

Distr.: General
3 June 2024
Arabic
Original: English

الجمعية العامة



مجلس حقوق الإنسان

الدورة السادسة والخمسون

18 حزيران/يونيه - 14 تموز/يوليه 2024

البند 9 من جدول الأعمال

العنصرية والتمييز العنصري وكره الأجانب وما يتصل بذلك من

أشكال التعصب: متابعة وتنفيذ إعلان وبرنامج عمل ديربان

الأشكال المعاصرة للعنصرية والتمييز العنصري وكره الأجانب وما يتصل بذلك من تعصب

تقرير المقررة الخاصة المعنية بالأشكال المعاصرة للعنصرية والتمييز العنصري وكرهية
الأجانب وما يتصل بذلك من تعصب، أشويني ك. ب. *

موجز

في هذا التقرير، توجز المقررة الخاصة المعنية بالأشكال المعاصرة للعنصرية والتمييز العنصري وكره الأجانب وما يتصل بذلك من تعصب، أشويني ك. ب.، ما اضطلعت به من أنشطة خلال العام الماضي وتتناول الطرق التي يسمح بها الافتراض السائد المتمثل في أن التكنولوجيا تتسم بالموضوعية والحياد للذكاء الاصطناعي بإدامة التمييز العنصري. وتبحث المقررة الخاصة أربعة طرق متداخلة يمكن للذكاء الاصطناعي أن يسهم من خلالها في مظاهر التمييز العنصري، وهي: مشاكل البيانات، والمسائل ذات الصلة بتصميم الخوارزميات، والاستخدام التمييزي المتعمد للذكاء الاصطناعي، وقضايا المساءلة. ثم تقدم أمثلة على تطبيق الذكاء الاصطناعي عبر مختلف المجالات المجتمعية وآثاره التمييزية العنصرية. وتحلل المقررة الخاصة الجهود الناشئة الرامية إلى إدارة وتنظيم الذكاء الاصطناعي قبل تقديم لمحة عامة عن المعايير ذات الصلة التي ينص عليها القانون الدولي لحقوق الإنسان. وتختتم بتقديم توصيات بشأن السبل التي ينبغي بها للدول أن تتناول إدارة وتنظيم تكنولوجيات الذكاء الاصطناعي لمنع التمييز العنصري والتصدي له.

* أُنق على نشر هذه الوثيقة بعد تاريخ النشر المعتاد لظروف خارجة عن إرادة الجهة التي قدمتها.



الرجاء إعادة الاستعمال

أولاً - مقدمة

- 1- يقدّم هذا التقرير عملاً بقرار مجلس حقوق الإنسان 36/52، الذي طلب فيه المجلس إلى المقررة الخاصة المعنية بالأشكال المعاصرة للعنصرية والتمييز العنصري وكره الأجانب وما يتصل بذلك من تعصب أن تقدم إليه تقريراً سنوياً. وتصف المقررة الخاصة في التقرير الأنشطة التي اضطلعت بها في إطار ولايتها وتبحث موضوع الذكاء الاصطناعي والتمييز العنصري.
- 2- وقد وجهت المقررة الخاصة إلى الدول الأعضاء في الأمم المتحدة والجهات المعنية الأخرى، بما في ذلك منظمات المجتمع المدني والمنظمات الدولية والمؤسسات الوطنية لحقوق الإنسان، نداءً لتقديم مساهمات بغية الاسترشاد بها في إعداد هذا التقرير. وتعرب المقررة الخاصة عن خالص امتنانها لجميع الدول الأعضاء والجهات المعنية الأخرى التي قدمت معلومات. فقد اعتمدت المقررة الخاصة على المساهمات الواردة في إعداد هذا التقرير ولا تزال مستعدة لإجراء حوار مستمر مع جميع الجهات المعنية بشأن هذا الموضوع الهام⁽¹⁾.

ثانياً - موجز الأنشطة

- 3- في تشرين الأول/أكتوبر 2023، قدمت المقررة الخاصة إلى الجمعية العامة، في دورتها الثامنة والسبعين⁽²⁾، تقريرها عن محاربة تمجيد النازية والنازية الجديدة والممارسات الأخرى التي تساهم في إثارة الأشكال المعاصرة من العنصرية والتمييز العنصري وكرهية الأجانب وما يتصل بذلك من تعصب، وعن خطاب الكراهية العنصرية على شبكة الإنترنت. وفي الفترة ما بين 31 تشرين الأول/أكتوبر و14 تشرين الثاني/نوفمبر 2023، أجرت المقررة الخاصة زيارة قطرية إلى الولايات المتحدة الأمريكية⁽³⁾.
- 4- وشاركت المقررة الخاصة في الدورة التاسعة لفريق الخبراء البارزين المستقلين المعني بتنفيذ إعلان وبرنامج عمل ديربان، المعقودة في آب/أغسطس 2023. وفي كانون الثاني/يناير 2024، حضرت المقررة الخاصة الاجتماع الإقليمي لمنطقة آسيا والمحيط الهادئ بشأن العقد الدولي للمنحدرين من أصل أفريقي. وفي شباط/فبراير 2024، حضرت المقررة الخاصة المؤتمر الدولي حول العدالة الغذائية من منظور حقوق الإنسان، الذي عُقد في قطر تحت شعار "تحديات الواقع ورهانات المستقبل". وفي نيسان/أبريل 2024، حضرت الدورة الثالثة للمنتدى الدائم المعني بالمنحدرين من أصل أفريقي، حيث قدمت عرضاً عن التغلب على العنصرية النظامية والضرر التاريخي في مجال التعليم.

ثالثاً - الذكاء الاصطناعي والتمييز العنصري

- 5- اختارت المقررة الخاصة التركيز، في هذا التقرير، على الذكاء الاصطناعي والتمييز العنصري. ويتمشى هذا الموضوع مع تركيزها الاستراتيجي على العلاقة بين التكنولوجيات الرقمية والتمييز العنصري، على النحو المبين في تقريرها المقدم إلى مجلس حقوق الإنسان، في دورته الثالثة والخمسين، الذي يبين

(1) تلقت المقررة الخاصة دعماً في مجالي البحث والتحليل من المركز الاستشاري الدولي لحقوق الإنسان التابع لكلية الحقوق بجامعة هارفارد ومن المركز الاستشاري الدولي لحقوق الإنسان وتسوية النزاعات ومركز ستانفورد للعدالة العرقية التابع لكلية الحقوق بجامعة ستانفورد. وهي تعرب عن خالص شكرها لجميع من ساهموا بدعمهم القيم في إعداد التقرير.

(2) A/78/302 وA/78/538.

(3) انظر A/HRC/56/68/Add.1.

رؤيتها الاستراتيجية وأولوياتها الأولية⁽⁴⁾. وتستند المقرر الخاصة إلى ما اضطلعت به المكلفة السابقة بالولاية من عمل بشأن التكنولوجيات الرقمية الناشئة والتمييز العنصري⁽⁵⁾ وتستجيب لاهتمام مجلس حقوق الإنسان ومنظومة الأمم المتحدة عموماً بحوكمة الذكاء الاصطناعي⁽⁶⁾.

6- ولا تزال التطورات الأخيرة في مجال الذكاء الاصطناعي التوليدي وأوجه التطبيق المتزايد للذكاء الاصطناعي تثير شواغل جدية في مجال حقوق الإنسان، بما في ذلك مخاوف من التمييز العنصري. فالذكاء الاصطناعي التوليدي آخذ في تغيير العالم ولديه القدرة على إحداث تحولات مجتمعية جذرية بشكل متزايد في المستقبل. وتعرب المقررة الخاصة عن قلقها البالغ من تسارع وتيرة انتشار تطبيق الذكاء الاصطناعي في مختلف المجالات. ولا يعني هذا أنه ليس للذكاء الاصطناعي فوائد محتملة. والواقع أنه يمكن أن يتيح فرصاً للابتكار والإدماج. غير أن هذه التكنولوجيات تنمو وتتطور بلا حسيب ولا رقيب إلى حد كبير. وتعرب المقررة الخاصة عن قلقها لأن التدابير السياسية والقانونية التي تسعى إلى إدارة الذكاء الاصطناعي وتنظيمه لا تواكب وتيرة نمو هذه التكنولوجيات، ولأن الجهود الناشئة الرامية إلى إدارته وتنظيمه لا تولي اهتماماً كافياً لما ترخر به هذه التكنولوجيات من قدرات حالية هائلة واحتمال أن تؤدي في المستقبل إلى إدامة التمييز العنصري النظامي وعميقه، فضلاً عن توسيع رقعة انعدام المساواة داخل المناطق والبلدان والمجتمعات وفيما بينها.

7- وعلى نحو ما أوضحتها المكلفة السابقة بالولاية، هناك اعتقاد ضار سائد مفاده أن التكنولوجيات محايدة وموضوعية:

وعادة ما يتصور عامة الناس أن التكنولوجيات محايدة وموضوعية بطبيعتها، وقد أشار البعض إلى أن افتراض الموضوعية والحياد التكنولوجيين ما زال شائعاً حتى في أوساط منتجي التكنولوجيات. بيد أن التكنولوجيات ليست محايدة بتاتاً. فهي تعكس قيم ومصالح الأشخاص الذين يؤثرون في تصميمها واستخدامها، وهي تتشكل أساساً في ضوء ذات هياكل عدم المساواة التي تحكم المجتمع⁽⁷⁾.

8- وتتناول المقررة الخاصة في هذا التقرير السبل التي يسمح بها الافتراض السائد أن التكنولوجيات موضوعية ومحايدة للذكاء الاصطناعي بإدامة التمييز العنصري.

ألف - الطرق المتداخلة التي يمكن للذكاء الاصطناعي أن يسهم بها في مظاهر التمييز العنصري

9- إن الذكاء الاصطناعي ليس نوعاً واحداً. والواقع أن هناك أنواعاً عدة من الذكاء الاصطناعي. ويعتبر الذكاء الاصطناعي التنبؤي شكلاً "تقليدياً" من أشكال هذه التكنولوجيات، وتستخدم نماذج البيانات والأنماط والاتجاهات السائدة منذ وقت طويل في استنباط تنبؤات مستنيرة بالأحداث أو النتائج المستقبلية.

(4) A/HRC/53/60، الفقرات من 50 إلى 53.

(5) انظر A/75/590 و A/HRC/44/57 و A/HRC/48/76.

(6) انظر، على سبيل المثال، High-level Advisory Body on Artificial Intelligence، "Interim report: governing AI for humanity" (December 2023)؛ وقرار مجلس حقوق الإنسان 29/53؛ وقراري الجمعية العامة 213/78 و 265/78.

(7) A/HRC/44/57، الفقرة 12.

10- أما الذكاء الاصطناعي الذي يُستخدم للتعرف على الأحرف المطبوعة والوجوه البشرية والأشياء والمعلومات الأخرى، فهو شكل آخر من أشكال الذكاء الاصطناعي "التقليدي"، ويشمل تقنيات مختلفة للتعرف على الأشياء والأفراد والأنماط وتمييزها في البيانات التي يُزود بها.

11- وأما نظم الذكاء الاصطناعي التوليدي، فهي أشكال جديدة من الذكاء الاصطناعي. وتتسم بتعدد استخداماتها وإمكانية استخدامها في طائفة من الأغراض. وهي تشمل فئة من نظم الذكاء الاصطناعي المصممة لإنتاج محتوى متنوع استناداً إلى مجموعات البيانات الواسعة النطاق التي تستخدم في تدريبها، وشبكات العصبية، وخاصيات التعلم العميق المدمجة فيها، والتعليمات التي تتلقاها من مستخدميها. ويمكن لنماذج الذكاء الاصطناعي التوليدي أن تنتج محتوى متنوعاً للغاية يشمل الصور والنصوص والصوت والفيديو والبيانات الاصطناعية. وعلى العكس من نماذج الذكاء الاصطناعي التي تركز على التعرف على الأنماط في البيانات الموجودة، يُدرَّب الذكاء الاصطناعي التوليدي على استحداث نقاط بيانية جديدة تحاكي أنماط وخصائص البيانات المستخدمة في تدريب نماذج التعلم الآلي. وسيؤدي ظهور الذكاء الاصطناعي التوليدي إلى العديد من التطبيقات الجديدة، كما سي طرح العديد من الأسئلة الجديدة المتعلقة بحقوق الإنسان⁽⁸⁾.

12- وهناك تطبيقات عديدة لأنواع الذكاء الاصطناعي المختلفة هذه. وتتناول المقررة الخاصة بإسهاب، في الفقرات التالية، المزيد من الأمثلة المحددة لاستخدامات الذكاء الاصطناعي وما يرتبط بها من آثار على صعيد التمييز العنصري. غير أنها تود أن تشدد على أن من المهم للغاية دراسة القواسم المشتركة بين الطرق التي يمكن أن يدمج بها الذكاء الاصطناعي التمييز العنصري، ولا سيما في إطار المناقشات القانونية والسياساتية المتعلقة بإدارته وتنظيمه. وفي سياق هذه المناقشات، ينبغي تناول آثار الذكاء الاصطناعي من منظور العنصرية النظامية التي عُرِّفت على أنها "العمل بنظام معقد ومتشابك من القوانين والسياسات والممارسات والمواقف في مؤسسات الدولة والقطاع الخاص والهيكل المجتمعية التي تؤدي، مجتمعة، بشكل مباشر أو غير مباشر، مقصود أو غير مقصود، بحكم القانون أو بحكم الواقع، إلى التمييز أو الاستبعاد أو التقييد أو التفضيل على أساس العرق أو اللون أو النسب أو الأصل القومي أو الإثني"⁽⁹⁾. وعلى نحو ما يتجلى في هذا التعريف، فإن العنصرية النظامية ظاهرة معقدة غالباً ما تكون خبيثة ومنتشرة على نطاق المجتمع. وتتسم مظاهر العنصرية النظامية في كل مجال بترابطها وتشابكها وتعزيزها بعضها بعضاً وتأزرها مع مظاهر العنصرية النظامية في المجالات الأخرى. ومن شأن بحث الطرق المتداخلة التي يسهم بها الذكاء الاصطناعي في التمييز العنصري أن يساعد في تحديد الطرق التي يتفاعل بها الذكاء الاصطناعي مع مظاهر العنصرية النظامية ويوطدها، والطرق التي يعزز بها بشكل كلي القمع المنهجي في المجتمع على أسس عرقية وإثنية⁽¹⁰⁾.

