أخلاقيات الذكاء الاصطناعي

Artificial intelligence ethics

بلقصير مصطفى

جامعة الدكتور مولاي الطاهر سعيدة (الجزائر)، moos743@gmail.com

تاريخ الاستلام: 2023/10/10 تاريخ القبول 27 / 03 / 2024 تاريخ النشر: 2024/12/29

ملخص:

يسعى هذا البحث إلى مناقشة مجمل القضايا المرتبطة بمبادئ أخلاقيات الذكاء الاصطناعي، تهدف إلى تفادي الأخطار التي تهدد الإنسان حاضرا ومستقبلا، وفهم العلاقة الجديدة التي تربط الإنسان بالآلة، لاستخلاص فوائد وعيوب الذكاء الاصطناعي من خلال عدسة الأخلاق، ونوع النظريات التي نختارها، ومن ثم صياغتها في قوانين تبرمج في الآلات

الذكاء الاصطناعي، هو إعادة إنتاج حصيلة سلوكاتنا الذكية ودمجها في الآلات، كي تقوم بتصرفات مكافئة للإنسان، وهذا بجعل منها مطلب إنساني مهم لكنه محفوف بالمخاطر، مما يستدعي مبادئ أخلاقية رصينة ومستدامة.

كلمات مفتاحية: أخلاق ، الذكاء الإصطناعي، الربوت، القيم، القانون.

Abstract:

This research aims to discuss the overall issues related to the ethics of artificial intelligence. It seeks to avoid the dangers that threaten present and future humanity, and to understand the new relationship between humans and machines. This is done in order to extract the benefits and drawbacks of artificial intelligence through an ethical lens, considering the types of theories we choose, and then formulating them into programming laws for machines.

This research aims to discuss the overall issues related to the ethics of artificial intelligence. It seeks to avoid the dangers that threaten present and future humanity, and to understand the new relationship between humans and machines. This is done in order to extract the benefits and drawbacks of artificial intelligence through an ethical lens, considering the types of theories we choose, and then formulating them into programming laws for machines

Keywords: Ethics; Artificial Intelligence; Robot; Values; Law.

1- مقدمة

يغزو الذكاء الاصطناعي حياتنا تدريجيًا، وغالبًا ما يكون ذلك بشكل غير محسوس ولكنه منتشر بشكل متزايد. وسواء كان الأمر يتعلق بالوكلاء الشخصيين الأذكياء، أو برامج البحث، أو التشغيل الآلي ، أو التعرف على الوجه أو

المساعدة التشخيصية الطبية، إضافة إلى وسائل النقل، أو الروبوتات، أو استخدام الطائرات العسكرية بدون طيار، أو حتى في التمويل أو الشرطة أو التعليم، فإن الذكاء الاصطناعي يتسلل إلى كل مكان. إنه يبهر ويقلق في نفس الوقت.

وفي الوقت نفسه، تضاعفت الأسئلة الأخلاقية: فبعيدًا عن المخاوف التي يغذيها الخيال العلمي بشأن تهديداته للبشرية، تميل الأفكار إلى التركيز على تصميم الخوارزميات التي قد تكون لها بالفعل عواقب وخيمة على حياتنا اليومية. إن عمليات الحساب وحل المشكلات التي يقوم عليها الذكاء الاصطناعي هي في الواقع غامضة بشكل خاص ويمكن أن تنتج التحيزات والتمييز وتؤدي إلى ظهور وجهات نظر عالمية، من دون أن نتمكن دائما من فتح "الصندوق الأسود" الذي يصنعها.

من هنا جاءت فكرة أخلاقيات الذكاء الاصطناعي، وكيفية وضع برمجة قوانين تضبط بها الآلات، لكن مفهوم الأخلاق والقيم هو في حد ذاته يحمل إشكاليات عديدة في المجالين الفلسفي والعلمي، ففي المجال الأول لا توجد نظرية موحدة بين الأخلاقيين تضبط سلوك الأفراد، وتزداد الإشكالية تعقيدا في المجال العلمي، بمعنى صياغة نظرية أخلاقية تطبق على الآلات، وهذا البحث الذي نقدمه يندرج ضمن هذا الإطار للمستوى العلمي أي الذكاء الإصطناعي، فارتأينا طرح الإشكالية التالية :

هل يمكن إعتماد مبادئ أخلاقيات في مجال الذكاء الإصطناعي، في ظل التعقيدات القيمية والتحديات العلمية؟

أما المنهج الذي نتبعه في مقاربتنا للإشكالية، ووفقا لطبيعة البحث هو المنهج التحليلي، من خلال طرح نظريات وأفكار المفكرين والعلماء في مجال الذكاء الإصطناعي، لنقوم بتحليلها ثم الوصول إلى مقاربة تسهم في معالجة الإشكالية الأساسية المرتبطة بأخلاقيات الذكاء الإصطناعي

2.ما هو الذكاء الإصطناعي

1.2 مقدمة إلى الذكاء الاصطناعي

لقد تطور مجال الذكاء الاصطناعي (AI) من بدايات متواضعة إلى مجال له تأثير عالمي. يصف الخبراء مجال الذكاء الاصطناعي هو أنه كل ما لا تستطيع أجهزة الكمبيوتر القيام به حاليًا. على الرغم من أنه أمر ملفت للنظر ظاهريًا، إلا أن هناك شعورًا بأن تطوير أجهزة الكمبيوتر والروبوتات الذكية يعني خلق شيء غير موجود اليوم. الذكاء الاصطناعي هو هدف متحرك.

في الحقيقة، حتى تعريف الذكاء الاصطناعي نفسه متقلب وقد تغير بمرور الوقت. يعرّف كابلان Kaplan وهاينلاين Haenlein الذكاء الاصطناعي بأنه "قدرة النظام على تفسير البيانات الخارجية بشكل صحيح، والتعلم من هذه البيانات، واستخدام تلك الدروس المستفادة لتحقيق أهداف ومهام محددة من خلال التكيف المرن"[1]. كما يعرفه عرّفه بول Poole وماكوورث Mackworth بأنه "المجال الذي يدرس تركيب وتحليل العوامل الحسابية التي تعمل بذكاء[2]". ويعرفون الوكيل بأنه " شيء يتصرف في بيئة – يقوم بفعل شيء. الوكلاء يشملون الديدان والكلاب والمراوح والطائرات

والروبوتات والبشر والشركات والدول[3]". ويهتمون بما يفعله الوكيل؛ أي، كيف يتصرف من خلال أفعاله، حيث يكون تصرف الوكيل ذكيا عندما:

" •يكون ما يفعله مناسبًا لظروفه وأهدافه،

•يكون مرنًا تجاه التغييرات في البيئات والأهداف المتغيرة،

•يتعلم من الخبرة،

•يتخذ خيارات مناسبة مع محدودياته الإدراكية والحسابية[4]".

