دبي الذكية

مبـادئ وإرشـادات أخلاقيات الذكاء الاصطناعي

دبـي الذكية
SMART DUBAI

دبـــي الذكية
SMART DUBAI

المحتويات

"رؤيتنا لدبي هي التميز في استخدام وتطوير الذكاء الاصطناعي لسعادة ونفع البشر"

د. عائـشة بنت بطي بن بـشـر
مدير عام - مكتب دبي الـذكية

المسؤولية

لا يتحمل مكتب دبي الذكية أي مسؤولية قانونية قد تنجم عن سوء استخدام مبـادئ وإرشـادات أخلاقيات الذكاء الاصطناعي، ويتحمل المستخدم كافة تبعات هذا الاستخدام.

الترخيص

مبادئ دبي للذكاء الاصطناعي

مبادئ دبي للذكاء الاصطناعي

الأخلاقيات

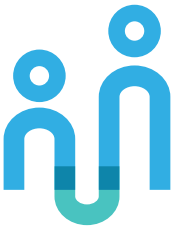
يجب أن تكون أنظمة الذكاء
الاصطناعي عادلة وتطبق الشفافية
وخاضعة للمساءلة وقابلة للفهم

الأمان

يجب أن تكون أنظمة الذكاء
الاصطناعي آمنة كما يجب أن تسخر
في خدمة وحماية الإنسانية

الشمولية

يجب أن ينفع الذكاء الاصطناعي كافة
أفراد المجتمع، كما يجب أن تُطبق
عليه الحوكمة عالمياً، مع احترام
كرامة الأفراد وحقوقهم



البشرية

يجب أن يكون الذكاء الاصطناعي
نافعًا للبشرية وأن ينسجم مع القيم
الإنسانية، على الأمدين القصير والبعيد

مبادئ دبي للذكاء الاصطناعي

الأخلاقيات

سنجعل أنظمة الذكاء الاصطناعي عادلة

- يجب أن تمثّل البيانات التي يتلقاها النظام الفئة المتأثرة، حيثما أمكن
- يجب أن تتجنب الخوارزميات التحيز غير التشغيلي*
- يجب اتخاذ الإجراءات التي تحد من وتقوم بتقييم أي انحياز في مجموعات البيانات
- يجب إثبات عدالة القرارات الهامة

سنجعل أنظمة الذكاء الاصطناعي قابلة للمساءلة

- لا تكمن المساءلة عن نتائج نظام الذكاء الاصطناعي في النظام بحد ذاته، بل إنها مقسّمة بين القائمين على التصميم والتطوير والتطبيق
- يجب أن يبذل المطوّرون جهدهم للحد من المخاطر المتأصلة في الأنظمة التي يقومون بتصميمها
- يجب أن تتوفر في أنظمة الذكاء الاصطناعي إجراءات مدمجة تتيح للمستخدمين الاعتراض على القرارات الهامة
- يجب أن تتولى فرق متنوعة أنظمة الذكاء الاصطناعي بحيث تتضمن خبراء في المجال الذي سيتم نشر النظام فيه

سنجعل أنظمة الذكاء الاصطناعي تتمتع بالشفافية

- يجب أن يقوم المطوّرون ببناء أنظمة يمكن تتبع وتشخيص حالات الإخفاق بها
- يجب أن يتم إعلام الناس كلما يتخذ الذكاء الاصطناعي قرارات هامة تخصهم
- ضمن حدود الخصوصية وحماية الملكية الفكرية، يجب أن يتمتع ناشرو أنظمة الذكاء الاصطناعي بالشفافية فيما يتعلق بالبيانات والخوارزميات التي يلجؤون إليها

سنجعل أنظمة الذكاء الاصطناعي قابلة للشرح تقنيا قدر الإمكان

- يجب أن نشرح للأفراد قرارات ومنهجيات أنظمة الذكاء الاصطناعي التي تؤثر تأثيرًا حيويًا عليهم، إلى الحد الذي تسمح به التكنولوجيا المتوفرة
- يجب توفير إمكانية التحقق من العوامل الأساسية التي تؤدي إلى اتخاذ أي قرار يمكن أن يؤثر تأثيرًا هاماً على الفرد
- في الحالة أعلاه، سنقوم بتوفير قنوات يمكن للأفراد التماس هذه الشروحات والتفسيرات عبرها

* المقصود بالتحيز غير التشغيلي هو التحيز غير المُصمم كسمة في نظام الذكاء الاصطناعي أو الذي لا يُعتبر مهمًا لتحقيق الغرض المعلن من النظام

مبادئ دبي للذكاء الاصطناعي

الأمان

ستكون أنظمة الذكاء الاصطناعي آمنة وخاضعة للتحكم من البشر

• سيكون أمن وأمان الناس، سواءً أكانوا مشغلين أم مستخدمين نهائيين أم أطراف ثالثة، على قمة سلم الأولويات في تصميم أي نظام من أنظمة الذكاء الاصطناعي

• يجب إتاحة إمكانية التحقق من أمن أنظمة الذكاء الاصطناعي وإمكانية التحكم بها طوال فترة حياتها التشغيلية، إلى الحد الذي تسمح به التكنولوجيا

• يجب مراعاة أمان وخصوصية المستخدمين عند إيقاف تشغيل أنظمة الذكاء الاصطناعي
يجب إيلاء العناية الواجبة بأنظمة الذكاء الاصطناعي التي تؤثر مباشرةً على حياة الناس تأثيرًا حيويًا أثناء مراحل تصميمها

• يجب مراعاة إمكانية إبطال مثل هذه الأنظمة أو إلغاء قراراتها بواسطة الأشخاص المعنيين المخولين

يجب عدم تمكين أنظمة الذكاء الاصطناعي من إلحاق أي أذى أو تخريب أو تضليل البشر

• يجب بناء أنظمة الذكاء الاصطناعي بهدف تقديم الخدمات والمعلومات وليس للخداع والتلاعب

• على الدول التعاون لتجنب سباقات التسلح بالأسلحة ذاتية التحكم المميتة، كما يجب فرض الرقابة الشديدة على هذه الأسلحة

• يجب تعزيز التعاون الفعال لتجنب تجاهل معايير السلامة

• يجب أن تقوم الأنظمة المُصممة لتوفير المعلومات للقرارات الهامة بهذه الوظيفة بموضوعية

مبادئ دبي للذكاء الاصطناعي
البشرية

سنضيف إلى أنظمة الذكاء الاصطناعي قيمًا إنسانية وسنجعلها مفيدة للمجتمع

- ستقوم الحكومة بتمويل بحوث الاستخدام النافع للذكاء الاصطناعي
- يجب تطوير الذكاء الاصطناعي ليتلاءم مع القيم الإنسانية ولكي يساهم في ازدهار البشر
- يجب أن تساهم الجهات المعنية في المجتمع كله في تطوير الذكاء الاصطناعي وحوكمته

سنقوم بالتخطيط لمستقبل سيتنامى فيه ذكاء أنظمة الذكاء الاصطناعي

- يجب إعداد نماذج الحوكمة للذكاء الاصطناعي العام (AGI) والذكاء الخارق
- سيسخر الذكاء الاصطناعي العام (AGI) والذكاء الخارق في خدمة الإنسانية ككل، إذا تم تطويرهما
- يجب تحديد المخاطر طويلة الأمد للذكاء الاصطناعي والاستعداد لها
- يجب الإفصاح عن أي تطوير لحلول الذكاء الاصطناعي بخاصية التحسين الذاتي المستمر ومراقبته عن كثب والتحكم بمخاطره