1- مشاكل البيانات

13- أدى ظهور نظم الذكاء الاصطناعي وخوارزميات التعلم الآلي إلى زيادة هائلة في رقمنة البيانات. وتستخدم الخوارزميات البيانات في اتخاذ القرارات والمشاركة في الإجراءات المتخذة عبر العديد من القطاعات. غير أن مجموعات البيانات التي تُدرَّب الخوارزميات عليها غالباً ما تتسم بعدم اكتمالها وعدم تمثيلها فئات معينة من الأشخاص تمثيلاً كافياً. فإذا كانت فئات معينة ممثلة تمثيلاً زائداً أو ناقصاً

(8) المساهمة الواردة من اللجنة الوطنية لحقوق الإنسان في أستراليا. وستُنشر جميع المساهمات على الموقع الشبكي لمفوضية الأمم المتحدة السامية لحقوق الإنسان.

(9) A/HRC/47/53، الفقرة 9.

(10) A/HRC/44/57، الفقرة 43.

في مجموعات البيانات التي تُدرّب الخوارزميات عليها، بما في ذلك على أسس عرقية وإثنية، فإنه يمكن أن ينتج عن ذلك تحيز خوارزمي. وعلى ذات المنوال، إذ كانت مجموعات البيانات التي تُدرّب الخوارزميات عليها تتضمن بيانات متحيزة بالفعل، فإنها يمكن أن تنتج نتائج متحيزة.

14- وإذا كانت مجموعات البيانات التي تدرّب الخوارزميات عليها غير كافية، فإن الخوارزميات قد تنتج تنبؤات تمييزية بشكل منهجي في حق الفئات غير الممثلة أو الممثلة تمثيلاً ناقصاً في مجموعات البيانات تلك. ولا يحدث التحيز الخوارزمي بسبب قلة البيانات فحسب، فالخوارزميات التي تستند إلى بيانات لا تمثل فئات معينة يمكن أن تنتج أيضاً نتائج متحيزة. فعلى سبيل المثال، أظهرت دراسة ركزت على قواعد بيانات الصور التابعة لأجهزة إنفاذ القانون في الولايات المتحدة أن المنحدرين من أصل أفريقي أكثر عرضة للاستهداف عن طريق الخطأ في الشبكات التي يستخدمها موظفو إنفاذ القانون للتعرف على الأشخاص من سمات وجوههم. ويعزى ذلك إلى أخطاء في خاصية التعرف على الأشخاص المنتمين إلى تلك الفئة من سمات وجوههم والتمثيل المفرط للمنحدرين من أصل أفريقي في قواعد بيانات الصور الفوتوغرافية التابعة للشرطة، مما يعكس أنماط التمييز النظامي التاريخية⁽¹¹⁾.

15- ويمكن أن تؤثر التحيزات التاريخية على البيانات نفسها. ويتمثل أحد العناصر الأساسية للتعلم الآلي في استنباط تنبؤات بالمستقبل استناداً إلى بيانات مستمدة من الماضي. وإذا كانت البيانات المستمدة من الماضي متحيزة ضد فئات معينة، بما في ذلك على أسس عرقية وإثنية، فإن نماذج الذكاء الاصطناعي يمكن أن تعيد إنتاج هذه التحيزات وتضخمها. ويمكن أن يؤدي استخدام البيانات المتحيزة أو المعيبة للاسترشاد بها في اتخاذ قرارات في الحياة اليومية إلى زيادة استهداف الفئات العرقية والإثنية المهمشة وإلحاق الضرر بها لأن استخدام هذه البيانات في سياق الذكاء الاصطناعي يولد المزيد من البيانات، التي يمكن استخدامها بعد ذلك للاسترشاد بها في اتخاذ القرارات المستقبلية. ويمكن لهذه النظم المزودة بخاصية التعزيز الذاتي أن تكرر الفوارق القائمة وتعمقها.

16- أما المشكلة الأخيرة من مشاكل البيانات، فهي مشكلة الخصوصية. فغالباً ما تتضمن البيانات المستخدمة في نظم الذكاء الاصطناعي المعلومات الشخصية للأفراد الذين تخصم تلك البيانات. غير أن جمع البيانات ومعالجتها من دون موافقة أصحابها ينتهك الحق في الخصوصية. وهناك أيضاً حالات يجري فيها تقاسم البيانات التي تجمع في أحد السياقات، مثل الرعاية الصحية، بما في ذلك من خلال استخدام تطبيقات الرعاية الصحية، من دون موافقة أصحابها، لاستخدامها في سياقات أخرى، كإنفاذ القانون مثلاً. وتثير عمليات اختراق البيانات والوصول غير المصرح به إلى المعلومات الشخصية من خلال القرصنة مخاوف إضافية فيما يتعلق بالخصوصية. وقد يؤدي ذلك إلى تقاوم شواغل حقوق الإنسان المتعلقة بحق الأشخاص المنتمين إلى الفئات المهمشة عنصرياً في الخصوصية. ويمكن أن تعرض انتهاكات الخصوصية تلك الفئات لخطر النبذ أو التمييز أو تشكل خطراً على سلامتها⁽¹²⁾.

2- مشاكل تصميم الخوارزميات

17- هناك شكل ثان من أشكال التحيز في أدوات الذكاء الاصطناعي، وهو شكل ينشأ عن الطريقة التي تُصمم بها الخوارزميات. فإذا كان التحيز مضمناً في خيارات التصميم، فإن الخوارزمية يمكن أن تسهم في توليد نتائج متحيزة، حتى إذا كانت البيانات التي تُزود بها الخوارزمية ممثلة تماماً لجميع الفئات.

(11) Nicol Turner Lee, Paul Resnick and Genie Barton, "Algorithmic bias detection and mitigation: best practices and policies to reduce consumer harms", Brookings Institution, 22 May 2019

(12) Samantha Lai and Brooke Tanner, "Examining the intersection of data privacy and civil rights", Brookings Institution, 18 July 2022. انظر أيضاً المساهمة المقدمة من المنظمة الدولية لحماية الخصوصية.

ويمكن أن تؤدي القرارات المتعلقة بمعايير الخوارزمية وعملها إلى إدراج تحيزات في الخوارزمية. ذلك أن مصممي الخوارزميات يتخذون قرارات بشأن المتغيرات التي ستستخدمها الخوارزمية، وكيفية تحديد فئات أو نطاقات المعلومات والبيانات التي ستستخدم في تصميمها. وتشمل الخيارات التي يتخذها المصممون كيفية قياس خصائص محددة وتحديد مقاييس نجاح الخوارزمية. وفي بعض الأحيان، قد يعتمد مصممو الخوارزميات، بسبب خلفياتهم أو وجهات نظرهم، إلى إدراج تحيزات لا واعية، بما فيها تحيزات عرقية، في تصميم خوارزمياتهم⁽¹³⁾. وتفيد التقارير بأن افتقار قطاعات التكنولوجيا الرقمية إلى التنوع يفاقمه عدم استناد تطوير نظم الذكاء الاصطناعي إلى عمليات تشاور شاملة للجميع، مما يسهم في مشاكل تصميم الخوارزميات⁽¹⁴⁾.

18- ويمكن أن تتجم عن خيارات تصميم الخوارزميات آثار تمييزية بالغة على أرض الواقع. فعلى سبيل المثال، قد تؤدي الطريقة التي يتم بها تعريف "المخاطر" وقياسها إلى نتائج تمييزية عند تصميم خوارزمية لتقييم مخاطر الإقراض. فإذا قرر مصمم الخوارزمية استخدام تقديرات الجدارة الائتمانية كبديل للمخاطر، فإن ذلك قد يؤدي إلى نتائج تمييزية ضد مجموعات الأشخاص الذين عادة ما تكون تقديرات جدارتهم الائتمانية ضعيفة. وقد أظهرت البحوث أنه يمكن أن يكون هناك ارتباط قوي بين تقدير الجدارة الائتمانية والعرق والمؤشرات الديمغرافية الأخرى وأن استخدام تقديرات الجدارة الائتمانية ينطوي على إجحاف في حق فئات معينة⁽¹⁵⁾. ويمكن اعتبار هذا الارتباط، في كثير من الحالات، بمثابة نتيجة عرضية للعنصرية والإقصاء التنظيميين القائمين. وقد يؤدي اختيار مصمم الخوارزمية استخدام تقديرات الجدارة الائتمانية في تقييم مخاطر الاقتراض إلى حرمان أفراد بعينهم من الحصول على هذه القروض، مع أنه ليس معياراً تمييزياً في ظاهره.

3- استخدام الذكاء الاصطناعي لأغراض تمييزية

19- يمكن استخدام الذكاء الاصطناعي، في بعض الحالات، لأغراض عنصرية صريحة من خلال استعماله استعمالاً انتقائياً ضد الفئات المستهدفة، مما يؤدي إلى نتائج تمييزية. فعلى سبيل المثال، تفيد تقارير بأن بعض أجهزة إنفاذ القانون تستخدم الذكاء الاصطناعي عمداً لرصد فئات معينة من المجتمع ومراقبتها مراقبة مفرطة على أسس تنطوي على تمييز عنصري⁽¹⁶⁾. وعلاوة على ذلك، يمكن أن يحدث التمييز المتعمد عندما تستغل الحكومات وغيرها من الجهات قدرات التكنولوجيا لرصد فئات معينة أو أفراد معينين وتصنيفهم واستهدافهم على أساس هوياتهم العرقية أو الإثنية⁽¹⁷⁾.

20- ويعد نشر المعلومات المضللة طريقة أخرى من الطرق التي يمكن من خلالها استخدام الذكاء الاصطناعي لأغراض عنصرية صريحة. ويمكن أن تستخدم الجهات الفاعلة السياسية الذكاء الاصطناعي في استحداث نصوص وصور ومقاطع فيديو بغرض التلاعب بالرأي العام والعمليات السياسية لصالحها

Ninareh Mehrabi and others, "A survey on bias and fairness in machine learning", *ACM Computing* (13)

Surveys, vol. 54, No. 6 (2022)؛ المساهمة الواردة من مؤسسة "London Story"؛ و *A/HRC/44/57*، الفقرة 17.

المساهمة الواردة من شبكة "NetMission.Asia". (14)

A.R. Lange and Natasha Duarte, "Understanding bias in algorithmic design", *Medium*, 6 September 2017 (15)

انظر *Amnesty International, Decode Surveillance NYC: Methodology* (London, 2022)؛ والمساهمة الواردة من شبكة "NetMission.Asia". (16)

المساهمة الواردة من شبكة "NetMission.Asia". (17)

وتقويض الثقة في المؤسسات، بما في ذلك على أسس عرقية. وتفيد تقارير أيضاً بأن بعض الحكومات استخدمت الذكاء الاصطناعي لبتث الفتنة وتيسير الرقابة على الإنترنت⁽¹⁸⁾.

4- مشاكل المساواة

21- تعني قدرة بعض أدوات الذكاء الاصطناعي على اتخاذ القرارات بشكل مستقل من دون تدخل بشري أن عملية صنع القرار تتم في الخفاء، أي كما لو كانت تتم في "صندوق أسود" معتم. وبالإضافة إلى ذلك، قد تتخذ الخوارزمية القرارات بشكل مستقل بمجرد وصولها إلى البيانات لأن خوارزميات الذكاء الاصطناعي تُحدّث نفسها تلقائياً باستمرار. وبمرور الوقت، قد تستخدم أداة الذكاء الاصطناعي، فيما تتخذ من قرارات، عوامل لم تكن الخوارزمية مبرمجة أصلاً للاعتماد عليها. بل تستمد الخوارزمية هذه العوامل من الأنماط التي حددتها الخوارزمية من تلقاء نفسها في البيانات. وبما أن الخوارزمية تدمج هذه الأنماط الجديدة في برمجتها وفي صنع القرارات، فإنه قد لا يعد بمقدور الأفراد الذين يعتمدون على الخوارزمية معرفة ما يجري في الخفاء وتحديد المعايير التي استخدمتها الخوارزمية في توليد نتائج معينة. وبناء على ذلك، فإن مشكلة عدم إمكانية معرفة ما يجري في الخفاء تجعل عملية اتخاذ القرار من جانب خوارزميات الذكاء الاصطناعي عملية خبيثة يكتنفها الغموض⁽¹⁹⁾. وبالإضافة إلى ذلك، لا يمكن التدقيق في العديد من الخوارزميات التي تطورها كيانات تجارية بسبب قوانين التعاقد والملكية الفكرية، مما يؤدي إلى مفارقة قضايا المساواة⁽²⁰⁾.

22- وتتطوي مشكلة انعدام الشفافية هذه على آثار تتعلق بوجه خاص بالعنصرية النظامية. وعلى النحو المبين أعلاه، فإن العنصرية النظامية هي آفة خبيثة ولكنها مدمرة للغاية على مستوى المجتمع. فالعنصرية النظامية، التي قلما يُعترف بالقوى المحركة لها، ظاهرة تقاومها الثغرات المستمرة في جمع البيانات المصنفة عرقياً وإثنية⁽²¹⁾. وبدون آليات مساواة فعالة، يتمتع الذكاء الاصطناعي بقدرة كبيرة على أن يكون عاملاً آخر من العوامل الموجبة لظاهرة العنصرية النظامية الخبيثة والمدمرة أصلاً.