كذلك عرّف راسل Russell ونورفيج Norvig الذكاء الاصطناعي بأنه "دراسة للعمليات التي يقوم بها الوكلاء الذين يستقبلون الإدراكات من البيئة ويقومون بتنفيذ إجراءات. يقوم كل وكيل من هذا النوع بتنفيذ وظيفة تختصر تسلسلات الإدراكات إلى إجراءات، ونغطي وسائل مختلفة لتمثيل هذه الوظائف، مثل الوكلاء التفاعليين، والمخططين المنطقيين، وأنظمة نظرية القرار، وأنظمة التعلم العميق[5]"، كما حدد راسل ونورفيج أيضًا أربع مدارس فكرية للذكاء الاصطناعي. في الأولى يركز الباحثين على إنشاء آلات تفكر مثل البشر. يسعى البحث داخل هذه المدرسة الفكرية إلى إعادة إنتاج عمليات وتمثيلات ونتائج التفكير البشري بطريقة ما على الآلة. وتركز المدرسة الثانية على إبداع آلات تتصرف مثل البشر. ويكون الإهتمام على العمل ، أي ما يفعله الوكيل أو الروبوت فعليًا في العالم، وليس على عملية وصوله إلى هذا الفعل. وتركز مدرسة ثالثة على تطوير الآلات التي تعمل بعقلانية. ترتبط العقلانية ارتباطًا وثيقًا بالمثالية. تهدف أنظمة الذكاء الاصطناعي هذه إلى القيام دائمًا بالشيء الصحيح أو التصرف بالطريقة الصحيحة. وأخيرًا، تركز المدرسة الرابعة على تطوير الآلات التي تفكر بعقلانية. من المفترض أن يكون التخطيط و/أو اتخاذ القرار الذي ستقوم به هذه الآلات هو الأمثل. المثلي هنا يرتبط بشكل طبيعي ببعض المشكلات التي يحاول النظام حلها.

لعل العنصر الأساسي المشترك بين التعريفات الثلاث السابقة جميعًا هو أن الذكاء الاصطناعي يتضمن دراسة وتصميم وبناء عملاء أذكياء يمكنهم تحقيق الأهداف. يجب أن تكون الاختيارات التي يتخذها الذكاء الاصطناعي مناسبة لقيوده الإدراكية والمعرفية. إذا كان الذكاء الاصطناعي مرنًا ويمكنه التعلم من الخبرة وكذلك الإحساس والتخطيط والتصرف على أساس تكوينه الأولي، فقد يقال إنه أكثر ذكاءً من الذكاء الاصطناعي الذي لديه مجموعة من القواعد التي توجه مجموعة ثابتة من المهام. أجراءات. ومع ذلك، هناك بعض السياقات التي قد لا ترغب فيها أن يتعلم الذكاء الاصطناعي قواعد وسلوكيات جديدة، أثناء أداء إجراء طبي، على سبيل المثال. ويميل أنصار الأساليب المختلفة إلى التأكيد على بعض هذه العناصر أكثر من غيرها. على سبيل المثال، يرى مطورو الأنظمة المتخصصة أن الذكاء الاصطناعي هو مستودع للمعرفة المتخصصة التي يمكن للبشر الرجوع إليها، في حين يرى مطورو أنظمة التعلم الآلي أن الذكاء الاصطناعي شيء قد يكتشف معرفة جديدة. وكما سنرى، فإن لكل نهج نقاط قوة ونقاط ضعف.

2.2 الذكاء الاصطناعي القوي والضعيف

قام جون سيرل John Searle بتقسيم الذكاء الاصطناعي إلى فئتين متميزين. الذكاء الاصطناعي الضعيف يقتصر على مهمة واحدة محددة بدقة. تصنف معظم أنظمة الذكاء الاصطناعي الحديثة ضمن هذه الفئة. تم تطوير هذه الأنظمة للتعامل مع مشكلة أو مهمة أو قضية واحدة وهي بشكل عام غير قادرة على حل المشكلات الأخرى، حتى تلك المرتبطة بها. وعلى النقيض من الذكاء الاصطناعي الضعيف، يعرّف سيرل الذكاء الاصطناعي القوي كالتالي: "إن الكمبيوتر المبرمج بشكل مناسب هو في الحقيقة عقل، بمعنى أنه يمكن القول حرفيًا أن أجهزة الكمبيوتر التي تم تزويدها بالبرامج الصحيحة تفهم وتمتلك حالات معرفية أخرى. في الذكاء الاصطناعي القوي، نظرًا لأن الكمبيوتر المبرمج لديه حالات معرفية، فإن البرامج ليست مجرد أدوات تمكننا من اختبار التفسيرات النفسية؛ بل إن البرامج هي نفسها التفسيرات[6]". في الذكاء الاصطناعي القوي، اختار سيرل ربط إنجاز الذكاء الاصطناعي بتمثيل المعلومات في العقل البشري. في حين أن معظم الباحثين في مجال الذكاء الاصطناعي غير مهتمين بإنشاء وكيل ذكي يلبي شروط الذكاء الاصطناعي القوي التي وضعها سيرل، فإن هؤلاء الباحثين يسعون في النهاية إلى إنشاء آلات لحل مشكلات متعددة غير محددة بشكل ضيق. وبالتالي فإن أحد أهداف الذكاء الاصطناعي هو إنشاء أنظمة مستقلة تحقق مستوى معينًا من الذكاء العام. لم يحقق أي نظام ذكاء اصطناعي ذكاءً عامًا حتى الآن.

3.2 ما هو الروبوت؟

عادة ما يكون وكيل الذكاء الاصطناعي هو برنامج يعمل عبر الإنترنت أو في عالم محاكي، وغالبًا ما يولد تصورات و/أو يتصرف داخل هذا العالم الاصطناعي. ومن ناحية أخرى، يوجد الروبوت في العالم الحقيقي، مما يعني أن وجوده وتشغيله يحدث في العالم الحقيقي. وللروبوتات هيئة جسدية أيضًا، مما يعني أن لديهم جسدًا ماديًا. وغالبًا ما توصف عملية اتخاذ الروبوت للقرارات الذكية بأنها "إدراك–تخطيط– عمل"، مما يعني أن الروبوت يجب أن يستشعر البيئة أولاً، ويخطط لما يجب القيام به، ثم يتصرف في العالم.

3. ما هي الأخلاق؟

غالبًا ما يتم أخذ مصطلحي "الإتيقا " و"الأخلاق" كمرادفين. ومع ذلك، يتم التمييز بينهما في بعض الأحيان، بمعنى أن الأخلاق تشير إلى مجموعة معقدة من القواعد والقيم والأعراف التي تحدد أو من المفترض أن تحدد تصرفات الناس، في حين تشير الإتيقا إلى نظرية الأخلاق. ويمكن القول أيضًا أن الأخلاق تهتم بالمبادئ والأحكام العامة والمعايير أكثر من اهتمامها بالأحكام والقيم الذاتية أو الشخصية.

من الناحية اللغوية، تعود كلمة "أخلاق" إلى الكلمة اليونانية القديمة "Ethos". يشير هذا في الأصل إلى مكان الإقامة، والموقع، ولكن أيضًا إلى العادة والعرف والتقليد. كان شيشرون هو من ترجم المصطلح اليوناني إلى اللاتينية بكلمة

"الأعراف" (الروح والعادات)، والتي اشتُق منها المفهوم الحديث للأخلاق (شيشرون 44 قبل الميلاد). ووصف الفيلسوف الألماني إيمانويل كانط الأخلاق بأنها تتعامل مع سؤال "ماذا علي أن أعمل ؟"[7].

1.3 الأخلاق التطبيقية

تشير الأخلاقيات التطبيقية إلى مجالات الأكثر تحديدا حيث يتم إصدار الأحكام الأخلاقية، على سبيل المثال في مجالات الطب (أخلاقيات الطب)، أو التكنولوجيا الحيوية (أخلاقيات علم الأحياء)، أو الأعمال التجارية (أخلاقيات الأعمال). وبهذا المعنى، يمكن تمييز الاعتبارات المعيارية العامة عن الاعتبارات الأكثر تطبيقًا. ومع ذلك، لا ينبغي النظر إلى العلاقة بين الاثنين على أنها أحادية الاتجاه، بمعنى أن الاعتبارات العامة تأتي أولاً ثم يتم تطبيقها لاحقًا على العالم الحقيقي. بل يمكن أن يكون الاتجاه في كلا الاتجاهين، مع وجود ظروف خاصة للمنطقة المعنية تؤثر على المسائل الأخلاقية العامة. على سبيل المثال، قد يعني المبدأ الأخلاقي العام للتضامن أشياء مختلفة في ظل ظروف مختلفة. في مجموعة صغيرة، قد يعني ذلك مشاركة سلع معينة بشكل مباشر مع أصدقائك وعائلتك. ومع ذلك، في مجموعة أكبر أو في المجتمع بأكمله، قد يعني ذلك ضمنًا اتخاذ تدابير مختلفة تمامًا، مثل التنافس العادل مع بعضها البعض.

