الإمارات العربية المتحدة
مكتب وزير دولة للذكاء الاصطناعي
والاقتصاد الرقمي وتطبيقات العمل عن بعد

أخلاقيات الذكاء الاصطناعي

الإرشادات والمبادئ التوجيهية

ديسمبر ٢٠٢٣

جدول المحتويات

	الصفحة
1- نظرة عامة	**4**
1-1 مقدمة	5
1-2 النطاق	6
1-3 التوافق الاستراتيجي	6
1-4 المسؤولية	8
1-5 الترخيص	8
1-6 آليات المشاركة	8
2- التعاريف	**10**
3-I لمبادئ والإرشادات التوجيهية	**16**
1-3 العدالة والإنصاف	**17**
1-1-3 المبدأ	17
جعل أنظمة الذكاء الاصطناعي عادلة ومنصفة	
2-1-3 الإرشادات	17
2-3 ذكاء اصطناعي قابل للمساءلة	**20**
1-2-3 المبدأ	20
جعل أنظمة الذكاء الاصطناعي قابلة للمساءلة	
2-2-3 الإرشادات	20
3-3شفافية الذكاء الاصطناعي	**26**
1-3-3 المبدأ	26
جعل أنظمة الذكاء الاصطناعي تتسم بالشفافية	
2-3-3 الإرشادات	26

	الصفحة
4-3 قابلية تفسير الذكاء الاصطناعي	**29**
1-4-3 المبدأ	29
جعل أنظمة الذكاء الاصطناعي قابلة للتفسير بقدر الإمكان من الناحية التقنية	
2-4-3 الإرشادات	29
5-3 إحكام وأمان الذكاء الاصطناعي	**31**
1-5-3 المبدأ	31
جعل أنظمة الذكاء الاصطناعي قابلة للمساءلة	
2-5-3 الإرشادات	32
6-3 الذكاء الاصطناعي المتمحور حول الإنسان	**34**
1-6-3 المبدأ	34
منح أنظمة الذكاء الاصطناعي قيماً بشرية وجعلها مفيدة للمجتمع	
2-6-3 الإرشادات	35
7-3 الذكاء الاصطناعي المستدام والصديق للبيئة	**38**
1-7-3 المبدأ	38
تعزيز الذكاء الاصطناعي المستدام بطريقة صديقة للبيئة	
2-7-3 الإرشادات	38
8-3 الذكاء الاصطناعي المحافظ على الخصوصية	**40**
1-8-3 المبدأ	40
احترام خصوصية الناس	
2-8-3 الإرشادات	40
44- مصادر الكتب	**44**

01
نظرة عامة

1-1 مقدمة

يُعد التقدم السريع في مجال الذكاء الاصطناعي وإمكانات الابتكار في مجموعة من المجالات المختلفة أمرًا مثيرًا. ومع ذلك، باتت هناك حاجة إلى إجراء مناقشة شاملة ومفتوحة حول أخلاقيات الذكاء الاصطناعي، والمبادئ التي يجب على المنظمات التي تستخدم هذه التكنولوجيا أخذها بعين الاعتبار.

تهدف هذه الوثيقة - أخلاقيات الذكاء الاصطناعي: الإرشادات والمبادئ التوجيهية - إلى تلبية هذه الحاجة من أجل تحقيق الموازنة بين هذين الاعتبارين المحوريين للذكاء الاصطناعي. ولقد تمت صياغتها بغية تقديم إرشادات مفصّلة لمساعدة الجهات الفاعلة في مجال الذكاء الاصطناعي على الالتزام بثمانية مبادئ من أخلاقيات الذكاء الاصطناعي.

نرى أن المبادئ الأساسية لأنظمة الذكاء الاصطناعي تتمثل فيما يلي:

العدالة والإنصاف	**القابلية للمساءلة**	**الشفافية**	**القابلية للتفسير**
الإحكام والأمان	**التمحور حول الإنسان**	**الإستدامة وصديق للبيئة.**	**المحافظة على الخصوصية.**

إن هذه المبادئ التوجيهية غير ملزمة، ولقد تمت صياغتها كتضافر للجهود التعاونية لأصحاب المصلحة المتعددين، مع وعيهم الكامل باحتياجات المنظمات للابتكار وحماية ملكيتها الفكرية. إن هذه عملية تعاونية يتم فيها دعوة جميع أصحاب المصلحة ليكونوا جزءًا من الحوار المفتوح والمستمر. نطمح إلى رؤية تطور للمبادئ التوجيهية لأخلاقيات الذكاء الاصطناعي ووصولها إلى مستوى عالمي وعملي قابل للتطبيق للتعريف بالمتطلبات والمبادئ الأخلاقية لتصميم الذكاء الاصطناعي واستخدامه. يتمثل الهدف النهائي الوصول إلى اتفاق واسع النطاق واعتماد سياسات متفق عليها بشكل عام للتعريف بالاستخدام الأخلاقي للذكاء الاصطناعي على المستوى الوطني والعالمي ككل.

2-1 النطاق

تقدم هذه الوثيقة إرشادات غير إلزامية لتحقيق التصميم الأخلاقي ونشر أنظمة الذكاء الاصطناعي في كل من القطاعين العام والخاص. إن الذكاء الاصطناعي يحيط بنا بالفعل، ولكن بعض التطبيقات قد يكون أكثر وضوحًا وحساسية من غيرها.

تنطبق هذه الوثيقة على أنظمة الذكاء الاصطناعي التي تتخذ «قرارات مهمة» أو تتنبئ بها - أي تلك القرارات التي يمكن أن يكون لها تأثيرًا كبيرًا إما على الأفراد أو على المجتمع ككل. كما أنها تنطبق على «القرارات الحاسمة»، وهي مجموعة فرعية من القرارات المهمة ذات الطبيعة الحاسمة والحيوية بشكل خاص.

ترشدنا هذه الوثيقة إلى التطورات المستجدة على المتطلبات والمبادئ الخاصة بهذا القطاع، ليتمكن لكل كيان داخل قطاعه من تعديل أطره الحالية، أو استخدامها، أو استكمالها بما يتلائم بشكل مناسب مع ظروفه واحتياجات أصحاب المصلحة المعنيين.

تعتبر هذه الوثيقة وثيقة حية، وستخضع للمزيد من المراجعات والتحسينات المستقبلية.

3-1 التوافق الاستراتيجي

تأتي هذه الوثيقة تحقيقاً للأهداف والتطلعات المحددة في الاستراتيجية الوطنية للذكاء الاصطناعي لدولة الإمارات العربية المتحدة وأهداف التنمية المستدامة الدولية (SDGs). لطالما كانت دولة الإمارات العربية المتحدة في طليعة التنمية المستدامة. لقد خلقت هذه الغايات العالمية المشتركة حركة عالمية وتبنيًا لها في العديد من التخصصات والقطاعات. وبصفتها متبني مبكر، أنشأت دولة الإمارات العربية المتحدة لجنة وطنية معنية بأهداف التنمية المستدامة من أجل تحقيق أجندة أهداف التنمية المستدامة لدولة الإمارات العربية المتحدة لعام 2030 على المدى الطويل. ونحن نرى هذه الأهداف مترسخة بشكل أكبر وتنعكس في مبادئ أخلاقيات الذكاء الاصطناعي تحت مواضيع العدالة والإنصاف، والشمولية، والمساواة، ومنفعة الإنسان، والحقوق.

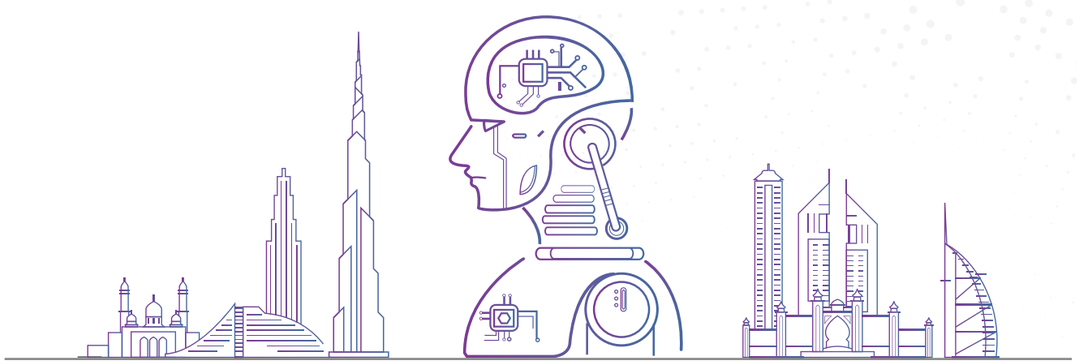
بالإضافة إلى ذلك، يعمل مكتب وزير الدولة للذكاء الاصطناعي في مكتب رئاسة مجلس الوزراء حاليًا على تطوير سياسات إضافية لأخذها بعين الاعتبار جنبًا إلى جنب مع هذه الوثيقة:

علامة الذكاء الاصطناعي

سيُستخدم ختم علامة الإمارات للذكاء الاصطناعي (UAI) التجارية لاستقطاب المواهب والشركات من جميع أنحاء العالم للقدوم إلى الإمارات العربية المتحدة لاختبار الذكاء الاصطناعي وتطويره. ويشمل ذلك علامة الإمارات للذكاء الاصطناعي (UAI) التي تعترف بشركات الذكاء الاصطناعي عالية الجودة والأخلاقية وتولي لها التقدير. ستكافئ تقنيات الذكاء الاصطناعي الآمنة والفعالة التي تم التحقق منها عن طريق منحها ختم الذكاء الاصطناعي الإماراتي (UAI) المعتمد. إن ختم علامة الإمارات للذكاء الاصطناعي (UAI) التجارية لا يزال قيد التطوير حاليًا على يد مكتب وزير الدولة للذكاء الاصطناعي بالتعاون مع القطاع الخاص.

إرشادات شراء أنظمة الذكاء

ستُستخدم إرشادات شراء أنظمة الذكاء الاصطناعي لتوجيه الكيانات الحكومية الاتحادية حول كيفية شراء أنظمة الذكاء الاصطناعي. ستُسرد المبادئ العامة لشراء منتجات الذكاء الاصطناعي ومراجعتها بين الكيانات الحكومية الاتحادية والموردين. ستأخذ هذه السياسة بعين الاعتبار البحث والتطوير بوصفهما تحديات محددة لا يوجد لها في الوقت الحالي منتجات حلول ذكاء اصطناعي جاهزة بالسوق. استنادًا إلى المعايير العالمية المتبعة لشراء أنظمة الذكاء الاصطناعي، تهدف هذه السياسة إلى تحفيز تبني الذكاء الاصطناعي داخل الكيانات الحكومية وتعزيز البحث والتطوير في مجال الذكاء الاصطناعي داخل القطاعات الخاصة لمواجهة تحديات حكومية محددة. تخضع سياسة الشراء في الوقت الحالي للتطوير على يد مكتب وزير الدولة للذكاء الاصطناعي بالتعاون مع وزارة المالية.



