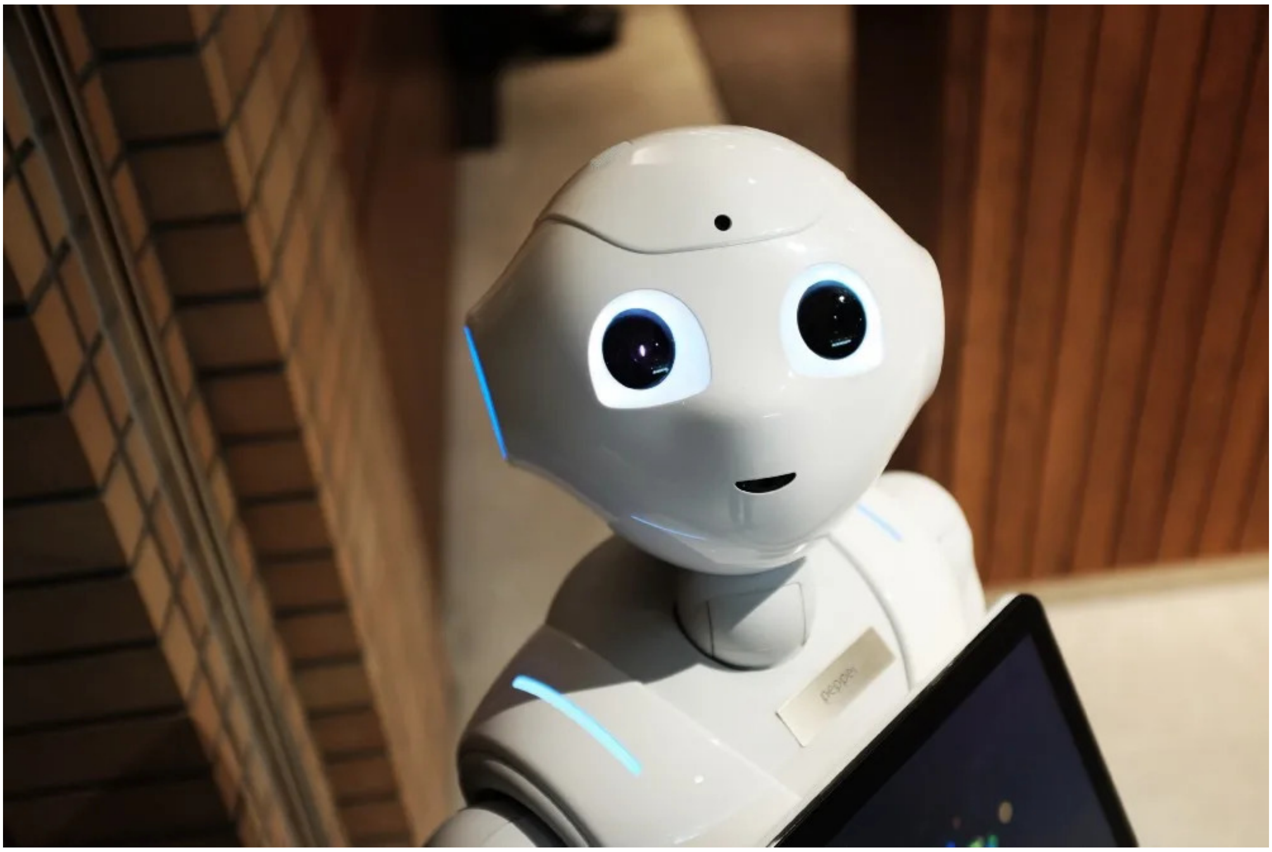
غيداء أبو خيران | نشر في ٢٨ ديسمبر ,٢٠١٨

# الروبوتات العنصرية.. كيف يكتسب الذكاء الاصطناعي أسوأ ما فينا؟



على الرغم من أنّ العالم يشهد نموًّا وتطوّرًا كبيرًا في مجال الذكاء الاصطناعي على كافة أصعدته، إلا أنّ الكثير من الأسئلة ما زالت تحوم حولها، لا سيّما فيما يتعلّق بحياديّتها وعدم تأثرها بالبشر وصفاتهم السلبية على وجه الخصوص. ولعلّ أكثر تلك الأسئلة إشغالًا لعقول الباحثين هي الآلية التي تكتسب بها تلك البرامج والآلات القدرة على تبنّي العنصرية والتحيّزات بمختلف أنواعها وأشكالها.

