BIAS AND UNFAIR DISCRIMINATION THEMATIC AREA NARRATIVE IN ENGLISH ARABIC FRENCH PORTUGUESE AND SPANISH

Rachel Adams , Kelly Stone

Rachel Adams , Kelly Stone

IDRC GRANT / SUBVENTION DU CRDI : - GLOBAL INDEX ON RESPONSIBLE ARTIFICIAL INTELLIGENCE

Global Index on Responsible AI

1 . 1. التحيّز والتمييز غير العادل

المؤشر العالمي للذكاء الاصطناعي المسؤول

البعد: حقوق الإنسان والذكاء الاصطناعي

البعد الفرعي: حقوق مدنية وسياسية

المجال المواضيعي: التحيّز والتمييز غير العادل

تعريف المصطلحات

يُعرّف **التحيّز** على أنه موقف لصالح أو ضدّ، فرد أو مجموعة، بطريقة غير عادلة في كثير من الأحيان. وتتكوّن الأحكام المسبقة نتيجة لميل الإنسان إلى تصنيف الناس إلى مجموعات وفقًا لخصائص معينة (مثل العرق، والجنس، والجنسية، والطبقة الاجتماعية، إلخ) وعلى أساس المستويات المختلفة للسلطة والمكانة والموارد التي يمتلكونها بشكل عام. إذ يمكن أن تؤدي هذه التصنيفات إلى صياغة أحكام قيميّة حول الأفراد تقوم على مستوى السلطة والمكانة والموارد التي يفترض أنهم يمتلكونها، وذلك اعتمادًا على هذه الخصائص والانتماء إلى مجموعات معينة. وغالباً ما تغذي العلاقات والمجموعات الاجتماعية والمؤسسات، بما فيها مؤسسات القطاع الخاص والعام، والمجتمع هذه الأحكام المسبقة وتعززها.

يُعرّف **التحيّز الخوارزمي** على أنه أخطاء منهجيّة ومتكرّرة في نظام مدعوم بالذكاء الاصطناعي تؤدي إلى نتائج غير عادلة، مثل تفضيل مجموعة على أخرى. ويمكن أن يكون التحيز الخوارزمي نتيجةً لاستخدام بيانات غير تمثيلية أو غير كاملة، أو بسبب استخدام معلومات خاطئة تعكس بدورها عدم مساواة تاريخي. ويمكن أن يؤدي ذلك إلى صياغة قرارات ذات تأثير تمييزيّ على أفراد أو مجموعات معيّنة من الأشخاص، بغض النظر عن وجود نية للتمييز أو لا. كما يمكن أن يؤدي ذلك إلى توزيع غير متوازن للفرص والموارد والمعلومات، بما في ذلك الحصول على الحقوق الاجتماعية والاقتصادية، وانتهاك حقوق الإنسان والحريات

المدنية، وفيما يتعلق بضمان أولويّة سلامة بعض الأفراد ورفاهيّتهم على حساب آخرين، وإضفاء الشرعيّة على الممارسات التمييزيّة.

يتمُّ تعريف **التمييز** على أنه معاملة فرد أو مجموعة من الأشخاص بشكل مختلف عن الآخرين بسبب العمر، أو الإعاقة، أو الجنس، أو العرق، أو الميول الجنسية، أو أية خصائص أخرى. **فالتمييز غير العادل** هو معاملة فرد أو مجموعة من الأشخاص بطريقة غير مبرّرة لأنه لا يهدف إلى تحقيق مستويات أكبر من المساواة أو لدعم أفراد مجموعة ما تعاني مسبقا من الحرمان. ومن المهمّ ملاحظة أنه ليست كلّ أشكال التمييز غير عادلة، حيث يمكن استخدام بعض أشكال التمييز لتعزيز مستويات أعلى من المساواة (مثل ضمان معاملة تفضيلية للمتقدمات في برامج العلوم والتكنولوجيا والهندسة والرياضيات).

في حين يمكن توظيف أشكال أخرى لإدامة أشكال معينة من الضرر (مثل مراقبة المعارضين السياسيين).

