

Distr.: General
29 August 2018
Arabic
Original: English

الجمعية العامة



الدورة الثالثة والسبعون

البند ٧٤ (ب) من جدول الأعمال المؤقت**

تعزيز حقوق الإنسان وحمايتها: مسائل حقوق
الإنسان، بما في ذلك النهج البديلة لتحسين التمتع
الفعلي بحقوق الإنسان والحريات الأساسية

تعزيز وحماية الحق في حرية الرأي والتعبير***

مذكرة من الأمين العام

يتشرف الأمين العام بأن يحيل إلى الجمعية العامة التقرير الذي أعده المقرر الخاص المعني بتعزيز
وحماية الحق في حرية الرأي والتعبير، دافيد كاي، المقدم وفقاً لقرار مجلس حقوق الإنسان ١٨/٣٤. وفي
هذا التقرير، يستكشف المقرر الخاص الآثار المترتبة على تكنولوجيات الذكاء الاصطناعي في حقوق
الإنسان في بيئة المعلومات، مع التركيز بوجه خاص على الحقوق في حرية الرأي والتعبير، والخصوصية،
وعدم التمييز.

* أعيد إصدارها لأسباب فنية في ٢٦ تشرين الأول/أكتوبر ٢٠١٨.

** A/73/150.

*** قُدم هذا التقرير بعد انقضاء الموعد النهائي، وذلك لتضمينه أحدث المستجدات.



تقرير المقرر الخاص المعني بتعزيز وحماية الحق في حرية الرأي والتعبير

المحتويات

الصفحة

| | | |
|----|-------|---|
| ٣ | | أولاً - مقدمة |
| ٣ | | ثانياً - فهم الذكاء الاصطناعي |
| ٣ | | ألف - ما هو الذكاء الاصطناعي؟ |
| ٧ | | باء - الذكاء الاصطناعي وبيئة المعلومات |
| ١١ | | ثالثاً - الإطار القانوني لحقوق الإنسان المتعلق بالذكاء الاصطناعي |
| ١١ | | ألف - نطاق الالتزامات المتعلقة بحقوق الإنسان في سياق الذكاء الاصطناعي |
| ١٢ | | باء - الحق في حرية الرأي |
| ١٤ | | جيم - الحق في حرية التعبير |
| ١٥ | | دال - الحق في الخصوصية |
| ١٦ | | هاء - الالتزام بعدم التمييز |
| ١٧ | | واو - الحق في الحصول على سبيل انتصاف فعال |
| ١٨ | | زاي - الردود التشريعية والتنظيمية والسياساتية على الذكاء الاصطناعي |
| ٢٠ | | رابعاً - اتباع نهج قائم على حقوق الإنسان في الذكاء الاصطناعي |
| ٢٠ | | ألف - المعايير الموضوعية لنظم الذكاء الاصطناعي |
| ٢٢ | | باء - عمليات لنظم الذكاء الاصطناعي |
| ٢٤ | | خامساً - الاستنتاجات والتوصيات |

أولاً - مقدمة

١ - يؤثر الذكاء الاصطناعي بصورة متزايدة على بيئة المعلومات في جميع أنحاء العالم. فهو يُمكن الشركات من تنظيم نتائج البحث والموجزات الإخبارية على وسائل التواصل الاجتماعي، ونشر الإعلانات فيها، وتنظيم ما يشاهده المستخدمون على منصاتها الإلكترونية، ومتى يشاهدوه. وتستخدم شركات وسائل التواصل الاجتماعي تكنولوجيات الذكاء الاصطناعي للمساعدة في تنقيح المحتوى المنشور على منصاتها الإلكترونية، عاملةً في كثير من الأحيان بمثابة خط الدفاع الأول لصد المحتوى الذي قد ينتهك قواعدها. ويوصي الذكاء الاصطناعي بشخص ما لمصادقته أو متابعة أخباره على منصات التواصل الاجتماعي، وبمقالات إخبارية لقراءتها، وبأماكن لزيارتها أو لتناول الطعام فيها أو للشراء من متاجرها أو للنوم في فنادقها. ويتيح السرعة والكفاءة والحجم، ويعمل على مساعدة الشركات الكبرى في قطاع تكنولوجيا المعلومات والاتصالات في إدارة الكم الهائل من المحتوى المحتمل على منصاتها الإلكترونية كل يوم. وتُمكن تكنولوجيات الذكاء الاصطناعي من تبادل المعلومات والأفكار على الصعيد العالمي على نحو أسرع وأوسع نطاقاً، مما يتيح فرصة هائلة لحرية التعبير والحصول على المعلومات. وفي الوقت نفسه، فإن غموض الذكاء الاصطناعي ينطوي أيضاً على مخاطر التدخل بتقرير مصير الأفراد، أو ما يشار إليه في هذا التقرير "الاستقلال الذاتي الفردي والقدرة الفردية"^(١). ويُواجه كل من يعمل على تعزيز حقوق الإنسان وسيادة القانون تحدياً عالمياً كبيراً وهو: كيف يمكن للدول والشركات والمجتمع المدني ضمان أن تعزز تكنولوجيات الذكاء الاصطناعي احترام حقوق الإنسان، بدلا من تقويضها وتعريضها للخطر؟

٢ - ولا يدعي هذا التقرير أنه يمثل الكلمة الأخيرة في مجال الذكاء الاصطناعي وحقوق الإنسان. بدلا من ذلك، فإنه يسعى إلى القيام بثلاثة أمور: تعريف المصطلحات الأساسية الضرورية لمناقشة حقوق الإنسان في سياق الذكاء الاصطناعي؛ وتحديد الإطار القانوني لحقوق الإنسان ذي الصلة بالذكاء الاصطناعي، وتقديم بعض التوصيات الأولية، في ضوء تطور التكنولوجيات التي تشمل الذكاء الاصطناعي، لكفالة إدماج اعتبارات حقوق الإنسان في صلب تلك العملية. وينبغي قراءة هذا التقرير بالاقتراح مع تقريره الأخير إلى مجلس حقوق الإنسان (A/HRC/38/35)، الذي يعرض نمحاً قائماً على حقوق الإنسان لتنقيح محتوى شبكة الإنترنت^(٢).

ثانياً - فهم الذكاء الاصطناعي

ألف - ما هو الذكاء الاصطناعي؟

٣ - كثيراً ما تستخدم عبارة الذكاء الاصطناعي بمثابة اختزال لزيادة الاستقلالية والسرعة والحجم المرتبطة بعملية اتخاذ القرار بواسطة التشغيل الآلي والحواسيب. والذكاء الاصطناعي ليس شيئاً واحداً فقط، بل يشير إلى "كوكبة" من العمليات والتكنولوجيات التي تُمكن الحواسيب من أن تُكمل أو تحلّ

(١) انظر Floridi Luciano و Mariarosaria Taddeo، "كيف يمكن لمنظمة 'الذكاء الاصطناعي الآن AI Now' أن تكون قوة من أجل الخير"، مجلة العلوم Science، المجلد رقم ٣٦١، العدد ٦٤٠٤ (٢٤ آب/أغسطس ٢٠١٨) متاح على الرابط التالي: <http://science.sciencemag.org/content/361/6404/751.full>.

(٢) استفاد هذا التقرير من مشاوره خبراء عقدت في جنيف في حزيران/يونيه ٢٠١٨، بمنحة من الاتحاد الأوروبي، واستفاد أيضاً من إسهامات من خبراء كانت قد قدمت في إطار إعداد الوثيقة A/HRC/35/38 في عامي ٢٠١٧ و ٢٠١٨. ويود المقرر الخاص بصفة خاصة أن يشكر كارلي نيسست وأموس توه، اللذين ساهما في أعمال البحث والصياغة الأساسية لهذا المشروع.

حل المهام المحددة التي بخلاف ذلك يؤديها البشر، مثل اتخاذ القرارات وحلّ المشاكل^(٣). ويمكن أن يكون مصطلح "الذكاء الاصطناعي" إشكالياً، لأنه يشير إلى أن الآلات يمكن أن تعمل وفقاً لنفس مفاهيم وقواعد الذكاء البشري. والواقع أنّها لا تستطيع أن تعمل وفقاً لذلك. فالذكاء الاصطناعي، عموماً، يعزز عمل المهام المحوسبة التي يعيّنّها البشر من خلال معاودة التكرار والمحاولة. ومع ذلك، فهو لغة الثقافة والشركات والحكومات هو مستخدم هنا.

٤ - وكثيراً ما تشير الثقافة الشعبية إلى أن المجتمع يتّجه صوب الذكاء الاصطناعي العام، وهو قدرة لا تزال بعيدة ("الوحدانية") بأن يقترب النظام الحاسوبي من الذكاء البشري أو يتجاوزه في مجالات متعددة^(٤). وفي المستقبل المنظور، ستظل هناك تطورات فيما يخص الذكاء الاصطناعي الضيق، التي تفيد بأن النظم الحاسوبية ستؤدي المهام المبرجة (رموز حاسوبية - خوارزميات - من وضع الإنسان) في مجالات محددة. فعلى سبيل المثال، يدعم الذكاء الاصطناعي الضيق المساعدة الصوتية في الأجهزة الالكترونية المحمولة، وخدمة محادثات العملاء عبر الإنترنت، وأدوات الترجمة المتاحة على الإنترنت، والسيارات ذاتية القيادة، ونتائج محركات البحث، وخدمات رسم الخرائط. ويعتبر التعلّم الآلي فئة من تقنيات الذكاء الاصطناعي الضيق المستخدمة في تدريب خوارزميات الحاسوب على استخدام مجموعات البيانات للتعرف على المسائل الحسابية والمساعدة في حلها. فعلى سبيل المثال، "تتعلم" الأجهزة الذكية المستخدمة في المنازل التي تعمل بطاقة الذكاء الاصطناعي باستمرار من البيانات التي تجمعها عن أنماط الكلام اليومي واللغة اليومية التي تسمعها من أجل تجهيزها والرد على الأسئلة المطروحة عليها من مستخدميها من الذكاء الاصطناعي. وفي جميع الظروف يؤدي البشر دوراً حاسماً في تصميم ونشر نظم الذكاء الاصطناعي، وتحديد أهداف تطبيقات الذكاء الاصطناعي، والقيام، تبعاً لنوع التطبيق، باختيار ووسم مجموعات البيانات وتصنيف النواتج. ويحدد البشر دائماً التطبيق واستخدام نواتج الذكاء الاصطناعي، ويشمل ذلك تحديد المدى الذي تُكمل فيه عمل البشر أو تحل محل البشر في عملية اتخاذ القرارات.

| المستقبل؟ | الحاضر | |
|---|---|---|
|  |  |  |
| الذكاء الاصطناعي العام: نظم الحاسوب التي تقارب الذكاء البشري في مجالات متعددة | الذكاء الاصطناعي الضيق: نظم خوارزميات الحاسوب المصممة من قبل البشر التي تحلل البيانات وتضع الحلول في مجالات محددة | خوارزميات الحاسوب: رموز حاسوبية تترجم البيانات إلى نواتج واستنتاجات ومعلومات |

(٣) انظر منظمة "الذكاء الاصطناعي الآن AI Now"، تقرير منظمة الذكاء الاصطناعي: الآثار الاجتماعية والاقتصادية المترتبة على تكنولوجيات الذكاء الاصطناعي على الأجل القريب، ٢٠١٦. متاح على الرابط التالي: https://ainowinstitute.org/AI_Now_2016_Report.pdf؛ اللجنة المختارة المعنية بالذكاء الاصطناعي التابعة لمجلس اللوردات في المملكة المتحدة لبريطانيا العظمى وأيرلندا الشمالية، "AI in the United Kingdom: ready, willing and able?", 2018, p.13.

(٤) المادة ١٩ "والمنظمة الدولية لحماية الخصوصية"، "Privacy and freedom of expression in an age of artificial intelligence", London, 2018, p. 8.

٥ - ويستند الذكاء الاصطناعي إلى قاعدة أساسها خوارزميات الحاسوب، وهي رموز حاسوبية يصممها ويكتبها البشر، وتنطوي على تعليمات بترجمة البيانات إلى استنتاجات أو معلومات أو نواتج. وما برحت خوارزميات الحاسوب منذ وقت طويل ضرورية لتشغيل نظم الاتصالات والهياكل الأساسية يومياً. ويمثل الحجم الهائل من البيانات المتاحة في الحياة العصرية والقدرة على تحليلها الوقود الذي يغذي الذكاء الاصطناعي. وينظر القطاع الخاص بالتأكيد إلى البيانات بتلك الطريقة: أي كلما ازداد حجم البيانات المتاحة لتغذية خوارزميات الحاسوب وتحسين نوعية تلك البيانات، لأصبحت خوارزميات الحاسوب أكثر قوة ودقة. وبإمكان نظم خوارزميات الحاسوب تحليل الكميات الضخمة من البيانات بسرعة، مما يُمكن برامج الذكاء الاصطناعي من أداء وظائف اتخاذ القرارات التي كانت في السابق مجالاً للبشر يتخذونها دون أدوات حاسوبية.

في مرحلة التصميم

يشترى البشر نظم الذكاء الاصطناعي ويحددون المراد منها.



في مرحلة النشر

يقوم البشر بتنظيف بيانات المدخلات ويضعون الوسوم عليها.



يحدد البشر المدخلات ويصممون خوارزميات الحاسوب.



في بعض الحالات، يحدد البشر نواتج نظم الذكاء الاصطناعي.



تحدد نظم الذكاء الاصطناعي النموذج المستخدم وتتعلم خوارزميات الحاسوب وتتكيف بصورة مستقلة.



في مرحلة التنفيذ

يقرر البشر كيفية تطبيق النواتج واستخدامها.



القدرة البشرية جزء لا يتجزأ من الذكاء الاصطناعي، ولكن الخصائص المميزة للذكاء الاصطناعي تستحق التمييز من منظور حقوق الإنسان فيما يتعلق بثلاثة على الأقل من جوانبه، هي: التشغيل الآلي، وتحليل البيانات، والقابلية للتكيف^(٥).

