

دور الأخلاق في الحد من التحيز في الذكاء الاصطناعي

The role of ethics in reducing bias in artificial intelligence

لوالبية فوزي

جامعة تامنغست، 11000، الجزائر، Loualbiafadi@yahoo.fr

تاريخ النشر: 2025/07/16

تاريخ القبول 28 / 10 / 2024

تاريخ الاستلام: 2023/08/27

ملخص:

هدفت هذه الدراسة إلى معرفة دور الأخلاق في الحد من التحيز في الذكاء الاصطناعي، من خلال استعراض مجموعة من المفاهيم حول الذكاء الاصطناعي مفهومه وأنواعه، ومفهوم التحيز في الذكاء الاصطناعي وأنواعه، وكذا مدى مساهمة الأخلاق في الحد منه.

أظهرت النتائج أن أنظمة الذكاء الاصطناعي تعمل على مساعدة الأفراد في اتخاذ القرارات التي تؤثر بشكل مباشر على حياتهم وحياة الآخرين، إلا أنها تشكل تهديدا نتيجة للاستخدام السلبي، مما يستدعي أخلاق الذكاء الاصطناعي لمواجهة التحيز في إدخال البيانات، من خلال إخضاعه إلى مجموعة من المبادئ المتعارف عليها، رغم أنها عملية معقدة بسبب أن البشر في حد ذاتهم لا يتفقون على قواعد أخلاقية واحدة، فما بالك بتعليمها للنماذج الذكاء الاصطناعي من أجل برمجتها بها.

كلمات مفتاحية: أخلاق، تحيز، ذكاء الاصطناعي، مبادئ تعارف عليها، استخدام سلبي.

Abstract:

This study aimed to know the role of ethics in reducing bias in artificial intelligence, by reviewing a set of concepts about artificial intelligence, its concept and types, the concept of bias in artificial intelligence and its types, as well as the extent to which ethics contribute to reducing it.

The results showed that artificial intelligence systems work to help individuals make decisions that directly affect their lives and the lives of others, but they pose a threat as a result of negative use, Which necessitates the ethics of artificial intelligence to confront bias in data entry, by subjecting it to a set of generally accepted principles, although it is a complex process because humans themselves do not agree on a single ethical rule, let alone teaching it to artificial intelligence models in order to program it with it.

Keywords: ethics, bias, artificial intelligence, generally accepted principles, negative use.

1. مقدمة :

يولي الأفراد اهتماما متسارعا في استخدام أنظمة الذكاء الاصطناعي لمحاكاة العمل البشري للأداء مهامها لتحسين نفسها بنفسها من خلال البيانات التي يتم جمعها، إن اللجوء الأفراد إلى الاهتمام الكبير بنماذج الذكاء الاصطناعي هو محاولة لمساعدتهم على اتخاذ القرارات المناسبة الذين يسعون إلى تحقيقها من خلال دراسة البيانات والمعلومات المتعلقة بالإشكالية المطروحة، حيث أن ضخامة المعلومات المتوفرة يجعل العقل البشري لا يستطيع أن يحصرها ودراستها بصفة دقيقة ، مما قد يؤدي إلى عدم حل الإشكالية بطرق منطقية وعدم اتخاذ القرار المناسب.

إن تقنيات الذكاء الاصطناعي ساعدت الأفراد على قيام بمهامه في حل الإشكاليات المطروحة أمامهم، إلا هناك جانب آخر مهم، وهو استخدام الذكاء الاصطناعي في اتجاه الأهداف السلبية، من خلال عملية التحيز في عملية تغذيته، مما يؤدي إلى وجود مجموعة من المخاوف للاستخدامات السلبية لتقنيات الذكاء الاصطناعي التي يمكن أن تسهم في الحد من تأثيراته السلبية .

انطلاقا مما سبق يمكن طرح الإشكالية التالية:

ما دور الأخلاق في الحد من التحيز في الذكاء الاصطناعي ؟

أهمية الدراسة

تكمن أهمية الدراسة في تناولها لموضوع الذكاء الاصطناعي، وهو أحد المشكلات التي تهدد الأفراد في استخدام تطبيقاته في شتى المجالات، من خلال المخاوف والتهديدات التي تتعارض مع ثقافة المجتمع خلال استخدامها، مما يستوجب وضع مجموعة من الضوابط القانونية والأخلاقية خلال عملية تأطير نماذج الذكاء الاصطناعي.

أهداف الدراسة

تهدف الدراسة إلى معرفة دور الأخلاق في الحد من تأثيرات التحيز في بناء نماذج الذكاء الاصطناعي من خلال:

- إلقاء الضوء على المفاهيم المختلفة للذكاء الاصطناعي وأهدافه.
- تحليل مفهوم التحيز في الذكاء الاصطناعي، أنواعه، أساليبه.
- إبراز العلاقة بين الأخلاق والتحيز في الذكاء الاصطناعي.

منهج الدراسة:

للإجابة على الإشكالية ، سوف نعتمد على المنهج الوصفي لتحليل الإطار النظري لموضوع التحيز في الذكاء الاصطناعي ، و دراسة الأخلاق في الحد منه.

هيكل الدراسة:

للإجابة على إشكالية هذه الدراسة، سنقوم في هذه الدراسة بالتطرق إلى:

المبحث الأول: الإطار النظري للتحيز في الذكاء الاصطناعي

المبحث الثاني: الأخلاق و التحيز في الذكاء الاصطناعي

المبحث الأول: الإطار النظري للتحيز في الذكاء الاصطناعي

المطلب الأول: مفهوم الذكاء الاصطناعي

الفرع الأول: نشأة الذكاء الاصطناعي

استخدم مصطلح الذكاء الاصطناعي أول مرة في سنوات الخمسينيات خلال مؤتمر جامعة " دارتموث " (Dartmouth) الأمريكية المتعلق بالذكاء الاصطناعي وذلك عام 1956 من خلال الباحث جون مكارثي، ومنذ ذلك الحين يقوم الباحثون والمبتكرون بتقديم براءات الاختراع والبحوث العلمية في مجال الذكاء الاصطناعي .