23- وتؤثر قضايا المساواة المتعلقة بالذكاء الاصطناعي تأثيراً كبيراً على قدرة من يتعرضون لأفعال التمييز العنصري على التماس سبل انتصاف فعالة. واليوم، عندما يواجه الأشخاص الذين ينتمون إلى الفئات العرقية والإثنية المهمشة نتائج مختلفة بسبب القرارات التي يتخذها البشر، يمكن للمحاكم والآليات المساواة الأخرى أن تنظر فيما إذا كانت تلك الإجراءات متعمدة ومبررة⁽²²⁾. وعندما تكون تلك القرارات من صنع البشر، غالباً ما تكون هناك أدلة يمكن استخدامها لإجراء تقييمات من هذا القبيل. وفي كثير من الحالات، لا يمكن اقتفاء أثر القرارات التي تتخذها الآلة من تلقاء نفسها كما هو الحال بالنسبة للقرارات

Tate Ryan-Mosley, "How generative AI is boosting the spread of disinformation and propaganda", (18) *MIT Technology Review*, 4 October 2023.

Yavar Bathaee, "The artificial intelligence black box and the failure of intent and causation", (19) *Harvard Journal of Law and Technology*, vol. 31, No. 2 (2018) و A/HRC/44/57، الفقرة 34، و Renata

M. O'Donnell, "Challenging racist predictive policing algorithms under the Equal Protection Clause", *New York University Law Review*, vol. 94, No. 3 (June 2019).

A/HRC/44/57، الفقرة 44. (20)

A/HRC/47/53، الفقرة 16. (21)

Bathaee, "The artificial intelligence black box" (22)

التي يتخذها البشر⁽²³⁾. وستؤدي المسائل المتعلقة بعدم شفافية الذكاء الاصطناعي إلى تفاقم العقبات الجمة أصلاً التي تمنع الأشخاص الذين يتعرضون للتمييز العنصري من الاحتكام إلى القضاء.

باء - استخدام الذكاء الاصطناعي وتأثيره التمييزي

24- تقدم المقررة الخاصة في هذا الفرع أمثلة على استخدام الذكاء الاصطناعي عبر مختلف المجالات المجتمعية وآثاره التمييزية العنصرية. وهذه الأمثلة هي أمثلة توضيحية وغير حصرية وهي مقدمة كدليل واضح على أن الذكاء الاصطناعي يساهم بالفعل في التمييز العنصري. وترى المقررة الخاصة أن هذه الأمثلة هي مظاهر مترابطة ومتعاضدة للتمييز العنصري تسهم في تعزيز القمع النظامي الشامل على نطاق المجتمع، على أسس عرقية وإثنية.

25- وقد اختارت المقررة الخاصة ثلاثة مجالات لتقديم أمثلة على الأثر التمييزي للذكاء الاصطناعي، هي: إنفاذ القانون والأمن ونظام العدالة الجنائية؛ والتعليم؛ والرعاية الصحية. وفيما يتعلق باستخدام الذكاء الاصطناعي في السياقات الأخرى، توصي المقررة الخاصة بالاطلاع على ما قدمته المكلفة السابقة بالولاية من تقارير عن ظهور الحدود الرقمية، وعن تحديد معالم التمييز العنصري وكره الأجانب في إنفاذ إجراءات الحدود الرقمية والهجرة، وعن استخدام التكنولوجيات الرقمية في إنفاذ إجراءات الحدود والهجرة⁽²⁴⁾. وتحيل المقررة الخاصة القراء أيضاً إلى تقريرها المقدم إلى الجمعية العامة، في دورتها الثامنة والسبعين، عن خطاب الكراهية العنصرية على شبكة الإنترنت، الذي يتناول استخدام الذكاء الاصطناعي في إدارة محتوى وسائل التواصل الاجتماعي⁽²⁵⁾، كما تحيل إلى التقرير الذي قدمه المقرر الخاص المعني بالفقر المدقع وحقوق الإنسان إلى الجمعية العامة في دورتها الرابعة والسبعين، والذي يقدم تحليلاً لاستخدام الذكاء الاصطناعي في أنظمة الحماية الاجتماعية⁽²⁶⁾.

1- إنفاذ القانون، والأمن، ونظام العدالة الجنائية

(أ) التعرف الآلي

26- تستخدم أجهزة إنفاذ القانون أدوات التعرف الآلي لربط ما تلاحظه في بيئة معينة بما يحتمل أن "يطابقه" في قواعد بياناتها. ومن أكثر أنواع أدوات التعرف الآلي شيوعاً تقنية التعرف على الأشخاص من سمات وجوههم. وتلتقط أدوات التعرف على الأشخاص من سمات وجوههم صور فيديو أو صوراً فوتوغرافية لشخص ما ثم تغذي بها الخوارزميات. ثم تقارن الخوارزميات الصور بقاعدة بيانات الصور الفوتوغرافية التابعة للشرطة أو بصور رخصة القيادة أو غيرها من الصور الفوتوغرافية بهدف تحديد هوية الشخص المعني⁽²⁷⁾. ويدرب مصممو هذه الأدوات النماذج التي تستند إليها من خلال عرض صور وجوه عليها باستخدام عملية التعلم الآلي. ويتمثل الهدف من ذلك في تدريب تلك النماذج على التعرف على السمات المميزة للوجوه البشرية⁽²⁸⁾. غير أن مجموعات بيانات الصور المستخدمة في تدريب هذه النماذج

(23) المرجع نفسه.

(24) A/75/590 وA/HRC/48/76.

(25) A/78/538.

(26) A/74/493.

(27) Marissa Gerchick and Matt Cagle, "When it comes to facial recognition, there is no such thing as a magic number", American Civil Liberties Union, 7 February 2024.

(28) Julia Dressel and Andrew Warren, "Breaking down data analytics and AI in criminal justice", Recidiviz, 8 March 2022.

لا تمثل دوماً جميع فئات السكان⁽²⁹⁾. ففي دراسة تناولت إحدى أكثر قواعد بيانات الصور شعبية، خلص الباحثون إلى وجود إفراط في تمثيل الرجال الذين تتراوح أعمارهم بين 18 عاماً و40 عاماً ونقص في تمثيل الأشخاص ذوي البشرة الداكنة⁽³⁰⁾. ووفقاً لدراسة أخرى للنظم التجارية المزودة بخاصية التعرف على سمات الوجه، تُدرّب خوارزميات تصنيف النوع الجنساني على مجموعات بيانات تتألف أغلبيتها الساحقة من صور لوجوه ذكور بيض⁽³¹⁾. ويؤدي افتقار مجموعات البيانات المستخدمة لتدريب أدوات الذكاء الاصطناعي إلى التنوع العرقي والجنساني والثقافي إلى إحدى مشكلات البيانات التقليدية المبيّنة أعلاه. فكلما كانت الفئات السكانية ممثلة تمثيلاً ناقصاً في بيانات تدريب الخوارزمية، بما في ذلك الفئات المعرضة لأشكال التمييز المتقاطعة، زاد احتمال مطابقتها بشكل خاطئ بواسطة الخوارزمية.

27- وتفيد التقارير بأن حالات الخطأ في تعرف هذه التكنولوجيات على الأشخاص من سمات وجوههم أدت إلى زيادة عدد حالات اعتقال الأشخاص المنحدرين من أصل أفريقي⁽³²⁾. وقد أشار المقرر الخاص المعني بتعزيز وحماية الحق في حرية الرأي والتعبير ومفوض الأمم المتحدة السامي لحقوق الإنسان إلى أن أدوات التعرف على الأشخاص من سمات وجوههم كثيراً ما تسهم في التمييز غير القانوني والتصنيف العنصري⁽³³⁾. وقد نشرت أجهزة إنفاذ القانون في عدد من البلدان نظماً تكفل التعرف على الأشخاص من سمات وجوههم على الرغم من هذه الشواغل المتعلقة بحقوق الإنسان. فعلى سبيل المثال، تفيد التقارير بأن حكومة الهند استثمرت استثماراً كبيراً في هذه النظم. وتفيد التقارير بأن نظام التعرف على الوجوه الذي تستخدمه شرطة دلهي دقيق في 2 في المائة فقط من الحالات، وأنه يجعل الأقليات أكثر عرضة من غيرها لخطر التعرف خطأً على وجوه أفرادها واعتقالهم اعتقالاً باطلاً⁽³⁴⁾. وتفيد التقارير بأن موظفي إنفاذ القانون في البرازيل اتهموا أفراداً خطأً ثم اعتقلوهم اعتقالاً باطلاً باستخدام أدوات تعرف معيبة. ووفقاً لدراسة أجريت في عام 2019، ينحدر 90 في المائة من الأشخاص الذين اعتقلوا في المدن البرازيلية باستخدام تكنولوجيا التعرف على الأشخاص من سمات وجوههم من أصول أفريقية⁽³⁵⁾.

28- ونظم الكشف عن الطلقات النارية هي نوع شائع آخر من أدوات التعرف الآلي التي يستخدمها موظفو إنفاذ القانون في عدد من البلدان. ويُعرف أحد هذه النظم باسم "ShotSpotter"، وهو ينطوي على وضع مجسات تحتوي على ميكروفونات مزودة بالنظام العالمي لتحديد المواقع وبذاكرة وبخاصية معالجة البيانات وببطارية في الأحياء السكنية⁽³⁶⁾. وعندما تكتشف المجسات صوتاً يشبه في أنه صوت طلقة نارية، تُنلّت الخوارزمية إحدائيات موقع أي شيء يحتمل أن يكون قد تسبب في ذلك الصوت. وتبرز

(29) المساهمة الواردة من مبادرة "AI for the People".

(30) Khari Johnson, "ImageNet creators find blurring faces for privacy has a 'minimal impact on accuracy'", VentureBeat, 16 March 2021.

(31) Joy Buolamwini and Timnit Gebru, "Gender shades: intersectional accuracy disparities in commercial gender classification", *Proceedings of Machine Learning Research*, vol. 81 (2018).

(32) أيضاً Gerchick and Cagle, "When it comes to facial recognition, there is no such thing as a magic number"، والمساهمة الواردة من منظمة "Internet Lab".

(33) Gerchick and Cagle, "When it comes to facial recognition, there is no such thing as a magic number".

(34) انظر A/HRC/41/35 وA/HRC/48/31.

(35) منظمة العفو الدولية، "Ban the scan: Hyderabad"، متاح في: <https://banthescan.amnesty.org/hyderabad/>.

(36) المساهمة الواردة من فريق الخبراء البرازيلي.

(37) Alisha Ebrahimji, "Critics of ShotSpotter gunfire detection system say it's ineffective, biased and costly", CNN, 24 February 2024.

الخوارزمية الأصوات المحتملة الأخرى قبل إرسال الصوت إلى مُراجع بشري⁽³⁷⁾. وتشير المعلومات المتاحة إلى أن نظم الكشف عن الطلقات النارية تنتشر بشكل غير متناسب في الأحياء التي تسكنها الفئات المهمشة عنصرياً وأنها تتسم بمعدل خطأ مرتفع للغاية⁽³⁸⁾. إن وضع نظم الكشف عن الطلقات النارية في الأحياء التي تعيش فيها الفئات العرقية والإثنية المهمشة وعدم دقة نظم الكشف عن الطلقات النارية يفاقم التحيزات النظامية السائدة في أجهزة إنفاذ القانون.

29- وهناك أمثلة عديدة على السبل التي أدى بها استخدام تكنولوجيا التعرف الآلي على سمات الوجه إلى عواقب أدت إلى قلب حياة أشخاص رأساً على عقب. ففي عام 2019، في الولايات المتحدة، أفادت الأنباء بأن رجلاً أسود في نيويورك اعتقل اعتقالاً باطلاً واحتجز في السجن لمدة 10 أيام بسبب خطأ في التعرف على سمات وجهه. وعلى الرغم من وجود أدلة تثبت براءته، لم تسقط السلطات القضائية المرفوعة ضده إلا بعد مرور عام تقريباً، وواجه عقوبة السجن لمدة تصل إلى 25 عاماً بسبب التهم الموجهة إليه. وقد أثر الحادث تأثيراً شديداً على حياة الرجل⁽³⁹⁾. وأفادت الأنباء بأن موظفي إنفاذ القانون في شيكاغو أطلقوا النار، في شباط/فبراير 2024، على طفل كان يشعل ألعاباً نارية بعد استجابتهم لإنذار كاذب من نظام "ShotSpotter"⁽⁴⁰⁾. ومن الأمثلة الأخرى على استخدام هذا النوع من تكنولوجيا الذكاء الاصطناعي، ما أفادت به التقارير من اعتماد جيش الدفاع الإسرائيلي نظام "Wolf Pack"، وهو عبارة عن قاعدة بيانات هائلة تحتوي على صور لفلسطينيين من الضفة الغربية وعلى جميع المعلومات المتاحة عنهم، وتحتوي أيضاً على برامج مراقبة شتى مثل "Blue Wolf" و"Red Wolf"⁽⁴¹⁾. وتفيد التقارير بأن جيش الدفاع الإسرائيلي ركب كاميرات مزودة بخاصية الذكاء الاصطناعي وقادرة على التعرف على الأشخاص من سمات وجوههم في جميع أنحاء مدينة الخليل القديمة، وهي كاميرات مربوطة ببرنامج "Blue Wolf"، وهو تطبيق متاح على الهواتف المحمولة يسمح للجنود بالكشف عن الفلسطينيين وتصنيفهم في جميع أنحاء الضفة الغربية عن طريق قاعدة بيانات بيومترية هائلة لم يوافق معظمهم على تسجيل بياناتهم فيها، مما يعني أن الفلسطينيين ما زالوا يخضعون للمراقبة حتى الآن. فتطبيق نظام "Wolf Pack" تطبيقاً صارماً من جانب جيش الدفاع الإسرائيلي يفاقم الفصل العنصري الممارس ضد الفلسطينيين ويديمه⁽⁴²⁾. وتبين هذه الأمثلة أن استخدام نظم الذكاء الاصطناعي في اتخاذ قرارات لها تبعات في السياقات التي تتسم بمخاطر عالية له آثار خطيرة على حقوق الإنسان.

