

## أخلاقيات الذكاء الاصطناعي

م.م: بنين حامد جبار  
كلية الآداب / جامعة بابل

### مقدمة:

تشهد مجالات الذكاء الاصطناعي تطوراً هائلاً في السنوات الأخيرة، حيث يعد الذكاء الاصطناعي من أهم العلوم المعرفية التي تهتم بدراسة وتطوير أنظمة حاسوبية ذكية قادرة على استيعاب البيئة المحيطة بها واتخاذ إجراءات تعزز فرص نجاحها. يتضمن مفهوم الذكاء الاصطناعي محاكاة قدرات الدماغ البشري في مجالات مثل الإدراك، والتعلم، والتخطيط، والتفاعل اللغوي. ورغم التقدم الكبير الذي تحقق في هذا المجال، إلا أن هناك تحديات ومشكلات يواجهها الباحثون والمهتمون بهذا المجال.

يتطرق هذا البحث إلى استكشاف مفهوم الذكاء الاصطناعي ومعالجة بعض المسائل والتحديات التي تواجه تطبيقاته. سنتناول في هذه الورقة مشكلة الذكاء الاصطناعي الكلاسيكي وتحدياته، ونناقش أيضاً الأساليب والتقنيات المستخدمة في تطوير نماذج الذكاء الاصطناعي.

### أهداف البحث:

تحقيقاً للهدف المذكور، فإن البحث يهدف إلى تحقيق الأهداف التالية:

تعريف مفهوم الذكاء الاصطناعي وتوضيح أبعاده المختلفة.

تسليط الضوء على المشكلة الرئيسية للذكاء الاصطناعي الكلاسيكي والتحديات التي

يواجهها.

دراسة الأساليب والتقنيات المستخدمة في تطوير نماذج الذكاء الاصطناعي الحديثة.

استكشاف التطورات الحديثة في مجال الذكاء الاصطناعي وتحليل أثرها على المجتمع

والتكنولوجيا.

توجيه التساؤلات المفتوحة حول مستقبل الذكاء الاصطناعي وتطبيقاته المحتملة.

#### مشكلة البحث:

تركز مشكلة البحث على القيود والتحديات التي تواجه الذكاء الاصطناعي، حيث يعتبر محدودًا في تمثيل الذكاء البشري وفهم البيئة المعقدة والتفاعل معها. ينبغي معالجة هذه المشكلة والبحث عن تقنيات جديدة قادرة على تعزيز قدرات الذكاء الاصطناعي وتطويره. وهذا بدوره يقودنا إلى التساؤل الرئيس:

ما هي التحديات الأخلاقية المتعلقة بقرارات الذكاء الاصطناعي المستندة إلى البيانات وكيف يمكن توجيه تطور الذكاء الاصطناعي لدعم الصالح العام والمصلحة العامة؟

#### تساؤلات البحث:

يتضمن البحث العديد من التساؤلات التي يتعين تناولها وتحليلها لتحقيق أهداف البحث، ومنها:

ما هو المفهوم الشامل للذكاء الاصطناعي وما هي أبعاده المختلفة؟

ما هي المشكلة الأساسية التي تواجه الذكاء الاصطناعي الكلاسيكي وما هي التحديات المرتبطة بها؟

ما هي الأساليب المستخدمة في تطوير نماذج الذكاء الاصطناعي الحديثة وكيف يمكن تطبيقها بفعالية؟

ما هي التطورات الحديثة في مجال الذكاء الاصطناعي وما هو تأثيرها على المجتمع والتكنولوجيا؟

ما هي التطبيقات المحتملة للذكاء الاصطناعي في المستقبل وما هي التحديات المتعلقة بها؟  
المنهج المستخدم:

سيعتمد البحث على منهجية تحليلية نقدية، حيث سيتم جمع المعلومات والمصادر ذات الصلة بمفهوم الذكاء الاصطناعي وتطبيقاته. سيتم دراسة الأدبيات العلمية المتاحة والأبحاث

السابقة في هذا المجال، بالإضافة إلى مراجعة الدراسات الحالية والتطورات الأخيرة. سيتم تحليل البيانات والنتائج المستخلصة بدقة للوصول إلى استنتاجات مدروسة وموثوقة تتعلق بمشكلة الذكاء الاصطناعي وتحدياته وتطبيقاته المحتملة.

### 1- مفهوم الذكاء الاصطناعي

الذكاء الاصطناعي هو التشغيل أو البرمجة الآتية لجميع جوانب الإدراك البشري، ويعد من أسس العلوم المعرفية خلال مباشرة العمل بالرمزية والرمزية الفرعية للذكاء الاصطناعي، ومعالجة اللغات الطبيعية، ورؤية الكمبيوتر، والنظم التطورية أو التكيفية. وهذه الطريقة فإن برمجة الذكاء الاصطناعي تختلف وبشكل ملحوظ من حيث مقاييس نهج هندسة البرمجيات، حيث إن البرمجة عادة ما تبدأ من مواصفات رسمية منفصلة، بينما برمجة الذكاء الاصطناعي، تقوم على الجهود التطبيقية الفعلية كجزء من مشكلة عملية التخصيص<sup>(1)</sup>

أو هو العلم الذي يشتغل بابتكار وتطوير خوارزميات مفيدة تسهم في المحاكاة الآلية لقدرات الدماغ البشري من إدراك للبيئة المحيطة والاستجابة المناسبة لمثيراتها، وتعلم، وتخطيط، وإيجاد الحلول للمسائل المستجدة، والتواصل اللغوي، وإدارة للتراكم المعرفي، ... إلخ، ويطلق البعض على هذه القدرات وأمثالها الملكات العليا « للإنسان، ويخرج من هذا التعريف المسائل المعلوم لها تعريف رياضي محكم والمعلوم لها حلول مفيدة مبرهنة رياضياً<sup>(2)</sup>.

الذكاء الاصطناعي هو العلم الذي يهتم بدراسة وتصميم أنظمة حاسوبية ذكية تستوعب بيئتها وتتخذ إجراءات تزيد من فرص نجاحها<sup>(3)</sup>.

(1) Neumann, G.: Programming Languages in Artificial Intelligence, German Research Center for Artificial Intelligence (LT-Lab, DFKI), 2002,p3.

(2) المعتز بالله السعيد: العربية والذكاء الاصطناعي دار جوه للنشر والتوزيع، الرياض، ٢٠١٩ م، ص ٢٩

(3) علي مولود فاضل: التزييف العميق: لغة الذكاء الاصطناعي في حروب السيبران الإعلامية، دار أمجد للنشر

والتوزيع، ٢٠٢٢، ص ٢٩.

ويعرفه مارفن مينسكي Marvin Minsky بأنه مقدر الآلة على أداء أعمالاً إذا قام بها الإنسان فإننا نصف هذه الأعمال بالذكاء<sup>(1)</sup>.

ربما يكون الذكاء الاصطناعي فريداً بين الموضوعات الهندسية من حيث أنه أثار أسئلة أساسية جداً حول طبيعة الحوسبة والادراك والاستدلال والتعلم واللغة والعمل والتفاعل والوعي والبشرية والحياة وما إلى ذلك ، وفي الوقت نفس أسهم بشكل كبير في الإجابة عن هذه الأسئلة في الواقع، يُنظر إليه أحياناً على أنه شكل من أشكال البحث التجريبي<sup>(2)</sup>.

