الدليل الاسترشادي لأخلاقيات استخدام الذكاء الاصطناعي

بالدول الأعضاء بمجلس التعاون لدول الخليج العربية

نسخة 1.0 | نوفمبر 2023م

الفهرس

مقدمة.....3
القيم والمبادئ.....4
القيم 4
القيمة الأولى: احترام كرامة وحرية واستقلالية الإنسان وحمايتها وتعزيزها..... 4
القيمة الثانية: احترام الشريعة الإسلامية والدستور وتعزيز اللحمة الخليجية 5
القيمة الثالثة: حماية البيئة وتعزيز الاستدامة 5
القيمة الرابعة: تعزيز الاستخدام السلمي لتحقيق رفاهية الإنسان الخليجي 5
المبادئ..... 5
المبدأ الأول: القرار للبشر 5
المبدأ الثاني: السلامة ومنع الضرر 6
المبدأ الثالث: العدالة والإنصاف وعدم التمييز..... 7
المبدأ الرابع: حماية الخصوصية وحماية البيانات 9
المبدأ الخامس: الشفافية والقابلية للشرح والتفسير..... 10
المبدأ السادس: المسؤولية والمسائلة والوعي..... 11
المبدأ السابع: النزاهة وعدم التزييف 12
استخدام هذا الدليل والانتفاع به وتطبيقه 12
الترويج لأخلاقيات الذكاء الاصطناعي 12
أحكام ختامية 13

مقدمة

يشير مصطلح الذكاء الاصطناعي (Artificial Intelligence - AI) إلى الأنظمة والأجهزة الحاسوبية التي تحاكي الذكاء البشري، حيث تتخذ سلوكاً خاصاً يجعلها تحاكي القدرات الذهنية البشرية وأنماط عملها، ومن أهم هذه القدرات التحليل والاستنتاج والتعلم التراكمي واتخاذ القرارات وردود الأفعال في أوضاع متغيرة لم تُبرمَج عليها الآلة مسبقاً وهذا هو المبدأ الرئيسي لعمل الذكاء الاصطناعي.

إن قدرة الذكاء الاصطناعي على محاكاة وتخطي الطريقة التي يستوعب ويتفاعل بها البشر مع العالم من حولنا من خلال ما تم تزويده به من أشكال عدة من التعلم الآلي التي تتعرف على أنماط البيانات أدت إلى نتائج حسنة عديدة؛ إلا أنها وفي ذات الوقت شكلت خطورة على المجتمعات البشرية والبيئة النظم الأيكولوجية والحياة البشرية وحتى على العقل البشري.

إن استخدام تقنيات الذكاء الاصطناعي يؤثر في العديد من قطاعات ومجالات الحياة، مثل الاقتصاد والعمل والتربية والتعليم والثقافة والصحة والرفاهية الاجتماعية وغيرها. وليس خافياً على أحد الطرائق التي يمكن أن يوفر بها الذكاء الاصطناعي فرصًا كبيرة للنهوض برفاهية المجتمع في العديد من مجالات الحياة، فعلى سبيل المثال، ساهم الذكاء الاصطناعي في تسهيل عملية التعليم، وتمكين وصول المعرفة لمختلف الفئات المجتمعية وخاصة خلال فترة الأزمة الوبائية التي كان من الصعب جدًا تخطيها دون اللجوء لتقنيات الذكاء الاصطناعي. إلا أن العديد من الإشكاليات والتحديات ينبغي أخذها بعين الاعتبار عند التعامل مع الذكاء الاصطناعي التي بإمكانها أن تقوض حماية حقوق الإنسان والبيانات الشخصية بشكل عام أو خاص. فمثلاً يمكن أن يهدد استخدام الذكاء الاصطناعي الحق في المساواة ومنع التمييز وهو أحد أهم الركائز التي تقوم عليها منظومة حقوق الإنسان بشكل عام. بالإضافة إلى ذلك، قد تتسبب تقنيات الذكاء الاصطناعي في تهديد وانتهاك الحقوق الخاصة كالحق في الخصوصية للأفراد أو أن تعززها بشكل ممنهج ومنظم كأن تعمل هذه التقنيات كصمام أمان للتمتع بالحقوق الأساسية والحرية الشخصية والسياسية، وعليه فإننا نتحدث هنا عن سلاح ذو حدين، وينبغي التعامل معه بحذر.

وعليه فإنّ المخاطر والتهديدات الناجمة عن انتشار استخدام الذكاء الاصطناعي - والتي من المتوقع أن تتزايد في السنوات القادمة - قد تؤثر على الحكومات والاقتصادات والمجتمعات محليًا وإقليميًا ودوليًا، مما يؤكد أهمية الاستجابة الوطنية والدولية العاجلة لتقليل احتمالية حدوثها أو الحد منها ومعالجة آثارها.

وبناءً على ما تقدم، تم تطوير الدليل الاسترشادي لأخلاقيات استخدام الذكاء الاصطناعي من أجل دعم الابتكار مع المحافظة على حماية مصالح دول مجلس التعاون الحيوية وأمنها الوطني من المخاطر والتهديدات التي قد تأتي من الذكاء الاصطناعي أو غيرها من التقنيات، وبالتالي الإسهام في تحقيق الاستفادة المثلى من التطورات التقنية وتمكين دول المجلس من الريادة عالميًا على جميع الأصعدة. كما كان الذكاء الاصطناعي الجدير بالثقة هو طموحنا الأساسي عند إنشاء هذه الوثيقة، حيث إنه من المهم بناء أنظمة ذكاء اصطناعي تستحق الثقة، لأن البشر لن يكونوا قادرين على جني فوائدها بثقة وبشكل كامل إلا عندما تكون التكنولوجيا - بما في ذلك العمليات والأشخاص الذين يقفون وراءها - جديرة بالثقة. وقد تم إنشاء هذه الوثيقة بالاعتماد على "التوصية الخاصة بأخلاقيات الذكاء الاصطناعي" الصادرة عن منظمة الأمم المتحدة للتربية والعلم والثقافة (يونسكو) عام 2021.

