مثال

عند تزويدك بنص محاط بعلامات اقتباس ثلاثية، إذا كان يحتوي على سؤال متعلق بالتعليم، فقم بتقديم إجابة، وإذا لم يكن يحتوي على سؤال متعلق بالتعليم، فقم بالرد بـ "السؤال لا علاقة له بالتعليم".

6 تقديم أمثلة ناجحة

تقديم أمثلة ناجحة للنموذج قبل تنفيذ مهام محددة يساعد في توجيه النموذج إلى النتيجة المطلوبة، وتعد هذه الطريقة مفيدة بصورة خاصة عندما يكون الطلب لمحاكاة نمط معين من الإجابات عن استفسارات المستخدم التي قد يكون من الصعب التعبير عنها بصورة مباشرة.

مثال

أُكتب عبارة ملهمة عن الأمل والتفاؤل بطريقة مشابهة للنص الآتي:

"""عش حياتك كل يوم كما لو كنت ستصعد جبلاً، وانظر بين الفينة والأخرى إلى القمة حتى لا تنسى هدفك، ولكن دون إضاعة الفرصة لرؤية المناظر الرائعة في كل مرحلة"""

7 تحديد الخطوات المطلوبة

تحديد الخطوات المطلوبة لإتمام طلبٍ ما يعزز موثوقية مخرجات النموذج ويزيد احتمالية إتمام المهمة بنجاح؛ إذ تُساعد النموذج على تقديم مخرجات شاملة ومنظمة وتحديداً عندما تكون المدخلات معقدة.

مثال

نفذ الإجراءات الآتية:

1. حلل التغريدات الآتية: """اذكر التغريدات"""
2. صنف كل تغريدة على أنها إيجابية أو سلبية.
3. ضع التصنيف في جدول.
4. وضح نسبة التغريدات الإيجابية والتغريدات السلبية.

8 التحقق من المخرجات

من المهم مطالبة النموذج بمراجعة مخرجاته لتقليل الأخطاء المحتملة وضمان جودة المخرجات ودقتها بالإضافة إلى زيادة الثقة في مخرجات النموذج؛ فقد يتجاهل النموذج الأخطاء أو التفاصيل المهمة بسبب طول المدخلات أو تشعبها.

مثال

هل نسيت أي تفاصيل مهمة في إجاباتك السابقة؟

9 استخدام مراجع محددة

إعطاء النموذج مراجع محددة وتوجيهه إلى استخدامها في صياغة ردود دقيقة وموثوقة، تُعد هذه طريقة فاعلة في تقليل هلوسة النموذج أثناء استخدامه لتوليد إجابات من المراجع، كما تعزز الشفافية والمساءلة؛ إذ تتيح للمستخدم الحصول على فهم أوضح لمصادر النتائج.

مثال

استخدم الملفات المرفقة للإجابة عن السؤال الآتي:
ما هي أبرز التحديات التي تواجه المدن الذكية؟

10 تطبيق طريقة التكرار

للمساعدة في تحسين المخرجات يمكن إعادة الطلب أكثر من مرة بصيغ مختلفة والتأكد من وضوح المدخلات عند الحصول على مخرجات غير مرضية، إذ يساعد التكرار على فهم توضيح المطلوب وتشكيل سلوك النموذج وتحسين مخرجاته.

مثال

الطلب الأول

كيف أستخدم الذكاء الاصطناعي التوليدي في مجال التعليم؟

الطلب الثاني

ما هي أفضل حالات استخدام الذكاء الاصطناعي التوليدي في مجال التعليم؟

This image was created using AI

التحديات والمخاطر

التحديات

تواجه عملية تطوير النماذج اللغوية الكبيرة عدة تحديات من أهمها:

ارتفاع التكلفة المادية

يتطلب بناء نماذج لغوية كبيرة وتدريبها تكاليف باهظة تتأثر بعدة عوامل من أبرزها: حجم البيانات، وعدد المُعاملات، وعدد مرات التدريب، وقدرت دراسة في عام 2020م تكلفة تدريب نموذج يحتوي على (1.5) مليار مُعامل بمبلغ يصل إلى (1.6) مليون دولار أمريكي (أي حوالي 6 ملايين ريال سعودي)[2].

