اشكالية التحيز الخوارزمي للذكاء الاصطناعي في الخدمات المالية

The problematic of AI Algorithmic Bias in Financial Services

سحنون خالد*

جامعة ابن خلدون – تيارت –الجزائر

Khaled.sahnoune@univ-tiaret.dz

بعلاش عصام

جامعة ابن خلدون – تيارت – الجزائر

assam.baalache@univ-tiaret.dz

تاريخ الاستلام: 2024/11/02 تاريخ القبول للنشر: 2024/11/21 تاريخ النشر: 2024/12/15

ملخص:

تهدف هذه الدراسة الى عرض تداعيات التحيز الخوارزمي على الخدمات المالية، كقضية أساسية معقدة وملحة تتطلب اهتماما متزايدا، وتفكيرا عميقا ومحايدا، وهذا من خلال مناقشة أشكال وحالات التحيز الخوارزمي في الخدمات الماليةوالاستراتيجيات التي تم اقتراحها وتطويرها للتخفيف منه.ومن بين النتائج المتوصل اليها: ان التحيز الخوارزمي مظهر من مظاهر التحيزات الاجتماعية في نتائج الخوارزميات، يجعل النتيجة في النهاية موضع شك وغير موثوقة أو ضارة أو تمييزية وأسبابه متعدد الأوجه فالعوامل المؤسسية والمجتمعية البشرية والنظامية تشكل مصادر مهمة للتحيز، كما أن مجرد فكرة أن الخوارزمية متحيزة في الخدمات المالية تؤدي إلى إدامة التفاوتات المجتمعية،وتحجب الأساس المنطقي وراء الاستشارات أو القرارات المالية المهمة، مما يقوض الثقة والمساءلة وتعريض سمعة هذه المؤسسات للخطر.وأن أحد المتطلبات الأساسية هو فهم العلاقة بين النظام والمعايير الأخلاقية والقانونية المعمول بها في السياقات ذات الصلة.

الكلمات المفتاحية: ذكاء اصطناعي، خوارزميات، تحيز، تحيز خوارزمي، خدمات مالية.

تصنيفات JEL:G39O33، G24.

Abstract:

This study aims to present the repercussions of algorithmic biason financial services, as a complex and urgent core issue that requires increased attention, and deep and neutral thinking.This is by discussing theforms and states of algorithmic bias in financial services and the strategies that have been proposed and developed to mitigate it.Among the findings : algorithmic bias is a manifestation of social biases in algorithmic outcomes, ultimately making the outcome questionable, unreliable, harmful, or discriminatory. Its causes are multifaceted. Human and systemic institutional and societal factors are important sources of bias. The mere idea that an algorithm is biased in financial services perpetuates societal inequalities, obscuring the rationale behind important financial consultations or decisions, undermining trust and accountability and jeopardizing the reputationof these institutions. And that one of the basic requirements is to understand the relationship between the system and the ethical and legal standards in force in the relevant contexts.

Keywords:Artificial Intelligence;Algorithms; Bias; Algorithmic Bias; Financial Services,

Jel Classification Codes:G39O33 ،G24

* المؤلف المراسل.

1. مقدمة

في السابق، كان مفهوم الذكاء الاصطناعي مقتصرا على الخيال العلمي، أما اليوم فقد أصبح متأصلا ومنتشرافي جوانب مختلفة لقدرته على إحداث ثورة بطرق لا حصر لها، ونظرا لما توصف به غالبا بأنها تقنيات موضوعية ومحايدة في اتخاذ قراراتها، فإن أحد التحديات الرئيسية التي تواجه تطوير ونشر أنظمة الذكاء الاصطناعي هو التحيز الخوارزمي وهو امكانية أن تقود النتائج أو التنبؤات أنظمة الذكاء الاصطناعي إلى قرارات غير عادلة أو غير متوازنة ولا تستند الى مبررات موضوعية. بالاظافة الى الضرر الذي سيتعرض له الكثيرين حول العالم في حياتهم ومصالحهم العملية والشخصية والاجتماعية، وأوضاعهم القانونية، وهذا ما أثار جدلا ومخاوف كبيرة كقضية أساسية معقدة تتطلب اهتماما متزايدا وتفكيرا عميقا كونه يتجاوز التوقعات مما أدى إلى تكثيف المناقشات حول العدالة والتحيز في انظمة لذكاء الاصطناعي حيث أصبحت التحيزات والتمييز المحتمل أكثر وضوحا، ومع ذلك، فإن عملية صنع القرار البشري في هذه المجالات وغيرها من الممكن أن تكون معيبة أيضا، وتتشكل بسبب التحيزات الفردية والمجتمعية التي غالبا ما تكون غير واعية.