إن جذور البحوث العلمية المتعلقة بالذكاء الاصطناعي تعود إلى أربعينات القرن الماضي من خلال انتشار أجهزة الإعلام الآلي واستخدامها ، وخاصة بعد التأسيس لعلم الشبكات العصبية خلال سنة 1943 ، حيث تم صياغة مصطلح " الروبوتات " من قبل "اسحق اسيموف" (Isaac Asimov) ¹

إن أول بحث علمي في الذكاء الاصطناعي كان للباحث Alan Turing عام 1950 ، حيث قام بتقديم بما يعرف باختبار تورينج (Turing test) ، الذي يهتم بتقييم الذكاء لجهاز الإعلام الآلي وتصنيفه ذكيا في حال قدرته على محاكاة العقل البشري. ²

كما أن محاولة لمحاكاة العقل البشري من خلال الذكاء الاصطناعي كانت عام 1957 للعالم فرانك روزنبلات ، حينما قام بوضع نموذج مبسط للشبكة العصبية تشبه إلى حد ما الخلايا العصبية في الدماغ البشري. ³

وفي عام 1958 جون مكارثي يخترع لغة البرمجة LISP للذكاء الاصطناعي ⁴ ، وفي عام 1979 قدمت Stanford Cart أول سيارة مستقلة مسيرة عن طريق جهاز الإعلام الآلي ⁵.

في حين أن عام 1997 شهد لأول مرة تفوق للذكاء الاصطناعي على العقل البشري من خلال تمكن أول جهاز للإعلام الآلي من التفوق على منافس بشري في لعبة الشطرنج ⁶، حيث تغلب برنامج يسمى Deep Blue Chess على بطل العالم "كاسبروف" (Garry Kasparov) ⁷.

وفي نفس السياق الذي قام به العالم فرانك روزنبلات خلال عام 1957 ، قام البروفيسور كيفن وارويك أستاذ علم التحكم الآلي بجامعة ريدينج البريطانية في عام 1998 بدراسة مدى تفاعل جهاز الإعلام الآلي والجهاز العصبي ، من خلال زرع شريحة الكترونية في ذراعه وتوصيلها لاسلكيا بجهاز الإعلام الآلي، بهدف إرسال إشارات من الدماغ يستقبلها جهاز الإعلام الآلي ويجولها حركة ، فكانت تفتح الأبواب وتضيء المصابيح بمجرد أن يتجول بالجامعة.⁸

وفي القرن الحالي شهدت أبحاث وبراءات الاختراع في مجال الذكاء الاصطناعي وتيرة سريعة وتطور ملحوظ في التحكم في الآلة، حيث أنه أصبح يستعمل في مختلف المجالات ولاسيما التشخيص الطبي، كما أنه حل في العديد من النشاطات مكان العامل البشري.

الفرع الثاني : تعريف الذكاء الاصطناعي

الذكاء الاصطناعي مصطلح يتكون من كلمتين، **الذكاء (Intelligence)** بمعنى القدرة على الفهم أو التفكير، **الاصطناعي (Artificial)** بمعنى شيء مصنوع أو غير طبيعي.⁹

الذكاء الاصطناعي يهتم بالمحاكاة للأنشطة الفكرية للإنسان من خلال استخدام وسائل مصنوعة وغير طبيعية، حيث أن البحوث العلمية في هذا المجال تهتم بتطوير مناهج الإعلام الآلي تتصرف بذكاء، بغرض جعل الآلات والمعالجة الآلية أكثر ذكاء (أكثر استخداما).¹⁰

إن الذكاء الاصطناعي مبني على أساس محاكاة الذكاء البشري في أنظمة وأجهزة الإعلام الآلي من خلال دراسة وتصميم تطبيقات وبرامج تتصرف وتحاكي التصرفات البشرية.¹¹

يعرف الذكاء الاصطناعي على أنه " فرع من فروع علم الإعلام الآلي يبحث في فهم استخدام التكنولوجيا من خلال الاعتماد على المحاكاة لجهاز الإعلام الآلي لخصائص السلوك الذكي."¹²

كما يعرفه Grewal على أنه " نظام المحاكاة الميكانيكية الذي يقوم جمع المعرفة والمعلومات التي تتعلق بمختلف القطاعات في العالم والعمل على معالجتها ونشرها للاستفادة منها على شكل ذكاء عملي. كما يعرف بأنه " أحد جوانب علم الإعلام الآلي الذي يقوم على توفير مجموعة مختلفة من الأدوات والأساليب والتقنيات لإنشاء تطبيقات ونماذج بغرض إيجاد حلول للمشكلات وذلك من خلال محاكاة سلوك الأشخاص."¹³

إن مصطلح الذكاء الاصطناعي لا يطلق على أي نظام أو برنامج أو تطبيق في جهاز الإعلام الآلي يحاكي تصرفات الأفراد، بل يجب أن تتوفر ثلاث صفات رئيسية تتمثل في:¹⁴

● **القدرة على التعلم:** أي اكتساب المعلومات ووضع قواعد استخدام هذه المعلومات.