غياب الحوكمة

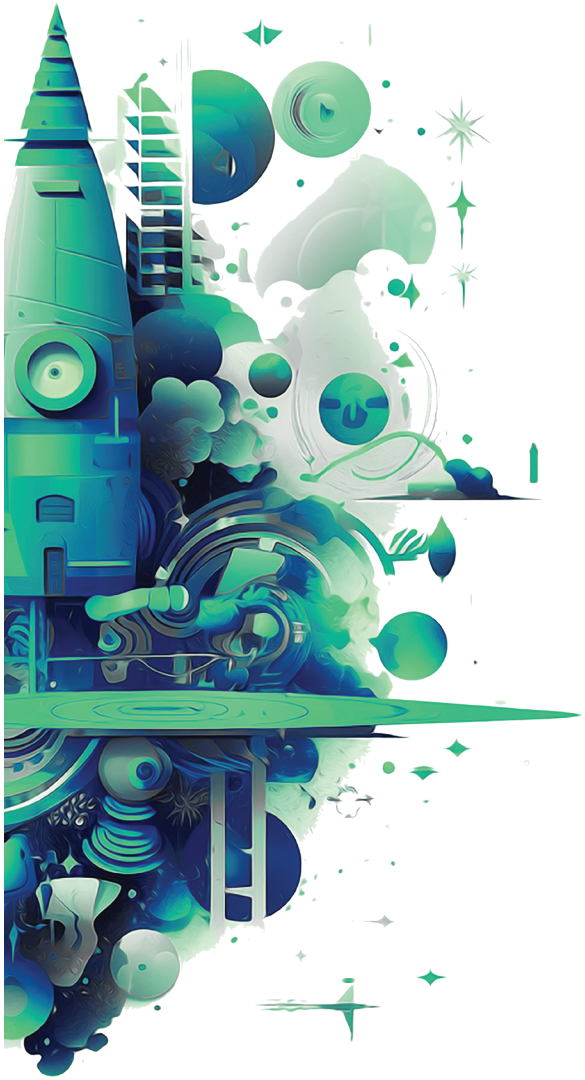
تشمل بعض القوانين والسياسات المتعلقة بحقوق النشر وخصوصية البيانات ودقة الخوارزميات النماذج اللغوية الكبيرة والمحتوى الذي تنتجه، إلا أنها لا تغطيها بشكل كامل، ولا تزال هذه النماذج تفتقر إلى وجود قوانين مخصصة بحوكمتها بشكل رئيسي.

شح الموارد والكفاءات

يتطلب تطوير النماذج اللغوية الكبيرة وتشغيلها توفير كميات ضخمة من البيانات النصية ذوات الجودة العالية، وموارد حوسبية هائلة للمعالجة، وخبرات وكفاءات متخصصة في المجال لبناء النماذج واختبارها ونشرها.

التعدي على الملكية الفكرية

تُدرب النماذج اللغوية الكبيرة على مجموعات بيانات ضخمة بعضها متاح بتراخيص تتطلب الحصول على أذونات عند إعادة استخدامها مثل بعض الأكواد البرمجية، على سبيل المثال واجهت شركة مايكروسوفت (Microsoft) دعوى قضائية جماعية عند بناء نظام جيت هب كوبايلت (GitHub Copilot).



This image was created using AI

[2] http://arxiv.org/abs/2004.08900

المخاطر

يرافق تبني النماذج اللغوية الكبيرة وانتشارها عدة مخاطر من أهمها:

التحيز

قد تعاني النماذج اللغوية الكبيرة من التحيز في مخرجاتها، مما قد يسبب أضراراً اجتماعية كالتفضيل في المُعاملات أو السماح بالوصول إلى بعض الخدمات بناءً على بعض السمات كالجنس، والدين، والعمر. على سبيل المثال أشار باحثون من شركة أوبن أي آي (OpenAI) إلى وجود تحيز في نموذج (جي بي تي-3) ضد بعض الفئات كالعرق الأسود وديانة الإسلام[3].

