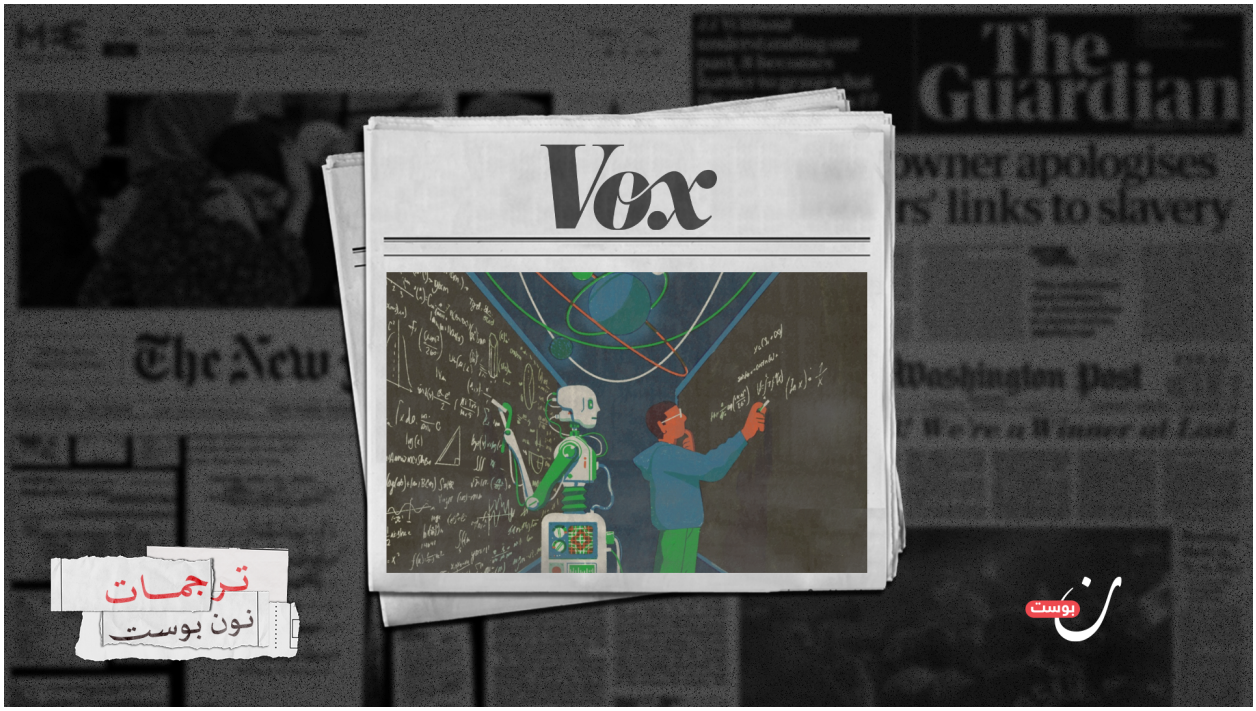
هل يدرك الذكاء الاصطناعي ما يقوله أم أنه مجرد مقلد بارع؟

كتبه سيغال صامويل | 1 أبريل 2025,



ترجمة وتحرير: نون بوست

يتحرك عالم الذكاء الاصطناعي بسرعة هائلة لدرجة أنه من السهل أن يضيع المرء وسط سيل المنتجات الجديدة اللامعة. تعلن "أوبن إيه آي" عن أحد هذه المنتجات، ثم تطلق شركة "ديب سيك" الصينية الناشئة آخر، ثم تعود "أوبن إيه آي" سريعًا بمنتج جديد. وكل واحد منها مهم، لكن التركيز المفرط على أي منها قد يجعلك تفوت القصة الكبرى الحقيقية للأشهر الستة الماضية.

القصة الكبرى هي أن شركات الذكاء الاصطناعي تزعم الآن أن نماذجها قادرة على الاستدلال الحقيقي؛ أي ذلك النوع من التفكير الذي نمارسه أنت وأنا عندما نحاول حل مشكلة.

والسؤال الكبير هو: هل هذا صحيح؟

إن الرهان كبير، لأن الإجابة على هذا السؤال ستحدد كيف ينبغي – أو لا ينبغي – للجميع، من والدتك إلى حكومتك، اللجوء إلى الذكاء الاصطناعي طلبًا للمساعدة.

