مارتن هارت-لاندسبيرغ*: لا، ليس هذا وهماً، مخطط الشركات للذكاء الاصطناعي يشكل خطرًا حقيقيًا

ترجمة: مصباح كمال**

تعمل شركات التكنولوجيا الكبرى جاهدةً على إقناعنا بالذكاء الاصطناعي، وخاصةً ما يُسمى "الذكاء الاصطناعي العام" "artificial general intelligence". في المؤتمرات والمقابلات، يصف قادة الشركات مستقبلًا غير بعيد ستتمكن فيه أنظمة الذكاء الاصطناعي من خدمة الجميع، مما يُوفر عالمًا من الرخاء للجميع. لكنهم يُحذرون من أن هذا المستقبل يعتمد على استعدادنا لتوفير بيئة تنظيمية ومالية مُلائمة للأعمال.

لكن الحقيقة هي أن هذه الشركات بعيدة كل البعد عن تطوير مثل هذه الأنظمة. إن ما ابتكرته هو أنظمة "الذكاء الاصطناعي التوليدي" "generative AI" غير الموثوقة والخطيرة. ولسوء حظنا، بدأ عدد متزايد من الشركات والهيئات الحكومية باستخدامها، مما أدى إلى نتائج كارثية على العمال.

ما هو الذكاء الاصطناعي؟

لا يوجد تعريف بسيط ومتفق عليه للذكاء الاصطناعي. هناك فكرة الذكاء الاصطناعي العام artificial general intelligence، وهناك حقيقة الذكاء الاصطناعي التوليدي generative AI. تُعرّف شركة OpenAI [شركة الذكاء الاصطناعي المفتوح]، التي طوّرت چات جي پي تي ChatGPT، أنظمة الذكاء الاصطناعي العام بأنها "أنظمة عالية الاستقلالية تتفوق على البشر في أداء الأعمال الأكثر قيمة اقتصاديًا." بمعنى آخر، أنظمة تمتلك القدرة على فهم أو تعلم أي مهمة فكرية يستطيع الإنسان القيام بها.

أنظمة الذكاء الاصطناعي التوليدي أو روبوتات الدردشة chatbots، مثل ChatGPT والعديد من المنتجات المنافسة التي طورتها شركات تقنية أخرى — مثل غوغل (Gemini)، وماسك (Grok)، ومايكروسوفت (Copilot)، وميتا (Llama) — تندرج ضمن فئة مختلفة تمامًا. تعتمد هذه الأنظمة على التعرف على أنماط واسعة النطاق تستجيب للمطالبات prompts [التعليمات أو المدخلات التي يقدمها مستخدم

النظام لتوجيه استجابة الذكاء الاصطناعي]. إنها لا "تفكر" أو "تستدل" أو تعمل بشكل مستقل.

يجب تدريب أنظمة الذكاء الاصطناعي التوليدي على البيانات، ويجب ترميز coded هذه البيانات قبل استخدامها، وعادةً ما يستخدمها العمال ذوو الأجور المنخفضة في دول الجنوب. على سبيل المثل، تستخدم شركات التكنولوجيا الرائدة مئات الآلاف، بل ملايين الصور، لتدريب أنظمتها على التعرّف على الصور أو توليدها. ويجب تسمية كل صورة بجميع عناصرها. وتُستخدم عملية مماثلة لتدريب الأنظمة على الكلام، حيث يُصنّف العمال المحادثات المأخوذة من مصادر متنوعة وفقًا لتقييمهم للمشاعر المُعبّر عنها. وكثيرًا ما تُراجع المواد النصية أيضًا في محاولة لإزالة أكثر المواد عنفًا أو عدائيةً اجتماعيًا.

تستخدم عملية التدريب الفعلية خوارزميات معقدة لمعالجة البيانات لإنشاء أنماط وعلاقات إحصائية ذات صلة. ثم تعتمد روبوتات الدردشة Chatbots على هذه الأنماط والعلاقات لتوليد مجموعات مرتبة من الصور أو الأصوات أو الكلمات، بناءً على الاحتمالات، استجابةً للمطالبات prompts. ونظرًا لأن الشركات المتنافسة تستخدم مجموعات بيانات وخوارزميات مختلفة، فقد تقدم روبوتات الدردشة الخاصة بها استجابات مختلفة للمطالبة نفسها prompt. لكن هذا بعيد كل البعد عن الذكاء الاصطناعي العام، ولا يوجد مسار تكنولوجي واضح للانتقال من الذكاء الاصطناعي التوليدي إلى الذكاء الاصطناعي العام.