تركزت المناقشات النظرية الكلاسيكية على القضايا المتعلقة بما إذا كان الذكاء الاصطناعي ممكناً على الإطلاق (غالباً ما توضع على أنها هل يمكن للآلات التفكير ؟) أو ما إذا كان بإمكانها حل مشاكل معينة (هل يمكن للآلة أن تفعل x)<sup>(3)</sup>.

نحن الآن في مرحلة يمكننا فيها أن نرى بشكل أوضح ما البدائل، كان الرأي «الحسابي» الكلاسيكي هو أن الإدراك هو الحساب على التمثيلات، والذي قد يحدث بالتالي في أي نظام حسابي، طبيعي أو اصطناعي، من وجهة النظر الكلاسيكية هذه، يعد الذكاء الاصطناعي والعلوم المعرفية وجهين لعملة واحدة ، وقد غذى هذا الرأي جزء كبيراً من الاهتمام الفلسفي والنظري بالذكاء الاصطناعي، ومع ذلك، فإن معظم السمات المميزة لهذا التوافق القديم مهددة الآن، إذن هل يجب أن نتخلى عن وجهة النظر الكلاسيكية، هل يجب أن نكملها، أم يجب أن ندافع عنها في مواجهة النقد العصري<sup>(4)</sup>؟

(1) رأفت الكمار: الحاسوب وعصر شغيلة المعرفة، دار الكتب العلمية للنشر والتوزيع، ٢٠٠٨، ص ٩٤.

(2) Müller VC (2012) Introduction: philosophy and theory of artificial intelligence, Mind Mach 22

(2):67-69, p67

(3) Ibid .p67.

189 (4) Ibid.p67.

الهدف من الذكاء الاصطناعي بناء آلات قادرة على أداء المهام التي تتطلب ذكاء ، مثل التفكير والتعلم والتخطيط وحل المشكلات والإدراك<sup>(1)</sup> فهذا العلم لا يهدف إلى مشابهة العقل البشري بالآلة ، بل يهدف إلى فهم العمليات الذهنية المعقدة التي يقوم بها العقل البشري أثناء ممارسته (التفكير) ، ومن ثم ترجمة هذه العمليات الذهنية إلى ما يوازنها من عمليات حاسب تزيد من قدرة الحاسب على حل المشاكل المعقدة، ويهدف أيضاً إلى فهم الذكاء الإنساني لعمل برامج حاسب قادرة على محاكاة السلوك الإنساني المتسم بالذكاء، وتعني قدرة حل مسألة، أو اتخاذ قرار ، ويجد البرنامج نفسه الطريقة التي يجب أن تتبع لحل المسألة ، أو للتوصل إلى القرار بالرجوع إلى العديد من العمليات الاستدلالية المتنوعة التي غذي بها البرنامج<sup>(2)</sup>.

## 2- الأخلاق

الأخلاق هي تخصص شامل يتسم بقدر كبير من التفهم والتعددية. على الرغم من وجود دعوات ومزيد من المؤلفات المتعلقة بالمنظور الأخلاقي للتقاليد غير الغربية في مؤلفات أخلاقيات الذكاء الاصطناعي (على سبيل المثال، يطالب تصميم الأخلاق المتوازن [2017] التابع لمعهد مهندسي الكهرباء والإلكترونيات بضرورة دمج نظم أخلاقية غير غربية ويبرز بعض هذه النظم، بما في ذلك الأخلاق الناشئة في اليابان وأفريقيا)، إلا أن الحوار السائد حول أخلاقيات الذكاء الاصطناعي يوجد ضمن سياقات غربية وأوروبية وشمال أمريكية، مما ينبغي أيضاً قراءته من حيث وصولنا إلى الأدب الإنجليزي. على هذا النحو، مع الأخذ في الاعتبار التحدي المتزايد للأطر الأخلاقية التي تمثل فقط قانون الأدب الذي نكتبه ونبحث فيه، فإن «الأخلاقي» في أخلاقيات الذكاء الاصطناعي يمثل مفاهيم رئيسة من هذا القانون الفلسفي. على سبيل المثال، نفهم عمومًا الأخلاق

(1) Luxton, D. D.: An Introduction To Artificial Intelligence In Behavioral And Mental Health Care, Academic Press, United Kingdom, 2016, P2.

(2) عبد الحميد بسيوني الذكاء الاصطناعي والوكيل الذي ، دار الكتب العلمية للنشر والتوزيع ، القاهرة ،

على أنها الدراسة العقلانية والمنهجية لمعايير الصواب والخطأ، والأخلاقي كمصطلح يستخدم عادة لمفاهيم الخير والشر في الاستخدام الشائع للغة الإنجليزية<sup>(1)</sup>.

بالإضافة إلى الأخلاق والأخلاقي، نضمن موضوع القانون ضمن مفهومنا الأوسع لأخلاقيات الذكاء الاصطناعي. نقوم بذلك لأننا نحتفظ بالموقف الفلسفي بأن القانون والأخلاق مترابطان بشكل وثيق ولأننا نلاحظ أن مواضيع الحوكمة والمساءلة والشفافية تستند بشكل كبير إلى الاجتهاد القضائي. نعتبر القانون أنه القواعد والإرشادات المرموقة في نطاق قانوني معين. والأمر المهم هو أن القانون قابل للنفاذ، أي أن هناك نواة ملزمة (عادة من قبل السلطة التنفيذية للحكومة). في الأخلاق الفلسفية، نفكر في ذلك من حيث القيود الخارجية، مما يقارن بالحالات الداخلية الموصفه بمفهوم الفضيلة<sup>(2)</sup>.

هذه المفاهيم مرتبطة ترابطاً وثيقاً، وفي استخدام اللغة الطبيعية، يمكن استبدال هذه المصطلحات واعتبارها مترادفة في المجال العام<sup>(3)</sup>.

يعد نطاق الأخلاق الفلسفية شاسعاً، مع وجود مدارس وتقاليد دراسية مختلفة، تتميز كل منها بمجتمعها الخاص وتنوع داخلي كبير (مثل الوجودية، النفعية، الطبيعة، الأنانية/المنفعة، أخلاقيات التوطين/الحقوق، إلخ). هناك نهجان مهيمنان، يمكن قراءتهما على أنهما أساس القانون العام والقانون القاري، وهما «النفعية» (يشار إليها غالباً باسم النتيجة) والأخلاق القائمة على الحقوق. إضافة إلى هذه، وأقل انعكاساً في سياق القانون المعاصر، هي أخلاقيات الفضيلة.

<sup>(1)</sup>Kazim, E, E Koshiyama, A's. A high - Level overview of ethics. Patterns, 2 (g) 2021.p3.

<sup>(2)</sup>Ibid.p3.

هذه هي النظريات الأخلاقية الثلاث السائدة في الفلسفة الأكاديمية للأخلاق. وبذلك يتم استكشاف هذه الثلاثة أدناه<sup>(1)</sup>.

1 النفعية: صياغة المبادئ خلال النظر في عواقب الأفعال التي تنتج عن تلك القواعد، حيث يُسعى إلى تحقيق أقصى قدر من المتعة/تقليل الإزعاج. هناك تفسيرات كثيرة للنفعية (مثل النفعية من الفعل والنفعية من القاعدة) بالإضافة إلى تساؤلات حول كيفية فهم وقياس المصطلحات "المتعة" و"الإزعاج". وبغض النظر عن هذه المخاوف العلمية، المفهوم العام هو أن الأخلاق تتعلق بتقدير عواقب الأفعال. واحدة من المجالات التي تهيمن فيها هذه النظرية الأخلاقية هو تبرير سياسات الحكومة واتخاذ القرارات<sup>(2)</sup>.

