تدقيق التحيز في الذكاء الاصطناعي في ضوء اطار عمل تدقيق الذكاء الاصطناعي لمعهد المدققين الداخليين (IIA) - دراسة نظرية تحليلية

Auditing bias in artificial intelligence in light of the framework of artificial intelligence auditing Institute of Internal Auditors (IIA) - Analytical Theoretical Study

ناظم حسن رشيد [1]♦، مي ابلحد أفرام [2]

[1] كلية الإدارة والاقتصاد، جامعة الحمدانية (العراق)، dr.nadhim1962@uohamdaniya.edu.iq

[2] كلية الإدارة والاقتصاد، جامعة الحمدانية (العراق)، may-a-ablahad@uohamdaniya.edu.iq

تاريخ الاستلام: 2022/12/08 تاريخ القبول: 2022/12/30 تاريخ النشر: 2023/01/15

ملخص:

يهدف البحث إلى التعرف على مفهوم التحيز في الذكاء الاصطناعي، وعلى إطار عمل تدقيق الذكاء الاصطناعي لمعهد المدققين الداخليين(IIA)، ولفت نظر المدققين الداخليين الى أهمية تدقيق تحيز الذكاء الاصطناعي من أجل تقليل المخاطر المتعلقة بالتحيز المحتمل للخوارزميات.

ولتحقيق أهداف البحث قام الباحثان باستخدام المنهج الاستدلالي في الدراسة والتحليل من خلال الاستعانة بالأطاريح والرسائل الجامعية والدوريات والكتب والمواقع الالكترونية التي تتناول موضوع الدراسة ولا سيما فيما يتعلق بمجالات: الذكاء الاصطناعي والتحيز في الخوارزميات ودور التدقيق الداخلي في تدقيق التحيز في الذكاء الاصطناعي في ضوء اطار عمل تدقيق الذكاء الاصطناعي لمعهد المدققين الداخليين(IIA) .

الكلمات المفتاحية: الذكاء الاصطناعي، التحيز في الذكاء الاصطناعي، الخوارزميات، الصندوق الاسود.

Abstract:

The research aims to identify the concept of bias in artificial intelligence, and the framework of artificial intelligence auditing of the Institute of Internal Auditors (IIA), and to draw the attention of internal auditors to the importance of auditing artificial intelligence in order to reduce the risks related to the potential bias of algorithms.

To achieve the objectives of the research, the researchers used the constructive approach in the study and analysis through the use of dissertations, theses, periodicals, books and Websites that deal with the subject of the study, especially with regard to the areas of: artificial intelligence and bias in algorithms and the role of internal audit in auditing bias in artificial intelligence in the light of the framework of artificial intelligence auditing. Institute of Internal Auditors (IIA)

Keywords: artificial intelligence, bias in artificial intelligence, algorithms, black box.

♦.المؤلف المرسل

المقدمة:

على مدى السنوات القليلة الماضية، بدأ المجتمع يتصارع مع مدى قدرة التحيزات البشرية، مع عواقبها المدمرة، على إيجاد طريقها من خلال أنظمة الذكاء الاصطناعي. إن إدراك هذه التهديدات بعمق والسعي لتقليلها هو أولوية ملحة عندما تتطلع العديد من الشركات إلى نشر حلول الذكاء الاصطناعي. يمكن أن يتخذ التحيز الخوارزمي في أنظمة الذكاء الاصطناعي أشكالاً متنوعة مثل التحيز الجنسي والتحيز العنصري والتمييز على أساس السن.

هناك تركيز متزايد على التحيز في الذكاء الاصطناعي، وعلى الرغم من عدم وجود سبب للذعر حتى الآن، فإن بعض القلق يعد معقولاً. يتم تضمين الذكاء الاصطناعي في أنظمة من الجدار إلى الجدار هذه الأيام، وإذا كانت هذه الأنظمة متحيزة، فإن نتائجها تكون كذلك. قد يفيدنا هذا أو يضرنا أو يفيد شخصاً آخر.

القضية الرئيسة هي أن التحيز نادراً ما يكون واضحاً. جميع البيانات متحيزة. هذه حقيقة قد لا يكون التحيز متعمداً. قد يكون لا مفر منه بسبب الطريقة التي يتم بها إجراء القياسات، ولكن هذا يعني أنه يجب علينا تقدير الخطأ (فترات الثقة) حول كل نقطة بيانات لتفسير النتائج.

لا يمكن لمهنة التدقيق الداخلي أن تقف على أمجادها خوفاً من التخلف عن الركب فيما قد تكون الثورة الرقمية القادمة –الذكاء الاصطناعي حاضرة، يجب أن يفهم المدققون الداخليون أساسيات الذكاء الاصطناعي، والدور الذي يجب أن يلعبه التدقيق الداخلي فيه ، فضلاً عن الفوائد والمخاطر الأساسية. يمكننا مساعدة وظائف التدقيق الداخلي في التغلب على تحديات الذكاء الاصطناعي وخاصة ما يتعلق بالتحيز من خلال إعادة ضبط خططهم وخبراتهم ومنهجياتهم لتلبية ذلك. سيضمن ذلك أنهم في وضع أفضل لإجراء تقييم نقدي لفعالية إدارة مخاطر الذكاء الاصطناعي وعمليات الرقابة وغيرها.

أولاً: مشكلة البحث

وتتمثل المشكلة الاساسية للبحث في تساؤل الرئيسي وهو: هل يمكن للتدقيق الداخلي القيام بتدقيق التحيز في الذكاء الاصطناعي؟ ويتفرع من هذا التساؤل الاسئلة الفرعية التالية:

1. ما هو السبب الجذري لإدخال التحيز في أنظمة الذكاء الاصطناعي، وكيف يمكن منعه؟ في أشكال عديدة، قد يتغلغل التحيز في الخوارزميات؟

2. مع استمرار المؤسسات في تبني الذكاء الاصطناعي(AI) ، هل من المتوقع أن يقدم التدقيق الداخلي تأكيداً على تطبيقه واستخدامه وعدم تحيزه؟

3. هل يمتلك فريق التدقيق الداخلي معاييراً وإطاراً مناسباً لتدقيق الذكاء الاصطناعي لتقييم المشكلات بفعالية بخصوص التحيز في الذكاء الاصطناعي؟

4. إلى أي مدى يجب تدقيق خوارزميات الذكاء الاصطناعي وهل أن الذكاء الاصطناعي متوافق مع اللوائح المعمول بها وغير متحيز؟

ثانياً: أهمية البحث

تكمن أهمية الموضوع الذي تتناوله هذا البحث في كونه موضوعاً حيوياً ومهماً من موضوعات الممارسات المهنية للتدقيق الداخلي بالوصف والتحليل لدور التدقيق الداخلي في تدقيق التحيز في الذكاء الاصطناعي وفقاً لاطار عمل تدقيق الذكاء الاصطناعي لمعهد المدققين الداخليين (IIA) باعتبار إن الذكاء الاصطناعي يتعلق بالقدرة على التفكير الفائق وتحليل البيانات أكثر من تعلقه بشكل معين أو وظيفة معينة ودوره في حل المشكلات واتخاذ القرارات بأسلوب منطقي وبطريقة تفكير العقل الانساني نفسها, الأمر الذي يجعل المدققين الداخليين أمام ضرورة تدقيق خوارزميات الذكاء الاصطناعي من أجل تقليل المخاطر المتعلقة بالتحيز المحتمل للخوارزميات ، فضلاً عن تقديم المشورة للشركة لمراقبة أداء الخوارزمية بانتظام ضد مجالات التمييز المحتملة.

وعليه تكمن الأهمية العلمية لهذا البحث في حداثة موضوعه، إذ لم يتم تناوله من قبل الباحثين بشكل كاف والحاجة للبحث فيه، وبالتالي فإن هذا البحث يعد من المحاولات العلمية الهادفة لإثراء النتاج الفكري حول هذا الموضوع.

ثالثاً: أهداف البحث

1. التعرف على الإطار النظري للتحيز في الذكاء الاصطناعي.
2. التعرف على اطار عمل تدقيق الذكاء الاصطناعي لمعهد المدققين الداخليين (IIA) .
3. تحديد أهداف وأنشطة الأخلاق والتحيز ذات الصلة بتدقيق التحيز في الذكاء الاصطناعي.

4. لفت نظر المدققين الداخليين الى أهمية تدقيق تحيز الذكاء الاصطناعي من أجل تقليل المخاطر المتعلقة بالتحيز المحتمل للخوارزميات.

5. دراسة نظرية تحليلية لدور التدقيق الداخلي في تدقيق التحيز في الذكاء الاصطناعي وفقا لاطار عمل تدقيق الذكاء الاصطناعي لمعهد المدققين الداخليين (IIA) .

رابعاً: فرضيات البحث

ولتحقيق أهداف البحث تم الاعتماد على الفرضيات الآتية :

1. يمكن للتدقيق الداخلي القيام بتدقيق التحيز في الذكاء الاصطناعي وفقاً لاطار عمل تدقيق الذكاء الاصطناعي لمعهد المدققين الداخليين(IIA)

2. هناك دور للتدقيق الداخلي في تقديم ضمان بأن نتائج أنشطة الذكاء الاصطناعي الخاصة بالمنظمة خالية من التحيزات المحتملة للخوارزميات.

خامساً: منهجية البحث

يقوم البحث على استخدام المنهج الاستدلالي في الدراسة والتحليل من خلال الإستعانة بالأطاريح والرسائل الجامعية والدوريات والكتب التي تتناول موضوع الدراسة ولا سيما فيما يتعلق بمجالات: الذكاء الاصطناعي والتحيز في الخوارزميات ودور التدقيق الداخلي في تدقيق التحيز في الذكاء الاصطناعي في ضوء اطار عمل تدقيق الذكاء الاصطناعي لمعهد المدققين الداخليين(IIA) .

وعليه فقد تم تقسيم البحث على وفق الخطة الآتية:

المبحث الأول: الاطار النظري للتحيز في الذكاء الاصطناعي

المبحث الثاني: تدقيق الذكاء الاصطناعي –اعتبارات لمهنة التدقيق الداخلي

المبحث الثالث: تدقيق التحيز في الذكاء الاصطناعي وفقاً لاطار معهد المدققين الداخليين (IIA)

المبحث الأول: الاطار النظري للتحيز في الذكاء الاصطناعي

إذا كنت تعمل في منظمة أو مؤسسة ترغب في تحسين فعالية عمليات إدارة المخاطر والحوكمة أو ترغب فقط في تحسين الشركة بشكل عام، فمن المحتمل أنك على دراية بالتدقيق الداخلي.

التدقيق الداخلي هو "نشاط مستقل وتأكيد موضوعي ذو طبيعة استشارية يهدف إلى إضافة قيمة للشركة وتحسين عملياتها، ويساعد التدقيق الداخلي الشركة على تحقيق أهدافها من خلال انتهاج

مدخل موضوعي لتقويم وتحسين فعالية إدارة المخاطر وفعالية الرقابة وفعالية عملية إدارة حوكمة الشركات (IIA, 2012).

أولاً: مفهوم الذكاء الاصطناع وأبعاده

الذكاء الاصطناعي (AI) هو محاكاة لعمليات الذكاء البشري بواسطة أجهزة الحاسوب. السمة الرئيسة لنظام الذكاء الاصطناعي، هي أنه يتعلم من كل دورة تشغيل (حلقة تغذية مرتدة) ، وبالتالي يحسن أدائه (يصبح أكثر ذكاءً) خلال كل تكرار متتالي عن طريق تصحيح أخطائه وتحسينها) Mach, 2019).

ويطلق عليه ايضاً ذكاء الآلة يرمز إلى تكامل الذكاء الشبيه بالبشر في الآلات. الفكرة الأساسية في الذكاء الاصطناعي هي فهم السياق واتخاذ قرارات ذكية بناءً على المعلومات المتوفرة. ينظر إلى الذكاء الاصطناعي باعتباره مرادفاً للتكنولوجيا المعرفية أو الحوسبة المعرفية بمستوى الذكاء المناسب لأداء المهام المعرفية.

هناك تعاريف عديدة للذكاء الاصطناعي أهمها:

1. بأنه تلك التقنية المتطورة المستخدمة، والتي تسهم في إدارة العمليات والمهام بآليات أكثر تطوراً وذكاء من الإنسان الذي صنعها ومنحها المعرفة والمقومات الحسية، بما يساعدها على التعلم التلقائي والتطور الذاتي. (الجابر,18,2020)

2. عرّف الذكاء الاصطناعي بأنه برنامج كمبيوتر قادر على اتخاذ قرارات متوازنة بناءً على السياق الحالي. النتيجة الإجمالية لاستخدام مثل هذا النظام هي تعزيز أهداف القرار. لتحقيق هذا الإنجاز، يجب أن يكون نظام الذكاء الاصطناعي قادرا على محاكاة الإجراءات البشرية مثل تحديد الصورة. وأن التشغيل الصحيح للذكاء الاصطناعي يتطلب من النظام قدرة تشغيل عالية وكميات كبيرة من البيانات (Issa et al, 2016,1).

وعليه فان الذكاء الاصطناعي مصطلح واسع يشمل أنشطة مختلفة مثل التعرف على الأنماط بواسطة أجهزة الكمبيوتر والأنظمة الخبيرة والتعلم العميق والتفكير بواسطة أجهزة الكمبيوتر، استخدام اللغة الطبيعية بواسطة أجهزة الكمبيوتر وما شابه. يوصف الذكاء الاصطناعي أيضاً بأنه "برنامج كمبيوتر يمكنه اتخاذ قرارات متوازنة ومراقبة بيئته واتخاذ الإجراءات يزيد من فرصه في تحقيق الهدف.

