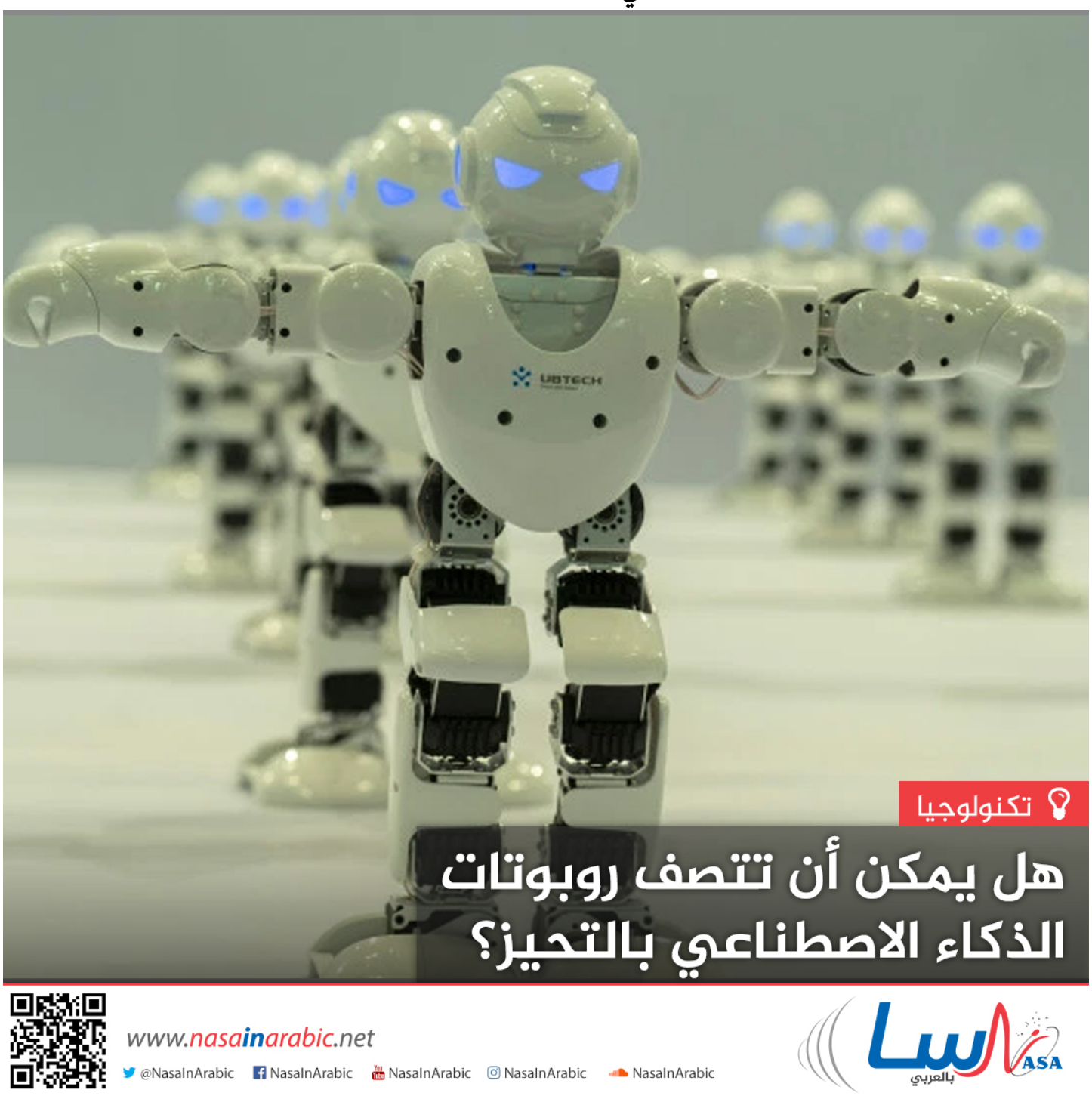
# هل يمكن لروبوتات الذكاء الاصطناعي أن تكون متحيزة؟




**لا يتطلب إظهار التحيّز ضد الآخرين مستوى عالٍ من القدرة الإدراكية ويمكن إظهاره ببساطة من قبل آلات الذكاء الصنعي، كما يقترح بحثٌ جديد.**

أظهر خبراء في علوم الحاسب وعلم النفس من جامعة كارديف **Cardiff University** ومعهد ماستشوستس للتكنولوجيا **MIT** بأن مجموعات من الآلات ذاتية التحكم قادرة على إظهار التحيّز ببساطة من خلال تحديد ونسخ وتعلم هذا السلوك من بعضها البعض، على الرغم من أن التحيّز قد يبدو وكأنه ظاهرة بشرية بحتة تتطلب الإدراك البشر لتشكيل رأي أو صورة نمطية حول شخص أو مجموعة معينة.