الحقوق: هي الامتيازات التي يحصل عليها الفرد بناءً على انتمائه إلى فئة معينة. وفي هذا السياق الأخلاقي، يتم منح مجموعة من "الحقوق"، أي المزايا المشار إليها في التعريف أعلاه، للشخص فقط بوجوده في فئة معينة. حيث يتم فهم الفئة على أنها فئة عامة للهوية. وهناك نوعان رئيسيان من الحقوق، وهي حقوق الإنسان والحقوق المدنية. الحقوق الإنسانية هي الحقوق الممنوحة لأي شخص ينتمي إلى فئة "إنسان"، والحقوق المدنية هي الحقوق الممنوحة لأعضاء الجماعة السياسية في نطاق جغرافي معين. بينما يعتبر الحقوق الإنسانية لا تتجاوز ولا تنتهك وتعتبر أساسية ببساطة بوجود الإنسان، فإن الحقوق المدنية تُمنح لأعضاء المجتمع السياسي<sup>(3)</sup>.

الفضيلة: هي تطوير شخصية الفرد والأفعال التي تنتج كنتيجة لشخصية صالحة. الأخلاق الفضيلية (التي تُعرف أيضًا بالأخلاق الطبيعية) هي موقف كلاسيكي يستند إلى تعاليم أرسطو قبل التنوير. إنه نهج للأخلاق يؤكد على تطوير الشخصية - وهو مرتبط ارتباطاً وثيقاً بـ

(1)Ibid.p3.

(2)Ibid.p3.

"الكمال"، حيث يتطور الشخص مع مرور الوقت نحو نموذج مثالي للذات (يوصف في كثير من الأحيان بمصطلحات دينية كالتقرب من الإله أو الاندماج مع الإله). يُفهم الشخص الصالح في سياق القيم (أي الفضائل) مثل الصدق وضبط النفس والنزاهة والشجاعة والسخاء والعدل<sup>(1)</sup>.

### 3- تكنولوجيا الذكاء الاصطناعي وكرامة الانسان

هناك عدد لا يحصى من الاعتبارات الأخلاقية المصاحبة لتطوير الذكاء الاصطناعي واستخدامه وأثاره. وتتراوح هذه الآثار من الآثار المحتملة للذكاء الاصطناعي على حقوق الإنسان الأساسية للمواطنين داخل المجتمع إلى أمن البيانات المجمعة واستخدامها؛ من التحيز والتمييز المدمجين عن غير قصد في الذكاء الاصطناعي من قبل مجموعة متجانسة من المطورين، إلى نقص الوعي العام والفهم حول عواقب اختياراتهم واستخدامهم لأي ذكاء اصطناعي معين مما يؤدي إلى قرارات غير مستنيرة وضرر لاحق<sup>(2)</sup>.

يعتمد الذكاء الاصطناعي على الثورات السابقة في تكنولوجيا المعلومات والاتصالات والحوسبة، وبالتالي سيواجه عددًا من المشاكل الأخلاقية المماثلة. في حين أن التكنولوجيا قد تستخدم من أجل الخير، فمن المحتمل أن يساء استخدامها. قد نقرط في تجسيم الاصطناعي وإضفاء الطابع الإنساني عليه، مما يطمس الخطوط الفاصلة بين الإنسان والآلة<sup>(3)</sup>.

بمرور الوقت، تحدى الذكاء الاصطناعي الدماغ البشري في الألعاب التنافسية، مما رفع التوقعات إلى ما هو أبعد من التصور AlphaGo الذي طورته Google DeepMind مؤخرًا، وهو برنامج يجمع بين الشبكات العصبية العميقة التي يتم تدريبها خلال مزيج جديد من التعلم الخاضع للإشراف (الخبراء البشريون) والتعلم المعزز (اللعبة الذاتي) في مارس ٢٠١٦ هزم AlphaGo بطل

(1)Ibid.p4.

(2)European Parliament Study: The ethics of artificial intelligence: Issues and initiatives, 13 March 2020,p37.

العالم Sedol Lee 14: درجة، مما جذب موجة جديدة من الاهتمام العالمي بقدرة النظام على تجاوز ذكاء البشر بشكل تراكمي، تم إحراز تقدم كبير نحو الذكاء الاصطناعي باستخدام أنظمة التعلم الخاضعة للإشراف وفي الوقت الحالي تطور إلى أشكال غير خاضعة للإشراف، حيث أصبحت Alphao معلمها الخاص<sup>(1)</sup>.

تعد كرامة الإنسان التي لا يجوز المساس بها والتي ينبغي لكل إنسان أن يتمتع بها القاعدة التي تقوم عليها منظومة حقوق الإنسان والحريات الأساسية العالمية المترابطة والمتشابكة وغير القابلة للتجزئة والتصرف، ولذلك لا بد من احترام كرامة الإنسان وحقوق الإنسان وحمايتها وتعزيزها وفقاً لأحكام القانون الدولي<sup>(2)</sup>.

تقوم الكرامة الإنسانية على طريقة قائمة على احترام الحقوق الأساسية والتي قد تؤدي إلى تقييم التكنولوجيا المتقدمة، وإذا كان جميع البشر يمتلكون الكرامة الإنسانية، فمن الواجب عدم استغلالهم كأدوات، ومن هنا يمكن القول بقيدين واضحين هما<sup>(3)</sup>:

- عدم مشروعية أية ممارسة خاضعة بما في ذلك روبوتات الرعاية التي تحاكي العواطف والتعلق بالمستخدم؛ لأن خداع المستخدم المتعمد سيظل غير مشروع.
- الرعاية هي بالضرورة علنكية، ونظراً لأن الآلات تتفاعل فقط مع البشر بدلاً من الارتباط بهم، فلا يمكن استخدام أية آلة لتحل محل البشر في رعاية أي فرد وخاصة الأفراد الضعفاء.

(1) João Reis, Paula Espirito Santo, and Nuno Melão, (2020) Artificial Intelligence Theory in Service Management, in book: Exploring Service Science, 10th International Conference, IESS 2020, Porto, Portugal, 138.

(2) اليونيسكو: التوصية الخاصة بأخلاقيات الذكاء الاصطناعي، اليونيسكو، باريس، نوفمبر، 2022، ص.9.

(3) Custers, B., Villarrong, A.E.F. (2022): Law and artificial intelligence: Regulating al and Appling al in

وينبغي ألا يتعرض أي إنسان أو أية جماعة بشرية لأي شكل من أشكال الأذى أو الضرر أو الإذلال، سواء أكان ذلك جسدياً أم اقتصادياً أم اجتماعياً أم سياسياً أم ثقافياً أم نفسياً، خلال أية مرحلة من مراحل دورة حياة نظم الذكاء الاصطناعي، وينبغي السعي، طوال ضرورة حياة تنظم الذكاء الاصطناعي، إلى تسخير هذه النظم لتحسين نوعية حياة البشر، مع ترك تعريف "نوعية الحياة" للأفراد أو الجماعات ما دام التعريف الموضوع لهذا المفهوم لا ينطوي على أي انتهاك أو اعتداء يمس بحقوق الإنسان والحريات الأساسية أو بكرامة الإنسان<sup>(1)</sup>.