يحلل الذكاء الاصطناعي كميات كبيرة من البيانات ويعالجها ويطور "خوارزمية."

الخوارزمية هي اسم خيالي لمجموعة من القواعد التي يجب أن تتبعها الآلة والتي تسمح للآلة بتولي المهام التي تستغرق وقتاً طويلاً والمتكررة والمملة التي قد يضطر المحترفون إلى أداؤها يومياً.

يؤدي ذلك إلى توفير الوقت وزيادة الإنتاجية، كفاءة العمل للمهنيين والمحاسبين الماليين والمدققين. يمكن لخوارزمية الذكاء الاصطناعي أيضاً معالجة كميات هائلة من البيانات بسرعة ودون تعب أو ارتكاب أخطاء بسبب قلة الانتباه.

خوارزمية الذكاء الاصطناعي هي عملية تعلم ذاتي وتصبح أكثر ذكاءً في تحديد أنماط معينة وتصبح أسرع في حل المشكلات، بمرور الوقت حيث تعالج صحة كل تكرار وتقوم بإجراء التعديلات قبل الدورة التالية من العملية.

في عالم تكنولوجي سريع ودائم التغير، يعد الذكاء الاصطناعي أداة فعالة يتم تبنيها بسرعة في العديد من الشركات والمؤسسات لتحسين بيئة العمل وحياة العديد من الأشخاص في جميع أنحاء العالم كل يوم. كل يوم يظهر المزيد من التحسن والتقدم الذكاء الاصطناعي هو المستقبل (Mach, 2019).

مع تطورات الروبوتات، والتعلم الآلي، والذكاء الاصطناعي، والتعلم العميق وما إلى ذلك، تستثمر معظم المنظمات في الحلول المتعلقة بالذكاء الاصطناعي لأنها تتيح لها تكليف المزيد من المهام بالآلات، وتحرير موظفيها للتركيز على الأولويات الاستراتيجية. وظائف التدقيق الداخلي ليست استثناءات من ذلك.

من خلال التدقيق الداخلي للذكاء الاصطناعي سيكون قادراً على تحسين اختبار التوقيت وزيادة الرؤية حيث يقوم الذكاء الاصطناعي بمراجعة البيانات وتحليلها على أساس مستمر.

مثل أي عملية تعتمد على الخوارزمية والبيانات، يقدم الذكاء الاصطناعي للتدقيق الداخلي دوراً واضحاً في ضمان الدقة والموثوقية (Omosa, 2020).

ثانياً: فهم الذكاء الاصطناعي بعمق أكبر

يهدف الذكاء الاصطناعي إلى جعل أجهزة الكمبيوتر متطورة بما يكفي لمحاكاة الوظائف الإدراكية البشرية. نتيجة لذلك، يعد الذكاء الاصطناعي مجالاً واسعاً يشمل رؤية الكمبيوتر ومعالجة اللغة والإبداع والتلخيص. التعلم الآلي هو تخصص الذكاء الاصطناعي الذي يتعامل مع الجوانب الإحصائية للذكاء الاصطناعي. إنه يدرب الكمبيوتر على حل المشكلات من خلال دراسة مئات أو آلاف الحالات، والتعلم منها وتطبيقها على الظروف الجديدة. التعلم العميق هو مجموعة فرعية من

التعلم الآلي الذي يسمح لأجهزة الكمبيوتر بالتعلم واتخاذ القرارات بشكل مستقل. مقارنة بمعظم طرق التعلم الآلي، يتضمن التعلم العميق مستوى أعلى من الأتمتة. (Kumari , 2022)

قدرات الذكاء الاصطناعي ليست مخفية. يمر ذكاء الآلات بتحول كبير من خلال تحسينات التعلم الذاتي المستمر. ومع ذلك، لا تزال نماذج الذكاء الاصطناعي هذه عبارة عن صندوق أسود. وغالباً ما يتم التشكيك في قراراتهم من قبل الزبائن. يتحرك البحث بشكل أسرع من أي وقت مضى في تحسين الخوارزميات وتحسينها، لكن هذا وحده لن يكفي. غالباً ما تكون المحادثات حول بناء الثقة في الذكاء الاصطناعي نقطة اهتمام للمطورين وفرق المبيعات والتسويق الذين يعملون مباشرة مع الزبائن لذلك ، من المهم النظر فيها (Mahmood, 2022).

تحتاج أنظمة الذكاء الاصطناعي إلى كميات كبيرة من البيانات لتزداد دقتها في تنفيذ المهمات المطلوبة، وبعد الجمع والتخزين يتم معالجة تلك البيانات بواسطة خوارزميات النظام الذكية حتى يتعلم النظام من أنماط وخصائص البيانات، ومن ثم تطوير قدرة الآلة على إنجاز المهام التي صُنعِت من أجلها .

يعد ضمان استخدام الذكاء الاصطناعي لإفادة الجميع تحدياً بالفعل، ومن الأهمية بمكان أن نحققه بالشكل الصحيح. عندما يصبح الذكاء الاصطناعي أكثر قوة، يتزايد تأثيره على اقتصادنا وسياستنا وثقافتنا. هذا من المحتمل أن يكون جيداً للغاية أو سيئاً للغاية. من ناحية أخرى، يمكن أن يساعدنا الذكاء الاصطناعي على تحقيق تقدم في العلوم والتكنولوجيا يسمح لنا بمعالجة أهم مشاكل العالم. من ناحية أخرى، يمكن أن تؤدي أنظمة الذكاء الاصطناعي القوية ولكن الخارجة عن السيطرة (نظم الذكاء الاصطناعي المنحرفة) إلى كارثة للإنسانية. نظراً للمخاطر، يُعد العمل نحو الذكاء الاصطناعي المفيد سبباً ذا أولوية عالية نوصي بدعمه، خاصة إذا كنت تهتم بحماية المستقبل على المدى الطويل.

ثالثاً: مفهوم التحيز في الذكاء الاصطناعي

على الرغم من التطور الضخم الذي يساهم فيه الذكاء الاصطناعي في الكثير من المجالات، إلاّ أن واحدة من أهم الاشكاليات التي ظهرت في السنوات السابقة هي عنصرية بعض تطبيقات الذكاء الاصطناعي على أساس جنس أو عرق معين؛ للدرجة التي تجعل بعض التطبيقات تنتهك حقوق الأفراد وتتحكم في مسار حياتهم

.( )https://masaar.net/ar/

تطور الذكاء الاصطناعي بسرعة خلال السنوات القليلة الماضية. قبل عقد من الزمن، كان الذكاء الاصطناعي مجرد مفهوم له عدد قليل من التطبيقات الواقعية، ولكنه اليوم أحد أسرع التقنيات نمواً ، ويجتذب اعتماداً واسع النطاق وخاصة في مجال المحاسبة والتدقيق والمجالات الاخرى . يتم استخدام الذكاء الاصطناعي بعدة طرق، من اقتراح المنتجات إلى تحليل المعلومات المعقدة من مصادر البيانات المتعددة لتوجيه قرارات الاستثمار والتداول.

نشأت مخاوف بشأن الأخلاق والخصوصية والأمن في الذكاء الاصطناعي، ولكن نظراً لوتيرة النمو السريع للتكنولوجيا، لم تحظ هذه المخاوف دائماً بالاهتمام الأكبر. أحد المجالات الرئيسة للقلق هو التحيز في أنظمة الذكاء الاصطناعي. يمكن للتحيز أن يحرف بشكل غير لائق الناتج من الذكاء الاصطناعي لصالح مجموعات بيانات معينة؛ لذلك، من المهم أن تحدد المنظمات التي تستخدم أنظمة الذكاء الاصطناعي كيف يمكن للتحيز التسلل ووضع ضوابط داخلية مناسبة لمعالجة القلق Sutaria (,2022,1-4) .

يعمل الذكاء الاصطناعي على تغيير العالم الذي نعيش فيه بشكلٍ جذري، وتتوفّر العديد من التطبيقات التي تهدف إلى جعل حياتنا أسهل مع هذا النوع من التكنولوجيا، وهو اتجاهٌ سيزداد بلا شكّ في السنوات المقبلة. لكن، إلى جانب هذا التطوّر الملموس، لا ينبغي أن نتجاهل جوانب مرتبطة بالأخلاقيات، وتحديداً مفهوم التحيز أو ما يُعرف بـ التحيز الخوارزمي.

يُعد التحيّز الخوارزمي ظاهرة جديدة في عالم التقنية، أثارت المخاوف من أن الكثيرين حول العالم قد يتعرضون للضرر في حياتهم ومصالحهم العملية والشخصية والاجتماعية، وأوضاعهم القانونية، بسبب هذه الظاهرة الوليدة المصاحبة لانتشار أنظمة الذكاء الاصطناعي في مجالات مختلفة، منها التوظيف ومكافحة الجريمة والائتمان، والمعاملات المالية والبنكية، والإفراج المشروط عن المسجونين، وغيرها. وقال باحثون إن المرأة والأقليات وبعض العقائد والتنويعات الاجتماعية في مقدمة ضحايا التحيّز الخوارزمي

.(https://www.emaratalyoum.com)

مصطلح التحيز يحمل معان كثيرة في الذكاء الاصطناعي أهمها:

1. هو خطأ يقع عندما يتمّ تغذية نماذج تدريب الذكاء الاصطناعي ببيانات مغلوطة تعكس أحكاماً مسبقة، أو يتم ترجيح بيانات على أخرى، ما يؤدي إلى نتائج منحرفة، وأخطاء تحليلية. ولا يقتصر هذا على الجانب التكنولوجي فحسب، بل علينا كبشر أيضا. (https://fihm.ai/)

2. أنه خطأ منهجي يتم إدخاله في أخذ العينات أو الاختبار عن طريق اختيار أو تشجيع نتيجة أو إجابة واحدة على غيرها. (Merriam–Webster,2021)

3. يعد انحياز الذكاء الاصطناعي حالة شاذة في ناتج خوارزميات التعلم الآلي، بسبب الافتراضات المتحيزة التي تم إجراؤها أثناء عملية تطوير الخوارزمية أو الأحكام المسبقة في بيانات التدريب.

(Dilmegan, 2020)

4. يحدث التحيز في الذكاء الاصطناعي عندما لا يمكن تعميم النتائج على نطاق واسع. غالبًا ما نفكر في التحيز الناتج عن التفضيلات أو الاستثناءات في بيانات التدريب، ولكن يمكن أيضًا تقديم التحيز من خلال كيفية الحصول على البيانات وكيفية تصميم الخوارزميات وكيفية تفسير مخرجات الذكاء الاصطناعي.

(Siwicki,2021)

وعليه يمكن القول أن تحيز نموذج الذكاء الاصطناعي يحدث عندما لا تعكس بيانات التدريب التي تعتمد عليها خوارزمية أو نموذج الذكاء الاصطناعي الواقع الذي من المفترض أن يعمل فيه الذكاء الاصطناعي. بعبارة أخرى، على الرغم من استخدام مصطلح التحيز النموذجي، فإن النموذج ليس متحيزاً في حد ذاته؛ بدلاً من ذلك، إنها بيانات التدريب مما يجعل النموذج متحيزاً.

لقد بدأنا للتو في فهم مقدار التحيزات البشرية التي تشق طريقها إلى الذكاء الاصطناعي. على سبيل المثال، قد يتم اختيار البيانات التي نستخدمها لتدريب الذكاء الاصطناعي بطريقة متحيزة. ولكن حتى إذا كانت الشركات التي تبني أنظمة الذكاء الاصطناعي لا تنوي التمييز، فإن الأدوات التي تستخدمها يمكن أن يكون لها نتائج تمييزية. ونظراً لأن البرامج تتحكم في الكثير من حياتنا اليومية، فإن النتيجة هي التحيز المنهجي الذي قد يكون من الصعب القضاء عليه (Heather & Amit 2020) .

يحدث التحيز في الذكاء الاصطناعي عندما لا يتم اعتبار مجموعتين من البيانات متساويتين، ربما بسبب الافتراضات المتحيزة في عملية تطوير خوارزمية الذكاء الاصطناعي أو التحيزات المضمنة في بيانات التدريب.

تشمل الأمثلة الحديثة على التحيز ما يلي:

1. اضطرت شركة تقنية رائدة إلى إلغاء أداة توظيف قائمة على الذكاء الاصطناعي أظهرت تحيزاً ضد النساء. (Dastin, J, 2018,1)

2.اضطرت شركة برمجيات رائدة إلى إصدار اعتذار بعد أن بدأ حساب Twitter المستند إلى الذكاء الاصطناعي في التغريد بتعليقات عنصرية. (Lee, D ,Tay, 2018)

3.كان على مؤسسة تقنية رائدة أن تتخلى عن أداة التعرف على الوجه لإظهار التحيز تجاه أعراق معينة (BBC News, 2020) .

4.اعتذرت إحدى منصات التواصل الاجتماعي الرائدة عن خوارزمية اقتصاص الصور التي أظهرت العنصرية من خلال التركيز تلقائيًا على الوجوه البيضاء على الوجوه الملونة. (Hern, A,2020)

عندما نتحدث عن الذكاء الاصطناعي، فإننا نتحدث عن الناس. يصمم البشر خوارزميات الذكاء الاصطناعي، ولا يزال البشر هم المستفيدون الرئيسيون من جميع تطبيقات الذكاء الاصطناعي التي نستخدمها يومياً.