إنّ الحق في عدم التعرّض للتّمييز غير العادل مضمون بموجب الحق في المساواة، وهو مبدأ أساسي منصوص عليه في العديد من الصكوك الدولية المتعلقة بحقوق الإنسان والمعايير القانونية. فالمادة 2 من الإعلان العالمي لحقوق الإنسان لسنة 1948، تنصّ على أنه "لكل إنسان حقُّ التمتُّع بجميع الحقوق والحرِّيات المذكورة في هذا الإعلان، دونما تمييز من أيِّ نوع، ولا سيما التمييز بسبب العنصر، أو اللون، أو الجنس، أو اللغة، أو الدِّين، أو الرأي سياسيًّا كان أوغير سياسيّ، أو الأصل الوطنيّ، أو الاجتماعي، أو الثروة، أو المولد، أو أيِّ وضع آخر". كما تشمل الصكوك الأساسيّة أيضًا الاتفاقـيّة الدوليّة للقضاء على جميع أشكال التمييز العنصري التي تمنح جميع البشر الحقّ في الحماية المتساوية ضد أي تمييز وضد أي تحريض على التمييز (المادة 1)، والعهد الدولي الخاص بالحقوق المدنية والسياسية، الذي يدعو إلى التمتّع، بشكل كامل ومتساوٍ، بجميع الحقوق المدنية والسياسية، بما في ذلك الحق في عدم التمييز (المادة 26)؛ والعهد الدولي الخاص بالحقوق الاقتصادية والاجتماعية والثقافية، الذي ينصّ على الحق في المساواة في العمل (المادة 6)، والتعليم (المادة 13)، وحق الحصول على الخدمات الأساسية والبنية التحتية والإدماج الرقمي والتمتّع بفوائد التقدّم العلمي وتطبيقاته مثل الإدماج الرقمي (المادة 15). وبالإضافة إلى ما سبق، توضّح الأطر القانونية الوطنية مفهوم الحق في عدم التمييز في كل بلد؛ ما هي تعريفات/ مفاهيم المساواة الموجودة في البلد وما هي سبل الانتصاف المتاحة للأفراد في حالة انتهاك حقوقهم.

الأسس (النظرية)

لقد أحرزت بعض البلدان تقدمًا كبيرًا نحو حماية الحق في المساواة وعدم التمييز، إلا أنّ تأثير التحيّزات المتجذرة في نظم الذكاء الاصطناعي يمكن أن يعرقل هذا التقدّم، ممّا يعرض الأشخاص المهمشين لخطر المزيد من الاستبعاد من مختلف جوانب الحياة الاقتصادية والاجتماعية والسياسية. وقد خلصت العديد من

الدراسات والتقارير إلى أن الخوارزميات المدرّبة على مجموعة من البيانات المتحيّزة، أو تلك المصمّمة بطريقة غير شاملة، يمكن أن تؤدي إلى نتائج تمييزيّة. فقد أفاد مثلا، مشروع GenderShades الذي أنجزه مختبر الوسائط - Media Lab - بمعهد ماساتشوستس للتكنولوجيا سنة 2018، بوجود نسبة خطأ تبلغ 34.7% في مجال التعرف على النساء ذوات البشرة الداكنة، مقارنة بنسبة 0.8% للرجال ذوي البشرة الفاتحة، وذلك ضمن نتائج ثلاثة أنظمة مختلفة للتعرف على الوجه مدعومة بالذكاء الاصطناعي. لم يساهم هذا البحث فقط في الكشف عن تحيّزات متجذرة في أنظمة التعرف على الوجه على أساس العرق والجنس، بل سلّط الضوء أيضًا على حقيقة أن العدالة الخوارزمية تعتمد على عدد من العوامل التي يحددها السياق.

قد لا يكون من الممكن القضاء على التحيز ضمن أنظمة الذكاء الاصطناعي، ولكن يمكن التخفيف من حدّة الضرر الناتج عنه من خلال معرفة: (1) مصادر التحيز وتجلياته، و(2) مَن المتضرر ومن المستفيد من الممارسات المتحيزة، و(3) كيف يظهر التحيز على مستوى البيانات و(4) كيف يمكن للتحيز أن يزيد من حدّة الممارسات التمييزية. إن الإلمام بهذه المعطيات سيسمح بفهمٍ أفضلَ للتدابير التي يمكن للحكومات اتخاذها للاستفادة من إمكانيات الذكاء الاصطناعي مع التخفيف في الوقت نفسه من حدّة المخاطر، من أجل ضمان الاستخدام المسؤول والأخلاقي للذكاء الاصطناعي.

ومن ذلك مثلا، أن تشمل التدابير التي تتخذها الشركات والمنظمات إعداد تعريفات واضحة وأنظمة تقييم للعدالة الخوارزمية، مثل استخدام تقنية العدالة المغايرة للواقع، ووضع أنظمة للتخفيف من حدة التحيّز، مثل تلك التي أوصت بها Google A؛ والاستثمار في البحوث المتعلقة بالتحيز، وتنويع نطاق الذكاء الاصطناعي. ومن ناحية أخرى، سلّطت الأدبيّات الحديثة الضوء على قدرة خوارزميات الذكاء الاصطناعي على مواجهة التحيزات الاجتماعية لصالح مجموعات طالما كانت مهمشة تقليدياً، وتحسين عمليات صنع القرار، وخاصة في مكان العمل، التي كان من الممكن أن تظلّ أكثر عرضة للتحيز لولا هذه الخوارزميّات. وبما أنّ الشمولية، وعدم التمييز في تطوير تقنيات الذكاء الاصطناعي يعتبران أمران أساسيّان لضمان أخلاقيّات هذا الأخير، فإنّ تقييم هذا الإطار المفاهيمي أمر ضروريّ لدراسة الذكاء الاصطناعي المسؤول