Jay Stanley, "Four problems with the ShotSpotter gunshot detection system", American Civil Liberties Union, 24 August 2021. (37)

MacArthur Justice Center, "ShotSpotter is deployed overwhelmingly in Black and Latinx neighborhoods in Chicago" المرجع نفسه؛ و <https://endpolicesurveillance.com/burden-on-Latinx-neighborhoods-in-Chicago-communities-of-color/>، متاح في: (38)

Gerchick and Cagle, "When it comes to facial recognition, there is no such thing as a magic number"؛ و Khari Johnson, "How wrongful arrests based on AI derailed 3 men's lives", *Wired*, 7 March 2022. (39)

Adam Schwartz, "Responding to ShotSpotter, police shoot at child lighting fireworks", *Electronic Frontier Foundation*, 22 March 2024. (40)

Amnesty International, *Automated Apartheid: How Facial Recognition Fragments, Segregates and Controls Palestinians in the OPT* (London, 2023), pp. 41–45. (41)

Sophia Goodfriend, "Algorithmic State violence: automated surveillance and Palestinian dispossession in Hebron's Old City", *International Journal of Middle East Studies*, vol. 55, No. 3 (2023). (42)

(ب) استخدام الخوارزميات التنبؤية في مجال ضبط الأمن

30- تعد الشرطة التنبؤية من أشكال تكنولوجيا الذكاء الاصطناعي التي يشجع استخدامها لدى أجهزة إنفاذ القانون. وتجري أدوات الشرطة التنبؤية تقييمات تسمح بالتنبؤ بالأشخاص الذين يحتمل أن يرتكبوا جرائم في المستقبل ومكان وقوع تلك الجرائم في المستقبل، استناداً إلى بيانات مواقعهم الجغرافية وبياناتهم الشخصية.

31- ويمكن أن تؤدي الشرطة التنبؤية إلى تفاقم إفراط الشرطة المعروف منذ أمد طويل في استهداف أحياء سكنية معينة استناداً إلى اعتبارات عرقية وإثنية لأغراض ضبط الأمن⁽⁴³⁾. وبما أن موظفي إنفاذ القانون كثيراً ما يركزون اهتمامهم على هذه الأحياء، فإن أفراد المجتمعات التي تقطن تلك الأحياء ممثلون تمثيلاً مفرطاً في سجلات الشرطة. ويؤثر هذا بدوره على الموقع الذي تنتبأ الخوارزميات بحدوث جريمة فيه مستقبلاً، مما يؤدي إلى زيادة نشر الشرطة في تلك المناطق⁽⁴⁴⁾. ويمكن أن تكون الشرطة التنبؤية أيضاً انعكاساً لجوانب مشكلة انعدام شفافية الخوارزميات، بما في ذلك من حيث طبيعة البيانات التي تخضع للتحليل وأساليب استخدام التنبؤات⁽⁴⁵⁾.

32- وتعتمد خوارزميات الشرطة التنبؤية القائمة على الموقع الجغرافي على العوامل التي تربط الأماكن والأحداث بالبيانات المتعلقة بالجرائم المرتكبة في الماضي في التنبؤ بمكان الجرائم المحتملة في المستقبل وزمانها⁽⁴⁶⁾. ثم تخطط قوات الشرطة لدورياتها وفقاً لذلك. وعندما يسجل الضباط في الأحياء التي تراقبها الشرطة مراقبة مفرطة جرائم جديدة، تنشأ حلقة تغذية راجعة تولد من خلالها الخوارزمية تنبؤات متحيزة تحيزاً متزايداً تجاه سكان هذه الأحياء. ويعني هذا، باختصار، أن التحيز في الماضي يؤدي إلى تحيز في المستقبل. وفي المملكة المتحدة لبريطانيا العظمى وأيرلندا الشمالية، أظهرت دراسة أجريت بتكليف من الحكومة عن التحيز الخوارزمي في أعمال الشرطة أن تحديد مواقع جغرافية معينة باعتبارها "بؤراً ساخنة" للجريمة يمكن أن يدفع الضباط إلى توقع ارتكاب المزيد من الجرائم في تلك المناطق. ونتيجة لذلك، فإن احتمال أن يكون الأشخاص الذين يتعرضون للتوقيف أو الاعتقال على أساس تحيز ضباط الشرطة أكبر من احتمال توقيفهم أو اعتقالهم على أساس مقتضيات السلامة العامة الفعلية⁽⁴⁷⁾. وفي أوروغواي، خلص الباحثون إلى أن البيانات المستخدمة في خوارزميات الشرطة التنبؤية القائمة على الموقع الجغرافي يمكن أن تكون متحيزة. ذلك أن متغير الموقع الجغرافي يمكن أن يستخدم كبديل للخلفية الاجتماعية والاقتصادية أو الإثنية، مما يؤدي إلى تمييز⁽⁴⁸⁾.

Jon Fasman, "The و Tim Lau, "Predictive policing explained", Brennan Center for Justice, 1 April 2020 (43)

.black box of justice: how secret algorithms have changed policing", *Fast Company*, 9 February 2021

Kristian Lum and William Isaac, "To predict and serve?", *Significance*, vol. 13, No. 5 (2016) (44)

والمساهمة الواردة من اللجنة الأسترالية لحقوق الإنسان.

Lau, "Predictive policing explained" (45)

Will Douglas Heaven, "Predictive policing algorithms are racist. They need to be dismantled", *MIT* (46)

.*Technology Review*, 17 July 2020

Government of the United Kingdom of Great Britain and Northern Ireland, المرجع نفسه. انظر أيضاً (47)

"Report commissioned by CDEI calls for measures to address bias in police use of data analytics",

.16 September 2019

Juan Ortiz Freuler and Carlos Iglesias, "Algorithms and artificial intelligence in Latin America: a (48)

study of implementation by governments in Argentina and Uruguay", World Wide Web Foundation,

September 2018; and Eticas Foundation, "Uruguay's Ministry of the Interior invests in predictive

.policing", 13 September 2021

33- وتوفر أدوات الشرطة التنبؤية القائمة على البيانات الشخصية طريقة للتنبؤ بالأشخاص الذين يحتمل أن يرتكبوا جرائم في المستقبل استناداً إلى البيانات الأساسية المتاحة عنهم. ويمكن أن تتضمن البيانات الأساسية سن الشخص ونوعه الجنساني وحالته الاجتماعية وسجل تعاطيه المخدرات وسجله الجنائي. وعلى غرار الأدوات القائمة على الموقع الجغرافي، يمكن أن تشوه البيانات المتاحة عن حالات الاعتقال السابقة، التي غالباً ما تكون مشوبة بالعنصرية النظامية في نظام العدالة الجنائية، التنبؤات المستقبلية التي تنتجها تلك الخوارزميات. ويمكن أن يؤدي استخدام متغيرات من قبيل الخلفية الاجتماعية والاقتصادية ومستوى التعليم والموقع الجغرافي بدلاً من العرق إلى إدامة التحيزات التاريخية⁽⁴⁹⁾. ففي أستراليا، استخدمت قوة شرطة نيو ساوث ويلز خطة قائمة على الخوارزميات لإدارة استهداف الأشخاص المشتبه فيهم من خلال تحديد الأشخاص الذين يحتمل أن يرتكبوا جرائم جنائية. وأدت التقارير التي أشارت إلى أن استخدام تلك الخطة تسبب في تفاعل الشرطة بشكل غير متناسب مع أفراد مجتمعات السكان الأصليين وسكان جزر مضيق توريس إلى وقف العمل بتلك الخطة⁽⁵⁰⁾.

(ج) خوارزميات تقييم احتمالات معاودة الإجرام

34- تُستخدم أدوات تقييم احتمالات معاودة الإجرام لتوجيه القرارات في مختلف مراحل نظام العدالة الجنائية، بما في ذلك فيما يتعلق بسندات الكفالة وإصدار العقوبات وقرارات الإفراج المشروط⁽⁵¹⁾. وتستند أدوات تقييم احتمالات معاودة الإجرام إلى البيانات التاريخية في تقييم مدى احتمال تصرف المدعى عليهم بطرق معينة، ولا سيما مدى احتمال ارتكابهم جرائم جديدة في المستقبل. وتنتج هذه الأدوات تقيماً للاحتتمالات استناداً إلى المعلومات المستمدة من مصادر من قبيل السجلات الجنائية والدراسات الاستقصائية للمدعى عليهم⁽⁵²⁾.

35- وتطرح أدوات التنبؤ باحتمالات معاودة الإجرام تحديات متعددة في مجال الذكاء الاصطناعي، وهي تحديات تساهم في التمييز العنصري. فأولاً، هناك مشاكل في البيانات التي تستند إليها هذه الأدوات. ذلك أن البيانات التي يستخدمها نظام العدالة الجنائية في تدريب تلك الخوارزميات تتسم بأوجه إجحاف منهجية تعزى إلى سلوك الشرطة العنصري منذ أمد طويل⁽⁵³⁾. وبالإضافة إلى ذلك، يمكن أن تسهم خيارات التصميم، مثل أساليب قياس المتغيرات أو تقييمها، في التمييز الخوارزمي⁽⁵⁴⁾. وعلاوة على ذلك، يمكن أن تؤثر المعايير التي يختار مصمم الخوارزمية تعريف "نجاح الخوارزمية" بناء عليها على الجوانب التي تسعى الخوارزمية إلى إبرازها. فإذا كانت الخوارزمية مبرمجة للتنبؤ بحد أدنى من الجرائم الجديدة، فإنها قد تربط طول أمد العقوبة بانخفاض احتمال معاودة الجرم، لأن الشخص لا يمكن أن يعاود الجرم وهو في السجن. ومن ثم، يمكن استخدام هذه الأنماط للتوصية بإصدار عقوبات أطول أمداً.

36- وأشار الباحثون إلى عدم دقة تنبؤات تلك الخوارزميات باحتمالات معاودة الجرم وإلى أن أخطاءها تؤثر على الفئات المهمشة عرقياً أكثر من غيرها. فعلى سبيل المثال، خلصت دراسة في الولايات

(49) Heaven, "Predictive policing algorithms are racist".

(50) المساهمة الواردة من اللجنة الأسترالية لحقوق الإنسان.

(51) Julia Angwin and others, "Machine bias", ProPublica, 23 May 2016.

(52) المرجع نفسه.

(53) Heaven, "Predictive policing algorithms are racist"; and Michael Mayowa Farayola and others, انظر

"Fairness of AI in predicting the risk of recidivism: review and phase mapping of AI fairness techniques", in *Proceedings of the 18th International Conference on Availability, Reliability and Security* (Association for Computing Machinery, 2023).

(54) Mehrabi and others, "A survey on bias and fairness in machine learning".

المتحدة إلى أن تقديرات احتمالات معاودة الجرم كانت غير موثوقة إلى حد كبير في تنبؤها بالجرائم العنيفة. وتقيد التقارير بأن معدل الخطأ في تصنيف الأشخاص المنحدرين من أصل أفريقي باعتبارهم مجرمين في المستقبل يفوق بمعدل الضعف معدل تصنيف البيض.

(د) منظومات الأسلحة الذاتية التشغيل

37- تشمل منظومات الأسلحة الذاتية التشغيل أي منظومات أسلحة مزودة بخاصية التشغيل الذاتي لوظائفها الحيوية، بما في ذلك الأسلحة الفتاكة الذاتية التشغيل والأسلحة الأقل فتكاً. وتستخدم هذه المنظومات في سياق إنفاذ القانون وكذلك في السياق العسكري، ولا تزال غير خاضعة للرصد إلى حد كبير. وهذه المنظومات قادرة على اختيار الأهداف واكتشافها وتحديدتها ومهاجمتها من دون تدخل بشري. ويُفعل السلاح الذاتي التشغيل بواسطة مجسات وبرامجيات تطابق الشخص المستهدف بـ "سمات الهدف" التي تحددها خوارزمية المنظومة. ولمنظومات الأسلحة الذاتية التشغيل آثار جسيمة للغاية على حقوق الإنسان، بما في ذلك فيما يتعلق بالحق في الحياة، وحظر التعذيب وغيره من ضروب المعاملة السيئة، وحق الفرد في الأمان على شخصه⁽⁵⁵⁾.

38- وقد أبلغت اللجنة الأولى للجمعية العامة بأن فرصة وضع ضوابط تكفل الحماية من مخاطر الأسلحة الذاتية التشغيل والتطبيقات العسكرية للدكاء الاصطناعي تكاد تضيع في الوقت الذي يستعد فيه العالم "لحدوث ثورة تكنولوجية"⁽⁵⁶⁾. وقد سبق للمقرر الخاص المعني بحالات الإعدام خارج القضاء أو بإجراءات موجزة أو تعسفاً أن أوصى بأن يدعو مجلس حقوق الإنسان جميع الدول إلى إعلان وتنفيذ وقف اختياري وطني على الأقل لاختبار الروبوتات الفتاكة المستقلة وإنتاجها وتجميعها ونقلها وحيازتها ونشرها واستخدامها⁽⁵⁷⁾.