ينبغي الإقرار بأن وسائل تكنولوجيا الذكاء الاصطناعي، في حد ذاتها، لا تضمن بالضرورة ازدهار المجتمعات البشرية والبيئة والنظم الإيكولوجية، ويجب ألا تتجاوز أية عملية من العمليات المرتبطة بدورة حياة أي نظام من نظم الذكاء الاصطناعي، فضلاً عن ذلك، الأمور الضرورية لتحقيق الأغراض أو الأهداف المشروعة، وينبغي لهذه العمليات أن تكون ملائمة للسياق. وينبغي ضمان تطبيق إجراءات لتقييم المخاطر، وضمان اتخاذ تدابير لدفع الضرر المحتمل، عندما يكون من المحتمل إلحاق أي ضرر بالبشر، أو بحقوق الإنسان والحريات الأساسية، أو بالمجتمعات المحلية والمجتمع بوجه عام، أو بالبيئة والنظم الإيكولوجية<sup>(2)</sup>. يعني وضع خطة لقياس مخاطر السلامة المحتملة التي يتعرض لها أي طرف بسبب سوء استخدام التكنولوجيا، ولضمان ذلك يجب اتباع الآتي<sup>(3)</sup>:

أ- اتباع نهج وقاية من المخاطر يسير بطريقة تجعلها تسلك سلوكاً موثقاً.

(1) اليونيسكو، ص 8.

(2) المرجع نفسه، ص 9.

(3) Geovf Artificial intelligence (2002): Structures and strategies for complex problem solving,

ب - يجب أن يضمن مطوري الذكاء الاصطناعي ألا تتسبب هذه الأنظمة في وقوع ضرر غير مقصود وحدوث أية آثار سلبية.

ج - قدرة أنظمة الذكاء الاصطناعي على الصمود أمام أي هجمات أمنية .

د - وضع خطة بديلة من قبل الذكاء الاصطناعي في حالة حدوث أي مشكلات.

#### 4- التحديات الأخلاقية للذكاء الاصطناعي

##### أ- الشفافية

"أدى التطوير السريع لنشر تقنيات الذكاء الاصطناعي إلى الحاجة إلى إرشادات أخلاقية لضمان الاستخدام المسؤول والمفيد. وضعت منظمات ومؤسسات مختلفة مبادئ توجيهية لأخلاقيات الذكاء الاصطناعي لمعالجة مخاوف مثل الإنصاف والشفافية والمساءلة والخصوصية. ومع ذلك، هناك نقص في آليات التنسيق والإنفاذ عبر هذه المبادئ التوجيهية، مما يسلط الضوء على الحاجة إلى التعاون والتنسيق العالميين لمواجهة التحديات الأخلاقية التي يطرحها الذكاء الاصطناعي.<sup>(1)</sup>"

كثيراً ما تكون شفافية نظم الذكاء الاصطناعي وقابليتها للشرح شرطين أساسيين لا بد منهما لضمان احترام حقوق الإنسان والحريات الأساسية والمبادئ الأخلاقية وحمايتها وتعزيزها. ولا بد أيضاً من الشفافية لتطبيق أحكام النظم الوطنية والدولية المتعلقة بالمسؤولية تطبيقاً فاعلاً. ويمكن أن يحد الافتقار إلى الشفافية من إمكانية الطعن بفعالية في القرارات المتخذة بناء على نتائج مستمدة من الذكاء الاصطناعي، وقد يؤدي هذا الأمر بالتالي إلى انتهاك الحق في المحاكمة العادلة وفي التمتع الفعلي بسبل الإنصاف ، وكذلك إلى الحد من المجالات التي يمكن استخدام هذه النظم فيها استخداماً قانونياً<sup>(2)</sup>.

(1) Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.

(2) اليونيسكو: ص 11.

يسهم تعزيز الشفافية، من منظور اجتماعي في بناء مجتمعات أكثر سلاماً وعدلاً وديمقراطية وشمولاً، وتتيح الشفافية الاضطلاع بالمراقبة العامة التي يمكن أن تحد من الفساد والتمييز، ويمكن أن تساعد . على الكشف عن العواقب السيئة على حقوق الإنسان وتفاديها. وترمي الشفافية إلى توفير المعلومات الملائمة اللازمة للمتلقين المعنيين لتمكينهم من الفهم والتعزيز اللغة، ويمكن أن تتيح الشفافية، فيما نقص الظلم الذكاء الاصطناعي، تمكين الناس من فهم كيفية الاضطلاع بكل مرحلة من مراحل دورة حياة أي نظام من تعلم الذكاء الاصطناعي وفقاً لسياق ومقدار حساسية النظام المستخدم ويمكن أن تتيح الشفافية أيضاً الوقوف على العوامل المؤثرة في تنبؤ معين أو في قرار محدد وعلى وجود أو عدم وجود ضمانات ملائمة كالتدابير الخاصة بالسلامة والأمن أو بالعدالة والإنصاف).<sup>(1)</sup>

تكمن الشفافية من تفعيل المعاملة المتساوية بين الموظفين مثل المساواة في الأجور بين الجنسين والتي تؤدي إلى جذب عديد من المواهب، فالشفافية في مجال الذكاء الاصطناعي تعني معلومات عن مجموعة البيانات والمعلومات والعمليات والمخرجات التي يتم مشاركتها بين أصحاب المصلحة. فليست الشفافية هنا جودة أو أداء فقط ، ولكنها أيضاً الطريقة التي تتبادل بها المنظمة وتعزز فهم مكونات النظام ووظيفته بين مختلف الأطراف.<sup>(2)</sup>

ويشير مفهوم القابلية للشرح إلى إمكانية شرح نتائج نظم الذكاء الاصطناعي وجعلها مفهومة، وإمكانية تقديم معلومات عن تلك النتائج. ويشير مفهوم القابلية للشرح أيضاً، فيما يخص نظم الذكاء الاصطناعي، إلى إمكانية فهم المدخلات والمخرجات وطريقة عمل كل لبنة خوارزمية وكيفية إسهامها في نتائج هذه النظم، ولذلك ترتبط القابلية للشرح ارتباطاً وثيقاً

<sup>(1)</sup> المرجع نفسه، ص 12.

<sup>(2)</sup> Ammana Thb., Trustwo Th A.L. (2022): A business guider for navigating trust and ethics in al, Joha

بالشفافية، إذ ينبغي للنتائج وللعمليات الفرعية المؤدية إلى النتائج أن ترمي إلى أن تكون مفهومة وقابلة للتتبع وفقاً لسياق الاستخدام، وينبغي للجهات الفاعلة المعنية بالذكاء الاصطناعي أن تلتزم بالعمل على ضمان قابلية الخوارزميات التي يجري إعدادها للشرح.<sup>(1)</sup>

#### ب- القابلية للشرح والتفسير:

تشير الشفافية إلى مفاهيم ذات صلة مثل القابلية للشرح والتفسير والفهم والتواصل ويمكن أن تشير إلى مواقف أخلاقية مختلفة<sup>(2)</sup>.

يجب أن تراعي مؤسسات تشغيل الذكاء الاصطناعي اطلاع الأشخاص بلغة مفهومة على معلومات حول البيانات التي تلقاها النظام، وأنواع الخوارزميات المستخدمة، وأهم العوامل التي تؤثر في نتائج القرارات.

يجب لضمان تحقيق مبدأ الشرح أن تتيح مؤسسات الذكاء الاصطناعي اطلاع الأشخاص على تفسير عام يفسر كيف تعمل أنظمة الذكاء الاصطناعي الخاصة بهم، وأن تُسهل سبل التوصل السريع والمجاني لها بطريقة مبسطة وسهلة للمستخدمين<sup>(3)</sup>.