2.3 العلاقة بين الأخلاق والقانون.

في كثير من الأحيان، يُنظر إلى الأخلاق والقانون على أنهما مختلفان بشكل واضح عن بعضهما البعض، بل وأحيانًا على أنهما متضادان، بمعنى أن الأخلاق، على سبيل المثال، تبدأ حيث ينتهي القانون. تبعا لذلك سيكون على الأشخاص أو الشركات واجبات قانونية وواجبات أخلاقية، والتي لا علاقة لها ببعضها البعض. ومع ذلك، يمكن تحدي وجهة النظر هذه بعدة طرق. أولاً، غالباً ما يكون للقواعد القانونية جانب أخلاقي أيضاً. على سبيل المثال، لا تزال المعايير القانونية التي تجعل التلوث البيئي غير قانوني بمثابة معايير أخلاقية أيضا. إن الكثير من الإطار القانوني للمجتمع (مثل قوانين مكافحة الاحتكار) له أهمية أخلاقية كبيرة للمجتمع. ثانيا، من الممكن أن تصبح الأخلاقيات (وقد أصبحت بالفعل) إلى حد ما نوعا من "القانون المرن"، بمعنى أن الشركات تحتاج إلى اتباع معايير أخلاقية معينة حتى لو كان القانون في بلد معين لا يتطلب ذلك بشكل صارم. خوفًا من الإضرار بسمعتها، أو خفض قيمة أسهمها، على سبيل المثال، تلتزم الشركات في كثير من الحالات بالقواعد الأخلاقية التي لها تقريبًا نفس العواقب والتأثير مثل القواعد القانونية ("القانون الصارم"). في بعض الأحيان، يتم استخدام العملية الأخلاقية المحددة للشركة كحجة للمبيعات فريدة، مثل الشركات التي تبيع قهوة "التجارة العادلة" بدلاً من ذلك أو مجرد قهوة قانونية عادية.

3.3 أخلاقيات الآلة

تحاول أخلاقيات الآلة الإجابة على السؤال التالي: ما الذي يتطلبه بناء ذكاء اصطناعي أخلاقي يمكنه من اتخاذ قرارات أخلاقية؟ والفرق الرئيسي بين البشر الذين يتخذون قرارات أخلاقية والآلات هو أن الآلات لا تملك "ظواهر" أو "مشاعر" بنفس الطريقة التي يمتلكها البشر " الأخلاق هي ببساطة تعبير عاطفي، ولا يمكن للآلات أن تمتلك عواطف[8]". ليس لديهم "حدس أخلاقي" أو "تثاقف" أيضًا. يمكن للآلات معالجة البيانات التي تمثل المشاعر، ومع ذلك، لا أحد، حتى

الآن، يفترض أن أجهزة الكمبيوتر يمكن أن تشعر بالفعل وتكون واعية مثل البشر. لقد تم تطوير روبوتات نابضة بالحياة، لكنها لا تمتلك وعيًا استثنائيًا أو مشاعر فعلية بالمتعة أو الألم.

كما أن الهدف من إنشاء آلات تتخذ قرارات أخلاقية لا يخلو من المنتقدين، ويؤكدون على فشل علماء الروبوتات بشكل عام في تقديم أسباب قوية لتطوير الروبوتات الأخلاقية.

العنصر الفلسفي هو نظرية أخلاقية مفصلة. وسوف يزودنا بوصف لسمات العالم التي لها أهمية أخلاقية وإجراءات اتخاذ القرار التي تمكننا من تحديد الأفعال الصحيحة والخاطئة في موقف معين.

4. المبادئ الأخلاقية للذكاء الاصطناعي

تقترح المبادئ الأخلاقية الأوروبية للذكاء الاصطناعي، التي قدمتها مجموعة AI4People [9] في عام 2018، خمسة مبادئ لأخلاقيات الذكاء الاصطناعي[10]، والتي يمكن ربطها بالثقة والعدالة على النحو التالي:

1.4 عدم الإيذاء

ينص مبدأ عدم الإيذاء على أن الذكاء الاصطناعي لن يؤذي الناس. تشكل أنظمة الذكاء الاصطناعي التي تضر بالناس (والحيوانات أو الممتلكات) خطرا، سواء على الأفراد أو الشركات. ومن حيث الثقة والإنصاف، يمتد هذا المبدأ إلى قضايا مثل التنمر وخطاب الكراهية التي تعد أمثلة بارزة على انتهاك مبدأ عدم الإيذاء عبر الإنترنت.

1.1.4 التنمر والتحرش عبر وسائل التواصل الاجتماعي

تتعدد حالات تعرض أطفال المدارس للتخويف من قبل زملاء الدراسة أو الموظفين الذين يتعرضون للتخويف في مكان العمل. هناك حالات أدى فيها هذا إلى الانتحار. وفي حين أن هذا النوع من التنمر كان موجودًا بالتأكيد قبل ظهور التقنيات الرقمية، فقد اكتسب، من خلال شبكات التواصل الاجتماعي، مستوى جديدًا من القلق، كما أشارت الدراسات الاستقصائية. وقد ارتفع خطر التعرض لمجموعات كبيرة أو للجمهور بشكل كبير. ونتيجة لذلك، صدرت في عدد من البلدان قوانين ضد التنمر في وسائل الإعلام الرقمية (كانت السويد أول دولة تفعل ذلك في عام 1993).

2.1.4 خطاب الكراهية

حظي خطاب الكراهية في السنوات الأخيرة بتغطية إعلامية كبيرة. خطاب الكراهية هو نوع من الخطاب الذي لا يقتصر فقط على توجيهه بشكل محدد ضد شخص معين، ولكن يمكن أيضًا توجيهه ضد مجموعات أو أجزاء بأكملها من السكان. فقد تهاجم، مثلا مجموعات على أساس أصلها العرقي، أو ميولها الدينيةا، أو جنسها، أو هويتها، أو إعاقتها، وغيرها. ويتضمن العهد الدولي الخاص بالحقوق المدنية والسياسية، المعمول به منذ عام 1976، بياناً ينص على أنه "تحظر بالقانون أية دعوة إلى الكراهية القومية أو العنصرية أو الدينية تشكل تحريضاً على التمييز أو العداوة أو العنف".

ومنذ ذلك الحين، تم إقرار قوانين ضد خطاب الكراهية في عدد من البلدان. فعلى سبيل المثال قد أصدرت بلجيكا قانونًا في عام 1981 ينص على أن التحريض على التمييز أو كراهية أو العنف ضد الأشخاص أو المجموعات من العرق أو اللون أو

الأصل أو المعارضة القومية أو العرقية يعد أمرًا غير قانوني ويعاقب عليه وفقًا لقانون العقوبات البلجيكي. ولدى دول أخرى ككندا ونيوزيلندا والمملكة المتحدة قوانين مماثلة تختلف في التفاصيل والعقوبات.