المعلومات الخاطئة

توّلد النماذج اللغوية الكبيرة كميات هائلة من النصوص دون التحقق من دقتها، مما قد يتسبب في نشر معلومات خاطئة، فعلى سبيل المثال من الممكن أن تقوم هذه النماذج بإنشاء أخبار أو تغريدات أو مقالات خاطئة تؤثر في الرأي العام. وقد قام باحثون من جامعة أكسفورد (University of Oxford) وشركة أوبن أي آي (OpenAI) بقياس مدى صواب إجابات النماذج اللغوية الكبيرة، ووجدوا أن نسبة صواب أفضل نموذج كانت (58%) بينما وصلت نسبة صواب الأداء البشري إلى (94%)[4].

تسرب البيانات

قد تتسبب النماذج اللغوية الكبيرة في تسرب معلومات شخصية أو سرية، وذلك في حال استخدم بيانات حساسة في مرحلة التدريب، أو قامت النماذج باستنتاجها عن شخص ما بناءً على وجود بيانات مرتبطة بأشخاص آخرين، أو عبر الهجمات السيبرانية على هذه النماذج، على سبيل المثال أعلنت شركة أوبن أي آي (OpenAI) عن تسرب بيانات شخصية لـ(1.2%) من المشتركين في (ChatGPT Plus) خلال فترة زمنية محددة وتتضمن البيانات المسربة: الاسم والبريد الإلكتروني ومعلومات الدفع[5].

أضرار بيئية

تعتمد معالجة النماذج اللغوية الكبيرة على موارد حوسبة ضخمة تستهلك كميات هائلة من الموارد الطبيعية. وقدّرت دراسة من جامعة ماساتشوستس أمهيرست (University of Massachusetts Amherst) أن تدريب النموذج اللغوي الكبير بيرت (BERT) ينتج قدراً كبيراً من الكربون يعادل ما تنتجه رحلة طيران من شرق الولايات المتحدة الأمريكية إلى غربها[6].

[3] http://arxiv.org/abs/2005.14165

[4] https://aclanthology.org/2022.acl-long.229

[5] https://openai.com/blog/march-20-chatgpt-outage

[6] http://arxiv.org/abs/1906.02243

أفضل الممارسات

نشرت شركة أوبن أي آي (OpenAI) بالتعاون مع شركتي كوهير (Cohere) وأي آي 21 لابز (AI21 Labs) أفضل الممارسات لتطوير النماذج اللغوية الكبيرة واستخدامها للمساهمة في تطوير النماذج اللغوية الكبيرة والتخفيف من مخاطرها[7]:

التأكيد على حسن الاستخدام

- الحرص على نشر إرشادات وشروط استخدام واضحة للنماذج بهدف حماية الأفراد والمجتمع من المعلومات المضللة كرسائل البريد الإلكتروني العشوائية أو التسويق الزائف.
- تحديد المجالات التي تتطلب توخي مزيد من الحيطة والحذر عند استخدام النماذج، بالإضافة إلى حظر حالات الاستخدام غير المناسبة، مثل: تصنيف الفرد بناءً على صفاته الشخصية كالعمر أو العرق.
- بناء أنظمة وبنية تحتية تفرض إرشادات الاستخدام التي قد تتضمن: تصفية المحتوى، والموافقة على التطبيق قبل مرحلة الإنتاج، ومراقبة النشاط غير الطبيعي.

تخفيف الضرر غير المقصود

- إجراء تقييم شامل للنماذج لمعرفة القيود والتقليل من مصادر التحيز المحتملة في بيانات التدريب.
- استخدام تقنيات تقلل من السلوك غير الآمن كالتعلم من المراجعات البشرية.
- توثيق أفضل ممارسات السلامة في استخدام النماذج، ونقاط الضعف المكتشفة.

التعاون مع أصحاب المصلحة

- تكوين فريق عمل من خلفيات متنوعة، إذ تساعد وجهات النظر المختلفة على تحديد كيفية عمل النماذج على أرض الواقع، وتساهم في التقليل من التحيزات.
- الإفصاح عن الدروس المستفادة فيما يتعلق بسلامة النموذج وسوء الاستخدام لتمكين تبنيها على نطاق واسع.
- التعامل مع جميع الموظفين المعنيين بتطوير النماذج اللغوية باحترافية، ويشمل ذلك توفير بيئة عمل بمعايير عالية.