يعد التعلم الآلي أداة رائعة، ولكن من الصعب شرح العمليات الداخلية للتعلم الآلي، مما يعرف عادة بـ "الصندوق الأسود". يجعل "الصندوق الأسود" الخوارزميات غامضة حتى بالنسبة لمبتكريها. يحد من قدرة الناس على فهم التكنولوجيا، ويؤدي إلى تواجد اختلافات معلوماتية كبيرة

<sup>(1)</sup>اليونسكو:ص12.

<sup>(2)</sup>Stahi, B. C. (2021): Artificial intelligence for a Better Future: An ecosystem perspective on the ethics of Alan emerging digital technologies, spring, Switzerland, p. 28.

<sup>(3)</sup>Wong, A. (2021): Artificial intelligence for knowledge management, stifflw6 international

بين خبراء الذكاء الاصطناعي والمستخدمين، ويعيق الثقة البشرية في التكنولوجيا ووكلاء الذكاء الاصطناعي. الثقة أمر بالغ الأهمية في جميع أنواع العلاقات وهو سبب أساسي للقبول<sup>(1)</sup>. علاوة على ذلك، قد يتطور الذكاء الاصطناعي بدون رقابة وإرشاد من البشر بسبب "الصندوق الأسود" الذي لا يستطيع البشر تفسيره. على سبيل المثال، في عام 2017، أغلق فيسبوك محرك الذكاء الاصطناعي الخاص به لأنه اكتشف أن الذكاء الاصطناعي قد أنشأ لغة فريدة من نوعها ولم يستطع البشر فهمها. مسألة ما إذا كان يمكن للبشر التحكم في وكلاء الذكاء الاصطناعي هي قلق كبير. يفضل البشر أن يقوم وكلاء الذكاء الاصطناعي بالقيام دائماً بما نريده منهم بالضبط. ومع ذلك، عندما ننظر إلى هذه المشكلة من منظور آخر، إذا قمنا بمعاملة وكلاء الذكاء الاصطناعي بشكل أخلاقي، هل من الأخلاقي أن نتحكم في الإجراءات التي يتخذونها وكيفية اتخاذهم للقرارات؟<sup>(2)</sup>

### ج- أمان البيانات والخصوصية

يُعد التمتع بالخصوصية حقاً ضرورياً لصون كرامة الإنسان والذود عن استقلاليته وحماية أعماله، ولذلك يجب احترام الخصوصية وصونها وتعزيزها طوال دورة حياة نظم الذكاء الاصطناعي. ومن المهم جمع البيانات المتعلقة بنظم الذكاء الاصطناعي واستخدامها وتبادلها وتشاطرها وحفظها وحذفها بطريقة تتوافق مع القانون الدولي، وكذلك مع القيم والمبادئ المنصوص عليها في هذه التوصية، وفي ظل التقيد بالأطر القانونية الوطنية والإقليمية والدولية المتعلقة بهذا الموضوع<sup>(3)</sup>.

(1) siau, K. & Wang, w.p79.

(2) Ibid.p80.

(3) اليونيسكو: ص 10.

يعتمد تطوير وكلاء الذكاء الاصطناعي بشكل كبير على كمية ضخمة من البيانات، بما في ذلك البيانات الشخصية والبيانات الخاصة. تقريبًا جميع مجالات التطبيق التي يحقق فيها التعلم العميق نجاحًا، مثل Siri من Apple و Google Home، تتمتع بوصول إلى كميات كبيرة من البيانات. مع زيادة توليد البيانات في المجتمعات والأعمال، يزداد احتمال استغلال هذه البيانات. على سبيل المثال، يحتوي سجل الصحة دائمًا على معلومات حساسة، وإذا لم يتم حمايتها بشكل كافٍ، يمكن أن تحصل مؤسسة غير مشروعة على وصول إلى تلك المعلومات وتضر بالمرضى من الناحية الشخصية والمالية. لذا، يجب إدارة البيانات بشكل صحيح لمنع سوء الاستخدام الخبيث. وتعتمد أهمية حفظ البيانات بشكل آمن، كتفصيلة الإجراءات المتخذة فيما يتعلق بالبيانات وتسجيلها، ومن يتولى القيام بتسجيل العمليات، ومن يمكنه الوصول إلى البيانات والسجلات<sup>(1)</sup>.

يجب على أنظمة الذكاء الاصطناعي احترام الخصوصية، وأن تستخدم الحد الأدنى من التطفل الضروري، وكذلك عدم نشر المراقبة أو غيرها من التقنيات المعتمدة على الذكاء الاصطناعي بطريقة تنتهك الخصوصية وكرامة الإنسان، ولضمان تحقيق مبدأ الخصوصية يمكن اتباع الآتي<sup>(2)</sup>:

أ- تقييم الأشخاص المخولين بالوصول إلى البيانات.

ب - تحديد مسؤوليات وتحديد أدوار بعينها لمسؤول حماية البيانات.

ج- منع الضرر الذي يلحق بالخصوصية خلال حوكمة البيانات بشكل دائم.

د- السماح فقط للموظفين المؤهلين أصحاب الكفاءة للوصول إلى بيانات الأفراد

والاطلاع عليها.

(1) siau, k. & wang, w. Artificial intelligence ethics: ethics of AI and ethical AL. Journal of Database Management (JDM) & 1(2). 2020.

(2) نوم ستوينز: ما بعد المعلومات والتاريخ الطبيعي للذكاء، ترجمة: مصطفى إبراهيم فهمي، المجلس الأعلى

للثقافة القاهرة ٢٠٠٥، ص ١٢٣.

ينبغي السعي إلى تفادي الأضرار غير المقصودة (المخاطر المتعلقة بالسلامة) ومواطن الضعف أمام الهجمات (المخاطر المتعلقة بالأمن) وكذلك إلى التصدي لها طوال دورة حياة نظم الذكاء الاصطناعي من أجل ضمان سلامة وأمن البشر والبيئة والنظم الإلكترونية، وتطلب التمكن من ضمان سلامة وأمن نظم الذكاء الاصطناعي وضع أطر مستدامة للانتفاع بالبيانات تتيح صون الخصوصية وتحسن كيفية تدريب نماذج الذكاء الاصطناعي وكيفية الموافقة عليها عن طريق استخدام بيانات جيدة لهذا الغرض<sup>(1)</sup>.

د- المسؤولية والمساءلة تعد المسألة هي التصرف بنزاهة وتحديد إسناد المسؤولية والالتزام القانوني، والتأكيد على الحاجة إلى الإبلاغ عن الضرر المحتمل والسعي لتعزيز التنوع وإدخال الأخلاق في تعليم العلوم والتكنولوجيا، ثم تحديد جهات فاعلة مسؤولة عن سلوكيات الذكاء الاصطناعي وقراراته<sup>(2)</sup>.

عندما يفشل وكيل الذكاء الاصطناعي في مهمة محددة، من ينبغي أن يتحمل المسؤولية. قد يؤدي هذا إلى ما يُشار إليه بـ "مشكلة الأيدي الكثيرة". عند استخدام وكيل الذكاء الاصطناعي، قد تسبب البرمجيات أو البيانات المدخلة أو العملية غير السليمة أو عوامل أخرى في حدوث نتيجة غير مرغوب فيها. من يجب أن يكون الجهة المسؤولة عن النتيجة غير المرغوب فيها - المبرمج، مالك البيانات، أو المستخدم النهائيون؟<sup>(3)</sup>

ينبغي للجهات الفاعلة المعنية بالذكاء الاصطناعي وللدول الأعضاء أن تحترم حقوق الإنسان والحريات الأساسية وأن تزدود عنها وتعززها، وينبغي لها أيضاً أن تعمل على تعزيز حماية

(1) اليونسكو: ص 9.