وإذا سبق لك استخدام "شات جي بي تي"، فأنت تعلم أنه مصمم لتقديم إجابات سريعة على أسئلتك. لكن "النماذج المنطقية" الحديثة – مثل أو1 من "أوبن إيه آي" أو آر1 من "ديب سيك"

– مصممة "للتفكير" لبعض الوقت قبل الرد، وذلك عبر تفكيك المشكلات الكبيرة إلى أجزاء أصغر ومحاولة حلها خطوة بخطوة. ويطلق القطاع على هذه العملية اسم "التفكير المتسلسل".

وتُحقق هذه النماذج نتائج مذهلة للغاية، فهي قادرة على حل ألغاز منطقية معقدة، والتفوق في اختبارات الرياضيات، وكتابة أكواد برمجية مثالية من المحاولة الأولى. ومع ذلك، فهي تفشل بشكل كارثي في حل مشكلات بسيطة جدًا.

وينقسم خبراء الذكاء الاصطناعي حول كيفية تفسير ذلك. فالمتشككون يعتبرون ذلك دليلًا على أن النماذج "المنطقية" لا تمارس الاستدلال الحقيقي على الإطلاق. وفي المقابل، يؤكد المؤيدون أن هذه النماذج تمارس نوعًا من الاستدلال، وحتى إن لم يكن بمرونة تفكير الإنسان حاليًا، فإنه في طريقه للوصول إلى ذلك.

إذن، من على حق؟ إن أفضل إجابة ستكون مقلقة لكل من المشككين المتشددين في الذكاء الاصطناعي والمؤمنين به تمامًا.

ما الذي يُعتبر منطقيًا؟

لنعد خطوة إلى الوراء.. ما هو المنطق بالضبط؟

تستخدم شركات الذكاء الاصطناعي، مثل "أوبن إيه آي"، مصطلح "التفكير المنطقي" للإشارة إلى قدرة نماذجها على تفكيك المشكلة إلى أجزاء أصغر، ثم معالجتها خطوة بخطوة، مما يؤدي في النهاية إلى حل أكثر دقة.

ولكن هذه تعريف ضيق جدًا لمصطلح التفكير المنطقي مقارنة بما قد يتبادر إلى أذهان الكثيرين. فعلى الرغم من أن العلماء لا يزالون يحاولون فهم كيفية عمل الاستدلال في الدماغ البشري – فضلًا عن فهمه في الذكاء الاصطناعي – إلا أنهم متفقون على وجود أنواع مختلفة من الاستدلال.

وهناك الاستدلال الاستنباطي، حيث تبدأ ببيان عام وتستخدمه للوصول إلى استنتاج محدد. وهناك الاستدلال الاستقرائي، حيث تعتمد على ملاحظات محددة لصياغة تعميم أوسع. بالإضافة إلى ذلك، يوجد الاستدلال القياسي، والاستدلال السببي، والاستدلال المنطقي... وباختصار، الاستدلال ليس شيئًا واحدًا!

والآن، إذا قدم لك شخص ما مسألة رياضية صعبة ومنحك الفرصة لتحليلها والتفكير فيها خطوة بخطوة، فستحقق نتائج أفضل بكثير مقارنة بمحاولة إعطاء الإجابة فورًا دون تفكير. لذا، فإن القدرة على القيام بـ"الاستدلال المتسلسل" التأملي مفيدة بلا شك، وقد تكون عنصرًا ضروريًا لإنجاز أي مهمة معقدة. ومع ذلك، فهي ليست كل ما يشكل الاستدلال.

إن إحدى السمات المهمة للاستدلال التي نهتم بها كثيرًا في العالم الحقيقي هي القدرة على "استنباط

قاعدة أو نمط من بيانات أو خبرة محدودة وتطبيق هذه القاعدة أو النمط على مواقف جديدة وغير مألوفة"، كما تكتب ميلاني ميتشل، الأستاذة في معهد سانتا في، مع زملائها في ورقة بحثية حول قدرات الذكاء الاصطناعي في الاستدلال. "حتى الأطفال الصغار جدًا بارعون في تعلّم القواعد المجردة من خلال أمثلة قليلة فقط."

بعبارة أخرى، يمكن للطفل الصغير أن يعمم. لكن هل يمكن للذكاء الاصطناعي ذلك؟

يدور الكثير من النقاش حول هذا السؤال. المشككون في قدرة الذكاء الاصطناعي على التعميم متشككون للغاية؛ فهم يعتقدون أن هناك شيء آخر يحدث.