كما مثل نشر الذكاء الاصطناعي في القطاعات المختلفة والمجالات الحساسة، حقبة تحولية تتميز بالكفاءة المعززة والابتكار. فالمؤسسات المالية تتم إعادة اختراعها وإعادة تعريفها بواسطة انظمة الذكاء الاصطناعي نظرا لطبيعتها القائمة على الدقة في اتخاذ القرار، وأصبحت تدرك بشكل متزايد اهمية تنوع التطبيقات والفوائد المحتملة لاستخدام خوارزميات الذكاء الاصطناعي والتي تتراوحبين تحسين وتخصيص تجارب العملاءوالمساهمة في أمن واستقرار الأنظمة المالية، وصولا إلى تطوير نماذج متطورة تسمح بتوفير الائتمان للعملاء الذين لم يكن من الممكن النظر إليهم في السابق. ومع ذلك، فإن هذا التكامل وطوفان البيانات الموجهة إليها يبرز أيضا اعتبارات أخلاقية معقدة تتحدى القيم الأساسية للخصوصية والإنصاف والشفافية وتظهر مخاوف التحيز الخوارزمي، والسبب الأكثر وضوحا هو أن مجرد فكرة أن خوارزمية الذكاء الاصطناعي متحيزة في الخدمات المالية يمكن أن تؤدي إلى إبعاد العملاء عن منتج أو خدمة تقدمها وتعريض سمعة هذه المؤسسات للخطر.

1.1 **إشكالية الدراسة:** بناءا على ماسبق يمكن طرح الاشكالية التالية:

ماهي تداعيات التحيز الخوارزمي لنظم ذكاء الإصطناعي على الخدمات المالية؟.

2.1 . أهداف الدراسة:

من أجل الإحاطة بأبعاد مشكلة الدراسة والتعمق أكثر في هذا المجال من علوم البيانات. فإن الدراسة تهدف إلى تسليط الضوء على المفاهيم الاساسية للتحيز الخوارزمي في نظم الذكاء الاصطناعي ووصفه وصفا دقيقا ملازما لطبيعته، وصولا الى الكشف عن المصادر الرئيسية المرتبطة بالتحيز والحالات الموجودة في المجالات البارزة مثل الصناعة المالية. واستنادا إلى الأدبيات المتنامية حول التحيز الخوارزمي تركز الدراسة بشكل خاص على تداعيات التحيز الخوارزميعلى الخدمات المالية ومناقشة الإستراتيجيات التي تم اقتراحها وتطويرها للحد منه.

3.1. أهمية الدراسة

لعل من أهم أسباب اختيارنا لهذا الموضوع هو ارتباطه الوثيق بالمواضيع الحديثة والمعاصرة. كما أن للدراسة أهمية خاصة حيث انه بالرغم من التطور الضخم الذي يساهم فيه الذكاء الاصطناعي في الكثير من المجالات، وعلى الرغم من أن عملية اتخاذ القرار القائمة على الخوارزميات أظهرت أداءا متفوقا في الدقة والقدرة على التعامل مع معلومات أكثر تعقيدا إلا أن واحدة من أهم المخاوف التي ظهرت في السنوات الاخيرة هي اشكالية التحيز الخوارزمي لنظم الذكاء الاصطناعي

واصبح العالم تتشابك فيه القرارات بين البشر والذكاء الاصطناعي. ومن المرجح بشكل كبير ان تأثيره بشكل أساسي سيكون على جميع مناحي حياة الفرد والمجتمع.