٦ - **التشغيل الآلي** - التشغيل الآلي يزيل التدخل البشري من أجزاء من عملية اتخاذ القرار، وينجز مهام محددة بأدوات حاسوبية. ويمكن أن تكون لهذا الأمر آثار إيجابية من منظور حقوق الإنسان إذا كان التصميم يحدّ من تحيز البشر. فعلى سبيل المثال، قد يُقرّر نظام دخول الحدود الآلي أفراداً للتدقيق بوضعهم على أساس السمات الموضوعية مثل التاريخ الجنائي أو وضع التأشيرة التي يحملونها، مما يحدّ من الاعتماد على التقييمات الذاتية (المعرضة للتحيز) لمظهرهم الخارجي أو الأصل العرقي أو السن أو الدين. كما يُمكن التشغيل الآلي من تجهيز كميات هائلة من البيانات بسرعة ونطاق لا يمكن للبشر تحقيقهما، مما يمكن أن يخدم السلامة العامة والصحة والأمن الوطني. بيد أن نظم التشغيل الآلي تعتمد على مجموعات بيانات يمكن أن تتيح، في مرحلة التصميم أو التنفيذ، إمكانية التحيز، وتتسبب بالتالي بإحداث آثار تمييزية. وعلى سبيل المثال، قد ينطوي التاريخ الجنائي نفسه أو بيانات التأشيرة نفسها المشار إليهما أعلاه على تحيز. وقد يقوض الاعتماد المفرط على القرارات المتخذة بواسطة التشغيل الآلي والثقة المفرطة به وعدم الاعتراف بهذه النقطة الأساسية، كل منها بدوره، التدقيق بنتائج الذكاء الاصطناعي ويشلّ قدرة الأفراد على الوصول إلى سبل الانتصاف من قرارات سلبية متخذة بواسطة الذكاء الاصطناعي. وقد يؤدي التشغيل الآلي إلى إعاقة شفافية العملية وقابليتها للتدقيق، مما يمنع حتى السلطات ذات النية الحسنة من تقديم توضيح للنتائج^(٦).

٧ - **تحليل البيانات** - تدعم مجموعات البيانات الضخمة معظم تطبيقات الذكاء الاصطناعي. ويمكن أن تشكل أي مجموعة بيانات الأساس لنظام الذكاء الاصطناعي، بدءاً من عادات التصفح على شبكة الإنترنت وانتهاءً ببيانات عن حركة المرور على الطرق السريعة. وتتضمن بعض مجموعات البيانات بيانات شخصية، في حين يشمل العديد من مجموعات البيانات الأخرى بيانات غفل لا تحدّد هوية أصحابها. ويثير استخدام الذكاء الاصطناعي لهذه المجموعات من البيانات شواغل خطيرة، بما في ذلك مصدرها ودقتها وحقوق الأفراد في التصرف فيها؛ وقدرة نظم الذكاء الاصطناعي على كشف هوية أصحاب البيانات المجهولة المصدر؛ والتحيز الذي يحتمل أن يكون متأصلاً في صميم مجموعات البيانات أو المغروس فيها من خلال التدريب البشري أو عملية وسم البيانات. وقد يبين تقييم الذكاء الاصطناعي للبيانات وجود روابط ولكنها ليست بالضرورة روابط سببية، الأمر الذي قد يؤدي إلى نتائج متحيزة ومشوبة بالعيوب يصعب تدقيقها.

٨ - **القابلية للتكيف** - نظم الذكاء الاصطناعي للتعليم الآلي قابلة للتكيف بسبب قدرة خوارزميات الحاسوب التي تغذيها على القيام تدريجياً بتحديد المشاكل الجديدة وتطوير إجابات جديدة عليها. وقد تحدد النظم، حسب مستوى الإشراف، الأنماط وتضع الاستنتاجات التي لم يتوقعها البشر الذين

(٥) Council of Europe, *Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications*, Council of Europe study, No. DGI (2017) 12, 2018. Available at <https://www.coe.int/en/web/freedom-expression/-/algorithms-and-human-rights-a-new-study-has-been-published>, p. 5

(٦) Council of Europe, *Algorithms and Human Rights*, p. 8

برمجوها أو كلفوها بمهام. وهذا الافتقار إلى القدرة على التنبؤ يحمل في طياته الوعد الحقيقي بأن يكون الذكاء الاصطناعي بمثابة تكنولوجيا تحويلية، ولكنه يلقي الضوء أيضا على مخاطره، وهي: كلما تم استبعاد البشر تدريجيا عن تحديد أهداف ونواتج نظم الذكاء الاصطناعي، لأصبح ضمان الشفافية والمساءلة والوصول إلى سبل الانتصاف الفعالة أكثر صعوبة، وكذلك الأمر بالنسبة لإمكانية التنبؤ بالآثار السلبية المترتبة على حقوق الإنسان والتخفيف منها.

باء - الذكاء الاصطناعي وبيئة المعلومات

٩ - الذكاء الاصطناعي له عواقب هامة بوجه خاص، وأحيانا إشكالية، على بيئة المعلومات، والنظم الإيكولوجية المعقدة للتكنولوجيات والمنصات الالكترونية والجهات الفاعلة الخاصة والعامة التي تيسر الوصول إلى المعلومات ونشرها من خلال الوسائل الرقمية. وتوجد خوارزميات الحاسوب وتطبيقات الذكاء الاصطناعي في كل ركن من أركان الإنترنت، وفي الأجهزة الرقمية والنظم التقنية، وفي محركات البحث، وفي منصات وسائل التواصل الاجتماعي وتطبيقات التراسل والآليات الإعلامية. وتمشيا مع تركيز الولاية، يشير المقرر الخاص أدناه إلى تطبيقات الذكاء الاصطناعي الثلاثة المثيرة للقلق في بيئة المعلومات.

١٠ - عرض المحتوى والسمات الشخصية - تسيطر وسائل التواصل الاجتماعي ومنصات البحث بشكل متزايد على الكيفية التي يمكن بها للأفراد الوصول إلى المعلومات والأفكار وتبادلها والكيفية التي يتم بها نشر الأخبار. وتُحدّد خوارزميات الحاسوب وتطبيقات الذكاء الاصطناعي مدى مشاركة المحتوى مع الجمهور والأفراد ومع أي منهم تتم مشاركة هذا المحتوى ومتى يتم ذلك. وتُعزدي مجموعات البيانات الضخمة التي تجمع بين تاريخ التصفح والخصائص الديمغرافية للمستخدم والتحليلات الدلالية وتحليلات المشاعر والعديد من العوامل الأخرى، بصورة متزايدة، نماذج خوارزميات الحاسوب الشخصية التي تُستخدم في تصنيف وتنظيم المعلومات، أي إظهار المعلومات للأفراد أو استبعادها ضمناً. وقد يتم الترويج للمحتوى المدفوع الأجر أو المكفول أو المرمّز (هاشتاغ) بغرض استبعاد محتوى آخر أو خفض رتبته. وتعرض صفحة الموجزات الإخبارية على وسائل التواصل الاجتماعي المحتوى حسب التقييمات الذاتية لكيفية عرض محتوى مثير للاهتمام أو جذاب للمستخدم؛ ونتيجة لذلك، قد يتم عرض قليل من القصص الاجتماعية والسياسية الهامة على الأفراد أو عدم اطلاعهم عليها وعلى المحتوى المنشور على منصاتهم^(٧). ويرسم الذكاء الاصطناعي صورة عالم المعلومات بطريقة مبهمة للمستخدم بل ومبهمة في كثير من الأحيان حتى للمنصة الالكترونية التي تعالجها.

١١ - البحث على شبكة الإنترنت هو أحد أكثر الأشكال انتشاراً لعرض المحتوى والسمات الشخصية المعدّين بتقنيات الذكاء الاصطناعي. وتوفر محركات البحث نتائج للاستفسارات (وإكمال الاستفسارات أو التنبؤ بها) باستخدام نظم الذكاء الاصطناعي التي تجهز بيانات مستفيضة عن المستخدمين فرادى وجماعات. ونظراً لأنه من غير المحتمل رؤية المحتوى الرديء التصنيف أو المحتوى الذي تم استبعاده تماماً من نتائج البحث، فإن تطبيقات الذكاء الاصطناعي للبحث لها تأثير هائل على نشر المعارف^(٨). وتقوم

(٧) مؤسسة الشبكة العالمية، "The invisible curation of content: Facebook's News Feed and our information diets"،

٢٢ نيسان/أبريل ٢٠١٨. متاح على الرابط التالي: [https://webfoundation.org/research/the-invisible-curation-of-](https://webfoundation.org/research/the-invisible-curation-of-content-facebooks-news-feed-and-our-information-diets/)

[content-facebooks-news-feed-and-our-information-diets/](https://webfoundation.org/research/the-invisible-curation-of-content-facebooks-news-feed-and-our-information-diets/)

(٨) Council of Europe, *Algorithms and Human Rights*, p. 17

الجهات المعنية بتجميع المحتوى والمواقع الإخبارية بالمثل باختيار المعلومات التي تتيحها للفرد لا على أساس التطورات الهامة الأخيرة بل على أساس تطبيقات الذكاء الاصطناعي التي تتوقع اهتمامات المستخدمين وأنماط الإخبار استناداً إلى مجموعات البيانات الواسعة النطاق^(٩). وبناء على ذلك، يلعب الذكاء الاصطناعي دوراً كبيراً ولكنه دائماً ما يكون مخفياً في تشكيل المعلومات التي يستهلكها الأفراد أو حتى المعلومات التي يعرفون أنهم يستهلكونها.

١٢ - ويدفع الذكاء الاصطناعي في مجال عرض المحتوى نحو المزيد من إضفاء السمات الشخصية على تجربة كل فرد عبر شبكة الإنترنت؛ وفي عصر وفرة المعلومات، يُنذر إضفاء السمات الشخصية بيث الفوضى على شبكة الإنترنت، مما يتيح للأفراد العثور على المعلومات المطلوبة^(١٠). وقد تشمل المنافع القدرة على الحصول على المعلومات والخدمات بطائفة أوسع من اللغات أو المعلومات وفي الوقت المناسب وتكون ذات صلة بالتجربة الشخصية أو الأفضليات الشخصية^(١١). وقد يقلل الذكاء الاصطناعي القائم على السمات الشخصية أيضاً التعرض إلى أدنى حد من الآراء المتنوعة التي تندخل بالقدرة الفردية على التماس الأفكار وتبادل الآراء حول الانقسامات الأيديولوجية أو السياسية أو المجتمعية. وقد يؤدي إضفاء السمات الشخصية إلى تعزيز التحيزات وتخصيص الترويج للمحتوى التحريضي أو للمعلومات المضللة والتوصية بهما من أجل الحفاظ على التواصل مع المستخدمين عبر الإنترنت^(١٢). ومن المؤكد أن جميع أنواع الأوساط الاجتماعية والثقافية قد تحد من تعرض الفرد للمعلومات. ولكن من خلال تحقيق المستوى الأمثل للتواصل والتفاعل على نطاق واسع، قد يؤدي إضفاء السمات الشخصية المعد بمساعدة الذكاء الاصطناعي إلى تقويض خيار الفرد العثور على أنواع معينة من المحتوى. وهذا صحيح خصوصاً لأن خوارزميات الحاسوب تحفض عادة أولوية المحتوى إلى مستويات أدنى من التواصل، مما يؤدي إلى إبعاد المحتوى المستقل الذي ينشئه المستخدم إلى طيات النسيان^(١٣). ويمكن للجهات الفاعلة البارعة أن تستغل أنظمة الذكاء الاصطناعي القائمة على القواعد

(٩) انظر، على سبيل المثال MIT Technology Review “How Reuters’s revolutionary AI system gathers global news,” ٢٧ تشرين الثاني/نوفمبر ٢٠١٧. متاح على الرابط التالي: <https://www.technologyreview.com/s/609558/how-reuters-ai-system-gathers-global-news/>؛ بول أرمسترونغ، ويو وانغ “China’s \$11 billion news aggregator Jinri Toutiao is no fake,” *Forbes*، ٢٦ أيار/مايو ٢٠١٧. متاح على الرابط التالي: <https://www.forbes.com/sites/ywang/2017/05/26/jinri-toutiao-how-chinas-11-billion-news-aggregator-is-no-fake/#1d8b97804d8a>

(١٠) Carly Nyst and Nick Monaco, *State-Sponsored Trolling: How Governments are Deploying Disinformation as Part of Broader Digital Harassment Campaigns* (Palo Alto, Institute for the Future, 2018), p. 8

(١١) مؤسسة الشبكة العالمية World Wide Web Foundation، “الذكاء الاصطناعي: الطريق إلى الأمام في البلدان المنخفضة والمتوسطة الدخل”، واشنطن العاصمة، ٢٠١٧.

(١٢) Zeynep Tufekci, “YouTube, the great radicaliser,” *Nieman Journalism Center*, ١٠ آذار/مارس ٢٠١٨. متاح على الرابط التالي: <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>; James Williams, *Stand Out of our Light: Freedom and Resistance in the Attention Economy* (Cambridge, Cambridge University Press, 2018).

(١٣) في الآونة الأخيرة، أبدت بعض منصات التكنولوجيا عزمها على الابتعاد عن “المشاركة” - المبنية على السمات الشخصية والعمل بدلاً من ذلك بالسمات الشخصية التي تعطي الأولوية لنوعية تجربة المستخدم على شبكة الإنترنت؛ انظر *The Guardian*، ١١ كانون الثاني/يناير ٢٠١٨. متاح على الرابط التالي: <https://www.theguardian.com/technology/2018/jan/11/facebook-news-feed-algorithm-overhaul-mark-zuckerberg>

والتي تم تحسينها لأغراض التواصل من أجل الحصول على مستويات أعلى من التعرض، كما يمكنها من خلال الاستيلاء على علامات الوسم الشهيرة (هاشتاغ) أو استخدام برامج التتبع، تحقيق الوصول الواسع النطاق عبر شبكة الإنترنت على حساب تنوع المعلومات.