وهذا يفسر لماذا يجب أن نبدأ في اعتبار الذكاء الاصطناعي المتحيز ليس فقط مشكلة تقنية، بل كمشكلة بشرية، وبالتالي نتبنى منظوراً جديداً.

عندما يتعلق الأمر بالتحيز الخوارزمي، هناك قضيتان رئيسيتان يجب معالجتهما: البيانات وتعريف النجاح. هل البيانات المتوفرة كاملة؟ هل هو ممثل لكل الناس؟ إذا لم يكن الأمر كذلك، فإن التنبؤ الذي تقوم به الخوارزمية سيكون حتماً متحيزاً.

لكن في الحقيقة، المشكلة الحقيقية هي أننا، كبشر، وبصفتنا مسؤولين عن تصميم أنظمة الذكاء الاصطناعي، فإننا منحازون بشكل طبيعي ولا شعوري. وإن تحيز البشر أصعب من إضعاف أنظمة الذكاء الاصطناعي(Seneory& Mezzanotte, 2022) .

رابعاً: أسباب التحيز في الذكاء الاصطناعي

تحتوي أنظمة الذكاء الاصطناعي على تحيزات لسببين: (Dilmegan, 2020)

1. التحيزات المعرفية: وهي أخطاء غير واعية في التفكير تؤثر على أحكام الأفراد وقراراتهم. تنشأ هذه التحيزات من محاولة الدماغ تبسيط معالجة المعلومات حول العالم. تم تحديد وتصنيف أكثر من 180 تحيزاً بشرياً من قبل علماء النفس. يمكن أن تتسرب التحيزات المعرفية إلى خوارزميات التعلم الآلي عبر أي منهما

المصممين يعرّفونهم عن غير قصد بالنموذج مجموعة بيانات تدريبية تتضمن تلك التحيزات.

2. نقص البيانات الكاملة: إذا لم تكن البيانات كاملة، فقد لا تكون متمثله وبالتالي قد تتضمن تحيزاً. على سبيل المثال، تتضمن معظم الدراسات البحثية في علم النفس نتائج من الطلاب الجامعيين الذين يمثلون مجموعة محددة ولا يمثلون جميع السكان.

يمكن القول أن المشكلة الحقيقية في البيانات، فهي المنحازة وليست التقنية، ويبدو الأمر كأن البشر يعلِّمون التقنية أن تكون عنصرية، فإذا كان ثمة قسم علمي بإحدى الكليات يتم ملؤه ببيانات للرجال فقط أو تجري فلترة المتقدمين وفق الجنس، ويتم رفض النساء معظم الوقت، فهو يستنتج من البيانات السابقة أن النساء لا يعملن بذلك القسم، ولذا إذا تقدمت إحداهن فبنسبة كبيرة سيتم رفضها أيضاً، وكذلك الحال مع مؤشر البحث كوكل، فإذا كتبنا كلمة لاعب –على سبيل المثال– فإن أغلب الصور ستكون للاعبي كرة القدم من الرجال فقط، دون ظهور أي صور تتعلق بالألعاب الأخرى أو النساء.

(https://www.scientificamerican.com)

مما سبق يتضح:

1. أحد الاهتمامات الأساسية عندما يتعلق الأمر بالذكاء الاصطناعي هو حقيقة أن التحيز مبرمج بشكل غير مقصود. هل من الممكن تصميم خوارزميات خالية من هذه المشكلة؟

2. يمكن لأنظمة الذكاء الاصطناعي غير الخاضعة للرقابة، وغير المنظمة، وفي بعض الأحيان غير المرغوب فيها أن تضخم العنصرية والتمييز على أساس الجنس وأشكال أخرى من التمييز.

3. مثل جميع التقنيات الأخرى، يتطلب الذكاء الاصطناعي خوارزميات التعلم الآلي للإدخال البشري والتي تعمل على التفكير وجمع وتقديم المعلومات للمستخدمين.

4. لا يعتقد معظم الخبراء أن الذكاء الاصطناعي سيكون خالياً من التحيز في أي وقت قريب. في الواقع، وافق البعض حتى على أن تحيز الخوارزمية أمر لا مفر منه. الخبر السار هو أننا نمتلك القدرة على تقليل التحيز في الخوارزميات إذا بذل أولئك الذين يبرمجونها جهوداً كبيرة.

5. تُستخدم أنظمة الذكاء الاصطناعي في أمور كثيرة من أوجه حياتنا اليومية، وفي حين أن تلك الاستخدامات تبشر بالخير، إلا أنها يمكن أن تضر الأشخاص المستضعفين والمهمشين، وتهدد الحقوق المدنية.

6. مع بدء الأفراد والشركات في الاستفادة من مزايا الذكاء الاصطناعي بشكل متكرر في الحياة اليومية، لا يزال أحد الشواغل الرئيسة يطفو على السطح فيما يتعلق بطريقة عمله.

7. لقد باتت تقنيات الذكاء الاصطناعي تتحكم في أغلب معاملات البشر، غير أنه كلما اتسعت رقعة استخدامها، تزايدت المخاوف تجاهها؛ ما جعل بعض الباحثين والمطورين يركزون على الوجه الآخر لتلك الخوارزميات، وكيف باتت تتحيز تجاه البعض وسط تحذيرات.

يمكن للمجموعة التالية من الإرشادات أن تساعد المنظمات على توقع تحيز نموذج الذكاء الاصطناعي عبر السياقات. يمكن أن تساعد مثل هذه الإرشادات المنظمة على نشر نماذج الذكاء الاصطناعي بطرق عادلة وشفافة.

1.تثقيف الجميع داخل المنظمة حول احتمالية مخاطر التحيز في نموذج الذكاء الاصطناعي. حتى بين أولئك الذين يشاركون بشكل مباشر في تطوير ونشر نماذج الذكاء الاصطناعي.

2.إنشاء لغة مشتركة لمناقشة المخاطر النموذجية وطرق التخفيف منها. يشترك الذكاء الاصطناعي الجدير بالثقة، والمعروف أيضاً باسم الذكاء الاصطناعي الأخلاقي أو المسؤول، في موضوعات مشتركة في تطوير واستخدام تطبيقات الذكاء الاصطناعي. تتضمن هذه الموضوعات العدالة والشفافية والموثوقية والمساءلة والسلامة والأمن والخصوصية. توفر مثل هذه الموضوعات لغة مشتركة وعدسة لتقييم مخاطر الذكاء الاصطناعي والتخفيف من حدتها، بما في ذلك التحيز في النموذج .(Deloitte,2021)

3.التأكد من أن الأشخاص الأكثر تأثراً بالنموذج هم في الفريق عند تطوير النموذج. أن البشر يميلون إلى الإيمان بدقة قرارات نموذج الذكاء الاصطناعي دون أي فهم حقيقي لكيفية عمل النموذج أو تطويره.(Bucinca, 2021, 1-20)

4.القيام بتضمين العملية والتكنولوجيا أيضاً. " التحيز هو التحدي. ستكون دائماً هناك أعتقاد أن أفضل طريقة لحلها هي اتباع نهج الأشخاص والعملية والتكنولوجيا ". هكذا يلعب البشر دوراً أساسياً في دورة حياة تطوير الذكاء الاصطناعي وتخفيف التحيز. لكن البشر ليسوا سوى جزء من مخطط أكبر ومتكامل يجعل الذكاء الاصطناعي الجدير بالثقة ممكناً (Deloitte,2021) .

خامساً : طرق التحيز في الذكاء الاصطناعي

يمكن أن يحدث تحيز البيانات بعدة طرق، ولكن غالباً ما ينشأ بطريقة منهجية قبل جمع البيانات. يحدث أيضاً في مراحل أخرى من تدريب الذكاء الاصطناعي. الاعتبار النقدي الأخلاقي هو أن تحيز البيانات، تماماً مثل تحيزاتنا، يؤدي إلى مشاكل اجتماعية كبيرة. الأهم من ذلك، أن الذكاء

الاصطناعي المتحيز، بعد كل شيء، ليس مفيدا جداً في أحسن الأحوال ويمكن أن يؤدي حتى إلى تفاقم المشكلات التي تم تصميمه لحلها أو معالجتها. في أسوأ الأحوال، يؤدي إلى تضخيم أسوأ ميولنا البشرية مثل العنصرية أو قلة الفرص للمرأة. وقد لوحظ كلاهما في عمليات نشر الذكاء الاصطناعي من قسم شرطة لوس أنجلوس لمشاكل الشرطة التنبؤية ، (Puente, 2019,10) إلى الذكاء الاصطناعي للرعاية الصحية الذي تستخدمه Optimum Health الذي يصنف المرضى البيض أعلى من المرضى السود الأكثر مرضاً لتقديم الخدمة على أساس التحيز التضمين Powers & (Mullainathan, 2019,446).

نشر الذكاء الاصطناعي في تصفية السير الذاتية لممارسات التوظيف أدى إلى استبعاد النساء إلى حد كبير لأن مجموعة البيانات التي تم تدريب الذكاء الاصطناعي عليها تتكون في الغالب من الرجال .

في جميع هذه الحالات، لم يتم تصميم الذكاء الاصطناعي عن قصد لتحقيق هذه النتائج، ولكن بدلاً من ذلك تم تطويره بناءً على تحيزات البيانات المضمنة في المنهجية والبيانات نفسها. فيما يلي ثلاث مراحل رئيسية منهجية حيث يمكن أن يحدث التحيز :(Brown & Chun,2020,3)

1. تأطير المشكلة:عندما يقرر عالم البيانات / المصممون ما يريدون تحقيقه ، يمكن أن يحدث التحيز في تأطير المشكلة. على سبيل المثال، لنتحدث عن الهدف الغامض المتمثل في "الجدارة الائتمانية" عندما تريد شركة بطاقات الائتمان التنبؤ بما يحدد الجدارة الائتمانية. قد يبدو هذا هدفًا نهائيا ومنفصلاً، لكنه في الواقع غامض تماماً من الناحية المنهجية، يجب أن تقرر الشركة ما إذا كانت "الجدارة الائتمانية" تعني تعظيم هوامش الربح أو عدد القروض المسددة.

2. جمع البيانات: عندما يتم جمع البيانات وتكون إما غير ممثلة للواقع أو تعكس تحيزاً قائماً، فإن جمع البيانات نفسه يكون متحيزاً. قد تكون البيانات غير ممثلة للواقع إذا تم تدريب خوارزمية التعلم العميق على صور الوجوه ذات البشرة الفاتحة، وبالتالي واجهت صعوبة في التعرف على الوجوه ذات درجات ألوان البشرة المختلفة (والعكس صحيح).

3. إعداد البيانات: عند اختيار السمات للخوارزمية للنظر في التحيز يمكن إدخالها. يُشار إلى ذلك باسم إعداد البيانات أو اختيار البيانات الوصفية المراد استخدامها. في مثال الائتمان، يمكن أن تكون السمات هي عمر الزبون أو دخله أو القروض المدفوعة أو الرمز البريدي.

سادساً: كيف يمكن للذكاء الاصطناعي أن يكون مُتحيزاً؟

رغم أنه قد تمّ تحديد التحيّز في أنظمة التعرّف على الوجه، وبرامج التوظيف، والخوارزميات الكامنة وراء عمليات البحث على الويب وغيرها، إلّا أنّ السؤال الذي يطرح نفسه هو حول كيفية دخول هذه التحيزات في الأنظمة. ولا بدّ من الإشارة إلى أنّه غالباً ما نختزل تفسيرنا لتحيّز الذكاء الاصطناعي من خلال إلقاء اللوم على بيانات التدريب. لكن الواقع الأكثر دقّة، أنّه يمكن للتحيّز أن يتسلّل قبل عملية جمع البيانات، حيث يظهر من خلال هذه العمليات الثلاث: بناء النموذج، أو جمع البيانات أو إعداد مجموعة بيانات تحكمها سمات معيّنة. إلّا أننا نجد التحيّزات الأكثر شيوعاً تتمثّل أثناء جمع البيانات. ويمكن أن يحدث هذا في حالتين: إمّا أن تكون البيانات التي تمّ الحصول عليها غير ممثلة للواقع، أو أنها تعكس التحيّزات القائمة. وقد تحدث الحالة الأولى مثلاً، إذا تمّ تغذية خوارزمية التعلّم العميق بصورٍ للوجوه ذات البشرة الفاتحة أكثر من الوجوه ذات البشرة الداكنة، وهذا من شأنه أن يجعل نظام التعرّف على الوجوه سيئاً في التعرّف على ذوي البشرة الداكنة.

()https://fihm.ai/

بشكل أساسي تكون عنصرية الذكاء الاصطناعي بسبب البيانات المُتحيّزة التي يتم تدريب الذكاء الاصطناعي بواسطتها، أيضاً وجود افتراضات عنصرية وانعكاس لمشكلات المجتمع العنصرية في خوارزميات الذكاء الاصطناعي؛ تجعله يتصرف بطرق تعكس عدم التسامح أو التمييز الموجود في المجتمع.

أحد الأمثلة الشهيرة على تحيز الذكاء الاصطناعي كانت شركة أمازون في عام 2014، بدأ فريق من المهندسين في أمازون العمل في مشروع لأتمتة التوظيف في شركتهم. كانت مهمتهم هي بناء خوارزمية يمكنها مراجعة السير الذاتية وتحديد المتقدمين الذين يجب على أمازون إحضارهم للمساعدة في عملية التوظيف بالشركة.