4-1 المسؤولية

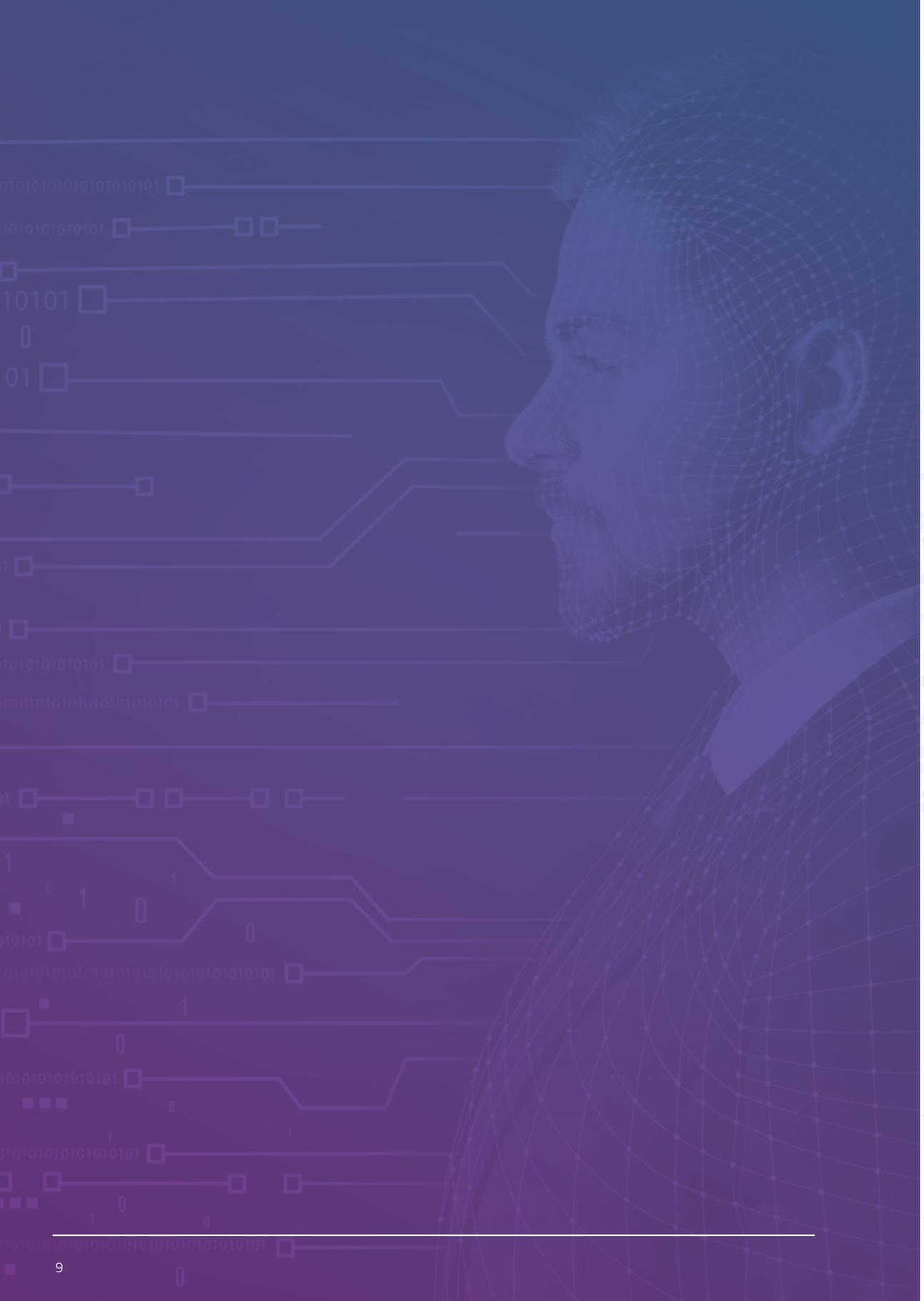
لن يكون مكتب وزير الدولة للذكاء الاصطناعي مسؤولاً عن أي إساءة استخدام لمبادئ وإرشادات أخلاقيات الذكاء الاصطناعي، ويتحمل المستخدم وحده جميع عواقب استخدامها.

5-1 الترخيص

نُشرت هذه الوثيقة بموجب شروط ترخيص مؤسسة المشاع الإبداعي 4.0 - رخصة عمومية دولية من أجل التسهيل على الحكومات الأخرى ومؤسسات القطاع الخاص لإعادة استخدامها. وبكل مختصر، فإن هذا يعني أن لك الحرية في مشاركة المواد وتكييفها، بما في ذلك للأغراض التجارية، شريطة أن تولي بالفضل لمجلس الإمارات للذكاء الاصطناعي والبلوك تشين بصفته المالك، وألا تقترح أن يصادق مجلس الإمارات للذكاء الاصطناعي والبلوك تشين على استخدامك.

6-1 آليات المشاركة

هذه الوثيقة هي جزء من آليات مشاركة ذات طبيعة مستقلة تعزز الوعي وتمكّن المؤسسات الحكومية والخاصة من متابعة حالات الاستخدام المبتكرة مع الحفاظ على القيم والمبادئ الإنسانية.



02 التعاريف

1-2 المؤسسة المطورة للذكاء الاصطناعي

هي أي مؤسسة تضطلع بأي مما يلي:

أ. **تحديد الغرض من نظام الذكاء الاصطناعي.**

ب. **أو تصميم نظام ذكاء اصطناعي.**

ت. **أو تطوير نظام ذكاء اصطناعي.**

ث. **أو إجراء الصيانة الفنية أو الضبط لنظام ذكاء اصطناعي.**

ملاحظة:

ينطبق هذا التعريف بغض النظر عما إذا كانت هذه المؤسسة هي المستخدم النهائي للنظام، أو ما إذا قامت ببيعه، أو تسليمه.

على سبيل المثال: تطور شركة ما نظام ذكي للتعرف على الوجه، وتبيعه لقوات الحدود في بلد ما، والذين يستخدمونه للتعرف على الموظفين المشبوهين، تكون الشركة في هذه الحالة هي المؤسسة المطورة للذكاء الاصطناعي، وتكون قوة الحدود هي المؤسسة المشغلة للذكاء الاصطناعي.

2-2 المؤسسة المشغلة للذكاء الاصطناعي

هي أي مؤسسة تضطلع بأي مما يلي:

أ. **استخدام أنظمة الذكاء الاصطناعي في العمليات الرسمية، أو غير الرسمية، أو اتخاذ القرارات.**

ب. **أو استخدام نظام الذكاء الاصطناعي لتقديم خدمة للشخص المعني بالذكاء الاصطناعي.**

ت. **أو صاحب عمل لنظام ذكاء اصطناعي.**

ث. **أو شراء البيانات ومعالجتها لاستخدامها في نظام الذكاء الاصطناعي.**

ج. **أو تقييم حالة الاستخدام لنظام ذكاء اصطناعي، وتقرير ما إذا كان يجب المضي قدمًا به.**

ملحوظات:

1- ينطبق هذا التعريف بغض النظر عما إذا كان نظام الذكاء الاصطناعي قد تم تطويره داخليًا أو تم شراؤه.

2- من الممكن أن تكون المؤسسات مؤسسة مطورة للذكاء الاصطناعي ومؤسسة مشغلة للذكاء الاصطناعي في آن واحد.

3-2 الذكاء الاصطناعي (AI)

يعني قدرة الوحدة الوظيفية على أداء الوظائف التي ترتبط عمومًا بالذكاء البشري؛ مثل التفكير المنطقي والتعلم وتحسين الذات.[1]

4-2 نظام ذكاء اصطناعي

هي منهجية منتج، أو خدمة، أو عملية، أو اتخاذ القرار تتأثر عمليتها أو مخرجاتها جوهريًا بوحدات وظيفية ذكية اصطناعيًا.

ملحوظات:

1- ليس من الضروري تحديد مخرجات النظام من خلال وحدات وظيفية ذكية اصطناعيًا فقط ليكون النظام عبارة عن نظام ذكاء اصطناعي.

2- من السمات الخاصة لأنظمة الذكاء الاصطناعي أنها تتعلم السلوك والقواعد غير المبرمجة فيها تعلمًا واضحًا.

مثال 1 : تستخدم محكمة المطالبات الصغيرة حزمة برامج ذكاء اصطناعي لجمع الأدلة المتعلقة بالقضية، ومقارنتها بقضايا مماثلة في الماضي، وتقديم بشأنها قرار للقاضي. يُقدم القاضي القرار النهائي. تتأثر منهجية صنع القرار هذه ماديًا بوحدة وظيفية ذكية اصطناعيًا، وبالتالي يتم تصنيفها كنظام ذكاء اصطناعي.

مثال 2 : يستخدم كيان حكومي نظام المحادثة الآلية «chatbot روبوت الدردشة «، والذي يسمح للعملاء بطرح الأسئلة الروتينية، وحجز المواعيد، وإجراء المعاملات المالية البسيطة. يستجيب نظام المحادثة الآلية « chatbot روبوت الدردشة « لاستفسارات العملاء عن طريق تزويدهم بردود مكتوبة مسبقًا بناءً على قواعد القرارات المبرمجة مسبقًا. لذلك، فإن نظام المحادثة الآلية « chatbot روبوت الدردشة « ليس نظام ذكاء اصطناعي. ومع ذلك، إذا قام نظام المحادثة الآلية « chatbot روبوت الدردشة « بتعديل تعامله مع العملاء بناءً على نتائج الحالات السابقة، فسيكون هذا نظام ذكاء اصطناعي.

[1] متوافق مع تعريف ISO/IEC 2382:2015

5-2 التحيز (للنظام)

يعني الميل، أو التحيز لصالح شخص، أو مجموعة واحدة، أو ضدها، لا سيما بطريقة غير عادلة.[2]

6-2 القرار الحاسم

هو قرار فردي مهم إما أن يكون له تأثيرًا كبيرًا جدًا على الفرد أو له مخاطر عالية بشكل خاص. يمكن أن تكون قرارات حساسة بشكل خاص، أو لديها القدرة على التسبب في حدوث خسارة أو ضرر كبير، أو أن تكون ذات أهمية اجتماعية، أو تشكل سابقة هامة.

ملاحظة:
إن أنواع القرارات المشار إليها هنا هي نفسها تلك الموجودة في تعريف القرارات المهمة على نطاق واسع، باستثناء أنه في هذه الحالة يُنظر إلى أثارها كنتيجة لقرار فردي بدلاً من مجموعة من العديد من القرارات.

مثال: تبت المحكمة فيما إذا كان المدعى عليه مذنبًا بتهمة جنائية، ومحكومًا عليه بعقوبة السجن مدى الحياة. هذا قرار حاسم لأن له تأثيرًا كبيرًا للغاية على حياة المدعى عليه، ويمثل أيضًا سابقة لقضايا مماثلة في المستقبل.

7-2 الوحدة الوظيفية

تعني وحدة متكاملة من الأجهزة أو البرامج، أو كليهما معًا، قادرة على تحقيق غرض محدد.[3]

8-2 قرار هام بشكل فردي

هو أي قرار يحتمل أن يكون له تأثيرًا كبيرًا على ظروف فرد واحد على الأقل، أو على سلوكه، أو اختياراته، أو له آثار قانونية عليه، أو ذو أهمية مماثلة له.

مثال: تقرر شركة ما أن موظفًا زائداً عن حاجتها، يعتبر هذا قراراً مهماً بشكل فردي بسبب تأثيره المحتمل على الوضع المالي لهذا الموظف.

9-2 التحيز غير التشغيلي (للنظام)

وهو التحيز الذي يكون إما:

أ. **ليس سمة من سمات التصميم.**

ب. **أو ليس مهمًا في تحقيق الغرض المعلن للنظام.**

10-2 مجموعة القرارات الهامة على نطاق واسع

تعني مجموعة من القرارات التي يتخذها نفس النظام أو المؤسسة والتي، عند اتخاذها بشكل إجمالي، يكون لها تأثيرًا كبيرًا على المجتمع ككل أو المجموعات المكونة له.

ملحوظات:

1- لا يلزم أن تكون القرارات مهمة بشكل فردي لكي تعتبر، بشكل إجمالي، مجموعة من القرارات المهمة على نطاق واسع.

2- من الأمثلة على المجالات التي لها تأثير كبير على المجتمع على سبيل المثال لا الحصر: التخصيص الواسع النطاق للموارد أو الفرص بين المجموعات؛ وهيكل الحكومة؛ وتقسيم السلطة بين الكيانات أو المجموعات الكبيرة؛ والقانون وتفسيره وتطبيقه؛ والصراعات والحروب؛ والعلاقات الدولية، إلخ.