إضافة إلى ذلك، تم التوصل إلى مستوى جديد من قوانين خطاب الكراهية في عام 2017 في ألمانيا، عندما أقر البوندستاغ الألماني مشروع قانون يجرم خطاب الكراهية على وجه التحديد على وسائل التواصل الاجتماعي. يصر القانون أيضًا على أنه قد يتم تغريم الشبكات الاجتماعية (مثل الفيسبوك Facebook أو غيره) بمبالغ كبيرة جدًا تصل إلى 50 مليون يورو في حالة عدم سعيها بنشاط إلى إزالة محتوى معين بنجاح في غضون أسبوع. وكان إقرار هذا القانون مثيرًا للجدل، حيث ذكر عدد من المعلقين الألمان، و المعلقين الدوليين، أن مثل هذا القانون بعيد المدى للغاية وسيكون له عدد من العواقب غير المقصودة والتي تؤدي إلى نتائج عكسية. ومنذ ذلك الحين، بذلت شبكات التواصل الاجتماعي مثل تويتر أو فيسبوك الكثير من الجهود للامتثال لهذا القانون الجديد. وعلى الرغم من أنهم نجحوا بالتأكيد في إزالة الكثير من المحتوى غير القانوني (كما اعترفت المفوضية الأوروبية في عام 2019، فقد كانت هناك أيضًا العديد من المشكلات المتعلقة بإزالة المحتوى إما عن طريق الصدفة أو بسبب الإفراط في تفسير بعض البيانات. كما بدأت الشبكات الاجتماعية المهمة في تنفيذ ذلك.

يتم تحديد التوازن بين خطاب الكراهية وحرية التعبير داخل الدول وفيما بينها. لقد غمرت وسائل التواصل الاجتماعي هذه المناقشات بعناصر من وجهات النظر الثقافية والوطنية. فقد تضطر الشركات الكبيرة مثل Facebook وGoogle إلى تعديل محتواها. ويرى البعض ذلك على أنه وسيلة ضرورية لمكافحة العنصرية، كما يرى البعض الآخر ذلك على أنه هجوم على حرية التعبير.

2.4 الربح

ينص مبدأ الإحسان على أن الذكاء الاصطناعي يجب أن يفيد الناس. وهو مبدأ في البيوإتيقا، يقضي بأن فوائد العلاج يجب أن تفوق الأضرار المحتملة. كذلك يحتاج نظام الذكاء الاصطناعي إلى مراعاة القواعد الأخلاقية ليصبح جديرًا بالثقة وعادلاً مما يمكنه من تحسين الحياة. ومن الأمثلة على الطرق التي يمكن لنظام الذكاء الاصطناعي من خلالها إثبات فائدته فيما يتعلق بالمشاكل المجتمعية نذكر:

- الحد من الوفيات والحوادث الناجمة عن المركبات ذاتية القيادة.
- توفير الدعم والرعاية الآلية لمجتمع المسنين.
- استخدام التطبيب عن بعد في المناطق النائية .
- تحسينات الاستدامة القائمة على الشبكة الذكية.
- إدخال تحسينات على التنوع البيولوجي، على سبيل المثال، من خلال تطبيقات الذكاء الاصطناعي لحفظ الأنواع المهددة بالانقراض
- استخدام الروبوتات والذكاء الاصطناعي في التعليم، باستخدام الذكاء الاصطناعي للتعلم الفردي أو دعم الطلاب خارج الفصل الدراسي.

3.4 مبدأ الإستقلالية

ينص مبدأ الاستقلالية على أن الذكاء الاصطناعي يجب أن يحترم أهداف الناس ورغباتهم. و مبدأ الاستقلالية له عدة معانٍ. تقليديا في الذكاء الاصطناعي والروبوتات، يشير المصطلح إلى قدرة نظام الذكاء الاصطناعي أو الروبوت على العمل دون تدخل بشري. وفي سياق البيوإتيقا، يشير إلى أن المرضى لديهم الحق في أن يقرروا بأنفسهم ما إذا كانوا سيخضعون للعلاج أم لا. ويستلزم ذلك منح المرضى الحق في رفض الإجراءات المنقذة للحياة واتخاذ قرار بعدم تناول الأدوية التي تقلل المخاطر. على سبيل المثال، هناك حالات معروفة توفي فيها أشخاص نتيجة رفض نقل الدم لأسباب دينية. بشكل عام، وجدت المحاكم أن الآباء ليس لديهم الحق في فرض آرائهم، مثل رفض نقل الدم، على أطفالهم[11]. ومع ذلك، بمجرد أن يصبح الطفل بالغًا، يمكنه رفض العلاج. وبشكل أعم، تشير الاستقلالية إلى قدرة الشخص على اتخاذ القرارات. يمكن للناس أن يقرروا ما إذا كانوا يريدون المخاطرة أم لا لكسب المزيد من المال أو الحصول على المزيد من المتعة.

هناك حدود أخلاقية للاستقلالية. لنفترض مثلا أن طفلاً قال للروبوت: "اسحب شعر أختي، لقد كانت لئيمة معي وأنا أكرهها!" في هذه الحالة هل يجب على الروبوت أن يطيع الطفل؟ إذا قال شخص بالغ: "اضرب ابنتي، لقد كانت شقية"، فهل يجب على الروبوت أن يطيعها؟ بشكل عام، لا ينبغي للأنظمة أن تساعد الأشخاص على تحقيق أهداف غير قانونية أو غير أخلاقية. علاوة على ذلك، لا ينبغي استخدام الأنظمة للقيام بأعمال غير قانونية أو غير أخلاقية. لا توجد حجة قوية للأنظمة التي تسمح للمستخدمين باستخدام الأنظمة لإيذاء الآخرين ما لم يكن هناك سبب وجيه. وينبغي للمرء أيضًا أن يكون حذرًا للغاية عند التفكير في تصميم أنظمة تسمح للآلات بإيذاء البشر الذين يطلبون الأذى. على سبيل المثال، قد يرفض المرء تطوير روبوت للقتل الرحيم. هناك خطر أن مثل هذا الروبوت قد لا يشعر بالمرض العقلي. كما قد يتردد المرء في تفويض مثل هذه القرارات الخطيرة للآلات على الإطلاق.

ومع ذلك، هناك بعض الحالات التي تم فيها تصميم الروبوتات لاستخدام القوة ضد البشر. غقد تحتاج الشرطة إلى إيذاء الأشخاص في بعض الحالات مثل القبض على مرتكبي الجرائم العنيفة. مثال على ذلك ما ورد في صحيفة الغاردن سنة 2016، من استخدام الشرطة الأمريكية لأول مرة غي التاريخ روبوتًا في استعراض للقوة المميتة. حيث استخدمت شرطة دالاس روبوتًا لإزالة القنابل مزودًا بجهاز متفجر على ذراعه لقتل مشتبه به بعد مقتل خمسة من ضباط الشرطة وإصابة سبعة آخرين. مما آثار قلق الخبراء القانونيين[12]. كان البشر يتحكمون في هذا الروبوت بالذات. في المجال العسكري، كانت الروبوتات القادرة على القيام بأعمال عنيفة ضد البشر موجودة منذ بعض الوقت. ومع ذلك، وبصرف النظر عن العنف القانوني الذي تمارسه الدولة، هناك حالات قليلة نسبيًا يشعر فيها الناس بالارتياح إزاء تصميم الروبوتات والذكاء الاصطناعي لإيذاء البشر.

