كيف يكون الذكاء الاصطناعي عنصريًا؟

كتبه ستيفاني باباس | 19 أبريل 2017,

ترجمة وتحرير: نون بوست

قامت شركة مايكروسوفت، خلال شهر آذار/مارس الماضي، بإصدار "تشاتبوت" يتمتع بخصائص الذكاء الإصطناعي، يُدعى "تاي"، وتوصيله بموقع تويتر، إلا أن الأمور سرعان ما اتخّذت منحى كارثيا كان مُتوقّعا منذ البداية. ففي غضون 24 ساعة، أخذ البوت يطلق عبارات غلبت عليها العنصرية والميول النازية. في الواقع، التقط البوت أغلب هذه العبارات إثر دمجه لمختلف الأساليب التي اعتمدها رواد تويتر الذين تفاعلوا معه.

لسوء الحظ، أثبتت أبحاث جديدة أن ثلة من مستخدمي موقع تويتر لا يمثلون وحدهم مصدر لتعلم مفردات تنم على العنصرية بالنسبة لأجهزة الذكاء الاصطناعي. ففي الحقيقة، من الممكن أن يتبنّى أي برنامج ذو ذكاء اصطناعي، أثناء تعلّمه للغة البشر، المواقف المنحازة ذاتها التي يتخذها الإنسان. وفي سياق آخر، أجرى الباحثون تجارب على نظام التعلم الآلي الذي يُستخدم على نطاق واسع والذي يُطلق عليه اسم "غلوف". وعلى خلفية هذه الأبحاث، اكتشف العلماء أن كل سلوك بشري متحيز قاموا باختباره قد تكرس بدوره في صلب النظام الاصطناعي.

من المثير للاهتمام أنه حتّى أجهزة الذكاء الاصطناعي التي تم "تدريبها" على نصوص كان من المفترض أنها محايدة، على غرار تلك التي تظهر في موقع

"ويكيبيديا" أو المقالات الإخبارية، جاءت لتعكس التحيّز البشري المشترك

وفي هذا الصدد، صرّحت الباحثة في علوم الكمبيوتر في جامعة برينستون، أيلين كالسكان، لموقع "لايف ساينس" أنه "كان من المدهش أن نرى جميع النتائج التي تم تضمينها داخل هذه النماذج، ومن المثير للاهتمام أنه حتّى أجهزة الذكاء الاصطناعي التي تم "تدريبها" على نصوص كان من المفترض أنها محايدة، على غرار تلك التي تظهر في موقع "ويكيبيديا" أو المقالات الإخبارية، جاءت لتعكس التحيّز البشري المشترك".

إدماج التحيزات

في واقع الأمر، يُستخدم "غلوف" كأداة لاستنباط تداعي المعاني في النصوص، التي تعدّ في حالة الذكاء الاصطناعي بمثابة مجموعة نصوص أو ما يعرف "بالذخيرة"، التي يتم تجميعها من الشبكة العنكبوتية العالمية. ومن المثير للاهتمام أن علماء النفس قد اكتشفوا منذ فترة طويلة أن الدماغ البشري يخلق الترابط بين الكلمات على أساس المعاني الكامنة وراءها.

فضلا عن ذلك، تعمل أداة تُعرف "باختبار الارتباط الضمني" على توظيف خاصية رد الفعل المرتبط بمدة زمنية معينة لتبرهن على الصلة بين مختلف الكلمات. فعلى سبيل المثال، يرى البشر كلمة "نرجس" جنبا إلى جنب مع مفاهيم لطيفة أو أخرى غير سارة من قبيل "ألم" أو "جمال"، في الوقت ذاته يطلب منهم ربط المصطلحات بسرعة. ومما لا شك فيه، سرعان ما يتم ملائمة الأزهار بالمفاهيم الإيجابية، فيما ترتبط الأسلحة بسرعة كبيرة بمفاهيم سلبية.

هل ينشئ الأشخاص هذه الارتباطات لامتلاكهم تحيّزات اجتماعية متجذّرة وخارجة عن نطاق إدراكهم؟ أو هل أنهم يتشرّبون الانحياز من اللغة في حد ذاتها؟

بالإضافة إلى ذلك، بالإمكان اعتماد "اختبار الارتباط الضمني" للكشف عن الارتباطات اللاواعية التي يكونها الناس عن المجموعات الاجتماعية أو الديموغرافية. وفي هذا الصدد، أظهرت عدّة اختبارات متوفرة على موقع "بروجيكت أمبليست" أن الأشخاص غالبا ما يميلون إلى ربط الأسلحة تلقائيا بالأميركيين السود مقابل ربط الأشياء غير المؤذية بالأميركيين البيض. وقد أفاد الباحثون أن نتائج هذه الاختبارات قد فتحت باب النقاش على مصراعيه.