مبادئ دبي للذكاء الاصطناعي
الشمولية

سنقوم بحوكمة الذكاء الاصطناعي كجهد تعاوني عالمي

- يجب تشجيع جهود التعاون الدولي لضمان حوكمة الذكاء الاصطناعي بشكل آمن
- ستساهم الحكومة في إرساء أفضل الممارسات والمعايير المعترف بها دوليًا الخاصة بالذكاء الاصطناعي، والالتزام بها بعد إرسائها

سنقوم بمشاركة منافع الذكاء الاصطناعي مع جميع أفراد المجتمع

- سيتم تنسيق تطوير أنظمة الذكاء الاصطناعي عن طريق الاستجابة لأثره على التوظيف
- سيتم استخدام الذكاء الاصطناعي لمساعدة البشر على تحقيق ذاتهم وعلى الازدهار عقليًا وعاطفيًا واقتصاديًا إلى جانب الذكاء الاصطناعي
- سيتم توفير التدريب والفرص والأدوات للجميع
- يجب أن يتطور التعليم ليعكس أحدث المستجدات في مجال الذكاء الاصطناعي مما سيسمح للناس بالتكيف مع التغييرات المجتمعية الناجمة عنه

سنعزز قيم الإنسانية والحرية والاحترام

- يجب أن يطوّر الذكاء الاصطناعي المجتمع، ويجب إشراك أفراد المجتمع بشكل يمثلهم لتشكيل ملامح تطور الذكاء الاصطناعي
- يجب أن تحتفظ البشرية بقدرتها على أن تحكم نفسها وتتخذ القرار النهائي في شؤونها، على أن يكون الذكاء الاصطناعي مسانداً لذلك
- يجب أن تتوافق أنظمة الذكاء الاصطناعي مع الأعراف والمعايير الدولية فيما يتعلق بالقيم الإنسانية وحقوق الأفراد والسلوكيات المقبولة

سنحترم خصوصية الأفراد

- يجب أن تحترم أنظمة الذكاء الاصطناعي الخصوصية وتستخدم الحد الأدنى الضروري من التدخل
- يجب أن تتبنى أنظمة الذكاء الاصطناعي أعلى معايير أمن وحوكمة البيانات لحماية المعلومات الشخصية
- يجب عدم استعمال تقنيات المراقبة أو تقنيات الأخرى بشكل ينتهك المعايير المتعارف عليها دوليًا و/أو فيما يخص الخصوصية والقيم الإنسانية وحقوق الأفراد

مبادئ دبي للذكاء الاصطناعي

سجل التغييرات

الإصدار	المرحلة	تاريخ الإتمام	ملخص التغييرات
1.0	داخلية	2018.9.6	المسودة الأولى
1.1	داخلية	2018.9.6	تحسين
1.2	استشارات	2018.9.10	إضافة شرح تحت كل مبدأ
1.3	استشارات	2018.9.10	التغييرات عقب الاجتماع التوجيهي الثاني: المواءمة مع البشرية
1.4	استشارات	2018.9.25	التغييرات عقب الاجتماع التوجيهي الثاني: المساواة مع الشمولية
1.5	آراء وملاحظات	2018.10.9	التغييرات عقب الجولة الأولى من الآراء والملاحظات
1.6	مراجعة داخلية	2018.12.30	تعديلات ومراجعة لصياغة بعض أجزاء المحتوى وإضافة المقدمة

هذه الوثيقة من إعداد دبي الذكية بالتعاون مع فالكون وشركائها.

إرشادات أخلاقيات الذكاء الاصطناعي

إرشادات أخلاقيات الذكاء الاصطناعي

المقدمة:

التطوّر السريع وفرص الابتكار التي تشهدها تقنية الذكاء الاصطناعي في مختلف المجالات مشوّقة. ورغم ذلك إلا أن المؤسسات التي تستخدم هذه التقنية لم تناقش بعد بشكل عميق وتفصيلي شامل المبادئ والأخلاقيات التي يجب مراعاتها أثناء استخدام الذكاء الاصطناعي، والعالم يحتاج هذه المبادئ والأخلاقيات بصفة عاجلة.

ولذلك أوجدنا منظومة أخلاقيات الذكاء الاصطناعي هذه لتكون دعماً عملياً عند تبني الذكاء الاصطناعي عبر منظومة المدن. وهي توفر لخبراء التقنية والمهتمين من الأكاديميين والأفراد دليلاً لكيفية استخدام تقنية الذكاء الاصطناعي بشكل مسؤول. وهذه المنظومة تتضمن مبادئ وإرشادات وأداة تقييم ذاتي تتيح للمطورين تقييم أنظمة الذكاء الاصطناعي التي يطورونها.

إن إرشادات دبي لأخلاقيات الذكاء الاصطناعي منبثقة من مبدأ الأخلاقيات ضمن مجموعة مبادئ دبي للذكاء الاصطناعي:

«يجب أن تكون أنظمة الذكاء الاصطناعي عادلة وتطبق الشفافية وخاضعة للمساءلة وقابلة للشرح»

توفر الإرشادات اقتراحات عملية لمساعدة الجهات والأفراد المعنيين بالذكاء الاصطناعي من الالتزام وتطبيق ماينص عليه مبدأ الأخلاقيات، وهي دليل تفصيلي يغطي المبادئ الفرعية الأربعة لمبدأ الأخلاقيات وهي:

- سنجعل أنظمة الذكاء الاصطناعي **عادلة**
- سنجعل أنظمة الذكاء الاصطناعي **قابلة للمساءلة**
- سنجعل أنظمة الذكاء الاصطناعي تتمتع **بالشفافية**
- سنجعل أنظمة الذكاء الاصطناعي **قابلة للشرح** تقنيا قدر الإمكان

إن هذه إرشادات غير مُلزمة، وتمت صياغتها كجهد تعاوني ومشترك بين الجهات المعنية، مع الإدراك التام لأهمية وحاجة المؤسسات إلى الابتكار وحماية حقوق ملكيتهم الفكرية.

[1] بما يتفق مع ISO/IEC 2382:2015

نطاق العمل

تقدّم هذه الوثيقة الإرشادات التي تُعنى بتحقيق التصميم والتطبيق الذي يراعي أخلاقيات الذكاء الاصطناعي للأنظمة والحلول في القطاعين العام والخاص. وعلى وجه التحديد، تغطي هذه الوثيقة مسائل العدالة والمساءلة والشفافية وقابلية التفسير والفهم التقني. ولا تغطي المسائل المتعلقة بالتوظيف أو الأمن أو أي جوانب أخرى تتعلق بحوكمة الذكاء الاصطناعي باستثناء تلك المسائل المذكورة أعلاه.

الذكاء الاصطناعي يحيط بنا أينما تواجدنا، ولكن بعض تطبيقاته وأنظمته ظاهرة أكثر من غيرها لنا. وتنطبق هذه الاتفاقية على أنظمة الذكاء الاصطناعي التي تزوّد معلومات هامة أو تتخذ قرارات هامة ومصيرية سواءاً كانت تمس الأفراد أو المجتمع ككل. كما يمكن أن تنطبق على «القرارات الحرجة» التي تتفرع من القرارات الهامة وتتمتع بطبيعة حرجة بشكل خاص. راجع قسم التعريفات للاطلاع على تعريف «القرار الهام» و«القرار الحرج».