مثال: نظام ذكاء اصطناعي يستخدمه موقع ويب لتحديد المحتوى الذي يجب إظهاره للمستخدمين. هذا القرار ليس مهمًا بشكل فردي، لأن المستخدم لا يتأثر تأثرًا كبيرًا بما إذا كان يعرض جزء معين من الوسائط عليهم. ومع ذلك، إذا كان موقع الويب هذا معروفًا، فقد يتخذ نظام الذكاء الاصطناعي مجموعة من القرارات المهمة على نطاق واسع، لأن أي تحيزات في النظام ستؤثر على عدد كبير من المستخدمين.

11-2 القرار الهام

هو أي قرار مهم بشكل فردي، أو هو جزء من مجموعة من القرارات المهمة على نطاق واسع.

12-2 الشخص المعني بنظام الذكاء الاصطناعي

هو شخص طبيعي يكون أيًا مما يلي:

أ. **مستخدم نهائي لنظام ذكاء اصطناعي**

ب. **أو متأثر تأثيرًا مباشرًا بتشغيل أو نتائج نظام ذكاء اصطناعي.**

ت. **أو مستفيد من خدمة أو توصية يقدمها نظام ذكاء اصطناعي**

03
المبادئ والإرشادات التوجيهية

3-1 العدالة والإنصاف (المبدأ1)

3-1-1 المبدأ

جعل أنظمة الذكاء الاصطناعي عادلة ومنصفة

لضمان استيفاء هذا المعيار، يجب:

- تكون البيانات المتداولة، حيثما أمكن ذلك، دقيقة وممثلة للأشخاص المعنيين.
- تتجنب الخوارزميات التحيز غير التشغيلي.
- اتخاذ خطوات للتخفيف والكشف عن التحيزات المتأصلة في مجموعات البيانات.
- تكون القرارات الهامة عادلة بشكل يمكن إثباته.
- يكون لجميع الموظفين المشاركين في تطوير ونشر واستخدام أنظمة الذكاء الاصطناعي دورًا ومسؤولية لتفعيل عدالة الذكاء الاصطناعي، ويجب تدريبهم وفقًا لذلك.

3-1-2 الإرشادات

3-1-2-1

يجب أن تكون فوائد أنظمة الذكاء الاصطناعي متاحة ومتوفرة للجميع. ينبغي ألا تميز أنظمة الذكاء الاصطناعي ضد الأشخاص أو الجماعات بطريقة قد يكون لها تأثيرًا سلبيًا. يجب أن تكون القرارات المهمة والحاسمة التي يتخذها الذكاء الاصطناعي أو يساعد فيها عادلة بشكل يمكن إثباته.

أ. بالنظر إلى أن تعريف «العدالة والإنصاف» قد ينطوي على العديد من التعريفات المختلفة، ينبغي للمؤسسة توثيق تعريفها الخاص للعدالة والإنصاف على السياق الذي ستنفذ فيه نظام الذكاء الاصطناعي. يجب على المؤسسات توثيق ما يمثله هدف الإنصاف المنفذ، ولماذا يعتبر هذا الاختيار هو الخيار الأنسب للسيناريو المحدد.

ب. يوصى بتحديد وتوثيق المجموعات الديمغرافية التي قد تتأثر سلبًا بذلك، مع التخفيف من المخاطر، حيثما أمكن ذلك.

ت. يمكن لمطوري الذكاء الاصطناعي ومشغلي الذكاء الاصطناعي النظر في تطبيق الإجراءات الرسمية مثل تقييمات أثر التمييز كوسيلة لضمان تحقيق العدالة والإنصاف. يجب توثيق هذا التقييم بالإضافة إلى غيره من تدابير التخفيف التي تنفذها المؤسسة. يجب إجراء تقييم الأثر قبل إصدار النظام وبعد ذلك بانتظام كتقييم مستمر، يُوصى كذلك بتوثيق النتائج.

2-2-1-3

ينبغي إيلاء الاعتبار فيما إذا كانت البيانات المتداولة دقيقة وتمثل الأشخاص المعنيين أم لا.

ث. للعدالة والإنصاف تعريفات عديدة ومختلفة في ثقافات مختلفة وسياقات مختلفة. وبالتالي، فإن تعزيز النظام البيئي المتنوع والشامل للذكاء الاصطناعي هو أمر بالغ الأهمية لضمان عدم تعارض أي تعريف للعدالة والإنصاف مع أي تعريف آخر، وكذلك لضمان أن تكون عملية تعريف العدالة والإنصاف نفسها عادلة، مع تمثيل المجموعات الممثلة تمثيلاً ناقصًا في المناقشة.

ج. على مطوري الذكاء الاصطناعي ومشغلي الذكاء الاصطناعي إجراء استكشاف و / أو اختبار معقول للبيانات لغرض تحديد اتجاهات صنع القرار المحتمل أن يكون لها أثارًا ضارة على أنظمة الذكاء الاصطناعي الناشئة عن عدم دقة البيانات والتحيزات في البيانات.

ح. يضطلع مطورو ومشغلو الذكاء الاصطناعي بتقييم جميع مجموعات البيانات لغرض تقييم شمولية المجموعات الديموغرافية المحددة وجمع البيانات لسد أي فجوات.

خ. يمتنع مطورو الذكاء الاصطناعي ومشغلو الذكاء الاصطناعي عن تدريب أنظمة الذكاء الاصطناعي على البيانات التي لا يحتمل أن تكون ممثلة للأشخاص المعنيون بالذكاء الاصطناعي، أو من غير المحتمل أن تكون دقيقة، سواء كان ذلك بسبب العمر، أو الإهمال، أو طريقة الجمع، أو أي عوامل أخرى.

مثال: بعد وقوع كارثة طبيعية، تستخدم وكالة الإغاثة الحكومية نظام الذكاء الاصطناعي للكشف عن المجتمعات التي تكون في حاجة ماسة للإغاثة من خلال تحليل بيانات الوسائط الاجتماعية من مجموعة من المواقع الإلكترونية. ومع ذلك، فإن المجتمعات التي يكون فيها استخدام الهواتف الذكية قليلاً تكون أقل وجودًا على وسائل التواصل الاجتماعي، وبذلك تكون عرضة لخطر عدم الحصول على الاهتمام المناسب. لذلك، تضيف وكالة الإغاثة الخيرية لأداة الذكاء الاصطناعي الخاصة بها التقنيات التقليدية لتحديد الأشخاص الأكثر حاجة في أماكن أخرى.

د. ينظر مطوري الذكاء الاصطناعي فيما إذا كان من المتوقع أن تعمل أنظمة الذكاء الاصطناعي الخاصة بهم بشكل جيد عند تعرضها لبيانات غير مرئية سابقًا، خاصة عند تقييم الأشخاص الذين لم يجر تمثيلهم تمثيلاً جيدًا في بيانات التدريب.

3-2-1-3

ينبغي النظر فيما إذا كانت عمليات صنع القرار تنطوي على تحيز أم لا

أ. عند إخضاع مجموعات مختلفة لعمليات صنع قرار مختلفة، يجب على مطوري الذكاء الاصطناعي النظر فيما إذا كان ذلك سيؤدي إلى تحيز غير تشغيلي أم لا.

ب. عند تقييم عدالة وانصاف نظام ذكاء اصطناعي، ينظر مطورو الذكاء الاصطناعي ومشغلو الذكاء الاصطناعي فيما إذا كان الأشخاص المعنيون بالذكاء الاصطناعي يتلقون معاملة متساوية وعادلة في نفس الظروف.

مثال: تستخدم مؤسسة ما أداة ذكاء اصطناعي لأتمتة الفحص المسبق للمرشحين لشغل وظيفة شاغرة، والتي يتم تدريبها على بيانات الموظفين الحاليين للشركة، والذين يأتي معظمهم من نفس الخلفية العرقية. لذلك، يتعلم النظام استخدام الاسم والجنسية كعوامل تمييزية في تصفية المتقدمين للوظيفة. كان يمكن تحديد ذلك من خلال إجراء الاختبارات وتصحيحها، على سبيل المثال، عن طريق موازنة بيانات التدريب أو استخدام حقول البيانات ذات الصلة للتدريب فقط.

4-2-1-3

ينبغي أن تكون القرارات المهمة التي تستند إلى استخدام الذكاء الاصطناعي عادلة

أ. ينظر مطورو الذكاء الاصطناعي ومشغلو الذكاء الاصطناعي في الإجراءات الرسمية مثل تقييمات أثر التمييز كوسيلة لضمان الإنصاف.

4-2-1-3

ينظر مشغلو الذكاء الاصطناعي في إمكانية الوصول إلى أنظمة الذكاء الاصطناعي الخاصة بهم واستخدامها بطريقة عادلة عبر مجموعات المستخدمين

2--1-3

5 ينبغي النظر كذلك في تأثير التنوع على عمليات التطوير والنشر

ب. تسعى المؤسسات إلى إشراك أشخاص من خلفيات ديموغرافية متنوعة عبر نمط الحياة الكامل لتضمين عمليات التصميم والتطوير والنشر. يجب أن تسعى المؤسسات إلى إشراك مجموعات داخلية وخارجية متنوعة أيضًا.

ت. ينظر مطورو الذكاء الاصطناعي فيما إذا كانت الافتراضات التي يضعونها حول الأشخاص الذين يخضعون للذكاء الاصطناعي يمكن أن تكون خاطئة أو من المحتمل أن تؤدي إلى تحيز غير تشغيلي؛ إذا كان الأمر كذلك، فيجب عليهم التفكير في استشارة الأشخاص الذين يخضعون للذكاء الاصطناعي بطريقة تمثيلية أثناء التصميم والتطوير والنشر لتأكيد مثل هذه الافتراضات.

2-3 ذكاء اصطناعي قابل للمساءلة (المبدأ 2)

1-2-3 المبدأ

جعل أنظمة الذكاء الاصطناعي قابلة للمساءلة

لضمان استيفاء ذلك، يجب أن:

- ألا تقع المساءلة عن نتائج نظام الذكاء الاصطناعي على عاتق النظام نفسه، بل توزع بين من يصممونه ويطورونه وينشرونه.
- يبذل المطورون جهودًا للتخفيف من حدة المخاطر الكامنة في الأنظمة التي يصممونها.
- تحتوي أنظمة الذكاء الاصطناعي على إجراءات استئناف مدمجة تمُكّن المستخدمين من الطعن على القرارات المهمة.
- تطوير أنظمة الذكاء الاصطناعي على أيدي فرق متنوعة تضم خبراء في المنطقة التي سيجري نشر النظام فيها.
- تخضع أنظمة الذكاء الاصطناعي للتدقيق الخارجي لضمان جودة القرار.

2-2-3 الإرشادات

1-2-2-3
عدم تحمل النظام نفسه للمساءلة عن نتائج نظام الذكاء الاصطناعي

أ. لا ينبغي أن تُعزى المسؤولية عن الخسائر أو الأضرار الناتجة عن تطبيق أنظمة الذكاء الاصطناعي إلى النظام نفسه.

ب. ينظر مشغلو الذكاء الاصطناعي ومطوري الذكاء الاصطناعي في تعيين موظفين ليكونوا مسؤولين عن التحقيق في أسباب الخسارة أو الضرر الناجم عن توزيع أنظمة الذكاء الاصطناعي وتصحيحها.

2-2-2-3
بذل جهود إيجابية لتحديد أي مخاطر كبيرة متأصلة في أنظمة الذكاء الاصطناعي المصممة والتخفيف من حدتها.

أ. يستخدم مشغلو الذكاء الاصطناعي فقط أنظمة الذكاء الاصطناعي المدعومة ببحوث أكاديمية وصناعية قائمة على الأدلة، وأن يبني مطوري الذكاء الاصطناعي تطويراتهم على مثل هذا البحث.