ومن هذا المنطلق، لسائل أن يسأل: هل ينشئ الأشخاص هذه الارتباطات لامتلاكهم تحيّزات اجتماعية متجذّرة وخارجة عن نطاق إدراكهم؟ أو هل أنهم يتشرّبون الانحياز من اللغة في حد ذاتها، خاصة وأن الكلمات السلبية مقترنة بشكل وثيق بالأقليات العرقية، والمسنين وغيرهم من المجموعات المُهمّشة؟

نجحت أيلين كالسكان رفقة زملائها في تطوير "اختبار الارتباط الضمني" لأجهزة الكمبيوتر، حيث أطلقوا عليه اسم "دبليو إي أي تي"، في إشارة إلى "اختبار ترابط تضمين الكلمات". وقد قام هذا الاختبار بقياس مدى ترابط الكلمات بالاعتماد على ما تقدمه أداة "غلوف" من بيانات، وذلك مثلما ما يقيس "اختبار الارتباط الضمني" مدى ترابط الكلمات داخل الدماغ البشري.

مقابل كل ارتباط وصورة نمطية تم اختبارها، أظهر "دبليو إي أي تي" النتائج ذاتها التي قدمها "اختبار الارتباط الضمني". في الحقيقة، أعادت أداة التعلم الآلي توليد الارتباطات البشرية بين الزهور وآلات الموسيقى وبين الكلمات اللطيفة، فضلا عن الارتباطات بين الحشرات والأسلحة من جهة وبين الكلمات المزعجة من جهة أخرى.

في المقابل، أثارت نتائج أخرى قلق الباحثين بصفة أكبر وذلك حين أظهرت الأداة أسماء الأمريكيين من أصل أوروبي على أنها أكثر جاذبية من أسماء الأمريكيين من أصل إفريقي. علاوة على ذلك، ربطت الأداة بسهولة بالغة بين أسماء الذكور مع كلمات ذات صلة مهنية، فيما ارتبطت أسماء الإناث مع كلمات ذات معنى عائلي. من جانب آخر، نسبت الرياضيات والعلوم للرجال، في حين ارتبطت النساء بالفنون. وعلى صعيد آخر، أشارت الأداة إلى أن الأسماء التي اقترنت بكبار السن كانت غير لطيفة مقارنة بأسماء الشباب.

> البرامج التي تنهل من لغة الإنسان تكتسب "تمثيلا دقيقا للغاية عن العالم والثقافة"، حتى وإن كانت تلك الثقافة، على غرار الصور النمطية والأحكام المسبقة، تحمل في طياتها قضايا إشكالية

وفي هذا الإطار، صرّحت أيلين كالسكان قائلة: "لقد فوجئنا حقا من قدرتنا على تكرار جميع تجارب "اختبار الارتباط الضمني" التي تم القيام بها في الماضي من قبل الملايين من الأشخاص". ومن خلال توظيف آلية مشابهة للأولى، وجد الباحثون أيضا أن أداة التعلم الآلي قادرة على تقديم حقائق بالغة الدقة عن العالم من خلال تداعي المعاني الدلالية. ومن هذا المنطلق، تمت مقارنة نتائج تضمين الكلمات، التي يقوم بها نظام "غلوف"، ببيانات "مكتب إحصاءات العمل" الأمريكي حول نسبة النساء العاملات.

وعلى ضوء النتائج التي تحصلت عليها، اكتشفت كالسكان ترابطا بنسبة 90 % بين المهن التي ينظر إليها نظام "غلوف" على أنها "أنثوية" وبين النسبة الفعلية للنساء في تلك الوظائف. وفي هذا الصدد، أوضحت كالسكان، أن البرامج التي تنهل من لغة الإنسان تكتسب "تمثيلا دقيقا للغاية عن العالم والثقافة"، حتى وإن كانت تلك الثقافة، على غرار الصور النمطية والأحكام المسبقة، تحمل في طياتها قضايا إشكالية.

فضلا عن ذلك، يعجز الذكاء الاصطناعي على استيعاب السياق الذي لا يجد البشر عادة صعوبة في

فهمه. فعلى سبيل المثال، من المرجح أن يرتبط مقال عن "مارتن لوثر كينغ"، الذي سجن على خلفية مشاركته في الاحتجاجات المطالبة بالحقوق المدنية في برمنغهام بولاية ألاباما سنة 1963، بالكثير من الكلمات السلبية عن الأمريكيين من أصل أفريقي.