4.1. منهجية الدراسة

لتحقيق أهداف الدراسة تم الاعتماد على المنهج الوصفي والتحليلي باعتباره من أنسب المناهج لتقرير الحقائق والتعريف بمختلف المفاهيم ذات الصلة بالدراسة، ورد الفروع والجزئيات إلى أصولها العامة الواردة في الأبحاث والدراسات ذات الصلة، وبهدف الالمام والاحاطة بجوانب هذه الدراسة تم تقسيمها إلى ثلاثة محاور وهي:

- الذكاء الاصطناعي مقابل الخوارزميات.
- مفهوم التحيز والتحيز الخوارزمي.
- طبيعة التحيز الخوارزمي في الخدمات المالية.

2.الذكاء الاصطناعي مقابل الخوارزميات

مع استمرار مجتمعنا في التعقيد، وتوسع رقمنة البيانات على نطاق هائل،أصبح الذكاء الاصطناعي (AI) سمة مشتركة في الأعمال التجارية للعديد من العمليات التشغيلية(Varsha P. S, 2023). وعلى الرغم من التطورات في استخدام الذكاء الاصطناعي وإمكانية أتمتة القرارات التي كان البشر مسؤولين عن اتخاذها.ومع ذلك، فقد احتدم الجدل حول لما يعرف باسم التحيز الخوارزمي نتيجة لاستخدام الذكاء الاصطناعي.(Ming, Hui Huang & Roland , T. Rust;, 2021) وتعرضت عمليات الخوارزمية لانتقادات بسبب ميلها إلى تكثيف وإعادة إنتاج التحيز، وتشويه الحقائق، وعدم تناسق المعلومات، وغموض العملية(Ananny, M & Crawford, K, 2018).مما أدى إلى تفاقم الظلم الخوارزمي وإدامة الأنماط غير العادلة والتمييزية(Shin, D, Zaid, B, Biocca, F, & Rasul, A, 2022).

1.2. تطور مفهوم أنظمة الذكاء الاصطناعي

اقترح العديد من الأكاديميين وخبراء الصناعة تعريفات مختلفة للذكاء الاصطناعي كآلات ذات قدرات شبيهة بالإنسان(Mc Gettigan, T, 2017). واختبارا لمعرفة ما إذا كان بإمكان الآلة محاكاة الوظائف المعرفية البشرية. ومناقشة أهمية أجهزة الكمبيوتر التي تستهلك البيانات وتحاكي السلوك البشري (G. Batra, A. Queirolo, & N. Santhanam, 2018).والقدرة على التفكير وحل المشكلات والتعلم ودمج مهارات متعددة مثل الإدراك أو الذاكرة أو اللغة أو التخطيط إلى الذكاء الاصطناعي(W.S. Sarle, 1994). تمكنها من التعامل مع المواقف المعقدةبهدف اتخاذ قرارات عقلانية في سياقات مختلفة(Malik, N, Tripathi, S.N., Kar, A.K, & Gupta, S, 2022).

2.2. مفهوم الخوارزميات

يشتق مصطلح الخوارزمية من اسم عالم الرياضيات الفارسي محمد بن موسى الخوارزمي، ويستخدم المصطلح منذ آلاف السنين للإشارة إلى مجموعة مفصلة من التعليمات خطوة بخطوة لحل مشكلة أو إكمال مهمة.(Tabsharani, F, 2023) امافي عالم ما قبل الخوارزميات، كان البشر والمنظمات يتخذون القرارات تحكمها في كثير من الأحيان القوانين التي تنظم عمليات صنع القرار من حيث العدالة والشفافية والمساواة(Nicol Turner Lee, Paul Resnick, & Genie Barton, 2019).