حجة المشككين

قالت لي شانون فالور، فيلسوفة التكنولوجيا في جامعة إدنبرة، عندما أطلقت "أوبن إيه آي" نموذج أو1 في سبتمبر/ أيلول: "إنه نوع من المحاكاة الفوقية".

وكانت تقصد أنه بينما يحاكي نموذج أقدم مثل "شات جي بي تي" العبارات التي كتبها البشر في بيانات تدريبه، فإن نموذجًا أحدث مثل أو1 يحاكي العملية التي ينخرط فيها البشر للتوصل إلى تلك العبارات.

بعبارة أخرى، تعتقد أنه لا يقوم بالاستدلال الحقيقي. وسيكون من السهل نسبيًا على أو1 أن يبدو وكأنه يقوم بالاستدلال، فبيانات تدريبه مليئة بأمثلة على ذلك، بدءًا من الأطباء الذين يحللون الأعراض لتشخيص المرض، وصولًا إلى القضاة الذين يقيّمون الأدلة للتوصل إلى حكم.

علاوة على ذلك، عندما قامت شركة "أوبن إيه آي" ببناء نموذج أو1، أجرت بعض التعديلات على نموذج "شات جي بي تي" السابق لكنها لم تعيد تصميم المعمارية بشكل جذري – وكان "شات جي بي تي" قد ارتكب أخطاء في إجابة أسئلة سهلة السنة الماضية، مثل تقديم إجابة سخيفة تمامًا عن كيفية نقل رجل وماعز عبر نهر.

لذا، تساءلت فالور: لماذا نفترض أن أو1 يقوم بشيء جديد وسحري تمامًا – خاصةً أنه هو الآخر يخطئ في الإجابة على أسئلة بسيطة؟ وأضافت: "في الحالات التي يفشل فيها، أرى أدلة مقنعة بالنسبة لي على أنه لا يقوم بالاستدلال على الإطلاق."

وشعرت ميتشل بالدهشة من أداء أو3 – أحدث نماذج الاستدلال من "أوبن إيه آي"، الذي أُعلن عنه في نهاية السنة الماضية كخليفة لـ أو1 – في الاختبارات. لكنها تفاجأت أيضًا بكمية الحسابات الهائلة التي استخدمها لحل المشكلات. ولا نعرف بالضبط ما الذي يفعله بكل هذه العمليات الحسابية، لأن "أوبن إيه آي" لا تُظهر شفافية حول ما يجري داخل النموذج.

وقالت لي: "لقد أجريتُ تجارب خاصة بي على أشخاص وهم يفكرون بصوت عالٍ حول هذه

المشكلات، ولم يستغرقوا ساعات من وقت الحساب، كما تعلم. إنهم فقط يقولون بضع جمل ثم يضيفون: "نعم، فهمت كيف يعمل الأمر"، لأنهم يستخدمون أنواعًا معينة من المفاهيم. ولا أعرف ما إذا كان أو3 يستخدم تلك الأنواع من المفاهيم."

وقالت ميتشل إنه بدون قدر أكبر من الشفافية من الشركة، لا يمكننا التأكد من أن النموذج يقوم بتفكيك المشكلة الكبيرة إلى خطوات ويحصل على إجابة أفضل نتيجة لهذا النهج، كما تزعم شركة "أوبن إيه آي".

وأشارت إلى ورقة بحثية بعنوان "دعونا نفكر نقطة بنقطة"، حيث لم يُطلب من النموذج تفكيك المشكلة إلى خطوات وسيطة، بل طُلب منه ببساطة توليد نقاط. كانت تلك النقاط بلا معنى تمامًا – وهو ما أطلق عليه مؤلفو الورقة "رموز حشو" – لكن تبيَّن أن مجرد وجود رموز إضافية أتاح للنموذج قدرة حسابية أكبر، مما مكّنه من حل المشكلات بشكل أفضل.

يشير ذلك إلى أنه عندما يُنتج النموذج خطوات وسيطة – سواء كانت عبارة مثل "دعونا نفكر في ذلك خطوة بخطوة" أو مجرد "...."؛ فإن تلك الخطوات لا تعني بالضرورة أنه يقوم بالاستدلال بطريقة مشابهة للبشر كما قد يُظن.