١٣ - **تنقيح المحتوى وإزالته** - يساعد الذكاء الاصطناعي شركات وسائل التواصل الاجتماعي على تنقيح المحتوى وفقاً للمعايير والقواعد المتعلقة بالمنصات الإلكترونية، بما في ذلك تكنولوجيا كشف الرسائل الإلكترونية التطفلية (spam) وتكنولوجيا مطابقة الرسائل المبعثرة (باستخدام البصمات الرقمية من أجل تحديد المحتوى الإرهابي أو محتوى استغلال الأطفال، على سبيل المثال)، ومرشحات الكلمات الرئيسية، ومعالجة اللغات الطبيعية (التي يجري بواسطتها تقييم طبيعة المحتوى لتحديد الكلمات أو الصور المحظورة) وغيرها من خوارزميات الكشف. ويمكن استخدام الذكاء الاصطناعي لإخضاع حسابات المستخدمين للتحذير أو تعليق الخدمة أو تعطيلها على أساس انتهاكاتهم لشروط الخدمة، أو يمكن أن يستخدم لعرقله أو لتنقية المواقع على أساس بيانات المجال أو المحتوى المحظور. وتستخدم شركات وسائل التواصل الاجتماعي الذكاء الاصطناعي لتنقية المحتوى عبر كامل نطاق قواعدها (بدءاً من التعري والتحرش وخطاب الكراهية وما إلى ذلك)، وإن كان من غير المعروف مدى اعتماد هذه الشركات على التشغيل الآلي من دون مدخلات بشرية بشأن حالات محددة^(١٤).

١٤ - يأتي الدعم والضغط الضروريين لزيادة دور الذكاء الاصطناعي من القطاعين الخاص والعام على السواء. وتدعي الشركات بأن حجم المحتوى غير القانوني والضار وغير الملائم الموجود على شبكة الإنترنت يتجاوز بكثير قدرات التنقيح البشرية وتحتاج بأن الذكاء الاصطناعي هو أداة يمكن أن تساعد في تحسين التصدي لهذا التحدي. وحسب بعض المنصات الإلكترونية، يُعدُّ الذكاء الاصطناعي أكثر كفاءة في تحديد المحتوى غير المناسب (حسب قواعدها) وغير القانوني الواجب إزالته (بفعل منقح بشري عادة) إضافة إلى أنه يتسم أيضاً بدقة أعلى من المعدل البشري في اتخاذ القرارات. وفي الوقت نفسه، تضغط الدول من أجل تطبيق التنقيح الآلي السريع والكفؤ على طائفة من التحديات المنفصلة، بدءاً من الاعتداء الجنسي على الأطفال والمحتوى الإرهابي، حيث يطبق الذكاء الاصطناعي بالفعل على نطاق واسع عليهما، وانهاءً بحقوق التأليف والنشر وإزالة المحتوى "المتطرف" و"خطاب الكراهية"^(١٥). وتُهيئ

بيد أنه من دون شفافية كاملة، وإعداد تقارير ومقاييس شاملة حول كيفية قيام نظم الذكاء الاصطناعي بإجراء مثل هذه التقييمات وتنفيذها، من الصعب تقييم ما إذا كان هذا التغيير له أثر واضح على تجربة مستخدمي الإنترنت.

(١٤) تحاول أداة في تطبيقات "إنستغرام" يطلق عليها اسم "النص العميق Deep Text"، البت بمدى "سمية" السياق، والسماح للمستخدمين بتكليف كلماتهم وتدوين الصور الرمزية "إيموجي emoji"، وتقييم علاقة المستخدمين أيضاً في محاولة أخرى لتحديد السياق (مثل ما إذا كان التعليق مجرد مزاح بين أصدقاء)، أندرو هاتشيسون، "إنستغرام ينشر أدوات جديدة لإزالة التعليقات المسببة"، وسائل التواصل الاجتماعي اليوم Social Media Today، ٣٠ حزيران/يونيه ٢٠١٧. متاح على الرابط التالي: <https://www.socialmediatoday.com/social-networks/instagrams-rolling-out-new-tools-remove-toxic-comments>.

(١٥) تفيد التقارير بأن المملكة المتحدة طورت أداة تكشف وتزيل تلقائياً المحتوى الإرهابي في نقطة التحميل. انظر، على سبيل المثال، وزارة الداخلية، "كُشف النقاب عن تكنولوجيا جديدة تساعد في مكافحة المحتوى الإرهابي على شبكة الإنترنت"، نشرة صحفية، ١٣ شباط/فبراير ٢٠١٨. انظر المفوضية الأوروبية، مقترح من أجل استصدار توجيه من البرلمان الأوروبي والمجلس بشأن حقوق التأليف في السوق الرقمية الموحدة، المفوضية (٢٠١٦)، المادة ١٣، الصيغة النهائية ٥٩٣، رسالة موجهة من المقرر الخاص إلى رئيس المفوضية الأوروبية، الرقم المرجعي OL OTH 41/2018، ١٣ حزيران/يونيه ٢٠١٨. متاحة على الرابط التالي: <https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-OTH-41-2018.pdf>.

المفوضية الأوروبية في توصيتها المتعلقة بالتدابير الرامية إلى زيادة تحسين فعالية مكافحة المحتوى غير القانوني عبر الإنترنت، الصادرة في آذار/مارس ٢٠١٨، بمنصات الإنترنت استخدام المرشحات الآلية لكشف المحتوى الإرهابي وإزالته، مع اقتراح بقيام البشر في بعض الحالات باستعراض المحتوى ليكون بمثابة قوة توازن ضرورية لتصويب الأخطاء الحتمية الناجمة عن نظم التشغيل الآلي^(١٦).

١٥ - وقد تؤدي الجهود الرامية إلى تنقيح المحتوى بالتشغيل الآلي ثمارها على حساب حقوق الإنسان (انظر A/HRC/38/35، الفقرة ٥٦). ويُقيّد تنقيح المحتوى المستند إلى الذكاء الاصطناعي عدّة قيود، بما في ذلك التحدي المتمثل في تقييم السياق ومراعاة التفاوت الواسع النطاق في التلميحات اللغوية، والمعاني والخصائص اللغوية والثقافية. ونظراً إلى أن تطبيقات الذكاء الاصطناعي غالباً ما تتجذر في مجموعات البيانات التي تتضمن افتراضات تمييزية^(١٧)، وفي ظل ظروف تكون فيها تكلفة الإفراط في التنقيح منخفضة، فإن هناك احتمالاً كبيراً بأن تفترض هذه النظم من لدنها إزالة المحتوى الإلكتروني أو تعليق الحسابات التي لا تثير مشاكل وبأن يتم حذف المحتوى استناداً إلى مفاهيم متحيزة أو تمييزية. ونتيجة لذلك، فإن الفئات الضعيفة هي الأكثر عرضة للحرمان بسبب نظم تنقيح محتوى الذكاء الاصطناعي. فعلى سبيل المثال، حدد برنامج "النص العميق DeepText" في تطبيق إنستغرام "المكسيكي" كوصمة لأن مجموعات البيانات لدى البرنامج مليئة ببيانات تربط "المكسيكي" بوضع "غير قانوني"، وهو مصطلح مشفر بصورة سلبية في خوارزميات الحاسوب^(١٨).

١٦ - الذكاء الاصطناعي يجعل من الصعب التدقيق في المنطق الكامن وراء أعمال المحتوى. وحتى عندما يُستكمل تنقيح محتوى خوارزميات الحاسوب باستعراض من قبل البشر - وهو ترتيب تُحاجج منصات وسائل التواصل الاجتماعي بأنه أضحى بشكل متزايد غير قابل للتطبيق على النطاق الذي تعمل فيه - فإن نزعة لاتخاذ القرارات بالاعتماد على الآلة (على أساس افتراضات الموضوعية المشار إليها أعلاه) تعوق استجواب نتائج تنقيح المحتوى، لا سيما عندما يحجب التصميم التقني للنظام هذا النوع من الشفافية.

١٧ - التمييز والإعلان والاستهداف - استفاد التقدم المحرز في مجال الذكاء الاصطناعي من نموذج الأعمال المستند إلى البيانات على الإنترنت وحفزه بشكل أكبر، على السواء، ومفاده أن الأفراد يقايضون

(١٦) توصية المفوضية المؤرخة ١ آذار/مارس ٢٠١٨ بشأن التدابير الرامية إلى التصدي بفعالية للمحتوى غير القانوني على الإنترنت ((C(2018) 1177 final)) متاحة على الرابط التالي: <https://ec.europa.eu/digital-single-market/en/news/commission-recommendation-measures-effectively-tackle-illegal-content-online>؛ انظر أيضاً: دافني كيلر، "تعليق رداً على توصية المفوضية الأوروبية الصادرة في آذار/مارس ٢٠١٨ بشأن التدابير الرامية إلى زيادة تحسين فعالية مكافحة المحتوى غير القانوني على الإنترنت"، كلية ستانفورد للقانون، مركز الإنترنت والمجتمع، ٢٩ آذار/مارس ٢٠١٨. متاح على الرابط التالي: <http://cyberlaw.stanford.edu/publications/comment-response-european-commissions-march-2018-recommendation-measures-further>.

(١٧) انظر Aylin Caliskan, Joanna Bryson and Arvind Narayanan, "Semantics derived automatically from language corpora contain human-like biases", *Science*, المجلد ٣٥٦ العدد ٦٣٣٤ (١٤ نيسان/أبريل ٢٠١٧)؛ وانظر أيضاً Solon Barocas and Andrew Selbst, "Big data's disparate impact", *California Law Review*, المجلد ١٠٤ العدد ٦٧١ (٢٠١٦).

(١٨) نيكولاس تومبسون، "Instagram's Kevin Systrom wants to clean up the @! Internet", *Wired*, ١٤ آب/أغسطس ٢٠١٧. متاح على الرابط التالي: <https://www.wired.com/2017/08/instagram-kevin-systrom-wants-to-clean-up-the-internet/>.

بياناتهم الشخصية بالحصول مجاناً على المحتوى والخدمات. ونظراً لتكدس موارد ضخمة من البيانات التي جمعت على مدى سنوات من الرصد والتنميط على شبكة الإنترنت فقد أصبح بمقدور الشركات تزويد نظم الذكاء الاصطناعي بمجموعات غنية من البيانات تمكنها من وضع نماذج التنبؤ والاستهداف بدقة أكبر من أي وقت مضى. وفي هذا الوقت، تستطيع الإعلانات التي تعدها الجهات الفاعلة الخاصة والعامّة أن تتحقق على المستوى الفردي؛ وأصبح المستهلكون والناخبون موضوع "الاستهداف المُحكّم" المصمم للاستجابة للخصائص الفردية واستغلالها.

١٨ - ويعمل الاستهداف المستند إلى الذكاء الاصطناعي على تحفيز جمع البيانات الشخصية واستغلالها على نطاق واسع، ويزيد من احتمال التحكم بفرادى المستخدمين من خلال نشر معلومات مضللة. والاستهداف يمكن أن يكرس التمييز، فضلاً عن استبعاد المستخدمين من الاستفادة من المعلومات أو الفرص من خلال السماح، على سبيل المثال، بنشر إعلانات عن فرص العمل والإسكان تستهدف استبعاد العمال المسنين، أو النساء أو الأقليات الإثنية^(١٩). وبدلاً من أن يتعرض الأفراد إلى التكافؤ والتنوع في الرسائل السياسية، على سبيل المثال، يؤدي نشر الاستهداف المُحكّم من خلال منصات التواصل الاجتماعي إلى خلق رؤية عالمية منظمة مناوئة لخطاب سياسي تعددي.

ثالثاً - الإطار القانوني لحقوق الإنسان المتعلق بالذكاء الاصطناعي

ألف - نطاق الالتزامات المتعلقة بحقوق الإنسان في سياق الذكاء الاصطناعي

١٩ - يجب أن يتم تصميم أدوات الذكاء الاصطناعي وتطويرها ونشرها، شأنها في ذلك شأن جميع التكنولوجيات، بحيث تكون متسقة مع التزامات الدول ومسؤوليات الجهات الفاعلة الخاصة بموجب القانون الدولي لحقوق الإنسان. ويفرض القانون الدولي لحقوق الإنسان على الدول التزامات سلبية بالامتناع عن تنفيذ التدابير التي تعوق ممارسة حرية الرأي والتعبير، والتزامات إيجابية بتعزيز الحق في حرية الرأي والتعبير وحماية ممارسته.

٢٠ - وفيما يتعلق بالقطاع الخاص، فإن الدول ملزمة بضمان احترام حقوق الأفراد، ولا سيما الحق في حرية الرأي والتعبير، بما في ذلك عن طريق حماية الأفراد من أعمال التعدي التي ترتكبها الأطراف الخاصة (المادة ٢ (١) من العهد الدولي الخاص بالحقوق المدنية والسياسية)^(٢٠) وتستطيع الدول الوفاء بهذا الالتزام من خلال اتخاذ تدابير قانونية ترمي إلى الحد من أو التأثير على وضع وتنفيذ تطبيقات الذكاء الاصطناعي، من خلال السياسات المتعلقة بشراء الجهات الفاعلة في القطاع العام تطبيقات الذكاء الاصطناعي من

(١٩) Julia Angwin و Noam Scheiber و Ariana Tobin، "عشرات الشركات تستخدم فيسبوك بغرض استبعاد العمال المسنين من إعلانات الوظائف"، ProPublica، ٢٠ كانون الأول/ديسمبر ٢٠١٧. متاح على الرابط التالي: <https://www.propublica.org/article/facebook-ads-age-discrimination-targeting>; Julia Angwin, Ariana Tobin, and Madeleine Varner, "Facebook (still) letting housing advertisers exclude users by race", ProPublica ٢١ تشرين الثاني/نوفمبر ٢٠١٧. متاح على الرابط التالي: <https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin>.