مثال: دولة أجنبية تقدم خدمة حكومية للتعرف على الآباء المدينون بدفع نفقة إعالة أطفالهم. غالبًا ما تكون عملية مطابقة البيانات غير صحيحة بسبب الأسماء التي يوجد بها أخطاء إملائية أو بيانات ناقصة، مما يؤدي إلى قيام النظام باستهداف بعض الأفراد تلقائيًا بشكل غير صحيح، مما يؤدي إلى تكاليف كبيرة وضعف التصنيف الائتماني وتجميد الأجور. إن الملاحقة القانونية للأفراد المستهدفين بشكل غير صحيح يستغرق وقتًا طويلاً وليس بسيطاً . إذا تم تقييم التأثير المحتمل للقرارات غير الصحيحة، كان من الممكن وضع تدابير التخفيف مثل إجراء المراجعة سهل الاستخدام.

ب. يُحدد مشغلو الذكاء الاصطناعي التأثير المحتمل للقرارات الآلية غير الصحيحة على الأشخاص الذين يخضعون للذكاء الاصطناعي، وفي حالة احتمال أن تتسبب القرارات غير الصحيحة في تكلفة كبيرة أو إزعاج، يجب النظر في اتخاذ تدابير التخفيف.

ت. ينظر مشغلو الذكاء الاصطناعي في تقييمات المخاطر الداخلية أو الأطر الأخلاقية كوسيلة لتسهيل تحديد المخاطر وتدابير التخفيف.

ث. عند تصميم أنظمة الذكاء الاصطناعي للاسترشاد بها في اتخاذ القرارات المهمة، ينظر مطوري الذكاء الاصطناعي في اتخاذ تدابير للحفاظ على دقة البيانات بمرور الوقت، بما في ذلك:

. اكتمال البيانات.

. تحديث البيانات في الوقت المناسب.

. ما إذا كان السياق الذي جمعت فيه البيانات يؤثر على مدى ملاءمتها لحالة الاستخدام المقصودة.

مثال: من الممكن لمسح كاميرا مراقبة الحدود الذي يبحث عن مؤشرات للخطر أن يسيء تفسير «التشنج اللاإرادي» لشخص مصاب بمتلازمة توريت على أنه شخص مشبوه. من الممكن أن يظهر هذا بأشكال متنوعة، ولا ينبغي أن يُخضع هذا الشخص لفحص ثانوي في كل مرة يمر فيها عبر الحدود . إذا تم تحديث البيانات بعد مروره في المرة الأولى، فسيؤدي ذلك إلى تجنب التسبب في أي إزعاج له في الزيارات اللاحقة.

[4] مكتب مجلس الوزراء (المملكة المتحدة)، إطار العمل الأخلاقي لعلوم البيانات، الإصدار رقم 1.0، مرخص بموجب ترخيص الحكومة المفتوحة v3.0، صفحة 13.
[5] حكومة كندا. (2018). الذكاء الاصطناعي المسؤول بحكومة كندا. سلسلة المستندات التعريفية للاضطرابات الرقمية. الإصدار 2.0، صفحة 26. مأخوذ من:
https://docs.google.com/document/d/1Sn-qBZUXEUG4dVk909eSg5qvfbpNlRhzlefWPtBwbxY/edit

ج. يضبط مطورو الذكاء الاصطناعي ومشغلو الذكاء الاصطناعي نماذج الذكاء الاصطناعي ضبطًا دوريًا لتلبية التغييرات التي تطرأ على البيانات و / أو النماذج بمرور الوقت.

ح. ينظر مشغلو الذكاء الاصطناعي فيما إذا كانت أنظمة الذكاء الاصطناعي المدربة في بيئة ثابتة ستظهر عدم استقرار للنموذج عند نشره في بيئات ديناميكية.

مثال: يجب أن تكون أنظمة الذكاء الاصطناعي قادرة على التكيف مع التغييرات في البيئة التي يتم توزيعها فيها. على سبيل المثال، يجب أن تكون السيارة ذاتية القيادة قادرة إلى التكيف بسرعة مع الطرق الخطرة وغير المتوقعة من خلال التعلم في الوقت الفعلي من السيارات الأخرى التي تعاملت بنجاح مع هذه الظروف. بالإضافة إلى ذلك، يجب أن تتعامل هذه التطبيقات ذات المهام الحرجة مع المدخلات العشوائية والدفاع ضد الجهات الخبيثة.

خ. يتعاون مطورو الذكاء الاصطناعي مع مشغلي الذكاء الاصطناعي لتدريب النماذج باستخدام البيانات التاريخية من المشغل.

د. ينظر مشغلو الذكاء الاصطناعي في العمل مع موردهم (مطوري الذكاء الاصطناعي) لمراقبة الأداء باستمرار.

ذ. يُخضِع مشغلو الذكاء الاصطناعي أنظمة الذكاء الاصطناعي التي ترشدنا إلى اتخاذ القرارات المهمة لفحوصات الجودة بنفس القدر من الصرامة الذي قد يكون مطلوبًا من الإنسان الذي يتخذ نفس القرار.

3-2-2-3
يجب إخضاع أنظمة الذكاء الاصطناعي التي ترشدنا في اتخاذ القرارات الحاسمة للتدقيق الخارجي المناسب

أ. عند استخدام أنظمة الذكاء الاصطناعي لاتخاذ قرارات حاسمة، ينبغي استخدام التدقيق الخارجي لأنظمة الذكاء الاصطناعي المعنية كوسيلة لضمان التقيد بمعايير هادفة للشفافية والمساءلة.

ب. في حالة ما إذا كانت القرارات الحاسمة ذات مصلحة مدنية، ينبغي النظر في تقديم إعلان عام عن نتائج التدقيق كوسيلة لضمان استمرار خضوع العمليات العامة للمساءلة أمام المعنيين بها.

ت. في حالة كان القرار الحاسم هو قرار يتعلق بالحياة والموت، ينبغي دعم هذه القرارات بمزيد من المصادقة والتحقق من خلال مشغل بشري.

ث. تسهيل تتبع أنظمة الذكاء الاصطناعي وقابليتها للتدقيق، لا سيما في السياقات أو المواقف الحرجة.

ج. الوضع في الاعتبار بأنه قد تكون هناك توترات أساسية بين المبادئ والمتطلبات المختلفة. يجب العمل باستمرار على تحديد وتقييم وتوثيق وإبلاغ هذه المفاضلات وحلولها.

ح. اعتماد قائمة تقييم موثوقة للذكاء الاصطناعي عند تطوير أنظمة الذكاء الاصطناعي أو نشرها أو استخدامها، وتكييفها مع حالة الاستخدام المحددة التي يطبق فيها النظام.

خ. الوضع في الاعتبار بأن قائمة التقييم هذه لن تكون شاملة على الإطلاق. لا يتعلق ضمان موثوقية الذكاء الاصطناعي بوضع علامة في مربعات الاختيار، بل يتعلق بتحديد المتطلبات وتنفيذها بشكل مستمر، وتقييم الحلول، وضمان تحسين النتائج طوال دورة حياة نظام الذكاء الاصطناعي، وإشراك أصحاب المصلحة في ذلك.

4-2-2-3
يجب أن يكون الأشخاص المعنيون بالذكاء الاصطناعي قادرين على تحدي القرارات المؤتمتة الهامة المتعلقة بهم، وعند الاقتضاء، قادرين على الانسحاب من مثل هذه القرارات

أ. يُحدد مشغلو الذكاء الاصطناعي الذين يستخدمون أنظمة الذكاء الاصطناعي لإرشادهم إلى اتخاذ القرارات المهمة إجراءات يمكن من خلالها للأشخاص المعنيين بالذكاء الاصطناعي الطعن على قرار معين يتعلق بهم.

ب. يراعي مشغلو الذكاء الاصطناعي هذه الإجراءات حتى للقرارات غير المهمة.

ت. يُطلع مشغلو الذكاء الاصطناعي الأشخاص المعنيين بالذكاء الاصطناعي على هذه الإجراءات، ويصمموها بطريقة تكون مريحة وسهلة الاستخدام.

ث. ينظر مشغلو الذكاء الاصطناعي في تعيين مقيمين للحالات البشرية لمراجعة أي تحديات من هذا القبيل، وعند الاقتضاء، إلغاء القرار المطعون فيه.

ج. ينظر مشغلو الذكاء الاصطناعي في إنشاء آلية إلغاء للقرارات الآلية المهمة.

مثال: يسمح البنك للعملاء بتقديم طلب للحصول على قرض عبر الإنترنت عن طريق إدخال بياناتهم. يستخدم البنك نظام الذكاء الاصطناعي لتحديد ما إذا كان سيتم منح القرض لهم تلقائيًا أم لا، وكم يجب أن يكون سعر الفائدة. إنهم يزودون المستخدمين بخيار للاعتراض على القرار ومراجعته على يد بشري . كما يطلبون من العملاء تبرير هذا الطعن من خلال ملء نموذج، مما يساعد مراجع الحالة ويمنع العملاء من الطعن على القرار دون سبب وجيه.[7]

[6] ستوكيبا وآخرون 2017، وجهة نظر بيركلي لتحديات الأنظمة للذكاء الاصطناعي، صفحة. 2، https://arxiv.org/pdf/1712.05855.pdf

ج) يمكن لمشغلي الذكاء الاصطناعي النظر في آليات «الطعن الجماعي» حيث إن العدد الكبير من الشكاوى يؤدي إلى إجراء تحقيق في نزاهة و / أو دقة عملية اتخاذ القرار ككل.

5-2-2-3

يجب ألا تحاول أنظمة الذكاء الاصطناعي التي ترشدنا في اتخاذ القرارات الهامة إصدار أحكام قيمّة نيابة عن الأشخاص دون الحصول على موافقة مسبقة

أ) عند إبلاغ أحد الأشخاص المعنيين بالذكاء الاصطناعي بالخيارات الهامة التي سيقومون بها، يجب ألا تقيد أنظمة الذكاء الاصطناعي الخيارات المتاحة بشكل غير معقول أو محاولة التأثير على أحكامهم القيمّة دون الحصول على موافقة صريحة من الأشخاص المعنيين بالذكاء الاصطناعي.

6-2-2-3

يجب تطوير أنظمة الذكاء الاصطناعي التي ترشدنا في اتخاذ القرارات الهامة على أيدي فرق متنوعة ذات خلفيات مناسبة.

أ) يستعين مطورو الذكاء الاصطناعي الذين يطورون أنظمة الذكاء الاصطناعي المستخدمة للمساعدة في اتخاذ القرارات الحاسمة في عملياتهم بخبراء لديهم خلفية في العلوم الاجتماعية أو السياسة أو أي موضوع آخر يهيئهم لتقييم التأثير المجتمعي الأوسع لعملهم.

ب) تشمل عملية تطوير أنظمة الذكاء الاصطناعي التي ترشدنا إلى اتخاذ القرارات الهامة التشاور مع الخبراء في المجال الذي سينشر فيه النظام.

