نورمان .. أول نظام ذكاء اصطناعي مريض نفسيًا في العالم

الاثنين 11 يونيو 2018 02:06 م

بتكر الباحثون في معهد ماساتشوستس للتكنولوجيا MIT نظام ذكاء اصطناعي تم تدريب خوارزمياته من خلال عرض نصوص وصور مزعجة عن الموت بطرق بشعة والتي تم العثور عليها على موقع ريديت Reddit، مما أثر على هذا النظام وجعله يتصرف كأنه مريض نفسي.

أطلق الباحثون على أول نظام ذكاء اصطناعي مصاب بمرض نفسي في العالم اسم نورمان Norman، نسبةً إلى الشخصية الأساسية في فيلم سايكو Psycho للمخرج ألفريد هيتشكوك في عام 1960.

وكجزء من التجربة قام الباحثون بتدريب نظام نورمان باستخدام نصوص من مجموعات fموقع ريديت مكرسة لتوثيق ومراقبة وقائع الموت البشعة وبسبب اعتبارات فنية وأخلاقية استخدم فريق البحث في معهد ماساتشوستس للتكنولوجيا وصف وشرح للصور ولم يستخدموا الصور الحقيقية لأشخاص يموتون، لمعرفة ما إذا كانت ستغير هذه البيانات والمعلومات من سلوك نظام الذكاء الاصطناعي.

في حين تتمثل القاعدة الأساسية لمجموعات موقع ريديت -المتخصصة في توثيق وقائع الموت- في أنه يجب أن يكون هناك مقطع فيديو لشخص يموت بالفعل في المنشور الذي يتم مشاركته، ويجب أن تكون عناوين الإرسال وصفية ودقيقة بما فيه الكفاية لفهم بالضبط ما هو محتوى الموضوع مثل" طعن شاب".

بعد ذلك أخضع الباحثون نظام نورمان لسلسلة اختبار رورشاخ، هو اختبار سيكولوجي للشخصية والذكاء وضعه الطبيب النفسي السويسري هرمان رورشاخ عام 1921. والذي يختبر ما يراه الأشخاص عندما ينظرون إلى سلسلة من بقع الحبر تمثل أنماطًا تجريدية نموذجية ويطلب منهم تحليلها أو تفسيرها، ثم يتم تحليل الإجابات المقدمة من المشاركين نفسياً من أجل قياس اضطرابات التفكير المحتملة.

وقد تم إخضاع نظام ذكاء اصطناعي قياسي لنفس اختبار رورشاخ الذي تعرض له نظام نورمان لمقارنة النتائج، التي جاءت كالتالي:

في أحد اختبارات بقع الحبر التي عُرضت على النظامين، رآها نظام الذكاء الاصطناعي القياسي أنها "لقطة مقرّبة لمزهرية مزينة بالورود"، بينما رآها نظام نورمان "رجل يُقتل بالرصاص". وفي مثال آخر وصف نظام الذكاء الاصطناعي القياسي بقعة الحبر بأنها "صورة سوداء وبيضاء لطائر صغير"، بينما وصفها نظام نورمان "رجل يتم سحبه إلى آلة العجين".

وللإجابة على سؤال لماذا أنشئ معهد ماساتشوستس للتكنولوجيا نظام الذكاء الاصطناعي المريض نفسيًا؟ قال الباحثين أنهم يحاولون معرفة كل شيء عن الخوارزميات وأنه عندما تسير الأمور هكذا يكون من الصعب إلقاء اللوم على الآلة، حيث أن البيانات التي تُستخدم لتعليم خوارزميات التعلم الآلي يمكن أن تؤثر بشكل كبير على سلوكها".

لذا عندما يتحدث الناس عن أن خوارزميات الذكاء الاصطناعى متحيزة وغير عادلة، فإن الجاني في الغالب ليس الخوارزمية نفسها ولكن البيانات المتحيزة التي تم تغذيتها بها.

وقال فريق البحث الذي يضم كل من بينار ياناردغ ومانويل سيبريان وإياد رهوان من معهد ماساتشوستس للتكنولوجيا، إن الدراسة أثبتت أن البيانات المستخدمة في تدريب خوارزميات التعلم الآلى يمكن أن تؤثر بشكل كبير على سلوكها.

وأوضحت الدراسة أن نتائج التجربة تظهر أنه عندما يتم اتهام الخوارزميات بأنها "منحازة أو غير عادلة" فإن الجاني في الغالب ليس الخوارزمية نفسها ولكن البيانات المتحيزة التي تم إدخالها لها.