(٢٠) انظر المبدأ ٣ من المبادئ التوجيهية المتعلقة بالأعمال التجارية وحقوق الإنسان: تنفيذ إطار الأمم المتحدة المعنون "الحماية والاحترام والانتصاف" A/HRC/17/31 و A/HRC/38/35، الفقرات ٦-٨.

شركات القطاع الخاص من خلال برامج ذاتية أو مشتركة وبناء قدرات شركات القطاع الخاص على الاعتراف بالحق في حرية الرأي والتعبير في مساعيها العامة ومنحها الأولوية.

٢١ - وتقع على الشركات أيضا مسؤوليات بموجب القانون الدولي لحقوق الإنسان ينبغي أن تسترشد بها في بناء تكنولوجيات الذكاء الاصطناعي الخاصة بها وتعبئتها وتكييفها (A/HRC/38/35، الفقرة ١٠). وتوفر المبادئ التوجيهية المتعلقة بالأعمال التجارية وحقوق الإنسان: تنفيذ إطار الأمم المتحدة "الحماية والاحترام والانتصاف" "معيارا عالميا لقواعد السلوك المتوقع من جميع المؤسسات التجارية أيا كان مكان عملها" (المبدأ ١١)، بما في ذلك وسائل التواصل الاجتماعي وشركات البحث. ولتكييف الاستنتاجات المستخلصة من المبادئ التوجيهية مع مجال الذكاء الاصطناعي (المرجع نفسه، الفقرة ١١)، تدعو المبادئ التوجيهية الشركات إلى الاعلان، على الأقل، عن التزامات رفيعة المستوى بسياسات احترام حقوق الإنسان لمستخدمي تطبيقات الذكاء الاصطناعي (المبدأ ١٦)؛ وعدم التسبب أو الإسهام في إلحاق آثار ضارة بحقوق الإنسان من خلال استخدامها لتكنولوجيا الذكاء الاصطناعي، ومنع وقوع الآثار الضارة المرتبطة بعملياتها أو التخفيف منها (المبدأ ١٣)؛ وبذل العناية الواجبة بشأن نظم الذكاء الاصطناعي لتحديد ومعالجة الآثار الفعلية والمحتملة على حقوق الإنسان (المبادئ ١٧ إلى ١٩)؛ والمشاركة في استراتيجيات الوقاية والتخفيف (المبدأ ٢٤)؛ وإجراء استعراض مستمر للأنشطة ذات الصلة بالذكاء الاصطناعي، بسبل تشمل التشاور مع أصحاب المصلحة والجمهور (المبدأ ٢٠ و ٢١)، وتوفير سبل الانتصاف المتاحة لتصحيح الآثار الضارة بحقوق الإنسان الناجمة عن نظم الذكاء الاصطناعي (المبادئ ٢٢ و ٢٩، ٣١)

باء - الحق في حرية الرأي

٢٢ - حرية اعتناق الآراء دون أي تدخل هو حق مكرس في المادة ١٩ (١) من العهد الدولي الخاص بالحقوق المدنية والسياسية وفي المادة ١٩ من الإعلان العالمي لحقوق الإنسان. ولا يُسمح بأي استثناء له أو قيد عليه سواء "أكان ذلك بموجب القانون أو أي سلطة أخرى" (٢١). ولاحظ المقرر الخاص، في تقريره لعام ٢٠١٥ المقدم إلى مجلس حقوق الإنسان عن مسألة التشفير وإخفاء الهوية في الاتصالات الرقمية (A/HRC/29/32)، لاحظت أن السبل التي تُخزن المعلومات بواسطتها أو تُرسل أو تُؤمن في العصر الرقمي تؤثر بشكل فريد على ممارسة الحق في اعتناق الآراء. وتُشكل الأنشطة والسجلات الرقمية التي تشمل استفسارات البحث وأنشطة التصفح والبريد الإلكتروني والرسائل النصية والوثائق والتذكارات المحفوظة في السحابة - إذا ما أُخذت هذه كلها معاً، نسيج الآراء التي يعتنقها المستخدمون (المرجع نفسه، الفقرة ١٢). وقد تتداخل الجهات الفاعلة الحكومية وغير الحكومية على السواء مع هذه الآليات والعمليات لتشكيل واعتناق الآراء.

٢٣ - وثمة عنصر أساسي من عناصر الحق في اعتناق الرأي هو "الحق في تكوين رأي وتطويره عن طريق الاستدلال" (٢٢). وخُلصت اللجنة المعنية بحقوق الإنسان إلى أن هذا الحق يقتضي عدم تعرض

(٢١) اللجنة المعنية بحقوق الإنسان، التعليق العام رقم ٣٤ (٢٠١١) بشأن حرية الرأي والتعبير، الفقرة ٩، متاح على الرابط التالي: www2.ohchr.org/english/bodies/hrc/docs/GC34.pdf؛ مانفريد نواك: العهد الخاص بالحقوق المدنية والسياسية

CCPR commentary. (١٩٩٣).

(٢٢) نواك، العهد الخاص بالحقوق المدنية والسياسية.

الفرد لقسر لا موجب له في تطوير معتقداته وإيديولوجياته وردود فعله ومواقفه^(٢٣). وبناء على ذلك، فإن التدخلات القسرية العصبية، وبرامج التلقين (مثل "معسكرات إعادة التأهيل") أو التهديد بالعنف الرامي إلى إجبار الأفراد على تكوين آراء معينة أو تغيير آرائهم تنتهك أحكام المادة ١٩ (١) من العهد. كما خلّصت اللجنة إلى أن الطبيعة القسرية الكامنة في "الترويج بالمعاملة التفضيلية" قد يرقى إلى مستوى الإقناع الذي يتعارض مع الحق في تكوين واعتناق الآراء (انظر الوثيقة (CCPR/C/78/D/878/1999)

٢٤ - ويثير التداخل بين التكنولوجيا وتنظيم المحتوى أسئلة جديدة حول أنواع القسر أو الترويج التي يمكن اعتبارها تدخلا في الحق في تكوين الرأي. ولطالما أغنى تنظيم المحتوى قدرة الفرد على تكوين الآراء: فعلى سبيل المثال، ترفع المنافذ الإعلامية قصصاً معينة على صدر الصفحة الأولى بقصد تشكيل معرفة الفرد بالأخبار الهامة في اليوم والتأثير عليه. وتسعى الإعلانات التجارية أيضاً إلى غرس آراء موالية بشأن منتجات وخدمات خاصة والترويج بها.

٢٥ - ويمدد استخدام الذكاء الاصطناعي نطاق تقاليد تنظيم المحتوى على شبكة الإنترنت، ويعززها، ويوفر للمستخدم وسائل أكثر تطوراً وفعالية لإضفاء السمات الشخصية على المحتوى وتنظيمه على نطاق يتجاوز حدود وصول وسائل الإعلام التقليدية. وتثير هيمنة الطرائق الخاصة لتنظيم المحتوى بمساعدة الذكاء الاصطناعي قلقاً بشأن أثرها على قدرة الفرد على تكوين وتطوير الآراء. فعلى سبيل المثال، يستأثر عدد قليل من شركات التكنولوجيا بمعالجة الغالبية العظمى من استفسارات البحث التي تطرح عبر الإنترنت. واحتكار الشركات لسوق البحث يجعل من الصعوبة بمكان على المستخدمين الانسحاب من تصنيف خوارزميات الحاسوب وتنظيم نتائج البحث، وقد يدفع المستخدمين إلى الاعتقاد أيضاً (على نحو ما تقصده الشركات) بأن النتائج المتولدة عن البحث هي أهم المعلومات المتاحة بشأن موضوع محدد أو أكثرها موضوعية. وقد يعزز أيضاً الافتقار إلى الشفافية بشأن كيفية وضع وتنفيذ معايير البحث من خلال استخدام الذكاء الاصطناعي الافتراض القائل بأن نتائج البحث المتولدة في منصة إلكترونية معينة هي عرض موضوعي لمعلومات وقائعية.

٢٦ - ولذلك، فإن المسائل التي تثيرها مسألة الهيمنة على السوق في مجال تنظيم المحتوى بمساعدة الذكاء الاصطناعي تختبر أنماط الفهم التاريخية لكيفية تأثير أو عدم تأثير تنظيم المحتوى على القدرة على تكوين رأي. وتطرح المسائل الجديدة التي أثيرت، بالافتقار مع الافتقار عموماً إلى الاجتهاد القضائي بشأن التدخلات في الحق في الرأي، المزيد من الأسئلة أكثر من الأجوبة حول تأثير تنظيم المحتوى بمساعدة الذكاء الاصطناعي على حقوق الإنسان في البيئة الرقمية المعاصرة. غير أن هذه المسائل ينبغي أن تقود البحوث الموجهة نحو الحقوق إلى دراسة الآثار الاجتماعية والاقتصادية والسياسية لتنظيم المحتوى بمساعدة الذكاء الاصطناعي. وينبغي للشركات أن تقدم، على الأقل، معلومات مفيدة عن كيفية التي تضع وتنفذ بموجبها معايير تنظيم وإضفاء السمات الشخصية على المحتوى المنشور على منصات الإلكترونية، بما في ذلك السياسات والعمليات للكشف عن التحيزات الاجتماعية أو الثقافية أو السياسية الكامنة في تصميم وتطوير نظم الذكاء الاصطناعي ذات الصلة.

(٢٣) قضية "وو - كانغ ضد جمهورية كوريا"، اللجنة المعنية بحقوق الإنسان البلاغ رقم ١٩٩٩/٨٧٨، (CCPR/C/78/D/878/1999) ١٦ تموز/يوليه ٢٠٠٣.

جيم - الحق في حرية التعبير

٢٧ - تكفل المادة ١٩ (٢) من العهد حقاً موسعاً في "التماس وتلقي ونقل المعلومات والأفكار بجميع أنواعها"، التي يجب حمايتها واحترامها بغض النظر عن الحدود أو نوع وسائل الإعلام. ويعتبر التمتع بالحق في حرية التعبير أمراً وثيق الصلة بممارسة الحقوق الأساسية الأخرى للأداء الفعال للمؤسسات الديمقراطية، ومن ثم تنطوي مسائل حماية واحترام وتعزيز الحق في حرية التعبير على واجب إدراج تعزيز تنوع وسائل الإعلام واستقلالها وحماية إمكانية الحصول على المعلومات فيها^(٢٤).

٢٨ - وخلافاً للحق في تكوين واعتناق الآراء، يمكن أن تخضع الحقوق في التعبير والحصول على المعلومات والأفكار لقيود في حالات محدودة (المادة ١٩ (٣) من العهد). ويجب أن تستوفي القيود المفروضة معايير المشروعية، بمعنى أن يتم توفيرها بشكل علني بموجب قانون يستوفي معايير الوضوح والدقة وتفسر من قبل سلطات قضائية مستقلة؛ ومعياري الضرورة والتناسب، بمعنى أن تكون التدابير قليلة التدخل إلا ما تقتضيه الضرورة لتحقيق المصلحة المشروعة قيد البحث وأن لا تهدد جوهر الحق؛ ومعياري الشرعية، بمعنى أنها يجب أن تتوخى السعي لتحقيق مصلحة مشروعة معدودة، وهي حماية حقوق الآخرين أو سمعتهم، أو الأمن الوطني أو النظام العام، أو الصحة العامة أو الآداب العامة (A/HRC/38/35، الفقرة ٧) وفي هذا الإطار، يمكن أيضاً تقييد الحقوق في التعبير عملاً بالمادة ٢٠ (٢) من العهد، التي تقتضي من الدول أن تحظر "الدعوة إلى الكراهية الوطنية أو العنصرية أو الدينية التي تشكل تحريضاً على التمييز أو العداوة أو العنف"، ولكن لا يزال يتعين أن تستوفي القيود الشروط التراكمية للشرعية والضرورة والشرعية^(٢٥).

٢٩ - يمكن أن تتفاقم درجة تعقيد عملية اتخاذ القرار الكامنة في عملية تنظيم المحتوى نتيجة أعمال عمليات التشغيل الآلي. وخلافاً للبشر، ليس في استطاعة خوارزميات الحاسوب اليوم تقييم السياق الثقافي، والكشف عن السخرية أو إجراء التحليل النقدي اللازم للتعرف بدقة، على سبيل المثال، على المحتوى "المتطرف" أو خطاب الكراهية، ومن ثمّ من الأرجح أن تفترض من لدنها حجب المحتوى وتقييده، مما يقوض حقوق فرادى المستخدمين في الاستماع إليهم، فضلاً عن تقويض حقهم في الحصول على المعلومات دون قيود أو رقابة^(٢٦).

٣٠ - وفي نظام يحكمه الذكاء الاصطناعي، يحكم نشر المعلومات والأفكار قواتٌ مبهمه قد تتعارض أولوياتها مع بيئة مواتية لتنوع وسائل الإعلام والأصوات المستقلة. وينبغي للدول الأطراف، من ناحية

(٢٤) المقرر الخاص المعني بتعزيز وحماية الحق في حرية الرأي والتعبير، وممثل منظمة الأمن والتعاون في أوروبا بشأن حرية وسائل الإعلام، ومنظمة الدول الأمريكية، والمقرر الخاص المعني بحرية التعبير واللجنة الأفريقية لحقوق الإنسان والشعوب، والمقرر الخاص المعني بحرية التعبير والحصول على المعلومات، "الإعلان المشترك بشأن حرية التعبير 'أخبار زائفة' والتضليل والدعاية" ٣ آذار/مارس ٢٠١٧. متاح على الرابط التالي: <https://www.osce.org/fom/302796>؛ انظر أيضاً اللجنة المعنية بحقوق الإنسان، التعليق العام رقم ٣٤ (٢٠١١) بشأن حرية الرأي والتعبير. الوثيقتان A/HRC/32/38، الفقرة ٦١ و A/HRC/29/32، الفقرة ٨٦.