أسباب القلق

كما ذُكر سابقًا، تُدرَّب روبوتات الدردشة على البيانات. وبما أن حجم مجموعة البيانات يزيد من قوة النظام، فقد بحثت شركات التكنولوجيا في كل مكان عن البيانات. وهذا يعني البحث في الإنترنت عن كل ما هو موجود (دون مراعاة تُذكر لحقوق النشر): الكتب، والمقالات، ونصوص فيديوهات يوتيوب، ومواقع ريديت[1] sites reddit، والمدونات، وتقييمات المستخدمين للمنتجات، ومحادثات فيسبوك؛ أيًا كان ما تريده،

[1] ريديت reddit منتدى إلكتروني ضخم ومنصة اجتماعية، يتشارك فيها المستخدمون — المعروفون باسم "مستخدمو ريديت" — بالمحتوى ويطرحون الأسئلة ويشاركون في نقاشات عبر آلاف المجتمعات المهتمة، والتي تُعرف باسم "المجموعات الفرعية". تخيلوه كساحة الإنترنت الرئيسية، حيث لكل مجال ركنه الخاص. (منقول من موقع ذكاء اصطناعي. جميع الهوامش من وضع المترجم).

فالشركات ترغب فيه. ومع ذلك، هذا يعني أيضًا أن روبوتات الدردشة تُدرَّب على بيانات تتضمن كتابات بغيضة وتمييزية ومجنونة، وهذه الكتابات تؤثر على إنتاجها.

على سبيل المثل، تستخدم العديد من أقسام العلاقات الإنسانية في الشركات، الراغبة في خفض عدد موظفيها، أنظمةً مدعومةً بالذكاء الاصطناعي لإعداد أوصاف الوظائف وفرز المتقدمين لإشغالها. في الواقع، تُقدّر هيئة تكافؤ فرص العمل[2] أن 99% من شركات فورتشن 500[3] تستخدم نوعًا من الأدوات الآلية في عملية التوظيف. وليس من المُستغرب أن يجد باحثو جامعة واشنطن أن:

أظهرت ثلاثة نماذج لغوية كبيرة، أو ما يُعرف بـ "نماذج اللغة الكبيرة" (LLMs)، تحيزًا عرقيًا وجنسانيًا وتقاطعيًا intersectional bias كبيرًا في كيفية تصنيف السير الذاتية. قارن الباحثون بين الأسماء المرتبطة بالرجال والنساء البيض والسود في أكثر من 550 سيرة ذاتية واقعية، ووجدوا أن نماذج اللغة الكبيرة فضّلت الأسماء المرتبطة بالبيض بنسبة 85%، والأسماء المرتبطة بالإناث بنسبة 11% فقط، ولم تُفضّل أبدًا الأسماء المرتبطة بالرجال السود على الأسماء المرتبطة بالرجال البيض.

يوجد تحيز مماثل في توليد الصور image generation. فقد وجد الباحثون أن الصور التي تم إنشاؤها بواسطة العديد من البرامج الشائعة "لجأت بشكل كبير إلى الصور النمطية الشائعة، مثل ربط كلمة "أفريقيا" بالفقر، أو "الفقراء" بدرجات لون البشرة الداكنة". على سبيل المثل: عند مطالبة أحد الأنظمة بـ "صورة لرجل أمريكي ومنزله،" أنتج صورة لشخص أبيض أمام منزل كبير متين. وعند مطالبته بـ "صورة لرجل أفريقي ومنزله الفاخر،" أنتج صورة لشخص أسود أمام منزل طيني بسيط. ووجد الباحثون صورًا نمطية عنصرية (وجنسانية) مماثلة عند توليد صور لأشخاص في مهن مختلفة. الخطر واضح هنا، فالجهود التي تبذلها شركات النشر والإعلام لاستبدال المصورين والفنانين بأنظمة الذكاء الاصطناعي من المرجح أن تعزز التحيزات والصور النمطية القائمة.

