SAFETY ACCURACY AND RELIABILITY THEMATIC AREA NARRATIVE IN ENGLISH ARABIC FRENCH PORTUGUESE AND SPANISH

Rachel Adams , Kelly Stone

Rachel Adams , Kelly Stone

IDRC GRANT / SUBVENTION DU CRDI : - GLOBAL INDEX ON RESPONSIBLE ARTIFICIAL INTELLIGENCE

1. 3. 1 السلامة والدقة والمصداقية

المؤشر العالمي للذكاء الاصطناعي المسؤول

البعد: الحوكمة المسؤولة للذكاء الاصطناعي
البعد الفرعي: المعايير التقنية
المجال المواضيعي: السلامة والدقة والمصداقية

تعريف المصطلحات

يمكن تعريف السلامة بصفة عامّة على أنها حالة الشخص المحميّ ضدّ أيّ خطر أو مخاطرة أو إصابة، أو الذي من المحتمل ألّا يكون السبب لهذ الخطر أو المخاطرة أو الإصابة. وخلال تطبيقها على سياقات الذكاء الاصطناعي، فإنّ السلامة تركّز بشكل خاص على الحلول التقنية للتثبّت من أنّ أنظمة الذكاء الاصطناعي تشتغل بأمان ومصداقية ولا تسبّب أضرارا جديدة ولا تؤدي إلى تزايد المخاطر الموجودة حاليّا.

وترتبط مبادئ الدقة والمصداقية بشدّة بمفهوم سلامة الذكاء الاصطناعي. إذ يمكن تعريف الدّقّة بأنها "جودة أو حالة الدقّة والصواب". وعندما يكون الشيء دقيقا فهو **صائب**، أي أنّ الأمر لا يتعلّق بتقريب أو تقدير وبالتالي، فإنّ نتيجته تكون خالية من الأخطاء أو العيوب. وفي سياق الذكاء الاصطناعي، يتطلّب مبدأ الدقة اتّخاذ إجراءات آنية لإزالة أو إصلاح البيانات غير الدقيقة لضمان أنّ نتائج الخوارزميات صحيحة ودقيقة.

كما يمكن تعريف المصداقية بأنّها"جودة استحقاق الثقة أو تقديم نتائج جيدة باستمرار". وبعبارة أخرى، إنّ المصداقية هي الدرجة التي يمكن الوثوق بها في **دقّة** نتيجة قياس أو حساب أو مواصفات

خاصة. وفي سياق الذكاء الاصطناعي، فإنّ المصداقية هي إحتمالية قيام أداة أو نظام للذكاء الاصطناعي بوظيفته بطريقة صحيحة خلال مدة زمنية محددة وتحقيق نتائج متسقة وقابلة للإعادة.

وبالتالي، يمكن اعتبار الدقة والمصداقية مبدأين متميزين ولكن مترابطين يساهمان في سلامة الذكاء الاصطناعي، وهما من المتطلبات الأساسية لإنشاء أنظمة وأدوات ذكاء اصطناعي موثوق به.

الأسس (النظرية)

يثير ظهور التقنيات المعتمدة على الذكاء الاصطناعي في مختلف الأنشطة البشرية تساؤلات جديدة حول سلامة هذه الأدوات ومصداقية قراراتها ونتائجها. وقد عزّزت الحركات العالمية الجهود الرامية إلى زيادة سلامة الذكاء الاصطناعي، وخاصة عبر الاهتمام بقيمة المعايير الصارمة فيما يتعلق بدقة ومصداقية أنظمة الذكاء الاصطناعي باعتبارها وسيلة لتخفيف الأضرار وتقليل المخاطر التي تواجه الأفراد والجماعات والمجتمعات التي تتطلع إلى استغلال قوة الذكاء الاصطناعي.

تحدّد منظمة التعاون الاقتصادي والتنمية سلامة الذكاء الاصطناعي باعتباها أحد مبادئها الخمسة القائمة على قيم. إذ ينصّ المبدأ 1. 4 على أنّ "أنظمة الذكاء الاصطناعي يجب أن تعمل بقوة وأمان طوال حياتها، كما يجب تقييم المخاطر المحتملة وإدارتها باستمرار". ولذلك تدعو منظمة التعاون الاقتصادي والتنمية المنتفعين في مجال الذكاء الاصطناعي إلى "ضمان المتابعة بما يشمل مجموعات البيانات والعمليات والقرارات المتخذة خلال دورة حياة نظام الذكاء الاصطناعي"، وذلك لغاية ضمان أن تكون القرارات والنتائج ملائمة للسياق ومتماسكة بشأن الهدف المرجوّ من التكنولوجيا. ولتحقيق ذلك، يجب إدماج معايير الدقة والمصداقية طوال دورة حياة نظام الذكاء الاصطناعي، ليس فقط لضمان دقة واتساق نتائج الذكاء الاصطناعي وإنما أيضا لضمان أدائه المناسب في سياق معيّن.

إنّ ضمان سلامة وحماية هذه التقنيات أمر مهمّ للغاية لتعزيز الثقة في الذكاء الاصطناعي. فيجب ألا تشكل أنظمة الذكاء الاصطناعي مخاطر غير متناسبة على السلامة والحماية، بما في ذلك الحماية الجسدية في ظل ظروف الاستخدام العادي أو المتوقّع أو سوء الاستخدام طوال دورة حياة الذكاء الاصطناعي. ولأنّ تقنيات الذكاء الاصطناعي مستمدة من أنظمة وسياقات خاضعة للخطأ البشري، ومن تقاطع التحيّزات اللاواعية وأشكال اللامساواة، فإنّه من الواجب والضروري ضمان مصداقية أنظمة الذكاء الاصطناعي وأدواته.