القيم والمبادئ

إن الهدف من القيم والمبادئ التوجيهية هو تعزيز تكنولوجيا الذكاء الاصطناعي بدول مجلس التعاون على كافة الأصعدة التقنية او الاجتماعية أو حتى السياسية منها؛ إذ أن وجود النوايا الحسنة للاستخدام لا يحول دون احتمالية أن تسبب أنظمة الذكاء الاصطناعي ضررا غير مقصودا. وعليه ينبغي على دول المجلس الاسترشاد بهذه الوثيقة لتحقيق الذكاء الاصطناعي الجدير بالثقة، والموائم لتوصيات الأمم المتحدة الخاصة بأخلاقيات الذكاء الاصطناعي.

وتحدد هذه القيم والمبادئ التوجيهية إطاراً عاماً لتطوير أنظمة الذكاء الاصطناعي تتمحور حول حماية كرامة وحقوق الإنسان، والمحافظة على البيئة والتعايش مع جميع الكائنات الحية، وترتكز على الالتزام باستخدام هذه التكنولوجيا في خدمة الإنسانية والصالح العام؛ وذلك بهدف تحسين رفاهية الإنسان وحريته. إذ وبالرغم من توفير أنظمة الذكاء الاصطناعي فرصًا عظيمة، فإنها تؤدي أيضًا إلى ظهور بعض المخاطر التي ينبغي التعامل معها بشكل مناسب ومتناسب.

ويعد هذا النهج هو ما ينبغي على دول مجلس التعاون تبنيه حتى تصبح إقليما رائدا للتكنولوجيا المتطورة والأخلاقية، إذ أن الذكاء الاصطناعي الآمن سيعود بالنفع على مواطني دول المجلس بما يتماشى مع القيم الأساسية المتمثلة في احترام الشريعة الإسلامية وحقوق الإنسان وسيادة القانون.

وفي هذا الإطار، تسعى دول مجلس التعاون من خلال هذه الوثيقة إلى زيادة الفرص الاستثمارية لمنتجي أنظمة الذكاء الاصطناعي وحصولهم على ميزة تنافسية وذلك من خلال دمج الذكاء الاصطناعي بمنتجات وخدمات تخدم الإنسانية .كما تسعى إلى تحسين ثقة المستخدمين بمنتجات الذكاء الاصطناعي؛ إذ أن عدم وجود إطار أخلاقي واضح وشامل لدعم التطبيق الآمن لهذه التكنولوجيا يحول دون ثقة البشر والمجتمعات في تطور التكنولوجيا وتطبيقاتها.

وعليه ينبغي احترام القيم والمبادئ المذكورة في هذه الوثيقة، وتعزيزها عند الضرورة وذلك من خلال إعداد تشريعات ونظم ومبادئ توجيهية جديدة، وإجراء التعديلات على التشريعات والنُظم والمبادئ التوجيهية القائمة بما يتوافق مع التشريعات والقوانين المحلية للدول الأعضاء

القيم

القيمة الأولى: احترام كرامة وحرية واستقلالية الإنسان وحمايتها وتعزيزها

تقوم الكرامة الإنسانية على فكرة أن كل إنسان يمتلك "قيمة جوهرية"، والتي لا ينبغي أبدًا التقليل منها أو المساس بها من قبل الآخرين بأي صورة كانت بما في ذلك التقنيات الجديدة مثل أنظمة الذكاء الاصطناعي. وفي هذا السياق، فإن احترام الكرامة الإنسانية يعني معاملة جميع الناس بالاحترام الواجب لهم كذوات أخلاقية مصونة، وليس مجرد أشياء يجب فرزها، أو تسجيلها، أو تربيتها، أو تكييفها. وعليه، ينبغي تطوير أنظمة الذكاء الاصطناعي بطريقة تحترم وتخدم وتحمي السلامة الجسدية والعقلية للبشر، وإحساسهم الشخصي والثقافي بالهوية، وتلبي احتياجاتهم الأساسية، وتمنح البشر الذين يتفاعلون مع أنظمة الذكاء الاصطناعي القدرة على الحفاظ على تقرير المصير الكامل والفعال لأنفسهم بما في ذلك اختيار وتحديد رغباتهم الخاصة.

القيمة الثانية: احترام الشريعة الإسلامية والدستور وتعزيز اللحمة الخليجية

ينبغي أن تعمل أنظمة الذكاء الاصطناعي على الحفاظ على مبادئ الشريعة الإسلامية السمحة، والدستور والتشريعات الوطنية والقوانين السائدة وكذلك المعايير والإجراءات المحلية السارية بدول مجلس التعاون وتعزيزها واحترام تعدد القيم وخيارات الحياة للأفراد. وألا تقوض تلك الأنظمة العمل بهذه المبادئ والتشريعات او تحد من فاعليتها. كما ينبغي - متى كان مناسبا - أن تدعم هذه الأنظمة تعزيز التعاون والتكامل بين دول المجلس. وأن تتضمن التزامًا بضمان عدم عملها بطرق تقوض الالتزامات الأساسية التي تقوم عليها سيادة القانون، والقوانين واللوائح الإلزامية، وضمان الإجراءات القانونية الواجبة والمساواة أمام القانون بين جميع أفراد المجتمع.