وقالت ميتشل: "أعتقد أن الكثير مما يفعله النموذج أقرب إلى كونه مجموعة من القواعد التقريبية أكثر من كونه نموذجًا للاستدلال." والقاعدة التقريبية هي اختصار ذهني، وهي وسيلة تتيح لك غالبًا تخمين الإجابة الصحيحة على مشكلة ما، ولكن دون التفكير فيها بشكل فعلي.

إليك مثالًا كلاسيكيًا: درّب الباحثون نموذج رؤية ذكاء اصطناعي لتحليل الصور بحثًا عن سرطان الجلد. ففي البداية، بدا أن النموذج يحدد بالفعل ما إذا كانت الشامة خبيثة، لكن اتضح أن الصور في بيانات تدريبه التي تحتوي على شامات خبيثة غالبًا ما كانت تتضمن مسطرة، لذا فقد تعلم النموذج للتو استخدام وجود المسطرة كمُستدل لتحديد ما إذا كانت الشامة خبيثة.

ويعتقد الباحثون المشككون في الذكاء الاصطناعي أن النماذج المتطورة قد تقوم بشيء مشابه: فهي تبدو وكأنها "تستدل" لحل مسألة رياضية، لكنها في الواقع تعتمد فقط على مزيج من المعلومات المحفوظة والاختصارات الذهنية (القواعد التقريبية).

حجة المؤيدين

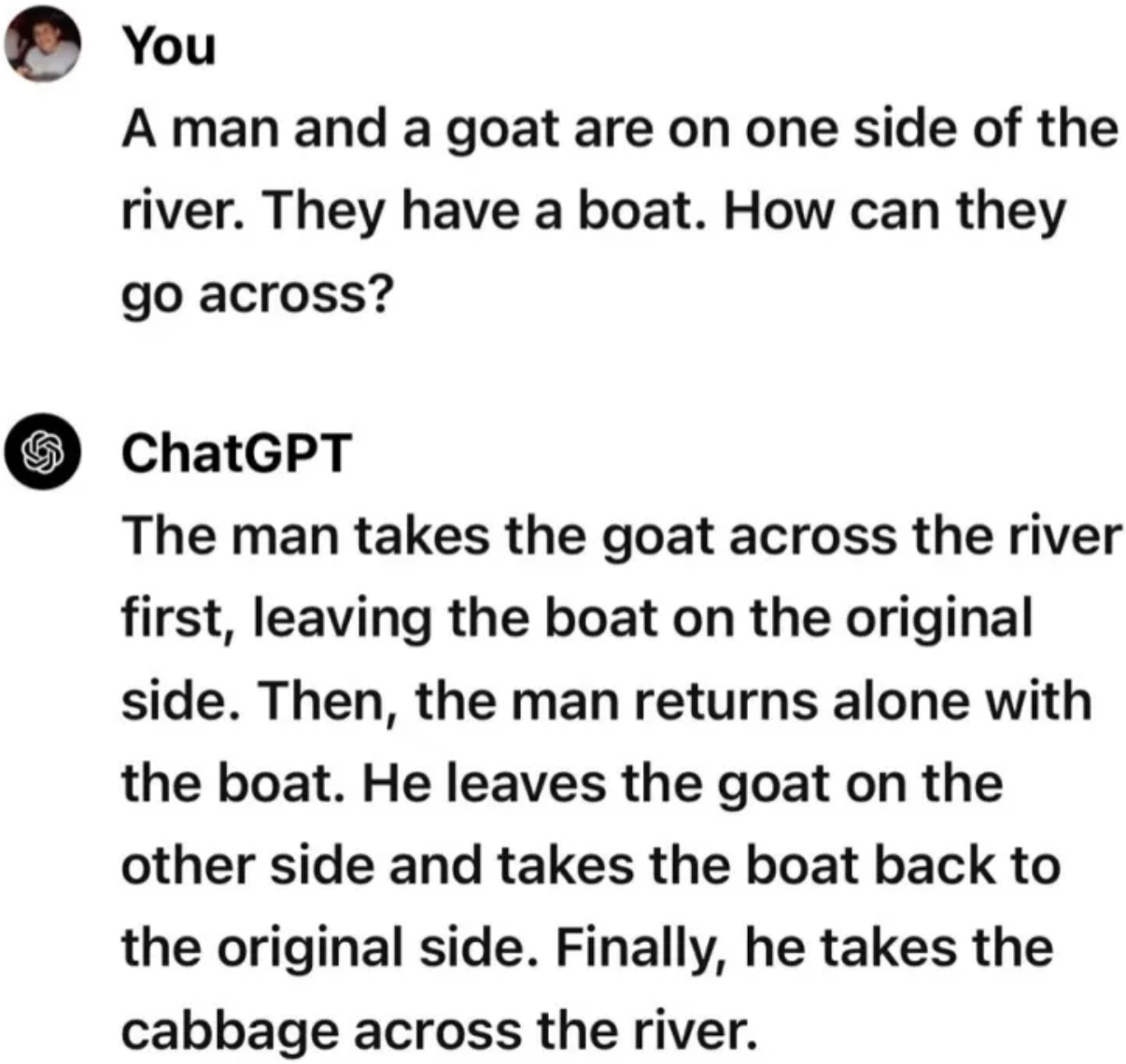
هناك خبراء آخرون أكثر تفاؤلًا بشأن نماذج الاستدلال؛ حيث يعتقد رايان غرينبلات، كبير العلماء في معهد "ريدوود ريسيرش"، وهو منظمة غير ربحية تهدف إلى الحد من مخاطر الذكاء الاصطناعي المتقدم، أن هذه النماذج تقوم بوضوح بنوع من الاستدلال.