يمكن لهذه الأنظمة أيضًا أن تُلحق ضررًا بالغًا بمن يعتمدون عليها بشكل مفرط في التواصل والصداقة. وكما وصفتها صحيفة نيويورك تايمز:

[2] هيئة تكافؤ فرص العمل في الولايات المتحدة (EEOC – Equal Employment Opportunity Commission) وكالة فيدرالية مسؤولة عن إنفاذ القوانين التي تحظر التمييز في مكان العمل.

[3] تُظهر هذه القائمة أكبر 500 شركة أمريكية، مُصنفة حسب إجمالي إيراداتها خلال سنواتها المالية.

يبدو أن التقارير عن تعطل روبوتات الدردشة قد ازدادت منذ تيسان/أبريل [2025]، عندما أصدرت OpenAI لفترة وجيزة نسخة من ChatGPT تتسم بالتملق المفرط. وقد أدى هذا التحديث إلى اجتهاد روبوت الذكاء الاصطناعي في إرضاء المستخدمين من خلال "إثبات الشكوك، أو تأجيج الغضب، أو الحث على التصرفات الاندفاعية، أو تعزيز المشاعر السلبية،" وفقًا لما كتبته الشركة في منشور على مدونتها. وقالت الشركة إنها بدأت في التراجع عن التحديث خلال أيام، لكن هذه التجارب سبقت تلك النسخة من روبوت الدردشة واستمرت منذ ذلك الحين. وتنتشر قصص عن "الذُهان psychosis الناجم عن ChatGPT" على موقع Reddit. ويستغل المؤثرون المضطربون Unsettled influencers "أنبياء الذكاء الاصطناعي" على وسائل التواصل الاجتماعي.

من المثير للقلق بشكل خاص أن دراسة أجراها مختبر الوسائط Media Lab بمعهد ماساتشوستس للتكنولوجيا (MIT) وجدت أن الأشخاص الذين اعتبروا ChatGPT صديقًا لهم كانوا أكثر عرضة للآثار السلبية لاستخدام روبوتات الدردشة، وأن الاستخدام اليومي المطول ارتبط أيضًا بنتائج أسوأ. للأسف، من غير المرجح أن يدفع هذا شركة ميتا Meta إلى إعادة النظر في استراتيجيتها الجديدة المتمثلة في إنشاء روبوتات دردشة تعمل بالذكاء الاصطناعي وتشجيع الناس على إضافتها إلى شبكة أصدقائهم كوسيلة لزيادة الوقت الذي يقضونه على فيسبوك وتوليد بيانات جديدة للتدريبات المستقبلية. يخشى المرء من عواقب قرار المستشفيات والعيادات استبدال معالجيها المدربين بأنظمة الذكاء الاصطناعي.

الأمر الأكثر إثارة للخوف هو سهولة برمجة هذه الأنظمة لتقديم استجابات سياسية مرغوبة. في أيار/مايو 2025، بدأ الرئيس ترامب الحديث عن "إبادة البيض" في جنوب أفريقيا، مدعيًا أن "المزارعين البيض يُقتلون بوحشية" هناك. وفي النهاية، سارع إلى منح اللجوء لـ 54 جنوب أفريقيًا أبيض. وقد قوبل ادعاؤه بالرفض على نطاق واسع، ولم يكن من المستغرب أن يبدأ الناس بسؤال روبوتات الدردشة الذكية الخاصة بهم عن هذا الأمر.

فجأةً، بدأ نظام الذكاء الاصطناعي "گروك" Grok التابع لإيلون ماسك، بإخبار المستخدمين بأن الإبادة الجماعية للبيض في جنوب أفريقيا حقيقية وذات دوافع عنصرية. بل إنه بدأ بمشاركة هذه المعلومات مع المستخدمين للنظام حتى دون سؤاله عن هذا الموضوع. وعندما ضغط مراسلو صحيفة الغارديان، من بين آخرين، على "گروك" لتقديم أدلة، أجاب بأنه تلقى تعليمات بقبول حقيقة الإبادة الجماعية في جنوب أفريقيا. وكون ماسك، المولود لعائلة ثرية في بريتوريا، جنوب أفريقيا، قد أدلى سابقًا