ويعرف عالم التكنولوجيا الخوارزميات على أنها ببساطة تعليمات آلية، أو "وصفة مشفرة يتم تنفيذها عندما تواجه محفزا"، فهي عملية أو مجموعة من القواعد المحددة التي يتعين اتباعها في حل المشكلات وأداء هدف محدد بطريقة منظمة وتسلسل منطقي(Sales Philip, 2021).استنادا لنماذج رياضية ومجموعة من تقنيات الحوسبة (Arpan Kumar Kar,

(Shweta Kumari Choudhary, & Vinay Ku, 2022.فهي سلسلة معقدة من العمليات المستخدمة في التعلم الآلي لتحليل مجموعات البيانات الكبيرة لاستخلاص استنتاجات وإجراء التنبؤات.

كما تعتمد الخوارزميات على بيانات متعددة (بيانات تدريب)، ومن بيانات التدريب هذه، تتعلم الخوارزميات بعد ذلك نموذجا يمكن تطبيقه على أشخاص أو أشياء أخرى وتتنبأ بما يجب أن تكون عليه المخرجات الصحيحة لهم (Nicol Turner Lee, Paul Resnick, & Genie Barton, 2019).باستخدام الحساب ومعالجة البيانات والاستدلال الآلي"والتعلم من تلقاء نفسها إذا أعطيت بضعة تعليمات، لهذا تعد الخوارزميات أدوات شائعة وفعالة تستخدمها الشركات (Möller, J, Trilling, D, Helberger, N., & van Es, B, 2018)، لما تقدمه لصناع القرار منمعلومات أساسية و توقعات أو احتمالات أو بعض المجهولات الرئيسية الأخرى من أجل تحسين جودة القرارات.

3.2. الفروقات الجوهرية بين الخوارزميات والذكاء الاصطناعي

بالرغم ان الخوارزميات والذكاء الاصطناعي ليسا نفس الشيء، إلا أنهما مرتبطان ارتباطا وثيقاويتم تداول مصطلحي بشكل متكرر وأحيانا يستخدم الباحثون مفاهيم مترادفة مختلفة دلالياوبشكل متبادل، ويطلق على كل ما يتعلق بالخوارزميات إسم "الذكاء الاصطناعي"، ومع ذلك، فإن فهم الاختلافات الدقيقة بينهما أمر بالغ الأهمية للتنقل بين تعقيدات التكنولوجيا الحديثة.علاوة على ذلك، فالخوارزميات ليست حكرا على مجال علوم الكمبيوتر في تطوير الذكاء الاصطناعيفهي موجودة في العديد من التخصصات، من الرياضيات والفيزياء إلى علم الأحياء والاقتصاد، ويكمن جوهر الخوارزمية في قدرتها على تبسيط العمليات وتعزيز الكفاءة وتوفير نهج لحل المشكلات في تطوير الذكاء الاصطناعي. .

فالذكاء الاصطناعي هو كلمة شائعة، ومصطلح تسويقي إلى حد كبير، لان ما يسمى اليوم بالذكاء الاصطناعيهو في الواقع تعلم آلي(Aaron Patzer, 2023).ولكن بعضهم يخفي الخوارزميات تحت مسمى الذكاء الاصطناعي , (Kaya Ismail 2018).فالخوارزميات هي اللبنات الأساسية للذكاء الاصطناعي.تعمل في انسجام للعثور على إشارة بين ضجيج البيانات وإيجاد مسارات للحلول بمستويات من التعقيد كبديل عن المعالجة البشرية للمعلومات، وغالبا ما تكون مستوحاة من الظواهر البيولوجية أو المعرفية أو الاجتماعية، وتستخدم تقنيات مثل التعلم الآلي، والتعلم العميق، ومعالجة اللغة الطبيعية، أو التعلم التعزيزي.