بحلول عام 2015 لاحظ المهندسون أن الخوارزمية التي تم اختبارها كأداة توظيف من قبل شركة أمازون العملاقة على الإنترنت متحيزة ضد النساء، ، كان من الواضح أن النظام لم يقم بتصنيف المرشحين بطريقة محايدة ومع مراجعة السير الذاتية التي تم تغذية النظام بها لتدريبه على اختيار الأنسب للوظائف، اكتشفت الشركة أن أغلب هذه السير الذاتية كانت لذكور، وبالتالي أعطى نظام الذكاء الاصطناعي تقييماً منخفضاً للسير الذاتية التي تحتوي على أسماء مؤنثة،). https://masaar.net/ar/)

يبدو أن أحداً لم يحاول الإجابة على السؤال حول كيف يمكن لشركة من مستوى أمازون –بموارد لا حصر لها على ما يبدو –أن تتعثر بهذه السوء. هل تقنية الذكاء الاصطناعي شريرة بطبيعتها؟ هل جميع مهندسي البرمجيات وعلماء البيانات متحيزون؟ هل التكنولوجيا غير ناضجة للغاية بالنسبة لمشاكل العمل المعقدة؟ أم أن هناك شيئاً محدداً بشأن مشروعات الذكاء الاصطناعي يجعل هذه المشروعات صعبة –حتى بالنسبة لأفضل الشركات؟ (Lauret, Julien , 2019)

سابعاً : طرق تخفيف التحيز في الذكاء الاصطناعي والحد منه

تطوير أنظمة ذكاء اصطناعي تتسم بالإنصاف وتنأى عن التحيز ليست مهمة سهلة لعدة أسباب؛ ربما تكون أهمها هو تغذية نماذج التعلم الآلي بالبيانات التي تُجمع من العالم الحقيقي، وبالتالي فإن أكثر الأنظمة دقة يمكن أن تتعلم أو تضخم التحيزات الموجودة مسبقاً في هذه البيانات، والتي يمكن أن تحتوي على تحيّز قائم على أساس العرق أو الجنس أو الدين أو أي خصائص أخرى. يمكن أيضاً أن يكشف النظام عن نقاط عمياء غير مقصودة بعد إطلاقه، هذه النقاط قد تكون إشكاليات تظهر قبل أو أثناء أو بعد تطوير نظام الذكاء الاصطناعي نتيجة وجود تحيزات أو أحكام مسبقة أو تفاوتات هيكيلية في المجتمع، وتحدث حتى مع تدريبه واختباره بشكل صارم.

لا يوجد تعريف موحد للإنصاف، سواء كان اتخاذ القرار يتم بواسطة البشر أو الآلات، فتحديد معايير الإنصاف المناسبة لنظام ما يتطلب مراعاة تجربة المستخدم والاعتبارات الثقافية والاجتماعية والتاريخية والسياسية والقانونية والأخلاقية، والتي قد يكون للعديد تفضيلات مختلفة اتجاهها.

هناك عدة تقنيات لفرض قيود الإنصاف للحد من التحيز في نماذج الذكاء الاصطناعي أهمها (Kumari , 2022) :

1. يتضمن المعالجة المسبقة للبيانات لضمان الدقة قدر الإمكان مع التخلص من أي ارتباط بين النتائج والميزات المحمية أو إنشاء تمثيلات البيانات التي لا تتضمن معلومات السمة الحساسة. تتضمن هذه المجموعة تقنيات الإنصاف المضاد، والتي تأسست على فرضية أنه عندما يتم تعديل سمة حساسة في عالم واقعي غير واقعي، يجب أن يظل القرار كما هو.

2. تقنيات ما بعد المعالجة هي النهج الثاني. لتلبية قيود الإنصاف، هذه تغير بعض تنبؤات النموذج بعد إجرائها.

3. تستخدم الطريقة الثالثة إما خصماً لتقليل قدرة النظام على التنبؤ بالميزة الحساسة أو تفرض قيوداً على الإنصاف على عملية التحسين نفسها .

في محاولة للتغلب على مشكلة تحيز الذكاء الاصطناعي نشرت شركة Innodata المتخصصة في هندسة البيانات 5 ممارسات يمكن الاعتماد عليها للحد من تحيز الذكاء الاصطناعي والوصول إلى نماذج تعلم آلي أكثر إنصافًا وشمولًا، والممارسات هي(https://masaar.net/ar):

1.اختيار مجموعة البيانات

في حين أن التخفيف من حدة تحيز الذكاء الاصطناعي والتعلم الآلي يمكن أن يمثل تحدياً، إلا أن هناك تقنيات وقائية يمكن أن تساعد في حل هذه المشكلة. التحدي الأكبر في تحديد التحيز يتمثل في فهم كيف تقوم بعض خوارزميات التعلم الآلي بتعميم البيانات التي تم تدريبها عليها. لذلك من المهم أن تتسم البيانات المستخدمة لتدريب النموذج بالشمول.

2. تنوع الفريق

بناء فريق متنوع يساهم إلى حد كبير في القضاء على التحيز. يمكن أن يكون لتنوع الفريق تأثير إيجابي على نماذج التعلم الآلي من خلال إنتاج مجموعات بيانات مُمثّلة ومتوازنة .كما أنه يساعد في التخفيف من التحيز الضار في بنية مجموعات البيانات وكيفية تطبيق التصنيفات على تلك البيانات.

3. تقليل الإقصاء

يعد اختيار الميزة (Feature selection) أمراً أساسياً للمساعدة في تقليل الإقصاء في الذكاء الاصطناعي، وهي عملية تقوم على تقليل عدد المتغيرات المُدخلّة إلى نماذج الذكاء الاصطناعي بهدف تحسين أداء التنبؤ بالنتائج. تستبعد هذه الخطوة عناصر البيانات التي لا تحتوي على تباين كافٍ للتأثير على النتائج.

4. الخوارزميات وحدها ليست كافية

هناك طريقة أخرى للبدء في حل مشكلة التحيز في الذكاء الاصطناعي، وهي عدم الاعتماد فقط على الخوارزميات، بل إبقاء الأفراد القائمين على تطوير الذكاء الاصطناعي على اطلاع بكل ما يتعلق بالنظام ليتمكنوا من التعرف بفعالية على أنماط التحيز غير المقصود.

5. بيانات ممثلة

يجب على المؤسسات فهم كيف تبدو البيانات الممثلة قبل جمع البيانات التي سيتم تدريب نموذج التعلم الآلي عليها. كما يجب أن يحتوي جوهر وخصائص البيانات المستخدمة أقل قدر من التحيز. وإلى جانب تحديد التحيز المحتمل في مجموعات البيانات، يجب على المؤسسات أيضاً توثيق أساليبها في اختيار البيانات وتنقيتها، للقضاء جذرياً على أسباب التحيز.

عند استخدام الذكاء الاصطناعي، من الأهمية بمكان مراعاة المناطق المعرضة للتحيز غير العادل، مثل تلك التي لها تاريخ من أنظمة التحيز أو البيانات المنحرفة. يمكن مراعاة الطرق التالية لتقليل التحيز في أنظمة الذكاء الاصطناعي: (Kumari , 2022)

1. وضع إجراءات لاكتشاف وتخفيف التحيز في أنظمة الذكاء الاصطناعي. جمع البيانات المحسن من خلال أخذ عينات هادفة أكثر واستخدام " أطراف ثالثة لتدقيق البيانات والنماذج هي أمثلة على التقنيات التشغيلية.

2.الانخراط في مناقشات قائمة على الحقائق بشأن تحيزات صنع القرار البشري. يجب على المنظمات استكشاف كيف يمكن تحسين العمليات التي يقودها الإنسان في المستقبل إذا أظهرت النماذج المدربة على القرارات أو السلوك البشري الحالي التحيز.

3. يجب إنفاق المزيد من الأموال على أبحاث التحيز، ويجب توفير المزيد من البيانات للتحليل (مع الحفاظ على الخصوصية)، ويجب اتباع نهج متعدد التخصصات. ستكون المشاركة متعددة التخصصات، بما في ذلك علماء الأخلاق وعلماء الاجتماع والخبراء الذين يفهمون بشكل أفضل تعقيدات كل مجال تطبيق في العملية، مطلوبة لتحقيق المزيد من التقدم.

4. يجب الاستثمار في تنويع منطقة الذكاء الاصطناعي. سيكون مجتمع الذكاء الاصطناعي الأكثر تنوعاً أكثر قدرة على التنبؤ والتعرف والتحقيق في حالات التحيز غير العادل وإشراك السكان الذين من المحتمل أن يتأثروا بالتحيز.

هناك العديد من الخطوات يمكن أن يفعله الرؤساء التنفيذيون وفرق الإدارة العليا لقيادة الطريق نحو التحيز والإنصاف؟ من بين أمور أخرى ، أهمها : (Manyika, & Presten, 2019)

1. سيحتاج قادة الأعمال إلى البقاء على اطلاع دائم على هذا المجال البحثي سريع الحركة. توفر العديد من المنظمات الموارد لمعرفة المزيد، مثل التقارير السنوية لمعهد AI Now ، والشراكة حول الذكاء الاصطناعي ، ومجموعة العدالة والشفافية والخصوصية التابعة لمعهد آلان تورينج.

2. عندما تقوم مؤسستك بنشر الذكاء الاصطناعي، قم بإنشاء عمليات مسؤولة يمكن أن تخفف من التحيز. ضع في اعتبارك استخدام مجموعة من الأدوات الفنية، بالإضافة إلى الممارسات التشغيلية.

3. الانخراط في محادثات قائمة على الحقائق حول التحيزات البشرية المحتملة. لطالما اعتمدنا على الوكلاء مثل الفحوصات الإجرائية عند تقرير ما إذا كانت القرارات البشرية عادلة.

4. وضع في الأعتبار كيف يمكن للبشر والآلات العمل معاً للتخفيف من التحيز. يمكن أن تساعد الشفافية حول ثقة هذه الخوارزميات في توصياتها البشر على فهم مقدار الوزن الذي يجب منحه إياها.

5. استثمار أكثر، وتوفير المزيد من البيانات، واتباع نهجاً متعدد التخصصات في أبحاث التحيز (مع احترام الخصوصية) لمواصلة التقدم في هذا المجال. جهود مهمة لجعل اختيارات المصممين أكثر شفافية وتضمين الأخلاقيات في مناهج علوم الكمبيوتر، من بين أمور أخرى، تشير إلى الطريق إلى الأمام في التعاون. ستكون هناك حاجة إلى المزيد.

6، استثمار أكثر في تنويع مجال الذكاء الاصطناعي نفسه. سيكون مجتمع الذكاء الاصطناعي الأكثر تنوعاً مجهزاً بشكل أفضل لتوقع التحيز ومراجعته وتحديده وإشراك المجتمعات المتأثرة.

ثامناً: أنواع تحيز البيانات

تعد أنظمة الذكاء الاصطناعي جيدة بقدر ما صُممت لتكون، ولسوء الحظ، يمكن أن تتسلل تحيزاتنا البشرية في كثير من الأحيان إلى الذكاء الاصطناعي. بدأ المشرعون والمنظمون والناشطون المدنيون في التركيز على تحيزات الذكاء الاصطناعي وكيف يمكن أن تؤثر على مجتمعنا. وأثناء قيامهم بذلك، طالبوا بمساءلة الشركات عن استخدامها للذكاء الاصطناعي(Heather & Amit 2020).

ببساطة، أصبح التحيز في الذكاء الاصطناعي الآن مشكلة قانونية يجب على الشركات معالجتها.

هناك طرق مختلفة يمكن من خلالها إدخال التحيز في نظام التعلم الآلي على الرغم من أن هذه القائمة ليست شاملة، إلا أنها تحتوي على أمثلة شائعة لتحيز البيانات في المجال، إلى جانب أمثلة على مكان حدوثه.

(Smith, 2018)

1. انحياز العينة (التحيز في الاختيار): يحدث تحيز العينة عندما لا تعكس مجموعة البيانات واقع البيئة التي سيتم فيها تشغيل النموذج. مثال على ذلك هو بعض أنظمة التعرف على الوجه المدربة

بشكل أساسي على صور الرجال البيض. هذه النماذج لديها مستويات أقل بكثير من الدقة مع النساء والأشخاص من مختلف الأعراق.

2. تحيز الاستبعاد: يعد تحيز الاستبعاد أكثر شيوعاً في مرحلة معالجة البيانات.

يحدث هذا عندما تُترك نقطة بيانات مهمة خارج البيانات المستخدمة وهو أمر يمكن أن يحدث إذا لم يتعرف المصممون على نقطة البيانات على أنها تبعية.

3.تحيز القياس. كما يوحي الاسم، ينشأ هذا التحيز بسبب المشاكل الأساسية المتعلقة بدقة البيانات وكيفية قياسها أو تقييمها. يمكن أن يكون استخدام صور العمال السعداء لتدريب نظام يهدف إلى تقييم بيئة مكان العمل متحيز إذا علم العمال في الصور أنهم يقيسون السعادة؛ سيكون النظام الذي يتم تدريبه لتقييم الوزن بدقة متحيزًا إذا تم تقريب الأوزان الواردة في بيانات التدريب باستمرار.

4. تحيز الاسترجاع: هذا نوع من تحيز القياس، وهو شائع في مرحلة وسم البيانات للمشروع. ينشأ تحيز الاسترجاع عندما تقوم بتسمية أنواع متشابهة من البيانات بشكل غير متسق. ينتج عن هذا دقة أقل. على سبيل المثال، لنفترض أن لديك فريقاً يصنف صور الهواتف على أنها تالفة أو تالفة جزئياً أو غير تالفة. إذا قام شخص ما بتسمية صورة واحدة على أنها تالفة، ولكن صورة مشابهة لها على أنها تالفة جزئياً، فستكون بياناتك غير متسقة.