(٢٥) اللجنة المعنية بحقوق الإنسان، التعليق العام رقم ٣٤ (٢٠١١) بشأن حرية الرأي والتعبير، الفقرة ٥٠.

(٢٦) Council of Europe, *Algorithms and Human Rights*, p. 21.

وثيقة الصلة بهذا الموضوع، أن تتخذ التدابير المناسبة لمنع أي هيمنة غير ملائمة لوسائل الإعلام أو أي مركز مجموعات إعلامية مملوكة للقطاع الخاص في أوضاع احتكارية قد تضر بتنوع المصادر والآراء^(٢٧).

٣١ - ويفتقر المستخدمون أيضاً إلى الوصول إلى قواعد اللعبة عندما يتعلق الأمر بالمنصات الإلكترونية والمواقع الشبكية المستندة إلى الذكاء الاصطناعي. ويؤدي الافتقار إلى الوضوح بشأن مدى ونطاق الذكاء الاصطناعي والتطبيقات الخوارزمية على شبكة الإنترنت إلى منع الأفراد من معرفة متى يتم نشر مقاييس المعلومات أو تقييدها أو تحديد أهدافها، وبناء على أي مقياس إعلامي يتم ذلك. وقد تُسهم التنازلات الصغيرة لتيسير التصدي لهذه المشكلة من قبيل التحديد الانتقائي لنتائج البحث المدعومة، أو منصات وسائل التواصل الاجتماعي التي تبين متى دفعت الجهات السياسية الفاعلة أجور نشر الإعلان، إسهاماً طفيفاً في مساعدة المستخدمين على فهم قواعد بيئة المعلومات، لكن هذه المعلومات لا تعكس ولا تعالج الشواغل المتعلقة بالنطاق الذي تُشكّل فيه العمليات الخوارزمية تلك البيئة^(٢٨).

٣٢ - وحتى عندما يكون الأفراد على علم بوجود نظم الذكاء الاصطناعي، وبنطاقها وسير عملها، فإن هذه النظم ذاتها قد تحبط الجهود الرامية لتعزيز الشفافية والملاءمة. وحتى الآن، لم يتم استحداث وسائل متطورة وقابلة للقياس من أجل التدقيق في الأسس التقنية للقرارات المتخذة بواسطة التشغيل الآلي في الفضاء الإلكتروني وتحقيق الشفافية فيها^(٢٩). وهذا يعني أن حقوق الأفراد في التعبير كثيراً ما تتأثر سلباً دون أن يتمكنوا من التحقيق في الأمر أو فهم السبب أو كيف أو على أي أساس حصل ذلك.

دال - الحق في الخصوصية

٣٣ - يعتبر الحق في الخصوصية في كثير من الأحيان بمثابة مدخل إلى التمتع بحرية الرأي والتعبير^(٣٠). وتحمي المادة ١٧ من العهد الفرد من "التدخل تعسفياً أو بشكل غير قانوني في خصوصياته أو شؤون أسرته، أو منزله أو مراسلاته" وتحميه أيضاً من "الاعتداءات غير القانونية على شرفه وسمعته" وتنص على أن "لكل شخص الحق في حماية القانون ضد هذا التدخل أو هذه الاعتداءات" وشددت مفوضية الأمم المتحدة لحقوق الإنسان ومجلس حقوق الإنسان على أن أي تدخل في الحق في الخصوصية يجب أن يستوفي معايير المشروعية والضرورة والتناسب (A/HRC/27/37، الفقرة ٢٣، وقرار مجلس حقوق الإنسان ٧/٣٤، الفقرة ٢)

٣٤ - وتعتمد نظم اتخاذ القرارات المستندة إلى الذكاء الاصطناعي على جمع البيانات واستغلالها، وتتراوح من البيانات المحيطة وغير الشخصية إلى المعلومات المحددة للهوية الشخصية، مع كون الغالبية العظمى من البيانات المستخدمة في تغذية نظم الذكاء الاصطناعي موجودة في مكان ما في وسط هذه

(٢٧) اللجنة المعنية بحقوق الإنسان، التعليق العام رقم ٣٤ (٢٠١١) بشأن حرية الرأي والتعبير، الفقرة ٤٠.

(٢٨) صفيه أوموجا نوبل، خوارزميات القمع: كيف تعزز محركات البحث العنصرية *Algorithms of Oppression: How Search Engines Reinforce Racism*، نيويورك، مطبعة جامعة نيويورك، (٢٠١٨).

(٢٩) مايك أتاني وكيت كراوفورد، "رؤية دون معرفة: القيود المفروضة على مثل الشفافية وتطبيقها على مساءلة خوارزميات الحاسوب *Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability*"، *New Media and Society*، المجلد ٢٠، العدد ٣ (١٣ كانون الأول/ديسمبر ٢٠١٦) متاحة على الرابط التالي: <http://journals.sagepub.com/doi/abs/10.1177/1461444816676645?journalCode=nmsa>.

(٣٠) انظر A/HRC/29/32، الفقرة ١٦، وقرار الجمعية العامة ١٦٧/٦٨ وقرار مجلس حقوق الإنسان ٨/٢٠.

المعادلة - البيانات التي يتم الاستدلال عليها أو استخراجها من البيانات الشخصية أو البيانات التي تم إخفاء هوية أصحابها (غالباً بصورة غير دقيقة). وتستخدم الشركات البيانات المستمدة من الترميز عبر الإنترنت والبصمات الرقمية وتشتري مجموعات البيانات من أطراف ثالثة مثل سمسرة البيانات وتستمد بيانات جديدة من مجموعات البيانات المجمعة الواسعة النطاق لتغذية نظم الذكاء الاصطناعي. وكثيراً ما تكون المنتجات الاستهلاكية المستندة إلى الذكاء الاصطناعي والنظم المستقلة مزودة بأجهزة استشعار تقوم بتوليد وجمع كميات كبيرة من البيانات عن الأفراد الموجودين في محيطها^(٣١) وتستخدم أساليب الذكاء الاصطناعي الموجودة على منصات وسائل التواصل الاجتماعي للاستدلال أو لتوليد معلومات حساسة عن أشخاص لم يقدموها أو يؤكدوها هم أساساً، مثل التوجه الجنسي، أو العلاقات الأسرية، أو الآراء الدينية، أو الظروف الصحية، أو الانتماء السياسي.

٣٥ - ويتحدى الذكاء الاصطناعي الأفكار التقليدية المتعلقة بالموافقة والغرض وقيود الاستخدام، والشفافية والمساءلة - وهي الركائز التي تقوم عليها المعايير الدولية لحماية البيانات^(٣٢). ونظراً إلى أن نظم الذكاء الاصطناعي تعمل من خلال استغلال مجموعات البيانات القائمة وتنشئ بيانات جديدة، فإن قدرة الأفراد على معرفة الكيفية التي تستخدم فيها بياناتهم، وفهمها والسيطرة عليها خالية من المعنى العملي في سياق الذكاء الاصطناعي. وعندما يتم تغيير القصد من البيانات في نظم الذكاء الاصطناعي، فإنها تفقد سياقها الأصلي، مما يزيد من خطر أن تصبح البيانات المتعلقة بالأفراد غير دقيقة أو متقدمة، ويحرم الأفراد من القدرة على تصحيح أو حذف البيانات. ويجري حالياً استخدام النظم المستندة إلى الذكاء الاصطناعي في اتخاذ القرارات التي تترتب عليها تبعات باستخدام تلك البيانات، التي يؤثر بعضها تأثيراً عميقاً في حياة الناس^(٣٣)، ومع ذلك لا يمتلك الأفراد سوى القليل من السبل لممارسة السيطرة على البيانات المستمدة من بياناتهم الشخصية، في الوقت الذي ما زالت تقنيات إخفاء الهوية تعاني فيه من أوجه القصور.

هاء - الالتزام بعدم التمييز

٣٦ - عدم التمييز هو مبدأ أصيل في قانون حقوق الإنسان، وهو موجود ليس لتوصيف التزامات الدول بضمان التمتع بجميع حقوق الإنسان دون تمييز فحسب وإنما ليكون، على النحو المنصوص عليه في المادة ٢٦ من العهد، بمثابة ضمان قائم بذاته لكفالة المساواة أمام القانون والمساواة في التمتع بحماية القانون. ويقع على عاتق الدول التزام واضح بأن "تحظر أي تمييز وأن تكفل لجميع الأشخاص على السواء حماية فعالة من التمييز لأي سبب كان، كالعرق أو اللون أو الجنس أو اللغة أو الدين أو الرأي السياسي أو غير السياسي، أو الأصل القومي أو الاجتماعي، أو الثروة أو النسب، أو غير ذلك من الأسباب". ولذلك، تنص المادتان ١٧ و ١٩ على حقوق الأفراد بعدم التعرض للتمييز في اعتناق وتكوين الآراء، والتعبير عن الأفكار والمعلومات والوصول إليهما، وممارسة الخصوصية وحماية البيانات الشخصية.

(٣١) المادة ١٩ والمنظمة الدولية لحماية الخصوصية، "Privacy and freedom of expression"

(٣٢) اللجنة المعنية بحقوق الإنسان، التعليق العام رقم ١٦: المادة ١٧ (١٩٨٨) بشأن الحق في الخصوصية، الفقرة ١٠.

(٣٣) المادة ١٩ والمنظمة الدولية لحماية الخصوصية، "Privacy and freedom of expression", p. 19.

٣٧ - وتمتد إمكانية قيام الذكاء الاصطناعي بتضمين التحيز والتمييز وإدامته، وصولاً إلى التمييز في ممارسة الحق في حرية الرأي والتعبير. وقد لا يراعي تنقيح خوارزميات الحاسوب السياقات والحساسيات الثقافية، أو اللغوية، أو الجنسانية، أو المصلحة العامة في المحتوى^(٣٤). وقد تُكرس الموجزات الإخبارية المستندة إلى الذكاء الاصطناعي المواقف التمييزية وتعززها، بينما يُيسر التنميط والإعلان المبنيان على الذكاء الاصطناعي، بكل وضوح، التمييز على أساس العرق والدين ونوع الجنس^(٣٥). وأسفرت مهام "الإكمال التلقائي Autocomplete" في نظم الذكاء الاصطناعي أيضاً عن نتائج تمييزية عنصرية^(٣٦).

٣٨ - وهناك عدد من العوامل التي تغرس التحيز في نظم الذكاء الاصطناعي، مما يزيد من إمكاناتها التمييزية. وتشمل هذه العوامل الطريقة التي تصمم بها نظم الذكاء الاصطناعي والقرارات المتعلقة بمصدر ونطاق مجموعات البيانات التي يتم تدريب هذه النظم عليها، والتحيزات الاجتماعية والثقافية التي قد يدرجها واضعو النظم في مجموعات البيانات تلك، ونماذج الذكاء الاصطناعي ذاتها والطريقة التي تنفذ بموجبها نواتج نموذج الذكاء الاصطناعي. فعلى سبيل المثال، تعاني تطبيقات التعرف على الوجوه من كونها متحيزة في مجموعات البيانات التي تركز في الغالب على الذكور ذوي البشرة البيضاء، الأمر الذي يتسبب في وقوع ما تصل نسبته إلى ٢٠ في المائة في كل مرة بالنسبة للنساء والأشخاص ذوي البشرة السوداء^(٣٧). وعندما تُستخدم هذه النظم، على سبيل المثال، لتصنيف الصور المتاحة عن طريق محرك بحث، فإن إمكاناتها التمييزية يمكن أن تُترجم إلى تدخلات في ممارسة الأفراد لحقوقهم في التماس وتلقي ونقل المعلومات وفي حرية التجمّع أو تكوين الجمعيات.

واو - الحق في الحصول على سبيل انتصاف فعال

٣٩ - يضمن قانون حقوق الإنسان للأفراد سبيل انتصاف تقرره سلطات قضائية أو إدارية أو تشريعية مختصة (المادة ٢ (٣) من العهد). ويجب أن تكون سبيل الانتصاف معروفة ومتاحة لأي شخص تُنتهك

(٣٤) أدى ذلك، على سبيل المثال، إلى إزالة الصور التاريخية ذات الأهمية الثقافية الخاصة. انظر جوليا كاري وونغ "مارك زكريبرغ متهم بإساءة استعمال السلطة بعد أن حذف فيسبوك من صفحاته صورة 'فناة النابالم'"، صحيفة الغارديان، ٩ أيلول/سبتمبر ٢٠١٦. متاح على الرابط التالي: <https://www.theguardian.com/technology/2016/sep/08/facebook-mark-zuckerberg-napalm-girl-photo-vietnam-war>؛ انظر أيضاً الفقرة ٢٩ من الوثيقة A/HRC/38/35.

(٣٥) جوليا أنغوين ومادلين فارنر وأريانا توبين، "فيسبوك مكنّ المعلنين من بلوغ 'كارهي اليهود'"، "Facebook enabled advertisers to reach 'Jew haters'", ProPublica، ١٤ أيلول/سبتمبر ٢٠١٧. متاح على الرابط التالي: <https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters>؛ أريانا توبين، "لماذا كنا نتقبل العنصرية والتحيز الجنسي وكرهية الأجانب والاعلانات التمكينية والاعلانات الفظيعة من نواح أخرى المنشورة على فيسبوك"، ProPublica، ٢٧ تشرين الثاني/نوفمبر ٢٠١٧. متاح على الرابط التالي: <https://www.propublica.org/article/why-we-had-to-buy-racist-sexist-xenophobic-ableist-and-otherwise-awful-facebook-ads>.