وعلى الرغم من أن الذكاء الاصطناعي ينطوي على قدرات أكثر تقدما تتجاوز مجرد اتباع التعليمات، فالخوارزمية تحدد العملية التي يتم من خلالها اتخاذ القرار، ويستخدم الذكاء الاصطناعي بيانات التدريب لاتخاذ مثل هذا القرار (Kaya Ismail , 2018). من ناحية أخرى، ليست كل خوارزميات الذكاء الاصطناعي متماثلة (Harshit Baluja, 2023) ، تعمل جميع المهام التي يؤديها الذكاء الاصطناعي على خوارزميات محددة(Harshit Baluja, 2023). كما يهدف الذكاء الاصطناعي إلى إنشاء أجهزة كمبيوتر قادرة على معالجة المعلومات واتخاذ القرارات دون الحاجة إلى تعليمات من البشر. تمثل خوارزميات الذكاء الاصطناعي جميع التعليمات اللازمة للاستجابة للبيانات المقدمة إلى الآلة .(Coursera Staff, 2024)

كما ان الذكاء الاصطناعي هو تكوين من الخوارزميات التي يمكنها تعديل نفسها وإنشاء خوارزميات جديدة "استجابة للمدخلات والبيانات المكتسبة بدلا من الاعتماد فقط على المدخلات التي تم تصميمها للتعرف عليها كمحفزات"، هذه القدرة على "التغيير والتكيف والنمو بناءا على البيانات الجديدة" هي الذكاء الاصطناعي(Sales Philip, 2021).

4.2. الترابط بين الذكاء الاصطناعي والخوارزميات

يصبح التفاعل بين الذكاء الاصطناعي والخوارزميات واضحا عند النظر في قدرات أنظمة الذكاء الاصطناعي، فالعلاقة بين الذكاء الاصطناعي والخوارزميات علاقة تكافلية، تمثل تفاعلا ديناميكيا بين الأنظمة الذكية والعمليات القائمة على القواعد. في حين يستغل الذكاء الاصطناعي قوة الخوارزميات المعقدة، فإن هذه الخوارزميات بدورها تعمل كأساس للوظائف المتنوعة التي يعرضها الذكاء الاصطناعي، وستفيد أنظمة الذكاء الاصطناعي من قوة الخوارزميات المتطورة لتحليل البيانات والتعرف على الأنماط والتنبؤ وتمكين الأنظمة من التكيف وتحسين أدائها بمرور الوقت، بناءا على الخبرة، يسمح هذا التآزر الديناميكي بإنشاء أنظمة ذكية قادرة على التكيف والتعلم من بيئتها .

وعليه، يشكل الذكاء الاصطناعي والخوارزميات قوتين متكاملتين تدفعان الابتكار التكنولوجي، ويمثل الذكاء الاصطناعي المفهوم الأوسع للأنظمة الذكية، في حين توفر الخوارزميات الأطر المنظمة التي تمكن هذه الأنظمة من أداء مهام محددة، فإدراك الاختلافات بينهما أمر ضروري لكل من المطورين والمستخدمين، مما يعزز الفهم الدقيق للتكنولوجيات التي تشكل عصرنا الرقمي ولكن بسبب تعقيد العمليات التي تنفذها، لا يمكن للبشر دائما التنبؤ بنتيجة البرمجة، مهما كانت درجة اطلاعهم (Sales Philip, 2021).

5.2. أنواع خوارزميات نظم الذكاء الاصطناعي

تعمل جميع خوارزميات الذكاء الاصطناعي من خلال ثلاث فئات رئيسية والفرق بينهما هو كيفية تدريبها وطريقة عملها.وتشكل هذه الأنواع من الخوارزميات معا مجالات مختلفة من الذكاء الاصطناعي ويمكن ذكرها كالاتي:

1.5.2.خوارزميات التعلم الخاضع للإشراف

وهي الأكثر شيوعا، يشير إسم "الإشراف" إلى العمل تحت إشرافريق من الخبراء المتخصصين وعلماء البيانات لاختبارها والتحقق من الأخطاء. ويقوم المطورون بتدريب البيانات مصنفة ومسماة بوضوح، لتحقيق الأداء الأقصى ثم اختيار النموذج(Harshit Baluja, 2023)حيث ترتبط بيانات الإدخال بالإخراج الصحيح. وتشمل أنواع خوارزميات التعلم الخاضع للإشراف ما يلي:(Coursera Staff, 2024)