القيمة الثالثة: حماية البيئة وتعزيز الاستدامة

تواجه مبادرات التحول الرقمي تحديات من أهمها التأثير السلبي على البيئة والنظم الأيكولوجية للبشر وسائر الكائنات الحية. لذا ينبغي أن تراعي دول مجلس التعاون الخليجي المخاطر البيئية المتولدة من تشغيل وتدريب أنظمة الذكاء الاصطناعي باستخدام الطاقة الكهربائية ومنها على سبيل المثال الانبعاثات الكربونية أو استغلال المياه العذبة للتبريد، بما يضمن عدم تدهور النظم البيئية وبالتالي ضمان استدامتها للأجيال القادمة.

القيمة الرابعة: تعزيز الاستخدام السلمي لتحقيق رفاهية الإنسان الخليجي

رفاهية الإنسان تعني أن أنظمة الذكاء الاصطناعي ينبغي أن تعمل على تحسين سبل الحياة للفرد والجماعة؛ فعند تغذية أنظمة الذكاء الاصطناعي بتدفق ضخم من البيانات ستتولى هذه الأنظمة المهام المتكررة بمستوى عالٍ من الدقة دون ملل أو تعب. وينبغي على الجهات المسئولة عن تطبيق أنظمة الذكاء الاصطناعي التحقق من أن هذه الأنظمة لا تستخدم في مجالات تهدد تعايش البشر أو أمنهم أو سلامتهم داخل مجتمعات دول مجلس التعاون. إذ لا ينبغي لأنظمة الذكاء الاصطناعي أن تسبب أو تفاقم الضرر أو أن تؤثر سلباً على البشر. وهذا يعني أن تكون أنظمة الذكاء الاصطناعي والبيئات التي تعمل فيها آمنة ومأمونة، وأن تكون قوية من الناحية الفنية وضمان أنها ليست عرضة للاستخدام الضار. كما توصي هذه الوثيقة بأن يحظى الأشخاص الضعفاء من ذوي الإعاقة أو كبار السن أو غيرهم باهتمام أكبر وأن يتم تضمين احتياجاتهم في تطوير أنظمة الذكاء الاصطناعي.

المبادئ

المبدأ الأول: القرار للبشر

ينبغي لأنظمة الذكاء الاصطناعي أن تدعم استقلالية الإنسان وصنع القرار، على النحو المنصوص عليه في القيم المذكورة مسبقا. وهذا يتطلب أن تعمل أنظمة الذكاء الاصطناعي كعوامل تمكين لمجتمع ديمقراطي ومزدهر وعادل من خلال دعم الفرد والجماعة وتعزيز الحقوق الأساسية، والسماح بالرقابة البشرية على أنظمة الذكاء الاصطناعي.

كأي تقنية، يمكن لتقنيات الذكاء الاصطناعي تمكين الحقوق الأساسية وإعاقتها في آن واحد. إذ يمكنها أن تفيد الناس على سبيل المثال من خلال مساعدتهم على تتبع بياناتهم الشخصية، أو من خلال زيادة إمكانية الوصول إلى التعليم، أو سرعة اكتشاف الأمراض، أو تقديم العلاجات المناسبة وبالتالي دعمهم ومنحهم الحق في التعليم والعلاج. إلا أنها ونظراً لمدى وصول أنظمة الذكاء الاصطناعي للبيانات المتوفرة وقدراتها على التحليل، فإنها يمكن أن تؤثر أيضًا سلبًا على الحقوق الأساسية. وفي الحالات التي توجد فيها مثل هذه المخاطر، ينبغي إجراء تقييم لأثر الحقوق الأساسية. وينبغي أن يتم ذلك قبل تطوير النظام ويتضمن تقييم مستوى الضرر والمخاطر على جميع فئات المجتمع لضمان سلامة وأمن الإنسان قبل الشروع بالتطبيق.

وعليه، ينبغي أن تدعم أنظمة الذكاء الاصطناعي حرية الأفراد في اتخاذ خيارات أفضل وأكثر استنارة بما يتوافق مع أهدافهم، وأن تتجنب هذه الأنظمة تشكيل السلوك البشري والتأثير عليه من خلال آليات قد يكون من الصعب اكتشافها، لأنها قد تستغل عمليات اللاوعي، بما في ذلك أشكال مختلفة من التلاعب غير العادل والخداع والتكييف، وكلها قد تهدد الاستقلالية الفردية. إذ ينبغي أن تكون استقلالية المستخدم هي المبدأ الأساسي في وظائف النظام، والمفتاح لذلك هو منحه الحق في عدم الخضوع لقرار يعتمد على المعالجة الآلية فقط إذا كان من شأن ذلك أن يؤدي إلى آثار قانونية على المستخدمين أو يؤثر عليهم بشكل كبير.

توصيات ومقترحات لتحقيق هذا المبدأ

- **تفعيل الحوكمة والرقابة البشرية:** توصي هذه الوثيقة بتفعيل الرقابة البشرية لضمان ألا يؤدي نظام الذكاء الاصطناعي إلى تقويض استقلالية الإنسان أو التسبب في آثار سلبية أخرى. ويمكن تحقيق الرقابة من خلال آليات حوكمة والتي قد تشمل قرار عدم استخدام نظام الذكاء الاصطناعي في موقف معين، أو تحديد مستويات من السلطة التقديرية البشرية أثناء استخدام النظام، أو ضمان القدرة على تجاوز القرار الذي اتخذه النظام. علاوة على ذلك، ينبغي التأكد من أن الجهات المنفذة العامة لديها القدرة على ممارسة الرقابة بما يتماشى مع صلاحياتها. وقد تكون هناك حاجة إلى آليات إشراف بدرجات متفاوتة لدعم تدابير السلامة والتحكم الأخرى، اعتمادًا على مجال تطبيق نظام الذكاء الاصطناعي والمخاطر المحتملة. وكلما قلت الرقابة التي يمكن أن يمارسها الإنسان على نظام الذكاء الاصطناعي، كلما كانت هناك حاجة إلى اختبارات تقييم أكثر شمولاً وحوكمة أكثر صرامة.

