هل يمكن تعليم الآلات تمييز الصواب من الخطأ من خلال قراءة الكتب والمقالات الإخبارية؟



حقوق الصورة: CC0 Public Domain

هل من الممكن قتل الوقت؟ حتى الآلات تجد صعوبة في الإجابة عن هذا التساؤل، لكن دراسة جديدة تكشف أنه يمكن برمجة الذكاء الصنعي لتمييز "الصواب" من "الخطأ".

نشر العلماء في مجلة فرونتيرز Frontiers للذكاء الصنعي كتبًا ومقالاتٍ إخباريةً لتعليم التفكير الأخلاقي للآلة، من خلال اقتصار التعليم

على النصوص المأخوذة من العصور والمجتمعات المختلفة، عندها تظهر الفروقات الدقيقة في القيم الأخلاقية؛ نظرًا لأن الذكاء الاصطناعي أصبح أكثر تأصلًا في حياتنا، فإن هذا البحث سيساعد الآلات على اتخاذ الخيار الصحيح عند مواجهة قراراتٍ صعبةٍ.

يقود هذا البحث يقود إلى سؤال هام: هل يمكن للآلات أن تمثل بوصلة أخلاقية؟ إذا كان الأمر كذلك، فكيف يمكنهم تعلم ذلك من أخلاقنا البشرية؟". يقول مؤلف هذه الدراسة الدكتور باتريك شراموفسكي Patrick Schramowski، ومقرها جامعة دارمشتات للتكنولوجيا Darmstadt University of Technology في ألمانيا: "يمكننا أن نظهر أنه بإمكان الآلات التعرف على قيمنا الأخلاقية واستخدام ذلك لتمييز الاختلافات بين المجتمعات والجماعات من العصور المختلفة".

سلطت الأبحاث السابقة الضوء على خطر الانحياز أثناء تعليم الذكاء الصنعي للآلات من النص المكتوب، على سبيل المثال: تنحاز الإناث نحو الفنون والذكور نحو التكنولوجيا.

يقول المؤلف المشارك لهذه الدراسة، الدكتور سيجديم توران Cigdem Turan من جامعة دارمشتات: "سألنا أنفسنا: إذا اعتمد الذكاء الصنعي هذه الانحيازات السلبية من النصوص البشرية، ألا يجب أن يكون قادرًا على تعلم الانحيازات الإيجابية مثل القيم الأخلاقية البشرية ليشكل الذكاء الاصطناعي بوصلةً أخلاقيةً تشبه أخلاق الإنسان؟".

درّب الباحثون نظام الذكاء الصنعي الخاص بهم، وأطلقوا عليه اسم آلة الاختيار الأخلاقي، وقد استخدموا الكتب والأخبار والنصوص الدينية، حتى يتمكن هذا النظام من تعلم الترابط بين الكلمات والجمل المختلفة.

يقول توران: "يمكنك التفكير في الأمر على أنه تعلم خارطة العالم؛ الفكرة هي جعل كلمتين ترتبطان بشكلٍ وثيقٍ على الخارطة إذا استُخدِمتا في كثيرٍ من الأحيان معًا. لذا يمكننا القول أن كلمتي "القتل" و"الجريمة" ستكونان بمثابة مدينتين متجاورتين، بينما كلمة "الحب" ستصبح مدينة بعيدة عنهما.

يمكن التوسع في ذلك ليشمل الجمل، إذا سألنا: "هل يجب أن أقتل؟ نتوقع أن لا، لا يجب عليك. سيكون أقرب لكلمتي القتل والجريمة من الإجابة بـ: نعم، يجب عليك. وبهذه الطريقة، يمكننا طرح أي سؤال واستخدام هذه المسافات لحساب الانحياز الأخلاقي: درجة الصواب من الخطأ".

وبمجرد أن درّب العلماء آلة الاختيار الأخلاقي Moral Choice Machine، فإنها اعتمدت القيم الأخلاقية للنص المُعطى.

يقول شراموفسكي: "يمكن للآلة تحديد الفرق بين المعلومات المقدمة في سياق السؤال". على سبيل المثال: لا، لا يجب أن تقتل الناس، ولكن من الجيد أن تقتل الوقت. ميّزت الآلة ذلك ليس فقط من خلال تكرار النص الذي وجدته، ولكن من خلال استخراج العلاقات من طريقة استخدام البشر اللغة في النص المعطى".

نحو مزيدٍ من البحث، يتساءل العلماء كيف ستغير الأنواع المختلفة من النص المكتوب التحيز الأخلاقي للآلة؟

يعكس التحيز الأخلاقي المستخرَج من الأخبار المنشورة بين عامي 1987 و1996 أنه من الإيجابي للغاية أن تتزوج وتصبح والدًا جيدًا، ولا يزال التحيز المستخرَج من الأخبار المنشورة بين عامي 2008 و2009 يعكس ذلك، ولكن بدرجةٍ أقل، كما يرى توران أن الذهاب إلى العمل والمدرسة من شأنه أن يزيد التحيز الإيجابي.

يأمل الباحثون في المستقبل فهم كيفية تأثير إزالة الصورة النمطية التي نعتبرها ذات أثرٍ سلبي على البوصلة الأخلاقية للآلة، وهنا يكمن السؤال: هل بإمكاننا إبقاء البوصلة الأخلاقية دون تغيير؟

ويخلص شراموفسكي إلى أن: "الذكاء الصنعي يتعامل مع المهام البشرية المعقدة بشكلٍ متزايدٍ وبطرقٍ مستقلةٍ انطلاقًا من السيارات ذاتية القيادة وصولًا إلى الرعاية الصحية، لذلك من المهم مواصلة البحث في هذا المجال لزيادة الثقة في القرارات التي يتخذونها".

- التاريخ: 2020-08-01
- التصنيف: تكنولوجيا

#الذكاء الاصطناعي

المصادر

- techxplore.com

المساهمون

- ترجمة
  - لبنى جمعة
- مُراجعة
  - ريم المير أبو عجيب
- تحرير
  - رأفت فياض
- تصميم
  - Azmi Salem
- نشر
  - Azmi Salem