- **شجرة القرار**:عبارة عن مخطط ذو شكل متفرع يمثل جميع النتائج الممكنة، حيث يمثل كل انقسام أو عقدة اختبار تصنيف مختلف.
- **الغابة العشوائية** : تستخدم خوارزمية الغابة العشوائية العديد من أشجار القرار، حيث تقوم كل منها باختبار مدخلات مختلفة. وتقوم بعمل تنبؤ بناءا على النتائج المجمعة لجميع أشجار القرار .
- **الانحدار الخطي** : يعد الانحدار الخطي أحد أكثر خوارزميات الذكاء الاصطناعي الأساسية، حيث يقوم بالتنبؤ بناءا على متغير مستقل يحدده مشغل الخوارزمية. على سبيل المثال، يمكن للانحدار الخطي التنبؤ بأسعار بيع المنازل باستخدام بيانات العقارات التاريخية في الحي والعقار الفردي المعروض للبيع .

2.5.2. خوارزميات التعلم غير الخاضع للإشراف

في التعلم غير الخاضع للإشراف، وهو مجال يتطور بسرعة جزئيا بسبب تقنيات الذكاء الاصطناعي التوليدي الجديدة تتعلم الخوارزمية من مجموعة بيانات غير مسماة من خلال تحديد الأنماط أو الارتباطات أو المجموعات داخل البيانات يستخدم هذا النهج عادة لمهام مثل التجميع وتقليل الأبعاد واكتشاف الشذوذ. تتضمن أمثلة خوارزميات التعلم غير الخاضعة للإشراف التجميع باستخدام طريقة k-means، وتحليل المكونات الأساسية(PCA) ، والمشفرات التلقائية.(Tabsharani, F, 2023)

3.5.2. التعلم التعزيزي

باستخدام التعلم التعزيزي، يمكن للخوارزمية أن تقرر أفضل طريقة لإنجاز المهمة بشكل مستقل، حيث تتدرب الخوارزمية وتتعلم من البيئة وتتلقى ردود الفعل في شكل مكافآت أو عقوبات لتعديل أفعالها في النهاية بناءا على ردود الفعل وهذا يسمح باتباع نهج التجربة والخطأ لحل المشكلات، هذا الشكل من خوارزميات هو الأكثر ملاءمة عندما لا تكون أفضل طريقة ممكنة لحل مشكلة واضحة. ويضع مبرمجو الكمبيوتر قواعد المكافأة والعقاب، لكن الخوارزمية تقرر الطريقة المثلى للعمل مع مجموعة البيانات .(Harshit Baluja, 2023)

3. مفهوم التحيز والتحيز الخوارزمي

1.3.الدلالة اللغوية لمصطلح التحيز

"التحيز" هو مصطلح مثقل ويعني أشياء مختلفة بشكل ملحوظ في سياقات مختلفة، وهو مصدر الفعل «تحيز»، ولقد ورد في المعاجم المعاصرة بمعنى التنحي بالقول وبمعنى الانضمام والموافقة على الرأي، وتبني رؤية ما، ولا يبدو في تلافيفه وتفاصيله أية معان سلبية، فهو الانتقال من شيء إلى آخر، هذا الشيء قد يكون موقفا، مكانا أو انتماء(عبد الوهاب المسيري ، 2024)، اما في اللغة الإنجليزية غالبا ما يكون لكلمة "التحيز" دلالة سلبية، التحيز هو شيء يجب تجنبه، أو أن هذا بالضرورة يمثل مشكلة(London & Danks, 2017). وعلى نفس المنوال، قد تختلف مفاهيم التحيز عبر الأديان والثقافات والمنظمات والأنظمة القانونية.