5.تحيز المراقب: المعروف أيضاً باسم تحيز التأكيد، هو تأثير رؤية ما تتوقع رؤيته أو تريد رؤيته في البيانات. يمكن أن يحدث هذا عندما يدخل الباحثون في مشروع بأفكار ذاتية حول دراستهم، سواء كانت واعية أو غير واعية. يمكننا أيضاً رؤية ذلك عندما يسمح المصممون لأفكارهم الذاتية بالتحكم في عاداتهم في وضع العلامات، مما يؤدي إلى بيانات غير دقيقة.

6.التحيز العنصري: على الرغم من أنه ليس تحيزاً في البيانات بالمعنى التقليدي، إلا أن هذا لا يزال يستدعي ذكره نظراً لانتشاره في تقنية الذكاء الاصطناعي مؤخراً. يحدث التحيز العنصري عندما تنحرف البيانات لصالح التركيبة السكانية الخاصة. يمكن ملاحظة ذلك في تقنية التعرف على الوجه والتعرف التلقائي على الكلام التي تفشل في التعرف على الأشخاص الملونين بدقة مثل القوقازيين.

7.انحياز الرابطة: يحدث هذا التحيز عندما تعزز البيانات الخاصة بنموذج التعلم الآلي أو تضاعف التحيز الثقافي. قد تحتوي مجموعة البيانات الخاصة بك على مجموعة من الوظائف التي يعمل فيها جميع الرجال أطباء وجميع النساء ممرضات. هذا لا يعني أن النساء لا يمكن أن يصبحن طبيبات

والرجال لا يمكن أن يكونوا ممرضين. ومع ذلك، فيما يتعلق بنموذج التعلم الآلي الخاص بك، لا توجد طبيبات وممرضات ذكور. يُعرف التحيز الجماعي بخلقه تحيزاً بين الجنسين.

المبحث الثاني: تدقيق الذكاء الاصطناعي – اعتبارات لمهنة التدقيق الداخلي

مع استمرار المؤسسات في تبني الذكاء الاصطناعي، من المتوقع أن يقدم التدقيق الداخلي تأكيداً على تطبيقه واستخدامه. ولكن هل يمتلك فريق التدقيق الداخلي الخاص بك إطاراً مناسباً لتدقيق الذكاء الاصطناعي لتقييم المشكلات بفعالية وتقديم تحدٍ قوي؟

رغم الجهود المبذولة لإزالة التحيّز من أنظمة الذكاء الاصطناعي لكنها لا تزال صعبة، ويرجع ذلك جزئياً إلى أنها لا تزال تعتمد على البشر في تدريبها من خلال تحديد نوعية البيانات المدخَلة. وعندما نتعمّق أكثر، نرى أن هناك الكثير من الأشياء التي يجب أخذها بعين الاعتبار. إلّا أنّ بناء ثقافة الإبلاغ والمساءلة والأخلاقيات والتدقيق في مجال الذكاء الاصطناعي، يعني أنه سيكون هناك فرصة مستقبلية لتحديد وإيقاف تحيّز البيانات أو الخوارزميات أو الأنظمة ويكون للتدقيق الداخلي دور في ذلك ، بما قد يساعد في ضمان أن تعمل هذه التقنية كقوةٍ لتحقيق العدالة والإنصاف إلى حدّ ما ( ). https://fihm.ai

أولاً : أهمية التركيز على تدقيق الذكاء الاصطناعي

على الرغم من التغير السريع في السوق والطبيعة الضبابية في بعض الأحيان للذكاء الاصطناعي، يمكن للمدققين الداخليين تقديم ضمانات قيمة وسليمة بشكل أساسي بأن المنظمات التي يخدمونها توجه استثماراتها في الذكاء الاصطناعي في الاتجاه الصحيح وغير متحيز، مع مراعاة المخاطر والفرص، وتنفيذ أهداف أعمالهم المتعلقة بالذكاء الاصطناعي. كلما أسرع المدققون الداخليون في القيام بذلك، كان ذلك أفضل، وذلك ببساطة لأن الذكاء الاصطناعي، بجميع أشكاله المختلفة، لن يختفي. إنها تكتسب زخماً.

لا تزال ممارسة تدقيق الذكاء الاصطناعي قيد التطوير. فيما يتعلق بتقنيات الذكاء الاصطناعي المتقدمة، مثل التعلم الآلي، لا توجد معايير عالمية تحكم الذكاء الاصطناعي، جنباً إلى جنب مع أطر للتدقيق، لا تزال قيد الكتابة. أصدرت جمعية تدقيق ومراقبة نظم المعلومات (ISACA) إرشادات حول تطبيق إطار عمل الحالي يستخدم لإدارة تكنولوجيا المعلومات الخاصة بالمؤسسة وإدارتها على الذكاء الاصطناعي (COBIT)، كما اقترح معهد المدققين الداخليين إطار عمل تدقيق الذكاء الاصطناعي.

وفي الوقت نفسه، تقوم مجموعة من المنظمات، بما في ذلك NIST و ISO و ASTM ، بصياغة معايير للذكاء الاصطناعي.

(Alvero , 2019)

التأثير على التدقيق الداخلي

مثل أي عملية تعتمد على الخوارزمية والبيانات، يقدم الذكاء الاصطناعي للتدقيق الداخلي دوراً واضحاً في ضمان الدقة والموثوقية. يمكن للذكاء الاصطناعي أن يعمل بشكل صحيح فقط عندما يقوم بتحليل البيانات الجيدة وتقييمها مقابل معايير صالحة، وهي المجالات التي يمكن أن يوفر فيها التدقيق الداخلي تأثيراً إيجابياً. مهم للمجالس وأصحاب المصلحة الآخرين فهم أن التدقيق الداخلي يمكن أن يركز على ما يلي :

(IIA, 2017,2)

1. مراحل دورة حياة الذكاء الاصطناعي:

أ. كما تم بناء الذكاء الاصطناعي. يتم إنشاء أنظمة الذكاء الاصطناعي من قبل البشر، الذين يمكن أن يكونوا متحيزين وحكميين وغير منصفين. بالنظر إلى التدقيق الداخلي لإجراء الاختبارات التي تحدد أن النتائج التي ينتجها الذكاء الاصطناعي تعكس الهدف الأصلي ولم تتأثر بتحيزات مبتكري التقنية – ويمكن أيضاً اقتراح طرق التخفيف من التحيز والظلم والأضرار المحتملة الأخرى في أنظمة صنع القرار الآلية.

ب. كما يؤدي الذكاء الاصطناعي: ان أهم دور للمدققين الداخليين في أداء الذكاء الاصطناعي هو تحديد وتقييم وإبلاغ الإدارة ومجلس الإدارة بمخاطر الذكاء الاصطناعي الكبيرة والجهود المبذولة للتصدي لتلك المخاطر. تشمل بعض المخاطر المحتملة سيتم تضمين أخطاء المنطق البشري في تقنية الذكاء الاصطناعي.

ج. حيث يتم إدارة الذكاء الاصطناعي والتحكم فيه. مثل أي تقنية ناشئة، سيتطلب الذكاء الاصطناعي إعادة فحص خطوط المساءلة والرقابة، ومراجعة أو تطوير السياسات والإجراءات الحاكمة.

2. دعم مجلس الإدارة: يجب على مجلس الإدارة دعم الرئيس التنفيذي للتدقيق في أنشطة مثل:

أ. التأكد من فهم المدققين الداخليين للأهداف الاستراتيجية للمؤسسة وعمليات الذكاء الاصطناعي أنشأت لتحقيق تلك الأهداف.

ب. تقييم كيف يمكن للذكاء الاصطناعي أن يكمل جهود المدققين الداخليين من خلال أداء الجوانب الإدارية للتدقيق الداخلي.

مجلس الادارة: تحتاج مجالس الإدارة إلى التأكد من طرح الأسئلة حول تقاطع الذكاء الاصطناعي / التدقيق الداخلي والإجابة عليها من قبل الهيئات المناسبة –أسئلة مثل:

أ. هل تم وضع خطة تدقيق للذكاء الاصطناعي؟

ب. هل يتم تدقيق تطبيقات الذكاء الاصطناعي بشكل مختلف عن تطبيقات تكنولوجيا المعلومات العامة؟ قد يحتاج المدققون الداخليون إلى التحدث إلى أشخاص مختلفين لتدقيق تطبيقات الذكاء الاصطناعي –علماء البيانات، على سبيل المثال، على عكس قسم تكنولوجيا المعلومات.

ج.هل تم تقييم مخاطر تطبيقات الذكاء الاصطناعي بالكامل؟

د.هل يقوم التدقيق الداخلي بتقييم ما إذا كانت تطبيقات الذكاء الاصطناعي تساعد المؤسسة في تحقيق الأهداف؟

هـ.هل تفهم هذه المنظمة النموذج يمثل التحول الذكاء الاصطناعي للبحث عن إجابة صحيحة مقابل احتمال الإجابة الصحيحة؟ لديه عتبة الاحتمالية التي يمكن أن تتحملها المنظمة تم تحديده؟

3. التحقق مما إذا كان المدققون الداخليون لديهم المهارات والتدريب للتعرف على المخاطر المتعلقة بالذكاء الاصطناعي وتقييمها وتقديمها التأكيد على أنشطة إدارة المخاطر في الإدارة.

4. تحديد ما إذا كان التدريب يقوم بمساعدة المدققون الداخليون بصقل مهارات التفكير النقدي التي ستأتي في الطليعة بمجرد أن يتولى الذكاء الاصطناعي المهام الإدارية.

5. تقييم معايير التوظيف . للتأكد من أن الموظفين الجدد يمكنهم ذلك تصفية الخزين الأكبر من البيانات التي ستكون متاحة من خلال الذكاء الاصطناعي وتحديد علاقات البيانات المهمة بالنسبة إلى منظمة.

مقتطفات من التقارير التي تم إصدارها مؤخرًا من معهد AI Now التابع لجامعة نيويورك حول صناعة التكنولوجيا التي تحاول إعادة تشكيل المجتمع وفقاً لخطوط الذكاء الاصطناعي دون أي ضمان نتائج موثوقة وعادلة. خلص أحد التقارير إلى أن "الجهود المبذولة لإخضاع الذكاء الاصطناعي للمعايير الأخلاقية حتى الآن كانت فاشلة، وأن الأطر الأخلاقية الجديدة للذكاء الاصطناعي بحاجة إلى تجاوز

المسؤولية الفردية لمساءلة المصالح الصناعية والحكومية والعسكرية القوية أثناء تصميمها للذكاء الاصطناعي واستخدامه. " في رأي المؤلفين، يتم تقديم أنظمة الذكاء الاصطناعي في مختلف والمناطق المعرضة للخطر ، مثل الشرطة والتعليم والرعاية الصحية وغيرها البيئات ، حيث يمكن أن يؤدي "الخلل في الخوارزمية إلى تدمير الحياة." اطار(IIA, 2017, GPAI-AI-Part-III ,1).

ثانياً: ماذا يعني الذكاء الاصطناعي للتدقيق الداخلي؟

ستحتاج وظائف التدقيق الداخلي إلى الاعتراف بالتحول العالمي نحو الذكاء الاصطناعي والنظر في الدور الاستشاري / التأكيد الذي سيكون له في هذه الحركة. سيكون من الأهمية بمكان مواكبة التطبيق العملي للذكاء الاصطناعي في الأعمال التجارية وتطوير الكفاءات التي ستمكنهم من تقديم خدمات الاستشارات والضمان المتعلقة بالذكاء الاصطناعي لمؤسستهم. نظرًا لأن الذكاء الاصطناعي يعتمد على تحليل البيانات ومعالجة البيانات الضخمة، فإن بعض المخاطر الرئيسية المرتبطة بالذكاء الاصطناعي ستنشأ من المخاطر المتعلقة بإدارة البيانات والجودة والأمان، فضلاً عن المخاطر الخاصة بالذكاء الاصطناعي.

(Rodriquez, 2018)

يمكن للتدقيق الداخلي المشاركة من خلال ما لا يقل عن خمسة أنشطة مهمة ومتميزة تتعلق بالذكاء الاصطناعي: (IIA, 2017, GPAI-AI-Part-I ,5) (Mahmood, 2022)

1.بالنسبة لجميع المنظمات، يجب أن يشمل التدقيق الداخلي الذكاء الاصطناعي في تقييم المخاطر الخاص به والنظر فيما إذا كان يجب تضمين الذكاء الاصطناعي في خطة التدقيق القائمة على المخاطر.

2. بالنسبة للمنظمات التي تستكشف الذكاء الاصطناعي، يجب أن يشارك التدقيق الداخلي بنشاط في مشاريع الذكاء الاصطناعي منذ بدايتها، وتقديم المشورة والبصيرة التي تساهم في التنفيذ الناجح ، يجب ألا يمتلك التدقيق الداخلي ولا يكون مسؤولاً عن تنفيذ عمليات أو سياسات أو إجراءات الذكاء الاصطناعي.

3.بالنسبة للمنظمات التي طبقت بعض جوانب الذكاء الاصطناعي، إما ضمن عملياتها أو دمجها في منتج أو خدمة ، يجب أن يوفر التدقيق الداخلي ضماناً بشأن إدارة المخاطر المتعلقة بموثوقية الخوارزميات الأساسية والبيانات التي تستند إليها الخوارزميات.