المبدأ الثاني: السلامة ومنع الضرر

أحد المميزات التي يتمتع بها الذكاء الاصطناعي هو القوة التقنية على توليد القرارات والتحكم بالمخرجات، والتي تتطلب أن يتم تطوير أنظمة الذكاء الاصطناعي ضمن نهج وقائي تجاه المخاطر وبطريقة تجعلها تتصرف بشكل موثوق على النحو المنشود مع تقليل الضرر غير المقصود وغير المتوقع، ومنع الضرر غير المقبول. وينبغي أن ينطبق هذا أيضًا على التغييرات المحتملة من تشغيل هذه الأنظمة ضمن بيئة والتي قد تتفاعل مع النظام بطريقة تهدد السلامة الجسدية والعقلية للإنسان وجميع الكائنات الحية.

وتحمل أنظمة الذكاء الاصطناعي خصائص مشابهة لجميع أنظمة البرمجيات، فهي قد تعاني من نقاط الضعف التي يمكن أن تسمح لقراصنة الأنظمة باستغلالها، وقد تستهدف الهجمات البيانات او نموذج الذكاء الاصطناعي او حتى بنيته التحتية. وقد ينتج عن تلك الهجمات تغييرات في سلوك النظام، مما يدفعه إلى اتخاذ قرارات مختلفة عن التوقعات، أو تتسبب في إغلاقه

بشكل يهدد سلامة البشر وسائر الكائنات الحية. وعليه توصي هذه الوثيقة بضرورة توفير الحماية الأمنية الكافية للبيانات المستخدمة بأنظمة الذكاء الاصطناعي، ونماذج التدريب والتشغيل، وحماية النظام من المجموعات المتطرفة أو الإرهابية التي قد تسخر استخدام النظام من قبل جهات ضارة بالمجتمع، وينبغي اتخاذ خطوات لمنع ذلك والتخفيف منه.

توصيات ومقترحات لتحقيق هذا المبدأ

- **وضع نقطة تراجع:** توصي هذه الوثيقة بضرورة وضع نقطة تراجع لأنظمة الذكاء الاصطناعي في حال حدوث اضرار عالية الخطورة تهدد الانسان أو الكائنات الحية أو البيئة. وهذا يشمل تقليل العواقب والأخطاء غير المقصودة. بالإضافة إلى ذلك، ينبغي إنشاء عمليات لتوضيح وتقييم المخاطر المحتملة المرتبطة باستخدام أنظمة الذكاء الاصطناعي، عبر مجالات التطبيق المختلفة، وتحديد ما يعرف بنظام الاسترجاع (Rollback) كتدابير للسلامة ومنع استمرار الضرر. وعندما يكون من المتوقع أن تشكل عملية التطوير أو النظام نفسه مخاطر عالية بشكل خاص، فمن الضروري تطوير واختبار تدابير الاسترجاع والسلامة بشكل استباقي.

- **تقييم دقة التنبؤات والتوصيات والقرارات:** توصي هذه الوثيقة بتحري مستوى عال من الدقة بقدرة نظام الذكاء الاصطناعي على إصدار أحكام صحيحة، أو قدرته على تقديم تنبؤات أو توصيات أو قرارات صحيحة بناءً على البيانات أو النماذج المتوفرة. ويمكن لعملية التقييم الواضحة والمصممة بشكل جيد أن تدعم وتخفف وتصحح المخاطر غير المقصودة الناجمة عن التنبؤات غير الدقيقة. وفي الحالات التي يصعب فيها تجنب التنبؤات غير الدقيقة، فإنه من الضروري تقدير معدل الخطأ المحتمل ليساعد متخذي القرار في تحديد مصير النظام وما إذا كان من الضروري إطلاقه أو منعه بحسب نسبة الخطأ، بالأخص في المواقف التي يؤثر فيها نظام الذكاء الاصطناعي بشكل مباشر على حياة الإنسان.

- **التجربة والتحقق من الموثوقية:** توصي هذه الوثيقة بضرورة التحقق من موثوقية أنظمة الذكاء الاصطناعي وقدرتها على تكرار النتائج، إذ يعتبر نظام الذكاء الاصطناعي موثوقا إذا كان يعمل بشكل صحيح مع مجموعة من المدخلات وفي مجموعة من المواقف. وهذا ضروري خلال مرحلة تقييم أداء نظام الذكاء الاصطناعي من أجل منع الأضرار غير المقصودة. وينبغي أن يتأكد المطورون خلال تجربة الذكاء الاصطناعي أن سلوك النظام لا يعمل بشكل شاذ عند تكرار نفس المدخلات ضمن نطاق الشروط ذاتها.