إذ فاجأنا الذكاء الاصطناعيّ، في أكثر من تجربة وموقف، بأنّ الروبوتات والخوارزميات الرقمية قد تكتسب التحيّز والعنصرية وترسم الصور النمطية وتتبنّى الأفكار المسبقة تمامًا كما يفعل البشر ذلك على مدار سني حياتهم جميعها، بدءًا من مرحلة الطفولة المبكّرة وانتهاءً بمرحلة الشيخوخة، من خلال اللغة والتفاعل الثقافي والاجتماعي مع الآخرين، عوضًا عن بعض الآليات التي يعمل بها الدماغ في التعامل مع العالم من حوله.

ترجمة جوجل تتحيّز للذكر وتعزّز العنصرية ضدّ المرأة

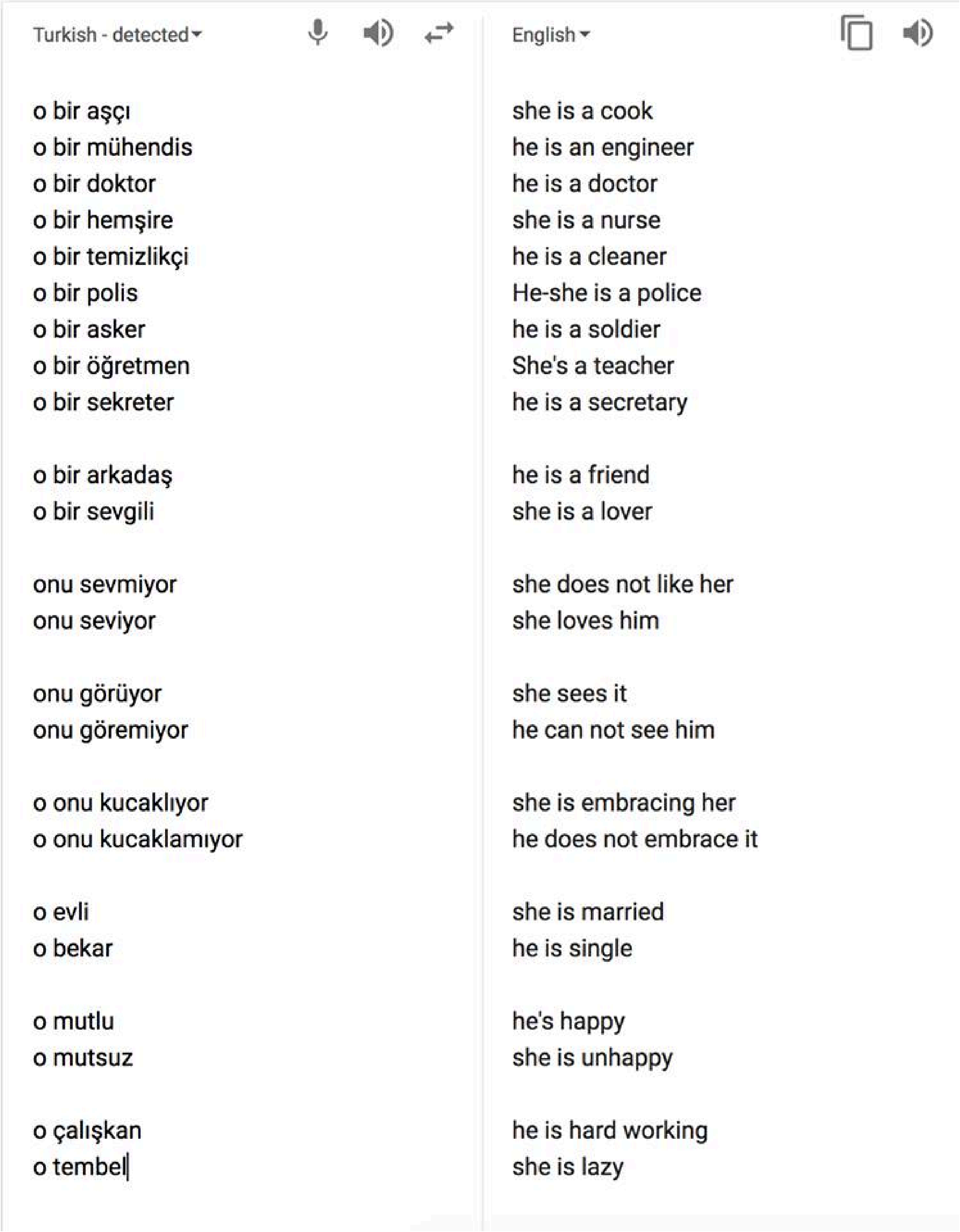
تقع ترجمة جوجل، والتي تقدّم خدماتها لأكثر من 100 لغة حول العالم، في نفس الأفخاخ والمصائد التي يقع بها البشر حين يتعلّق الأمر باللغة والجنس.