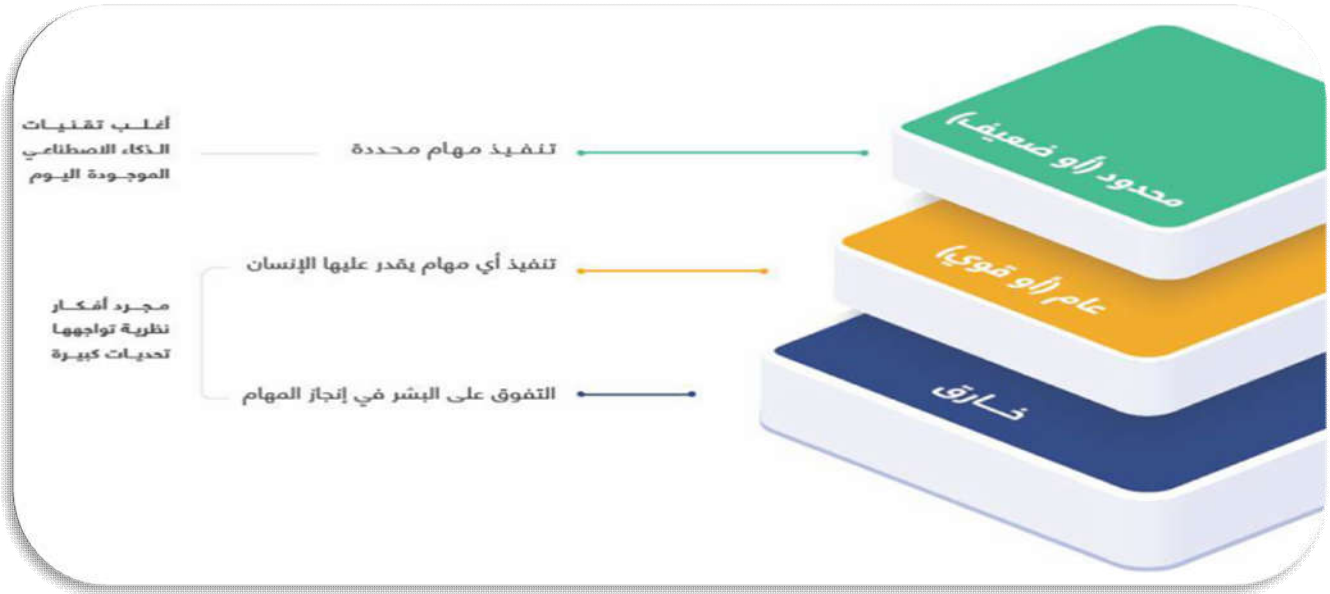
• إمكانية جمع وتحليل هذه البيانات والمعلومات وخلق علاقات فيما بينها: يساعد في ذلك الانتشار المتزايد للبيانات الضخمة (Big Data).

• اتخاذ قرارات بناء على عملية تحليل المعلومات، وليس فقط مجرد خوارزمية تحقق هدفاً معيناً.

الفرع الثالث : أنواع الذكاء الاصطناعي

يمكن تقسيم أنواع الذكاء الاصطناعي إلى ثلاثة أنواع رئيسية، كما هو موضح في الشكل رقم 01 وذلك على النحو التالي:

الشكل رقم 01: أنواع الذكاء الاصطناعي



المصدر: سعاد بوبجة، مرجع سبق ذكره، ص 95

• **الذكاء الاصطناعي محدود (أو الضعيف):** يسمح هذا النوع من الذكاء للآلة بفهم الأوامر ، وتطبيقها والامثال لها ،مثل الطائرات المسييرة،السيارات ذاتية القيادة،برامج التعرف على الوجوه والصور ، لعبة الشطرنج ، وغيرها ، وما يميزها هو محدودية نطاقها بحيث لا يمكنها الخروج عن المهام التي برمجت عليها مسبقاً، وبالتالي تحت نطاق التنبؤ بأفعالها، وإمكانية السيطرة عليها.¹⁵

أي لديها ذكاء محدد يحاكي السلوك الذكي في منطقة محددة¹⁶، بمعنى أن هذا النوع من الذكاء الاصطناعي مبرمج لأداء مهمة واحدة ، أو تحليلات البيانات الخام ويعمل ضمن نطاق محدد مسبقاً.¹⁷

• **الذكاء الاصطناعي عام (أو قوي):** عبارة عن آلات يمكن أن تؤدي بنجاح مهام فكرية إنسانية¹⁸، حيث يمكن إجراء الأبحاث، بأفكار مختلفة، وذلك للحالات التي تواجهها لأول مرة¹⁹، بمعنى أن جهاز الإعلام الآلي يمكن برمجته ليكون عقلاً بشرياً ذكياً بمعنى الكلمة، وأن يكون له الإدراك والفهم ، حتى يتمكن من

اكتساب الخصائص الرئيسية التي يتميز بها الإنسان ، ولا سيما : القدر على التفكير والتفاعل الذكي، حل الألغاز ،إصدار الأحكام،التخطيط والتعلم و التواصل .²⁰

● **الذكاء الاصطناعي الخارق:** وهو الذي من المفترض أن يتجاوز الذكاء البشري من حيث الإبداع وحل المشكلات بفاعلية ²¹، حيث يستخدم في انجاز المهام التي تتفوق فيها القدرة الفكرية للآلة على قدرة البشر ²² ، كالطبيب الجراح ذو الخبرة، حيث يسمح الذكاء للآلة في هذا النوع بالتقدم وتطوير قدراتها المعرفية من خلالها تجربتها الخاصة،بفضل تقنية التعلم الخاصة لدى هذا النوع من الذكاء الاصطناعي ، بحيث تمكنها من القدرة على التعلم والتخطيط ، والتواصل التلقائي، بحيث تقوم بإصدار القرارات والأحكام بسرعة وبصفة مستقلة .²³

المطلب الثاني: التحيز في الذكاء الاصطناعي

الفرع الأول: مفهوم التحيز في الذكاء الاصطناعي

يعد التحيز في الذكاء الاصطناعي ظاهرة جديدة في تكنولوجيات الحديثة، حيث يثير مخاوف العديد من الأفراد حول العالم، قد يمسهم الضرر في حياتهم ومصالحهم العملية والشخصية والاجتماعية.

يعرف التحيز في الذكاء الاصطناعي على أنه " انحراف في نتائج خوارزميات التعلم الآلي يحدث بسبب وجود فرضيات متحيزة أثناء عملية تطوير الخوارزمية، وتمثل انعكاسا لعنصرية المجتمع وتحيزه ضد فئة معينة، أو قد يكون نتيجة تحيز في بيانات التدريب التي يتم تغذية نظام الذكاء الاصطناعي بها .²⁴

وعليه يمكن القول بأن التحيز في الذكاء الاصطناعي في خطأ منهجي يتم تغذية نماذج تدريب الذكاء الاصطناعي ببيانات خاطئة تعكس أحكاما مسبقة ، أو يتم ترجيح بيانات أخرى ، مما يؤدي إلى نتائج منحرفة ، وأخطاء تحليلية ، ولا يقتصر هذا على الجانب التكنولوجي بل على الأفراد أيضا.