المبدأ الثالث: العدالة والإنصاف وعدم التمييز

ينبغي إشراك جميع أصحاب المصلحة طوال دورة حياة نظم الذكاء الاصطناعي وتمكين الجميع في تطويرها بالإضافة إلى تمكين الأفراد والمجتمعات من الحصول على منافع تطبيقات نظم الذكاء الاصطناعي بشكل متساو وعادل، مما يعني ضرورة ضمان المساواة في الوصول من خلال عمليات التصميم الشامل بالإضافة إلى المعاملة المتساوية. ويرتبط هذا المطلب ارتباطا وثيقا بمبدأ العدالة. كما ينبغي للدول الأعضاء تبني مبدأ النفاذ للجميع خلال مراحل تصميم النظام بطريقة تسمح لجميع الأشخاص باستخدام منتجات أو خدمات الذكاء الاصطناعي، بغض النظر عن أعمارهم أو جنسهم أو قدراتهم أو صفاتهم. كما ينبغي أن تتاح إمكانية الوصول لهذه التكنولوجيا إلى الأشخاص ذوي الإعاقات أو الضعفاء من كبار السن وغيرهم.

توصيات ومقترحات لتحقيق هذا المبدأ

- **التدقيق على مجموعات البيانات المستخدمة في تدريب الأنظمة:** توصي هذه الوثيقة بضرورة التأكد من خلو مجموعات البيانات التي تستخدمها أنظمة الذكاء الاصطناعي للتدريب والتشغيل من أي تحيزات غير مقصودة قد تؤثر على فئات معينة وتخلق تحيزًا أو تمييزًا مباشرًا ضد مجموعات أو أشخاص معينين، مما قد يؤدي إلى تفاقم التحيز والتهميش بالمخرجات. ويمكن أن ينجم الضرر أيضًا عن الاستغلال المتعمد للتحيزات أو عن طريق الانخراط في منافسة غير عادلة. ولا يقتصر التحييز أو التمييز على القصور بالبيانات، ولكن من الممكن أن تكون الطريقة التي يتم بها تطوير أنظمة الذكاء الاصطناعي (مثل برمجة الخوارزميات) تعاني أيضًا من تحيز غير عادل.

- **وضع عمليات رقابية لضمان العدالة والإنصاف وعدم التمييز:** توصي هذه الوثيقة بضرورة وضع عمليات رقابية لتحليل ومعالجة وتقييم غرض نظام الذكاء الاصطناعي والقيود والمتطلبات والقرارات بطريقة واضحة وشفافة قبل الشروع بإطلاق النظام. علاوة على ذلك، فإن التوظيف من خلفيات وثقافات وتخصصات متنوعة يمكن أن يضمن تنوع الآراء وينبغي الحرص على تشجيعه.

- **التصدي للفجوات الرقمية:** توصي هذه الوثيقة بضرورة التصدي إلى الفجوات الرقمية والمعرفية داخل الدول وفيما بينها طوال دورة حياة أي نظام من أنظمة الذكاء الاصطناعي، ولا سيّما فيما يخص الانتفاع بالتكنولوجيا والبيانات وجودة سُبل الانتفاع بها، وفقًا للأُطر القانونية الوطنية المتعلقة بهذا الموضوع، وكذلك فيما يخص المعارف والمهارات وإمكانية الترابط والاتصال ومشاركة المجتمعات المحلية المعنية في المساعي المبذولة في هذا الصدد مشاركة فعالة من أجل معاملة كل شخص معاملة منصفة.

- **إشراك أصحاب المصلحة:** توصي هذه الوثيقة بضرورة مشاركة أصحاب المصلحة بشكل مسبق والتشاور معهم طوال دورة حياة أنظمة الذكاء الاصطناعي، وذلك لكونهم أكثر عرضة للتأثر بهذه الأنظمة سواء بشكل مباشر أو غير مباشر. ومن المفيد أيضاً التماس ردود أفعال أصحاب المصلحة بصورة منتظمة حتى بعد إطلاق أنظمة الذكاء الاصطناعي طوال عملية التنفيذ والتشغيل لأنظمة الذكاء الاصطناعي.

- **ضمان التنوع في سبل إتاحة أنظمة الذكاء الاصطناعي:** توصي هذه الوثيقة بضرورة الحرص على التنوع في سبل إتاحة أنظمة الذكاء الاصطناعي وألاّ تقتصر على أساليب أو تقنيات أو قنوات محدودة، وينبغي أن تؤخذ بعين الاعتبار مبادئ التصميم التي تتناول أوسع نطاق ممكن من المستخدمين، وفقًا لمعايير إمكانية الوصول والنفاذ للجميع. وهذا سيمكن من الوصول العادل والمشاركة النشطة لجميع الناس.

المبدأ الرابع: حماية الخصوصية وحماية البيانات

ترتبط الخصوصية ارتباطًا وثيقًا بمبدأ منع الضرر، وهي حق أساسي يتأثر بشكل خاص بأنظمة الذكاء الاصطناعي. وينبغي على دول مجلس التعاون منع الإضرار بالخصوصية من خلال حوكمة استخدام البيانات التي تحفظ جودة وسلامة البيانات المستخدمة وأهميتها في ضوء المجال الذي سيتم نشر أنظمة الذكاء الاصطناعي فيه وبروتوكولات الوصول الخاصة بها مع القدرة على معالجة البيانات بطريقة تحمي الخصوصية.

توصيات ومقترحات لتحقيق هذا المبدأ

- **تقنين استخدام السجلات الرقمية:** توصي هذه الوثيقة بضرورة تقنين استخدام السجلات الرقمية التي تعرّف أنظمة الذكاء الاصطناعي بالسلوك البشري لاستنتاج تفضيلات أو توجهات الأفراد مثل استخدام العمر، الجنس، الدين، لكيلا تستخدم للتمييز غير القانوني أو غير العادل ضدهم. ينبغي أن تضمن أنظمة الذكاء الاصطناعي حماية الخصوصية والبيانات طوال دورة حياة النظام بأكملها، بما فيها المعلومات التي قدمها المستخدم في البداية، بالإضافة إلى المعلومات التي تم إنشاؤها حول المستخدم على مدار تفاعله مع النظام.