ترتبط الفلسفة الأخلاقية ارتباطًا مباشرًا بالاستقلالية. إذا لم يكن لدى الشخص الحكم الذاتي أو الإرادة الحرة، فيمكن القول بأن هذا الشخص ليس لديه مسؤولية أخلاقية أيضا. من وجهة نظر أخلاقية، يمكن إثبات العلاقة بين الاستقلالية والأخلاق بشكل أفضل من خلال النظرية الأخلاقية لإيمانويل كانط والتي تتكون من ثلاث صيغ رئيسية:

1. افعل كما لو كان على مُسلَّمة فِعلِك أن ترتفع عن طريق إرادتك إلى قانونٍ طبيعي عام[13].

2. افعل الفعل بحيث تُعامل الإنسانية في شخصك وفي شخص كل إنسان سواك بوصفها دائمًا وفي نفس الوقت غايةً في ذاتها، ولا تعاملها أبدًا كما لو كانَت مجرد وسيلة[14].

3. ينتج عن هذا المبدأ العملي الثالث للإرادة بوصفه الشرط الأعلى لموافقته للعقل العملي العام؛ أعني فكرة الإرادة عند كل كائنٍ عاقل كإرادةٍ تضع تشريعًا عامّا[15].

تخبرنا صيغة القانون العالمي أنه لتبرير فعل ما أخلاقيا، يجب على المرء أن يكون قادرا على تعميم القاعدة الأخلاقية أو "المبدأ". إحدى الطرق للتفكير في هذا هي طرح السؤال: ماذا لو اتبع الجميع هذه القاعدة؟ كيف سيبدو العالم حينها؟ يخبرنا مبدأ الإنسانية أننا يجب أن نحترم أهداف (أهداف، رغبات) الأشخاص الآخرين. لا ينبغي لنا أن نتعامل مع البشر الآخرين على أنهم "مجرد وسائل" فقط. بل يجب علينا أن نفكر في الغايات التي لديهم وكذلك غاياتنا.

كما صاغ كانط فلسفته الأخلاقية في ما يسمى بنسخة "مملكة الغايات" من الأمر الذي ينص على ما يلي: "يجب على الكائن العاقل أن يَعُد نفسه دائمًا مُشرِّعًا في مملكةٍ الغايات مُمكنةٍ عن طريق حرية الإرادة[16]". يخبرنا مبدأ الاستقلالية وصياغة مملكة الغايات أنه يجب علينا أن نسير في حديثنا الأخلاقي بأنفسنا. وهذا يعني أننا ملزمون بإطاعة القواعد الأخلاقية التي نتوقع من الآخرين اتباعها. يرى كانط أننا، ككائنات مستقلة، ملزمون بالنظر بشكل عقلاني في معاييرنا الأخلاقية. إن مجرد اتباع القانون ليس جيدًا بما فيه الكفاية. ويترتب على ذلك أن نظام الذكاء الاصطناعي يجب أن يتمتع بالاستقلالية بالمعنى الكانطي حتى يتمكن من التصرف بطريقة أخلاقية.

يجمع العديد من الكتاب بين المسؤولية الأخلاقية واتخاذ القرار الأخلاقي معًا في تعريفاتهم لما هو "الوكيل الأخلاقي". ويفصل البعض بين هاتين الفكرتين، حيث يرى أن الذكاء الاصطناعي يمكنه اتخاذ قرارات أخلاقية دون أن يكون مسؤولاً عن مثل هذه القرارات. لأن " الروبوتات الحالية لا تملك إرادة حرة. وبالتالي، لا يمكن تحميل الروبوتات المسؤولية الأخلاقية عن أفعالها بنفس الطريقة التي يتحملها البشر بموجب القانون[17]". وفقًا لمفهوم كانط، فإن النظام الذي تمت برمجته ليتبع ببساطة قواعد مثل قوانين أسيموف الثلاثة للروبوتات لا يمكن اعتباره عاملًا أخلاقيًا.

4.4 العدالة

ينص مبدأ العدالة على أن الذكاء الاصطناعي يجب أن يتصرف بطريقة عادلة وغير متحيزة. غالبًا ما يتم توضيح العدالة من خلال تمثال للإلهة الرومانية جوستيتيا. في كثير من الأحيان، يتم تصويرها بالسيف والمقاييس ومعصوبة العينين. عصابة العينين تمثل الحياد. والموازين تمثل وزن الأدلة. السيف يمثل العقوبة.

يمثل تعريف "العدالة" على المستوى البشري تحديًا كبيرًا للذكاء الاصطناعي. المشكلة الرئيسية التي تواجه الذكاء الاصطناعي هي أن النظرية الأخلاقية موضع جدل شديد. يقول ستيوارت مل "أسس الأخلاق كان يعتبر أهم مشكلة قي التفكر النظري الذي شغل بال المفكرين الموهوبين، فقسمهم طوائف ومدارس وأعلن بينهم حربا شعواء بعضهم ضد بعض، وبعد أكثر من ألفي سنة تواصلت نفس النقاشات والفلاسفة مازالوا تخت نفس الراية المتقابلة ولا أحد من هؤلاء المفكرين ولا الجنس البشري بدا قادرا على تحقيق الإجماع حول هذا الموضوع، مثلما كان الشاب سقراط يصغي إلى الشيخ بروتاغوراس[18]". يُظهر الاستطلاع الذي أجراه دفيد بورجيه David Bourget ودفيد تشالمرز David J. Chalmers أن أياً من المدارس الكبرى للنظرية الأخلاقية لا تتمتع بدعم أغلبية قوي. حوالي ربع الفلاسفة "يقبلون" أو "يميلون" إما إلى "نظرية الواجب" أو "المذهب النفعي"، وحوالي الثلث يقبل أو يميل نحو أخلاق الفضيلة. وبالتالي فإن تعريف "العدالة" أو "الأخلاق" بشكل عام من حيث ما يمكن للآلات معالجته هو أمر صعب. لا يوجد اتفاق على النظرية الأخلاقية[19].

وعلى النقيض من ذلك، يتمتع البشر "بالحدس الأخلاقي" وقد تعلموا بعض الأشياء عن الصواب والخطأ على مدار عدد من السنوات. كثيراً ما يقول البعض أن الحدس الأخلاقي البشري هو "صندوق أسود" للشفرة البيولوجية "المبهمة". نحن لا نفهم تمامًا كيف يتخذ البشر قرارات أخلاقية. نحن لا نفهم حتى كيف تقوم العقول البشرية بتخزين المعلومات. ومع ذلك، في حين أن هناك القليل من الاتفاق على النظرية الأخلاقية، عندما يتعلق الأمر بالممارسة الأخلاقية، هناك العديد من الأفعال التي تعتبر بشكل عام "صحيحة" والعديد منها تعتبر "خاطئة". وبينما تحتدم الخلافات الأخلاقية حول موضوعات مثل الإجهاض، والقتل الرحيم، والعصيان المدني، وعقوبة الإعدام، هناك العديد من المشاكل الأخلاقية الأقل صعوبة بكثير. إذا تم تقليص نطاق التطبيق وإمكانية الحصول على المعلومات التي يتم على أساسها اتخاذ القرارات الأخلاقية، فمن الممكن للذكاء الاصطناعي أن يتخذ بعض القرارات الأخلاقية المحددة للغاية. في كثير من الأحيان، هناك لوائح وقوانين واضحة يمكن اعتبارها نهائية من الناحية المعيارية في تطبيقات محددة للغاية. ومن الناحية العملية، أدى الذكاء الاصطناعي بالفعل إلى عدد من التطبيقات المنفذة التي قد تكون لها آثار معنوية وأخلاقية.