[7] https://openai.com/blog/best-practices-for-deploying-language-models

توقعات مستقبلية

%50

من الأكواد البرمجية لمواقع وتطبيقات جديدة ستستخدم الذكاء الاصطناعي التوليدي بحلول عام 2026م[8]

10-5 سنوات

ستصل النماذج اللغوية الكبيرة إلى مرحلة الإنتاجية وسينخفض مستوى مخاطرها بصورة كبيرة[10]

10 أضعاف

ستزيد حالات استخدام النماذج اللغوية الكبيرة في عام 2026م مقارنة بعام 2021م[11]

+37 مليار ريال سعودي

ستُستثمر في الشركات الناشئة التي تعتمد على النماذج التأسيسية حتى عام 2026م[9]

This image was created using AI

[8] https://www.gartner.com/document/code/757883

[9] https://www.gartner.com/document/4022922

[10] https://www.gartner.com/interactive/hc/4016433

[11] https://www.gartner.com/document/4022570

مراجع

- Ecosystem Graphs for Foundation Models. https://crfm.stanford.edu/ecosystem-graphs/index.html.
- Reflections on Foundation Models. https://hai.stanford.edu/news/reflections-foundation-models (2021).
- Saudi Data & AI Authority. Data and AI Glossary. https://sdaia.gov.sa/ar/MediaCenter/KnowledgeCenter/ResearchLibrary/SDAIAPublications15.pdf (2022).
- What's in the Chatterbox? Large Language Models, Why They Matter, and What We Should Do About Them | Science, Technology and Public Policy (STPP). https://stpp.fordschool.umich.edu/research/research-report/whats-in-the-chatterbox (2022).
- Zhao, W. X. et al. A Survey of Large Language Models. http://arxiv.org/abs/2303.18223 (2023).

This image was created using AI

ملحق: أبرز النماذج والتطبيقات

الاسم	النوع	المطور	تاريخ الإطلاق	بيانات التدريب	لغة بيانات التدريب	عدد المُعاملات	الإتاحة
GPT-3	نموذج	OpenAI	يونيو 2020م	نصوص	الإنجليزية ولغات أخرى	175 مليار مُعامل	متاح تجارياً لمعظم الدول عبر واجهة برمجة التطبيقات (API)
Wu Dao 2.0	نموذج	BAAI 智源研究院	يناير 2021م	نصوص صور	الإنجليزية الصينية	1.75 تريليون مُعامل	مفتوح المصدر
Switch-C	نموذج	Google	يناير 2021م	نصوص	الإنجليزية ولغات أخرى	1.6 تريليون مُعامل	مفتوح المصدر
PanGu-Alpha	نموذج	HUAWEI	أبريل 2021م	نصوص	الصينية ولغات أخرى	200 مليار مُعامل	غير معروف
LaMDA	نموذج	Google	مايو 2021م	حوارات نصوص	الإنجليزية ولغات أخرى	137 مليار مُعامل	غير معروف
HyperCLOVA	نموذج	NAVER	سبتمبر 2021م	نصوص	الكورية ولغات أخرى	82 مليار مُعامل	متاح تجارياً لعملاء الشركة عبر أستوديو (HyperCLOVA Studio) بنسخة بيتا
MT-NLG	نموذج	Microsoft NVIDIA	أكتوبر 2021م	نصوص أكواد برمجية	الإنجليزية ولغات أخرى	530 مليار مُعامل	متاح للمنظمات عبر واجهة برمجة التطبيقات (API)