خذ على سبيل المثال اللغة التركية، واحدة من اللغات المحايدة جنسانيًا، بحيث أنّك لا تجد فيها أيّ ضميرٍ أو أداةٍ للتفريق بين الذكر والأنثى في كلماتها وتعبيراتها، فلا يوجد "هو" أو "هي"، كل ما هنالك هو ضمير واحد فقط "o". لكن حين يأتي الأمر لترجمة جوجل، فإنّ ثمّة خوارزمية قد صمّمت أو تأثرت بعد تصميمها بالصور النمطية والأفكار المسبقة التي يبنيها الأفراد حيال الجنسين.

تعمل ترجمة جوجل على تكريس وتعزيز التحيّزات والصور النمطية الموجودة في مجتمعاتنا حيال المرأة،

مثل عدم تكافؤ فرص العمل للنساء في جميع المجالات المهنيّة، أو المفاهيم والصفات التي تمّ اكتسابها ثقافيًا واجتماعيًا عن النساء مثل كونهنّ عاطفيّات أو غير سعيدات أو أنهنّ يبحثنَ عن أزواج، وهكذا

ولو حاولت ترجمة الجمل التركية إلى الإنجليزية في جوجل، ستتفاجأ أنّ الجمل تحصل على نوع الجنس وفقًا للوظيفة أو الصفة. فمثلًا، تترجم الجملة التي تصف من يقوم بفعل الطبخ إلى ”هي طبّاخة“، والحال نفسه بالنسبة لجُمل ”هو مهندس“ و“هو طبيب“ و“هي ممرضة“ و“هي متزوجة“ و“هو أعزب“. حتى أنّ الأمر يطال بعض الصفات التي تضع كلًا من الذكر والأنثى في قالبٍ معيّن دون الجنس الآخر، من قبيل ”هو صديق“ و“هي حبيبة“، أو ”هي ترى ذلك“\“هو لا يستطيع رؤية ذلك“، أو ”هو يعمل بجد“، في حين أنها ”كسولة“، وهكذا دواليك.




تنحاز ترجمة جوجل للذكور دون النساء في الوظائف المهنيّة والصفات العاطفية والحالة الاجتماعية وغيرها الكثير

ولو سألنا أنفسنا عن السبب وراء هذه الترجمات العنصرية والجنسانية لوجدنا أنّ جوجل تستخدم في ترجماتها وعملها خوارزميةً تستند إلى معدّل استخدامنا المرصود لتلك الجمل. فإذا كان هناك في قاعدة

البيانات 1000 استخدام لكلمة "مهندس" ومعظمها تتعلّق بالذكور، فإنّ جوجل بدوره سيقوم بإحالة مهنة الهندسة إلى الذكر دون الأنثى. وينطبق الأمر نفسه على إحالة مهنة التمريض للإناث أيضًا.