1.4.4 تحديد الإستحقاق الائتماني

تستخدم البنوك والمؤسسات الائتمانية الأخرى بالفعل، في عدد من البلدان، أنظمة الذكاء الاصطناعي لفرز طلبات الائتمان مسبقًا على أساس البيانات المتاحة حول مقدم الطلب. ومن المؤكد أن هذا له عدد من المزايا، أحدها هو القدرة على التوصل إلى قرار بسرعة أكبر وعلى أساس المزيد من المعلومات، مما يجعله أكثر عقلانية من الناحية النظرية. ومع ذلك، قد ينطوي هذا أيضًا على عيوب، ولا سيما ما يؤدي إلى بعض التحيزات. على سبيل المثال، ستحتوي المعلومات الشخصية لمقدم طلب الائتمان في معظم الحالات على معلومات حول الحي الذي يقيم فيه. وعلى هذا الأساس،

وبالاستفادة من البيانات الإضافية المتاحة للعامة (أو المجمعة بشكل خاص)، قد يحدث تحيز منهجي ضد أشخاص من مناطق سكنية معينة.

يُظهر الفحص المتعمق للتحيز العنصري المرتبط باستخدام الخوارزميات لإدارة الرعاية عالية المخاطر العديد من التحديات والقضايا المرتبطة بالمفاهيم المدمجة للعدالة وإنصاف الخوارزميات والدقة. حيث قام الباحثون بفحص المدخلات والمخرجات والوظيفة الموضوعية للخوارزمية المستخدمة لتحديد المرضى الذين لديهم خطر كبير في احتياجهم إلى رعاية حادة. لقد أظهرت الفحوص أن الخوارزمية المستخدمة على نطاق واسع، والتي تعتبر نموذجية لهذا النهج على مستوى الصناعة والتي تؤثر على ملايين المرضى، تظهر تحيزًا عنصريًا كبيرًا: عند درجة مخاطرة معينة، يكون المرضى السود أكثر مرضًا بكثير من المرضى البيض، كما يتضح من علامات الأمراض غير المنضبطة. ومن شأن معالجة هذا التفاوت أن يزيد نسبة المرضى السود الذين يتلقون مساعدة إضافية من 17.7 إلى 46.5%.

على الرغم من أن الخوارزمية تستبعد العرق على وجه التحديد من الاعتبار، فإن النظام يستخدم بشكل معقول معلومات حول تكاليف الرعاية الصحية للتنبؤ باحتياجات الرعاية الصحية. يستخدم النظام التعلم الآلي لإنشاء نموذج للتنبؤ بتكاليف الرعاية الصحية المستقبلية. ويفترض أن هؤلاء الأفراد الذين سيتحملون أعلى تكاليف الرعاية الصحية سيكونون نفس الأفراد الذين يحتاجون إلى أكبر قدر من الرعاية الصحية، ومع ذلك، فإن هذا الافتراض يقدم فوارق تنتهي في نهاية المطاف بالارتباط بالعرق. فعلى سبيل المثال، يواجه المرضى الفقراء تحديًا أكبر في الوصول إلى الرعاية الصحية لأنهم قد لا يستطيعون الوصول إلى وسائل النقل أو رعاية الأطفال أو لديهم متطلبات منافسة متعلقة بالعمل. وخلصوا إلى أن القضية المركزية هي صياغة المشكلة: التحدي المتمثل في تطوير خوارزميات حسابية دقيقة تعمل على مفاهيم غير متبلورة. ومن المحتم أن تتضمن أنواع التدابير الدقيقة اللازمة لمثل هذه الخوارزميات تشوهات تعكس غالبا التفاوتات البنيوية والعوامل المرتبطة بها. قد تكون هذه المشكلات مستوطنة بين العديد من خوارزميات الصناعة في العديد من الصناعات[20].

2.4.4 استخدام الذكاء الاصطناعي في المحكمة

يتم في بعض البلدان استخدام برامج الذكاء الاصطناعي في المحاكم. أحد الأمثلة لذاك هو استخدام تحديد الترتيب الذي يتم بموجبه عرض القضايا على القاضي، والاستفادة من المعلومات حول خطورة القضايا، والإدانات السابقة، والمزيد، من أجل جعل عمل المحكمة أكثر كفاءة. الاستخدام الأكثر تأثيرًا هو دعم القضاة لتحديد ما إذا كان سيتم إطلاق سراح السجين تحت المراقبة أم لا. وجدت دراسة أجرتها ProPublica[21] عام 2016 أن نظام COMPAS[22] أظهر تحيزًا منهجيًا ضد المتهمين الأمريكيين من أصل أفريقي في مقاطعة بروارد بولاية فلوريدا من حيث تقييم خطر العودة إلى الإجرام. أثارت هذه القضية جدلا كبيرا. جادل مطورو COMPAS، في ردهم على التحليل الإحصائي لـ ProPublica بأنها " ركز على إحصائيات التصنيف التي لم تأخذ في الاعتبار المعدلات الأساسية المختلفة لعودة الإجرام بين السود والبيض. وقد أدى استخدامهم لهذه الإحصائيات إلى تأكيدات كاذبة في مقالتهم، والتي تكررت لاحقًا في المقابلات وفي المقالات في وسائل الإعلام الوطنية[23]". وهذه حجة تقنية للغاية. لاحظ العديد من المعلقين أن هناك مفاهيم متعددة لـ

"العدالة" في أدبيات التعلم الآلي. مع الإشارة بشكل خاص إلى معدلات العودة إلى الإجرام في مقاطعة بروارد، فقد ثبت رياضيًا أن "أي أداة تحقق التكافؤ التنبؤي عند عتبة معينة يجب أن يكون لديها معدلات أخطاء إيجابية كاذبة أو سلبية كاذبة غير متوازنة عند تلك العتبة. لفهم سبب استبعاد التكافؤ التنبؤي وتوازن معدل الخطأ في تحديد معدل انتشار العودة إلى الإجرام غير المتكافئ، فمن المفيد التفكير في كيفية ارتباط هذه الكميات جميعًا[24]".

وفي كثير من هذه الحالات، لا يكون من السهل وضع التدابير المضادة. هناك مفاهيم متعددة للعدالة والتي قد تتعارض. وقد يواجه المرء مقايضة محتملة بين العدالة من ناحية، والدقة من ناحية أخرى. في الواقع، يرى البعض أن الجهود الرامية إلى "حجب" الخوارزميات عن المعلومات الفئوية المرفوضة، مثل العرق، قد تكون ضارة وأن النهج الأفضل هو تغيير كيفية استخدامنا للتعلم الآلي والذكاء الاصطناعي[25]. بينما يريد المرء بالتأكيد تجنب التحيز المنهجي في برمجة الذكاء الاصطناعي، يجب أن تعكس البيانات أيضًا الوضع الفعلي، أي "ما هو الحال"، وإلا فإن هذه البيانات لن تكون مفيدة جدًا لاستخلاص المزيد من الاستنتاجات والتصرف على أساسها.

ومن الأهمية بمكان أن نقدر حدود التصنيفات القائمة على النماذج الإحصائية. في حين أن الكثير من الجدل حول العودة إلى الإجرام في مقاطعة بروارد تركز على الفرق بين النتائج الإيجابية الكاذبة بين السود والبيض، فقد وجدت إحدى الدراسات أن على أن الدقة التنبؤية الأساسية لنظام COMPAS كومباس كانت حوالي 65٪ فقط. ولوحظ أن هذه التوقعات ليست دقيقة كما كان يأمل، لا سيما من وجهة نظر المدعى عليه الذي يقع مستقبله على الميزان[26].