(2) Park, A. F. (2022): Mapping the public voice of development natural language processing of social media text data, Asian development bank, p.125.

البيئة والنظم الإيكولوجية، وأن تتحمل المسؤولية الأخلاقية والقانونية التي تقع على عاتقها، وفقاً لأحكام القانون الوطني وأحكام القانون الدولي، ولا سيما الأحكام الخاصة بواجبات الدول الأعضاء المتعلقة بحقوق الإنسان، ومنها الواجبات المرتبطة بالجهات الفاعلة المعنية بالذكاء الاصطناعي الموجودة فعلاً داخل أراضيها والخاضعة لسيطرتها الفعلية، وكذلك وفقاً للإرشادات الخاصة بالأخلاقيات طوال دورة حياة نظم الذكاء الاصطناعي، وينبغي أن يتسنى دائماً عزو المسؤولية الأخلاقية عن القرارات والإجراءات القائمة بأي شكل من الأشكال على أي نظام من نظم الذكاء الاصطناعي إلى الجهات الفاعلة المعنية بالذكاء الاصطناعي في نهاية المطاف وفقاً لدور كل جهة منها في دورة حياة نظام الذكاء الاصطناعي<sup>(1)</sup>.

"الهدف النهائي لأخلاقيات الآلة هو إنشاء آلة تتبع بنفسها مبدأً أخلاقياً مثاليًا أو مجموعة من المبادئ". من الناحية النظرية سهل، ولكن من الناحية العملية صعب، صياغة المبادئ الأخلاقية لوكلاء الذكاء الاصطناعي. بدون معايير أخلاقية شاملة وغير متحيزة، كيف يمكن للبشر تدريب الآلة لتكون أخلاقية؟ وعلاوة على ذلك، كيف يمكننا التأكد من أن الآلات الذكية تفهم المعايير الأخلاقية بنفس الطريقة التي نفهمها نحن؟ على سبيل المثال، إذا قمنا ببرمجة الروبوتات لعدم إلحاق الضرر، فيجب علينا أولاً التأكد من أن الروبوتات تفهم ما هو الضرر. ينتج عن هذا مشكلة أخرى - ما هي المعايير الأخلاقية للضرر؟ وثمة حاجة إلى مستوى عالٍ من الأخلاقيات.<sup>(2)</sup>

##### 5- الأثر الاجتماعي للقضايا الأخلاقية

بدون تدريب في قوانين حقوق الإنسان، يمكن لمهندسي البرمجيات كتابة رموز تنتهك حقوق الإنسان الرئيسية دون حتى معرفة ذلك. من المهم تعليم قوانين حقوق الإنسان لمهندسي

(1) اليونسكو: ص12.

البرمجيات. ضمان الخصوصية حسب التصميم مهم وأكثر كفاءة من حيث التكلفة من البدائل. يمكن أن تساعد المعرفة الأفضل بقوانين حقوق الإنسان مصممي ومهندسي الذكاء الاصطناعي على القضاء على التمييز وانتهاك قضايا الخصوصية في الذكاء الاصطناعي أو على الأقل التخفيف من حدتها<sup>(1)</sup>.

#### أ- التأثير المترتب على الأتمتة واستبدال الوظائف

لا يزال الجدل حول ما إذا كان الذكاء الاصطناعي والصناعة ستخلق وظائف جديدة أم ستقضي على بعض الوظائف مستمرًا. تتوارد القصص عن عمال المصانع الذين يتم استبدالهم بأنظمة وروبوتات آلية. يقول البعض إن الذكاء الاصطناعي سيخلق أيضًا ملايين من الوظائف الجديدة، وكثير من هذه الوظائف غير موجودة في الوقت الحالي. ومع ذلك، لا يزال هناك قلق بشأن اضطرابات القوى العاملة المستقبلية في عصر الذكاء الاصطناعي، مثل التعاون بين البشر ووكلاء الذكاء الاصطناعي. سيتعرض سوق العمل لاضطراب وتحول مع التطورات المرتبطة بالذكاء الاصطناعي. ما ليس واضحًا تمامًا هو سرعة ونطاق التغيير. كانت النية الأصلية لتطوير التكنولوجيا هي مساعدة البشر وتحسين حياتهم. إذا تسببت الأتمتة والذكاء الاصطناعي في استبدال وظائف كبيرة وبطالة، هل يجب علينا الاحتفاظ بوتيرة التطور التكنولوجي السريع؟ وأيضًا، كيف يمكننا حماية حقوق الإنسان ورفاهية الإنسان في حين نواكب التطورات والثورات التكنولوجية السريعة؟<sup>(2)</sup>

#### ب- التوفر والوصولية

تعد الوصولية، كمبدأ أخلاقي، هي قدرة الأنظمة والمنتجات والخدمات على أن تكون متاحة ومناسبة لجميع الأشخاص، بما في ذلك كبار السن وذوي الإعاقة. نظرًا لتعقيد التكنولوجيا الجديدة والمنتجات عالية التقنية، بالإضافة إلى تزايد السكان المسنين في بعض البلدان، فإن

(1)Ibid.p81.

وصول التكنولوجيا الجديدة سيؤثر مباشرة على رفاهية الإنسان. يجب أن يعود تطوير التكنولوجيا لصالح البشر. ولكن إذا كانت الفوائد محصورة في جزء من الناس، هل هذا أمر أخلاقي وعادل؟ يجب أن يُولى اهتمام لتطوير أنظمة ومنتجات وخدمات يمكن الوصول إليها من قبل الجميع، ويجب توزيع فوائد التكنولوجيا المتقدمة بشكل عادل على الجميع<sup>(1)</sup>.

### ج- الديمقراطية وحقوق المدنيين

يؤدي الذكاء الاصطناعي غير الأخلاقي إلى تشتت الحقيقة وفقدان الثقة الناتجة عنه، وفقدان الدعم المجتمعي لتكنولوجيا الذكاء الاصطناعي. وبما أن الديمقراطيات تعاني وتُعزز التحيزات الهيكلية، فإن ممارسة حرية حقوق المدنيين لم تعد متاحة بشكل متساوٍ للجميع. يجب أن تأخذ الأخلاقيات الخاصة بالذكاء الاصطناعي في الاعتبار الديمقراطية وحقوق المدنيين. سيساعد حل هذه المسائل بشكل صحيح في وضع إطار لبناء ذكاء اصطناعي أخلاقي<sup>(2)</sup>.

### د- التلاعب

تتجاوز المسائل الأخلاقية للذكاء الاصطناعي في المراقبة مجرد تجميع البيانات وتوجيه الاهتمام. تشمل استخدام المعلومات لأغراض مشكوك فيها. واحدة من هذه الأغراض هي التلاعب بالسلوك، سواء عبر الإنترنت أو في الحياة الواقعية، وغالبًا ما تستهدف أموال المستخدم. بالطبع، الجهود المبذولة للتلاعب بالسلوك بطريقة تضعف الاختيار الرشيد المستقل قديمة، ولكنها قد تكتسب جودة جديدة عند استخدام أنظمة الذكاء الاصطناعي. نظرًا للتفاعل المكثف بين المستخدمين وأنظمة البيانات والمعرفة العميقة حول الأفراد التي توفرها، نحن ضعفاء أمام "الدفعات" والتلاعب والخداع. باستخدام البيانات السابقة الكافية، يمكن استخدام خوارزميات

(<sup>1</sup>)Ibid.p82.