39- وهناك خطر جدي أن يؤدي استخدام هذه المنظومات إلى تمييز عنصري جسيم، بل وفتاك في بعض الحالات. فعادة ما تشمل المعايير المستخدمة لاختيار الأهداف النوع الجنساني والسن والعرق⁽⁵⁸⁾. وتتضمن سمات الأهداف أيضاً معايير محايدة ظاهرياً، مثل الوزن أو البصمة الحرارية، ولكن الآلات غالباً ما تعكس تحيزات مبرمجيتها والمجتمع الذي ينتمون إليه. ويمكن برمجة تلك الأسلحة أيضاً من خلال تصنيف الأهداف عمداً وفقاً لسمات تمييزية⁽⁵⁹⁾. فعلى سبيل المثال، تقيد التقارير بأن إسرائيل تستخدم منظومات أسلحة فتاكة ذاتية التشغيل وشبه ذاتية التشغيل. وتقيد التقارير بأن ذلك يشمل استخدام الطائرات رباعية المرواح التي يتم التحكم فيها عن بعد لاستهداف الفلسطينيين، بالإضافة إلى نظم توليد الأهداف الآلي، التي تعمل بسرعة وحجم لا مثيل لهما، في إصدار "قوائم الاعتقالات"⁽⁶⁰⁾. وتقيد التقارير بأن

Amnesty International, "Autonomous weapons systems: five key human rights issues for consideration" (April 2015), p. 5 (55)

United Nations, "Without adequate guardrails, artificial intelligence threatens global security in evolution from algorithms to armaments, speaker tells First Committee", 24 October 2023 (56)

A/HRC/23/47، الفقرة 113. (57)

Ray Acheson, "Gender and bias", متاح في: <https://www.stopkillerrobots.org/wp-content/uploads/2021/09/Gender-and-Bias.pdf> (58)

Bonnie Docherty, "Expert Panel on the Social and Humanitarian Impact of Autonomous Weapons at the Latin American and Caribbean Conference on Autonomous Weapons", Human Rights Watch, 8 March 2023 (59)

Marwa Fatafta and Daniel Leufer, "Artificial genocidal intelligence: how Israel is automating human rights abuses and war crimes", Access Now, 9 May 2024 (60)

نظامي "Gospel" و"Lavender"، وهما نظامان من نظم تكنولوجيا الذكاء الاصطناعي يستخدمهما جيش الدفاع الإسرائيلي، قد كثفا مستويات الدمار في غزة، مما أسفر عن وقوع إصابات كبيرة، ولا سيما في أوساط النساء والأطفال الفلسطينيين⁽⁶¹⁾.

2- الرعاية الصحية

(أ) تقديرات المخاطر الصحية

40- يمكن استخدام الذكاء الاصطناعي في تقدير المخاطر الصحية واستعمال تلك التقديرات في أغراض شتى في مجال الرعاية الصحية، منها التشخيص الطبي وتخطيط الرعاية الصحية. ويمكن أن يؤدي تصميم هذه الخوارزميات والبيانات المستخدمة لتدريب نظم الذكاء الاصطناعي إلى آثار تمييزية في حال استخدام خوارزميات من هذا القبيل في تخصيص موارد الرعاية الصحية. وهناك حالات صُممت فيها خوارزميات من هذا القبيل لتحديد الأشخاص المؤهلين للحصول على رعاية إضافية، وذلك باستخدام تكاليف الرعاية الصحية السابقة بدلاً من الاحتياجات الطبية. ويمكن أن تتأثر البيانات التي تستند إليها هذه القرارات بعوامل من قبيل افتقار الأشخاص المنتمين إلى الفئات العرقية والإثنية المهمشة في سياق العنصرية النظامية إلى إمكانية الحصول على الرعاية الصحية الكافية فيما مضى، فضلاً عن اختلاف أنماط الأمراض بسبب تباين العوامل الاجتماعية والاقتصادية المحددة للصحة.

41- وفي الولايات المتحدة، استُحدثت حاسبة لمساعدة مقدمي الرعاية الصحية في تقييم احتمالات نجاح حالات الولادة المهبلية بعد الولادة القيصرية. وكشفت دراسة أجريت في عام 2019 عن تحيز في الخوارزمية التي استندت إليها الحاسبة. فقد استندت الحاسبة إلى عاملين تصحيحيين قائمين على العرق، مما أدى إلى انخفاض معدلات نجاح الولادات المهبلية المتوقعة للنساء المنحدرات من أصل أفريقي والنساء المنحدرات من أصل هسباني مقارنة بالنساء البيض ذوات السمات المماثلة. وبسبب عوامل التصحيح هذه، قد تؤدي الحاسبة إلى تقاوم أوجه التفاوت العرقي في نتائج صحة الأم من خلال ثني الأطباء عن توفير إمكانية الولادة المهبلية للنساء المنحدرات من أصل أفريقي والنساء المنحدرات من أصل هسباني، مما يؤدي إلى ارتفاع معدلات العمليات القيصرية في أوساطهن⁽⁶²⁾.

(ب) الكشف عن الأمراض

42- من التطبيقات الأخرى لتكنولوجيات الذكاء الاصطناعي، الكشف عن الأمراض، بما في ذلك الكشف عن الإصابة بمرض السرطان⁽⁶³⁾. فنظم الذكاء الاصطناعي المدربة على مجموعات بيانات واسعة تضم آلاف أو ملايين الصور، بما في ذلك صور الأشعة والصور المستمدة من خلال تحليلات علم الأمراض والصور الفوتوغرافية قادرة على تعلم التمييز بين الآفات الطبيعية والآفات السرطانية⁽⁶⁴⁾. واستخدام الذكاء الاصطناعي في هذا المجال يمكن أن يساعد إلى حد كبير في الكشف المبكر عن الإصابة بالسرطان، ومن ثم، إلى إنقاذ الأرواح وتحسين كفاءة نظام الرعاية الصحية. غير أن الأشخاص

Yuval Abraham, "‘Lavender’: the AI machine directing Israel’s bombing spree in Gaza", (61) +972 Magazine, 3 April 2024.

Darshali A. Vyas and others, "Challenging the use of race in the Vaginal Birth after Cesarean (62) .Section Calculator", *Women’s Health Issues*, vol. 29, No. 3 (2019).

المساهمة الواردة من المنظمة الدولية لحماية الخصوصية. (63)

Likhitha Kolla and Ravi B. Parikh, "Uses and limitations of artificial intelligence for oncology", (64) *Cancer*, 30 March 2024.

المنتمين إلى فئات عرقية وإثنية مهمشة قد لا يستفيدون من هذه التطورات على قدم المساواة مع غيرهم بسبب عدم إمكانية تعميم الخوارزميات على فئات المرضى غير الممثلة تمثيلاً كافياً في بيانات التدريب. وأشار الباحثون إلى أن أداء خوارزميات الذكاء الاصطناعي المستخدمة في الكشف عن الإصابة بسرطان الجلد أضعف عندما يتعلق الأمر بالأشخاص ذوي البشرة الداكنة نظراً إلى تحيز العديد من مجموعات بيانات الصور المتاحة للعموم والمستخدم لتدريب الخوارزميات وافتقار مجموعات البيانات تلك إلى التنوع من حيث ألوان البشرة والأصل الإثني⁽⁶⁵⁾. فعلى سبيل المثال، كشفت دراسة استقصائية شملت 21 مجموعة بيانات متاحة للعموم عن الأوقات الجلدية تحتوي على أكثر من 100 000 صورة، عن نقص كبير في تمثيل صور الأشخاص ذوي البشرة الداكنة⁽⁶⁶⁾.

(ج) الأجهزة الطبية التي تعتمد على الذكاء الاصطناعي

43- يواكب تطور الذكاء الاصطناعي واستخدامه التطورات الأخرى في مجال تكنولوجيايات الرعاية الصحية، بما في ذلك أجهزة الرعاية الصحية. وبما أن العديد من هذه الأجهزة يعتمد على الذكاء الاصطناعي، فإن التحيز العنصري يمكن أن يكون مدمجاً في خصائصها التشغيلية. ففي المملكة المتحدة مثلاً، أظهر تقرير أن التحيز كان مدمجاً في تشغيل الأجهزة الطبية في جميع مراحل تصنيعها، بما في ذلك في مرحلتها تطوير الخوارزميات والتعلم الآلي. وإبان جائحة مرض فيروس كورونا (كوفيد-19)، أدى استخدام أجهزة قياس النبض والأكسجين في الدم لقياس مستويات الأكسجين المنخفضة في الدم إلى المبالغة في تقدير مستويات الأكسجين في دم الأشخاص ذوي البشرة الداكنة⁽⁶⁷⁾.

3- التعليم

(أ) خوارزميات النجاح الأكاديمي والمهني

44- في بلدان مثل فنلندا والولايات المتحدة، تستخدم أدوات التحليلات التنبؤية في مجال التعليم لتحديد احتمالات النجاح في المستقبل استناداً إلى بيانات وخوارزميات إحصائية وإلى خاصية التعلم الآلي⁽⁶⁸⁾. وتتضمن البيانات المستخدمة في هذه الخوارزميات بيانات عن المواظبة على الدراسة والدرجات والسلوك والنشاط على شبكة الإنترنت. وهي مصممة لمساعدة المعلمين في توجيه الطلاب في اتخاذ القرارات المتعلقة بمسيرتهم التعليمية والمهنية. ومع أن أدوات التحليلات التنبؤية تهدف إلى مساعدة المعلمين في تحسين تحصيل الطلاب، فإنها غالباً ما تصنف الأقليات العرقية على أن احتمالات نجاح أفرادها في مسيرتهم التعليمية والمهنية أقل، بسبب تصميم الخوارزميات ومعايير اختيار البيانات المستخدمة لتدريبها. وعلى أساس هذه التصنيفات، قد يوجه المعلمون الطلاب المنتمين إلى المجموعات العرقية والإثنية المهمشة إلى الابتعاد عن الخيارات التعليمية والمهنية التي من شأنها أن تتيح لهم فرصة استغلال

(65) David Wen and others, "Characteristics of publicly available skin cancer image datasets: a systematic review", *The Lancet Digital Health*, vol. 4, No. 1 (2022)

(66) المرجع نفسه. انظر أيضاً المساهمة الواردة من المنظمة الدولية لحماية الخصوصية.

(67) المساهمة الواردة من المنظمة الدولية لحماية الخصوصية.

(68) Stina Westman and others, "Artificial intelligence for career guidance – current requirements and prospects for the future", *International Academic Forum Journal of Education*, vol. 9, No. 4 (2021); and Kelli A. Bird, Benjamin L. Castleman and Yifeng Song, "Are algorithms biased in education? Exploring racial bias in predicting community college student success", *Journal of Policy Analysis and Management*, 31 January 2024

إمكاناتهم إلى أقصى حد وتوفر لهم أفضل الفرص لكسر دوامة الإقصاء، أو قد يستثمرون موارد أقل في هؤلاء الطلاب.

(ب) خوارزميات تحديد الدرجات

45- عادة ما تستخدم هذه الخوارزميات بيانات الدرجات السابقة في تقييم أداء الطلاب. ويمكن أن تؤدي أنماط العنصرية النظامية السائدة في مؤسسات التعليم إلى تحيز تلك البيانات. وتكرر الخوارزميات التي تنتج درجات الطلاب هذا التحيز في البيانات، ولا سيما في حال استبعاد التعليقات الواردة من المعلمين⁽⁶⁹⁾. ويمكن أن تؤدي خوارزميات تقدير الدرجات دوراً هاماً للغاية في تحديد الفرص السانحة للطلاب، بما في ذلك فيما يتعلق بالوصول إلى التعليم الجامعي أو الحصول على فرص العمل بعد انتهاء المسيرة التعليمية. وبناء على ذلك، فإن القرارات الآلية المتحيزة عنصرياً قد تحد من الفرص السانحة للطلاب المنتمين إلى الفئات العرقية والإثنية المهمشة وتقوض إمكانات التعليم باعتباره أداة قادرة على تعطيل العنصرية النظامية.

46- وتقدم المملكة المتحدة مثلاً يدعو إلى الحذر من نشر خوارزميات تقدير الدرجات. ففي عام 2020، أُلغيت امتحانات المستوى المتقدم (A-level) بسبب جائحة كوفيد-19. وكبديل لدرجات الامتحانات، طُلب من المعلمين التنبؤ بنتائج الطلاب. ثم نشرت الوكالة التنظيمية الوطنية خوارزمية لتوحيد الدرجات المتوقعة على أساس بيانات الدرجات السابقة لطلاب كل مدرسة. ونتيجة لذلك، خُفضت درجات 40 في المائة من الطلاب، وهم طلاب كان العديد منهم ملتحقين بمدارس تقع في مناطق منخفضة الدخل. وعلى العكس من ذلك، رفعت الخوارزمية درجات عدد كبير للغاية من الطلاب الملتحقين بالمدارس المستقلة غير المجانية. وردت الحكومة على الجدول الذي أعقب ذلك بإلغاء توحيد الدرجات الذي أجرته الخوارزمية. غير أن هذه الواقعة تسببت في اضطرابات كبيرة في إجراءات القبول بالجامعات⁽⁷⁰⁾.