4.5 القابلية للتفسير

ينص مبدأ القابلية للتفسير على أنه من الممكن تفسير سبب وصول نظام الذكاء الاصطناعي إلى نتيجة أو نتيجة معينة. لا ينبغي مساواة القابلية للتفسير بالشفافية. في حين يدعو البعض إلى أقصى قدر من الشفافية للبرامج والرموز، عندما يتعلق الأمر بأنظمة الذكاء الاصطناعي، فإن هذا قد لا يحل مشكلة وقد يخلق مشاكل جديدة. لنفترض أن البرنامج الذي يحتوي على ملايين الأسطر من التعليمات البرمجية أصبح شفافًا، فما الفائدة من ذلك؟ أولاً، ربما لن يكون البرنامج واضحًا لغير الخبراء، وحتى الخبراء سيواجهون صعوبة في فهم ما يعنيه ذلك. ثانياً، قد تؤدي الشفافية القصوى للبرمجيات إلى خلق خطر في مواجهة المنافسين، وتعيق المزيد من الاستثمار في هذا القطاع. ونظراً لاعتبارات كهذه، تحولت بعض أجزاء المناقشة إلى مفهوم "القابلية للتفسير".

القابلية للتفسير، كما يرى فلوريدي " تُفهم على أنها تتضمن كلا من الوضوح والمساءلة[27]". من المهم في التطبيقات الأخلاقية أن يتمكن أولئك الذين يستخدمون أنظمة الذكاء الاصطناعي أو الذين تتأثر مصالحهم بأنظمة الذكاء الاصطناعي من "فهم" الطريقة الدقيقة التي يتخذ بها الذكاء الاصطناعي قرارًا معينًا. الوضوح يعني أن طريقة عمل الذكاء الاصطناعي يمكن أن يفهمها الإنسان. النظام ليس "صندوقا أسود" غامضا لا يمكن فهم مكوناته الداخلية. يستطيع شخص ما، حتى لو كان مبرمجًا ذا خبرة، أن يفهم طريقة عمل النظام ويشرحها للقضاة وهيئات المحلفين والمستخدمين.

نفذ الاتحاد الأوروبي مطلبًا قانونيًا بشأن "الحق في الحصول على المعلومات" (الذي كان يُسمى في الأصل "الحق في التوضيح") في إطار اللائحة العامة لحماية البيانات. وهذا يعني أن الأشخاص الذين تأثرت اهتماماتهم بقرار خوارزمي، لديهم الحق في أن تشرح لهم الخوارزمية القرار. من الواضح أن هذا يشكل تحديًا لبعض أساليب التعلم الآلي "المبهمة" مثل الشبكات العصبية. هناك بعض الأشخاص الذين لا ينزعجون من "الغموض" في التعلم الآلي. ويرون أنهم قادرون على اختبار النظام تجريبيا. وطالما أنها تعمل في الممارسة العملية، فإنهم لا يهتمون إذا لم يتمكنوا من فهم أو شرح كيفية عملها في الممارسة العملية [28].

بالنسبة للعديد من تطبيقات التعلم الآلي، قد يكون هذا جيدًا. ومع ذلك، في المواقف المشحونة أخلاقيا والتي قد تخضع للمراجعة القضائية، سيكون هناك شرط لتفسير القرار. قد يكون هناك أيضًا شرط لتبرير القرار. إن العنصر الأساسي في الأداء الأخلاقي ليس مجرد القيام بالشيء الصحيح، بل إن تبرير ما فعله الفاعل هو الصواب. ولا يمكن أن يرتكز التبرير الأخلاقي على صندوق أسود "غامض". ومع ذلك، هناك بحث مستمر حول "الذكاء الاصطناعي القابل للتفسير" والذي يسعى إلى توليد تفسيرات لسبب اتخاذ الشبكات العصبية للقرارات التي تتخذها. وربما يمكّن مثل هذا البحث في نهاية المطاف التعلم الآلي من توليد تفسيرات كافية للقرارات التي يتخذها.

ومع ذلك، ومن الناحية العملية، فإن استخدام نظام كومباس لتقييم مخاطر العودة إلى الإجرام وبالتالي التأثير على احتمالات المراقبة قد تم اختباره في المحكمة. وفي قضية لوميس ضد ويسكونسن، زعم المدعي أنه حُرم من "الإجراءات القانونية الواجبة" بسبب طبيعة ملكية خوارزمية كومباس التي تعني أن دفاعه لا يستطيع تحدي الأساس العلمي الذي تم حساب نتيجته عليه. إلا أن مناشداته باءت بالفشل. رأى القضاة أن قرارات الأحكام لم تكن مبنية بالكامل على درجات المخاطر الخاصة بنظام كومباس. لا يمكن للقضاة أن يبنوا الأحكام على درجات المخاطر وحدها، ولكن يمكنهم أن يأخذوا في الاعتبار هذه الدرجات إلى جانب عوامل أخرى في تقييمهم لخطر العودة إلى الإجرام.

يمكن توفير المساءلة في أبسط صورها في شكل ملفات سجل. على سبيل المثال، يتعين على شركات الطيران التجارية أن يكون لديها جهاز تسجيل طيران يمكنه النجاة من أي حادث[29].

في حالة تحطم طائرة، يمكن استعادة مسجلات الرحلة. أنها تسهل التحقيق في الحادث من قبل السلطات. تسمى مسجلات الطيران أحيانًا "الصناديق السوداء". ملفات السجل التي أنشأها النظام والتي يمكن أن تشرح سبب قيامه بما فعله. غالبًا ما تُستخدم هذه البيانات لتتبع الخطوات التي اتخذها النظام للوصول إلى السبب الجذري، وبمجرد العثور على هذا السبب الجذري، قم بإلقاء اللوم عليه.

5. خاتمة:

خلاصة القول، أنه مع حرية الاختيار، سيقبل المستخدمون الأنظمة إذا كانوا يثقون بها، ويجدونها مفيدة ويستطيعون تحمل تكاليفها. ولذلك فإن الشركات لديها حافز لجعل الأنظمة يثق بها الناس. تتضمن الثقة مجموعة معقدة للغاية من المفاهيم.

فلكي يثق المستخدمون في النظام، عليهم أن يكونوا واثقين من أن النظام سيفيدهم،أي أنه سيظهر الإحسان إليهم. سيحتاجون إلى أن يكونوا واثقين من أن النظام لن يفعل أشياء سيئة لهم مثل إيذائهم أو السرقة منهم أو الإضرار بمصالحهم بطريقة أخرى (على سبيل المثال انتهاك خصوصيتهم أو التسبب في إحراجهم).

يجب أن يكونوا واثقين من أن الذكاء الاصطناعي لن يعرض استقلاليتهم للخطر. هذا لا يعني أن الروبوتات والذكاء الاصطناعي لن يقولوا أبدًا لا للبشر. إنه مجرد القول بأنهم لن يقولوا "لا" إلا إذا كان هناك سبب وجيه. على سبيل المثال، إذا أراد الإنسان أن يفعل شيئًا خاطئًا مع الروبوت، فيمكن للروبوت الذكي أخلاقيًا (نظريًا) أن يرفض طاعة الإنسان.

سيحتاج الناس إلى أن يكونوا واثقين من أن النظام يمكنه التصرف بشكل عادل في نطاقه الوظيفي. إذا كانت هناك قوانين وأنظمة واضحة ولم تكن هناك خلافات أخلاقية، فسيكون تنفيذها أسهل بكثير. وفي الوقت الحالي، وبسبب عدم الاتفاق على النظرية الأخلاقية، فإن هذه النطاقات الوظيفية ضيقة. ومع ذلك، هناك العديد من التطبيقات العملية المشحونة أخلاقيًا والتي يتم التعامل معها بشكل مناسب بواسطة الذكاء الاصطناعي.