(٣٦) باريس مارتينو، "يشير البحث في 'يوتيوب' إلى عنصرية الإكمال التلقائي autocompletes"، الموجز The Outline، ١٣ أيار/مايو ٢٠١٨. متاح على الرابط التالي: <https://theoutline.com/post/4536/youtube-s-search-autofill-suggests-racist-results?zd=1&zi=3ygz6hw>.

(٣٧) جوي بولاموني، "مخاطر البيانات البيضاء للغاية والنظرة المشفرة" "The dangers of supremely white data and the coded gaze"، معروضة على ويكيமானيا Wikimania 2018، كيب تاون. متاح على الرابط التالي: <https://www.youtube.com/watch?v=ZSJKoD6mA8&feature=youtu.be>.

حقوقه؛ ويجب أن تنطوي على إجراء تحقيق فوري وشامل ونزيه في الانتهاكات المزعومة^(٣٨)؛ ويجب أن تكون قادرة على إنهاء الانتهاكات الجارية (A/HRC/27/37، الفقرات ٣٩-٤١).

٤٠ - وكثيراً ما تتعارض نظم الذكاء الاصطناعي مع الحق في الانتصاف. أولاً، إخطار الفرد غير متاح بطبيعته تقريباً. وفي معظم تطبيقات تكنولوجيا الذكاء الاصطناعي في بيئة المعلومات لا يعرف الأفراد نطاق أو مدى أو حتى وجود العمليات الخوارزمية في اتخاذ القرارات التي قد تؤثر على تمتعهم بالحق في الرأي والتعبير. والجانب الثاني والأكثر صعوبة هو إمكانية فهم نظام الذكاء الاصطناعي نفسه. فقد لا يكون المنطق الكامن وراء قرار خوارزميات الحاسوب واضحاً حتى للخبراء المدربين على التقنيات الأساسية للنظام. وعلى الرغم من أنه من المنطقي الافتراض بأن المزيد من الشفافية بشأن نظم الذكاء الاصطناعي يتيح مزيداً من التدقيق، فإن شفافية خوارزميات الحاسوب لا تساوي بالضرورة التفسيرات المفهومة لعمليات اتخاذ القرار. وقد تحجب خوارزميات الحاسوب واقعة اتخاذ قرارٍ تترتب عليه تبعات أو يتسم بتعقيدات أو يتوقف فهمه على السياق، إلى حد يصعب معه تفسيره. وتزداد الحالة تعقيداً بسبب قيام الشركات التي تعمل في بيئة المعلومات في كثير من الأحيان بتحديث خوارزميات الحاسوب الخاصة بها؛ وبالمثل، قد تقوم تطبيقات التعلم الآلي بتغيير قواعدها وخوارزميات الحاسوب الخاصة بها على مر الزمن^(٣٩).

٤١ - ومما يزيد من تفاقم هذه الشواغل هو التحول نحو التشغيل الآلي لنظم الانتصاف نفسها، التي تفيد بأن شكاوى فرادى المستخدمين، إما بشأن قرارات تنقيح المحتوى وإما بشأن الآثار الضارة بحقوق الإنسان الناجمة عن تكنولوجيا الذكاء الاصطناعي، تُنظر فيها وتحُدُّها تكنولوجيا الذكاء الاصطناعي^(٤٠). وتثير عمليات الاستجابة بواسطة التشغيل الآلي القلق بشأن ما إذا كانت آليات معالجة الشكاوى تشكل وسيلة انتصاف فعالة، بالنظر إلى أن عوامل الافتقار إلى السلطة التقديرية، وإمكانية تحليل السياق والقدرة على البت بصورة مستقلة، كامنة بأجمعها في صلب هذه العمليات^(٤١).

زاي - الردود التشريعية والتنظيمية والسياساتية على الذكاء الاصطناعي

٤٢ - يعكف العديد من الدول الآن على وضع استراتيجيات وطنية للذكاء الاصطناعي من أجل استكشاف وتطوير السياسات والمبادرات الرامية إلى تحقيق الحد الأقصى من الفوائد المحتملة من الذكاء الاصطناعي لمواطنيها^(٤٢). وعلى الرغم من عدم قيام أي دولة حتى الآن باقتراح قانون شامل أو لائحة

(٣٨) اللجنة المعنية بحقوق الإنسان، التعليق العام رقم ٣١ (٢٠٠٤) بشأن طبيعة الالتزام القانوني العام المفروض على الدول الأطراف في العهد، الفقرة ١٥.

(٣٩) باري شوارتز، "غوغل: تُنفذ الآلاف من التحديثات على خوارزميات البحث كل سنة"، اجتماع المائدة المستديرة المعني بمحركات البحث، ٥ حزيران/يونيه ٢٠١٥. متاح على الرابط التالي: <https://www.seroundtable.com/google-updates-thousands-20403.html>

(٤٠) Council of Europe, *Algorithms and Human Rights*, p. 24.

(٤١) بي جانغ، وصوفي ستالا بورديلون وليستر غيلبرت، "نموذج ربط المحتوى بالسياق لأغراض إجراءات الإخطار والاعتماد"، *WebSci*، ١٦ أيار/مايو ٢٠١٦. متاح على الرابط التالي: <http://takedownproject.org/wp-content/uploads/2016/04/ContentLinkingModelZhangStallaGilbert.pdf>

(٤٢) تيم دوتون، "لحة عامة عن الاستراتيجيات الوطنية المتعلقة بالذكاء الاصطناعي"، مجلة Medium، ٢٨ حزيران/يونيه ٢٠١٨. متاح على الرابط التالي: <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>.

شاملة بخصوص الذكاء الاصطناعي، فثمة أسباب تدعو للحذر إزاء اتباع نهج من هذا القبيل، الذي قد يكون غير مناسب لمثل هذا المجال المبتكر والذي قد يعوض عن الافتقار إلى التفاصيل بفرض أحكام تقييدية أو متساهلة بشكل مفرط. وعلى الرغم من أن إصدار لوائح لكل قطاع بمفرده قد تكون أفضل، فإنه يمكن القول إن القوانين والأنظمة القائمة، في مجال حماية البيانات على سبيل المثال، يمكن أن تكون مرنة ومتاحة دون الحاجة إلى سن تشريعات خاصة بها.

٤٣ - وفي الوقت نفسه، ينبغي على الدول ضمان تطوير الذكاء الاصطناعي بما يتماشى مع معايير حقوق الإنسان. وينبغي لأي جهود تُبذل لوضع سياسة أو لوائح للدولة في مجال الذكاء الاصطناعي أن تكفل مراعاة الشواغل المتعلقة بحقوق الإنسان^(٤٣). وكثيراً ما يُستبعد الحق في حرية الرأي والتعبير، على وجه الخصوص، من المناقشات العامة والسياسية حول الذكاء الاصطناعي، والتي تميل، إلى الحد الذي تعالج فيه قضايا حقوق الإنسان، إلى التركيز على التحيز والتمييز في تقديم الخدمات.

٤٤ - ونظراً لأن تطوير نظم فعالة للذكاء الاصطناعي يتوقف على اقتناء مجموعات البيانات الكبيرة فضلاً عن الاستثمار الطويل الأجل في القدرات التكنولوجية، فمن المرجح أن تهيمن كيانات القطاع الخاص على مجالات التطوير والإنتاج والقدرة، مما يؤدي إلى زيادة اعتماد القطاع العام على الشركات للحصول على نظم الذكاء الاصطناعي. وهذا يزيد من احتمال وتصور الجمهور بأن تصبح مصالح الشركات والحكومات أكثر تشابكاً. وهذا هو الحال بوجه خاص في بيئة المعلومات، التي كثيراً ما تكون الحكومات من المستخدمين فيها - في منصات وسائل التواصل الاجتماعي وفي محركات البحث والتكنولوجيات الأخرى - بدلا من مقدمي الخدمات إليها. ولا تؤدي المواءمة بين المصالح العامة والخاصة في حد ذاتها إلى تدخلات في حقوق الإنسان ولكنها تثير القلق بشأن الشفافية والمساءلة. ومع مضي القطاع الخاص بتطوير الذكاء الاصطناعي، هناك خطر حقيقي جدا بأن تُفوض الدول الشركات بولايات الرقابة والإشراف المتزايدة التعقيد والمرهقة.

٤٥ - وينبغي لأي دولة تحاول صياغة قانون أو سياسة عامة في مجال الذكاء الاصطناعي أن تتحدث إلى القطاعين العام والخاص على السواء، بدلا من مجرد التركيز على تنظيم الذكاء الاصطناعي للقطاع العام. وبناء على ما خلص إليه مجلس أوروبا، فإن "القضايا المتعلقة بالحكومة الخوارزمية و/أو التنظيم الخوارزمي هي من اختصاصات السياسة العامة ولا ينبغي تركها للجهات الفاعلة الخاصة لوحدها"^(٤٤). وقد تنطوي النهج التي تتبعها الدول على تعزيز الشفافية وفرض التزامات الإفصاح على الشركات وسن تشريعات صارمة لحماية البيانات التي تتناول الشواغل المتصلة بالذكاء الاصطناعي.

٤٦ - وتشهد مبادرات القطاعين العام والخاص الرامية إلى استكشاف وإدماج الأخلاقيات في المشتريات والتصميم والنشر والتنفيذ المتعلقة بنظم الذكاء الاصطناعي مزيداً من الانتشار. ويشجع المقرر الخاص بقوة على إدماج الشواغل المتعلقة بحقوق الإنسان في هذه الجهود. وكثيراً ما ينطوي تركيز القطاع الخاص على الأخلاقيات ودفع القطاع العام نحوها على مقاومة وضع لوائح مبنية على حقوق

(٤٣) مما يثير القلق، على سبيل المثال، أن لجنة برلمانية في المملكة المتحدة لبريطانيا العظمى وأيرلندا الشمالية أصدرت تقريراً مؤلفاً من ٢٠٠ صفحة ولم يرد فيه، ولو مرة واحدة، أي ذكر لحقوق الإنسان. انظر اللجنة المختارة المعنية بالذكاء الاصطناعي التابعة لمجلس اللوردات في المملكة المتحدة لبريطانيا العظمى وأيرلندا الشمالية، "منظمة الذكاء الاصطناعي (AI) في المملكة المتحدة".

الإنسان^(٤٥). وفي حين أن الأخلاقيات توفر إطاراً هاماً جداً للعمل من خلال مواجهة تحديات خاصة في مجال الذكاء الاصطناعي، فإنها ليست بديلاً عن حقوق الإنسان، التي يقع على عاتق كل دولة التقيد بها بموجب القانون. وينبغي للشركات والحكومات أن تكفل إدماج الاعتبارات والمسؤوليات المتعلقة بحقوق الإنسان، بقوة في جميع جوانب عمليات الذكاء الاصطناعي التي تضطلع بها حتى في أثناء وضع المدونات والتوجيهات المتعلقة بالأخلاقيات^(٤٦).

رابعاً - اتباع نهج قائم على حقوق الإنسان في الذكاء الاصطناعي

٤٧ - في التقارير الأخيرة، حدد المكلفون بولايات مجموعة من الاعتبارات القانونية والعملية للشركات بأن تضع مبادئ حقوق الإنسان في صميم سياسات تنظيم المحتوى ووضعها بالتفصيل المعايير والعمليات الموضوعية التي تكفل أن تمثل الشركات لمسئولياتها في مجال حقوق الإنسان بموجب المبادئ التوجيهية المتعلقة بالأعمال التجارية وحقوق الإنسان في جميع جوانب عملياتها. وهذا الإطار نفسه هو الذي يُصمم شكل النهج المعروض هنا فيما يتعلق بتكنولوجيات الذكاء الاصطناعي. وتنطبق المعايير والعمليات الموضوعية المقترحة أدناه على الشركات، بصفتها الجهات الفاعلة التي تقوم بتصميم ونشر وتنفيذ نظم الذكاء الاصطناعي، وتنطبق أيضاً على الدول باعتبارها ملزمة بالامتناع عن التدخل في حقوق الإنسان عند اعتمادها واستخدامها لنظم الذكاء الاصطناعي. وهذه المعايير والعمليات مصممة لضمان أن يكون قانون حقوق الإنسان في صميم التطورات الحاصلة في مجال الذكاء الاصطناعي. وهناك مبدآن أساسيان يشكلان جزءاً لا يتجزأ من جميع المعايير والعمليات المقدمة: الحاجة إلى حماية واحترام القدرة الفردية والاستقلال الذاتي، وهما شرطان مسبقان أساسيان لممارسة الحق في حرية الرأي والتعبير؛ وأهمية الإفصاح المفيد من جانب الجهات الفاعلة في القطاعين العام والخاص، الذي يتجلى من خلال بذل جهود مفتوحة ومبتكرة ترمي إلى توضيح تكنولوجيات الذكاء الاصطناعي وتيسير تدقيقها.

ألف - المعايير الموضوعية لنظم الذكاء الاصطناعي

٤٨ - ينبغي للشركات أن توجه معاييرها وقواعدها وتصميم نظمها حول مبادئ حقوق الإنسان العالمية (A/HRC/38/35، الفقرات ٤١-٤٣). وينبغي أن تُستكمل المصطلحات والواجهات العامة والمبادئ التوجيهية بالتزامات على صعيد السياسة الداخلية بمراعاة تعميم اعتبارات حقوق الإنسان في جميع عمليات الشركة، ولا سيما فيما يتعلق بتطوير ونشر نظم الذكاء الاصطناعي وخوارزميات الحاسوب. وينبغي للشركات أن تنظر في كيفية وضع المعايير المهنية لمهندسي الذكاء الاصطناعي، وترجمة المسؤوليات في مجال حقوق الإنسان إلى توجيهات لأغراض التصميم التقني وخيارات العمليات. ويمكن أن يشكل وضع مدونات الأخلاقيات والهياكل المؤسسية المرافقة تكملة هامة، ولكنها ليست بديلاً عن الالتزامات المتعلقة بحقوق الإنسان. وينبغي للمدونات والمبادئ التوجيهية الصادرة عن هيئات القطاعين العام والخاص أن تؤكد على ضرورة أن ينص قانون حقوق الإنسان على القواعد الأساسية لحماية الأفراد

(٤٥) بن واغتر، "الأخلاقيات للفرار من اللوائح: من غسل الأيدي من الأخلاقيات إلى تسوّق الأخلاقيات؟" مبلغ، ميري (٤٥) *Being Profiling: Cogitas Ergo Sum*, Mireille Hildebrandt, ed. (Amsterdam University Press (forthcoming))

(٤٦) المادة ١٩، والمنظمة الدولية لحماية الخصوصية، p. 13 "Privacy and freedom of expression".