(ج) استخدام النماذج اللغوية الكبيرة في مجال التعليم

47- تعتمد أدوات الذكاء الاصطناعي التوليدي على النماذج اللغوية الكبيرة في إنتاج المحتوى الجديد، بما في ذلك النصوص والموسيقى والصور ومقاطع الفيديو. ويمكن أن تساعد النماذج اللغوية الكبيرة، التي تستخدم في السياقات التعليمية، في تحسين النتائج الأكاديمية للطلاب بمختلف أعمارهم. غير أن دراسات أظهرت أن النماذج اللغوية منحازة تجاه اللغة الإنكليزية، وهي أكثر اللغات استخداماً على شبكة الإنترنت واللغة التي يعمل بها معظم الباحثين والتقنيين في مجال الذكاء الاصطناعي. وعلاوة على ذلك، يملك عدد قليل للغاية من بين لغات العالم التي يقدر عددها بزهاء 6 000 لغة موارد بيانات عالية الجودة يمكن استخدامها لتدريب نماذج الذكاء الاصطناعي. ولسد هذه الفجوة، بدأت الشركات في تطوير نماذج لغوية متعددة اللغات. غير أن أداء النماذج متعددة اللغات لا يوازي أداء النماذج اللغوية الإنكليزية. ويمكن أن يؤدي استخدام النماذج اللغوية الكبيرة في السياقات التعليمية إلى الإضرار بالطلاب

Benjamin Herold, “Why schools need to talk about racial bias in AI-powered technologies”, (69) *Education Week*, 12 April 2022

Bryan Walsh, “How an AI grading system ignited a national controversy in the U.K.”, Axios, (70) 19 August 2020; and Daan Kolkman, “F**k the algorithm? What the world can learn from the UK’s A-level grading fiasco”, London School of Economics Impact Blog, 26 August 2020

المنحدرين من خلفيات لغوية غير ممثلة في موارد البيانات الأساسية، وهو ما قد يؤثر على مجموعة عرقية معينة أكثر من غيرها⁽⁷¹⁾.

48- وهناك جدل حول ما إذا كان ينبغي حظر استخدام الطلاب أدوات الذكاء الاصطناعي التوليدي القائمة على النماذج اللغوية الكبيرة بدلاً من دمجها في المناهج الدراسية. وتسعى بعض المؤسسات التعليمية أيضاً إلى تقييد استخدام الطلاب أدوات الذكاء الاصطناعي التوليدي التي تعتمد على النماذج اللغوية الكبيرة. وتستخدم بعض المؤسسات التعليمية أدوات الذكاء الاصطناعي للكشف عن استخدام الطلاب الذكاء الاصطناعي. وقد يؤدي استخدام هذه الأدوات، التي قد تحتوي على تحيز خوارزمي، في مراقبة الغش إلى المزيد من التحيزات التي تضر بالطلاب المنتمين إلى الفئات العرقية والإثنية المهمشة. وسيتفاقم هذا الضرر لا محالة في الحالات التي لم تضع فيها تلك المؤسسات إجراءات طعن منصفة⁽⁷²⁾.

(د) استخدام خاصية التعرف على سمات الوجه في المؤسسات التعليمية

49- تستخدم العديد من المؤسسات التعليمية تكنولوجيات التعرف على سمات الوجه، على الرغم من الأدلة التي تشير إلى وجود تحيز عنصري في عمل تلك التكنولوجيات، على النحو المبين أعلاه. وتستخدم نظم التعرف على سمات الوجه في تسجيل الحضور والغياب آلياً، وتعزيز أمن المدارس، وأداء مهام مراقبة الامتحانات، وحتى لتسجيل مشاعر الأطفال في المدارس لرصد مقدار ما يتعلمونه. وغالباً ما يتم ذلك من دون بذل العناية الواجبة في مجال حقوق الإنسان أو إخضاع تلك الأدوات للإشراف التنظيمي. ففي البرازيل مثلاً، يستعين عدد متزايد من المدارس بأدوات التعرف على سمات الوجه لتبسيط أساليب العمل وتتبع الحضور والغياب وتعزيز الأمن⁽⁷³⁾. غير أن التقارير تفيد بأن البلديات والولايات لم تجر أي دراسات لتقييم أثر برامج التعرف على سمات الوجه على حقوق الإنسان أو لتحليل مخاطر التمييز المرتبطة بتلك البرامج قبل تنفيذ تلك المشاريع⁽⁷⁴⁾.

50- وتتجم عن استخدام خاصية التعرف على الوجه في السياقات التعليمية آثار تمييزية عنصرية. فقد سُجلت حالات، منها حالة أُبلغ عنها في مملكة هولندا، اضطر فيها الطلاب المنحدرين من أصل أفريقي إلى تسليط الضوء على وجوههم حتى يتسنى لنظم الذكاء الاصطناعي المستخدمة في السماح للطلاب بخوض الامتحانات المهمة التعرف عليهم. ولا تؤثر هذه التجارب على حق الطلاب في الحصول على التعليم على قدم المساواة فيما بينهم فحسب، وإنما تؤدي أيضاً إلى انتهاكات وإلى حالات إقصاء عندما تعطي الطلاب المنحدرين من الجماعات العرقية والإثنية المهمشة انطباعاً بأن هذه النظم لم تُصمَّم

(71) Felix Richter, “The most spoken languages: on the Internet and in real life”, Statista, 21 February 2024; Emily M. Bender, “The #BenderRule: on naming the languages we study and why it matters”, The Gradient, 14 September 2019; Gabriel Nicholas and Aliya Bhatia, “Lost in translation: large language models in non-English content analysis”, Center for Democracy and Technology, 23 May 2023; A. Bergman and Mona Diab, “Towards responsible natural language technology”, in *The 60th Annual Meeting of the Association for Computational Linguistics: Findings of ACL 2022* (Association for Computational Linguistics, BigScience Workshop, “A 176B-parameter open-access multilingual language model” (ArXiv, 2022).

(72) Regina Ta and Darrell M. West, “Should schools ban or integrate generative AI in the classroom?”, نظر Brookings Institution, 7 August 2023; and Robert Topinka, “The software says my student cheated using AI. They say they’re innocent. Who do I believe?”, *The Guardian*, 13 February 2024.

(73) المساهمة الواردة من منظمة “Internet Lab”.

(74) المرجع نفسه.

بما يراعي احتياجاتهم. ويؤثر تسجيل ورصد مشاعر الأطفال في المدارس تأثيراً كبيراً على خصوصية جميع الطلاب ويمكن أن يديم التحيز العنصري. وقد تبين أن هذه النظم تقسر تعابير وجه الأفراد المنحدرين من أصل أفريقي والأفراد البيض تفسيراً مختلفاً، وأنها كثيراً ما تنسب المشاعر السلبية، مثل الازدراء والغضب، إلى الطلاب المنحدرين من أصل أفريقي⁽⁷⁵⁾.

جيم - المبادرات الناشئة الرامية إلى تنظيم الذكاء الاصطناعي وإدارته

51- بدأت الدول في اتخاذ خطوات واعدة لتنظيم الذكاء الاصطناعي وإدارته. وفي هذا الفرع، توجه المقررة الخاصة الانتباه إلى بعض هذه المبادرات. ويستند تحليلها غير الشامل إلى التقارير المقدمة من الدول وجماعات المجتمع المدني، فضلاً عن عملها القطري والبحوث التي أجرتها بغرض إعداد هذا التقرير.

1- المبادرات الوطنية

52- اتخذت دول خطوات لتنظيم وإدارة الذكاء الاصطناعي على الصعيد الوطني من خلال سن أحكام قانونية ملزمة ووضع معايير سياساتية طوعية، وكذلك من خلال مزيج من هذه الإجراءات في كثير من الحالات. فالبرازيل مثلاً بصدد سن أحكام تشريعية بشأن تنظيم مجال التكنولوجيا⁽⁷⁶⁾، بما في ذلك الذكاء الاصطناعي. واعتمدت الحكومة أيضاً عدداً من الوثائق السياساتية ذات الصلة، منها وثيقة بعنوان "العنصرية على شبكة الإنترنت: أدلة من أجل رسم سياسات رقمية"، وهي وثيقة تعيد التقارير بأن وزارة المساواة العرقية وضعتها وبأنها تتضمن تدابير لمعالجة التحيز الخوارزمي، بما في ذلك فيما يتعلق بالعرق⁽⁷⁷⁾. وترحب المقررة الخاصة بالجهود الرامية إلى وضع أحكام تنظيمية مكرسة وملزمة، تكملها معايير سياساتية ذات صلة. غير أنها تلقت معلومات مقلقة تعيد إجراء مشاورات فعلية مع السكان المنحدرين من أصل أفريقي وعدم إشراكهم في وضع الأحكام التشريعية المتعلقة بتنظيم الذكاء الاصطناعي، فضلاً عن افتقار مختلف المعايير والممارسات الحالية التي تتبعها الدولة للتناسق والاتساق عموماً⁽⁷⁸⁾.

53- وتقدم الولايات المتحدة، التي تعيد التقارير بأنها اتخذت خطوات لوضع مزيج من المعايير الملزمة والطوعية بشأن استخدام الذكاء الاصطناعي، مثلاً آخر. فقد رحبت المقررة الخاصة، عقب زيارتها البلد مؤخراً، بتوقيع الأمر التنفيذي رقم 14110 بشأن تطوير الذكاء الاصطناعي واستخدامه بشكل آمن ومأمون وموثوق، وبالإشارات الواردة فيه إلى مخاطر التحيز والتمييز في استخدام الذكاء الاصطناعي. وفي سياق إعداد هذا التقرير، تلقت المقررة الخاصة المزيد من المعلومات عن تنظيم الذكاء الاصطناعي في الولايات المتحدة، بما في ذلك عن العمل على صوغ شرعة حقوق الذكاء الاصطناعي، باعتبارها معياراً طوعياً، وعن الجهود المبذولة في ولايات مثل فرجينيا وكاليفورنيا ونيويورك لتنظيم الذكاء الاصطناعي، والمبادرات الرامية إلى تيسير تعهد الكيانات التجارية طوعاً بتطوير ذكاء اصطناعي آمن ومأمون وشفاف⁽⁷⁹⁾. وترحب المقررة الخاصة بهذه الجهود، غير أنها تعرب في الوقت ذاته عن قلقها لخلو

(75) المرجع نفسه.

(76) المساهمة الواردة من البرازيل.

(77) المساهمة الواردة من فريق الخبراء البرازيلي.

(78) المرجع نفسه.

(79) المساهمة الواردة من شبكة "NetMission.Asia". انظر أيضاً Kay Firth-Butterfield, Karen Silverman and Benjamin Larsen, "Understanding the US 'AI Bill of Rights' – and how it can help keep AI

الأمر التمييزي رقم 14110 من أي إشارة صريحة إلى التمييز العنصري والتحيز على الرغم من العدد الكبير من البحوث التي تشير إلى وجود تحيز عنصري عميق في المنتجات التجارية الرقمية في الولايات المتحدة⁽⁸⁰⁾.

54- وتفيد التقارير بأن كندا تكف على وضع مزيج من المعايير الملزمة والطوعية. ولم يعتمد البلد بعد مشروع قانون الذكاء الاصطناعي والبيانات، الذي تفيد التقارير بأنه من المقرر أن يتضمن ضوابط قانونية ملزمة لنظم الذكاء الاصطناعي عالية المخاطر. وبالإضافة إلى ذلك، وضعت كندا معايير طوعية تشمل مدونة قواعد السلوك الطوعية بشأن التطوير والإدارة المسؤولين لنظم الذكاء الاصطناعي التوليدي المتقدمة. واستحدثت البلد أيضاً أداة تقييم أثر الخوارزميات، وهي أداة مصممة لمساعدة الإدارات والأجهزة الحكومية على تقييم مخاطر الذكاء الاصطناعي، بما فيها المخاطر المتعلقة بالتمييز والتحيز، والتخفيف من حدتها⁽⁸¹⁾.

55- وتلقت المقررة الخاصة معلومات عن دول أخرى، مثل أستراليا والصين والهند واليابان، وهي معلومات تفيد بأن تلك الدول اتخذت خطوات لإدارة الذكاء الاصطناعي وتنظيمه، بما في ذلك من خلال التدابير السياسية، وفي بعض الحالات، من خلال التشريعات الملزمة⁽⁸²⁾.

2- المبادرات الإقليمية

56- فيما يتعلق بالمبادرات الإقليمية، ترحب المقررة الخاصة بما تلقت من الاتحاد الأوروبي ومن الدول الأعضاء فيه من معلومات عن اعتماد قانون الذكاء الاصطناعي⁽⁸³⁾. وتدرك المقررة الخاصة أن هذا القانون هو معيار تنظيمي ملزم سيكون له تأثير كبير في منطقة الاتحاد الأوروبي من خلال مواءمة المعايير القانونية الوطنية مع أحكامه. وترحب المقررة الخاصة بإشارة نص قانون الذكاء الاصطناعي إلى العرق، واحتوائه على ضمانات لحقوق الإنسان فيما يتعلق باستخدامات الذكاء الاصطناعي الشديد الخطورة، وحظره بعض استخدامات الذكاء الاصطناعي، ونصه على آليات انتصاف للأشخاص المتضررين من استخدام نظم الذكاء الاصطناعي الشديدة الخطورة. وترحب أيضاً بتناول خطة عمل الاتحاد الأوروبي لمكافحة العنصرية للفترة 2020-2025 التمييز العنصري الناشئ عن استخدام التكنولوجيات الجديدة، مثل الذكاء الاصطناعي، مما يشير إلى أن مختلف معايير الاتحاد الأوروبي تتسم بدرجة من الاتساق السياسي⁽⁸⁴⁾. غير أن المقررة الخاصة تلقت معلومات مقلقة للغاية تشير إلى وجود استثناءات من تدابير الحماية المنصوص عليها في القانون في سياقات الهجرة وإدارة الحدود وإنفاذ

United States, Office of Science and Technology Policy of the White House, “Blueprint for an AI bill of rights: making automated systems work for the American people”, white paper, October 2022
 “Fact sheet: Biden-Harris Administration secures voluntary commitments from eight additional artificial intelligence companies to manage the risks posed by AI”, 12 September 2023

A/HRC/56/68/Add.1، الفقرة 54. (80)

Canada, Innovation, Science and Economic Development Canada, “The Artificial Intelligence and Data Act (AIDA) – companion document”, 13 March 2023 (81)

المساهمة الواردة من شبكة “NetMission.Asia”. (82)

المساهماتن الواردتان من الاتحاد الأوروبي وإسبانيا. (83)

المساهمة الواردة من الاتحاد الأوروبي. انظر أيضاً https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html (84)

القانون⁽⁸⁵⁾. وتقيد التقارير بأن هذه الاستثناءات قائمة على الرغم من وجود تمييز عنصري شديد في هذين المجالين منذ أمد طويل ومن المخاطر المرتبطة بالسماح بوضع أطر قانونية موازية⁽⁸⁶⁾. ويهدد هذا النهج بتسيخ التسلسل الهرمي العنصري القائم وإدامة انتهاكات حقوق الإنسان إلى حد كبير في سياق الهجرة وإدارة الحدود وإنفاذ القانون في الدول الأعضاء في الاتحاد الأوروبي.