وقال غرينبلات: "إنهم يقومون بذلك بطريقة لا تعمم بنفس كفاءة الطريقة التي يفعلها البشر، فهم

يعتمدون على الحفظ والمعرفة أكثر مما يفعل البشر- لكنهم لا يزالون يقومون بالأمر نفسه. لا يبدو الأمر وكأنه لا يوجد تعميم على الإطلاق."

ففي نهاية المطاف، تمكنت هذه النماذج من حل مشكلات صعبة تتجاوز الأمثلة التي تدربت عليها، وغالبًا بطريقة مبهرة. وبالنسبة لغرينبلات، فإن أبسط تفسير لذلك هو أنها بالفعل تقوم ببعض عمليات الاستدلال.

ويمكن لنقطة الاختصارات الذهنية أن تُستخدم في الاتجاهين، سواء كنا نتحدث عن نموذج استدلالي أو نموذج أقدم مثل "شات جي بي تي". خذ على سبيل المثال مسألة "رجل، وقارب، وعنزة" التي أثارت سخرية العديد من المشككين تجاه "أوبن إيه آي" السنة الماضية:



فما الذي يحدث هنا؟ يقول غرينبلات إن النموذج أخطأ لأن هذا السؤال في الواقع هو لغز منطقي كلاسيكي يعود إلى قرون مضت، وقد يكون قد ظهر مرات عديدة في بيانات التدريب. وفي بعض صيغ

لغز عبور النهر، يجب على مزارع لديه ذئب وعنزة وملفوف أن يعبر بالقارب، ولكن القارب لا يمكنه حمل سوى المزارع وغرض واحد في كل مرة.

ويكمن التحدي في نقل الجميع دون أن يأكل أحدهم الآخر، فالذئب قد يأكل العنزة، والعنزة قد تأكل الملفوف. وهذا ما يفسر ذكر النموذج للملفوف في إجابته، إذ إنه "تعرف" فورًا على اللغز.

ويقول جرينبلات: "تخميني الأفضل هو أن لدى النماذج دافعًا قويًا للغاية للقول: 'آه، إنه هذا اللغز! أعرف هذا اللغز! يجب أن أجيب بهذه الطريقة لأنها أدت بشكل جيد في بيانات التدريب.' إنه أشبه بحيلة تعلمها النموذج." ما الذي يعنيه ذلك؟ "ليس الأمر أنها لا تستطيع حله. ففي كثير من الحالات، إذا أخبرتها أن السؤال خادع ثم قدمته لها، فإنها غالبًا ما تجيب بشكل صحيح تمامًا."