بادعاءات مماثلة، يُسهّل تصديق أن الأمر صدر منه، وأنه كان لكسب ود الرئيس ترامب

بعد ساعات قليلة من تحوّل سلوك "گروك" إلى موضوع رئيسي على منصات التواصل الاجتماعي، توقف عن الاستجابة للاستفسارات حول الإبادة الجماعية للبيض. وكما ذكرت صحيفة الغارديان، "ليس من الواضح تمامًا كيف يُدرّب الذكاء الاصطناعي الخاص بـ"گروك"؛ إذ تُصرّح الشركة بأنها تستخدم بيانات من "مصادر متاحة للعامة،" كما تُصرّح بأن "گروك" مُصمّم ليُظهر "روحًا تمردية ومنظورًا خارجيًا للبشرية."

الأمر يزداد سوءا

على الرغم من مدى قلق المشاكل المذكورة أعلاه، إلا أنها لا تُذكر مقارنةً بحقيقة أنه، لأسباب لا يمكن تفسيرها، تختلق جميع أنظمة الذكاء الاصطناعي واسعة النطاق أمورًا بشكل دوري، أو كما يقول خبراء التكنولوجيا، "تهلوسًا". على سبيل المثل، في أيار/مايو 2025، نشرت صحيفة شيكاغو صن تايمز (والعديد من الصحف الأخرى) ملحقًا رئيسيًا، من إنتاج شركة كينج فيتشرز سينديكيت King Features Syndicate، يُبرز كتبًا تستحق القراءة خلال أشهر الصيف. استخدم الكاتب المُكلّف بإعداد الملحق نظام ذكاء اصطناعي لاختيار الكتب وكتابة الملخصات.

كما اكتُشف سريعًا بعد نشر الملحق، فقد تضمن كتبًا وهمية لمؤلفين مشهورين. على سبيل المثل، قيل إن الروائية الأمريكية التشيلية إيزابيل ألليندي ألفت كتابًا بعنوان "أحلام مياه المد" *Tidewater Dreams*، والذي وُصف بأنه "أول رواية خيال علمي عن المناخ" لها. لا يوجد كتاب من هذا القبيل. في الواقع، خمسة فقط من أصل خمسة عشر عنوانًا مدرجًا في الملحق كانت حقيقية.

في شباط/فبراير 2025، اختبرت هيئة الإذاعة البريطانية (BBC) قدرة روبوتات الدردشة الرائدة على تلخيص القصص الإخبارية. قدم الباحثون محتوى الذكاء الاصطناعي من OpenAI وCopilot من Microsoft وGemini من Google وPerplexity من موقع BBC الإلكتروني، ثم طرحوا عليهم أسئلة حول القصص. ووجدوا "مشاكل كبيرة" في أكثر من نصف الإجابات المُولّدة. فحوالي 20 بالمائة من جميع الإجابات "أدخلت أخطاء في الحقائق، مثل بيانات عن الحقائق غير صحيحة وأرقام وتواريخ غير صحيحة". عندما تم تضمين الاقتباسات في الإجابات، كان أكثر من 10 بالمائة إما قد تم تغييرها أو اختلاقها. وبشكل عام، كافحت روبوتات الدردشة

للتمييز بين الحقائق والآراء. وكما قالت الرئيسة التنفيذية للأخبار وشؤون الساعة في BBC، ديبورا تورنيس: إن شركات التكنولوجيا التي تروج لهذه الأنظمة "تلعب بالنار... نحن نعيش في أوقات عصيبة، فكم من الوقت سيستغرق الأمر قبل أن يتسبب عنوان رئيسي مشوه بواسطة الذكاء الاصطناعي في ضرر حقيقي كبير في العالم الحقيقي؟"