تعريفات

لغايات العمل بهذه الإرشادات، تسري شروحات التعريفات التالية:

مؤسسة تطوير أنظمة الذكاء الاصطناعي

مؤسسة تؤدي أي من الوظائف التالية:

- تحديد الغرض من نظام الذكاء الاصطناعي
- أو تصميم نظام الذكاء الاصطناعي
- أو بناء نظام الذكاء الاصطناعي
- أو إجراء الصيانة الفنية أو ضبط وتعديل نظام الذكاء الاصطناعي

ملاحظة 1 على التعريف أعلاه: يسري هذا التعريف بغض النظر عما إذا كانت المؤسسة هي المستخدم النهائي للنظام أو كانت تبيعه أو توزعه دون مقابل.

مثال:

تقوم شركة ما بتطوير نظام ذكي اصطناعي للتعرف على الوجه، ثم تبيعه إلى قوات حرس الحدود في دولة ما والتي تستخدمه بدورها للتعرّف على الموظفين المشتبه بهم. فتكون الشركة هي مطوّر نظام الذكاء الاصطناعي وتكون قوات حرس الحدود هي مشغّل نظام الذكاء الاصطناعي.

مؤسسة تشغيل أنظمة الذكاء الاصطناعي

هي المؤسسة التي تؤدي أي من الوظائف التالية:

- تستخدم أنظمة الذكاء الاصطناعي في العمليات التشغيلية، أو العمليات الداخلية، أو اتخاذ القرارات
- تستخدم أنظمة الذكاء الاصطناعي لتقديم خدمة للأشخاص المشاركين بعمليات نظام الذكاء الاصطناعي
- أو هي المالك لنظام الذكاء الاصطناعي
- أو هي المسؤولة عن جمع البيانات ومعالجتها من أجل استخدامها في نظام الذكاء الاصطناعي
- أو هي المسؤولة عن تقييم حالات وأوجه استخدام نظام ذكاء اصطناعي وتتخذ قرار اعتمادها وتنفيذها

ملاحظة 1 على التعريف أعلاه: ينطبق هذا التعريف سواءاً تم تطوير نظام الذكاء الاصطناعي داخليًا في المؤسسات المذكورة بالفئات أعلاه أم تم شراؤه من مزوّد.

ملاحظة 2 على التعريف أعلاه: يمكن أن تكون المؤسسة ذاتها مسؤولة عن تطوير نظام الذكاء الاصطناعي وتشغيله معاً

الذكاء الاصطناعي

(يُشار إليه أيضًا بالاختصار «AI»)

قدرة وحدة تقنية على أداء مهام ووظائف ترتبط عادة بقدرات الذكاء البشري، مثل الربط المنطقي بين المعطيات والتعلّم وتطوير الذات[1].

نظام الذكاء الاصطناعي

(يُشار إليه أيضًا بالاختصار «نظام AI»)

منتج أو خدمة أو عملية أو منهجية لاتخاذ القرارات يعتمد تشغيلها أو نتيجة عملها على وحدات تقنية تعمل بتقنية الذكاء الاصطناعي

ملاحظة 1 على التعريف أعلاه: ليس ضرورياً أن تكون مخرجات نظام تقني ما ناتجة بالكامل عن عمليات وحدات تقنية عاملة بالذكاء الاصطناعي حتى نتمكن من تعريف النظام بأنه نظام ذكاء اصطناعي

ملاحظة 2 على التعريف أعلاه: من السمات المميزة لأنظمة الذكاء الاصطناعي هي أنها تتعلم سلوكيات وقواعد لم تكن مبرمجة فيها بشكل مسبق محدد.

[1] بما يتفق مع ISO/IEC 2382:2015

مثال:

تستخدم محكمة مختصة بالنظر في الدعاوى الصغيرة برمجيات ذكاء اصطناعي لجمع الأدلة التي تتعلق بإحدى القضايا ومقارنتها مع القضايا المماثلة السابقة، ومن ثم تقوم باقتراح حُكمٍ للقاضي الذي يصدر الحكم النهائي. ونظراً لأن منهجية اتخاذ القرارات تتم بتأثير من برمجة الذكاء الاصطناعي فهي تعتبر نظام ذكاء اصطناعي.

مثال:

تستخدم هيئة حكومية روبوتات الدردشة التي تتيح لجمهورها طرح الأسئلة الروتينية وحجز المواعيد وإجراء معاملات مالية ثانوية. وترد روبوتات الدردشة على استفسارات العملاء عن طريق إجابات مكتوبة مسبقًا وتعتمد على قواعد مبرمجة مسبقًا لاتخاذ قرارات معينة. وفي هذه الحالة لا تعتبر روبوتات الدردشة نظام ذكاء اصطناعي، ولكن إذا أصبحت روبوتات الدردشة قادرة على معالجة استفسارات العملاء ذاتياً بناءً على تحليلها لنتائج الحالات السابقة، فيمكن أن نعتبرها حينها نظام ذكاء اصطناعي.

تحيز

(النظام التقني)

ميل أو انحياز لمصلحة أو ضد مصلحة شخص أو مجموعة أشخاص، وخاصة بطريقة تُعتبر غير عادلة[2].

القرار الحرج

هو قرار فردي هام له تأثير كبير على الفرد أو يتضمن على مخاطر جسيمة بشكل خاص – سواءً كان حساسًا جدًا أو يمكن أن يسبب خسائر أو أضراراً كبرى – أو مؤثر وهام على مستوى المجتمع أو يسجّل سابقة مهمة.
ملاحظة ا على التعريف أعلاه: إن أنواع القرارات المشار إليها هنا هي نفس القرارات التي تظهر في تعريف القرارات الهامة من حيث حجم التأثير، ولكن في هذه الحالة يتم استشعار التأثيرات نتيجة لاتخاذ قرار فردي وليس نتيجة عدة قرارات معًا.

مثال:

تقرر المحكمة إذا كان المدعى عليه مذنبًا بتهمة جنائية، مع العلم أن عقوبة إثبات الإدانة هي السجن المؤبد مدى الحياة. إن هذا القرار هو قرار حرج لأن تأثيره سيكون ضخمًا على حياة المتهم كما أن الحكم في هذه القضية سيصبح نقطة مرجعية للحكم في القضايا المماثلة مستقبلاً.

[2] Oxford Dictionary ,2018 Oxford University Press, تاريخ الزيارة 4 أكتوبر 2018, <https://en.oxforddictionaries.com>

الأخلاقيات

(كما تنطبق على الذكاء الاصطناعي)

تعني مفاهيم العدالة والمساءلة والشفافية وقابليتها للشرح.

ملاحظة 1 على التعريف أعلاه: لغايات هذه الوثيقة، لا تتضمن أخلاقيات الذكاء الاصطناعي مسائل الخصوصية ولا دقة نموذج العمل (ما عدا ما يتعلق بالعدالة والإصلاح على سبيل المثال) أو التوظيف أو أي مسائل أخرى تتعلق بالذكاء الاصطناعي باستثناء تلك المذكورة في التعريف.

الوحدة الوظيفية

هي وحدة معدات تقنية أو برمجيات أو كلاهما معاً بحيث تكون قادرة على تحقيق غرض محدد[3].

القرار الهام على المستوى الفردي

هو قرار قد يكون له تأثير ضخم على ظرف واحد على الأقل من ظروف الفرد أو سلوكياته أو خياراته، أو له تداعيات قانونية أو هامة أخرى على الفرد.