وأشار غرينبلات إلى أن البشر يفشلون بنفس الطريقة طوال الوقت. فمثلًا، إذا قضيت شهرًا في دراسة نظرية الألوان – من الألوان المتكاملة إلى التأثيرات النفسية للألوان المختلفة وحتى الأهمية التاريخية لبعض الأصباغ في لوحات عصر النهضة – ثم واجهت سؤالًا في اختبار يقول: "لماذا رسم الفنان السماء باللون الأزرق في هذه اللوحة؟"... فقد تُخدع وتقدم إجابة معقدة بلا داعٍ! ربما تكتب عن كيف أن اللون الأزرق يرمز إلى السماوات الإلهية، أو أن الدرجة اللونية المحددة تشير إلى أن اللوحة رُسمت في ساعات الصباح الباكر، مما يرمز إلى البعث... بينما الحقيقة ببساطة هي: لأن السماء زرقاء!

وتتفق أجيا كوترا، كبيرة المحللين في مؤسسة "أوبن فيلانثروبي" والباحثة في مخاطر الذكاء الاصطناعي، مع غرينبلات في هذه النقطة. وقالت عن النماذج الأحدث: "أعتقد أنها تتحسن بالفعل في مجموعة واسعة من المهام التي قد يصنفها البشر ضمن مهام الاستدلال."

إنها لا تنكر أن النماذج تقوم بنوع من المحاكاة الفوقية، لكنها أوضحت أنه عندما يقول المشككون "إنه مجرد محاكاة فوقية"، فإن الجزء المثير للجدل هو كلمة "فقط". وأضافت: "يبدو أن ما يحاولون الإيحاء به غالبًا هو: "ولذلك، لن يكون له تأثير كبير على العالم" أو "ولذلك، لا يزال الذكاء الاصطناعي الفائق بعيدًا"، وهذا ما أعترض عليه.

ولتوضيح وجهة نظرها، قالت: تخيل أنك تُدرِّس فصلًا جامعيًا في الفيزياء. لديك أنواع مختلفة من الطلاب. أحدهم مجرد غشاش: ينظر إلى الإجابات في نهاية الكتاب ثم يكتبها. وطالب آخر عبقري إلى حد أنه لا يحتاج حتى إلى التفكير في المعادلات؛ فهو يفهم الفيزياء على مستوى عميق وبديهي، مثل أينشتاين، ويمكنه استنتاج المعادلات الصحيحة مباشرة. أما بقية الطلاب، فهم في مكان ما بين هذين النقيضين: لقد حفظوا قائمة من 25 معادلة ويحاولون معرفة أي معادلة يجب استخدامها في كل موقف.

وقالت كوترا إن نماذج الذكاء الاصطناعي، مثل غالبية الطلاب، تمزج بين بعض الحفظ وبعض الاستدلال.

وقالت: "نماذج الذكاء الاصطناعي تشبه طالبًا ليس شديد الذكاء ولكنه يتمتع باجتهاد يفوق

القدرات البشرية، لذا لم يحفظ 25 معادلة فقط، بل 500 معادلة، بما في ذلك تلك الخاصة بالمواقف النادرة التي قد تطرأ" وأوضحت أن هذه النماذج تمزج بين قدر كبير من الحفظ والقليل من الاستدلال، أي تحديد أي مجموعة من المعادلات يجب تطبيقها على المشكلة.

وأضافت: "وهذا وحده يقطع شوطًا طويلًا! فهي تبدو للوهلة الأولى مثيرة للإعجاب تمامًا مثل الشخص الذي يمتلك فهمًا عميقًا وبديهيًا."

وبالطبع، عند التدقيق، لا يزال من الممكن العثور على ثغرات لم تغطها معادلاتها الـ500. لكن هذا لا يعني أن عملية الاستدلال لم تحدث على الإطلاق.

وبعبارة أخرى، فإن النماذج ليست مقتصرة على الاستدلال فقط، ولا تقتصر على التكرار فقط.

وقالت كوترا: "إنه يقع في مكان ما بين الاثنين. أعتقد أن الناس يواجهون صعوبة في تقبّل ذلك لأنهم يرغبون في تصنيفه ضمن أحد المعسكرين. إما أن يكون مجرد حفظ، أو يكون استدلالًا عميقًا حقيقيًا. ولكن الحقيقة هي أن هناك طيفًا لمستوى عمق الاستدلال."

أنظمة الذكاء الاصطناعي تمتلك "ذكاءً غير متجانس"

لقد ابتكر الباحثون مصطلحًا لافتًا لوصف هذا النمط من الاستدلال: "الذكاء غير المتجانس". ويشير ذلك إلى الحقيقة الغريبة التي أوضحها عالم الحاسوب أندريه كاربايثي، حيث يمكن لنماذج الذكاء الاصطناعي المتقدمة "أن تؤدي مهام مذهلة للغاية (مثل حل مسائل رياضية معقدة) بينما تواجه في الوقت نفسه صعوبة في حل مشكلات بسيطة للغاية."