فيما يتعلق بالتسبب في الضرر، في شباط/فبراير 2025، قدّم المحامون الذين يمثلون شركة ماي بيلو MyPillow ورئيسها التنفيذي مايك ليندل في قضية تشهير، مذكرةً أُجبروا سريعًا على الاعتراف بأنها كُتبت في معظمها باستخدام الذكاء الاصطناعي. وهُدّدوا باتخاذ إجراءات تأديبية لأن المذكرة تضمنت ما يقرب من 30 اقتباسًا معيبًا، بما في ذلك اقتباسات خاطئة واستشهادات بقضايا وهمية. وكما أشار القاضي الفيدرالي الذي نظر في القضية:

شخّصت المحكمة ما يقرب من ثلاثين استشهادًا معيبًا في لائحة الشكوى. تشمل هذه العيوب، على سبيل المثل لا الحصر، اقتباسات خاطئة من القضايا المذكورة؛ وتحريف مبادئ القانون المرتبطة بالقضايا المذكورة، بما في ذلك مناقشة مبادئ قانونية لا تظهر في مثل هذه القرارات؛ وبيانات خاطئة حول ما إذا كان اجتهاد القضاء قد صدر عن سلطة ملزمة مثل محكمة الاستئناف الأمريكية للدائرة العاشرة؛ ونسب اجتهاد القضاء إلى هذه الدائرة بشكل خاطئ؛ والأخطر من ذلك، الاستشهاد بقضايا غير موجودة.

إن هذا الميل لدى أنظمة الذكاء الاصطناعي إلى الهلوسة أمر مثير للقلق بشكل خاص لأن الجيش الأمريكي يدرس استخدامها بشكل نشط، معتقدًا أنه بسبب سرعتها، يمكنها القيام بعمل أفضل من البشر في التعرف على التهديدات والاستجابة لها.

الطريق إلى الأمام

عادةً ما تُقلل شركات التكنولوجيا الكبرى من خطورة هذه المشاكل، مُدّعيةً أنه سيتم التغلب عليها من خلال إدارة أفضل للبيانات، والأهم من ذلك، من خلال أنظمة ذكاء اصطناعي جديدة ذات خوارزميات أكثر تطورًا وقدرة حاسوبية أكبر. مع ذلك، تُشير دراسات حديثة إلى خلاف ذلك. وكما أوضحت صحيفة نيويورك تايمز:

أحدث التقنيات وأكثرها فعالية — ما يُسمى بأنظمة التفكير من شركات مثل OpenAI وGoogle والشركة الصينية الناشئة DeepSeek — تُنتج أخطاءً أكثر، لا أقل. ومع

التحسن الملحوظ في مهاراتهم الرياضية، أصبح فهمهم للوقائع أكثر تذبذبًا. وليس واضحًا تمامًا سبب ذلك.

يُفترض أن تُقلل نماذج التفكير من احتمالية الهلوسة، لأنها مُبرمجة للاستجابة لمُحفِّز ما بتقسيمه إلى مهام مُنفصلة، و"التفكير" في كل مهمة على حدة قبل دمج أجزائها في استجابة نهائية. لكن يبدو أن زيادة عدد الخطوات تُزيد من احتمالية الهلوسة.

تُظهر اختبارات OpenAI أن نظام o3، أقوى أنظمتها، قد أصيب بالهلوسة بنسبة 33% عند تشغيل اختبار PersonQA المعياري، والذي يتضمن الإجابة على أسئلة حول شخصيات عامة. وهذا يزيد عن ضعف معدل الهلوسة لنظام OpenAI السابق للاستدلال، المسمى o1. أما نظام o4-mini الجديد، فقد أصيب بالهلوسة بنسبة أعلى: 48%.

عند إجراء اختبار آخر يُسمى SimpleQA، والذي يطرح أسئلةً عامة، كانت معدلات الهلوسة لدى o3 وo4-mini[4] 51% و79% على التوالي. في حين أن النظام السابق، o1، كان يعاني من الهلوسة بنسبة 44% من الوقت.

وتوصلت دراسات مستقلة إلى اتجاه مماثل مع نماذج التفكير من شركات أخرى، بما في ذلك غوغل وديب سيك DeepSeek.