- **تقييد الاستخدام المطلق للبيانات:** توصي الوثيقة باتباع المبادئ والممارسات الفضلى في التعامل مع البيانات و الآليات المناسبة لتشفيرها، مما يترتب عليه ضرورة تقييد الاستخدام المطلق للبيانات وبالأخص البيانات الشخصية بأنظمة الذكاء الاصطناعي، وينبغي تحديد الغرض من الاستخدام، والنتائج المتوقعة وتقييم المخاطر لكل نظام على حدة.

- **تقنين الوصول المطلق للبيانات:** توصي الوثيقة بضرورة اتباع مبادئ حوكمة البيانات بما يضمن أمنها وإتاحة المساءلة لمنتهكي خصوصية البيانات، حيث ينبغي تقنين الوصول المطلق للبيانات وبالأخص البيانات الشخصية المراد استخدامها بأنظمة الذكاء الاصطناعي. وينبغي وضع بروتوكولات تحكم الوصول إلى البيانات، بحيث يتم تحديد من يمكنه الوصول إلى البيانات وتحت أي ظروف، حيث ينبغي السماح فقط للموظفين المؤهلين الذين يتمتعون بالكفاءة والحاجة إلى الوصول إلى بيانات الأفراد، مع أخذ التدابير اللازمة للاستخدام الأمثل لسجلات المراجعة (Audit Trail) الخاصة بالوصول إلى البيانات واستخدامها.

- **مراقبة جودة وسلامة البيانات:** توصي الوثيقة بضرورة مراقبة جودة وسلامة البيانات المستخدمة بأنظمة الذكاء الاصطناعي، وتنفيذ الاختبارات الكافية للتحقق من جودة مجموعات البيانات المستخدمة، والذي يعد أمرًا بالغ الأهمية لأداء أنظمة الذكاء الاصطناعي. وينبغي العمل على تطوير منظومة مؤتمتة داخلية للمراقبة والتنبيه بحيث تشمل عمليات جمع البيانات مراقبة بشرية، إذ قد تتضمن تحيزات اجتماعية وعدم دقة وأخطاء وبيانات ضارة قد تسبب أضرارًا خلال تدريب أو تشغيل نماذج الذكاء الاصطناعي؛ كتغيير سلوكه، خاصة مع أنظمة التعلم الذاتي. وعليه ينبغي العمل على إجراء اختبارات وفق معايير واضحة للعمليات ومجموعات البيانات المستخدمة وتوثيقها في كل خطوة مثل التخطيط، والتدريب، والاختبار، والنشر. وينبغي أن ينطبق هذا أيضًا على أنظمة الذكاء الاصطناعي التي لم يتم تطويرها داخليًا، وإنما تم الحصول عليها من مؤسسات خارجية. كما ينبغي توفير قنوات مناسبة للتغذية الراجعة لتقنيات الذكاء الاصطناعي بشكل لا يسمح بالتلاعب بها من جهات أخرى.

- **وضع آليات وأُطر ملائمة لحماية البيانات:** توصي الوثيقة بضرورة وضع أُطر وآليات ملائمة لحماية البيانات تستند إلى المبادئ والمعايير الوطنية لحماية البيانات الشخصية فيما يخص جمعها واستخدامها ومعالجتها وإفشاءها وممارسة الأشخاص المعنيين بالبيانات لحقوقهم، وأن تضمن هذه الأطر والآليات وجود غرض أو هدف مشروع وأساس قانوني سليم للتعامل مع البيانات الشخصية بما فيها متطلبات الحصول على موافقة مستنيرة من الأشخاص المعنيين بتلك البيانات. كما ينبغي أن تحظى هذه الأطر والآليات بحماية النظم القضائية لها وأن يتم ضمان الأخذ بها طوال دورة حياة أنظمة الذكاء الاصطناعي. ومن الأمثلة التطبيقية على ذلك منح مستخدمي تكنولوجيا الميتافيرس أو من ينوب عنهم الحق الكامل في إدارة بياناتهم والتحكم فيها وفقاً لسياسات وتشريعات حماية البيانات الشخصية.

المبدأ الخامس: الشفافية والقابلية للشرح والتفسير

لا بدّ من بذل جهود لزيادة شفافية أنظمة الذكاء الاصطناعي وتعزيز قابليتها للشرح طوال دورة حياتها، بيد أنه ينبغي لمستوى الشفافية والقابلية للشرح أن يكون ملائماً دائماً للسياق والعواقب، حيث تعد القابلية للتفسير أمرًا بالغ الأهمية لبناء ثقة المستخدمين في أنظمة الذكاء الاصطناعي والحفاظ عليها. وينبغي للناس أن يكونوا على علم تام بأي قرار يجري اتخاذه بناءً على معلومات مستمدة من أنظمة الذكاء الاصطناعي واستنادًا إلى خوارزميات الذكاء الاصطناعي، بما فيها القرارات التي تمس بسلامتهم أو بحقوق الإنسان، ولهم في هذه الحالة أن يطلبوا من الجهات الفاعلة المعنية بالذكاء الاصطناعي أو من مؤسسات القطاع العام معلومات توضيحية بشأن أسباب اتخاذ تلك القرارات التي تمس بحقوقهم وحرياتهم. ويمكن أن يتحقق ذلك من خلال تقديم طلبات الطعن أو الاعتراض إلى موظف مكلف بتلقي طلبات الناس في هذه الحالات لإعادة النظر في القرارات المتخذة وتصحيحها، على ألا يتم التعامل مع آلية اتخاذ القرارات كخوارزميات "الصندوق الأسود" في مثل هذه الظروف. وينبغي للجهات الفاعلة المعنية بالذكاء الاصطناعي أن تُعلِم المستخدِمين إعلامًا سليمًا في الوقت المناسب بالمنتجات أو الخدمات المقدمة بطريقة مباشرة وبالمنتجات أو الخدمات المقدمة من خلال الاستعانة بنُظم الذكاء الاصطناعي. كما ينبغي ملاحظة أنه يمكن أن يتطلب الأمر إيجاد توازن بين مبدأ الشفافية والقابلية للشرح ومبادئ أخرى كمبدأ الحق في الخصوصية ومبدأ السلامة والأمن.