تخيل الأمر بهذا الشكل: إذا كان الذكاء البشري يشبه سحابة ذات حواف مستديرة ناعمة، فإن الذكاء الاصطناعي يشبه سحابة شائكة ذات قمم ووديان حادة متجاورة. ففي البشر، ترتبط العديد من القدرات على حل المشكلات ارتباطًا وثيقًا ببعضها، لكن الذكاء الاصطناعي قد يكون بارعًا في شيء معين بينما يكون ضعيفًا بشكل مثير للدهشة في شيء آخر يبدو لنا قريبًا منه؛ ضع في اعتبارك أن الأمر كله نسبي.

وقال غرينبلات: "بالمقارنة مع ما يجيده البشر، تبدو النماذج غير متسقة تمامًا. أعتقد أن مقارنة النماذج بالبشر قد تكون مربكة بعض الشيء. فمن منظور النموذج، قد يبدو الأمر وكأنّه يقول: 'يا لهؤلاء البشر! إنهم غير متسقين تمامًا! إنهم سيئون جدًا في توقع الرمز التالي!' لذا، ليس من الواضح ما إذا كان هناك معيار موضوعي يجعل الذكاء الاصطناعي أكثر تباينًا".

إن حقيقة أن النماذج المنطقية مُدرَّبة على أن تبدو وكأنها تفكر مثل البشر تجعلنا نميل إلى مقارنة ذكاء الذكاء الاصطناعي بالذكاء البشري، ولكن الطريقة الأفضل للتفكير في الذكاء الاصطناعي ليست باعتباره "أذكى من الإنسان" أو "أغبى من الإنسان"، بل ببساطة على أنه "مختلف".

ومع ذلك، تتوقع كوترا أنه عاجلًا أم آجلًا، سيكون ذكاء الذكاء الاصطناعي واسعًا لدرجة أنه سيحتوي بداخله على كل الذكاء البشري، بل وأكثر من ذلك.

وقالت كوترا: "أفكر في المخاطر التي قد تظهر عندما تصبح أنظمة الذكاء الاصطناعي أفضل من الخبراء البشريين في كل شيء. قد تظل غير متساوية في قدراتها، لكن ذكاءها المتفاوت سيشمل كامل الذكاء البشري وأكثر. أنا دائمًا أنظر إلى ذلك المستقبل وأستعد له".

وفي الوقت الحالي، النتيجة العملية بالنسبة لمعظمنا هي: تذكر ما يجيد الذكاء الاصطناعي فعله وما لا يجيده – واستخدمه وفقًا لذلك.

إن أفضل حالات الاستخدام هي المواقف التي يكون من الصعب عليك فيها التوصل إلى حل، ولكن بمجرد أن تحصل على حل من الذكاء الاصطناعي، يمكنك بسهولة التحقق مما إذا كان صحيحًا. وتُعدُّ كتابة الأكواد مثالًا مثاليًا. ومثال آخر هو إنشاء موقع إلكتروني: يمكنك رؤية ما أنجزه الذكاء الاصطناعي، وإذا لم يعجبك، يمكنك ببساطة أن تطلب منه إعادة المحاولة.

وفي مجالات أخرى – خاصة تلك التي لا يوجد فيها إجابة صحيحة موضوعية أو التي تكون المخاطر فيها عالية – يجب أن تكون أكثر حذرًا عند استخدام الذكاء الاصطناعي. يمكنك الاستفادة منه في الحصول على اقتراحات مبدئية، ولكن لا تعتمد عليه كثيرًا، خصوصًا إذا بدا لك أن ما يقوله غير مقنع. فعلى سبيل المثال، عند طلب نصيحة حول كيفية التعامل مع معضلة أخلاقية، يمكنك استخدامه لإثارة بعض الأفكار، ولكن دون اعتباره مصدرًا للحل النهائي.

وقالت كوترا: "كلما كانت الأمور ضبابية وتعتمد على الحكم الشخصي، كلما كان من الأفضل استخدام الذكاء الاصطناعي كشريك في التفكير، وليس كمرجع مطلق."

المصدر: فوكس

رابط المقال : https://www.noonpost.com/303126/