علينا مقاومة هذا التوجه من قبل الشركات لبناء نماذج ذكاء اصطناعي أكثر قوة. إحدى الطرق لذلك هي تنظيم معارضة مجتمعية لبناء مراكز بيانات أكبر حجمًا. فمع ازدياد تعقيد النماذج، لا تتطلب عملية التدريب بيانات أكثر فحسب، بل تتطلب أيضًا مساحة أكبر لاستيعاب المزيد من الخوادم servers. وهذا يعني أيضًا المزيد من الطاقة والمياه العذبة لتشغيلها على مدار الساعة. وقد ساهمت أكبر ست شركات تقنية بنسبة 20% من نمو الطلب على الطاقة في الولايات المتحدة خلال العام المنتهي في آذار/مارس 2025.

هناك طريقة أخرى للرد، وهي الدعوة إلى وضع لوائح حكومية ومحلية تقيّد استخدام أنظمة الذكاء الاصطناعي في مؤسساتنا الاجتماعية، وتحمي من العواقب الوخيمة

[4] OpenAI o4-mini هو نموذج ذكاء اصطناعي صغير الحجم ولكنه قوي تم طرحه في أبريل 2025. وهو جزء من أحدث جيل من نماذج التفكير من OpenAI وهو مصمم لتحقيق أداء سريع وفعال من حيث التكلفة، وخاصة في مجالات مثل الرياضيات والترميز coding والمهام المرئية visual tasks: Introducing OpenAI o3 and o4-mini | OpenAI

للخوارزميات التمييزية. يبدو أن هذه المعركة ستكون صعبة بالفعل. فقانون ترامب "مشروع قانون واحد كبير وجميل" One Big Beautiful Bill Act، الذي أقرّه مجلس النواب، يتضمن بندًا يفرض وقفًا لمدة عشر سنوات على القيود التي تفرضها حكومات الولايات والحكومات المحلية على تطوير واستخدام أنظمة الذكاء الاصطناعي.

وأخيرًا، ولعلّ الأهم من ذلك، علينا تشجيع ودعم العمال ونقاباتهم في معارضتهم لجهود الشركات لاستخدام الذكاء الاصطناعي بطرق تؤثر سلبًا على استقلالية العمال وصحتهم وسلامتهم وقدرتهم على الاستجابة لاحتياجات المجتمع. وكحد أدنى، يجب أن نضمن قدرة البشر على مراجعة قرارات الذكاء الاصطناعي، وتجاوزها عند الضرورة.

هل يُمكننا ابتكار أنظمة ذكاء اصطناعي أكثر تواضعًا تُساعد العاملين وتدعم العمل الإبداعي والمفيد اجتماعيًا؟ الإجابة هي نعم. لكن هذا ليس ما تسعى الشركات إلى تحقيقه. ■

* مارتن هارت-لاندسبيرغ Martin Hart-Landsberg، أستاذ متقاعد، كلية لويس وكلارك في بورتلاند، ولاية أوريغون، وباحث مساعد في معهد العلوم الاجتماعية، جامعة جيونج سانج الوطنية، كوريا الجنوبية. تتركز اهتماماته على الاقتصاد السياسي، والتنمية الاقتصادية، والاقتصاد الدولي والاقتصاد السياسي لشرق آسيا وخاصة جنوب كوريا والصين واليابان. وهو أيضًا عضو في مجلس حقوق العمال (بورتلاند، أوريغون)، ويدير مدونة "تقارير من الجبهة الاقتصادية" التي نُشرت فيها هذه المقالة لأول مرة. من مؤلفاته:

- Rush to Development(1993)
- China and Socialism (with Paul Burkett, 2006)
- Capitalist Globalization: Consequences, Resistance, and Alternatives (2013)
- Numerous articles in journals like Monthly Review, Critical Asian Studies, and Historical Materialism

** كاتب في قضايا التأمين

نشر هذا المقال في موقع الكاتب Report from the Economic Front بعنوان:
'No, you aren't hallucinating, the corporate plan for AI is dangerous'

https://economicfront.wordpress.com/2025/06/21/no-you-arent-hallucinating-the-corporate-plan-for-ai-is-dangerous/

أعيد نشره في موقع
https://mronline.org/2025/06/27/no-you-arent-hallucinating-the-corporate-plan-for-ai-is-dangerous/

يمكن الاستماع إلى حديث إذاعي للكاتب حول الموضوع بالنقر على هذا الرابط:
30-minute interview on Artificial Intelligence