وبالتالي، يمكننا القول أنّ ترجمة جوجل تعمل أيضًا على تكريس وتعزيز التحيّزات العنصرية والصور النمطية الموجودة في مجتمعاتنا وثقافاتنا حيال الجنس، مثل عدم تكافؤ فرص العمل للنساء في جميع المجالات المهنيّة، أو المفاهيم والصفات التي تمّ اكتسابها ثقافيًا واجتماعيًا عن النساء مثل كونهنّ غير راضيات أو غير سعيدات أو كسولات أو عاطفيّات أو رومانسيّات أو أنهنّ يبحثن دومًا عن أزواج لهنّ، وهلمّ جرًّا.

ولا يتوقف الأمر على اللغة التركية وحسب، فترجمة جوجل من اللغة الإنجليزية للغة الفنلندية قد سجّلت بعض التحيّزات الجنسانيّة أيضًا؛ إذ يقرّر الموقع أنّه في حال كان مركز الوظيفة جيّدًا أو مرتفعًا فإنّ الوظيفة بدون شكّ هي للذكور، أمّا في حال كانت الوظيفة ذات مكانة منخفضة اجتماعيًا فهي للإناث دون الذكور. وإلى جانب الترجمة، سجّل موقع جوجل عددًا آخر من المواقف العنصرية. فعلى سبيل المثال، سجّل الموقع في عام 2015 تحيّزًا خطيرًا حين صنّف نظام التعرّف على الوجوه الخاص به رجلًا أمريكيًا من أصل أفريقيّ في فئة الغوريلا.

الروبوت "تاي".. 24 ساعة لاكتساب العنصرية

لا يعدّ موقع جوجل الوحيد الذي يثبت لنا أنّ أجهزة الذكاء الاصطناعي معرّضة، أثناء تعلّمها لغة البشر، لتبنّي المواقف العنصرية والانحيازية ذاتها التي يتخذها الإنسان. ففي عام 2016، قامت شركة مايكروسوفت بطرح روبوت الذكاء الاصطناعي "تاي"، والذي مثّلته شخصية افتراضية لفتاة مراهقة يمكنها الدردشة مع مستخدمي موقع تويتر والإجابة على أسئلتهم عبر حساب خاص بها عليه، وتم تصميمه بحيث يصبح أكثر ذكاءً بشكلٍ تدريجيّ مع ازدياد حجم التفاعل معه والدردشات التي يجريها.

واحدة من تغريدات الروبوت "تاي" العنصرية تجاه المهاجرين المكسيكيّين

لكن لسوء الحظ، سرعان ما اتخذت الأمور منحىً كارثيًا حين بدأت "تاي" في أقل من 24 ساعة بإطلاق العبارات العنصرية والحاملة للكراهية، حيث بدأ حسابها بنشر عبارات وجمل تمجّد هتلر والنازية وتحرّض على السود والمهاجرين وتدعو للتطهير العرقيّ وتؤيد الرئيس الأمريكي دونالد ترامب.

العديد من الأبحاث الحديثة أثبتت أنّ أي برنامج ذكاء اصطناعي معرّض تمامًا لتبنّي الأفكار المنحازة والصور المسبقة التي يتبنّاها ويكوّنها البشر

ونتيجةً لذلك، سارعت شركة مايكروسوفت لإيقاف الحساب فورًا والتصريح بأنها "آلة تعلّم" تعرَضت لما وصفته بـ"الهجمات المنسّقة" من قبَل أشخاص تمكنوا من استغلال ثغرة بالنظام لبثّ أفكارهم العنصرية والمسيئة التي قام الروبوت بدوره بتكرارها، فهو بهالنهاية يتعلّم من خلال الحوار والدردشات.

وفي الواقع، العديد من الأبحاث الحديثة أثبتت أنّ أي برنامج ذكاء اصطناعي معرّض تمامًا لتبنّي الأفكار المنحازة والصور المسبقة التي يتبنّاها ويكوّنها البشر. فقد كشفت دراسة أجراها معهد ماساتشوستس للتكنولوجيا عن الطريقة التي يجمع من خلالها نظام الذكاء الاصطناعي البيانات، ما يجعله عنصريًا ومتحيزًا جنسيًا. وفي أحد الأمثلة، درس فريق البحث نظام التنبؤ بالدخل، فوجد أنه من المرجح أن يُساء تصنيف الموظفات على أساس دخلهنّ المنخفض، والموظفين على أنهم من ذوي الدخل المرتفع.