لاستهداف الأفراد أو الفرق الصغيرة بنفس نوع الإدخال الذي من المرجح أن يؤثر على هؤلاء الأفراد بشكل خاص<sup>(1)</sup>.

سيستخدم العديد من الإعلانين والتسويقيين والبائعين عبر الإنترنت أي وسيلة قانونية تتاح لهم، بما في ذلك استغلال التحيزات السلوكية والخداع وإثارة الإدمان - على سبيل المثال خلال "نماذج مظلمة" على صفحات الويب أو في الألعاب. هذا التلاعب هو نموذج الأعمال الأساسي في العديد من صناعات القمار والألعاب، ولكنه ينتشر، على سبيل المثال، إلى شركات الطيران ذات التكلفة المنخفضة. القمار وبيع المواد الادمانية يخضعان لتنظيم صارم، ولكن التلاعب والإدمان عبر الإنترنت ليسا كذلك. يصبح التلاعب بالسلوك على الإنترنت نموذج أعمال أساسي للإنترنت<sup>(2)</sup>.

بالإضافة إلى ذلك، أصبحت وسائل التواصل الاجتماعي الآن المواقع الأساسية للعداوة السياسية. يمكن استخدام هذا التأثير لتوجيه سلوك الاقتراع، كما حدث في "فضيحة" فيسبوك وكامبريدج أناليتيكا، وإذا نجح ذلك، فقد يضر بحرية الفرد.

تقنيات التلاعب بالذكاء الاصطناعي المحسنة تجعل ما كان يومًا ما دليلًا موثوقًا به دليلًا غير موثوق به وقد حدث ذلك بالفعل في الصور الرقمية وتسجيلات الصوت ومقاطع الفيديو ... وسيكون من السهولة قريبًا إنشاء (وليس تعديل) نصوص وصور ومواد فيديو "عميقة مزيفة" بأي محتوى مطلوب. قريبًا، ستمكّن أيضًا من إنشاء تفاعل متطور في الوقت الحقيقي مع الأشخاص عبر الرسائل النصية والهاتف أو الفيديو. لذا، لا يمكننا الوثوق في التفاعل الرقمي، في حين أننا في نفس الوقت نعتمد بشكل متزايد على هذا التفاعل<sup>(3)</sup>.

(1) Müller, vin Cent.C. P.6.

(2) Ibid.P.6.

تواجه السياسة في هذا المجال من الخصوصية والتلاعب صعوبة في مواكبة التطورات التقنية والاجتماعية. تتعرض الحريات المدنية وحماية حقوق الفرد لضغوط شديدة من تجار لضغط الأعمال والخدمات السرية ووكالات الدولة الأخرى التي تعتمد على المراقبة. لقد تناقشت الحماية القانونية الفعلية من المراقبة والتلاعب بشكل جذري مقارنةً بالعصر الرقمي السابق (الرسائل والهاتف التناظري والمحادثة الشفهية). في حين أن لائحة حماية البيانات العامة في الاتحاد الأوروبي (GDPR 2016) قد قوّت حماية الخصوصية إلى حد ما، إلا أن الولايات المتحدة والصين يفضلان النمو دون تنظيم أقل، على الأرجح على أمل أن يوفر ذلك ميزة تنافسية.<sup>(1)</sup>

#### هـ- التعقيم

يعد التعقيم والتحيز قضيتين مركبتين في التحديات الأخلاقية التي تواجه الذكاء الاصطناعي، فقد يكون من الصعب أحياناً على الشخص أن يعرف كيف وصل النظام إلى هذه النتيجة، أي أن النظام غامض بالنسبة لذلك الشخص، فإذا كان النظام يتضمن تعلقاً آلياً فسيكون عادة غامضاً بالنسبة للخبير الذي لن يعرف كيف يتم تحديد نمط ما، وهكذا يزداد التحيز في أنظمة الذكاء الاصطناعي بسبب هذا التعقيم.<sup>(2)</sup>

التعقيم والتحيز أيضاً هما قضيتان مركبتان في ما يسمى الآن أحياناً "أخلاقيات البيانات" أو "أخلاقيات البيانات الضخمة". تعمل أنظمة دعم القرار الآلي و"التحليل التنبؤي" التي تستخدم الذكاء الاصطناعي على البيانات وتنتج قراراً كنتيجة "تحليل البيانات". يمكن أن تتراوح هذه النتائج من الأمور الثانوية إلى الأمور ذات الأهمية الكبيرة: "هذا المطعم يتناسب مع تفضيلاتك"، "المريض في صورة الأشعة السينية هذه قد اكتمل نمو العظام لديه"، "تم رفض طلب بطاقة الائتمان"، "سيتم منح العضو المانح لمريض آخر"، "تم رفض كفالة"، أو "تم التعرف على

(1)Ibid.p7.

الهدف والمشاركة معه". يُستخدم التحليل البياني في العديد من المجالات مثل الأعمال والرعاية الصحية وغيرها للتنبؤ بالتطورات المستقبلية - حيث يكون التنبؤ أسهل مع الذكاء الاصطناعي وسيصبح أيضًا سلعة أرخص<sup>(1)</sup>.

يبدو أن نظم الذكاء الاصطناعي لدعم القرار التلقائي جزء من هيكل سلطة يكون فيه من المستحيل في كثير من الأحيان على الشخص المتأثر معرفة كيف وصل النظام إلى هذه النتيجة، أي أن النظام يكون "عتمًا" بالنسبة لذلك الشخص. إذا تضمن النظام التعلم الآلي، فإنه في العادة سيكون عتمًا حتى بالنسبة للخبير الذي لن يعرف كيف تم تحديد نمط معين، أو حتى ما هو هذا النمط. يتفاقم التحيز في أنظمة القرار ومجموعات البيانات بسبب هذه العتامة. لذا، على الأقل في الحالات التي يرغب في إزالة التحيز، يجب معالجة تحليل العتامة والتحيز، ويجب أن تتعامل الاستجابة السياسية مع كلا القضيتين معًا<sup>(2)</sup>.

---

(1) Müller, vin Cent.C. P7.

## والتحيز

تعمل أنظمة الذكاء الاصطناعي ودعم القرار الآلية والتحليلات التنبؤية على البيانات ، وتنتج قرارًا على شكل مخرجات ، قد تتراوح هذه النتيجة من عادية إلى ذات أهمية كبيرة. ينشأ التحيز عندما يتخذ الذكاء الاصطناعي قرارات بشكل أكثر دقة، فقد تكون القرارات غير عادلة للأفراد أو لمجموعات ما، و رغم وجود مشاكل التحيز في المجتمع من قبل إلا أن القلق من توسيع نطاقها قد ازداد في ظل الذكاء الاصطناعي<sup>(1)</sup>.