مثال:

قررت شركة من الشركات تسريح أحد موظفيها. يُعتبر هذا قراراً هاماً على المستوى الفردي لأنه سيؤثر على الوضع المالي للموظف.

التحيز غير التشغيلي

(للنظام)

هو التحيز الذي يكون إما:

1. غير مُصمم كسمة في النظام
2. أو غير مهم في تحقيق الغرض المعلن عنه من وجود النظام

[3] نقلًا عن ISO/IEC 2382:2015

مجموعة القرارات الهامة على المستوى الجماعي

هي مجموعة من القرارات التي يتخذها النظام نفسه أو المؤسسة نفسها، والتي عندما تتراكم يكون لها تأثير هام على المجتمع ككل أو على مجموعات معنية في المجتمع.

ملاحظة 1 على التعريف أعلاه: لا يلزم أن تكون القرارات هامة على المستوى الفردي لكي يتم اعتبارها – بعد تراكمها – على أنها قرارات هامة على المستوى الجماعي

ملاحظة 2 على التعريف أعلاه: تتضمن الأمثلة على المجالات التي تؤثر تأثيرًا بالغًا على المجتمع المجالات التالية على سبيل المثال لا الحصر: تخصيص الموارد أو توزيع الفرص بين المجموعات على نطاق واسع؛ تأثيرها على الهيكل الحكومي؛ التأثير على توازن القوى بين جهات أو مجموعات كبيرة؛ القانون وتفسيراته وإنفاذه؛ النزاعات والحروب؛ العلاقات الدولية؛ إلى آخره.

مثال:

يستخدم موقع إلكتروني نظاماً للذكاء الاصطناعي لتحديد أي المحتويات يتم عرضها للمستخدمين من زوّار الموقع. هذا القرار لا يعتبر هاماً على المستوى الفردي لأن المستخدمين لن يتأثروا تأثيرًا بالغًا إذا تم عرض خبر ما لهم أم لا. ولكن، إذا كان للموقع الإلكتروني شعبية كبيرة بين الناس، فقد يتخذ نظام الذكاء الاصطناعي مجموعة من القرارات الهامة على المستوى الجماعي لأن أي انحياز في نظام الذكاء الاصطناعي في الموقع الإلكتروني سيؤثر على عدد كبير من المستخدمين.

القرار الهام

هو قرار هام على المستوى الفردي، أو أنه جزء من القرارات الهامة على المستوى الجماعي.

الشخص المشارك في نظام الذكاء الاصطناعي

يُشار له أيضاً بالاختصار («موضوع AI»)

هو شخص طبيعي تتوفر فيه واحدة من الصفات التالية:

- المستخدم النهائي لنظام الذكاء الاصطناعي
- أو الشخص المتأثر مباشرةً بتشغيل أو بنتائج ومخرجات نظام الذكاء الاصطناعي
- أو متلقي لخدمة أو توصية يقدمها نظام الذكاء الاصطناعي

الإرشادات

1.1 سنحقق الاستخدام العادل لأنظمة الذكاء الاصطناعي

1.1.1 يجب مراعاة أن تكون البيانات التي يتم تغذية الأنظمة بها والواردة إليها تعكس واقع الفئة المتأثرة

1.1.1.1 يجب أن تقوم مؤسسات تشغيل وتطوير أنظمة الذكاء الاصطناعي ببحث منطقي عن البيانات و/أو بتقييم البيانات لكي تحدد احتمالية اتخاذ أية قرارات قد تكون مجحفة وتنشأ عن أي انحياز ناجم عن البيانات

مثال:

بعد وقوع كارثة طبيعية، تستخدم وكالة الإنقاذ الحكومية نظام ذكاء اصطناعي لكي تحدد أكثر المجتمعات احتياجًا للإغاثة عن طريق تحليل بيانات وسائل التواصل الاجتماعي من عدة مواقع إلكترونية. ولكن المجتمعات ذات الانتشار الأقل للهواتف الذكية يعتبر محدوداً على وسائل التواصل الاجتماعي، وهو ما قد ينجم عنه عدم تلقيهم للدعم المطلوب. ولهذا تجمع وكالات الإغاثة بين أدوات الذكاء الاصطناعي والأساليب التقليدية لكي تستطيع التعرف على المجتمعات الأكثر احتياجاً للإغاثة في مناطق أخرى.

1.1.1.2. يجب أن تتجنب مؤسسات تطوير وتشغيل أنظمة الذكاء الاصطناعي تدريب تلك الأنظمة على بيانات قد لا تمثّل الأشخاص المتأثرين والمستهدفين بالذكاء الاصطناعي، أو حتى البيانات التي يحتمل عدم دقتها بسبب التقادم أو الحذف أو أسلوب الجمع أو أية عوامل أخرى.

1.1.1.3. يجب أن تراعي مؤسسات تطوير أنظمة الذكاء الاصطناعي مدى كفاءة أنظمتها عند معالجتها لبيانات لم يتم تغذيتها بها مسبقًا، وخاصة عند استخدام الأنظمة لتقييم أشخاص غير ممثلين بشكل كافٍ في البيانات المستخدمة في تدريب النظام.

1.1.2. يجب التحقق من مدى وجود أي انحياز في عمليات اتخاذ القرارات

1.1.2.1. عند إخضاع مجموعات مختلفة إلى عمليات اتخاذ قرارات مختلفة، يجب أن تأخذ مؤسسات تطوير أنظمة الذكاء الاصطناعي بعين الاعتبار إذا ما كان هذا سيؤدي إلى انحياز غير تشغيلي

1.1.2.2. عند تقييم عدالة نظام الذكاء الاصطناعي، يجب أن تأخذ مؤسسات تطوير وتشغيل أنظمة الذكاء الاصطناعي بالاعتبار إذا ما كان الأشخاص المشاركون في الذكاء الاصطناعي يتلقون معاملة متساوية.

مثال:

تستخدم مؤسسة من المؤسسات أداة ذكاء اصطناعي لأتمتة الفرز المبدئي للمتقدمين لوظيفة ما. وقد تم تدريب الأداة على بيانات من موظفي الشركة الحاليين الذين يشترك معظمهم في الخلفية العرقية نفسها. ولهذا يتعلم نظام الذكاء الاصطناعي استخدام الاسم والجنسية كعوامل تمييزية عند ترشيح طلبات الوظيفة. وهو أمر كان يمكن تجنبه بإجراء عمليات تجريبية واختبار للنظام ومعالجة الخلل عن طريق إيجاد توازن في بيانات التدريب أو استخدام حقول البيانات ذات العلاقة في موضوع التدريب.

1.1.3 يجب توفر العدالة في أية قرارات هامة يتم اتخاذها بناءاً على الذكاء الاصطناعي

1.1.3.1. يمكن أن تعتمد مؤسسات تطوير وتشغيل أنظمة الذكاء الاصطناعي إجراءات رسمية مثل اختبار أثر الانحياز في النتائج كوسيلة لضمان العدالة في القرارات.