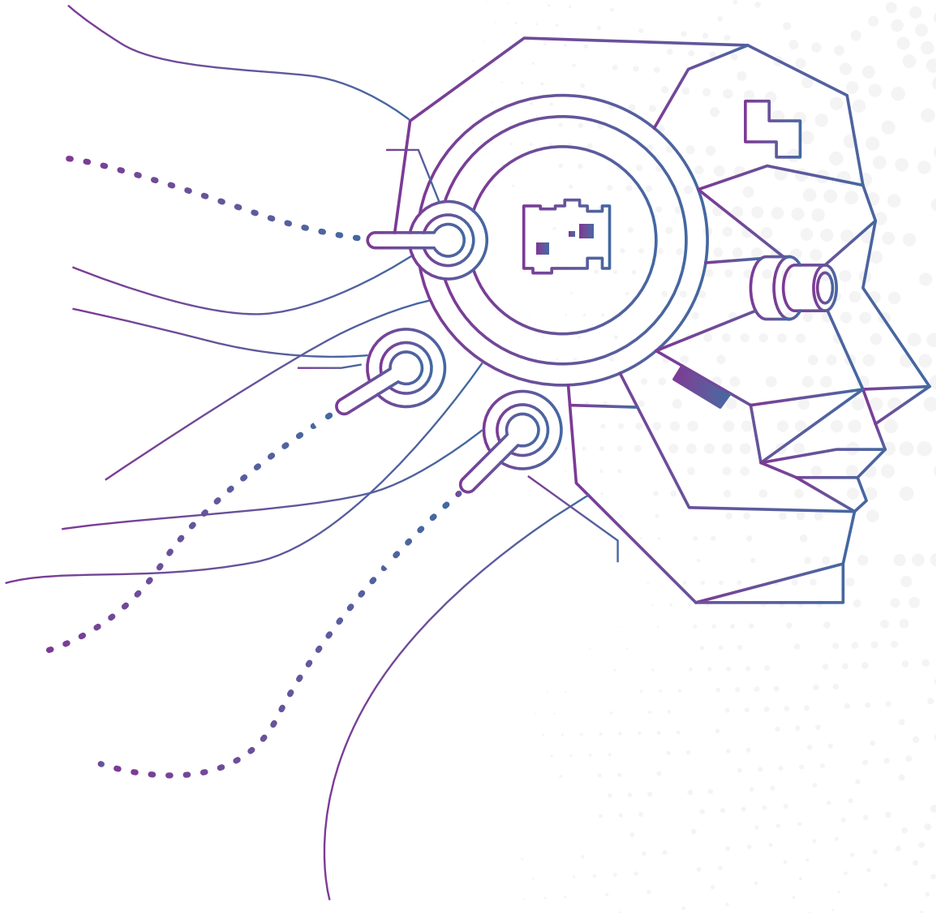
مثال: تطبيقًا يستخدم الذكاء الاصطناعي لتقييم الأعراض الطبية ولديه قاعدة مستخدمين كبيرة يجب أن يخضع لتدقيق تنظيمي بسبب عدد الشكاوى الواردة من الأطباء، والذين حذروا من أن التطبيق يمكن أن يفوت بعض العلامات الخطيرة للمرض. حدد الأطباء عددًا من أوجه القصور المختلفة بالتطبيق، والتي يمكن للشركة معالجة بعضها وحلها.[8]

7-2-2-3

يعي مشغلو الذكاء الاصطناعي بأنظمة الذكاء الاصطناعي التي يستخدمونها وعيًا كافيًا لتقييم مدى ملاءمتها لحالة الاستخدام ولضمان المساءلة والشفافية.

أ) في حالة القرارات الحاسمة، يتجنب مشغلو الذكاء الاصطناعي استخدام أنظمة الذكاء الاصطناعي التي لا يمكن أن تخضع لمعايير ذات مغزى من حيث المساءلة والشفافية.

ب) ينظر مطورو الذكاء الاصطناعي في إخطار العملاء ومشغلي الذكاء الاصطناعي بحالات الاستخدام التي صمم النظام من أجلها، وتلك التي لا يناسبها هذا النظام.



[7] معتمدة من مفوضية الاتحاد الأوروبي، هل يمكن أن أخضع لعملية صنع قرار فردي آلي، بما في ذلك التوصيف؟ مأخوذ من: 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

[8] فاينانشيال تايمز. 2018. تطبيق طبي رفيع المستوى يخضع للفحص بعد شكاوى الأطباء. مأخوذ من: https://www.ft.com/content/19dc6b7e-8529-11e8-96dd-fa565ec55929

3-3 شفافية الذكاء الاصطناعي(المبدأ 3)

3-3-1 المبدأ

جعل أنظمة الذكاء الاصطناعي تتسم بالشفافية

لضمان استيفاء هذا المعيار، يجب:

- **على المطورين بناء أنظمة يمكن تتبع حالات اخفاقها وتشخيصها.**
- **إبلاغ الناس عند اتخاذ نظام الذكاء الاصطناعي لقرارات مهمة بشأنهم.**
- **ضمن حدود الخصوصية والحفاظ على الملكية الفكرية، يجب أن يتحلّى أولئك الذين ينشرون ويوزعون أنظمة الذكاء الاصطناعي بالشفافية بشأن البيانات والخوارزميات التي يستخدمونها.**
- **تقديم الإفصاحات المسؤولة في الوقت المناسب، وتقديم مبررات معقولة لنتائج أنظمة الذكاء الاصطناعي. يتضمن ذلك المعلومات التي تساعد الناس على فهم النتائج، مثل العوامل الرئيسية المستخدمة في صنع القرار.**

3-3-2 الإرشادات

3-3-2-1

ينبغي النظر في إمكانية تتبع القرارات الهامة، لا سيما تلك التي يحتمل أن تؤدي إلى وقوع خسارة أو ضرر أو ضرر.

أ بالنسبة لأنظمة الذكاء الاصطناعي التي تتخذ قرارات هامة، لا سيما تلك التي يمكن أن تسبب حدوث خسارة أو ضررًا أو أذى، ينظر مطوري الذكاء الاصطناعي في إمكانية التتبع (أي القدرة على تتبع العوامل الرئيسية المؤدية إلى أي قرار محدد).

ب تضمن المؤسسات التحقيق في الأضرار التي تسببها أنظمة الذكاء الاصطناعي، وتضمن معالجتها كذلك، من خلال سن آليات إنفاذ قوية وإجراءات تصحيحية، للتأكد من احترام حقوق الإنسان وسيادة القانون في العالم الرقمي وفي العالم المادي.

ت لتسهيل النقطة الواردة أعلاه، ينظر مطورو الذكاء الاصطناعي ومشغلو الذكاء الاصطناعي في توثيق المعلومات التالية أثناء مراحل التصميم والتطوير والنشر والتوزيع، والاحتفاظ بهذه الوثائق لفترة زمنية مناسبة لنوع القرار أو مجال الصناعة:

- مصدر بيانات التدريب، وطرق جمعها، ومعالجتها، وكيفية نقل البيانات، والتدابير المتخذة للحفاظ على دقتها مع مرور الوقت.
- تصميم النموذج والخوارزميات المستخدمة.
- التغييرات الطارئة على قاعدة الشفرات، وصياغة تلك التغييرات.

ث حيثما أمكن بالنظر إلى تصميم النموذج، ينظر مطوري الذكاء الاصطناعي في إنشاء وسيلة يمكن من خلالها تسجيل «رحلة اتخاذ القرار» لنتيجة محددة (أي مكونات القرار المؤدية إليها).

مثال: تمتلك شركة تكنولوجيا منتجًا مصممًا للمساعدة في التشخيص الطبي، يوثق كل مرحلة من مراحل تحليلها، ويربطها مرة أخرى ببيانات الإدخال [9]

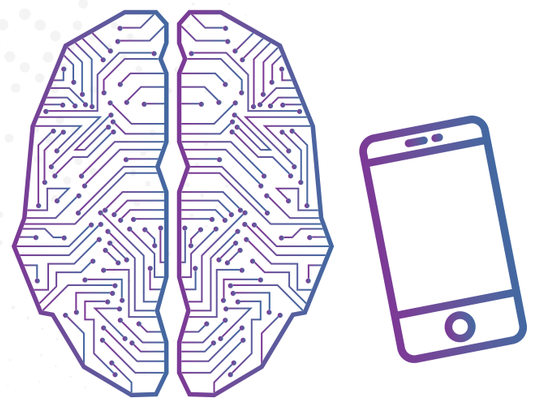
2-2-3-3
يجب إبلاغ الناس بمدى تفاعلهم مع أنظمة الذكاء الاصطناعي

أ يبلغ مشغلو الذكاء الاصطناعي الأشخاص المعنيين بالذكاء الاصطناعي عندما يتخذ نظام الذكاء الاصطناعي قرارًا مهمًا يؤثر عليهم.

مثال: تفصل محكمة الدعاوى الصغيرة في القضايا المدنية البسيطة مثل تحصيل الديون وعمليات الإخلاء. تُقدم المحكمة نظام ذكاء اصطناعي لاقتراح نتيجة الحكم. في وقت صدور الحكم، يجب إخطار المدعي والمدعى عليه أنه تم صدور القرار بمساعدة نظام الذكاء الاصطناعي. كما تقدم المحكمة تفسيرًا للقرار.

ب إذا كان نظام الذكاء الاصطناعي قادرًا على انتحال صفة إنسان بشكل مقنع، فلا ينبغي أن يفعل ذلك إلا بعد إخطار الشخص المعني بالذكاء الاصطناعي بأن هذا نظام ذكاء اصطناعي.

مثال: تنتج شركة تقنية وكيل ذكاء اصطناعي للمحادثة يمكنه إجراء بعض المكالمات الهاتفية نيابة عن مستخدميها. قد يعتقد أولئك الذين يتلقون المكالمات أنهم يتحدثون إلى إنسان. لذلك، تقوم الشركة ببرمجة هذا الوكيل للتعريف عن نفسه في بداية كل محادثة.



[10] الجارديان. يقول الخبراء إن رقابة نظام الذكاء الاصطناعي بحاجة إلى تنظيم عملية صنع القرار الآلي. مأخوذ من:
https://www.theguardian.com/technology/2017/jan/27/ai-artificial-intelligence-watchdog-needed-to-prevent-discriminatory-automated-decisions

3-4 قابلية تفسير الذكاء الاصطناعي (المبدأ 4)

1-4-3 المبدأ

جعل أنظمة الذكاء الاصطناعي قابلة للتفسير بقدر الإمكان من الناحية التقنية

لضمان استيفاء هذا المعيار، يجب أن:

- تكون قرارات ومنهجيات أنظمة الذكاء الاصطناعي التي لها تأثير كبير على الأفراد قابلة للتفسير، إلى الحد الذي تسمح به التكنولوجيا المتاحة.
- يكون من الممكن التأكد من العوامل الرئيسية التي تؤدي إلى اتخاذ أي قرار محدد يمكن أن يكون له تأثيرًا كبيرًا على أي فرد.
- في الحالة المذكورة أعلاه، سنوفر قنوات يمكن للناس من خلالها طلب مثل هذه التفسيرات.

2-4-3 الإرشادات

1-2-4-3
ينظر مشغلو الذكاء الاصطناعي في تزويد الأشخاص المعنيين بالذكاء الاصطناعي بشرح رفيع المستوى لكيفية عمل نظام الذكاء الاصطناعي الخاص بهم.

أ ينظر مشغلو الذكاء الاصطناعي في إبلاغ الأشخاص المعنيين بالذكاء الاصطناعي، بلغة مفهومة وغير تقنية، بما يلي:

. البيانات التي يستوعبها النظام.
. أنواع الخوارزميات المستخدمة.
. الفئات التي يمكن تصنيف الأشخاص وفقًا لها.
. أهم السمات التي تؤدي إلى نتائج القرارات.
. قائمة شاملة للميزات الهندسية والنماذج التي تم أخذها في الاعتبار أثناء مرحلة بناء النموذج.

مثال: من الممكن إخبار الشخص الذي رُفض منحه بطاقة ائتمان أن الخوارزمية أخذت سجله الائتماني وعمره ورمزه البريدي في الاعتبار، ولكن لا يعرف سبب رفض طلبه.[10]

ب بالنسبة لحالات الاستخدام غير الحساسة للقطاع العام المصممة للصالح العام، ينظر مشغلو الذكاء الاصطناعي في جعل شفرة المصدر، إلى جانب شرح طريقة عمل نظام الذكاء الاصطناعي، متاحة إما للجمهور أو عند الطلب (ينبغي القيام بذلك فقط إذا هناك مخاطر منخفضة من التلاعب بالنظام).

ت يحتفظ مشغل الذكاء الاصطناعي بالوثائق الضرورية التي تبين تفصيلاً وتوضيحًا لكيفية عمل الخوارزميات، على سبيل المثال: توثيق العمليات، ومخططات اتخاذ القرارات في النظام، وكيفية دمج عملية الطعن والاستئناف.