غالبًا ما يظهر التحيز عندما يتم اتخاذ أحكام غير عادلة بسبب تأثير الشخص القاضي بسمة غير ذات صلة بالمسألة المطروحة، وعادةً ما يكون التحيز التعليمي سمة معرفية تمت تعلمها للشخص، وغالبًا ما لا يتم توضيحها. قد لا يكون الشخص المعني على علم بوجود هذا التحيز - حتى قد يكون معارضًا بصراحة وبشكل صريح للتحيز الذي تم اكتشافه لديه) على سبيل المثال ، من خلال التنشيط المسبق. بصرف النظر عن الظاهرة الاجتماعية للتحيز المتعلم، يكون النظام الإدراكي البشري عرضة عمومًا لأنواع مختلفة من "التحيزات الإدراكية"، على سبيل المثال "تحيز التأكيد"، حيث يميل البشر إلى تفسير المعلومات على أنها تؤكد ما يعتقدونه بالفعل. يقال في كثير من الأحيان أن هذا النوع الثاني من التحيز يعرقل الأداء في الحكم الرشيد - على الرغم من أن بعض التحيزات الإدراكية تولد ميزة تطويرية ، على سبيل المثال الاستخدام الاقتصادي للموارد للحكم الحدسي. هناك سؤال ما إذا كانت أنظمة الذكاء الاصطناعي يمكن أن تكون لديها تحيزات إدراكية من هذا القبيل أو يجب أن يكون لديها<sup>(2)</sup>.

(1)Coeckelber, A.M. (2020): AI Ethics, Massa Chusets institute of technology, USA, p.125.

التحيز الثالث الموجود في البيانات هو التحيز النظامي، عندما يظهر خطأً نظامي في البيانات، على سبيل المثال إحدى أنواع "التحيز الإحصائي". بدقة، فإن أي مجموعة بيانات معينة ستكون غير متحيزة لنوع واحد فقط من المشكلة، لذلك فإن إنشاء مجموعة بيانات تنطوي على خطر أن يتم استخدامها لنوع مختلف من المشكلة، ومن ثم يتبين أنها متحيزة لهذا النوع. بالتالي، فإن التعلم الآلي بناءً على مثل هذه البيانات لن يتعرف فقط على التحيز، بل سيُعتبر تاريخياً ويتم تشفير التحيز. تم اكتشاف هذا التحيز التاريخي في نظام غريبله التوظيف الآلي في أمازون (الذي تم إيقافه في أوائل عام 2017) الذي قام بالتمييز ضد النساء - على الأرجح بسبب تاريخ الشركة في التمييز ضد النساء في التوظيف. المشكلة مع هذه الأنظمة هي التحيز الذي تحمله بالإضافة إلى الثقة المفرطة من البشر في هذه الأنظمة<sup>(1)</sup>.

هناك جهود تقنية كبيرة لاكتشاف وإزالة التحيز من أنظمة الذكاء الاصطناعي، ولكن من النادر أن تكون هذه الجهود في مراحل مبكرة. يبدو أن التحسينات التكنولوجية لها حدودها في أنها تحتاج إلى مفهوم رياضي للعدالة، والأمر ليس سهلاً<sup>(2)</sup>.

ينبغي للجهات الفاعلة المعنية بالذكاء الاصطناعي أن تسعى إلى تعزيز العدالة الاجتماعية، وكذلك إلى ضمان الإنصاف والحيولة دون ممارسة أي نوع من أنواع التمييز، وفقاً للقانون الدولي. ويتطلب تحقيق العدالة والإنصاف الأخذ بنهج شامل لضمان توفير منافع وسائل تكنولوجيا الذكاء الاصطناعي للجميع وضمان تمكين الجميع من الحصول عليها والتمتع بها، مع مراعاة الاحتياجات الخاصة لمختلف الفئات العمرية والمنظومات الثقافية والجماعات اللغوية والمعوقين والفتيات والنساء والفئات السكانية المحرومة والمهمشة والضعفاء أو المستضعفين<sup>(3)</sup>.

(1) Müller, vin Cent.C. P9.

(2) Ibid.p9.

(3) اليونسكو، ص11.

ينبغي للدول الأعضاء أن تسعى إلى التصدي للفجوات الرقمية، وكذلك إلى ضمان ارتفاع الجميع بالذكاء الاصطناعي ومشاركة الجميع في تطويره. وينبغي للدول الأعضاء، على الصعيد الوطني، أن تعزز العدالة والإنصاف بين المناطق الريفية والحضرية، وكذلك بين كافة الأشخاص بغض النظر عن العرق أو اللون أو النسب، أو الجنس أو السن، أو اللغة أو الدين، أو الآراء السياسية، أو الأصل القومي أو الاثني أو الاجتماعي، أو الوضع الاقتصادي أو الاجتماعي المرتبط بالمولد، أو الإعاقة، أو أي سبب من أسباب التمييز الأخرى، فيما يخص الانتفاع بنظم الذكاء الاصطناعي والمشاركة في العمليات المرتبطة بدورة حياة هذه النظم. وتقع على عاتق أكثر البلدان تقدماً من الناحية التكنولوجية، على الصعيد الدولي، مسؤولية التضامن مع أقل البلدان تقدماً لضمان تشاطر منافع وسائل تكنولوجيا الذكاء الاصطناعي لكي يسهم تمكين أقل البلدان تقدماً من الانتفاع بنظم الذكاء الاصطناعي والمشاركة في العمليات المرتبطة بدورة حياة هذه النظم في إيجاد نظام عالمي أكثر عدلاً وإنصافاً فيما يخص المعلومات والاتصالات والثقافة والتربية والتعليم والبحث والاستقرار الاجتماعي والاقتصادي والسياسي<sup>(1)</sup>.

### ز\_ الخداع والروبوتات

تفاعل الإنسان مع الروبوتات (HRI) هو مجال أكاديمي مستقل يولي اهتماماً كبيراً للقضايا الأخلاقية وديناميات الإدراك من الجانبين والمصالح المختلفة الموجودة وتعقيد السياق الاجتماعي، بما في ذلك التعاون<sup>(2)</sup>.

بينما يمكن استخدام الذكاء الاصطناعي لتلاعب البشر للإيمان والقيام بأشياء، يمكن أيضاً استخدامه لقيادة الروبوتات التي قد تكون مشكلة إذا تضمنت عملياتها أو مظهرها خداعاً، أو تهديداً لكرامة الإنسان، أو انتهاكاً لمتطلبات كانت الخاصة بـ "احترام الإنسانية". يهتم البشر

(1) المرجع نفسه، ص.11.

بسهولة بخصائص عقلية للأشياء ويتعاطفون معها، خاصة عندما يكون المظهر الخارجي لهذه الأشياء مشابهًا لتلك المتواجدة في الكائنات الحية. يمكن استخدام ذلك لخداع البشر (أو الحيوانات) لإسناد أهمية فكرية أكبر أو حتى عاطفية للروبوتات أو أنظمة الذكاء الاصطناعي مما يستحقه. بعض أجزاء الروبوتات ذات الهيئة البشرية مشكلة في هذا الصدد (على سبيل المثال الروبوتات التي يتم التحكم فيها عن بعد لهيروشي إيشيغورو)، وهناك حالات كانت خادعة بشكل واضح لأغراض العلاقات العامة (على سبيل المثال، "صوفيا" من إنتاج شركة هانسون روبوتيكس). بالطبع، تنطبق بعض القيود الأساسية إلى حد ما لأخلاقيات الأعمال والقانون على الروبوتات أيضًا: سلامة المنتج ومسؤوليته، أو عدم الخداع في الإعلان.

مع ذلك، هناك حالات يكون فيها للتفاعل بين الإنسان والإنسان جوانب تبدو بشرية على وجه التحديد بطرق ربما لا يمكن استبدالها بالروبوتات: الرعاية والحب والجنس<sup>(1)</sup>.

---

(<sup>1</sup>)Ibid.p8.