4. يجب أن يتأكد التدقيق الداخلي من القضايا الاخلاقية والتحيز التي قد تحيط باستخدام المنظمة للذكاء الاصطناعي يتم تناولها.

5. مثل استخدام أي نظام رئيسي آخر سليم يجب إنشاء هياكل الحوكمة و يمكن أن يوفر التدقيق الداخلي ضمانًا في هذا المجال.

يحتاج التدقيق الداخلي إلى تضمين المخاطر المتعلقة بالذكاء الاصطناعي عندما تحدد المؤسسات أهدافها وتثبت تحمل مخاطر المؤسسة، ومواجهة التغييرات في تطبيق الذكاء الاصطناعي. يجب أن يأخذوا في الاعتبار استراتيجية الذكاء الاصطناعي والاستراتيجية الرقمية وحوكمة الذكاء الاصطناعي والعامل البشري والتحيز.

ثالثاً: إطار **عمل تدقيق الذكاء الاصطناعي لمعهد المدققين الداخليين(IIA)**

لمساعدة التدقيق الداخلي على أداء هذا الدور، يمكن للمدققين الداخليين الاستفادة من إطار عمل تدقيق الذكاء الاصطناعي التابع لمعهد المدققين الداخليين (IIA) في تقديم خدمات استشارية / توكيدية ذات صلة بالذكاء الاصطناعي أو ضمان أو خدمات استشارية بما يتناسب مع المنظمة. يتكون الإطار من ثلاثة مكونات شاملة –استراتيجية **الذكاء الاصطناعي، والحوكمة ، والعامل البشري** – وسبعة عناصر: **المرونة السيبرانية. كفاءات الذكاء الاصطناعي؛ جودة البيانات؛ هندسة البيانات والبنية التحتية؛ قياس الأداء؛أخلاق مهنية التحيز؛ والصندوق الأسود.** (IIA, 2017, GPAI–AI– Part–III ,2)

يتم تعيين كل عنصر من عناصر إطار عمل تدقيق الذكاء الاصطناعي في سياق إستراتيجية للذكاء الاصطناعي ويتم توضيحها أدناه.

إستراتيجية

ستكون إستراتيجية الذكاء الاصطناعي الخاصة بكل منظمة فريدة بناءً على نهجها في الاستفادة من فرص الذكاء الاصطناعي. يجب أن يأخذ التدقيق الداخلي في الاعتبار استراتيجية الذكاء الاصطناعي الخاصة بالمنظمة في البداية ويجب أن يساعد التدقيق الداخلي الإدارة ويدرك مجلس الإدارة أهمية صياغة إستراتيجية الذكاء الاصطناعي مدروسة تتوافق مع أهداف المنظمة , Rodriquez) . 2018)

يجب أن يتضمن التدقيق الداخلي للذكاء الاصطناعي جهداً لتحديد ما إذا كانت المنظمة قد أوضحت بوضوح إستراتيجيتها للذكاء الاصطناعي وما إذا كانت تعبر بوضوح عن النتيجة المرجوة لأنشطة الذكاء الاصطناعي. من المهم أيضاً التعرف على ما إذا كانت الاستراتيجية واقعية. تأخذ الإستراتيجية الواقعية في الاعتبار الكفاءات الداعمة اللازمة لتنفيذ مبادرات الذكاء الاصطناعي ويجب تطويرها بشكل تعاوني بين قادة الأعمال والتكنولوجيا لضمان عدم توقع أي منهما للآخر غير الواقعية(Alvero & Randy, 2019).

أخيراً، يجب أن تكون استراتيجية الذكاء الاصطناعي متسقة مع مهمة المنظمة وقيمها. يجب أن يكون التدقيق الداخلي في حالة تأهب للتعارضات المحتملة (أو التعارضات المتصورة) بين استراتيجية الذكاء الاصطناعي وقيم المنظمة المتعلقة بالعدالة والشفافية والخصوصية والتمييز ومواطنة الشركة. (Thompson ,2019)

حوكمة الذكاء الاصطناعي

تشير إلى تقييم الحوكمة الحالية وإجراء التعديلات عند الحاجة، بحيث تتم مراقبة الذكاء الاصطناعي والتعبير عن قيم المنظمة.

تشير حوكمة الذكاء الاصطناعي إلى الهياكل والعمليات والإجراءات المنفذة لتوجيه وإدارة ومراقبة أنشطة الذكاء الاصطناعي في المنظمة. سيختلف هيكل الحوكمة والشكلية بناءً على الخصائص المحددة للمؤسسة. حوكمة الذكاء الاصطناعي تسهم الى: (IIA, 2017, GPAI-AI-Part-II ,5)

1. تؤسس حوكمة الذكاء الاصطناعي المساءلة والرقابة.

2. تساعد على ضمان أن المسؤولين لديهم المهارات والخبرات اللازمة لمراقبة الذكاء الاصطناعي بشكل فعال .

3. تساعد على ضمان انعكاس قيم المنظمة في أنشطة الذكاء الاصطناعي الخاصة بها. يجب أن تؤدي أنشطة الذكاء الاصطناعي إلى قرارات وإجراءات تتماشى مع المسؤوليات الأخلاقية والاجتماعية والقانونية للمنظمة.

هندسة البيانات والبنية التحتية

من المرجح أن تعكس جميع بنية البيانات والبنية التحتية بنية المؤسسة للتعامل مع البيانات الضخمة. يتضمن اعتبارات: (IIA, 2017, GPAI-AI-Part-I ,7)

1. طريقة الوصول إلى البيانات (البيانات الوصفية، التصنيف، المعرفات الفريدة واصطلاحات التسمية) .

2. خصوصية المعلومات وأمنها طوال دورة حياة البيانات (جمع البيانات واستخدامها وتخزينها وإتلافها).

3. الأدوار والمسؤوليات لملكية البيانات واستخدامها طوال دورة حياة البيانات.

مقياس الاداء

نظراً لأن منظمة تدمج الذكاء الاصطناعي في أنشطتها، يجب تحديد مقاييس الأداء لربط أنشطة الذكاء الاصطناعي بأهداف العمل وتوضيح ما إذا كان الذكاء الاصطناعي يساهم بشكل فعال في تحقيق هذه الأهداف. يجب أن تراقب الإدارة بنشاط أداء أنشطة الذكاء الاصطناعي الخاصة بها.

في المنظمات التي تم فيها تنفيذ الذكاء الاصطناعي، يجب أن يوفر التدقيق الداخلي ضمان على خط الدفاع الأول وخط الدفاع الثاني الرقابة المتعلقة بالذكاء الاصطناعي. ربما لا توجد طريقة أفضل للتوضيح الداخلي تدقيق الكفاءة في الذكاء الاصطناعي بدلاً من استخدام تقنيات الذكاء الاصطناعي، مثل العملية الروبوتية الأتمتة لتدقيق الذكاء الاصطناعي.

جودة البيانات

تعد جودة البيانات مجالاً آخر يجب الانتباه إليه ، حيث يجب أن يوفر التدقيق الداخلي تأكيداً على شموليتها ودقتها.

يعد اكتمال البيانات التي تُبنى عليها خوارزميات الذكاء الاصطناعي ودقتها وموثوقيتها أمراً بالغ الأهمية. ليس من غير المعتاد أن يكون لدى المنظمة بنية غير متماسكة وغير واضحة التحديد لبياناتها. في كثير من الأحيان ، لا تكون الأنظمة متكاملة ولا تتواصل مع بعضها البعض وتقوم بذلك فقط من خلال الوظائف الإضافية المعقدة أو التخصيصات. إن كيفية تجميع هذه البيانات وتوليفها والتحقق من صحتها أمر بالغ الأهمية. 37(Rodriquez , 2018)

جميع حلول الذكاء الاصطناعي والبيانات التي تعتمد عليها موجودة في البنية التحتية ، سواء كانت مستضافة في أماكن العمل أو على السحابة. على هذا النحو ، تحتاج هذه البنية التحتية إلى إدارتها وصيانتها وتأمينها بشكل صحيح.

لتحقيق ذلك ، يجب أن تمتلك الفرق التي تدعم البنية التحتية للذكاء الاصطناعي مجموعة المهارات المناسبة مع توفير تدريب إضافي حسب الحاجة. يجب إدارة أي جهات خارجية تستضيف أو تدعم البنية التحتية للذكاء الاصطناعي أو الحلول بشكل استباقي ، ويجب منح حق الوصول للأفراد المصرح لهم فقط.

(Thompson ,2019)

المبحث الثالث : تدقيق التحيز في الذكاء الاصطناعي وفقاً لاطار معهد المدققين الداخليين (IIA)

أولاً : العامل البشري والتحيز في الذكاء الاصطناعي

بغض النظر عن الأنشطة المحددة المؤداة ، التدقيق الداخلي مناسب تماماً ليكون مساهماً رئيساً في أنشطة المنظمة المتعلقة بالذكاء الاصطناعي وذلك :(IIA, 2017, GPAI-AI-Part-I ,5)

1. يفهم الأهداف الإستراتيجية للمنظمة والعمليات المنفذة لتحقيق تلك الأهداف.

2. قادر على تقييم ما إذا كانت أنشطة الذكاء الاصطناعي كذلك تحقق أهدافهم.

3.يمكن أن توفر ضمانات داخلية على أنشطة إدارة المخاطر ذات الصلة بمخاطر الذكاء الاصطناعي وخاصة التحيز.

4. يُنظر إليه على أنه مستشار موثوق به يمكن أن يكون إيجابياً في دعم اعتماد الذكاء الاصطناعي لتحسين الأعمال أو تحسين عروض المنتجات والخدمات.

يعالج مكون العامل البشري ، الذي يتضمن عناصر الأخلاق والصندوق الأسود ، مخاطر الخطأ البشري الذي يضر بقدرة الذكاء الاصطناعي على تحقيق النتائج المتوقعة.

لا يزال العامل البشري يمثل مخاطرة عالية ، في إشارة إلى الأخلاق والتحيز والشفافية والصندوق الأسود الذي قد يعيق التوقعات بشأن إدخال الذكاء الاصطناعي. يمكن تسجيل الصناديق السوداء المدرجة في الروبوتات جميع تصرفاتهم والإبلاغ في حالة وقوع حوادث الظروف التي أدت إلى سلوكهم.

تعد إدارة المواهب أحد المخاطر الرئيسة التي يواجهها التدقيق الداخلي عند تنفيذ الذكاء الاصطناعي في المنظمات ، حيث إن المواهب التقنية هم من يديرون الخوارزميات الخاصة بالتعلم الآلي ،

باستخدام البيانات الضخمة ، وجمعها ووضع نماذج لها ، لكنهم ليسوا ماهرين في تحليل المخاطر والحوكمة.

يحتاج المرشحون للتدقيق الداخلي إلى امتلاك مهارات لمكافحة المخاطر المستقبلية ، والتفكير النقدي ، والتواصل الجيد ، ومهارات الإبداع والابتكار ، ومهارات تحليل البيانات ، وفي الوقت نفسه ، يحتاجون إلى احترام خصوصية وأمان بيانات المستخدمين وإيجاد طرق لتقييم أداء استخدام الذكاء الاصطناعي. (Alina & Cerasela,2018,444)

تم تطوير الخوارزميات من قبل البشر. سيؤثر الخطأ البشري والتحيزات على أداء الخوارزميات. يأخذ مكون العامل البشري في الاعتبار ما إذا كان (Rodriquez , 2018):

1. تحديد وإدارة مخاطر التحيزات البشرية غير المقصودة في تصميم الذكاء الاصطناعي .
2. تم اختبار الذكاء الاصطناعي بشكل فعال للتأكد من أن النتائج تعكس الهدف الأصلي .
3. يمكن أن تكون تقنيات الذكاء الاصطناعي شفافة بالنظر إلى التعقيد الذي تنطوي عليه .
4. يتم استخدام مخرجات الذكاء الاصطناعي بشكل قانوني وأخلاقي ومسؤول.

في تقرير عن الذكاء الاصطناعي ، أدرج معهد المدققين الداخليين التحيز كأحد المخاطر المهمة التي يجب مراعاتها. كما حذرت الجمعية من مخاطر تضمين أخطاء المنطق البشري ، والاختبار والرقابة غير الملائمين ، والضرر المالي أو الضرر بالسمعة.

يجب على المؤسسات أن تنظر بعناية في استخدام الذكاء الاصطناعي في التدقيق ، مع مراعاة قيود التكنولوجيا.

حتى مع مثل هذه التطورات الكبيرة ، فإن التدخل البشري مطلوب لإكمال التدقيق.

هناك عوائق يجب التغلب عليها(https://www.wipfli.com) :

1. التحيز البشري. هو خطر كبير يجب مراعاته تم بناء الذكاء الاصطناعي من قبل البشر ،
2. خطر غرس هؤلاء البشر التحيز عن غير قصد في التكنولوجيا أعلى مما يشعر به معظم الناس.

يتوقع مسح أجندة CIO لعام 2018 الذي أجرته شركة Gartner أن 85٪ من مشاريع الذكاء الاصطناعي حتى عام 2022 ستحقق نتائج خاطئة بسبب التحيز في البيانات أو الخوارزميات أو فرق التطوير.