1.1.4. يجب أن تراعي مؤسسات تشغيل الذكاء الاصطناعي مدى إتاحة أنظمتها إمكانية الوصول والاستخدام لأنظمة الذكاء الاصطناعي بطريقة عادلة لمختلف مجموعات المستخدمين

1.1.5 يجب مراعاة تأثير التنوع الديموغرافي بمختلف مراحل عمليات التطوير وتطبيق حلول الذكاء الاصطناعي

1.1.5.1. ينبغي بذل الجهود لإشراك الناس من مختلف البيئات الديموغرافية في عمليات التطوير والتطبيق

1.1.5.2. يجب أن تضع مؤسسات تطوير الذكاء الاصطناعي احتمالية أن تكون الفرضيات التي كونتها عن الأشخاص المشاركين في الذكاء الاصطناعي تحتمل الخطأ أو أنها قد تؤدي إلى تحيز غير تشغيلي؛ وإن حدث هذا فينبغي أن تستشير الأشخاص المشاركين في الذكاء الاصطناعي بطريقة تضمن تمثيل العينة المشاركة أثناء تطوير ونشر الأنظمة لغايات تأكيد هذه الفرضيات.

1.2. سنجعل أنظمة الذكاء الاصطناعي قابلة للمساءلة

1.2.1 يجب ألا تكون المساءلة عن مخرجات نظام الذكاء الاصطناعي واقعة في النظام نفسه

1.2.1.1 ينبغي ألا تُعزى المساءلة عن الأضرار أو الخسائر التي تنتج عن تطبيق أنظمة الذكاء الاصطناعي إلى النظام نفسه

1.2.1.2 يجب أن تراعي مؤسسات تطوير وتشغيل أنظمة الذكاء الاصطناعي تعيين أشخاص ليكونوا مسؤولين عن التحقيق في أية خسائر أو أضرار قد تنشأ عن أنظمة الذكاء الاصطناعي وتصويبها

1.2.2 يجب المبادرة بجهود تُسهم مسبقاً في التعرّف على أية مخاطر هامة تتضمنها طبيعة النظام المُصمّم والحد من أثرها

1.2.2.1 يجب أن تستخدم مؤسسات تشغيل أنظمة الذكاء الاصطناعي سوى الأنظمة المدعومة ببحوث أكاديمية موثوقة ومبنية على أدلة علمية، ويجب على المؤسسات المطورة لهذه الأنظمة أن تعتمد في عمليات تطويرها على هذه البحوث

1.2.2.2 يجب على مؤسسات تشغيل أنظمة الذكاء الاصطناعي أن تحدد التأثير المحتمل للقرارات المؤتمتة الخاطئة على الأشخاص المشاركين في الذكاء الاصطناعي في حالة وجود احتمالات تشير إلى أن هذه القرارات الخاطئة يمكن أن تسبب فرض تكاليف باهظة أو إزعاج أو مضايقات، مع مراعاة تدابير الحد منها

مثال:

لدى دولة ما خدمة حكومية تتعرّف على الآباء المدينين بالمال لصندوق رعاية الأطفال. لكن عملية مطابقة بيانات المدينين غالبًا ما تكون خاطئة بسبب وجود أخطاء هجائية في الأسماء أو نقص في البيانات، مما يجعل النظام يستهدف بعض الأشخاص آليًا بطريق الخطأ ويرفع قيمة الدفعات المستحقة عليهم أو ضعف تقييمهم الائتماني أو حتى إيقاف الأجور. ويكون الرجوع إلى الأشخاص الذين تم استهدافهم بطريق الخطأ مضيعة للوقت والجهد والموارد[4]. وكل ذلك كان يمكن تجنبه لو تم إجراء تقييم الأثر المحتمل للقرارات الخاطئة، ووضع آليه سهلة الاستخدام من قبل الجمهور المعني لمراجعة نتائج الذكاء الاصطناعي كان سيجعل الحل أكثر سهولة.

1.2.2.3 يجب أن تعتمد مؤسسات تشغيل الذكاء الاصطناعي أطراً أخلاقية أو أدوات لتقييم المخاطر الداخلية كوسيلة للتعرف على المخاطر مسبقاً وتحديد سبل الحد منها.

[4] مكتب رئاسة مجلس الوزراء (المملكة المتحدة)، إطار أخلاقيات علوم البيانات، الإصدار 1.0، مرخص بموجب رخصة الحكومة المفتوحة، الإصدار 3.0، الصفحة 13

1.2.2.4 عند تصميم أنظمة الذكاء الاصطناعي التي تمد المعلومات المتعلقة بالقرارات الهامة، يجب على مؤسسات تطوير الأنظمة اتخاذ تدابير للحفاظ على دقة البيانات مع مرور الوقت، بما في ذلك:

- اكتمال البيانات
- وتحديث البيانات بشكل دوري مستمر
- وتقييم إذا ما كان السياق الذي تم جمع البيانات فيه يؤثر على ملائمتها لحالة الاستخدام المقصودة

مثال:

تقوم إحدى كاميرات النقاط الحدودية بمسح ومراقبة الأشخاص بحثًا عن إشارات تنبئ بخطر ما، وفي حالة أمراض مثل متلازمة توريت –التي تتضمن تقلّص عضلات الوجه - ستخلط الكاميرا بين أعراض هذا المرض ومعايير تحديد المشتبه بهم. وهو مثال على أخطاء قد تظهر بطرق متعددة، ولذا يجب ألا يتم إخضاع هذا الشخص إلى تفتيش جديد في كل مرة يعبر فيها الحدود بعد حصول الخلط الأوّل[5]. وإذا تم تحديث البيانات بعد مواجهة أول حالة مثل المذكورة، فيمكن تلافي التسبب بإزعاج هذا الشخص في زياراته اللاحقة.

1.2.2.5 يجب أن تأخذ مؤسسات تطوير وتشغيل أنظمة الذكاء الاصطناعي بالاعتبار ضبط وتعديل نماذج الذكاء الاصطناعي دوريًا لغايات استيعاب أية تغييرات تطرأ على البيانات و/أو نماذج العمل مع مرور الوقت.

1.2.2.6 يجب أن تأخذ مؤسسات تشغيل أنظمة الذكاء الاصطناعي بالاعتبار أن بعض الأنظمة التي يتم تدريبها في بيئات ثابتة نسبيًا ستكشف عدم استقرار في أدائها أو طريقة عملها عند نشرها في بيئات ديناميكية متغيرة.

مثال:

تحتاج أنظمة الذكاء الاصطناعي لأن تكون قادرة على التكيّف مع التغييرات الموجودة بالبيئة التي تُنشر بها. ومنها على سبيل المثال، يجب أن تكون المركبات ذاتية القيادة قادرة على التعامل الفوري اللحظي مع أية مخاطر أو ظروف تطرأ على الطريق من خلال استفادتها وتعلّمها من مواجهة المركبات ذاتية القيادة الأخرى التي واجهت هذه المخاطر أو الظروف قبلها. وكما يجب أن تكون تطبيقات الذكاء الاصطناعي المتعلقة بمهام حرجة مثل هذه الحالة قادرة على عزل أي معطيات تسبب تشويشاً والحماية من أية عوامل ضارة[6].

1.2.2.7 يجب أن تحرص مؤسسات تشغيل الذكاء الاصطناعي التعاون المستمر مع المزوّدين (مؤسسات تطوير أنظمة الذكاء الاصطناعي) لمراقبة أداء أنظمتها بشكل مستمر.