كما ان التحيز ليس بعيب أو نقيصة الا ان العقول تعاقدت على ذلك بل على العكس يمكن أن يجرد من معانيه السلبية(صديقي، علي، 2011)، فهو جزء لا يتجزأ من الطبيعة الإنسانية، فالتحيز حتمي ذلك لأنه مرتبط ببنية عقل الإنسان ذاتها، فكل ما هو إنساني يحوي على قدر من التفرد والذاتية ومن ثم التحيز(زيدان رغداء، 2010)، غير أنه رغم حتمية التحيز، وعدم إمكانية استبعاده بشكل تام، فإنه ليس نهائيا، ويمكن التقليل من آثاره، لأن ثمة مرجعية نهائية مشتركة، فهو «ليس نهاية المطاف، فالنهائي هو الإنسانية المشتركة (والقيم الأخلاقية) التي تسبق أي تحيز(صديقي، علي، 2011).

هناك معاني واستخدامات مختلفة لمصطلح "التحيز"، والتحيز هو موقف أو شعور تجاه مجموعة دون أسس عادلة وأدلة كافية في سياقات متعددة(Fiske, S. T, 1998)، ويمكن أن يشير إلى أي شكل من أشكال التفضيل، سواء كان عادلا أو غير عادل(Jake Silberg & James Manyika, 2019)، كما ان التعرف على التحيز والحد منهما مهمة صعبة لأنه يختلف بين الثقافات، ونتيجة لذلك، تتأثر معايير التحيز بتجربة المستخدم والعوامل الثقافية والاجتماعية والتاريخية والسياسية والقانونية والأخلاقية، فهو خطأ منهجي يغير سلوكيات الإنسان أو أحكامه على الآخرين بسبب انتمائه إلى مجموعة محددة بخصائص مميزة مثل الجنس أو العمر.(Tiago Palma Pagano , Rafael Bessa Loureiro, & and all, 2023)

ويشير "التحيز" ببساطة إلى الانحراف عن المعيار. وبالتالي، يمكن أن يكون لدينا تحيز إحصائي ينحرف فيه التقدير عن معيار إحصائي، التحيز الأخلاقي الذي ينحرف فيه الحكم عن معيار أخلاقي؛ وبالمثل بالنسبة للتحيز التنظيمي أو القانوني والتحيز الاجتماعي، والتحيز النفسي، وغيرها، بشكل عام، هناك العديد من أنواع التحيز اعتمادا على نوع المعيار المستخدم كما يمكن أن يكون الشيء نفسه متحيزا وفقا لمعيار واحد، ولكن ليس وفقا لمعيار آخر(London & Danks, 2017).

2.3. مفهوم التحيز الخوارزمي

يتم اتخاذ العديد من القرارات المهمة بمساعدة أنظمة المعلومات التي تستخدم الذكاء الاصطناعي ونماذج التعلم الآلي.ولكن زيادة تعقيد وغموض هذه الانظمة أصبح من الصعب بشكل متزايد فهم كيفية وصولها إلى قرارات مستقلة (London & Danks, 2017). واصبحت هناك مخاوف متزايدة بشأن التحديات الأخلاقية التي تواجهها نماذج صنع القرار الآلي ، وتتمثل إحدى هذه المخاوف في ضمان عدالة قرارات النموذج وخلوها من التحيز الخوارزمي(Tiago Palma Pagano , Rafael Bessa Loureiro, & and all, 2023)

فالتحيز الخوارزمي كمصطلح تم تعريفه لأول من طرف **Trishan Panch and Heather Mattie** في برنامج في جامعة هارفارد تي إتش (مدرسة تشان للصحة العامة). (Stephen J. Bigelow, Alexander S. Gillis, & Mary K. Pratt)لكن كمفهوم له جذوره في الظواهر الاجتماعية مثل التمييز والظلم الاجتماعي.وعندما يتم تصور التحيز الخوارزمي وقياسه بناءا على نهج بعض الفلاسفة واعتمادا على تعريف النموذج للمساواة والعدالة، قد يتم صياغة التحيز الخوارزمي بشكل مختلف.