ث يحق للفرد الطعن والتماس سبيل الانتصاف الفعال على القرارات التي تتخذها أنظمة الذكاء الاصطناعي. يجب على المجموعة أو الفريق المعني بدعم هذه النماذج مناقشتها.

ج وضع مؤشرات مناسبة لتقييم التدخلات التكنولوجية لنظام الذكاء الاصطناعي من وجهات نظر متعددة.

2-2-4-3

ينظر مشغلو الذكاء الاصطناعي في تزويد الأشخاص المعنيين بالذكاء الاصطناعي بوسائل لطلب تفسيرات لقرارات هامة محددة، إلى أقصى حد ممكن بالنظر إلى حالة البحث الحالي واختيار النموذج.

أ يوفر مشغلو الذكاء الاصطناعي وسيلة يمكن من خلالها للأشخاص المعنيين بقرار هام، أرشدهم إليه الذكاء الاصطناعي، الوصول إلى الأسباب الكامنة وراء هذا القرار.

مثال: يطلب مكتب الحماية المالية للمستهلك الأمريكي من الدائنين في الولايات المتحدة الذين يرفضون طلبات الائتمان أن يشرحوا لمقدم الطلب السبب (الأسباب) الرئيسية لرفض طلبه (على سبيل المثال، «مدة الإقامة» أو «عمر السيارة»)[11]. وعلى وجه الخصوص، « تشير البيانات إلى أن الإجراء السلبي استند إلى المعايير أو السياسات الداخلية للدائن أو أن مقدم الطلب، أو مقدم الطلب المشترك، أو الطرف المماثل أخفق في تحقيق درجة مؤهلة في نظام التصنيف الائتماني للدائن».

ب عندما لا تكون القابلية للتفسير هذه ممكنة في ضوء التكنولوجيا المتاحة، ينظر مشغلو الذكاء الاصطناعي في التوصل لحلول وسط مثل منطق المغايرة، أو تضمين العوامل الأكثر ترجيحًا التي تساهم في اتخاذ القرار.

مثال: طورت هيئة الخدمات الصحية الوطنية في المملكة المتحدة أداة تسمى «Predict»، والتي تسمح للنساء المصابات بسرطان الثدي بمقارنة حالتهن بالنساء الأخريات اللاتي كن يعانين من نفس الحالة المرضية في الماضي، وتوضح تصور لمعدل البقاء المتوقع في ظل خيارات العلاج المختلفة. يتضمن موقع الويب صفحة تفسيرية توضح الترجيحات الكامنة وراء العوامل المختلفة وتحتوي على وصف للحسابات الأساسية[12].

ت في حالة توفر مثل هذه التفسيرات، ينبغي أن تكون سهلة الوصول وسريعة ومجانية وسهلة الاستخدام.

5-3 إحكام وأمان الذكاء الاصطناعي (المبدأ 5)

1-5-3 المبدأ

جعل أنظمة الذكاء الاصطناعي قوية ومحكمة من الناحية الفنية

لضمان استيفاء هذا المعيار، يجب أن:

- **اتباع نهج وقاية من المخاطر يسير بطريقة تجعلها تسلك سلوكًا موثوقًا.**
- **يضمن مطورو الذكاء الاصطناعي ألا تسبب أنظمة الذكاء الاصطناعي في وقوع أي ضرر غير مقصود أو حدوث أي آثار سلبية.**
- **تكون أنظمة الذكاء الاصطناعي قادرة على الصمود أمام الهجمات الأمنية مثل تسمم البيانات (أي تقديم بيانات خالية من المعنى أو مؤذية بهدف التأثير على أداء نماذج التعلم الآلي وخوارزميات الذكاء الاصطناعي المختلفة) وتسريب النماذج.**
- **يكون لأنظمة الذكاء الاصطناعي ضمانات تمُكّن من وضع خطة احتياطية في حالة حدوث مشكلات.**
- **تكون نتائج نظام الذكاء الاصطناعي قابلة للتكرار.**

[11] مكتب الحماية المالية للمستهلك، تقرير الحماية المالية 12 الجزء 1002 - قانون تكافؤ الفرص الائتمانية (اللائحة ب)، الفقرة1002.9 الإشعارات، مأخوذ من: https://www.consumerfinance.gov/policy-compliance/rulemaking/regulations/1002/

[12] موقع بريدكت، مأخوذ من هذا الرابط: http://www.predict.nhs.uk/predict_v2.1/legal/algorithm

2-5-3 الإرشادات

1-2-5-3

يستمر مشغلو الذكاء الاصطناعي في إجراء تقييمات لنقاط الضعف، والتحقق من السلوكيات المختلفة لنظام الذكاء الاصطناعي في المواقف غير المتوقعة ولأي حالة من حالات الاستخدام المزدوج، لتشمل:

أ. وضع تدابير لضمان سلامة ومرونة نظام الذكاء الاصطناعي ضد الهجمات المحتملة.

مثال: قدم باحثو شركة آي بي إم طريقة جديدة لاستغلال الذكاء الاصطناعي لأغراض خبيثة، حيث يقوم نظامهم، ويسمى « DeepLocker» بتضمين قدرات الذكاء الاصطناعي داخل البرامج الخبيثة نفسها من أجل تحسين أساليب المراوغة. يستخدم نموذج الذكاء الاصطناعي لتحديد هدفه باستخدام مؤشرات مثل التعرف على الوجه وتحديد الموقع الجغرافي والتعرف على الصوت.

ب. التحقق من كيفية تصرف أنظمة الذكاء الاصطناعي في المواقف والبيئات غير المتوقعة.

ت. اتخاذ تدابير وقائية ضد أي حالة استخدام مزدوج محتملة للنظام.

2-2-5-3

يضع مشغلو الذكاء الاصطناعي خطة احتياطية مناسبة مع اختبارها للحفاظ على الجاهزية في المواقف غير المتوقعة وفي مستوى مخاطر السلامة العالية، لتشمل:

أ. وضع خطة لقياس وتقييم مخاطر السلامة المحتملة التي قد تتعرض لها أنت أو أي طرف أخر بسبب سوء استخدام التكنولوجيا العرضي أو الضار.

ب. تحديد عتبات للنتائج المقبولة للنظام وإجراءات الحوكمة للرجوع إلى الخطط البديلة المحددة والمختبرة.

ت. النظر في شراء وثيقة تأمين للتخفيف من المخاطر الناشئة عن الأضرار المحتملة.

3-2-5-3

يضمن مشغلو الذكاء الاصطناعي موثوقية النظام للمستخدمين النهائيين من خلال عمليات التوثيق والتشغيل للاختبار والتحقق من النتائج المرجوة، وعلى وجه التحديد:

أ. ينبغي تصميم نهج لنظام الذكاء الاصطناعي لمواصلة مراقبة ما إذا كان النظام يلبي الأهداف والأغراض والتطبيق المقصود منه أم لا.

ب. يُحدد مشغلو الذكاء الاصطناعي نهج لضمان إمكانية تكرار النتائج من خلال عمليتي التشغيل والتوثيق الواضحتين.

ت. يضمن ويوثق مشغلو الذكاء الاصطناعي قوة وإحكام النظام للمستخدمين النهائيين.

4-2-5-3

تكون أنظمة الذكاء الاصطناعي آمنة ويمكن للبشر التحكم فيها

أ. تكون سلامة وأمن الأشخاص، سواء كانوا مشغلين أو مستخدمين نهائيين أو أي أطراف أخرى، موضع اهتمام بالغ في تصميم أي نظام ذكاء اصطناعي.

ب. تكون أنظمة الذكاء الاصطناعي آمنة ويمكن التحقق منها والتحكم فيها طوال عمرها التشغيلي، إلى الحد الذي تسمح به التكنولوجيا.

ت. مراعاة استمرار أمان وخصوصية المستخدمين عند إيقاف تشغيل أنظمة الذكاء الاصطناعي.

ث. تتلقى أنظمة الذكاء الاصطناعي، التي تؤثر تأثيرًا مباشرًا وكبيرًا في حياة الناس، عناية متناسبة في تصميمها.

ج. تكون هذه الأنظمة قابلة للإلغاء أو التراجع عن قراراتها على أيدي أفراد معينين.

[13] شركة آي بي إم: https://documents.trendmicro.com/assets/white_papers/wp-malicious-uses-and-abuses-of-artificial-intelligence.pdf

5-2-5-3

ألا تكون أنظمة الذكاء الاصطناعي قادرة على إيذاء البشر، أو إلحاق الضرر بهم، أو خداعهم أليًا

أ إنشاء أنظمة الذكاء الاصطناعي لغرض تقديم الخدمة والإعلام، وليس لغرض الخداع والتلاعب.

ب للدول أن تتعاون مع بعضها البعض لتفادي حدوث سباق تسلح في الأسلحة الفتاكة ذاتية التشغيل، وينبغي مراقبة هذه الأسلحة بإحكام.

ت السعي إلى تفعيل التعاون النشط لتلافي الاختلاف في معايير السلامة.

ث النظم مصممة لإبلاغ القرارات المهمة القيام بذلك بنزاهة.

6-3 الذكاء الاصطناعي المتمحور حول الإنسان (المبدأ 6)

1-6-3 المبدأ

منح أنظمة الذكاء الاصطناعي قيمًا بشرية وجعلها مفيدة للمجتمع

ولضمان استيفاء هذا المعيار، يجب:

- **دعم الحكومة للبحث في الاستخدام المفيد للذكاء الاصطناعي.**
- **إشراك أصحاب المصلحة في المجتمع في تطوير الذكاء الاصطناعي وحوكمته.**
- **تصميم أنظمة الذكاء الاصطناعي لتبني وتعلم واتباع قواعد وقيم المجتمع الذي تخدمه.**
- **الحاجة إلى إجراء التحليلات المنهجية التي تدرس أخلاقيات تصميم النظم العاطفية لتحفيز البشر قبل نشرها.**
- **عدم تفويض القرارات المتعلقة بالقوة المميتة والحياة والموت لأنظمة الذكاء الاصطناعي، بل يجب اعتماد القواعد والمعايير لضمان السيطرة البشرية الفعالة على تلك القرارات.**

2-6-3 الإرشادات

1-2-6-3

تطوير الذكاء الاصطناعي ليتماشى مع القيم الإنسانية، ويساهم في ازدهار الإنسان ويفيد المجتمع ككل.

أ استشارة المجتمع بطريقة تمثيلية لإثراء تطوير الذكاء الاصطناعي.

ب إشراك أصحاب المصلحة في المجتمع في تطوير الذكاء الاصطناعي وحوكمته.

ت الاستعانة بأصحاب المصلحة بشكل استباقي في الإشراف المسؤول على الذكاء الاصطناعي الجدير بالثقة في السعي لتحقيق نتائج مفيدة للناس وكوكبنا.

ث إعطاء المنظمات الأولوية لتثقيف جميع أصحاب المصلحة حول مقاييس الرفاهية كمحدد محتمل لكيفية إنشاء تقنيات الذكاء الاصطناعي الخاصة بها، ونشرها، وتسويقها، ومراقبتها.

مثال: تحدد ممارسة[14] معهد مهندسي الكهرباء والإلكترونيات أو آي تربل إي بي 7010 الموصى بها مقاييس الرفاهية المتعلقة بالعوامل البشرية التي تتأثر تأثيرًا مباشرًا بالأنظمة الذكية والمستقلة والتي تضع أساسًا لأنواع البيانات الموضوعية والذاتية التي يجب على هذه الأنظمة تحليلها وتضمينها (في برامجها وأدائها) لزيادة رفاهية الإنسان بشكل استباقي.

2-2-6-3

الحاجة إلى إجراء التحليلات المنهجية التي تدرس أخلاقيات تصميم النظم العاطفية لتحفيز البشر قبل نشرها.

أ يكون للتحفيز في أنظمة الذكاء الاصطناعي سياسة نظام قبول بموجب موافقة صريحة.