1.2.2.8 يجب أن تُخضع مؤسسات تشغيل أنظمة الذكاء الاصطناعي هذه الأنظمة التي تمد القرارات الهامة بالمعلومات لإجراءات فحص جودة تُماثل تلك التي تتم على موظف إنسان يتخذ هذا النوع من القرارات

[5] حكومة كندا. (2018). الذكاء الاصطناعي المسؤول في حكومة كندا . سلسلة تقارير الثورة الرقمية. الإصدار 2.0، ص 26. متوفر على: 1Sn-qBZUXEUG4dVk909eSg5qvfbpNlRhzIefWPtBwbxY/edit/https://docs.google.com/document/d

[6] Stoica, I. et. al, 2017, p. 2, https://arxiv.org/pdf/1712.05855.pdf

1.2.3 معلّق - يجب أن تكون أنظمة الذكاء الاصطناعي المرتبطة بقرارات حرجة متاحة لإجراء التدقيق الخارجي عليها

1.2.3.1. عندما تستخدم أنظمة الذكاء الاصطناعي لاتخاذ قرارات حرجة، يجب أن يكفل إجراء التدقيق الخارجي مدى الالتزام بمعايير الشفافية والمساءلة

1.2.3.2. إذا كانت القرارات الحرجة التي تتخذها أنظمة الذكاء الاصطناعي تخص المدنيين، يجب إعلان نتائج التدقيق عليها للعامة حتى تكون العمليات العامة من هذا النوع خاضعة للمساءلة من قبل الجمهور المتأثر بها باستمرار.

ملاحظة حول البنود المعلقة:

تم تعليق الإرشادات رقم 1.2.3 حتى إشعار آخر. السبب: عدم وجود آلية للتدقيق الخارجي حتى الآن.

1.2.4 يجب أن يتاح للأشخاص المشاركين والمتأثرين بأنشطة الذكاء الاصطناعي الاعتراض على القرارات المؤتمتة الهامة المتعلقة بهم، وأن يتمكنوا من اختيار عدم المشاركة عند الإمكان.

1.2.4.1 يجب أن توفر مؤسسات تشغيل أنظمة الذكاء التي تمد القرارات الهامة بالمعلومات إجراءات يتمكن من خلالها الأشخاص المتأثرون بالذكاء الاصطناعي من الاعتراض على قرار معين يخصهم

1.2.4.2 يجب أن تراعي مؤسسات تشغيل أنظمة الذكاء الاصطناعي هذه الإجراءات حتى مع القرارات غير الهامة

1.2.4.3 يجب أن تُحيط مؤسسات تشغيل أنظمة الذكاء الاصطناعي الأشخاص المتأثرين بالذكاء الاصطناعي بهذه الإجراءات، ويجب أن تصمم هذه الإجراءات بطريقة سلسة وسهلة الاستخدام من الجمهور المعني.

1.2.4.4 يجب أن تراعي مؤسسات تشغيل أنظمة الذكاء الاصطناعي توظيف مقيّمين بشريين لمراجعة أية تحديات تطرأ في اتخاذ القرارات، وأن تتيح رفض القرار الذي تم الاعتراض عليه عند اللزوم.

1.2.4.5 يجب أن تراعي مؤسسات تشغيل أنظمة الذكاء الاصطناعي استحداث آلية تتيح للجمهور الانسحاب وإلغاء مشاركتهم من القرارات المؤتمتة الهامة.

مثال:

يسمح بنك من البنوك لعملائه بالتقدم للحصول على قرض عبر الإنترنت عبر إدخال بياناتهم. ويستخدم البنك نظاماً للذكاء الإصطناعي ليحدد آليًا قرار منح القرض ونسبة الفائدة التي سيفرضها. ويعطي البنك مستخدمي الخدمة خيار الاعتراض على القرار وطلب مراجعته من قبل موظف بشري[7]. ويطلب البنك من العميل إيضاح أسباب اعتراضه على القرار عن طريق ملء نموذج مخصص لهذا الغرض، مما يساعد الشخص المسؤول عن تدقيق الحالة ويقلل فرصة اعتراض العملاء على أي قرار دون أسباب وجيهة.

1.2.4.6 يجب أن تراعي مؤسسات تشغيل أنظمة الذكاء الاصطناعي تعويض الأشخاص المتأثرين بقرارات الذكاء الاصطناعي في حالات الخسارة أو الإزعاج أو المضايقات التي تسببت بها القرارات المؤتمتة الخاطئة.

1.2.4.7 يجب أن تنظر مؤسسات تشغيل أنظمة الذكاء الاصطناعي في آليات «الاعتراض الجماعي» بحيث تتطلب حالات الشكاوى الجماعية إجراء تحقيق حول مدى عدالة و/أو دقة عملية اتخاذ القرارات ككل.

1.2.5 يجب ألّا تقوم أنظمة الذكاء الاصطناعي بإصدار أحكام هامة بالنيابة عن الأشخاص المعنيين دون الحصول على موافقتهم المسبقة

1.2.5.1 عند إعلام الأشخاص المتأثرين بالذكاء الاصطناعي عن الخيارات الهامة التي هم بصدد اتخاذها، يجب ألا تقيد أنظمة الذكاء الاصطناعي بشكل متعمّد الخيارات المتاحة أمام هؤلاء الأشخاص بشكل غير منطقي أو أن تحاول التأثير على قراراتهم الحاسمة دون الحصول على موافقة صريحة منهم.

1.2.6 يجب تطوير أنظمة الذكاء الاصطناعي المرتبطة بالقرارات الهامة بواسطة فرق متعددة الخلفيات والخبرات تتمتع بالمعرفة والخبرات المناسبة

1.2.6.1 يجب على مؤسسات تطوير الذكاء الاصطناعي التي تطّور أنظمةً تتخذ قرارات حرجة أن توفّر ضمن آلية اتخاذ القرار خبراء في مجالات مثل العلوم الاجتماعية، والسياسات أو أي مجالات أخرى تمكن هذه المؤسسات من تقييم الأثر المجتمعي لعملهم وأنظمتهم.

1.2.6.2 يجب أن يتضمن تطوير أنظمة الذكاء الاصطناعي المستخدمة في القرارات الهامة بالمعلومات استشارة خبراء في مجال عمل النظام الذي سيتم تطبيقه.

[7] نقلًا عن مفوضية الاتحاد الأوروبي، هل يمكن أن أخضع لقرارات فردية مؤتمتة، بما في ذلك تحليل ملفي الشخصي؟ متوفر على:
https://ec.europa.eu/info/law/law-topic/data-protection/reform/rights-citizens/my-rights/can-i-be-subject-automated-individual-decision-making-including-profiling_en#example-

مثال:

يوجد تطبيق يستخدم الذكاء الاصطناعي لتقييم الأعراض المرضية لقاعدة عريضة من المستخدمين، لكن يجب إخضاعه لفحص تنظيمي محكم بسبب ورود العديد من الشكاوى من الأطباء. فقد حذر الأطباء من أن التطبيق قد تفوته أعراض أمراض خطيرة ولايتعرف عليها. وقد ساهم الأطباء في تحديد عدد من نقاط القصور التي تمكنت الشركة من معالجتها[8].