ومع ذلك ، فقد تطورت التكنولوجيا الرقمية بطريقة تقلل بشكل كبير من مخاطر الإبلاغ. يمكن لشركات التدقيق تقديم خدمة ذات جودة أفضل ، وتحسين الموارد ، وتوسيع التغطية ، واكتشاف العيوب في وقت مبكر من العملية وتتبع عملية التدقيق الكاملة بسهولة (Paramasivam ,2020).

ثانياً : تحيز الخوارزمية

ستؤثر الخوارزميات التي طورها البشر والتي تتضمن خطأ بشرياً وتحيزات (سواء كانت مقصودة أو غير مقصودة) على أداء الخوارزمية.

إلى أي مدى يجب مراجعة خوارزميات الذكاء الاصطناعي، ومن قبل من، لا يزال موضوع نقاش. ولكن ، من الآمن أن نقول إن المدققين الداخليين يجب أن يتطلعوا إلى توفير بعض التأكيد على أن خوارزميات الذكاء الاصطناعي مصممة بكفاءة ، وأنها تعمل على النحو المتوقع ، وأنها شفافة بما يكفي للمستخدمين ، وأنهم لا يعرضون المنظمة للمخاطر من خلال النتائج غير مقصودة . لا يحتاج المدققون الداخليون إلى امتلاك الخبرة الموضوعية لمبرمج الخوارزمية، ولكن يجب أن يفهموا بشكل كافٍ عملية تطوير نظام الذكاء الاصطناعي لفهم ماهية هدف الخوارزمية، وما هي البيانات التي تستخدمها كمدخلات وما هي المعايير التي تستخدمها لاتخاذ القرارات / تنبؤات.

يجب أن يركز المدققون الداخليون على الحوكمة والضوابط حول تصميم الخوارزمية وأدائها والسعي للإجابة على أسئلة مثل :(رشيد ,220,2022)

1. هل الخوارزمية متحيزة بطريقة لا تتوافق مع مهمة الشركة أو أخلاقياتها أو قيمها؟

2. هل ينتج عن ذلك نتائج يمكن أن تدفع الشركة إلى المخاطرة التي لا تتماشى مع قابليتها للمخاطرة؟

3. هل يمكن أن تعرض الخوارزمية الشركة لمخاطر قانونية تتعلق بالسمعة من حيث صلتها بالعدالة أو الشفافية؟

يجب أن يكون المدققون الداخليون على دراية بمخاطر استخدام الذكاء الاصطناعي، تفتح خوارزميات الذكاء الاصطناعي الحديثة العديد من الاحتمالات الجديدة للأتمتة. ومع ذلك، في ضوء

دورهم ، من المهم أن يكون المدققون الداخليون على دراية بالمخاطر التي ينطوي عليها تطبيق مثل هذه الأنظمة في الشركة.

غالباً ما يكون لدى الناس انطباع بأن النظام الآلي محصن ضد الخطأ ، على عكس البشر. في الواقع ، يعلم جميع المدققين الداخليين بوضوح أن الضوابط الآلية دائماً ما تكون أفضل من الضوابط اليدوية ، ولكن في حالة خوارزميات الذكاء الاصطناعي ، ليس الأمر بهذه السهولة. لماذا يقدر المدققون الضوابط الآلية؟ لأنه يمكنهم التحقق من القواعد المبرمجة في الضوابط والتأكد من أن النظام لا يثني هذه القواعد.

في حالة خوارزميات الذكاء الاصطناعي، لا نعرف حقاً القواعد التي تلتزم بها الخوارزمية، لأنها تحدد هذه القواعد من تلقاء نفسها، وحتى مطور الخوارزمية ليس على دراية بها. لذلك، إذا تم استخدام خوارزميات الذكاء الاصطناعي لاتخاذ قرارات مهمة، فقد يوصي المدققون الداخليون بإجراء استئناف بشأن قرار الخوارزمية أو تصعيد السؤال ذي الصلة إلى شخص مخول بتغيير قرار الخوارزمية.

ثالثاً : الخوارزميات والقرارات المتحيزة

بينما تتعلم الخوارزميات من القرارات البشرية، غالباً ما تذهب تحيزات الناس إلى نتائج الخوارزمية. لذلك، إذا كانت الشركة تميل إلى تعيين رجال في مناصب إدارية، فإن الخوارزمية سوف "ترى" هذا الاتجاه وسترفض أيضاً السير الذاتية للنساء عند البحث عن مرشحين للمناصب الإدارية.

من أجل تقليل المخاطر المتعلقة بالتحيز المحتمل للخوارزميات، يجب على المدققين الداخليين تقديم المشورة للشركة لمراقبة أداء الخوارزمية بانتظام ضد مجالات التمييز المحتملة.

لذلك، كما نرى ، يفتح الذكاء الاصطناعي العديد من الفرص الجديدة للشركات ، لكن المدققين الداخليين بحاجة إلى أن يكونوا على دراية بالمخاطر المحتملة لهذه التقنيات والتدابير التي يمكن أن تساعد في تقليل المخاطر ذات الصلة. (CAAU , 2019)

وفقًا لتقرير حديث لشركة McKinsey & Company ، تسارع الشركات في تطبيق التعلم الآلي على اتخاذ القرارات التجارية. تضع البرامج خوارزميات معقدة للعمل على مجموعات بيانات كبيرة يتم تحديثها بشكل متكرر.

ومع ذلك، فإن التحيز الخوارزمي هو عمل محفوف بالمخاطر ، لأنه يمكن أن يضر بالغرض من التعلم الآلي إذا تم التغاضي عنه ، وتركه دون رادع .على سبيل المثال ، في التصنيف الائتماني ، يتم

تحديد الزبون الذي لديه تاريخ طويل في الاحتفاظ بالقروض بدون تأخير أو تقصير بشكل عام على أنه "منخفض المخاطر."

ومع ذلك ، فإن ما قد يكون غير مرئي هو أن الرهون العقارية لهذا الزبون قد تمت صيانتها ودعمها بمزايا ضريبية كبيرة من المقرر أن تنتهي صلاحيتها. قد يكون الارتفاع المفاجئ في حالات التخلف عن السداد وشيكاً وغير محسوب في نموذج المخاطر الإحصائي لمؤسسة الإقراض. من خلال الوصول إلى البيانات الصحيحة والإرشادات من قبل خبراء الموضوع ، يمكن لنماذج التعلم الآلي التنبؤية العثور على الأنماط المخفية في البيانات وتصحيح مثل هذه الارتفاعات. فضلاً عن ذلك ، خارج قرارات العمل ، يمكن أن يتسبب التحيز الخوارزمي في حدوث أخطاء قد تثير بعض المشكلات الحقيقية والاضطرابات بين المواطنين. على سبيل المثال ، تُستخدم خدمة صور Google والخدمات الأخرى المشابهة لتحديد الأشخاص والأشياء والمشاهد ، ولكن يمكن أن يحدث خطأ فادح ، مثل عندما تغفل الكاميرا عن علامة الحساسية العرقية ، أو عند استخدام برنامج تقييمات المخاطر للتنبؤ بالمجرمين في المستقبل الذين أظهروا تحيزاً .(رشيد, 220,2022)

رابعاً : الأخلاق والتحيز البشري ودور المدقق الداخلي :

عند تطوير الحلول واختبارها ، بما في ذلك البيانات والنماذج التي تدعمها ، يجب على مطوري الذكاء الاصطناعي الانتباه جيداً (Thompson, 2020):

1. للأخلاقيات والتحيز البشري. يتضمن ذلك فهماً شاملاً للتحيزات ، بما في ذلك تحيزات الفريق والبيانات وطرق التنفيذ وتأثير التحيز اللاواعي ، والذي يصعب التخفيف منه.

2. يجب مراعاة الأخلاقيات والجدارة بالثقة لاستخدام الذكاء الاصطناعي للمساعدة في اتخاذ القرارات قبل اعتماد حل ، ويجب أن تكون القرارات شفافة وقابلة للتفسير وقابلة للتدقيق عندما يكون ذلك ممكناً.

3. يجب أن تكون قوية وذات نواتج موثوقة وقابلة للتكرار.

4. يجب إعلام جميع الأشخاص الخاضعين لعمليات الذكاء الاصطناعي بهذا الأمر وفقاً للائحة العامة لحماية البيانات ، ولا ينبغي معالجة بياناتهم بواسطة الذكاء الاصطناعي أو بأي طريقة أخرى ، لأي أسباب أخرى غير الغرض الذي تم الحصول عليه من أجله في الأصل.

ستكون وظائف التدقيق الداخلي مطلوبة لتوفير ضمانات بشأن المخاطر المرتبطة باستخدام منظمتهم للذكاء الاصطناعي وتصميم العمليات القائمة على الذكاء الاصطناعي وأدائها ومراقبتها وحوكمتها. في ضوء ذلك ، تحتاج وظائف التدقيق الداخلي إلى البدء في الاستعداد لإدخال الذكاء الاصطناعي. سيتضمن ذلك اكتساب فهم لكيفية عمل الذكاء الاصطناعي ورفع مهاراتهم حتى يتمكنوا من تقديم المشورة والضمانات للأعمال التجارية بشأن التحديات والمخاطر التي سيجلبها الذكاء الاصطناعي. نظراً لمدى سرعة تطور الذكاء الاصطناعي ، ستكون هذه مهمة مستمرة لوظائف التدقيق الداخلي. سيكون من المهم أن يتمكنوا من مواكبة التطورات في الذكاء الاصطناعي حتى يتمكنوا من الاستمرار في تقديم خدمة في الوقت المناسب تضيف قيمة أيضاً.

يشكل هذا تحدياً حقيقياً لوظائف التدقيق الداخلي حيث أن العديد منها ليسوا مجهزين جيداً حالياً لتقديم التأكيد الذي يبحث عنه لجان التدقيق ومجلس الإدارة. يعد هذا الأمر أكثر أهمية الآن ، نظراً لأن عدداً من المنظمات بدأت للتو في الشروع في رحلة الذكاء الاصطناعي الخاصة بهم. (Hall,2022)

أهداف وأنشطة أو إجراءات الأخلاق والتحيز ذات الصلة(IIA, 2017, GPAI–AI–Part–III ,3)

هدف (أهداف) المشاركة أو الرقابة	الأنشطة أو الإجراءات
تقديم ضماناً بأن نتائج أنشطة الذكاء الاصطناعي الخاصة بالمنظمة خالية من التحيزات غير المقصودة.	*مراجعة النتائج المرجوة من أنشطة الذكاء الاصطناعي (الأهداف الاستراتيجية) وقارن مع النتائج الفعلية، إذا تم الكشف عن تباين ، فحدد ما إذا كان التحيز هو السبب.
يمكن للمنظمة أن "تصنع معنى" لمخرجات الذكاء الاصطناعي.	*مراجعة نواتج الذكاء الاصطناعي والمعنى المشتق من المخرجات.

خامساً: الصندوق الأسود

الصندوق الأسود هو "جهاز إلكتروني معقد عادة ما تكون آليته الداخلية مخفية عن المستخدم أو غامضة بالنسبة له ؛ على نطاق واسع: أي شيء له وظائف أو آليات داخلية غامضة أو غير معروفة ".

مع تقدم المؤسسات لتطبيق تقنيات الذكاء الاصطناعي من النوع الثالث والرابع – باستخدام الآلات أو الأنظمة الأساسية التي يمكنها التعلم بمفردها أو التواصل مع بعضها البعض – تصبح كيفية عمل الخوارزميات أقل شفافية أو مفهومة. سيصبح عامل الصندوق الأسود أكثر فأكثر تحدياً كملف تصبح أنشطة الذكاء الاصطناعي الخاصة بالمنظمة أكثر تعقيداً. , IIA, 2017, GPAI-AI-Part-II) 7)

لا تشتمل الأهداف والأنشطة أو الإجراءات ذات الصلة التي حددها معهد المدققين الداخليين (IIA) على خطة تدقيق محددة ، ولكنها أمثلة يجب أن تكون مفيدة في تحديد أهداف المشاركة أو الرقابة ، وفي تخطيط وتنفيذ ارتباطات تدقيق الذكاء الاصطناعي.

يجب أن تتوافق عمليات تدقيق الذكاء الاصطناعي مع معيار معهد المدققين الداخليين الدولي(IIA) 2200 :التخطيط للمشاركة.

أهداف وأنشطة أو إجراءات الصندوق الأسود ذات الصلة

(IIA, 2017, GPAI-AI-Part-III ,4)

هدف (أهداف) المشاركة أو الرقابة	الأنشطة أو الإجراءات
تقييم فهم المنظمة من بيانات "الصندوق الأسود" (أي المصدر الأساسي الخوارزميات أو الوظائف الداخلية أو الآليات التي تمكن الذكاء الاصطناعي).	* مراجعة تطوير الذكاء الاصطناعي وتنفيذ السياسات والعمليات والإجراءات والتحقق من بيانات الصندوق الأسود تم التعرف عليه. * مقابلة المسؤولين عن الذكاء الاصطناعي النتائج والتحقق من أنهم يفهمون ويمكن أن يشرح بيانات الصندوق الأسود.