ب نوصي، عند الاقتضاء، أن يتمتع النظام العاطفي الذي يحفز البشر بالقدرة على التمييز بدقة بين المستخدمين، بما في ذلك اكتشاف الخصائص: مثل ما إذا كان المستخدم شخصًا بالغًا أم طفلاً. يجب وضع تدابير حماية إضافية للمستخدمين الأضعف، مثل الأطفال، عند صعوبة الحصول على الموافقة المستنيرة، أو عندما لا تكون هذه ضمانة كافية.

ت تقييم أنظمة الذكاء الاصطناعي مع استراتيجيات التحفيز، ومراقبتها بعناية.

3-2-6-3
تحتفظ البشرية بالقدرة على إدارة نفسها واتخاذ القرار النهائي، مع قيام الذكاء الاصطناعي بدور مساعد.

أ. عدم تفويض القرارات المتعلقة بالقوة المميتة والحياة والموت لأنظمة الذكاء الاصطناعي، بل يجب اعتماد القواعد والمعايير لضمان السيطرة البشرية الفعالة على تلك القرارات.

ب. تدريب الأطراف المسؤولة (مثل الآباء، والممرضات الممارسين، والأخصائيين الاجتماعيين، والحكومات) على اكتشاف التأثير الناتج عن الذكاء الاصطناعي وتدابير التخفيف غير الفعالة. في الحالات القصوى، يجب أن يكون من الممكن دائمًا إيقاف تشغيل نظام ذكاء اصطناعي ضار.

ت. تصميم الإجراءات التي يقوم بها نظام عاطفي، والتي من المرجح أن تولد استجابة عاطفية، بطريقة قابلة للتغيير بسهولة.

ث. يكون المستخدمون قادرون على اتخاذ قرارات مستنيرة بشأن أنظمة الذكاء الاصطناعي، عن طريق إلمامهم بالمعرفة والأدوات المناسبة لفهم أنظمة الذكاء الاصطناعي والتفاعل معها تفاعلاً مرضيًا، وحيثما أمكن إجراء تقييم ذاتي معقول أو رفض النظام.

4-2-6-3
تصميم أنظمة الذكاء الاصطناعي بطريقة تحترم سيادة القانون وحقوق الإنسان وقيم المجتمع، وأن تتضمن ضمانات مناسبة لضمان عدالة وانصاف المجتمع.

أ. يُصمم مطوري الذكاء الاصطناعي أنظمة الذكاء الاصطناعي لاعتماد، وتعلم، واتباع معايير وقيم المجتمع الذي يخدمونه.

ب. تحدد المنظمات معايير المجتمع المحدد الذي ستُنشر فيه أنظمة الذكاء الاصطناعي. وعلى وجه الخصوص، الانتباه للمعايير ذات الصلة بأنواع المهام المصممة أنظمة الذكاء الاصطناعي لأدائها. يمكن لنظام الذكاء الاصطناعي توثيقها وكذلك معالجة هذه المعايير.

ت. لغرض الاستجابة للتغيير الديناميكي للمعايير في المجتمع، يمكن لنظام الذكاء الاصطناعي أن يكون قادرًا على تعديل معاييره الحالية وتعلم معايير جديدة، مع التحلي بالشفافية بشأن هذه التغييرات.

ث. ينظر المصممون في النماذج والمقاييس بقصد تقييم التوافق المعياري لنظام ذكاء اصطناعي على مدى عمر النظام (مثل اتفاق الإنسان والآلة على القرارات الأخلاقية، وإمكانية التحقق من قرارات الذكاء الاصطناعي، والثقة المبررة).

ج. توثق عملية تحديد المعايير أوجه التشابه والاختلاف بين المعايير التي يطبقها البشر على البشر الآخرين والمعايير التي يطبقونها على أنظمة الذكاء الاصطناعي، مع تقييم تطبيق المعايير خاصة مقابل المعايير التي يتوقع المجتمع أن يتبعها نظام الذكاء الاصطناعي.

ح. في الحالات التي تكون فيها هناك حاجة لذلك، ينبغي النظر في تقييمات الأثر على حقوق الإنسان والعناية الواجبة بتلك الحقوق، أو ميثاق السلوك الأخلاقي للإنسان، أو علامات الجودة والشهادات التي تهدف إلى تعزيز القيم التي تتمحور حول الإنسان والإنصاف.

خ. وضع آليات لتلقي ردود الفعل الخارجية بشأن أنظمة الذكاء الاصطناعي التي من المحتمل أن تنتهك الحقوق الأساسية.

[14] معهد مهندسي الكهرباء والإلكترونيات أو آي تربل إي. https://sagroups.ieee.org/7010/

7-3 الذكاء الاصطناعي المستدام والصديق للبيئة (المبدأ 7)

1-7-3 المبدأ

تعزيز الذكاء الاصطناعي المستدام بطريقة صديقة للبيئة

ولضمان استيفاء هذا المعيار، يجب:

- **استخدام أنظمة الذكاء الاصطناعي لإفادة جميع البشر، بما في ذلك الأجيال القادمة، والبيئة أساسية لذلك.**
- **طوال دورة حياة نظام الذكاء الاصطناعي، إنشاء التطبيقات على أساس الفهم والاعتراف الكامل بالآثار المترتبة على الاستدامة والبيئة.**
- **تقييم تطوير نظام الذكاء الاصطناعي ونشره واستخدامه من خلال الفحص الدقيق لاستخدام الموارد واستهلاك الطاقة.**
- **إنشاء آليات لقياس الأثر البيئي بسبب نوع استخدام الطاقة وطاقة المعالجة التي توفرها مراكز البيانات.**

2-7-3 الإرشادات

1-2-7-3 طوال دورة حياة نظام الذكاء الاصطناعي، إنشاء التطبيقات دائمًا بعد الفهم والاعتراف الكامل بآثار الذكاء الاصطناعي على الاستدامة والبيئة.

أ. فهم تطبيق الذكاء الاصطناعي لمواجهة تحديات الاستدامة فهمًا جيدًا. تتزايد أهمية إنشاء وتشغيل أنظمة الذكاء الاصطناعي الخضراء والمستدامة، بالنظر إلى البصمة الكربونية الكبيرة التي يمكن أن تولدها، والسياق الأوسع للتعامل مع تغير المناخ.

ب. تطوير نموذج الذكاء الاصطناعي، من خلال التشغيل الموسع، للسعي إلى تحقيق التوازن بين الأداء التقني واستهلاك الطاقة والأثر البيئي.

ت. بذل الجهود لتقدير وفهم الاستدامة والأثر البيئي عبر دورة حياة نظام الذكاء الاصطناعي (على سبيل المثال، تكاليف استهلاك الطاقة المرتبطة بتدريب النموذج، وانبعاثات ثاني أكسيد الكربون، وتكاليف الحوسبة السحابية المرتبطة بنشر النظام وتشغيله).

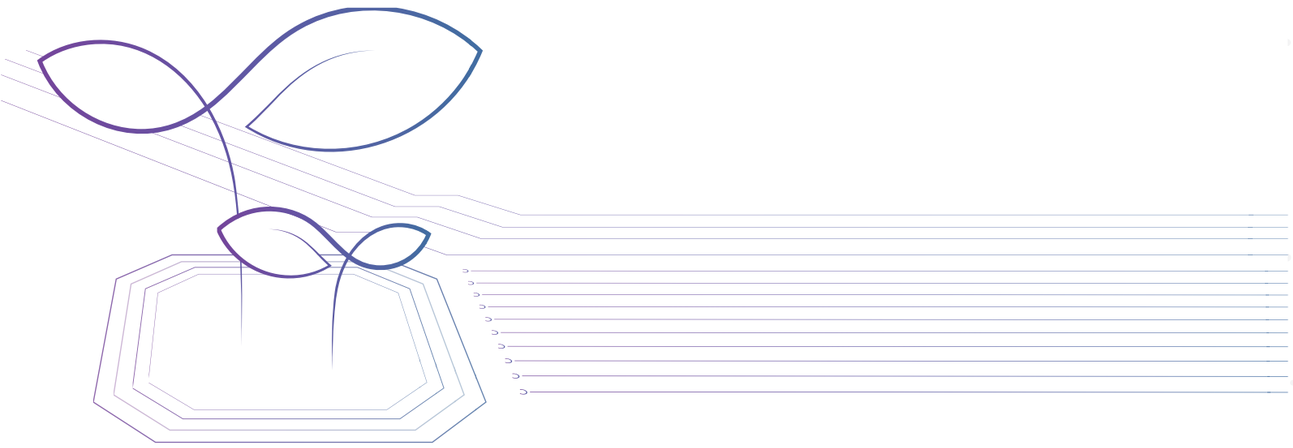
ث. في حال عدم وجود وسائل موحدة للإبلاغ عن التأثير البيئي لأنظمة الذكاء الاصطناعي، حيثما أمكن، يجب جعل تأثيرات الكربون اعتبارًا أساسيًا جنبًا إلى جنب مع المتطلبات الوظيفية والتجارية عند تطوير النماذج وتشغيلها.

ج. النظر في استخدام نماذج أصغر - تقليص حجم النموذج واستخدام دورات حساب أقل (لموازنة التكاليف المالية وتكاليف الأداء مع الأداء النهائي للنموذج).

ح. مراعاة الوعي بالكربون - تعديل المؤشرات التشغيلية لنظام الذكاء الاصطناعي وتحديد أفضل وقت ومكان ديناميكي لاستخدام الطاقة من الشبكة سيؤدي إلى تقليل انبعاثات الكربون.

مثال 1 : في مجال الزراعة، يمكن للذكاء الاصطناعي تحويل الإنتاج بشكل جيد من خلال مراقبة وإدارة الظروف البيئية وإنتاج المحاصيل. يمكن أن يساعد الذكاء الاصطناعي في تقليل كل من الأسمدة والمياه المستخدمة، مع تحسين إنتاج المحاصيل.

مثال 2 : طور قسم DeepMind في غوغل ذكاءً اصطناعيًا يُعلم نفسه تقليل استخدام الطاقة لتبريد مراكز بيانات غوغل. نتيجة لذلك، خفضت غوغل متطلبات الطاقة في مركز البيانات بنسبة 40 ٪.[15]



8-3 الذكاء الاصطناعي المحافظ على الخصوصية (المبدأ 8)

1-8-3 المبدأ

احترام خصوصية الناس

ولضمان استيفاء هذا المعيار، يجب:

- **على أنظمة الذكاء الاصطناعي احترام الخصوصية، وأن تستخدم الحد الأدنى من التطفل الضروري.**
- **التزام أنظمة الذكاء الاصطناعي بالمعايير العالية لإدارة البيانات وأمنها، وحماية المعلومات الشخصية.**
- **عدم نشر المراقبة أو غيرها من التقنيات المعتمدة على الذكاء الاصطناعي بطريقة تنتهك المعايير المقبولة دوليًا و / أو المعايير الإماراتية للخصوصية والكرامة الإنسانية وحقوق الأشخاص.**
- **تضمين الخصوصية حسب التصميم في أنظمة الذكاء الاصطناعي، وحيثما أمكن، يكون لخوارزمية الذكاء الاصطناعي تقييمات مناسبة لتأثير الخصوصية.**
- **إنشاء أطر حماية للبيانات وآليات الحوكمة المناسبة في نهج أصحاب المصلحة المتعددين وضمانها طوال دورة حياة أنظمة الذكاء الاصطناعي.**
- **يسعى مطورو ومشغلو الذكاء الاصطناعي لتحقيق التوازن بين متطلبات الخصوصية، والحقوق الفردية، وتطور الابتكار، ومصالح المجتمع.**

2-8-3 الإرشادات

1-2-8-3
إنشاء آلية للمستخدمين للإبلاغ عن القضايا المتعلقة بالخصوصية أو حماية البيانات

أ. يتولى مشغلو الذكاء الاصطناعي مراجعة النظام من أجل تسجيل إذن الدخول تسجيلاً مناسبًا، مع قدرة المستخدمين على إلغاء الإذن، كلما أمكن ذلك.