3- المبادرات الدولية

57- إن المقررة الخاصة على علم بما اتخذته الأمم المتحدة من تدابير للمساهمة في إدارة الذكاء الاصطناعي. وترحب المقررة الخاصة بإنشاء الأمين العام الهيئة الاستشارية الرفيعة المستوى المعنية بالذكاء الاصطناعي وينشر الهيئة تقريرها المرحلي مؤخراً. غير أنها تأسف لعدم الإشارة تحديداً في ذلك التقرير إلى خطر التحيز والتمييز العنصريين. وترحب المقررة الخاصة بعمل مفوضية الأمم المتحدة السامية لحقوق الإنسان الرامي إلى إدماج حقوق الإنسان في الحوارات الدولية المتصلة بالتكنولوجيات الناشئة، مثل الذكاء الاصطناعي، بما في ذلك من خلال مشروع "B-Tech". والمقررة الخاصة على علم أيضاً، بالإضافة إلى عمل الأمم المتحدة، بالمبادرات الدولية الأخرى الرامية إلى تعزيز الحوار وإدارة الذكاء الاصطناعي، بما في ذلك مبادرات منظمة التعاون والتنمية في الميدان الاقتصادي ومجموعة الدول السبع⁽⁸⁷⁾.

58- وبوسع المنظمات الدولية أن تيسر التعاون الدولي وتقديم المساعدة التقنية وإجراء البحوث الرامية إلى ضمان إدارة الذكاء الاصطناعي بطريقة لا تؤدي إلى تفاقم أوجه التفاوت الجسيمة أصلاً بين البلدان، التي تعد في كثير من الحالات تركبة من تركبات الاستعمار والعبودية. وقد تواجه البلدان تحديات مختلفة في تنفيذ أدوات إدارة الذكاء الاصطناعي بسبب الاختلافات الكبيرة في الهياكل الأساسية التكنولوجية. وقد انصب قدر كبير من التركيز على تكنولوجيا الذكاء الاصطناعي وما طرحه من تحديات تمييزية على بلدان شمال الكرة الأرضية، مما قد يؤدي إلى فجوات في فهم السبل التي يؤثر بها الذكاء الاصطناعي على الأقليات الثقافية والدينية وغيرها من الأقليات في بلدان الجنوب⁽⁸⁸⁾. ويحتمل أن تتمكن أكثر الدول تقدماً في شمال الكرة الأرضية من التأثير على النقاش والحوار المتعلقين بالذكاء الاصطناعي بطريقة تديم اختلافات موازين القوى العالمية وتحد من قدرة بلدان الجنوب على جني فوائده المحتملة.

دال - إطار القانون الدولي لحقوق الإنسان

59- ينبغي أن تقوم تكنولوجيا الذكاء الاصطناعي على معايير القانون الدولي لحقوق الإنسان. ويرد أشمل حظر للتمييز العنصري في الاتفاقية الدولية للقضاء على جميع أشكال التمييز العنصري. فعلى النحو الوارد في الفقرة 1 من المادة 1 من الاتفاقية، صاغت الدول الاتفاقية بغرض إدراجها تعريفاً واسعاً للتمييز العنصري على أنه يعني أي تمييز أو استثناء أو تقييد أو تفضيل يقوم على أساس العرق أو اللون أو النسب أو الأصل القومي أو الإثني ويستهدف أو يستتبع تعطيل أو عرقلة الاعتراف بحقوق الإنسان

(85) المساهمة الواردة من المنظمة الدولية لحماية الخصوصية؛ و Access Now, "The EU AI Act: a failure for human rights, a victory for industry and law enforcement", 13 March 2024.

(86) انظر A/HRC/48/76.

(87) المساهمة الواردة من شبكة "NetMission.Asia".

(88) Danni Yu, Hannah Rosenfeld and Abhishek Gupta, "The 'AI divide' between the Global North and the Global South", World Economic Forum, 16 January 2023.

والحريات الأساسية أو التمتع بها أو ممارستها، على قدم المساواة، في الميدان السياسي أو الاقتصادي أو الاجتماعي أو الثقافي أو في أي ميدان آخر من ميادين الحياة العامة.

60- وقد التزمت الدول الأطراف في الاتفاقية الدولية للقضاء على جميع أشكال التمييز العنصري بالسعي إلى تحقيق الهدف المتمثل في إيجاد مجتمع محلي ودولي خال من جميع أشكال العنصرية. ولتيسير تحقيق هدف المساواة العرقية فعلياً، تلزم المادة 2 من الاتفاقية الدول الأطراف بأن تكفل عدم مشاركتها في أي عمل من أعمال التمييز العنصري وعدم تأييدها أي برامج تؤدي إلى عدم المساواة العرقية. والدول الأطراف ملزمة، علاوة على ذلك، باتخاذ إجراءات فعالة وفورية في حالات العنصرية أو عدم المساواة العرقية أو التمييز العنصري. والالتزام باتخاذ إجراءات هو التزام مطلق. وتشمل التزامات الدول الأطراف بمنع عدم المساواة العرقية والتمييز العنصري التدابير الوقائية والعلاجية على حد سواء. وتتص الاتفاقية على أشمل حظر للتمييز العنصري، غير أن معاهدات أخرى توفر أيضاً الحماية من أشكال التمييز هذه.

61- وتمتد الالتزامات بتحقيق المساواة العرقية وضمن عدم التمييز إلى جميع مجالات سياسات الحكومات ونفوذها، بما في ذلك تصميم تقنيات الذكاء الاصطناعي وتطبيقها. وسواء كان التمييز العنصري الناجم عن الذكاء الاصطناعي متعمداً أم لا، فهذا أمر لا صلة له بواجب التصرف الواقع على عاتق الدول الأطراف في ضوء نطاق حظر التمييز العنصري بموجب الاتفاقية الدولية للقضاء على جميع أشكال التمييز العنصري وغيرها من معاهدات حقوق الإنسان. وللواجبات الواقعة على عاتق الدول الأطراف في الاتفاقية والمتمثلة في السعي إلى إيجاد مجتمع محلي ودولي خال من جميع أشكال التمييز العنصري أيضاً صلة بالطريقة التي تمنع بها الدول وتعالج أوجه التفاوت داخل البلدان وفيما بينها فيما يتعلق بتوزيع فوائد تكنولوجيات الذكاء الاصطناعي.

62- ويجب على الدول أيضاً أن تكفل تمتع جميع الجماعات العرقية والإثنية بحقوقها الإنسانية كاملة، على النحو المنصوص عليه في المادة 5 من الاتفاقية الدولية للقضاء على جميع أشكال التمييز العنصري. وتتص المادة 5 من الاتفاقية على المساواة أمام القانون، بما في ذلك، في جملة أمور، الحق في المساواة في المعاملة أمام المحاكم وسائر هيئات إقامة العدل؛ وحق الفرد في الأمان على شخصه وفي حماية الدولة له من أي عنف أو أذى بدني، يصدر سواء عن موظفين رسميين أو عن أية جماعة أو مؤسسة؛ والحق في حرية التجمع السلمي وتكوين الجمعيات؛ والحق في التمتع بخدمات الصحة العامة والرعاية الطبية والضمان الاجتماعي والخدمات الاجتماعية؛ والحق في التعليم والتدريب. وهذه الحقوق، شأنها في ذلك شأن الأحكام التي تضمن تطبيقها تطبيقاً لا تمييز فيه، هي حقوق تتص عليها أيضاً المعاهدات الأخرى لحقوق الإنسان، بما فيها العهد الدولي الخاص بالحقوق المدنية والسياسية والعهد الدولي الخاص بالحقوق الاقتصادية والاجتماعية والثقافية.

63- ويتضمن القانون الدولي لحقوق الإنسان أحكاماً أخرى تسند إلى الدول مسؤولية معالجة الآثار التمييزية لتكنولوجيات الذكاء الاصطناعي، على النحو المبين أعلاه. ويثير جمع البيانات واستخدامها من دون توفير ضمانات لحقوق الإنسان مخاوف شديدة فيما يتعلق بالخصوصية، وهي مخاوف يمكن أن تتفاقم عندما يتعلق الأمر بالأشخاص المنتمين إلى المجموعات العرقية والإثنية المهمشة. وبناء على ذلك، تود المقررة الخاصة أن تذكر الدول بأحكام المادة 17 من العهد الدولي الخاص بالحقوق المدنية والسياسية التي تتص على عدم التدخل التعسفي أو غير القانوني في خصوصية الشخص وتلزم الدول بضمان الحماية القانونية ذات الصلة. وتطبق أحكام أخرى من العهد أيضاً على مظاهر التمييز العنصري المرتبطة بتكنولوجيات الذكاء الاصطناعي. ويمكن أن يؤثر استخدام الذكاء الاصطناعي، بما في ذلك في سياقات من قبيل إنفاذ القانون، على حرية الشخص وأمنه، كما يمكن أن تنجم عنه عواقب تهدد حياة

الأشخاص المنتمين إلى المجموعات العرقية والإثنية المهمشة. وتحدد المادة 6 من العهد الحق في الحياة باعتباره حقاً ملازماً لكل إنسان وتلتزم الدول بتوفير الحماية القانونية في هذا الصدد. وتتص المادة 7 على ألا يُعرض أي شخص للتعذيب أو المعاملة أو العقوبة القاسية أو اللاإنسانية أو المهينة. وتتص المادة 9 على أن لكل فرد حق في الحرية وفي الأمان على شخصه وعلى أنه لا يجوز توقيف أحد أو اعتقاله تعسفاً. وتوضح المادة 14 أن الناس جميعاً سواء أمام القضاء. وتتص المادة 26 على حماية الأقليات من التمييز. وتتص الفقرة 1 من المادة 2 من العهد على التزام بضمان تطبيق جميع أحكام العهد تطبيقاً لا تمييز فيه. وينص إطار القانون الدولي لحقوق الإنسان أيضاً على أحكام تتعلق باستخدام الذكاء الاصطناعي في سياق الهجرة ومراقبة الحدود وفي وسائل التواصل الاجتماعي. وقد بحث التقريران السابقان المقدمان في إطار هذه الولاية هذه الأحكام⁽⁸⁹⁾.

64- وينص القانون الدولي لحقوق الإنسان على أن لجميع الأشخاص الذين قد يتعرضون للتمييز العنصري الحق في الوصول إلى سبل الانتصاف، وينطبق ذلك على الحالات التي يكون التمييز ناتجاً فيها عن الذكاء الاصطناعي. وتتص المادة 6 من الاتفاقية الدولية للقضاء على جميع أشكال التمييز العنصري على الحق في الحصول على حماية وسبل انتصاف فعالة عن طريق المحاكم الوطنية المختصة وغيرها من مؤسسات الدولة. وبالإضافة إلى ذلك، اعترفت الجمعية العامة بخمسة عناصر رئيسية لحق ضحايا الانتهاكات الجسيمة لحقوق الإنسان في الانتصاف والجبر، وهي: رد الحق، والتعويض، ورد الاعتبار، والترضية، وضمانات عدم تكرار انتهاكات الحقوق⁽⁹⁰⁾.

65- وتؤدي الكيانات التجارية دوراً هاماً في تصميم الذكاء الاصطناعي وتطبيقه. وهي الجهات الفاعلة الرئيسية في تطوير الذكاء الاصطناعي، وكثيراً ما تتعاقد الحكومات معها لنشر الذكاء الاصطناعي في مؤسسات القطاع العام. وتحدد المبادئ التوجيهية بشأن الأعمال التجارية وحقوق الإنسان الالتزامات ذات الصلة الواقعة على عاتق الحكومات والمسؤوليات ذات الصلة الواقعة على عاتق الحكومات والأعمال التجارية على حد سواء. وتتص المبادئ التوجيهية على أنه يجب على الدول أن تحمي من انتهاكات حقوق الإنسان التي ترتكبها أطراف ثالثة داخل أراضيها و/أو المناطق الخاضعة لولايتها القضائية، بما في ذلك مؤسسات الأعمال. وينبغي للدول أن توفر هذه الحماية من خلال اتخاذ إجراءات تشمل فيما تشمله ضمان سياسات وتشريعات ولوائح وأحكام قضائية فعالة. وتتص المبادئ التوجيهية على مسؤولية الشركات عن منع انتهاكات حقوق الإنسان التي تكون قد تسببت أو أسهمت فيها، وعن التخفيف من حدتها ومعالجتها، وعن بذل العناية الواجبة في مجال حقوق الإنسان فيما يتعلق بالأنشطة التجارية ذات الصلة⁽⁹¹⁾. وبالإضافة إلى ذلك، تحدد المبادئ التوجيهية التزامات الحكومات ومسؤوليات الكيانات التجارية فيما يتعلق بضمان إمكانية الوصول إلى سبل الانتصاف من انتهاكات حقوق الإنسان المتعلقة بالأعمال التجارية، باعتبارها مكملتها للحق في الانتصاف المنصوص عليه في معايير أخرى، على النحو المبين أعلاه. وفي إطار مشروع "B-Tech" الذي تنفذه المفوضية السامية

(89) A/75/590 وA/78/538.