توصيات ومقترحات لتحقيق هذا المبدأ

- **تطوير إمكانية تتبع الوصول إلى القرارات:** توصي الوثيقة بتطوير إمكانية التتبع بحيث يتم توثيق مجموعات البيانات والعمليات التي تؤدي إلى قرار نظام الذكاء الاصطناعي، بما في ذلك جمع البيانات وتصنيفها بالإضافة إلى الخوارزميات المستخدمة، وفقًا لأفضل المعايير الممكنة للسماح بإمكانية التتبع وزيادة الشفافية. وينطبق هذا أيضًا على القرارات التي يتخذها نظام الذكاء الاصطناعي. وينبغي أن تتيح تلك الآلية تحديد الأسباب التي جعلت قرار الذكاء الاصطناعي خاطئًا، والذي بدوره يمكن أن يساعد في منع الأخطاء المستقبلية.
- **تطوير أنظمة ذكاء اصطناعي قابلة للشرح:** توصي هذه الوثيقة بضرورة تطوير أنظمة الذكاء الاصطناعي بطريق قابل للشرح إلى أقصى حد ممكن. وتتعلق قابلية الشرح بالقدرة على شرح كل من العمليات التقنية لنظام الذكاء الاصطناعي والقرارات البشرية ذات الصلة مثل مجالات تطبيق النظام. وينبغي أن يتوفر الشرح أو التفسير في الوقت المناسب وبشكل يناسب المستويات الفكرية لخبرة أصحاب المصلحة المعنيين (مثل الشخص العادي أو المنظم أو الباحث).

- **إعلام المستخدمين بتفاعلهم مع أنظمة ذكاء اصطناعي:** توصي هذه الوثيقة بضرورة إعلام المستخدمين بأنهم يتفاعلون مع نظام ذكاء اصطناعي وينبغي أن تكون أنظمة الذكاء الاصطناعي قابلة للتعريف على هذا النحو. بالإضافة إلى ذلك، ينبغي توفير خيار اتخاذ قرار ضد هذا التفاعل لصالح التفاعل البشري عند الحاجة لضمان الامتثال للحقوق الأساسية. كما ينبغي إبلاغ ممارسي الذكاء الاصطناعي أو المستخدمين النهائيين بقدرات وقيود نظام الذكاء الاصطناعي بطريقة مناسبة لحالة الاستخدام الحالية. هذا يمكن أن يشمل الإبلاغ عن مستوى دقة نظام الذكاء الاصطناعي، فضلاً عن حدوده أو حتى المخاطر المرتبطة باستخدامه.

المبدأ السادس: المسؤولية والمسائلة والوعي

إنّ شرطي المسؤولية والمساءلة يرتبطان ارتباطا وثيقا بمبدأ العدالة. ويتطلب ذلك وضع آليات لضمان تحمل المسؤولية عن أداء النظام وإتاحة إمكانية المساءلة عن أنظمة الذكاء الاصطناعي ونتائجها، قبل وبعد تطويرها ونشرها واستخدامها. ويتطلب ذلك التركيز على وعي الأفراد ومسؤوليتهم الأخلاقية عن الممارسات والأفعال التي يقومون بها في البيئة الافتراضية نظرا لعدم وضوح المرجعية القانونية التي تحكمها بعد. كما ينبغي السعي إلى تعزيز الوعي والدراية بوسائل تكنولوجيا الذكاء الاصطناعي وأخلاقياته لدى عامة الناس عن طريق تعليمهم وتثقيفهم وتدريبهم مع مراعاة أوجه التنوع اللغوي والاجتماعي والثقافي بالمجتمع المحلي.

توصيات ومقترحات لتحقيق هذا المبدأ

- **تطوير تشريعات تضمن المحاسبة والمسائلة القانونية:** توصي هذه الوثيقة بضرورة تطوير تشريعات تؤكد على أن مسؤولية أداء أنظمة الذكاء الاصطناعي تقع على كيانات يمكن محاسبتها قانونيا. وعليه لا تترك المسؤولية على النظام، بل تحدد بشكل واضح، تتحمل فيه الجهات المعنية بتطوير أنظمة الذكاء الاصطناعي المسؤولية عن أداء النظام خلال دورة حياته.

- **تطوير تشريعات تضمن إجراء التدقيق والتقييم:** توصي الوثيقة بضرورة تطوير التشريعات والإجراءات المعنية بالتدقيق الفني على أداء النظام مع إمكانية التدقيق وتقييم الخوارزميات والبيانات وعمليات التصميم. وينبغي أن يكون النظام متاحًا دائمًا لأي عملية تقييم مصرّح بها من قبل المدققين الداخليين والخارجيين بشكل مستقل مع توفير تقارير التقييم بما يسهم في تحسين مصداقية التكنولوجيا.