ب. النظر في تدريب نموذج الذكاء الاصطناعي بدون أو بالاستعانة بالحد الأدنى من المستخدمين للبيانات الحساسة أو الشخصية.

ت. استخدام تدابير لتعزيز الخصوصية مثل؛ التشفير وإخفاء الهوية والتجميع.

2-2-8-3
إنشاء آلية إشراف لجمع البيانات، وتخزينها، ومعالجتها، واستخدامها عبر المؤسسة

أ. تقييم الأشخاص المخولين بالوصول إلى البيانات، وشروط ذلك.

ب. تعيين مسؤوليات ودور معين لمسؤولي حماية البيانات.

ت. منع الضرر الذي يلحق بالخصوصية من خلال حوكمة البيانات بشكل ملائم يغطي جودة وسلامة البيانات المستخدمة، ومدى أهميتها في ضوء المجال الذي ستنشر أنظمة الذكاء الاصطناعي فيه، وبروتوكولات الوصول الخاصة بها، والقدرة على معالجة البيانات بطريقة تحمي الخصوصية.

ث. تضمن أنظمة الذكاء الاصطناعي حماية الخصوصية والبيانات طوال دورة حياة نظام الذكاء الاصطناعي بالكامل. يتضمن ذلك المعلومات التي قدمها المستخدم في البداية، بالإضافة إلى المعلومات المستحدثة حول المستخدم على مدار تفاعله مع النظام (على سبيل المثال، النتائج التي أنشأتها أنظمة الذكاء الاصطناعي حول مستخدمين محددين، أو كيفية استجابة هؤلاء المستخدمين لتوصيات معينة). لن تسمح السجلات الرقمية للسلوك البشري لأنظمة الذكاء الاصطناعي باستنتاج تفضيلات الأفراد فقط، ولكن ستستنتج أيضًا أعمارهم، أو نوعهم، أو آرائهم الدينية، أو السياسية. ولمنح الأفراد الثقة في عملية جمع البيانات، يجب التأكد من عدم استخدام البيانات التي تم جمعها عنهم للتمييز ضدهم بشكل غير قانوني أو غير عادل.

ج. في أي منظمة تتعامل مع بيانات الأفراد (سواء كان مستخدم النظام شخصًا ما أم لا)، يجب وضع بروتوكولات للبيانات التي تحكم الوصول إلى البيانات بما يتماشى مع تشريعات الخصوصية الوطنية (سواء العالمية أو الخاصة بالقطاع). تُحدد هذه البروتوكولات الأشخاص المخولين بالوصول إلى البيانات وشروط ذلك. يسمح فقط للموظفين المؤهلين حسب الأصول ذوي الكفاءة والحاجة للوصول إلى بيانات الأفراد الاضطلاع بذلك.

ج تتطلب أنظمة الخوارزميات إجراء تقييمات مناسبة لتأثير الخصوصية، والتي تشمل أيضًا الاعتبارات المجتمعية والأخلاقية لاستخدامها والاستخدام المبتكر لنهج الخصوصية حسب التصميم. تضمن الجهات الفاعلة في مجال الذكاء الاصطناعي أنها مسؤولة عن تصميم وتنفيذ أنظمة الذكاء الاصطناعي بطريقة تعكس حماية المعلومات الشخصية طوال دورة حياة نظام الذكاء الاصطناعي.

خ وضع سياسات بيانات أو أطر عمل مكافئة، أو تعزيز السياسات القائمة لضمان الأمن الكامل للبيانات الشخصية والبيانات الحساسة، والتي، إذا تم الكشف عنها، قد تتسبب في حدوث أضرار أو إصابات أو صعوبات استثنائية للأفراد. تشمل الأمثلة على ذلك: البيانات المتعلقة بالجرائم والإجراءات الجنائية والإدانات والتدابير الأمنية ذات الصلة؛ والبيانات البيومترية والجينية والصحية؛ والبيانات الشخصية مثل تلك المتعلقة بالعرق، أو النسب، أو النوع، أو العمر، أو اللغة، أو الدين، أو الرأي السياسي، أو الأصل القومي، أو الأصل العرقي، أو الأصل الاجتماعي، أو الحالة الاقتصادية، أو الاجتماعية، أو الإعاقة وأي خصائص أخرى.

د تعزيز الآليات، مثل المستودعات المفتوحة للبيانات الممولة من القطاع العام أو المملوكة للجمهور، ورموز المصدر، وصناديق البيانات، لدعم التبادل الآمن والعادل والقانوني والأخلاقي للبيانات، من بين أمور أخرى.

ذ تعزيز وتسهيل استخدام مجموعات بيانات قوية وعالية الجودة للتدريب، وتطوير واستخدام أنظمة الذكاء الاصطناعي، وتوخي اليقظة في الإشراف على جمعها واستخدامها.

مثال: بمساعدة خوارزميات[16] الذكاء الاصطناعي والتعلم الآلي ، يمكن إنشاء بيانات اصطناعية لتعزيز الخصوصية. يكون لتلك البيانات المنشأة نفس الخصائص الإحصائية لبيئة الاختبار ومتى يمكن لطرف خارجي الوصول إليها.

16 https://research.aimultiple.com/privacy-enhancing-technologies/

04 __
مصادر الكتب

مايكروسوفت 2022. معيار غوغل للذكاء الاصطناعي المسؤول، الإصدار 2. المتطلبات العامة. يونيو 2022. مأخوذ من:
https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4ZPmV

الايزو. 2022. معيار الأيزو المقترح بشأن إدارة مخاطر الذكاء الاصطناعي: ما يجب أن تعرفه الشركات. مايو 2022. مأخوذ من:
https://unesdoc.unesco.org/ark:/48223/pf0000381137/PDF/381137eng.pdf.multi

الجمعية الدولية لشبكات الهاتف المحمول أو جي إس إم إيه.2022. دليل أخلاقيات الذكاء الاصطناعي. تطبيق المبادئ الأخلاقية في الأعمال اليومية. فبراير 2022. مأخوذ من:
The-Mobile-Industry-Ethics-/01/https://www.gsma.com/betterfuture/wp-content/uploads/2022
Playbook_Feb-2022.pdf

منظمة التعاون والتنمية في الميدان الاقتصادي. 2022. مرصد سياسات الذكاء الاصطناعي. مأخوذ من:
https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

اليونسكو. 2022. توصية بشأن أخلاقيات الذكاء الاصطناعي. اعتمد في نوفمبر 2021. مأخوذ من:
https://unesdoc.unesco.org/ark:/48223/pf0000381137/PDF/381137eng.pdf.multi

المفوضية الأوروبية. 2021. مقترح لتنظيم البرلمان الأوروبي والمجلس الأوروبي لوضع القواعد المنسقة بشأن الذكاء الاصطناعي (قانون الذكاء الاصطناعي) وتعديل بعض القوانين التشريعية للاتحاد. مأخوذ من:
https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

المفوضية الأوروبية. فريق الخبراء رفيع المستوى المعني بالذكاء الاصطناعي (AI HLEG). المبادئ التوجيهية الأخلاقية للذكاء الاصطناعي الجدير بالثقة. أبريل 2019. مأخوذ من:
https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

مركز أخلاقيات البيانات والابتكار. حكومة المملكة المتحدة: خارطة الطريق إلى نظام بيئي فعال لضمان الذكاء الاصطناعي. ديسمبر 2021. بالرجوع إليه من:
https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem

مجلس مدينة نيويورك. 2021. قانون محلي لتعديل القانون الإداري لمدينة نيويورك، فيما يتعلق بأدوات قرار التوظيف الآلي. مأخوذ من:
https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=%20B051915D-A9AC-&6596032FA3F9-451E-81F8

اليونسكو، توصية بشأن أخلاقيات الذكاء الاصطناعي. 2021. مأخوذ من:
https://unesdoc.unesco.org/ark:/48223/pf0000381137

معهد « AI NOW». المساءلة الحسابية للقطاع العام. أغسطس 2021. مأخوذ من:
/https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector

معهد مهندسي الكهرباء والإلكترونيات أو آي تربل إي 2017. تصميم أخلاقي: رؤية لإعطاء الأولوية لرفاهية الإنسان مع الأنظمة المستقلة والذكية. مأخوذ من:
https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

جوجل. مبادئ الذكاء الاصطناعي 2021. مأخوذ من:
https://ai.google/principles/

مايكروسوفت. الإصدار 2 من معيار الذكاء الاصطناعي المسؤول (يونيو 2022). مأخوذ من:
https://www.microsoft.com/en-us/ai/responsible-ai

مجموعة ايكونوميست 2022. تقرير مستقبل الذكاء الاصطناعي في الشرق الأوسط وشمال إفريقيا. مأخوذ من:
https://impact.economist.com/perspectives/sites/default/files/google_ai_mena_report.pdf

لجنة حماية البيانات الشخصية. (5 يونيو 2018). مناقشة بحثية حول الذكاء الاصطناعي (AI) والبيانات الشخصية. سنغافورة: لجنة حماية البيانات الشخصية في سنغافورة (PDPC). مأخوذ من:
https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-forOrganisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf

مجلس صناعة تكنولوجيا المعلومات. مبادئ سياسة الذكاء الاصطناعي. مأخوذ من:
https://www.itic.org/public-policy/ITIAIPolicyPrinciplesFINAL.pdf

مكتب مجلس الوزراء (2016، 19 مايو). الإطار الأخلاقي لعلوم البيانات. مأخوذ من:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524298/Data_science_ethics_framework_v1.0_for_publication__1_.pdf

البرلمان الأوروبي (16 فبراير 2017). قرار البرلمان الأوروبي بتاريخ 16 فبراير 2017 مع توصيات إلى لجنة قواعد القانون المدني بشأن الروبوتات (2103/2015 (INL)). مأخوذ من:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524298/Data_science_ethics_framework_v1.0_for_publication__1_.pdf

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

اللجنة الوطنية للمعلوماتية والحريات (2017). كيف يمكن للبشر أن يكون لهم اليد العليا؟ المسائل الأخلاقية التي تثيرها الخوارزميات والذكاء الاصطناعي. مأخوذ من:
https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf

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

المكتب التنفيذي لرئيس المجلس الوطني للعلوم والتكنولوجيا بالولايات المتحدة. لجنة التكنولوجيا. (2016). التحضير لمستقبل الذكاء الاصطناعي. مأخوذ من:
https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

مقار التنشيط الاقتصادي الياباني. (2015). استراتيجية روبوت جديدة. استراتيجية الروبوت في اليابان. الرؤية والاستراتيجية وخطة العمل. مأخوذ من:
http://www.meti.go.jp/english/press/2015/pdf/0123_01b.pdf

مجلس الخزانة التابع لمجلس الوزراء الكندي. (2018). الذكاء الاصطناعي المسؤول في حكومة كندا. سلسلة المستندات التعريفية للاضطرابات الرقمية. الإصدار 2.0
https://docs.google.com/document/d/1Sn-qBZUXEUG4dVk909eSg5qvfbpNlRhzIefWPtBwbxY/edit

إعلان تورنتو: حماية الحق في المساواة وعدم التمييز في أنظمة التعلم الآلي. تورنتو، كندا: منظمة العفو الدولية وأكسيس ناو. مأخوذ من:
https://www.accessnow.org/cms/assets/uploads/201808//The-Toronto-Declaration_ENG_082018-.pdf

http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/oral/73546.html

(مجلس اللوردات لجنة اختيار الأدلة الشفوية، 61)

مجلس اللوردات لجنة اختيار الذكاء الاصطناعي. (2018). الذكاء الاصطناعي في المملكة المتحدة: جاهزون وراغبون وقادرون؟

لجنة اختيار العلوم والتكنولوجيا بمجلس العموم. الخوارزميات في صنع القرار.