على الرغم من أن بعض الخوارزميات الحاسوبية قد أظهرت التحيّز سابقاً (مثل العنصرية أو التمييز على أساس الجنس) وذلك بناءً على

ما تعلمته من السجلات العامة وغيرها من البيانات المُولدة من البشر، فإن هذا البحث الجديد يثبت إمكانية تطوير الذكاء الاصطناعي لتحيزات جماعية بمفرده.

نُشرت هذه الاكتشافات الجديدة في مجلة ساينتيفيك ريبورتس **Scientific Reports**، وقد بُنيت على محاكاة حاسوبية أظهرت التماثل بين البشر والعملاء الافتراضيين في اختيارهم للمجموعات التي يتفاعلون معها، أي في تحيزهم ضد المجموعات الأخرى. في لعبة قائمة على مبدأ "الأخذ والعطاء"، يتخذ كل شخصٍ قراراً فيما إذا كان سيتبرع إلى شخص من مجموعته أو من مجموعة مختلفة بناءً على السمعة الشخصية لهذا الشخص (الذي سيتم التبرع له) بالإضافة الى استراتيجية التبرع الخاصة بالمُتبرِّع والتي تتضمن مستويات التحيّز تجاه الغرباء الخاصة به.

مع تقدم اللعبة وتخزين الآلاف من عمليات المحاكاة هذه من قبل حاسوب عملاق يبدأ كل فرد بتعلم استراتيجيات جديدة من خلال تقليده للآخرين ضمن مجموعته أو ضمن اللعبة بأكملها.

البروفيسور روجير ويتيكر **Roger Whitaker** المشارك في الدراسة من معهد أبحاث الجريمة والأمن ومدرسة علوم الحاسب والمعلوماتية في جامعة كارديف، يقول: "من خلال إجراء هذه المحاكاة آلاف وآلاف المرات بدأنا بفهم كيفية يتطور التحيّز والظروف التي تساهم في ترقيته أو إعاقته". ويتابع: "إن محاكاتنا تُظهر بأن التحيّز هو أحد القوى الطبيعية ومن خلال التطور يمكن بسهولة أن يُحفَّز ضمن الأنظمة الافتراضية ليحدّ من الاتصال الأوسع مع الآخرين. يمكن للحماية من الجماعات المتحيزة أن تؤدي عن غير قصد إلى تشكيل الأفراد لمجموعات متحيزة أخرى، مما يؤدي إلى انقسام في المجتمع، وهذا النوع من التحيّز واسع النطاق يصعب عكسه".

تضمن الاكتشافات تحديث الأفراد (الأنظمة الاصطناعية) بتحديث لمستويات تحيزهم من خلال تقليدهم بشكل تفضيلي لأولئك الذين حصلوا على أعلى مردود على المدى القصير، بمعنى أن هذه القرارات لا تتطلب بالضرورة قدرات إدراكية متقدمة.

يقول البروفيسور ويتيكر: "من المتوقع أن الآلات ذاتية التحكم التي أظهرت القدرة على التحيّز وتقليد الآخرين قد تتعرض في المستقبل لظاهرة التحيّز ذاتها التي نراها عند البشر. العديد من التطورات في مجال الذكاء الاصطناعي في هذ الأيام تتضمن الأتمتة والتحكم الذاتي، حيث أن سلوك الاجهزة يمكن أن يتأثر أيضاً بسلوك الأجهزة الأخرى المحيطة بها، ومن الأمثلة على ذلك السيارات ذاتية القيادة وأجهزة إنترنت الأشياء. تقدم دراستنا مفهوماً نظرياً حول الأنظمة الإلكترونية التي تعتمد على غيرها دورياً كنوع من الموارد."

من الاكتشافات الأخرى المثير للاهتمام في هذه الدراسة ظهر تحت ظروف معينة تضمنت مجتمعات فرعية (من الأنظمة) أكثر تخصصاً، وفي هذه الحالة كان ترسيخ التحيّز أكثر صعوبة.

اختتم البروفيسور ويتيكر: "مع وجود عدد أكبر من المجتمعات الفرعية يمكن أن تتعاون تحالفات المجموعات الغير متحيزة دون استغلالها، وهذا يقلّل أيضاً من وضعهم كأقليات مما يقلّل من القابلية للتحيّز ضدهم. ومع ذلك، فإن هذا يتطلب أيضاً وجود ظروف يكون فيها للأنظمة المدروسة ميل أعلى للتفاعل خارج مجموعاتها".

- التاريخ: 2019-02-06
- التصنيف: تكنولوجيا

#الروبوتات #الذكاء الاصطناعي

## المصادر

- sciencedaily

## المساهمون

- ترجمة
  - محمد الطويل
- مُراجعة
  - فرح درويش
- تحرير
  - ليلاس قزيز
  - زين صالح
- تصميم
  - عبد الرحمن محيي
- صوت
  - أهلة عبيد
- نشر
  - أمل أحمد