إن أنظمة الذكاء الاصطناعي تحتوي على التحيز بسبب:²⁵

■ **التحيزات المعرفية :** هي أخطاء غير واعية في التفكير تؤثر على أحكام الأفراد وقراراتهم. تنشأ هذه التحيزات من محاولة الدماغ تبسيط معالجة المعلومات حول العالم. تم تعريف وتصنيف أكثر من 180 تحيزا بشريا من قبل علماء النفس. يمكن أن تتسرب التحيزات المعرفية إلى خوارزميات التعلم الآلي عبر أي منهما:

○ المصممون يقدمونها إلى النموذج دون أن يلاحظونها.

○ مجموعة بيانات تدريبية تتضمن تلك التحيزات .

■ **عدم وجود بيانات شاملة:** إذا لم تكن البيانات شاملة، فقد لا تكون تمثيلية وبالتالي قد تتضمن تحيزا. على سبيل المثال، تتضمن معظم الدراسات البحثية في علم النفس نتائج من الطلاب الجامعيين الذين يمثلون مجموعة محددة ولا يمثلون جميع السكان.

الفرع الثاني: أنواع التحيز في الذكاء الاصطناعي

التحيز في الذكاء الاصطناعي يمكنه أن يؤثر على المجتمع ، مما يؤدي إلى وجوب دراسته ومعالجة جوانبه، وهذا انطلاقاً من مختلف الطرق التي يمكن من خلالها إدخال التحيز في الذكاء الاصطناعي، والتي يمكن نذكر بعضها ، ولاسيما :²⁶

■ **انحياز العينة (التحيز في الاختيار):** يحدث تحيز العينة عندما لا تعكس مجموعة البيانات واقع البيئة التي سيتم فيها تشغيل النموذج.

■ **تحيز الاستبعاد:** يحدث التحيز في مرحلة معالجة البيانات من خلال عدم استخدام مجموعة من البيانات في النموذج واستبعادها رغم أهميتها.

■ **تحيز القياس:** يحدث هذا التحيز بسبب المشاكل الأساسية المتعلقة بدقة البيانات وكيفية قياسها أو تقييمها.

■ **تحيز الاسترجاع:** يحدث هذا النوع عندما يقام بتسمية أنواع متشابهة من البيانات بشكل غير متسق، مما يؤدي إلى عنه عدم الدقة.

■ **تحيز المراقب :** يحدث عندما يدخل المصممون في النموذج أفكارهم الذاتية حول دراستهم ، من خلال ما يريدون أن يتحقق في معالجة البيانات ، مما يؤدي إلى بيانات غير دقيقة.

■ **التحيز العنصري:** وهو يحدث عندما يتم تنحرف البيانات لصالح فئة معينة.

■ **انحياز الرابطة :** يحدث هذا التحيز عندما يتم تصميم نموذج وفق بيانات خاصة ، مما يؤدي إلى فشل المعالجة الآلية للنموذج بسبب عدم التعرف على الفئات الأخرى.

المبحث الثاني: الأخلاق و التحيز في الذكاء الاصطناعي

المطلب الأول: أهمية الأخلاق في بناء الذكاء الاصطناعي

إن أهمية أخلاق الذكاء الاصطناعي جاءت نتيجة طرح مخاوف إمكانية خروجه عن السيطرة، و أن يتفوق مستقبلاً على الإنسان من خلال الاعتماد على برمجة نفسه بنفسه، مما يستدعي تغذية الذكاء الاصطناعي بجملة من القواعد الأخلاقية التي ينبغي أن يسير وفقها.

1-2 مخاطر التحيز في بناء الذكاء الاصطناعي

توجد عدة مخاطر التي يمكنها أن تهدد مستقبل الإنسان من خلال الاعتماد على تقنيات الذكاء الاصطناعي، وتنقسم هذه المخاوف إلى:²⁷

■ **مخاطر اقتصادية:** إن تقنيات الذكاء الاصطناعي تؤثر على حجم ونوعية الوظائف وفرص العمل المتاحة ، حيث أن يمكن من المتوقع أثناء بناء الذكاء الاصطناعي أن يؤثر بالسلب على الوظائف في مجال الصناعات التحويلية وصناعة السيارات والأدوات الكهربائية، بالإضافة إلى خدمة العملاء.

■ **مخاطر أمنية:** إن أحد المخاوف التي يطرحها بناء الذكاء الاصطناعي هو حق الإنسان في الحياة، وخاصة في حالة الأنظمة القتالية المستقلة (**Lethal Autonomous Weapons**) مثل الدرونز التي تحمل أسلحة. أو الروبوتات المقاتلة الموجودة، حيث تكمن الخطورة هنا في أن هذه الأجهزة مصممة من أجل التدمير أساساً ، فماذا يحدث إذا تم استغلالها من الشخص الخطأ ، أو تم التلاعب بها نتيجة لقصور أو خطأ بشري في إجراءات التأمين أو تم التلاعب بالخوارزميات التي تتحكم فيها. فهنا ستكون النتائج وخيمة.

■ **مخاطر اجتماعية:** تؤدي زيارة الاحتكاك مع الآلات إلى انفصال البشر تدريجاً عن محيطهم الاجتماعي البشري. وهو ما يفقد العلاقات الإنسانية مرونتها التقليدية، ويجعلها أكثر صلابة وجوداً. فتتحول طرق التفكير والتفاعلات البشرية من التعقيد المفيد، إلى لتنميط ولو كان منتجاً. ويصبح الهدف من العلاقات الإنسانية مادياً بعدما كان معنوياً بالأساس.

2-2 مبادئ الأخلاق في بناء الذكاء الاصطناعي

إن الاعتماد على تقنيات الذكاء الاصطناعي في تعزيز البحث والتطوير لتشجيع الدول والأفراد على تلبية حاجاتهم ، يخضع إلى وجوب وضع مجموعة من المبادئ في تصميم الذكاء الاصطناعي نذكر أهمها :²⁸

■ النزاهة والإنصاف

إن عملية بناء أو تصميم أو جمع أو بحث أو تطوير أو استخدام أنظمة الذكاء الاصطناعي تتطلب النزاهة والإنصاف من خلال اتخاذ الإجراءات اللازمة للقضاء على التحيز أو التمييز الذي يتعرض له الأفراد أو الجماعات أو الفئات ، بغرض ضمان معايير عادلة ومنصفة وغير متحيزة وموضوعية وشاملة ومتنوعة وممثلة لجميع شرائح المجتمع .