في سياق الذكاء الاصطناعي، في حين يمكن أن تساعد أطر الأخلاقيات على زيادة تطوير المحتوى وتطبيق حقوق الإنسان في ظروف محددة.

٤٩ - ويجب على الشركات والحكومات أن تكون صريحة مع الأفراد حول القرارات التي يتم اتخاذها في بيئة المعلومات من خلال نظم التشغيل الآلي والقرارات التي تكون مصحوبة بمراجعة بشرية، فضلا عن العناصر العامة للمنطق الذي تستخدمه تلك النظم. وينبغي أيضا إبلاغ الأفراد متى ستصبح البيانات الشخصية التي يقدمونها إلى جهة فاعلة من جهات القطاع الخاص (صراحة أو عن طريق استخدامهم للخدمة أو للموقع) جزءا من مجموعة البيانات المستخدمة من قبل نظام الذكاء الاصطناعي، لتمكينهم من عوامة تلك المعرفة في اتخاذ قرار بشأن ما إذا كان ينبغي الموافقة على جمع البيانات وعلى أي نوع من أنواع البيانات التي يرغبون في الكشف عنها^(٤٧). وعلى غرار الإخطارات العامة اللازمة لاستخدام كاميرات الدوائر التلفزيونية المغلقة، ينبغي لنظم الذكاء الاصطناعي أن تكشف للأفراد مباشرة (من خلال وسائل مبتكرة مثل الرسائل الالكترونية المنبثقة pop-up) وبطريقة واضحة ومفهومة بأنهم يخضعون لعملية اتخاذ صنع القرار بواسطة الذكاء الاصطناعي أو يساهمون ببيانات فيها، فضلا عن معلومات مفيدة عن المنطق المستخدم في العملية وأهمية العواقب المترتبة على ذلك بالنسبة للفرد.

٥٠ - ولا تتوقف الشفافية عند الإفصاح لفردى المستخدمين عن وجود تكنولوجيات الذكاء الاصطناعي في المنصات والخدمات الإلكترونية التي يستخدمونها. ولا بد من أن تتحلى الشركات والحكومات بالشفافية في كل جانب من جوانب سلسلة القيمة للذكاء الاصطناعي. ولا حاجة لأن تكون الشفافية معقدة كيما تكون فعالة؛ وحتى التفسيرات المبسطة للغرض من نظام الذكاء الاصطناعي والسياسات والمدخلات والنواتج المتصلة به يمكن أن تساهم في تثقيف الجمهور العام والمناقشة العامة بشأنه^(٤٨). وبدلا من التعامل مع مأزق جعل العمليات التقنية المعقدة مقروءة للجمهور العادي، ينبغي للشركات أن تسعى إلى تحقيق الشفافية من خلال توفير رؤى غير تقنية في النظام. وتحقيقا لهذه الغاية، ينبغي أن ينصب التركيز على تثقيف فرادى المستخدمين بوجود نظام الذكاء الاصطناعي، والغرض منه، ودستوره وأثره بدلا من أن ينصب التركيز على رمز المصدر والتدريب على البيانات والمدخلات والنواتج^(٤٩).

٥١ - وتقتضي الشفافية الجذرية بشأن أثر نظم الذكاء الاصطناعي في بيئة المعلومات الإفصاح، على سبيل المثال، عن البيانات المتعلقة بحجم المحتوى الذي تزيله نظم الذكاء الاصطناعي، وعن عدد المرات التي يوافق فيها منقح بشري على إزالة محتوى باقتراح من نظم الذكاء الاصطناعي، وعن عدد المرات التي تكون فيها عمليات إزالة المحتوى موضع طعن، وعدد المرات التي تقبل فيها الطعون في عمليات إزالة المحتوى. وينبغي أن تكون البيانات المجمعة التي تبين الاتجاهات في عرض المحتوى متاحة للمستخدمين من أجل تفتيشها، إلى جانب دراسات الحالات الفردية التي تبين السبب الذي يجعل محتوى معين يحظى بالأولوية على محتوى آخر. ويعتبر الإفصاح عن المصادر والمستفيدين من الإعلانات السياسية والتجارية عنصراً حاسماً للشفافية الجذرية. وينبغي أيضا للجهات الفاعلة في القطاعين العام والخاص التي تقوم بتنفيذ

(٤٧) اللجنة المعنية بحقوق الإنسان، التعليق العام رقم ١٦ (١٩٨٨) بشأن الحق في الخصوصية.

(٤٨) Aaron Rieke, Miranda Bogen and David Robinson, "Public scrutiny of automated decisions: early lessons and emerging methods" (Omidyar and Upturn, 2018), p. 5

(٤٩) Rieke, Bogen and Robinson, "Public scrutiny of automated decisions", p. 8

نظم الذكاء الاصطناعي أن تتحلى هي الأخرى بالشفافية بشأن حدود الذكاء الاصطناعي، بما في ذلك، على سبيل المثال، تدابير الثقة وسيناريوهات الفشل المعروفة، والقيود الملائمة المفروضة على استعمالها^(٥٠).

٥٢ - ويشكل التصدي لانتشار التمييز في نظم الذكاء الاصطناعي تحدياً وجودياً للشركات والحكومات؛ وسيؤدي الفشل في معالجة وحل العناصر والتأثيرات التمييزية إلى جعل التكنولوجيا ليست فعالة فحسب وإنما خطيرة أيضاً. وهناك قيادة وموارد فكرية مفتوحة أمام الشركات والحكومات للاستفادة منها في كيفية التعامل مع التحيز والتمييز في نظم الذكاء الاصطناعي؛ وبصفة عامة، فهي تتطلب القيام على مستوى المدخلات ومستوى النواتج على السواء بعزل التمييز والمحاسبة عليه. ويشمل ذلك، على الأقل، معالجة الأخطاء في أخذ العينات (حيث تكون مجموعات البيانات غير ممثلة للمجتمع)، وتنقية مجموعات البيانات لإزالة البيانات التمييزية منها، ووضع تدابير للتعويض عن البيانات التي "تتضمن بصمات أنماط التمييز التاريخية والميكانيكية" التي يرجح أن تُطوّر نظم الذكاء الاصطناعي منها سجلاتٍ تمييزية^(٥١). كما يشكل الرصد الفعال للنواتج التمييزية لنظم الذكاء الاصطناعي جزءاً لا يتجزأ من تجنب الآثار السلبية على حقوق الإنسان للأفراد وتخفيفها.

باء - عمليات لنظم الذكاء الاصطناعي

٥٣ - تقييمات الأثر على حقوق الإنسان - يستلزم اعتناق الشفافية الجذرية في جميع مراحل دورة حياة الذكاء الاصطناعي قيام الشركات والحكومات باتخاذ خطوات تتيح تدقيق النظم والطعن بمحتواها من مرحلة البداية إلى مرحلة التنفيذ. وتمثل تقييمات الأثر على حقوق الإنسان إحدى الأدوات التي يمكن أن تثبت وجود التزام بالتصدي للآثار المترتبة على نظم الذكاء الاصطناعي في حقوق الإنسان وينبغي الاضطلاع بها قبل شراء النظم أو تطويرها أو استخدامها وينبغي أن تشمل كلا من التقييم الذاتي والاستعراض الخارجي. وقد اقترح مجمع الفكر التابع لمنظمة "الذكاء الاصطناعي الآن AI Now" إجراء تقييم لأثر خوارزميات الحاسوب على وكالات القطاع العام ينص على ضرورة أن تجري الحكومات استعراضاً داخلياً لنظم الذكاء الاصطناعي، وأن تيسر كذلك إجراء عمليات استعراض خارجي للبحوث بهدف اختبار الافتراضات والاستنتاجات والتحقق منهما^(٥٢). وينبغي للشركات أيضاً أن تجري تقييمات على نفس الشاكلة.

٥٤ - ويجب أن تقرن مشتريات القطاع العام لتكنولوجيات الذكاء الاصطناعي من موردين من القطاع الخاص بمشاوره عامة لالتماس آراء وإسهامات المجتمع بشأن تصميم وتنفيذ نظام الذكاء الاصطناعي قبل شرائه. ويجب على الشركات والحكومات، على حد سواء، أن تجري مشاورات مفيدة ومستمرة مع المجتمع

(٥٠) منظمة العفو الدولية، ومنظمة الوصول الآن، "إعلان تورنتو: حماية الحق في المساواة وعدم التمييز في نظم التعلم الآلي"، المادة ٢٧ (د)، ٢٠١٨. متاح على الرابط التالي: <https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/>

(٥١) إياسون غابرييل، "الأسباب الموجبة لاعتماد خوارزميات أكثر إنصافاً"، مجلة Medium، ١٤ آذار/مارس ٢٠١٨. متاح على الرابط التالي: https://medium.com/@Ethics_Society/the-case-for-fairer-algorithms-c008a12126f8.

(٥٢) ديلون ريسمان وآخرون، "تقييمات أثر خوارزميات الحاسوب: إطار عملي لمسألة الوكالات العامة" (منظمة الذكاء الاصطناعي الآن، ٢٠١٨) متاح على الرابط التالي: <https://ainowinstitute.org/aiareport2018.pdf>.

المدني، وجماعات حقوق الإنسان، والمجتمعات المحلية المعنية وممثلي الفئات الممثلة تمثيلاً ناقصاً أو المهمشة تاريخياً قبل وضع أو تشغيل أو استخدام نظم وتكنولوجيات الذكاء الاصطناعي.

٥٥ - **عمليات المراجعة** - يوفر تيسير إجراء استعراض خارجي لنظم الذكاء الاصطناعي ضماناً بالغة الأهمية في سبيل تحلي الشفافية بالدقة والاستقلال. ولهذا السبب، ينبغي أن تُكتمل عمليات المراجعة المستقلة الجارية تقييمات الأثر على حقوق الإنسان قبل الشراء باعتبارها آلية هامة لضمان الشفافية والمساءلة في نظم الذكاء الاصطناعي. وقد أثارت الجهات الفاعلة في القطاع الخاص اعتراضات على إمكانية المراجعة في فضاء الذكاء الاصطناعي، بالنظر إلى ضرورة حماية التكنولوجيا المسجلة. وفي حين أن هذه المخاوف تستند إلى أسس سليمة، فإن المقرر الخاص يتفق مع منظمة "الذكاء الاصطناعي الآن" ولا سيما عندما يتم استخدام تطبيق الذكاء الاصطناعي من قبل وكالة من وكالات القطاع العام، بأن رفض البائع لكي يكون شفافاً بشأن تشغيل النظام لا يتفق مع التزامات المساءلة المعمول بها في الهيئات العامة.

٥٦ - وعلى أي حال، هنالك الكثير من الاقتراحات المبتكرة لمراجعة تكنولوجيا الذكاء الاصطناعي التي تسمح بسرية الملكية الخاصة: يمكن تصور إمكانية توليد براهين خالية من المعرفة من خلال خوارزميات الحاسوب لإثبات أنها تتوافق مع مملكتيات معينة، متفادية بذلك الحاجة إلى تدقيق خوارزميات الحاسوب الأساسية^(٥٣)، أو يمكن الكشف عن خوارزميات الحاسوب لأطراف ثالثة من الخبراء الذين يضعونها في حساب ضمان خاضع لشرط السرية، يتيح التدقيق للمصلحة العامة، ولكن دون السماح بأن تصبح خوارزميات الحاسوب في متناول الجمهور^(٥٤). ويمكن السماح للمنظمين الحكوميين من مجالات الاتصالات السلكية واللاسلكية أو المنافسة بالوصول إلى نظم الذكاء الاصطناعي على أساس السرية، كما يحدث بالفعل، على سبيل المثال، في تنظيم آلات القمار في أستراليا ونيوزيلندا، اللتين يتعين على الشركات فيهما أن تقدم نظمها الخوارزمية للمراجعة بموجب اللوائح التنظيمية^(٥٥). تتضمن المؤلفات الأكاديمية اقتراحات أخرى للأشكال المبتكرة من عمليات مراجعة نظم الذكاء الاصطناعي^(٥٦).

٥٧ - وقد تواجه كل آلية من هذه الآليات تحديات في التنفيذ، لا سيما في بيئة المعلومات، ولكن ينبغي للشركات أن تعمل على جعل مراجعة نظم الذكاء الاصطناعي ممكنة. وينبغي للحكومات أن تسهم في فعالية عمليات المراجعة عن طريق النظر في القيام بتدخلات سياسية أو تشريعية تلزم الشركات بجعل مدونات الذكاء الاصطناعي قابلة للمراجعة، بما يضمن وجود سجلات للمراجعة، ومن ثم إتاحة المزيد من فرص الشفافية للأفراد المتضررين.

(٥٣) Council of Europe, *Algorithms and Human Rights*, p. 36.

(٥٤) فرانك باسكال، مجتمع الصندوق الأسود: خوارزميات الحاسوب السرية التي تراقب الأموال والمعلومات *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, Harvard University Press) مطبعة جامعة هارفارد، ٢٠١٥.

(٥٥) Council of Europe, *Algorithms and Human Rights*, p. 34.