1.2.7 يجب أن تكون لمؤسسات تشغيل أنظمة الذكاء الاصطناعي دراية كافية بطبيعة أنظمة الذكاء الاصطناعي التي تستخدمها حتى تكون قادرة على معرفة ملاءمتها لحالة الاستخدام وذلك تحقيقاً لضمان المساءلة والشفافية

1.2.7.1 في حالة القرارات الحرجة، يجب أن تتجنب مؤسسات تشغيل أنظمة الذكاء الاصطناعي استخدام أية أنظمة لا يمكن إخضاعها لمعايير المساءلة والشفافية

1.2.7.2 يجب أن تراعي مؤسسات تطوير الذكاء الاصطناعي إمكانية إبلاغ العملاء ومؤسسات تشغيل الذكاء الاصطناعي بحالات الاستخدام التي صُمم النظام من أجلها، وتلك الحالات التي لا يناسبها استخدامه فيها

1.3 سنحقق الشفافية في أنظمة الذكاء الاصطناعي

1.3.1 يجب أن تضمن المؤسسات المشغلة لأنظمة الذكاء الاصطناعي وأن تتيح إمكانية تتبع جذور أي قرار هام اتخذته الأنظمة بشكل آلي، وخاصة القرارات التي قد تؤدي إلى وقوع خسائر أو أذى أو ضرر

1.3.1.1 في حالة أنظمة الذكاء الاصطناعي المرتبطة بإصدار قرارات هامة، وخاصة تلك التي قد تسبب بخسائر أو أضرار أو إتلاف، يجب على مؤسسات تطوير الذكاء الاصطناعي أن يتضمن تصميمها للنظام قابلية التتبع، وهنا نعني القدرة على تتبع العوامل الأساسية المؤدية إلى قرارات معينة.

1.3.1.2 لتسهيل ما ذكر أعلاه، يجب أن تراعي مؤسسات تشغيل وتطوير أنظمة الذكاء الاصطناعي توثيق المعلومات التالية أثناء مراحل التصميم والتطوير والتطبيق، والاحتفاظ بهذه التوثيقات لفترة زمنية ملائمة لنوع القرار أو الصناعة :

- مصدر البيانات المستخدمة في التدريب وأساليب جمعها ومعالجتها، وكيفية نقل البيانات، والتدابير المتخذة للمحافظة على دقة البيانات مع مرور الوقت
- وتصميم وخوارزميات النموذج المستخدم
- والتغييرات على قاعدة برمجة البيانات وأصحاب هذه التغييرات

[8] فايننشال تايمز. 2018. تطبيق طبي ذائع الانتشار يخضع للفحص بعد استلام شكاوى من الأطباء. متوفر على: https://www.ft.com/content/96dd-fa565ec55929-11e8-8529-19dc6b7e

1.3.1.3 بالاعتماد على طبيعة نموذج تصميم نظام الذكاء الاصطناعي يجب على مؤسسات تطوير الذكاء الاصطناعي أن تُدرج في تصاميمها «رحلة اتخاذ القرارات» الخاصة بنتائج محددة (أي سلسلة القرارات التي تؤدي إلى هذه النتيجة)

مثال:

لدى شركة تكنولوجيا منتج تم تصميمه للمساعدة في التشخيص الطبي. يقوم المنتج بتوثيق كل مرحلة من مراحل استدلاله ويقوم بربطها مع البيانات المُدخلة[9].

1.3.2 يجب إعلام الناس بمستوى تفاعلهم مع أنظمة الذكاء الاصطناعي

1.3.2.1. يجب أن تُخبر مؤسسات تشغيل الذكاء الاصطناعي الأشخاص المتأثرين بالذكاء الاصطناعي عندما يقوم نظام الذكاء الاصطناعي باتخاذ قرارات هامة تؤثر عليهم

مثال:

تقوم محكمة دعاوى صغيرة بالفصل في الشؤون المدنية البسيطة مثل تحصيل الديون والإخلاءات والحيازة. وقد استخدمت المحكمة نظام ذكاء اصطناعي لاقتراح نتيجة الأحكام. وعند إصدار الحكم، يتم إشعار المدعي والمدعى عليه بأن الحكم قد صدر بمساعدة نظام الذكاء الاصطناعي. وتقوم المحكمة بدورها بتقديم شرح للحكم.

1.3.2.2 إذا استطاع نظام الذكاء الاصطناعي أن يبدو بصورة بشرية، يتوجّب على النظام إخبار الجمهور أنه نظام ذكاء اصطناعي.

مثال:

تقوم شركة تكنولوجيا بإنتاج نظام ذكاء اصطناعي قادر على إجراء بعض المكالمات الهاتفية نيابة عن مستخدميه لدرجة أن مستقبلي المكالمات الهاتفية قد يعتقدون أنهم يتحدثون إلى إنسان. ولهذا، تقوم الشركة ببرمجة الوكيل ليعرّف عن نفسه في بداية كل محادثة.

[9] انظر IBM WatsonPaths

1.4 سنجعل أنظمة الذكاء الاصطناعي قابلة للشرح تقنيا قدر الإمكان

1.4.1 مكن أن تتيح مؤسسات تشغيل الذكاء الاصطناعي إطلاع الأشخاص المتأثرين بالذكاء الاصطناعي على تفسير عام يشرح كيف تعمل أنظمة الذكاء الاصطناعي الخاصة بهم

1.4.1.1. يمكن أن تراعي مؤسسات تشغيل الذكاء الاصطناعي إطلاع الأشخاص المتأثرين بالذكاء الاصطناعي بلغة مفهومة وغير تقنية على معلومات حول ما يلي:

- البيانات التي يتلقاها النظام
- وأنواع الخوارزميات المستخدمة
- والفئات التي يمكن تصنيف الناس إليها
- وأهم العوامل التي تؤثر في نتائج القرارات

مثال:

قدم شخص طلبًا للحصول على بطاقة ائتمانية ولكن رُفض طلبه، وتم إعلامه أن الخوارزميات قد أخذت تاريخه الائتماني وعمره ورمزه البريدي بالاعتبار ولكن لم يتم إعلامه عن سبب رفض طلبه[10].

1.4.1.2 في حالات استخدام القطاع العام غير الحساسة والمُصممة للمنفعة العامة، يجب على مؤسسات تشغيل أنظمة الذكاء الاصطناعي مراعاة أن يكون الأصل البرمجي للنظام مع شرح لآلية عمل نظام الذكاء الاصطناعي متوفراً للعامة أو عند الطلب (وهذا شريطة أن لا تؤدي اتاحة هذه المعلومات لسوء استخدام الناس للنظام أو التلاعب فيه)

1.4.2 يجب أن تتيح مؤسسات تشغيل الذكاء الاصطناعي للأشخاص المتأثرين بالذكاء الاصطناعي وسائل لطلب تفسيرات لقرارات هامة تمسهم قدر الإمكان، مع مراعاة حالة البحوث الحالية ونموذج العمل

1.4.2.1. يجب أن تراعي مؤسسات تشغيل الذكاء الاصطناعي توفير سبل تتيح للأشخاص المتأثرين بقرار هام استمد معلوماته من الذكاء الاصطناعي إمكانية الوصول إلى مسوغات القرار

[10] ذا جارديان يقول الخبراء إنه يلزم توفير مراقبين لتنظيم عملية اتخاذ القرارات المؤتمتة. متوفر على:
ai-artificial-intelligence-watchdog-needed-to-prevent-discriminatory-automated-decisions/27/jan/2017/https://www.theguardian.com/technology

مثال:

يطلب مكتب الحماية المالية للمستهلكين في الولايات المتحدة من المؤسسات الائتمانية التي ترفض طلبات الحصول على الائتمان أن تشرح إلى مقدم الطلب السبب/أو الأسباب الأساسية وراء رفض الطلب (مثل «طول مدة الإقامة» أو «عمر المركبة[11]») . وبشكل خاص، "يمكن التصريح بأن سبب الرفض كان مبنيًا على المعايير أو السياسات الداخلية السارية في المؤسسة الائتمانية، أو أن مقدم الطلب أو الطلبات المشتركة أو الطرف المماثل لم يجمع نقاطًا تؤهله للحصول على الائتمان حسب نظام النقاط المعتمد في المؤسسة".