■ الخصوصية والأمن

يعبر مبدأ الخصوصية والأمن عن القيم والمبادئ الشاملة التي يطلب بموجبها من أنظمة الذكاء الاصطناعي، أن تكون مصممة بطريقة آمنة وتراعي خصوصية أصحاب البيانات الشخصية التي يتم جمعها بحيث تكون على أعلى مستويات الأمن في جميع العمليات والإجراءات المتعلقة بالبيانات وسريتها، الأمر الذي يفرض بدوره إلى منع اختراق البيانات والنظام ، مما قد يؤدي إلى آثار جانبية سواء كانت تفسية أو مالية أو مهنية أو غيرها ، وعليه فإن بناء أنظمة الذكاء الاصطناعي يجب أن يراعي مجموعة من الضوابط والآليات توفر إمكانية إدارة ومراقبة نتائجها طوال فترة عملها لضمان امتثالها دائماً لقواعد الخصوصية والأمن.

■ الإنسانية

يقوم مبدأ الإنسانية على ضرورة بناء أنظمة الذكاء الاصطناعي باستخدام منهجية عادلة مسموح بها أخلاقياً تستند إلى حقوق الإنسان والقيم الثقافية الأساسية وذلك لإحداث أثر مفيد على الأطراف المعنية والمجتمعات المحلية والمساهمة من تحقيق الأهداف والغايات طويلة وقصيرة الأجل من أجل صالح البشرية، ومن الضروري أن يتم تصميم النماذج التنبؤية بحيث لا تتخذ، أو تتلاعب، أو تضع سلوكاً لا يقصد به تمكين أو تعزيز ، أو زيادة المهارات البشرية ، بل ينبغي لها أن تتبنى نهجاً تصميمياً أكثر تركيزاً على الإنسان يتيح له الاختيار واتخاذ القرار.

■ المنافع الاجتماعية والبيئية

يعزز مبدأ المنافع الاجتماعية والبيئية الأثر الإيجابي والمفيد للأولويات الاجتماعية والبيئية التي يجب أن تفيد الأفراد والمجتمع ككل والتي تركز على الأهداف والغايات المستدامة. لا ينبغي لأنظمة الذكاء الاصطناعي أن تسبب أو تسرع الضرر أو تؤثر سلباً على البشر، بل يجب أن تسهم في تمكين واستكمال التقدم الاجتماعي والبيئي مع معالجة التحديات الاجتماعية والبيئية المرتبطة بها، وهذا يستلزم حماية المنفعة الاجتماعية والاستدامة البيئية.

■ الشفافية والقابلية للتفسير

يعد مبدأ الشفافية والقابلية للتفسير عاملاً مهماً لبناء الثقة في أنظمة وتقنيات الذكاء الاصطناعي والحفاظ عليها، لذا يجب بناء أنظمة الذكاء الاصطناعي بدرجة عالية من الوضوح والقابلية للتفسير، مع وجود ميزات لتتبع مراحل اتخاذ القرارات المؤتمتة، ولاسيما تلك التي قد تؤدي إلى آثار ضارة تجاه أصحاب البيانات، وهذا يعني أن البيانات والخوارزميات والقدرات والعمليات والغرض من نظام الذكاء الاصطناعي تحتاج إلى أن تكون شفافة ومعقدة وقابلة للتفسير للمتأثرين بشكل مباشر وغير مباشر، وتعتمد الدرجة التي يكون فيها النظام قابلاً للتتبع والتدقيق والشفافية والقابلية للتفسير على سياق نظام الذكاء الاصطناعي والغرض منه والنتائج التي قد تنتج عن هذه التقنية، ويجب أن تكون أنظمة الذكاء الاصطناعي ومصمموها قادرين على تبرير أسس تصميمها وممارساتها وعملياتها وخوارزمياتها وقراراتها أو سلوكياتها المسموح بها أخلاقياً وغير ضار للعامة.

■ الموثوقية والسلامة

يضمن مبدأ الموثوقية والسلامة التزام نظام الذكاء الاصطناعي بالمواصفات المحددة وأن نظام الذكاء الاصطناعي يعمل بشكل كامل وفق الآلية التي كان يقصدها ويتوقعها مصمموه. تمثل الموثوقية مقياساً للثبات وتبعث الثقة بمدى قوة النظام كما تمثل مقياساً للاعتمادية التي يتوافق بها النظام من الناحية التشغيلية مع وظائفه المرجوة والنتائج التي يحققها، من ناحية أخرى تمثل السلامة مقياساً للكيفية التي لا يشكل بها نظام الذكاء الاصطناعي خطراً على المجتمع والأفراد. على سبيل التوضيح، يمكن لأنظمة الذكاء الاصطناعي مثل المركبات ذاتية القيادة أن تشكل خطراً على حياة الناس في حال عدم التعرف عليهم ككائنات حية أو في حالة عدم تدريب هذه المركبات على بعض

السيناريوهات أو تعطل النظام. يجب أن يكون نظام العمل الموثوق آمناً من خلال عدم تعريض المجتمع للخطر ويجب أن تكون لديه آليات مدمجة لمنع الضرر.

■ المساءلة والمسؤولية

يُحمّل مبدأ المساءلة والمسؤولية المصممين والموردين والقائمين على المشتريات والمطورين ومسؤولي ومقيمي أنظمة الذكاء الاصطناعي والتقنية نفسها المسؤولية الأخلاقية والمسؤولية عن القرارات والإجراءات التي قد تؤدي إلى مخاطر محتملة وآثار سلبية على الأفراد والمجتمعات، ويجب تطبيق الإشراف البشري والحوكمة والإدارة المناسبة عبر دورة حياة نظام الذكاء الاصطناعي بأكملها لضمان وجود آليات مناسبة لتجنب الضرر وإساءة استخدام هذه التقنية، وينبغي ألا تؤدي أنظمة الذكاء الاصطناعي إلى خداع الناس أو الإضرار بحرية اختيارهم دون مبرر، ويكون المصممون والمطورون والأشخاص الذين ينفذون نظام الذكاء الاصطناعي قابلين للتعرف عليهم وأن يتحملوا المسؤولية عن أي أضرار محتملة للتقنية على الأفراد أو المجتمعات، حتى لو كان التأثير السلبي غير مقصود. على الأطراف المسؤولة اتخاذ الإجراءات الوقائية اللازمة بالإضافة إلى وضع إستراتيجية تقييم المخاطر والتخفيف منها للحد من الضرر الناجم عن نظام الذكاء الاصطناعي، ويرتبط مبدأ المساءلة والمسؤولية ارتباطاً وثيقاً بمبدأ العدالة، ويجب على الأطراف المسؤولة عن نظام الذكاء الاصطناعي ضمان الحفاظ على عدالة النظام واستدامتها من خلال آليات الرقابة، وعلى جميع الأطراف المشاركة في دورة حياة نظام الذكاء الاصطناعي مراعاة هذه القيم عند اتخاذهم للقرارات.