- **تطوير آليات للإبلاغ عن مخرجات أنظمة الذكاء الاصطناعي:** توصي الوثيقة بضرورة تطوير آليات للإبلاغ عن الإجراءات أو القرارات التي تنتج عن أنظمة الذكاء الاصطناعي، وينبغي ضمان الاستجابة لهذه البلاغات بشكل مناسب يرضي المستخدمين. كما ينبغي توثيق البلاغات وتحليلها من قبل الجهات المتخصصة لتقييم وتقليل الآثار السلبية المحتملة لأنظمة الذكاء الاصطناعي وتوفير الحماية الواجبة للمبلغين عند الإبلاغ بشكل مشروع عن أداء نظام الذكاء الاصطناعي.

- **تعزيز الوعي ونشر المعرفة:** توصي الوثيقة بضرورة تعزيز الوعي بأنظمة الذكاء الاصطناعي عن طريق إنشاء مواد تعليمية لمختلف الفئات العمرية بالمجتمع. وتتاح هذه المواد التعليمية بشكل مفتوح للجميع لزيادة الوعي بمجالات الذكاء الاصطناعي وأثره على المجتمع وإيجابيات وسلبيات استخداماته، وأهمية البيانات خلال دورة حياة الذكاء الاصطناعي. كما أنه من

المناسب زيادة الحملات الإعلامية مع مراعاة التنوع اللغوي والاجتماعي والثقافي لضمان وصول الفائدة للجميع ومساعدة الفرد لاتخاذ القرارات المناسبة حول استخدام أنظمة الذكاء الاصطناعي. ويمكن تنظيم التدريبات المشتركة بين دول مجلس التعاون لتحقيق التكامل المعرفي حول آليات عمل أنظمة الذكاء الاصطناعي.

المبدأ السابع: النزاهة وعدم التزييف

ينبغي تبني مبدأ النزاهة في استخدام الذكاء الاصطناعي والذي يتمثل في الحرص على عدم تزييف الحقائق وعدم المبالغة أو التضخيم أو التشويه لقدرات وإمكانيات الذكاء الاصطناعي لأهداف وأغراض جانبية مثل جني الأرباح أو تحقيق ميزات تنافسية أو التأثير على الرأي العام بما يؤثر سلباً على المستفيدين الأساسيين، حيث ينبغي مراعاة المصالح الأساسية للأفراد والمتعاملين الأساسيين في المقام الأول.

توصيات ومقترحات لتحقيق هذا المبدأ
- **وضع آليات للتحقق من تطبيق أسس ومعايير محددة حول استحداث أو تعميم أنظمة الذكاء الاصطناعي:** توصي هذه الوثيقة بوضع آليات واضحة تتيح إمكانية التحقق من تطبيق أسس ومعايير محددة يتم الاستناد عليها عند اتخاذ قرارات حول استحداث أو تعميم أنظمة الذكاء الاصطناعي، كما ينبغي أن تكون الأسس والمعايير واضحة وموثقة ومتاحة لأي عملية تقييم مصرّح بها من قبل المدققين الداخليين والخارجيين.

استخدام هذا الدليل والانتفاع به وتطبيقه

ينبغي على الدول الأعضاء وسائر الجهات المعنية احترام وتعزيز ونشر وحماية القِيم والمبادئ والمعايير الأخلاقية المبيَّنة في هذه الوثيقة فيما يخص الذكاء الاصطناعي؛ وينبغي عليها أيضاً اتخاذ كل التدابير الممكنة لتطبيق بنود هذا الدليل.

كما ينبغي على الدول الأعضاء أن تسعى إلى توسيع نطاق عملها بشأن هذا الدليل، وتكميل هذا العمل عن طريق التعاون مع جميع المنظمات الوطنية والدولية الحكومية وغير الحكومية المعنية، وكذلك مع الشركات العابرة للحدود الوطنية ومع المنظمات العلمية التي تندرج أنشطتها في نطاق هذا الدليل وضمن أغراضه وأهدافه. ويمكن أن يكون وضع منهجية لتقييم العواقب الأخلاقية، وإنشاء لجان وطنية لأخلاقيات الذكاء الاصطناعي، وسيلتَين مهمتَين لهذا الأمر.

الترويج لأخلاقيات الذكاء الاصطناعي

سيقوم مجلس التعاون لدول الخليج العربية بالترويج لهذا الدليل من خلال نشره في الموقع الرسمي للأمانة العامة للمجلس وبوابة الحكومة الإلكترونية للمجلس، وكذلك نشره عبر بوابات الحكومة الإلكترونية ومواقع التواصل الاجتماعي الرسمية الخاصة بالدول الأعضاء بحسب ما تراه الدول الأعضاء مناسباً.

وينبغي أن تؤخذ هذه الوثيقة بعين الاعتبار عند استحداث سياسات أو تشريعات جديدة أو عند إعداد استراتيجيات وطنية للحكومة الإلكترونية والتحول الرقمي والاقتصاد الرقمي وأمن المعلومات، بما في ذلك البرامج والمبادرات والمشاريع المنبثقة منها.

أحكام ختامية

هذه وثيقة واحدة لا تتجزّأ، وينبغي أن تُعتبر القِيم والمبادئ الأساسية الواردة فيها مبادئ متكاملة ومترابطة.

لا يجوز تفسير أي حكم من أحكام هذه الوثيقة تفسيراً يرمي إلى إحلال أي حكم من أحكامه محلّ واجبات الدول أو حقوقها أو تفسيراً يؤدي إلى تغيير هذه الواجبات والحقوق أو إلى المساس بها بطريقة أخرى.