الاسم	النوع	المطور	تاريخ الإطلاق	بيانات التدريب	لغة بيانات التدريب	عدد المُعاملات	الإتاحة
RETRO	نموذج	Google DeepMind	ديسمبر 2021م	نصوص أكواد برمجية	10 لغات	7.5 مليارات مُعامل	غير معروف
PaLM	نموذج	Google	أبريل 2022م	حوارات نصوص أكواد برمجية	الإنجليزية ولغات أخرى	540 مليار مُعامل	غير معروف
NOOR	نموذج	TII Technology Innovation Institute	أبريل 2022م	نصوص	العربية	10 مليارات مُعامل	قريباً
OPT-175B	نموذج	Meta	مايو 2022م	نصوص	الإنجليزية ولغات أخرى	175 مليار مُعامل	متاح مجاناً للباحثين في المنظمات الحكومية والأوساط الأكاديمية، والمجتمع المدني.
BLOOM	نموذج	BigScience	يوليو 2022م	نصوص أكواد برمجية	46 لغة 13 لغة برمجية	176 مليار مُعامل	مفتوح المصدر
ChatGPT	تطبيق	OpenAI	نوفمبر 2022م	حوارات	الإنجليزية ولغات أخرى	175 مليار مُعامل	متاح تجارياً في معظم الدول عبر واجهة برمجة التطبيقات (API)
LLaMA	نموذج	Meta	فبراير 2023م	نصوص	20 لغة	65 مليار مُعامل	متاح مجاناً للباحثين الأكاديميين

الاسم	النوع	المطور	تاريخ الإطلاق	بيانات التدريب	لغة بيانات التدريب	عدد المُعاملات	الإتاحة
Bard	تطبيق	Google	فبراير 2023م	غير معروف	غير معروف	غير معروف	متاح للعامة
Falcon LLM	نموذج	TII Technology Innovation Institute	مارس 2023م	النصوص	الإنجليزية ولغات أخرى	40 مليار مُعامل	مفتوح المصدر
Claude	نموذج	ANTHROP\C	مارس 2023م	غير معروف	غير معروف	غير معروف	متاح تجارياً
GPT-4	نموذج	OpenAI	مارس 2023م	غير معروف	غير معروف	غير معروف	متاح تجارياً للمشتركين في (ChatGPT Puls
Tongyi Qianwen	نموذج	Alibaba	أبريل 2023م	غير معروف	غير معروف	غير معروف	متاح لعملاء الشركة في الصين بنسخة بيتا
HuggingChat	تطبيق	Hugging Face	أبريل 2023م	غير معروف	غير معروف	غير معروف	مفتوح المصدر
ALLAM	تطبيق	SDAIA الهيئة السعودية للبيانات والذكاء الاصطناعي Saudi Data & AI Authority	مايو 2023م	غير معروف	غير معروف	غير معروف	متاح بدعوات خاصة

الاسم	النوع	المطور	تاريخ الإطلاق	بيانات التدريب	لغة بيانات التدريب	عدد المُعاملات	الإتاحة
PaLM 2	نموذج	Google	مايو 2023م	نصوص أكواد برمجية	أكثر من 100 لغة	غير معروف	متاح في الولايات المتحدة عبر واجهة برمجة التطبيقات (API)
Noon	نموذج	Naseej نسيج الأنظمة العربية المتطورة Arabian Advanced Systems	يونيو 2023م	نصوص أكواد برمجية	العربية	7 مليار مُعامل	مفتوح المصدر
ERNIE 3.5	نموذج	Baidu 百度	يونيو 2023م	نصوص أكواد برمجية	غير معروف	غير معروف	غير معروف
Pangu Models 3.0	نموذج	HUAWEI	يوليو 2023م	نصوص صور أكواد برمجية	غير معروف	100 مليار مُعامل	غير معروف
Claude 2	نموذج	ANTHROP\C	يوليو 2023م	نصوص أكواد برمجية	الإنجليزية ولغات أخرى	غير معروف	متاح تجارياً للشركات عبر واجهة برمجة التطبيقات (API) متاح للعامة في الولايات المتحدة والمملكة المتحدة بنسخة بيتا
Llama 2	نموذج	Meta	يوليو 2023م	نصوص أكواد برمجية	الإنجليزية ولغات أخرى	70 مليار مُعامل	مفتوح المصدر
Qwen-7B	نموذج	Alibaba	أغسطس 2023م	نصوص أكواد برمجية	الصينية والإنجليزية ولغات أخرى	7 مليار مُعامل	مفتوح المصدر
Qwen-7B-Chat	تطبيق	Alibaba	أغسطس 2023م	نصوص أكواد برمجية	الصينية والإنجليزية ولغات أخرى	7 مليار مُعامل	مفتوح المصدر

SDAIA.GOV.SA SDAIA_SA SDAIA.SAUDI SDAIA-KSA