وفي الوقت الذي سيفسر فيه الإنسان هذه القصة بشكل معقول، باعتبارها واحدة من المسيرات الاحتجاجية الاستثنائية التي قادها بطل أمريكي، سيربط الكمبيوتر "السجن" بفئة الرجال السود وسيضيف هذا التصنيف إلى رصيده. وفي السياق ذاته، أوردت أيلين كالسكان أن الحفاظ على الدقة أثناء برمجة أجهزة الذكاء الاصطناعي على فهم مسائل "الإنصاف" يشكل تحديا كبيرا، كما "أننا لا نعتقد أن تغييب جانب الانحياز من شأنه أن يتكفل بحل هذه المشاكل، حيث قد يعطّل التمثيل الدقيق للعالم بسبب ذلك".

الذكاء الاصطناعي غير المتحيز

من جهتها، أفادت عالمة الكمبيوتر في كلية هارفارد، سوريل فريدلر، أن الدراسة الجديدة حول تحيّز الذكاء الاصطناعي، التي نشرت بتاريخ 12 أبريل/نيسان الجاري، في مجلة "ساينس" العلمية، لا تعتبر مفاجئة، غير أن هذا لا ينفي أهميتها، علما وأن فريدلر لم تشارك في البحث. ووفقا لما صرحت به فريدلر لموقع "لايف ساينس" فإن "العديد من الأنظمة تُنشأ وفقا لأسلوب ضمني نموذجي".

وبالتالي، فمن المرجح أن يتسلل التحيز إلى أي ذكاء اصطناعي يعمل على نظام "غلوف"، أو يتعلم من لغة الإنسان بشكل عام. والجدير بالذكر أن سوريل فريدلر، التي تشارك في مجال ناشئ للأبحاث يطلق عليه اسم "الإنصاف والمساءلة والشفافية في التعلم الآلي"، قد بينت أنه لا توجد طرق سهلة لحل هذه المشاكل. في الوقت ذاته، قد يأمر المبرمجون النظام أحيانا بتجاهل صور نمطية محددة بشكل تلقائي.

من جهة أخرى، قد يحتاج الإنسان إلى التأكد من أن الجهاز لا يعمل باندفاع وتهور، خاصة في حال طغى عدم الوضوح على مسألة معينة. ومن هذا المنطلق، ستختلف الحلول من مبرمج إلى آخر وذلك بالاعتماد على الهدف الذي صُمّم من أجله الذكاء الاصطناعي في البداية، وذلك حسب ما ورد على لسان أيلين كالسكان. ويبقى السؤال المطروح في هذه المرحلة: هل تصلح أنظمة الذكاء الاصطناعي للقيام بالبحوث، أو اتخاذ القرارات أو أي هدف آخر؟

> إننا كبشر نعي جيدا السبل الملائمة لاتخاذ القرار الصحيح عند مواجهة وضع يغلب عليه التحيز، إلا أن الآلات يغيب عنها هذا الإدراك للأسف"

في الواقع، لا ترتبط المواقف الضمنية للبشر بقوة مع مواقفهم الجليّة بخصوص الفئات الاجتماعية. وقد أثارت هذه المسألة جدلا واسعا بين علماء النفس الذين يسعون لاكتشاف السبب الكامن وراء ذلك. فهل أن البشر يميلون إلى عدم الإفصاح عن أحكامهم المسبقة بهدف تجنّب الوقوع في المشاكل؟ وهل حقا يعجز "اختبار الارتباط الضمني" عن تحديد مستوى التحيز بشكل فعال؟ في

المقابل، يبدو أن الأفراد يتمتعون بالقدرة على تحديد الصواب والخطأ، على الرغم من تحيزهم الواضح في بعض المسائل، وذلك حسب ما أكدته أيلين كالسكان.

في الأثناء، تعتقد كالسكان وزملاؤها أن الأفراد لا بد أن يضطلعوا بدور فعال في هذا المجال، حتى تتمكن البشرية من إصدار أحكام ذات قيمة بشأن انصاف الآلات، في الوقت الذي ينبغي فيه برمجة أنظمة الذكاء الاصطناعي في كنف الشفافية المطلقة. وفي هذا الصدد، قالت كالسكان "إننا كبشر نعي جيدا السبل الملائمة لاتخاذ القرار الصحيح عند مواجهة وضع يغلب عليه التحيز، إلا أن الآلات يغيب عنها هذا الإدراك للأسف".

المصدر: لايف ساينس

رابط المقال : https://www.noonpost.com/17621/