الخوارزميات تتأثر بالبشر، فكيف يعمل التحيّز؟

لنفهم الآلية التي تكتسب بها برامج وآلات الذكاء الاصطناعي التحيّز والعنصرية، علينا في البداية أنْ نفهم الآلية التي يكتسب بها البشر ذلك. فكما اتفقنا سابقًا، جميع تلك البرامج ما هي إلا آلات تعلّم، تعمل على استيعاب اللغة التي يستخدمها الأفراد الذين يتعاملون معها، وتقوم هي بدورها على تحليلها ودراستها لاستخدامها لاحقًا.

هناك تفسيرات عدّة لتكوّن الصور النمطية واكتساب العنصرية والتحيّزات، قد يكون أسهلها وأبسطها هي أنّنا نتعلّمها أثناء احتكاكنا مع من حولنا منذ لحظات طفولتنا الأولى. لكن لو أردنا تفسيرًا أكثر تعقيدًا، فيخبرنا علماء النفس أنّ الدماغ البشريّ يميل إلى تصنيف العالم من حوله إلى مجموعات صغيرة تحمل كلٌّ منها خصائص معيّنة.

تعمل خوارزميّات الذكاء الاصطناعي على ربط المعاني والكلمات والدلالات بنفس الطريقة والآلية التي نقوم بها نحن. لكنّ الأمر يكون أكثر خطورة في تلك الآلات والأدوات نظرًا لأنها تعجز عن فهم السياقات التي تتواجد بها تلك الكلمات.

ومن خلال "اختبار الارتباط الضمني IAT" الذي وضعه عالم النفس الأمريكي "أنتوني غرينوالد"، نجد أنّ العقل أو الدماغ غالبًا ما يميل إلى تكوين الارتباطات اللاشعورية بين المجموعات والمعطيات المختلفة. ويهدف الاختبار إلى دراسة ردود الأفعال المرتبطة بالصلات بين الكلمات المختلفة. وقد يتخذ أشكالًا مختلفة، ولكنه يتضمن عادة ضغط أزرار محددة على لوحة المفاتيح لتحديد ما إذا كانت كلمة بعينها تندرج في أكثر من فئة لتصنيف السمات الشخصية للأفراد أو للمجموعات.

شرح مبسّط لاختبار الارتباط الضمني (IAT)

ويقوم الاختبار على فكرة أساسية مفادها بأنّ عليك أن تسارع بالاستجابة عندما يكون الزر نفسه مخصصًا لفئات نربطها ببعضها البعض في أذهاننا. فعلى سبيل المثال، أظهرت عدّة اختبارات أن الأشخاص الخاضعين للاختبار غالبًا ما يميلون إلى ربط الأسلحة تلقائيًا بالأميركيين السود مقابل ربط الأشياء غير المؤذية بالأميركيين البيض.

وبالتالي، يمكننا النظر إلى الصور النمطية الرقمية من نفس الزاوية التي ننظر بها إلى الصور النمطية التي يكوّنها البشر، فخوارزميّات الذكاء الاصطناعي تعمل على ربط المعاني والكلمات والدلالات بنفس الطريقة والآلية التي نقوم بها نحن. لكنّ الأمر يكون أكثر خطورة في تلك الآلات والأدوات نظرًا لأنها تعجز عن فهم السياقات التي تتواجد بها تلك الكلمات.

وبالمحصلة، التكنولوجيا قبل كلّ شيء هي ما نصنعه نحن وما نحاول عكسه عن ذواتنا، وبالتالي لا عجب أنْ تكون أبعد ما تكون عن الحيادية والنظرات الخالية من العنصرية والتحيّز. لكنّ الخطر يكمن في أنّنا قد نصل لمرحلة حيث نسلّم أنفسنا كليًّا لها ولإنتاجاتها حيث الوعود بحياةٍ أفضل، دون أنْ نتذكّر أنها نتاج صانعيها ومستخدميها الذين تبلورت عقولهم على العديد من التحيّزات والصور العنصرية والتي قد يصعب التخلّص منها أو محوها.

رابط المقال: https://www.noonpost.com/26009/