سادساً : استخدام المعايير لتدقيق الذكاء الاصطناعي

يجب أن يلتزم المدققون الداخليون بجميع معايير معهد المدققين الداخليين المعمول بها عند التخطيط أو أداء ارتباطات الذكاء الاصطناعي. يتم تمييز معايير معهد المدققين الداخليين (IIA) ذات الصلة بشكل خاص بالذكاء الاصطناعي ، ولكن قد يتم تطبيق معايير أخرى أيضاً (IIA, 2017,) .GPAI-AI-Part-II ,13

تتضمن معايير IIA الدولية للممارسة المهنية للتدقيق الداخلي عدة معايير ذات صلة خاصة بالذكاء الاصطناعي ، بما في ذلك : (IIA, 2012)

*معيار: 1210 – الكفاءة

*معيار:2010 – التخطيط

* معيار :2030 – إدارة الموارد

*معيار :2100 – طبيعة العمل

*معيار :2110 – الحوكمة

* معيار :2130 –الرقابة

*معيار :2200 – تخطيط مهمة التدقيق الداخلي

*معيار :2201 – اعتبارات التخطيط

*معيار: 2210 – أهداف مهمة التدقيق الداخلي

*معيار: 2220– نطاق مهمة التدقيق الداخلي

*معيار :2230– تخصيص الموارد لمهمة التدقيق الداخلي

*معيار :2240 – برنامج عمل مهمة التدقيق الداخلي

*معيار: 2310– تحديد المعلومات

الاستنتاجات والتوصيات

أولاً : الاستنتاجات

1. التحيز سمة بشرية متأصلة ويمكن أن تنعكس في كل شيء نبتكره ، لا سيما عندما يتعلق الأمر بالذكاء الاصطناعي.

2. يمكن أن ينبع التحيز في خوارزميات الذكاء الاصطناعي من بيانات التدريب غير المكتملة أو الاعتماد على المعلومات المعيبة التي تعكس عدم المساواة التاريخية. إذا تركت الخوارزميات المتحيزة

دون رادع ، يمكن أن تؤدي إلى قرارات يكون لها تأثير جماعي متباين على مجموعات معينة من الناس حتى بدون نية المبرمج في التمييز.

3. غالباً ما يُنظر إلى التحيز في أنظمة الذكاء الاصطناعي على أنه مشكلة فنية، لكن في الحقيقة يقر بأن قدراً كبيراً من تحيز الذكاء الاصطناعي ينبع من التحيزات البشرية والتحيزات المنهجية والمؤسسية أيضاً.

4. أنظمة الذكاء الاصطناعي لا تعمل بمعزل عن غيرها. إنها تساعد الناس على اتخاذ القرارات التي تؤثر بشكل مباشر على حياة الآخرين. إذا أردنا تطوير أنظمة ذكاء اصطناعي جديرة بالثقة، فنحن بحاجة إلى النظر في جميع العوامل التي يمكن أن تقوض ثقة الجمهور في الذكاء الاصطناعي. تتعدى العديد من هذه العوامل التكنولوجيا نفسها إلى تأثيرات التكنولوجيا.

5. يتمثل دور التدقيق الداخلي في مساعدة المؤسسة على تقييم وفهم وإبلاغ الدرجة التي سيكون للذكاء الاصطناعي تأثير (سلبي أو إيجابي) على قدرة المؤسسة على خلق قيمة على المدى القصير أو المتوسط أو الطويل.

6. أدرج معهد المدققين الداخليين التحيزفي الذكاء الاصطناعي كأحد المخاطر المهمة التي يجب مراعاتها. كما حذرت الجمعية من مخاطر تضمين أخطاء المنطق البشري، والاختبار والرقابة غير الملائمين ، والضرر المالي أو الضرر بالسمعة.

7. يجب على المدققين الداخليين أن يتطلعوا إلى توفير بعض التأكيد على أن خوارزميات الذكاء الاصطناعي مصممة بكفاءة، وأنها تعمل على النحو المتوقع ، وأنها شفافة بما يكفي للمستخدمين ، وأنهم لا يعرضون المنظمة للمخاطر من خلال النتائج غير مقصودة .

8. لا يحتاج المدققون الداخليون إلى امتلاك الخبرة الموضوعية لمبرمجة الخوارزمية، ولكن يجب أن يفهموا بشكل كافٍ عملية تطوير نظام الذكاء الاصطناعي لفهم ماهية هدف الخوارزمية، وما هي البيانات التي تستخدمها كمدخلات وما هي المعايير التي تستخدمها لاتخاذ القرارات / تنبؤات.

9. من أجل تقليل المخاطر المتعلقة بالتحيز المحتمل للخوارزميات، يجب على المدققين الداخليين تقديم المشورة للشركة لمراقبة أداء الخوارزمية بانتظام ضد مجالات التمييز المحتملة.

10.هناك حاجة للمدققين الداخليين إلى أن يكونوا على دراية بالمخاطر المحتملة لتحيز تقنيات الذكاء الاصطناعي والتدابير التي يمكن أن تساعد في تقليل المخاطر ذات الصلة.

ثانيا: المقترحات

1. ضرورة قيام الجهات ذات العلاقة بمهنة التدقيق الداخلي بعقد الندوات وورش العمل المتخصصة للتعَّريف بنظم الذكاء الاصطناعي والتحيز في الخوارزميات ودور التدقيق الداخلي في ذلك وفقاً لاطار عمل تدقيق الذكاء الاصطناعي لمعهد المدققين الداخليين(IIA) .

2. ضرورة إصدار التعليمات والإرشادات من الجهات ذات العلاقة بالتدقيق بخصوص تحديات الذكاء الاصطناعي وخاصة ما يتعلق بالتحيز من خلال إعادة ضبط خططهم وخبراتهم ومنهجياتهم لتلبية ذلك.

3.ضرورة لفت نظر المدققين الداخليين الى اهمية تدقيق تحيز الذكاء الاصطناعي من أجل تقليل المخاطر المتعلقة بالتحيز المحتمل للخوارزميات في المنظمات المختلفة .

4. ضرورة قيام المنظمات المهنية والمعاهد المختصة بالتدقيق الداخلي بتطوير وتحديث مناهجها التدريبية والأكاديمية سنوياً وذلك لتمكين المدققين الداخلين القيام بمهام جديدة تتطلبها التقنيات مثل تدقيق الذكاء الاصطناعي وتحيزه.

المراجع

1. الجابر, غدير محمد عودة (2020) أثر الذكاء الإصطناعي على كفاءة النظم المحاسبية في البنوك الاردنية. رسالة ماجستير غير منشورة , جامعة الشرق الاوسط عمان الاردن.

2. رشيد, ناظم حسن (2022) الذكاء الاصطناعي – رؤية في التدقيق الداخلي , دار ابن الاثير للطباعة والنشر ,جامعة الموصل , العراق.

3- Alina, Carataș Maria & Cerasela, Spatariu Elena (2018) Internal Audit Role in Artificial Intelligence

"Ovidius" University Annals, Economic Sciences Series ,Volume XVIII, Issue

4 - Alvero ,Kevin M. (2019) Don't Let the Race to Embrace AI Overshadow the Needs to Audit Your Advancements, https://medium.com/nielsen-forward/dont-let-the-race-to-embrace-ai-overshadow-the-needs-to-audit-your-advancements -

5- Alvero, Kevin & Randy, Pierson (2019)Artificial Intelligence: Building the Foundation for Internal Audits that Deliver Value, https://www.corporatecomplianceinsights.com/ai-internal-audits-value/

6- Buçinca, Zana,& Malaya, Barbara, Maja, & Krzysztof Z. Gajos (2021)

To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making, Proceedings of the ACM on Human-Computer InteractionVolume 5Issue CSCW1April 2021 Article No.: 188 https://doi.org/10.1145/3449287

7- Brown, Charla Griffy& Chun, Mark (2020) Avoiding Bias and Identifying Risks when Deploying Artificial Intelligence, Journal Information Management/Technology (IT) VOLUME 23 ISSUE 1.https://gbr.pepperdine.edu

8- BBC News (2020)IBM Abandons 'Biased' Facial Recognition Tech, , https://www.bbc.com/news/technology-52978191

9- CAAU(2019) Risks of using artificial intelligence internal auditors should be aware of, (2019) https://www.pabu.com.ua/en/mediacentr-eng/proffesional-news/1451.

10- Deloitte, (2021) Building trust in AI: How to overcome risk and operationalize AI governance. https://www2.deloitte.com/content/dam/Deloitte/ca/Documents/financial-services/ca-omnia-ai-operation-trust-pov-aoda-en.pdf

11- Dastin ,Jeffrey (2018) "Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women," Reuters, 10 October 2018,

https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

12- Dilmegan, Cem (2020) Bias in AI: What it is, Types, Examples & 6 Ways to Fix it, https://research.aimultiple.com/ai-bias/

13- Hall, Patrick (2022) What We Learned Auditing Sophisticated AI for Bias , https://www.oreilly.com/radar/what-we-learned-auditing-sophisticated-ai-for-bias/

14- Hern, A.; (2020)Twitter Apologises for 'Racist' Image-Cropping Algorithm,", https://www.theguardian.com/technology/2020/sep/21/

15- Heather J. Meeker & Amit Itai(2020) Bias in Artificial Intelligence: Is Your Bot Bigoted https://news.bloomberglaw.com/tech-and-telecom-law/bias-in-artificial-intelligence-is-your-bot-bigoted

16- Institute of Internal Auditors IIA, (2017) "GLOBAL PERSPECTIVES AND INSIGHTS Artificial Intelligence – Considerations for the Profession of Internal Auditing, Part I

17 - Institute of Internal Auditors IIA, (2017) "GLOBAL PERSPECTIVES AND INSIGHTS The IIA's Artificial Intelligence Auditing Framework Practical Applications, Part II

18- Institute of Internal Auditors IIA, (2017) "GLOBAL PERSPECTIVES AND INSIGHTS The IIA's Artificial Intelligence Auditing Framework

, Practical Applications, Part II I .

19- Institute of Internal Auditors (IIA)2012, , «International Standards For Professional Practice Of Internal Auditing», USA

20- Issa, H., Sun, T., & Vasarhelyi , M. A. (2016). Research Ideas for Artificial Intelligence in Auditing: The Formalization of Audit and Workforce Supplementation. Journal of Emerging Technologies in Accounting, 13(2),

21- Kumari ,Annu (2022) Understanding Bias in Artificial Intelligence Models and Ways to Mitigate, https://www.marktechpost.com/2022/02/25/understanding-bias-in-artificial-intelligence-models-and-ways-to-mitigate.

22- Lee, D. Tay(2016) Microsoft Issues Apology Over Racist Chatbot Fiasco," BBC News, 25 March 2016, https://www.bbc.com/news/technology-35902104.

23- Lauret, Julien (2019) Amazon's sexist AI recruiting tool: how did it go so wrong? https://becominghuman.ai/amazons-sexist-ai-recruiting-tool-how-did-it-go-so-wrong-e3d14816d98e.

24- Merriam-Webster, "Bias ," accessed November 2, 2021.View in Article

25- Mach, Evalee ,(2019) How Artificial Intelligence Can Help Internal Auditing, https://avianaglobal.com/how-artificial-intelligence-can-help-internal-auditing/

26- Mahmood, Anam (2022) Tackling bias in machine learning models, https://developer.ibm.com/articles/tackling-bias-in-machine-learning-models/

27- Mahmood, Anam (2022) Tackling bias in machine learning model, https://developer.ibm.com/articles/tackling-bias-in-machine-learning-models

28 - Meeker ,Heather .J& Itai, Amit(2020) Bias in Artificial Intelligence: Is Your Bot Bigoted? https://news.bloomberglaw.com/tech-and-telecom-law/

29- Manyika, James, Silberg, Jake, & Presten, Brittany (2019) What Do We Do About the Biases in AI, HARVARD BUSINESS REVIEW, https://hbr.org

30- Omosa, Isaac (2020) Internal Audit and Assurance Manager at ActionAid International,
https://www.linkedin.com/pulse/artificial-intelligence-ai-from-internal-audit-i

31- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 25 Oct 2019, Vol. 366, Issue 6464

32- Puente, M. (2019). LAPD Predictive Policing Tool Raises Racial Bias Concerns. *Los Angeles Times*. https://www.govtech.com/public-safety/LAPD-Predictive-Policing-Tool-Raises-Racial-Bias-Concerns.html

33- Paramasivam , Kumar (2020), Reducing Audit Risk Through Artificial Intelligence and Automation, https://www.infosys.com/insights/ai-automation/reducing-audit-risk.html

34- Rodriquez ,Ojel (2018) Is your internal audit function artificially intelligent? https://www.grantthornton.pr/globalassets

35- Siwicki, Bill (2021) How AI bias happens – and how to eliminate it, https://www.healthcareitnews.com/news/how-ai-bias-happens-and-how-eliminate-it.

36- Sutaria, Niral (2022) Bias and Ethical Concerns in Machine Learning, ISACA JOURNA, VOLUME 4

37- Smith, Allen, J.D(2018) Audit Annually to Catch Bias in Artificial Intelligence,

https://www.shrm.org/resourcesandtools/legal-and-compliance/employment-law/pages/artificial-intelligence-diversity.aspx

38- Seneor, Abby& Mezzanotte, Matteo (2022) Open source data science: How to reduce bias in AI, https://www.weforum.org/agenda/2022/10/

39- **Thompson,** Stephen (2019) Artificial Intelligence: Building the Foundation for Internal Audits that Deliver Value, https://www.corporatecomplianceinsights.com/subscribe/

40- Thompson, Stephen (2020) Do you have an audit framework in place for AI? https://www.grantthornton.co.uk/insights/do-you-have-an-audit-framework-in-place-for-ai/

41 - https://masaar.net/ar/

42- https://www.emaratalyoum.com/technology/electronic-equipment/2018-11-02-1.1149993

43-https://www.scientificamerican.com/arabic/articles/news/how-artificial-intelligence-learns-racism-and-bias-from-human/

44- https://fihm.ai/