(90) المبادئ الأساسية والمبادئ التوجيهية بشأن الحق في الانتصاف والجبر لضحايا الانتهاكات الجسيمة للقانون الدولي لحقوق الإنسان والانتهاكات الخطيرة للقانون الإنساني الدولي، الفقرات من 15 إلى 23.

(91) انظر أيضاً United Nations Office on Genocide Prevention and the Responsibility to Protect and Economic and Social Research Council Human Rights, Big Data and Technology Project, University of Essex, "Countering and addressing online hate speech: a guide for policy makers and practitioners", policy paper, July 2023، وA/74/486، الفقرتين 44 و45.

لحقوق الإنسان، جرى العمل على توفير ما يلزم من توجيه وموارد لتنفيذ المبادئ التوجيهية في مجال التكنولوجيا، وشمل ذلك عملاً في مجال الذكاء الاصطناعي تحديداً⁽⁹²⁾.

رابعاً - الاستنتاجات والتوصيات

66- وجهت المكلّفة السابقة بالولاية إلى الدول والجهات الأخرى صاحبة المصلحة، بما في ذلك الكيانات التجارية، دعوة واضحة إلى رفض نهج "الحياد العرقي" في حوكمة وتنظيم التكنولوجيات الناشئة، بما فيها الذكاء الاصطناعي. وحثت الدول على تنظيم هذه التكنولوجيات في إطار نهج يعترف بالعنصرية الهيكلية ويستند إلى المعايير الأساسية لحقوق الإنسان. ومع ذلك، فإن إدارة وتنظيم الذكاء الاصطناعي لا يزالان غير كافيين إلى حد كبير ولا يعبران اهتماماً كافياً للتحيز العنصري ولا يجسدان معايير القانون الدولي لحقوق الإنسان. ولا يزال هذا الواقع قائماً على الرغم من وضوح وحسن توقيت الدعوات السابقة إلى النأي عن اتباع نهج "الحياد العرقي" وعلى الرغم من تزايد الوعي بالعنصرية النظامية خلال السنوات الأربع التي مضت منذ توجيه تلك الدعوات. ولا يزال افتراض موضوعية التكنولوجيا وحيادها شائعاً ولا يزال يشكل محركاً للتسابق في إدماج الذكاء الاصطناعي في المجتمع على الرغم من آثاره التمييزية العنصرية ومن دون إيلاء الاعتبار الواجب لمدى ضرورته. ومع أن الذكاء الاصطناعي يزخر فعلاً بإمكانات إيجابية، بما في ذلك على صعيد المساواة والشمول، فإنه ليس حلاً سحرياً لجميع القضايا المجتمعية ويجب إدارته بفعالية لتحقيق التوازن بين فوائده ومخاطره.

67- وتنظيم الذكاء الاصطناعي تنظيمياً فعالاً وشاملاً أمر أساسي لتحقيق هذا التوازن الدقيق. ومع أن التنظيم الفعال للذكاء الاصطناعي أمر لا غنى عنه، فإن هناك خطوات إضافية يمكن أن تتخذها الدول والجهات الأخرى لمعالجة الآثار التمييزية العنصرية لهذه التكنولوجيات على نحو فعال. ومن المهم للغاية أيضاً استحداث برامج قائمة على حقوق الإنسان لتثقيف عامة الجمهور بالتكنولوجيات الناشئة وتمكينه من الإلمام بالذكاء الاصطناعي. فعندما يفهم الأفراد والجماعات الذكاء الاصطناعي ويدركون حقوق الإنسان الخاصة بهم في الفضاء الرقمي، يصبحون قادرين على استخدام تلك المعرفة بمسؤولية باعتبارهم جمهوراً متبصراً يتحلى بالمسؤولية ويمكنه تحسين المساءلة عن نظم الذكاء الاصطناعي.

68- وينبغي أن تقوم الدول بما يلي:

(أ) التصدي للتحدي المتمثل في تنظيم الذكاء الاصطناعي انطلاقاً من شعورها المتزايد بالطابع الملح لذلك في ضوء سرعة تطور هذه التكنولوجيات وتعدد الأساليب التي تديم بها بالفعل التمييز العنصري عبر المجالات المجتمعية؛

(ب) وضع أطر تنظيمية للذكاء الاصطناعي تستند إلى فهم شامل للعنصرية النظامية وتستند إلى القانون الدولي لحقوق الإنسان، بما في ذلك إلى حظر التمييز العنصري. وينبغي ألا تستند هذه الأطر إلى نهج منعزلة، كما ينبغي أن تأخذ في اعتبارها مختلف الصكوك القانونية، بما في ذلك التشريعات المكرسة للذكاء الاصطناعي، وقوانين الخصوصية، والأحكام المتعلقة بحرية المعلومات، وتشريعات مكافحة التمييز، واللوائح القطاعية، بما يكفل تنظيم الذكاء الاصطناعي تنظيمياً شاملاً وفعالاً يمنع أثره التمييزي العنصري ويعالجه؛

(92) انظر المفوضية السامية لحقوق الإنسان، مشروع B-Tech: المفوضية السامية لحقوق الإنسان والأعمال التجارية وحقوق الإنسان، متاح في: <https://www.ohchr.org/en/business-and-human-rights/b-tech-project>.

(ج) النظر في الدور الذي يمكن أن تؤديه المعايير الطوعية في الأطر التنظيمية للذكاء الاصطناعي. ويمكن أن توفر المعايير الطوعية مبادئ توجيهية قابلة للتنفيذ بشأن التطبيق العملي للمعايير القانونية. غير أن تنظيم الذكاء الاصطناعي ينبغي ألا يعتمد على المعايير الطوعية وحدها، نظراً لأهمية آثار هذه التكنولوجيات على حقوق الإنسان، بما في ذلك فيما يتعلق بالتمييز العنصري؛

(د) النص، في إطار الأطر التنظيمية، على التزام ملزم قانوناً بإجراء تقييمات شاملة لبذل العناية الواجبة بحقوق الإنسان، بما في ذلك معايير صريحة لتقييم التحيز العنصري والإثني، في تطوير ونشر جميع تكنولوجيات الذكاء الاصطناعي. ويجب أن تشمل العناية الواجبة بحقوق الإنسان بروتوكولات وعتبات لاختبار البيانات بما يكفل الحماية من التحيز الخوارزمي، بما في ذلك التحيز العنصري والإثني. وينبغي إنجاز ذلك قبل نشر التكنولوجيات الجديدة، ولا سيما في مؤسسات القطاع العام، مثل المؤسسات التعليمية وأجهزة إنفاذ القوانين ومؤسسات الرعاية الصحية؛

(هـ) النظر في حظر استخدام نظم الذكاء الاصطناعي التي ثبت أنها تنطوي على مخاطر غير مقبولة على حقوق الإنسان، بما في ذلك النظم التي تنتهك حظر التمييز العنصري؛

(و) ضمان أن تنص الأطر التنظيمية على أحكام تضمن اتسام عمليات صنع القرارات الآلية بالشفافية التامة، بما في ذلك حقوق الأفراد في الوصول إلى المعلومات، في الحالات التي يعتبر فيها استخدام الذكاء الاصطناعي مسموحاً به، وذلك استناداً إلى عمليات تقييم شاملة لمدى بذل العناية الواجبة بحقوق الإنسان؛

(ز) وضع إجراءات طعن واضحة يمكن الوصول إليها وتخويل القائمين على تلك الإجراءات صلاحية تقييم ومعالجة الآثار التمييزية العنصرية للذكاء الاصطناعي، وضمان أن تشمل تلك الإجراءات المراجعة البشرية. وينبغي ضمان إمكانية الاحتكام المنصف إلى إجراءات الطعن هذه؛

(ح) إنشاء آليات لتمكين المتضررين، من أفراد وجماعات، من الوصول إلى سبل الانتصاف التي تكفل رد الحقوق والتعويض ورد الاعتبار والترضية وضمانات عدم التكرار في الحالات التي تؤدي فيها تكنولوجيات الذكاء الاصطناعي إلى انتهاك حقوق الإنسان، بما في ذلك إلى التمييز العنصري؛

(ط) الحيلولة دون النص في المعايير التنظيمية على أي استثناءات من شأنها أن تؤدي إلى انتهاكات لحظر التمييز العنصري بموجب القانون الدولي لحقوق الإنسان؛

(ي) ضمان التشاور الفعلي والفعال مع أصحاب المصلحة من جميع الفئات العرقية والإثنية المهمشة، وكذلك مع المهنيين في القطاعات المعنية، في وضع وتنفيذ اللوائح المنظمة للذكاء الاصطناعي، وكذلك في تطوير واستخدام تكنولوجيات الذكاء الاصطناعي؛

(ك) الاستثمار في جمع البيانات المصنفة عبر جميع القطاعات ذات الصلة بما يكفل الحصول على المعلومات اللازمة لكي يستند تنظيم الذكاء الاصطناعي إلى فهم العنصرية النظامية، ومعالجة مشاكل البيانات في نظم الذكاء الاصطناعي وتحسين رصد وتقييم تأثير تكنولوجيا الذكاء الاصطناعي على الأشخاص المنتمين إلى المجموعات العرقية والإثنية المهمشة؛

(ل) اتباع نهج يقوم على معايير حقوق الإنسان إزاء تصنيف البيانات وتخزينها من خلال ضمان تصنيفها وتوفير إمكانية التحديد الذاتي للهوية وكفالة الشفافية والخصوصية والمشاركة والمساءلة في هذا الصدد⁽⁹³⁾؛

(م) إنشاء آليات قوية للإشراف على أدوات الذكاء الاصطناعي ورصدها باستمرار، بما في ذلك عمليات تقييم منتظمة لآثارها، بما يضمن امتثالها للوائح ومعالجة أي مخاوف يثيرها الأفراد المتضررون أو المجتمعات المتضررة، إلى جانب التحيزات التي يحتمل أن تنتجها نماذج الذكاء الاصطناعي بمرور الوقت؛

(ن) الانخراط في التعاون الدولي وبناء القدرات والبحوث لضمان توزيع فوائد الذكاء الاصطناعي توزيعاً أكثر إنصافاً فيما بين البلدان لتجنب تعميق أوجه التفاوت القائمة والموروثة عن الاستعمار والعبودية؛

(س) استحداث برامج تثقيف عامة تركز على حقوق الإنسان بشأن الاستخدام المقبول والمسؤول لتكنولوجيا الذكاء الاصطناعي بما يكفل زيادة الإلمام بالذكاء الاصطناعي، بما يشمل مواد تثقيفية تدكي الوعي على وجه التحديد بالآثار التمييزية العنصرية الناجمة عن الذكاء الاصطناعي.

69- وينبغي أن تقوم الكيانات التجارية بما يلي:

(أ) إجراء تقييمات لمدى بذل العناية الواجبة بحقوق الإنسان في جميع مراحل تصميم منتجات الذكاء الاصطناعي وتطويرها ونشرها؛

(ب) ضمان التشاور الفعلي والفعال مع الأشخاص المنتمين إلى المجموعات العرقية والإثنية المهمشة والمهنيين العاملين في المجالات المجتمعية ذات الصلة والخبراء في مجال العنصرية النظامية فيما يخص تصميم منتجات الذكاء الاصطناعي، وتطويرها ونشرها؛

(ج) وضع بروتوكولات لضمان اتساق المنتجات التي تؤثر على حقوق الإنسان بالشفافية التامة وتبادل المعلومات عن عمليات اتخاذ القرار من جانب الخوارزميات التي تستند إليها تلك المنتجات؛

(د) ضمان رصد التحيز العنصري في منتجات الذكاء الاصطناعي رصداً مستمراً؛

(هـ) استحداث مواد تدريبية بشأن التمييز العنصري، بما في ذلك التحيز المستتر والعنصرية النظامية، لفائدة الأشخاص الذين يشاركون في تصميم الذكاء الاصطناعي وتطويره ونشره. وينبغي أن يستند استحداث هذه المواد التدريبية إلى معايير القانون الدولي لحقوق الإنسان والبحوث التي تتناول الأثر التمييزي العنصري لتكنولوجيات الذكاء الاصطناعي؛

(و) المساهمة في الجهود الرامية إلى استحداث مواد تثقيفية عامة تركز على حقوق الإنسان بما يعزز إمكانية الاستفادة من المواد التي تكفل الإلمام بالذكاء الاصطناعي.

70- وينبغي أن تقوم الأمم المتحدة وآلياتها المستقلة لحقوق الإنسان بما يلي:

(أ) تيسير الحوار والنقاش الفعالين بين أصحاب المصلحة بشأن تكنولوجيات الذكاء الاصطناعي وتنظيمها؛

- (ب) إدماج منظور يركز صراحة على الأثر التمييزي العنصري الناجم عن تكنولوجيا الذكاء الاصطناعي في عمل الهيئة الاستشارية الرفيعة المستوى المعنية بالذكاء الاصطناعي؛
- (ج) ضمان استناد المنشورات والإرشادات المتعلقة بتكنولوجيا الذكاء الاصطناعي إلى القانون الدولي لحقوق الإنسان، بما في ذلك حظر التمييز العنصري، والاعتراف صراحة بالتحيز العنصري في تصميم ونشر تكنولوجيا الذكاء الاصطناعي باعتباره قضية عالمية خطيرة؛
- (د) المشاركة في رصد آثار تكنولوجيا الذكاء الاصطناعي على حقوق الإنسان، بما في ذلك الآثار المرتبطة بالتمييز العنصري؛
- (هـ) دعم التعاون الدولي وبناء القدرات وإجراء البحوث سعياً إلى توزيع فوائد الذكاء الاصطناعي على نحو أكثر إنصافاً فيما بين البلدان.