(٥٦) كريستيان ساندفيغ وآخرون، "مراجعة خوارزميات الحاسوب: أساليب البحث للكشف عن حالات التمييز على منصات الإنترنت" "Auditing algorithms: research methods for detecting discrimination on Internet platforms" ورقة عرضت في مؤتمر عن "البيانات والتمييز: تحويل الشواغل البالغة الأهمية إلى استفسارات منتجة"، وهو مؤتمر تمهيدي للاجتماع السنوي الرابع والستين لرابطة الاتصالات الدولية، سياتل، ٢٢ أيار/مايو ٢٠١٤.

٥٨ - الاستقلال الذاتي الفردي - يجب ألا يحل الذكاء الاصطناعي بشكل خفي محل قدرة الأفراد على تكوين واعتناق آرائهم، أو يتلاعب بها أو يتدخل فيها أو في الوصول إلى الأفكار والتعبير عنها في بيئة المعلومات. ويعني احترام الاستقلال الذاتي الفردي، على الأقل، ضمان أن يمتلك المستخدمون المعارف وحرية الاختيار والتحكم. وتقوض تطبيقات الذكاء الاصطناعي المتفشية والخفية التي تحجب عمليات عرض المحتوى والسمات الشخصية والتنقيح والتنميط والاستهداف قدرة الأفراد على ممارسة حقوقهم في حرية الرأي والتعبير والخصوصية. وينبغي للشركات أن تضع في اعتبارها الآثار الضارة بحقوق الإنسان التي تنبع من تطبيقات الذكاء الاصطناعي التي تعطي الأولوية للمصالح التجارية أو السياسية على حساب الشفافية والاختيار الفردي.

٥٩ - الإخطار والموافقة - يجب أن تكفل الشركات إطلاع المستخدمين تماماً على الكيفية التي تصوغ بها عملية اتخاذ القرار المبني على خوارزميات الحاسوب استخدامهم لمنصة إلكترونية أو موقع أو خدمات. ويمكن تحقيق ذلك من خلال الحملات التثقيفية، والرسائل الإلكترونية المبنية، والإعلانات البينية وغيرها من وسائل الإشارة التي تنبعث عندما يحدد نظام الذكاء الاصطناعي تجربة المستخدمين لمحرك بحث أو لموقع إخباري أو منصة للتواصل الاجتماعي. وقد تكون متطلبات الإفصاح عن البيانات التي تفرضها الدولة وسيلة مناسبة لحماية الإخطار والموافقة. وللأفراد أيضاً الحق في معرفة الوقت الذي يقوم فيه تطبيق من تطبيقات الذكاء الاصطناعي بجمع بياناتهم وما إذا كانت تلك البيانات ستصبح جزءاً من مجموعة من البيانات التي سوف تُبلَّغ في وقت لاحق تطبيقاً من تطبيقات الذكاء الاصطناعي، فضلاً عن الشروط التي سوف يتم فيها استخدام البيانات وتخزينها وحذفها.

٦٠ - سبل الانتصاف - يجب أن تكون الآثار الضارة المترتبة على نظم الذكاء الاصطناعي في حقوق الإنسان قابلة للانتصاف والمعالجة من قبل الشركات المسؤولة. ويتمثل الشرط المسبق لإنشاء عمليات للانتصاف الفعال في ضمان معرفة الأفراد بأنهم موضع قرار خوارزمي (بما في ذلك قرار اقترحه أحد نظم الذكاء الاصطناعي ووافق عليه مُحاور بشري) وبأن يكونوا مزودين بمعلومات عن المنطق الكامن وراء ذلك القرار. وعلاوة على ذلك، ينبغي أن تكفل الشركات استعراض طلبات الانتصاف من قبل البشر من أجل التحقق على النحو المناسب من النظم وضمان المساءلة. ويجب نشر البيانات المتعلقة بالوتيرة التي يتم فيها تشغيل آليات الانتصاف من أجل القرارات التي تتخذها تكنولوجيات الذكاء الاصطناعي.

خامساً - الاستنتاجات والتوصيات

٦١ - في هذا التقرير، قام المقرر الخاص ببحث الآثار المترتبة على الذكاء الاصطناعي، الحالية والمحتملة، في الحق في حرية الرأي والتعبير، مشيراً إلى أن الذكاء الاصطناعي بات الآن جزءاً أساسياً من بيئة المعلومات، مما ينطوي على فوائد ومخاطر بالنسبة لتمتع الأفراد بحقوقهم. ولقد اقترح المقرر الخاص وضع إطار مفاهيمي للتفكير بشأن التزامات الدول ومسؤوليات الشركات باحترام هذه الحقوق تجاه توسع القدرات التكنولوجية واقترحت أيضاً تدابير ملموسة يمكن تنفيذها من قبل الحكومات والشركات على السواء لضمان احترام حقوق الإنسان مع تنامي قوة وباع ونطاق تكنولوجيا الذكاء الاصطناعي.

التوصيات الموجهة إلى الدول

٦٢ - عند اقتناء أو نشر نظم أو تطبيقات الذكاء الاصطناعي، ينبغي أن تكفل الدول بان تتصرف هيئات القطاع العام على نحو يتسق مع مبادئ حقوق الإنسان. وهذا يشمل، في جملة أمور، إجراء مشاورات عامة والاضطلاع بتقييمات الأثر على حقوق الإنسان أو بتقييمات أثر خوارزميات الحاسوب على الوكالات العامة قبل شراء أو نشر منظومات الذكاء الاصطناعي. وينبغي إيلاء اهتمام خاص إلى التباين في تأثير هذه التكنولوجيات على الأقليات العرقية والدينية، والنشطاء السياسيين والمعارضة السياسية. وينبغي أن يخضع نشر الحكومات لنظم الذكاء الاصطناعي لمراجعة منتظمة من قبل خبراء خارجيين ومستقلين.

٦٣ - وينبغي أن تضمن الدول بأن تكون حقوق الإنسان المحور الذي يستند إليه تصميم القطاع الخاص لنظم الذكاء الاصطناعي ونشرها وتنفيذها. ويشمل ذلك تحديث وتطبيق اللوائح التنظيمية القائمة، ولا سيما لوائح حماية البيانات على مجال الذكاء الاصطناعي، ومتابعة الخطط التنظيمية والمشاركة التنظيم التي تهدف إلى إلزام أوساط الأعمال بإجراء تقييمات الأثر وعمليات مراجعة تكنولوجيات الذكاء الاصطناعي وضمان فعالية آليات المساءلة الخارجية^(٥٧). وقد يكون التنظيم القطاعي لتطبيق معين من تطبيقات الذكاء الاصطناعي، عند الاقتضاء، ضرورياً وفعالاً لحماية حقوق الإنسان. وبقدر ما تتدخل مثل هذه القيود بحرية التعبير أو تُسهّل التدخل بها، يجب على الدول أن تضمن بأن تكون تلك القيود ضرورية ومتناسبة مع تحقيق هدف مشروع وفقاً لما هو منصوص عليه في المادة ١٩ (٣) من العهد. وينبغي أن يتم وضع اللوائح التنظيمية المتصلة بالذكاء الاصطناعي من خلال مشاورات عامة واسعة النطاق تشمل التفاعل مع المجتمع المدني، وجماعات حقوق الإنسان وممثلي الفئات الممثلة تمثيلاً ناقصاً أو المهمشة من المستخدمين النهائيين.

٦٤ - وينبغي أن تهيئ الدول بيئة سياسية وتشريعية تفضي إلى بيئة معلومات متنوعة وتعددية. ويشمل ذلك اتخاذ تدابير لكفالة وجود ميدان تنافسي في مجال الذكاء الاصطناعي. ويمكن أن تشمل هذه التدابير تنظيم الاحتكارات التكنولوجية لمنع تركيز خبرات وسلطات الذكاء الاصطناعي في أيدي عدد قليل من الشركات المهيمنة، ووضع لوائح تنظيمية ترمي إلى زيادة التشغيل المشترك للخدمات والتكنولوجيات، واعتماد سياسات تدعم حياد الشبكات وحياد الأجهزة التكنولوجية^(٥٨).

التوصيات الموجهة إلى الشركات

٦٥ - ينبغي لجميع الجهود الرامية إلى إعداد مبادئ توجيهية أو قوانين بشأن الآثار الأخلاقية المترتبة على تكنولوجيات الذكاء الاصطناعي أن تتأسس على مبادئ حقوق الإنسان. ويجب على القطاعين الخاص والعام، عند كل تطوير ونشر للذكاء الاصطناعي، أن يوفر للمجتمع المدني فرصاً للتعليق عليه. وينبغي للشركات أن تكرر في السياسات المؤسسية والتوجيهات التقنية

(٥٧) واغتر، "الأخلاقيات للفرار من اللوائح Ethics as an escape from regulation".

(٥٨) هيئة تنظيم الاتصالات والوظائف الإلكترونية، الأجهزة الإلكترونية، الحلقة الضعيفة في تحقيق شبكة إنترنت مفتوحة (٢٠١٨). متاح على الرابط التالي: https://www.arcep.fr/uploads/tx_gspublication/rapport-terminaux-fev2018-ENG.pdf

الموجهة للمهندسين والتقنيين، ومطوّري البيانات وماسحي البيانات، والمبرمجين وكل من يشارك في دورة عمر الذكاء الاصطناعي ضرورة أن يسترشدوا في تسيير جميع عملياتهم بالمسؤوليات المتعلقة بحقوق الإنسان وبأن المبادئ الأخلاقية يمكن أن تساعد في هذا المجال من خلال تيسير تطبيق مبادئ حقوق الإنسان على حالات محددة من عمليات تصميم الذكاء الاصطناعي ونشره وتنفيذه. وعلى وجه الخصوص، ينبغي أن تستند شروط الخدمة المتعلقة بالمنصات الإلكترونية إلى مبادئ حقوق الإنسان العالمية.

٦٦ - ويجب على الشركات أن توضح مكان وكيفية استخدام تكنولوجيات الذكاء الاصطناعي وتقنيات التشغيل الآلي على منصاتها الإلكترونية وخدماتها وتطبيقاتها. ويُعتبر استخدام وسائل مبتكرة لإبلاغ الأفراد عندما يكونون خاضعين لعملية اتخاذ القرار المبني على الذكاء الاصطناعي، أو عندما يؤدي الذكاء الاصطناعي دوراً في عرض أو تنقيح محتوى البيانات الشخصية، أو عندما يتم إدماج البيانات الشخصية للأفراد في مجموعة بيانات ستُستخدم في إبلاغ نظم الذكاء الاصطناعي، أمراً بالغ الأهمية بالنسبة لإعطاء المستخدمين الإخطار اللازم لفهم ومعالجة أثر الذكاء الاصطناعي على تمتعهم بحقوق الإنسان. وينبغي للشركات أن تنشر أيضاً بيانات عن إزالة المحتوى، بما في ذلك مدى تكرار التنازع على عمليات إزالة المحتوى ومدى تكرار رفض الطعن بالإزالة، وأن تنشر كذلك بيانات عن الاتجاهات في عرض المحتوى، إلى جانب دراسات الحالات الفردية والتنقيف بشأن التمييز التجاري والسياسي.

٦٧ - ويجب على الشركات منع التمييز والمحاسبة عليه على مستوى كل من مدخلات ونواتج نظم الذكاء الاصطناعي. وهذا ينطوي على ضمان أن تعكس الأفرقة التي تقوم بتصميم ونشر نظم الذكاء الاصطناعي مواقف متنوعة وغير تمييزية وأن تعطي الأولوية لتفادي التحيز والتمييز في اختيار مجموعات البيانات وفي تصميم النظام، بما في ذلك عن طريق معالجة الأخطاء التي تجري عند أخذ العينات، وتنقية مجموعات البيانات لإزالة البيانات التمييزية ووضع تدابير للتعويض عن تلك البيانات. ويُعدّ الرصد الفعال للنتائج التمييزية المنبثقة من نظم الذكاء الاصطناعي أمراً أساسياً أيضاً.

٦٨ - وينبغي الاضطلاع بتقييمات الأثر على حقوق الإنسان وبالمشاورات العامة خلال مرحلتَي تصميم ونشر نظم الذكاء الاصطناعي الجديدة، بما في ذلك نشر النظم القائمة في الأسواق العالمية الجديدة. وينبغي أن تعقد المشاورات العامة ومشاركة الجمهور قبل وضع الصيغة النهائية للمنتج أو الخدمة أو البدء في تنفيذهما، وذلك لكفالة أن تكون المشاورات مجددة، وينبغي أن تشمل المشاورات العمل مع المجتمع المدني والمدافعين عن حقوق الإنسان وممثلي الفئات الممثلة تمثيلاً ناقصاً أو المهمشة من المستخدمين النهائيين. وينبغي أن تكون نتائج تقييمات الأثر على حقوق الإنسان والمشاورات العامة علنية.

٦٩ - وينبغي للشركات أن تجعل جميع مدونات الذكاء الاصطناعي قابلة للتدقيق بشكل كامل، كما ينبغي لها أن تتبع وسائل مبتكرة من شأنها تمكين التدقيق الخارجي والمستقل لنظم الذكاء الاصطناعي، بمعزل عن المتطلبات التنظيمية. وينبغي أن توضع نتائج عمليات مراجعة الذكاء الاصطناعي ذاتها في متناول الجمهور.

٧٠ - ويجب أن يحصل فرادى المستخدمين على إمكانية الوصول إلى سبل الانتصاف من الآثار الضارة المترتبة على نظم الذكاء الاصطناعي في مجال حقوق الانسان. وينبغي للشركات أن تضع أنظمة تُمكن البشر من مراجعة شكاوى جميع المستخدمين وتوفير سبيل الانتصاف للرد عليها وكذلك على الطعون المقدمة ضد النظم المبنية على الذكاء الاصطناعي، في الوقت المناسب. وينبغي أن تنشر بانتظام بيانات عن وتيرة إخضاع نظم الذكاء الاصطناعي للشكاوى وعن طلبات الحصول على الانتصاف، فضلا عن أنواع سبل الانتصاف المتاحة وفعاليتها.