المطلب الثاني : مساهمة الأخلاق في الحد من التحيز في الذكاء الاصطناعي

الفرع الأول: طرق لحد من التحيز في الذكاء الاصطناعي

إن استخدام الأخلاق في الذكاء الاصطناعي، يلزم على المصممين أن يتبعون جملة من النقاط في بناء أنظمة الذكاء الاصطناعي، نذكر منها:²⁹

■ **توفير بيانات تدريب متنوعة وتمثيلية :** إن أهمية ضمان أنظمة الذكاء الاصطناعي على مجموعة بيانات متنوعة وتمثيلية للحد من التحيز ، بما فيها البيانات من مختلف التركيبة السكانية والخلفيات ووجهات النظر ، وهذا من خلال توسيع مجموعة البيانات، مما يجعل نماذج الذكاء الاصطناعي لها إمكانية معرفة اتخاذ القرارات أكثر عدلاً وشمولاً.

■ **إجراء تقييمات مستمرة واختبارات صارمة:** يجب أن تخضع أنظمة الذكاء الاصطناعي لفحوصات واختبارات متكررة وشاملة لتحديد وتصحيح التحيزات المحتملة، يمكن إجراء عمليات تدقيق مستقلة لتقييم الأداء والتحيزات المحتملة لنماذج الذكاء الاصطناعي ، مما يساعد على تحديد أي أنماط تمييزية غير مقصودة واتخاذ إجراءات تصحيحية . حيث يجب أن تتضمن هذه الإجراءات مراجعة التعليقات وتقارير المستخدمين وبيانات الأداء لضمان نتائج عادلة ومعلومات صحيحة.

■ **الرقابة والتدخل البشري:** يلعب هذا الإجراء دوراً هاماً في ضمان موثوقية وعدالة وأخلاق نتائج الناتجة عن لذكاء الاصطناعي ، بينما يمكن للذكاء الاصطناعي أن يعالج العمليات آلياً لتقديم نتائج فعالة، حيث يوقر تدخل البشري الضوابط والتوازنات اللازمة للحد من التحيزات وتقييم النتائج ومؤامة القرارات مع المبادئ الأخلاقية. إن التدخل البشري في بناء أنظمة الذكاء الاصطناعي من خلال فهم السياق و الخبرة و التفكير الأخلاقي يمكنهم من إجراء تقييم نقدي للنتائج الناتجة عن الذكاء الاصطناعي ، بغرض تحديد التحيزات والحد منها ، كما أنه يمكنه من حل كل المسائل المعقدة والجديدة التي قد تعاني منها نماذج الذكاء الاصطناعي وهذا من خلال إرساء المساءلة وتعزيز ثقة المستخدم وضمان تصميم واستخدام أنظمة الذكاء الاصطناعي بطريقة مسؤولة ومفيدة.

الفرع الثاني : تحديات الأخلاق في الحد من التحيز في الذكاء الاصطناعي

إن مشروع أخلاق الذكاء الاصطناعي ترجع جذورها إلى اسحق أزيوف Isaac Asimov الذي يعد من بين أبرز الكتاب الخيال العلمي، حيث أشار في مؤلفه « أنا ربات » إلى ثلاثة قوانين أخلاقية ينبغي للذكاء الاصطناعي أن يتبعها: ³⁰

- لا يجوز لأي ربات أن يؤدي أي إنسان أو يسمح من خلال عدم أخذ أر رد فعل، بحدوث أي أذى لأي إنسان.
- يتعين على نموذج للذكاء الاصطناعي يخضع للأوامر التي يصدرها له الإنسان، فيما عدا الأوامر التي تتعارض مع القانون الأول.
- يتعين على نماذج الذكاء الاصطناعي حماية وجودها ما دامت هذه الحماية لا تتعارض مع القانون الأول والثاني.
- أن النماذج الذكاء الاصطناعي لا ينبغي لها أن تؤدي الإنسانية بعدم القيام بأي عمل أو يسمح للإنسانية بإيذاء نفسها.

إلا أن هذا المشروع تواجه عدة تحديات لعل أهمها: ³¹

- أنه لكي تكون الأخلاقيات قابلة للحوسبة يجب أن نعرف النظرية الأخلاقية الصحيحة. والتي تبعاً لها سنقوم بعمليات الحوسبة.
- يجب أن تكون هذه النظرية كاملة، أي يجب أن نحبرنا بكيفية التصرف في مأزق أخلاقي قد يواجهنا ، بمعنى أن الأخلاق ما لم تكتمل بعد كنظرية ، وما لم تتحدد نوعية وطبيعة هذه الأخلاق فلا يمكن القيام بحوسبتها.

- نتائج الدراسة

أنظمة الذكاء الاصطناعي تعمل على مساعدة الأفراد في اتخاذ القرارات التي تؤثر بشكل مباشر على حياتهم وحياة الآخرين ، إلا أنها تشكل تهديداً نتيجة للاستخدام السلبي، مما قد تؤدي الإنسانية ، خاصة عند تغذية النماذج ببيانات خاطئة تعكس أحكاماً مسبقة ، أو من خلال ترجيح بيانات عن أخرى. مما يستدعي إيجاد الحلول اللازمة لمواجهتها أو الحد من أضرارها .