وأخيرًا، لكي نثق في الآلات، يجب أن تكون قابلة للتفسير. ولأسباب تتعلق بالوضوح والمساءلة، ستحتاج أنظمة الذكاء الاصطناعي إلى الاحتفاظ بالسجلات والتفسيرات حول سبب قيامها بما تفعله.

5. قائمة المراجع:

أ– بالعربية:

- إيمانويل كانط، **نقد العقل العملي**، المنظمة العربية للترجمة، ط1، بيروت–لبنان، 2008.
- إيمانوئيل كانط ، **تأسيس ميتافيزيقا الأخلاق**، مؤسسة هنداوي، 2020.
- جون ستيوارت ميل، **النفعية**، ط1، المنظمة العربية للترجمة، بيروت–لبنان، 2012.

ب– بالأجنبية:

- Alexandra Chouldechova, Fair predictionwith disparate impact Astudy of bias, **recidivism**, 2017.
- Andreas Kaplan, and Michael Haenlein, Siri, siri, in my hand: Who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence, **Business Horizons**, 2018.
- David Bourget, and David J. Chalmers. What do philosophers believe **Philosophical Studies**, N° 03, Vol. 170, 2014.
- David Weinberger, Don't make AI artificially stupid in the name of transparency. **Wired**, 2018

- J.H Moor, The nature, importance and difficulty of machine ethics, **Intelligent Systems**, 21 (4) 2006.
- John. R. Searle, Minds, brains, and programs. **Behavioral and Brain Sciences 3**, Cambridge University Press, 1980.
- Jon, Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan. **Algorithmic airness**, Aea papers and proceedings.
- Julia Dressel, and Hany Farid, The accuracy, fairness, and limits of predicting recidivism The accuracy, fairness, and limits of predicting recidivism.**Science Advances 4** (1).
- Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena. Ai4people an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations. **Minds and Machines**, 28 (4), 2018.
- S.Woolley, Children of Jehovah's witnesses and adolescent Jehovah's witnesses: what are their rights? **Archives of Disease inChildhood** 90 (7), 2005: 715–719.
- Sam Thielmann, Use of police robot to kill dallas shooting suspect believed to be first in us history. **The Guardian**, 2016, https://www.theguardian.com/technology/2016/jul/08/police-bomb-robot-explosive-killed-suspect-dallas
- Sean Welsh, **Ethics and security automata**, First published, Routledge, New York 2018.
- Stuart J.Russell, and Peter Norvig, **Artificial intelligence**: **a modern approach**, Fourth Edition, Pearson Education Limited, United Kingdom, 2022.
- William Dieterich, Christina Mendoza, and Tim. Brennan, **Compas risk scales**: : **Demonstrating accuracy equity and predictive parity**, Northpointe Inc, 2016.
- Ziad Obermeyer, Brian Powers,Christine Vogeli, and Sendhil Mullainathan, Dissecting racial bias in an algorithm used to manage the health of populations. **Science**, 366 (6464), 2019.

-David L. Poole, and Alan K. Mackworth, **Artificial intelligence Foundations**, Cambridge University Press, New York, 2010.

6. الهوامش:

[1] - Andreas Kaplan, and Michael Haenlein, Siri, siri, in my hand: Who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence, **Business Horizons**, 2018, p16.

[2] -David L. Poole, and Alan K. Mackworth, **Artificial intelligence Foundations**, Cambridge University Press, New York, 2010, p 3

[3] - Ibid, p 4

[4] - Ibid, p 4

[5] - Stuart J.Russell, and Peter Norvig, **Artificial intelligence**: **a modern approach**, Fourth Edition, Pearson Education Limited, United Kingdom, 2022, pp 7-8.

[6] - John. R. Searle, Minds, brains, and programs. **Behavioral and Brain Sciences 3**, Cambridge University Press, 1980, pp417-457.

[7] – إيمانويل كانط، **نقد العقل العملي**، المنظمة العربية للترجمة، ط1، بيروت-لبنان، 2008، ص 781

[8] - J.H Moor, The nature, importance and difficulty of machine ethics, **Intelligent Systems**, 21 (4) 2006, p 18.

[9] – AI4People هو المنتدى الأوروبي الأول الذي يجمع أصحاب المصلحة الرئيسيين مثل الأوساط الأكاديمية والصناعة ومنظمات المجتمع المدني والبرلمان الأوروبي لوضع الأسس لـ "مجتمع الذكاء الاصطناعي الجيد" الذي يشكل تأثير الذكاء الاصطناعي.

[10] - Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena. Ai4people an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations. **Minds and Machines**, 28 (4), 2018: 689–707.

[11] - S.Woolley, Children of Jehovah's witnesses and adolescent Jehovah's witnesses: what are their rights? **Archives of Disease inChildhood** 90 (7), 2005: 715–719.

[12] - Sam Thielmann, Use of police robot to kill dallas shooting suspect believed to be first in us history. **The Guardian**, 2016, https://www.theguardian.com/technology/2016/jul/08/police-bomb-robot-explosive-killed-suspect-dallas

[13] – إيمانوئيل كانط ، **تأسيس ميتافيزيقا الأخلاق**، مؤسسة هنداوي، 2020، ص 61.

[14] – المرجع نفسه، ص 70.

[15] – المرجع نفسه، ص 73.

[16] – المرجع نفسه، ص 76

[17] - Sean Welsh, **Ethics and security automata**, First published, Routledge, New York 2018.p33.

[18] – جون ستيوارت ميل، **النفعية**، ط1، المنظمة العربية للترجمة، بيروت–لبنان، 2012، ص27.

[19] - David Bourget, and David J. Chalmers. What do philosophers believe **Philosophical Studies**, N° 03, Vol. 170, 2014, pp. 465-500

[20] - Ziad Obermeyer, Brian Powers,Christine Vogeli, and Sendhil Mullainathan, Dissecting racial bias in an algorithm used to manage the health of populations. **Science**, 366 (6464), 2019, pp447–453.

[21] - منصة صحفية غير ربحية مقرها في الولايات المتحدة. تأسست في عام 2007 وتعتبر من أوائل وسائل الإعلام الرقمية التي تهتم بالصحافة التحقيقية على مستوى وطني. تتخذ من الصحافة التحقيقية والبيانات الصحفية مدخلاً رئيسيًا في تقاريرها. وتهدف إلى كشف الفساد والظلم والممارسات غير الأخلاقية في القطاعين العام والخاص، وتعزيز الشفافية والمساءلة.

[22] - هو أداة برمجية تستخدم في النظام القانوني في الولايات المتحدة. تم تصميمها لمساعدة القضاة ولجان الإفراج عن السجناء في اتخاذ قرارات حول العقوبات والإفراج من خلال تقييم خطر العودة إلى الجريمة (احتمالية ارتكاب جريمة جديدة) للأفراد الذين أدينوا بجرائم.

[23] - William Dieterich, Christina Mendoza, and Tim. Brennan, **Compas risk scales**: : **Demonstrating accuracy equity and predictive parity**, Northpointe Inc, 2016, p1.

[24] - Alexandra Chouldechova, Fair predictionwith disparate impact Astudy of bias, **recidivism**, 2017 P 5.

[25] - Jon, Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan. **Algorithmic airness**, Aea papers and proceedings 108: pp22–27.

[26] - Julia Dressel, and Hany Farid, The accuracy, fairness, and limits of predicting recidivism The accuracy, fairness, and limits of predicting recidivism.**Science Advances 4** (1), 208. P 3.

[27] - Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena. Ai4people–an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations, Op Cit, p 696

[28] - David Weinberger, Don't make AI artificially stupid in the name of transparency. **Wired**, 2018

[29] - Op Cit.