1.4.2.2 إذا لم تتمكن المؤسسة من تقديم هذه التفسيرات للمتعامل بسبب طبيعة التكنولوجيا المستخدمة، يجب على مؤسسات تشغيل الذكاء الاصطناعي تقديم تبسيط وبعض الخيارات البديلة مثل العوامل والمعطيات الأهم التي أدت لقرارها .

مثال:

طوّرت هيئة الخدمات الصحية الوطنية في المملكة المتحدة أداة اسمها «تنبّأ Predict» تسمح للنساء اللواتي يعانين من سرطان الثدي من مقارنة حالتهن مع حالات نساء أخريات مررن بنفس الحالة في الماضي، ومعرفة نسبة التعافي وفرص الحياة التي تتيحها الخيارات العلاجية المختلفة. ويتوفر على الموقع الإلكتروني صفحة توضيحية تبين الاحتمالات المرجحة للعوامل المختلفة وتضم وصفًا للعمليات الحسابية التي تم الاعتماد عليها[12].

1.4.3 في حالة توفر هذه التفسيرات، يجب إتاحة الوصول السهل والسريع والمجاني إليها، بطريقة مبسطة وسهلة للمستخدمين.

[11] مكتب الحماية المالية للمستهلكين، 12 سي إف آر، الجزء 1002، قانون الفرص الائتمانية المتكافئة (التشريع ب)، إشعارات § 1002.9، متوفر على: https://www.consumerfinance.gov/policy-compliance/rulemaking/regulations/1002/

[12] يمكن زيارة موقع «Predict» على http://www.predict.nhs.uk/predict_v2.1/algorithm/legal

مبادئ دبي للذكاء الاصطناعي

سجل التغييرات

الإصدار	المرحلة	تاريخ الإتمام	ملخص التغييرات
1.0	استشارات	2018.9.5	مسودة أولية
1.1	استشارات	2018.9.6	التعديلات قبل التعميم
1.2	استشارات	2018.9.10	تمت إضافة تعريفات، وإعادة صياغة الإرشادات بناءً عليها
1.3	استشارات	2018.9.10	أمثلة مضافة
1.4	استشارات	2018.9.30	إعادة تنسيق الأمثلة
1.5	استشارات	2018.10.1	إضافة تعريف للذكاء الاصطناعي
1.6	استشارات	2018.10.3	إعادة تنسيق وتبديل بعض الأمثلة
1.7	آراء وملاحظات	2018.10.9	دمج الجولة الأولى من الآراء والملاحظات
1.8	آراء وملاحظات	2018.10.9	أمثلة مُدققة؛ وتدقيقات أخرى متنوعة
1.9	آراء وملاحظات	2018.10.10	تدقيقات متنوعة
1.10	مراجعة داخلية	2018.10.10	مراجع إضافية
1.11	مراجعة داخلية	2018.10.30	تعديلات ومراجعة للصياغة وإضافة المقدمة

هذه الوثيقة من إعداد دبي الذكية بالتعاون مع فالكون وشركائها.

المراجع

1. هيئة حماية البيانات الشخصية (PDPC) في سنغافورة (5 يونيو 2018). ورقة نقاشية حول الذكاء الاصطناعي والبيانات الشخصية. سنغافورة: هيئة حماية البيانات الشخصية (PDPC) في سنغافورة. متوفرة على:
https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Dis
pdf.050618---cussion-Paper-on-AI-and-PD

2 . مجلس قطاع تكنولوجيا المعلومات (ITI). مبادئ سياسة الذكاء الاصطناعي. متوفرة على:
https://www.itic.org/public-policy/ITIAIPolicyPrinciplesFINAL.pdf

3. مكتب رئاسة مجلس الوزراء (2016/5/19) إطار أخلاقيات علوم البيانات. متوفر على:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_d
pdf._1__for_publication_Data_science_ethics_framework_v1.0/524298/ata/file

4.البرلمان الأوروبي. (2017/2/16) قرار البرلمان الأوروبي بتاريخ 2017/2/16 مع توصيات إلى هيئة تشريعات القانون المدني بخصوص الروبوتات (2103/2015(INL)). متوفر على:
0051-2017-TA-http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8
EN//DOC+XML+V0+0+

5. سي. فيلاني، (2018). نحو ذكاء اصطناعي منطقي من أجل استراتيجية فرنسية وأوروبية. متوفر على:
https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

6. اللجنة الوطنية للمعلوماتية والحريات (CNIL). (2017). كيف يمكن أن يسيطر البشر؟ المسائل الأخلاقية التي أبرزتها الخوارزميات والذكاء الاصطناعي. متوفر على:
https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf

7. المكتب التنفيذي لرئيس الولايات المتحدة الأمريكية. المجلس الوطني للعلوم والتكنولوجيا. اللجنة الفرعية لبحوث وتطوير الشبكات وتكنولوجيا المعلومات. (2016). الخطة الاستراتيجية الوطنية للبحث والتطوير في مجال الذكاء الاصطناعي. متوفرة على: https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf

8. المكتب التنفيذي لرئيس الولايات المتحدة الأمريكية، المجلس الوطني للعلوم والتكنولوجيا. لجنة التكنولوجيا. (2016). الاستعداد لمستقبل الذكاء الاصطناعي. متوفر على:
https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/
NSTC/preparing_for_the_future_of_ai.pdf

9. مجلس مدينة نيويورك. (2018). صياغة قانون محلي معني بأنظمة القرارات المؤتمتة التي تستخدمها الوكالات الحكومية. متوفرة على:
47-62E1-437A6A6D=GUID&3137815=http://legistar.council.nyc.gov/LegislationDetail.aspx?ID
461253F9C6D0-9C42-E2

10. المقر الرئيسي لإحياء الاقتصاد الياباني. (2015). استراتيجية الروبوتات الجديدة. استراتيجية الروبوتات في اليابان. الرؤية والاستراتيجية وخطة العمل. متوفرة على:
01b.pdf_0123/pdf/2015/http://www.meti.go.jp/english/press

11. مجلس أمانة الخزينة الكندي. (2018). الذكاء الاصطناعي المسؤول في حكومة كندا. سلسلة تقارير الثورة الرقمية. الإصدار 2.0
1Sn-qBZUXEUG4dVk909eSg5qvfbpNlRhzIefWPtBwb/https://docs.google.com/document/d
xY/edit

12. إعلان تورنتو: حماية الحق في العدالة وعدم التمييز في أنظمة تعلم الآلة. تورنتو، كندا: العفو الدولية وأكساس ناو. متوفرة على:
08_The-Toronto-Declaration_ENG/08/2018/https://www.accessnow.org/cms/assets/uploads
pdf.2018-

http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/arti.13
html (house of lords select.73546/cial-intelligence-committee/artificial-intelligence/oral
(committee oral evidence, Q61

14. لجنة الذكاء الاصطناعي في مجلس اللوردات. (2018). الذكاء الاصطناعي في المملكة المتحدة: هل جاهزون ومستعدون وقادرون؟

15. لجنة العلوم والتكنولوجيا في مجلس العموم. الخوارزميات في عالم اتخاذ القرارات.

دبـــي الذكيـة
SMART DUBAI