إن عملية أخلاق الذكاء الاصطناعي تعد حلاً لمواجهة التحيز في إدخال البيانات، من خلال إخضاعه إلى مجموعة من المبادئ المتعارف عليها التي يتم وضعها عند تصميم نماذج الذكاء الاصطناعي، مع وجوب مراعاة مجموعة من الإجراءات و القوانين التي يجب أن يتم إتباعها وتطبيقها بغرض الحد من التحيز في الذكاء الاصطناعي.

إن أخلاق الذكاء الاصطناعي عملية معقدة بسبب أن البشر في حد ذاتهم لا يتفقون على قواعد أخلاقية واحدة ، فما بالك بتعليمها للنماذج الذكاء الاصطناعي من أجل برمجتها بها ، و أن تتصرف لوحدها ، في حين أنها يجب أن تخضع إلى أوامر البشر و أن تتصرف وفق لبرامج معينة ، وبالتالي أفعالها مقيدة ، وبالتالي من يتخذ القرار ، النماذج أم البشر ؟.

- توصيات الدراسة

- وجوب القيام بندوات و ورشات عمل متخصصة للتعريف بالذكاء الاصطناعي ونماذجه والتحيز في إدخال البيانات.
 - وجوب وجود نصوص تنظيمية و إرشادية خاصة بكيفية استخدام نماذج الذكاء الاصطناعي وكل ما يتعلق بالتحيز.
 - وجوب توجيه المصممين إلى أهمية الحد من التحيز في تصميم نماذج الذكاء الاصطناعي من خلال إتباع المبادئ المتعارف عليها في أخلاقيات الذكاء الاصطناعي.
- التهميش.

¹ سعاد بوبحة، "الذكاء الاصطناعي: تطبيقات وانعكاسات"، مجلة اقتصاد المال والأعمال، الجزائر، المجلد 6 العدد 4، 2022، ص 92

² امل كاظم ميرة، تحرير جاسم كاطع، " تطبيقات الذكاء الاصطناعي في التعليم من وجهة نظر تدريسي الجامعة " ، مجلة العلوم النفسية، العراق، وقائع المؤتمر العلمي الدولي الأول للدراسات الانسانية: الذكاء والقدرات العقلية ، العدد 22 ، 2019، ص 298.

³ نرمين مجدي، "الذكاء الاصطناعي وتعلم الآلة"، سلسلة كتيبات تعريفية موجهة إلى الفئة العمرية الشابة في الوطن العربي، العدد 3، الصندوق النقد العربي، الإمارات العربية المتحدة، 2020، ص 5.

⁴ سعاد بوبحة، مرجع سبق ذكره، ص 92.

⁵ امل كاظم ميرة، تحرير جاسم كاطع، مرجع سبق ذكره، ص 298.

- 6 المرجع نفسه.
- 7 سعاد بوبجة، مرجع سبق ذكره، ص 92.
- 8 نرمين مجدي، مرجع سبق ذكره، ص 6.
- 9 الأسد صالح الأسد، "المخاوف الأخلاقية من الاستخدامات السلبية لتقنيات الذكاء الاصطناعي: تقنية التزييف العميق أمودجا"، مجلة الرسالة للدراسات الإسلامية، الجزائر، المجلد 06، العدد 02، 2022، ص 372.
- 10 Abdelhamid Abdouni, « Informatique, Intelligence et Intelligence Artificielle », Revue des sciences sociales et humaines, Algérie, N° 04, 1995, P 11.
- 11 جباري لطيفة، "دور نماذج الذكاء الاصطناعي في اتخاذ القرار"، مجلة العلوم الإنسانية، المركز الجامعي تندوف، الجزائر، العدد 01 2017، ص 122.
- 12 بدري جمال، «الذكاء الاصطناعي: بحث عن مقارنة قانونية»، المجلة الجزائرية للعلوم القانونية والسياسية، الجزائر، المجلد 59، العدد 4، 2022، ص 175.
- 13 نورة محمد عبد الله الغزام، "دور الذكاء الاصطناعي في رفع كفاءة النظم الإدارية لإدارة الموارد البشرية بجامعة تبوك"، مجلة التربوية، جامعة سوهاج، مصر، أبريل، ج 2021، 74، ص 477.
- 14 إيهاب خليفة، "الذكاء الاصطناعي: تأثيرات تزايد دور التقنيات الذكية في الحياة اليومية للبشر"، مركز المستقبل للأبحاث والدراسات المتقدمة، أوب طي، تحليل المستقبل، اتجاهات الأحداث، العدد 20، 2017، ص 63.
- 15 بدري جمال، مرجع سبق ذكره، ص 176.
- 16 سعاد بوبجة، مرجع سبق ذكره، ص 95.
- 17 ناصري محمد الشريف، خشايمية سلوى، "مدخل مفاهيمي للذكاء الاصطناعي وتطبيقاته في الإدارة الرياضية"، مجلة علوم الأداء الرياضي، الجزائر، المجلد 3، العدد (خاص 1)، 2021، ص 235.
- 18 المرجع السابق،
- 19 ندى بدر جراح، "تقنيات الذكاء الاصطناعي لتطوير التعلم الآلي الإحصائي"، المجلة العراقية لتكنولوجيا المعلومات، العارقي، المجلد 9، العدد 2019، 3، ص 44.
- 20 سعاد بوبجة، مرجع سبق ذكره، ص 95.
- 21 ناصري محمد الشريف، خشايمية سلوى، مرجع سبق ذكره، ص 95.
- 22 سعاد بوبجة، مرجع سبق ذكره، ص 95.
- 23 بدري جمال، مرجع سبق ذكره، ص 176.
- 1 ²⁴ <https://masaar.net/ar/06/07/2023>، كيف يمكن التغلب على تحيز الذكاء الاصطناعي... تقنيات وأدوات.
- ²⁵ Cem Dilmegani, « Bias in AI: What it is, Types, Examples & 6 Ways to Fix it in 2023 », 2022, <https://research.aimultiple.com/ai-bias/#extra-resources>.
- ²⁶ ناظم حسن رشيد، مي ابلحد افرام، "تدقيق التحيز في الذكاء الاصطناعي في ضوء إطار عمل تدقيق الذكاء الاصطناعي لمعهد المدققين الداخليين (IIA) - دراسة نظرية"، مجلة الدراسات التجارية والاقتصادية المعاصرة، الجزائر، المجلد 6، العدد 2023، 1، ص: 446-447.
- ²⁷ إيهاب خليفة، "تهديدات ذكية: مخاطر خروج "الذكاء الاصطناعي" عن السيطرة البشرية" 30 يوليو، 2017، متاح على الموقع مركز المستقبل <https://futureuae.com/ar-AE>، 2023/07/20.
- ²⁸ الهيئة السعودية للبيانات والذكاء الاصطناعي، «مبادئ أخلاقيات الذكاء الاصطناعي»، متاح على الموقع منصة استطلاع، <https://istitlaa.ncc.gov.sa/ar/Transportation/NDMO/AIEthicsPrinciples/Pages/default.aspx> 2023/07/22.
- ²⁹ Cecilia Cury « The Risks of Biased AI – And How To Avoid I », 22/06/2023، متاح على الموقع <https://rockcontent.com/blog/risks-of-biased-ai/>، 23/07/2023.
- ³⁰ مليكة مذكور، "الأخلاق والحوسبة"، مجلة الآداب والعلوم الاجتماعية، الجزائر، المجلد 19، العدد 2، 2022، ص 41.
- ³¹ المرجع نفسه، ص 42.