تعريفـــات

يقيّم هذا الإطار المفاهيمي التدابير التي اتخذتها الدول للوقاية والحدّ من خطر التمييز الناجم عن التحيز أثناء صياغة نظم الذكاء الاصطناعي وآلياته، وتطويرها واستخدامها. ويجب أن يأخذ التحليل بعين الاعتبار، على وجه الخصوص، (1) **الأطر القانونية** المتعلّقة بأنظمة الذكاء الاصطناعي، و(2) **الإجراءات الحكومية** لتنفيذ تلك الأطر أو معالجة الموضوع، و(3) **الجهات الفاعلة غير الحكومية** التي تعمل في هذا المجال في بلد ما.

يمكن للأطر القانونية في دولة ما، بما في ذلك الأطر الموجودة مسبقا أو تلك التي تم اقتراحها، أن تكون على شكل قوانين ولوائح وسياسات (بما في ذلك السياسات القطاعية و/أو سياسة أقسام خاصة) و/أو مشاريع قوانين و/أو توجيهات. بينما يمكن أن تشمل الإجراءات الحكومية إنشاء هيئات حكومية، بما في ذلك هيئات رقابة مسؤولة عن صياغة التوصيات السياسية بشأن هذه القضية و/أو تطبيق اللوائح التنظيمية، بالإضافة إلى تنفيذ برامج تهدف لمعالجة هذه الإشكالية و/أو التوعية أو جمع المزيد من البيانات حولها. من جهتها، يمكن أن للجهات الفاعلة غير الحكومية أن تكون منظمات غير حكومية، ولكن أيضًا شركات متعددة الجنسيات، أو منظمات عسكرية خاصة، أو وسائل اتصال، أو مجموعات عرقية منظمة، أو مؤسسات أكاديمية، أو مجموعات ضغط، أو نقابات عمالية، أو حركات اجتماعية

أمثلة:

الأطر القانونية

تدعو استراتيجية البيانات الوطنية في المملكة المتحدة، المنظماتَ الخاصة إلى مكافحة "التحيزات التي تنتج عن استخدام البيانات أو الخوارزميات" على وجه التحديد، وذلك بناءً على ما يحدده التقرير المؤقت لمركز أخلاقيات البيانات والابتكار (CDEI).

إجراءات حكومية

أسّست حكومة المملكة المتحدة سنة 2018، مركز CDEI لتقديم توصيات سياسية للاستخدام الأخلاقي للذكاء الاصطناعي والتقنيات الأخرى المعتمدة على البيانات. وفي السنة نفسها، كلفت المركز بإجراء مراجعتين للسياسات، بما في ذلك السياسات الخاصة بالتحيز الخوارزمي. ووقد قام مركز CDEI بموجب خطّة عمله، بالتحقيق في التحيز الخوارزمي في أربعة قطاعات، شملت الشرطة، والحكومة المحلية، والخدمات المالية، والتوظيف. وأطلق المركز في يونيو 2022، برنامجًا جديدًا استجابة لهذه النتائج، يركّز بشكل خاص على "استكشاف الإمكانيات التي تتيحها الأشكال الجديدة لإدارة البيانات لمساعدة المؤسسات في الوصول إلى البيانات الديموغرافية من أجل التحكم في منتجاتها وخدماتها فيما يتعلق بالتحيزات الخوارزمية".

جهات فاعلة غير حكومية

على مستوى القطاع غير الحكومي، تقرّ "رابطة العدالة الخوارزمية" (Algorithmic Justice League) ومقرها الولايات المتحدة، بقدرة الذكاء الاصطناعي على إدامة أشكال مختلفة من التمييز الناتج عن التحيّز وتعمل على رفع مستوى الوعي العام حول هذه المسألة من خلال استخدام مزيج من الفن

والبحث العلمي، وكذلك "لتزويد الأشخاص المناصرين لهذه القضية بموارد ضرورية لتعزيز الحملات، وتعزيز صوت المجتمعات الأكثر تضرراً واختياراتها، وتحفيز الباحثين وصانعي السياسات والمتخصّصين في هذا المجال على تفادي الأضرار الناجمة عن الذكاء الاصطناعي". وقد أصدرت المنظّمة، بالتوازي مع أبحاثها المستمرة، فيلمًا وثائقيًا بعنوان "التحيز المشفّر"، كما نظّمت ورشة عمل حول "Drag vs. AI" ، وأشرفت على مشروع "Algorithmic Community Reporting of System Harms" (الإبلاغ المجتمعي عن أضرار النظام الخوارزمي) بمشاركة العديد من الأطراف المعنية بهذه المسألة.