تعريفـــات

يُقيّم هذا المجال المواضيعي الإجراءات التي اتّخذتها البلدان لزيادة سلامة الذكاء الاصطناعي من خلال الالتزام بمبدأي الدقّة والمصداقية في تصميم وتطوير واستخدام تقنيات الذكاء الاصطناعي، وبصفة خاصة، يتطلّب النظر في تحليل (1) **الأطر القانونية** المتعلقة بسلامة الذكاء الاصطناعي وتحديد مبادئ و\أو متطلبات الدقة والمصداقية، و(2) **الإجراءات الحكومية** التي تعمل من أجل تعزيز سلامة الذكاء الاصطناعي وتطبيق الإجراءات لضمان دقة ومصداقية أنظمة الذكاء الاصطناعي وأدواته، و(3) **الجهات الفاعلة غير الحكومية** التي تعمل على تعزيز سلامة الذكاء الاصطناعي واعتماد معايير و\أو متطلبات الدقة والمصداقية في بناء أنظمة ذات مصداقية للذكاء الاصطناعي.

قد تأخذ الأطر القانونية في الدولة، شكل قوانين أو لوائح، أو سياسات معتمدة، أو مشاريع سياسات (حسب القطاع و\أو القسم) و\ أو توجيهات. وقد تشمل الإجراءات الحكومية صياغة مشاريع قوانين أو سياسات تتعلق بسلامة الذكاء الاصطناعي، أو إنشاء مجموعات عمل، بالإضافة إلى إنشاء هيئات مشرفة للتحقق من امتثال السياسات المصممة لتعزيز سلامة الذكاء الاصطناعي من خلال معايير الدقة والمصداقية. أمّا الجهات الفاعلة غير الحكومية (INE) فيمكن أن تكون منظمات غير حكومية (ONG)، ولكن يمكن أن تكون أيضا شركات متعددة الجنسيات، أو منظمات عسكرية خاصة، أو وسائل إعلام، أو مجموعات عرقية منظمة، أو مؤسسات أكاديمية، أو مجموعات ضغط، أو نقابات، أو حركات اجتماعية تعمل من أجل تحسين سلامة ودقة ومصداقية تقنيات الذكاء الاصطناعي.

أمثلة:

الأطر القانونية

نشرت وزارة الاقتصاد والتجارة والصناعة اليابانية في مارس 2021، بالتعاون مع وزارة الصحة والعمل والرعاية الاجتماعية، ووكالة إدارة الحرائق والكوارث، الطبعة الثانية من المبادئ التوجيهية بشأن تقييم مصداقية الذكاء الاصطناعي في مجال سلامة المصانع. وتهدف هذه المبادئ التوجيهية إلى مساعدة الشركات المالكة للمصانع في الحصول على نتائج دقيقة وذات مصداقية لتحسين السلامة والإنتاجية من خلال توفير إطار لتقييم المصداقية بإمكان الموردين استخدامه لتنفيذ معايير الصناعة في مجال الذكاء الاصطناعي الجدير بالثقة.

إجراءات حكومية

يعمل المعهد الوطني للعلوم الصناعية والتكنولوجيا المتقدمة (AIST)، والجمعية اليابانية للتقييس (JSA) بدعم من وزارة التجارة والصناعة على تطوير المعايير الصناعية للذكاء الاصطناعي. وبالإضافة إلى ذلك فإنّ هذين المعهدين يعدّان جزءا من الهيئة الدولية للجنة الفرعية ISO/IEC JTC 1/SC 42، والتي تعمل على تطوير معايير عالمية لضمان مصداقية الذكاء الاصطناعي، ومنها بصفة خاصة متطلبات الدقة والمصداقية في النمذجة والنتائج الخوارزمية. وبالإضافة إلى ذلك استضافت طوكيو مؤتمرات مهمة في إطار اللجنة الفرعية، مثل اجتماعات SC42، والمؤتمر الدولي للشراكة العالمية للذكاء الاصطناعي لسنة 2022 (GPAI)، لتعزيز التعاون في تطبيق مبادئ الذكاء الاصطناعي بما في ذلك الدقة والمصداقية.

جهات فاعلة غير حكومية

تعد شركة فوجيتسو، وهي شركة يابانية متعدّدة الجنسيات تعمل في مجال تكنولوجيا المعلومات والاتصالات (ICT)، واحدة من أوائل الشركات التي روّجت لأخلاقيات الذكاء الاصطناعي من خلال إنشاء اللجنة الاستشارية الخارجية لأخلاقيات الذكاء الاصطناعي. وفي سنة 2022، أنشأت الشركة مكتبا خاصا بأخلاقيات الذكاء الاصطناعي وحوكمته يركّز على تنفيذ إجراءات لتعزيز الذكاء الاصطناعي بطريقة فعالة عبر التزام مجموعة فوجيستو للذكاء الاصطناعي، والذي يتضمن الالتزام "بالاستفادة من خبرتها ومعارفها المتراكمة للتطوير المتواصل وتحسين مصداقية الذكاء الاصطناعي.