قائمة المراجع والمصادر:

المواقع الإلكترونية

1. [https://masaar.net/ar/تقنيات وأدوات/الاصطناعي...](https://masaar.net/ar/تقنيات وأدوات/الاصطناعي... كيف يمكن التغلب على تحيز الذكاء الاصطناعي) كيف يمكن التغلب على تحيز الذكاء الاصطناعي... تقنيات وأدوات 06/07/2023.
2. Cem Dilmegani, « Bias in AI: What it is, Types, Examples & 6 Ways to Fix it in 2023 »,2022, <https://research.aimultiple.com/ai-bias/#extra-resources>
3. إيهاب خليفة، "تهديدات ذكية: مخاطر خروج " الذكاء الاصطناعي " عن السيطرة البشرية"30، يوليو 2017، متاح على الموقع مركز المستقبل، <https://futureuae.com/ar-AE>، 2023/07/20.
4. الهيئة السعودية للبيانات والذكاء الاصطناعي، «مبادئ أخلاقيات الذكاء الاصطناعي»، متاح على الموقع منصة استطلاع <https://istitlaa.ncc.gov.sa/ar/Transportation/NDMO/AIEthicsPrinciples/Pages/default.aspx> 2023/07/22.
5. Cecília Cury « The Risks of Biased AI – And How To Avoid I »,22/06/2023, متاح على الموقع <https://rockcontent.com/blog/risks-of-biased-ai/> , 23/07/2023

المقالات

1. سعاد بوبحة، "الذكاء الاصطناعي: تطبيقات وانعكاسات"، مجلة اقتصاد المال والأعمال، الجزائر، المجلد 6 العدد4، 2022.
2. امل كاظم ميرة، تحرير جاسم كاطع، " تطبيقات الذكاء الاصطناعي في التعليم من وجهة نظر تدريسي الجامعة " ،مجلة العلوم النفسية، العراق، وقائع المؤتمر العلمي الدولي الأول للدراسات الإنسانية: الذكاء والقدرات العقلية ، العدد 22، 2019.
3. الأسد صالح الأسد، "المخاوف الأخلاقية من الاستخدامات السلبية لتقنيات الذكاء الاصطناعي: تقنية التزييف العميق أنموذجا"، مجلة الرسالة للدراسات الإسلامية، الجزائر، المجلد 06، العدد02، 2022.
4. جباري لطيفة، "دور نماذج الذكاء الاصطناعي في اتخاذ القرار "، مجلة العلوم الإنسانية، المركز الجامعي تندوف، الجزائر، العدد01، 2017
5. بدري جمال، «الذكاء الاصطناعي: بحث عن مقاربة قانونية»، المجلة الجزائرية للعلوم القانونية والسياسية، الجزائر، المجلد 59، العدد 4، 2022.
6. نورة محمد عبد الله الغزام، " دور الذكاء الاصطناعي في رفع كفاءة النظم الإدارية لإدارة الموارد البشرية بجامعة تبوك"، مجلة التربية، جامعة سوهاج، مصر، أبريل، ج1، 2021، 74.

7. ناصري محمد الشريف، خشايمية سلوى، "مدخل مفاهيمي للذكاء الاصطناعي وتطبيقاته في الإدارة الرياضية"، مجلة علوم الأداء الرياضي، الجزائر، المجلد 3، العدد (خاص 1)، 2021
8. ندى بدر جراح، "تقنيات الذكاء الاصطناعي لتطوير التعلم الآلي الإحصائي"، المجلة العراقية لتكنولوجيا المعلومات، العار، المجلد 9، العدد 3، 2019.
9. مليكة مذكور، "الأخلاق والحوسبة"، مجلة الآداب والعلوم الاجتماعية، الجزائر، المجلد 19، العدد 2، 2022.
10. ناظم حسن رشيد، مي ابلحد افرام، "تدقيق التحيز في الذكاء الاصطناعي في ضوء إطار عمل تدقيق الذكاء الاصطناعي لمعهد المدققين الداخليين (IIA) - دراسة نظرية"، مجلة الدراسات التجارية والاقتصادية المعاصرة، الجزائر، المجلد 6، العدد 1، 2023
6. Abdelhamid Abdouni, « **Informatique, Intelligence et Intelligence Artificielle** », Revue des sciences sociales et humaines, Algérie, N° 04, 1995

مصادر أخرى

1. نرمين مجدي، "الذكاء الاصطناعي وتعلم الآلة"، سلسلة كتيبات تعريفية موجهة إلى الفئة العمرية الشابة في الوطن العربي، العدد 3، الصندوق النقد العربي، الإمارات العربية المتحدة، 2020
2. ايهاب خليفة، "الذكاء الاصطناعي: تأثيرات تزايد دور التقنيات الذكية في الحياة اليومية للبشر"، مركز المستقبل للأبحاث والدراسات المتقدمة، أوب ظبي، تحليل المستقبل، اتجاهات الأحداث، العدد 20، 2017.