وكونه ظاهرة اجتماعية تقنية له تعريفات تختلف عبر النظم الاجتماعية والنماذج الفلسفية (Maddalena Favaretto, Eva De Clercq , & Bernice Simoe Elger, 2018). ويشمل جانبها الاجتماعي التحيزات التي كانت موجودة منذ فترة طويلة في المجتمع والتي تؤثر على فئات معينة مثل المجتمعات المحرومة والمهمشة، والوقوف ضد أو لصالح الأفراد أو الجماعات بناءا على هوياتهم الاجتماعية مثل العرق والجنس والجنسية. في حين أن جانبها التقني المرتبط بالنصوص الخوارزمية ينطوي على مظهر من مظاهر التحيزات الاجتماعية في نتائج الخوارزميات(Binns R, 2018).

يشار للتحيز الخوارزميأيضا باسم تحيز التعلم الآلي أو تحيز الذكاء الاصطناعي.فهو ظاهرة تشير إلى الأخطاء المنهجية التي تحدث في عمليات صنع القرار، مما يؤدي إلى نتائج غير عادلة. وتولدتنبؤاتغيرعادلة.او افتراضات متأصلة وخاطئة في عملية التعلم الآلي(Shashkina Victoria, 2024). تؤدي إلى تحريف المخرجات، أو التأثير على النتائج المقدمة أو الحد منها.

ونظرا لأن المجتمع متحيز، فإن الكثير من البيانات التي يتم تدريب الذكاء الاصطناعي عليها تعكس وتديم التحيزات البشرية داخل المجتمع وأحكامه المسبقة ، بما في ذلك عدم المساواة الاجتماعية التاريخية والحالية ، لذلك يتعلم تلك التحيزات وينتج نتائج تدعمها(Flori Needle, 2023) . ويمكن العثور على التحيز في بيانات التدريب الأولية أو الخوارزمية أو التنبؤات التي تنتجها الخوارزمية.وهذا يعني ان يكون إدخال البيانات متحيزا، ومن المرجح أن يكون الإخراج متحيزا (Ming, Hui Huang & Roland , T. Rust;, 2021)مما يؤدي إلى مخرجات مشوهة ونتائج تمييزية غير صحيحة ضد فئات معينة.بما في ذلك الفوارق العرقية والجنسانية والاجتماعية والاقتصادية، وهي قضية ملحة في التفاعل بين الإنسان والذكاء الاصطناعي، بسبب غموض أنظمة الذكاء الاصطناعي(Hou Tsung-Yu, Tseng Yu-Chia, Yuan Chien Wen (Tina), & , 2024).

فخوارزميات الذكاء الاصطناعي مثل البشر، بعيدة كل البعد عن الكمال كونها ترث التحيزات من البشر.و غالبا ما ينظر إلى التحيز باعتباره مشكلة إنسانية فهو نتاج أدمغة غير كاملة، وليس أنظمة ذكاء اصطناعي محايدة كما يفترض.لكن نماذج الذكاء الاصطناعي لا تعكس التحيزات البشرية فحسب، بل يمكنها تضخيمها على نطاق واسع، بطرق يصعب اكتشافها ومنعها.(Lev Craig, 2023) وبقدر ما يتم تحديد التحيز من خلال عملية صنع القرار وليس فقط من خلال النتائج، فإن هذا الغموض قد يتحدى فكرة الرصد البشري للتحيز.(London & Danks, 2017)

3.2. عوامل التحيز الخوارزمي ومصادره

التحيز يمكن أن يتسلل خلال جميع مراحل المشروع، سواء من خلال تحديد المشكلة التي يجب حلها بطرق تؤثر على الفئات بشكل مختلف، أو الفشل في التعرف على التحيزات الإحصائية أو معالجتها، أو إعادة إنتاج التحيز الماضي، أو النظر في مجموعة غير غنية بما فيه الكفاية من العوامل(Solon Barocas & Andrew D Selbst , 2016).

لهذا يعد تحيزالخوارزمي مشكلة معقدة، ومن الضروري فهم العوامل والمصادر، فالتصنيف الأكثر شيوعا للتحيز الخوارزمي يأخذ مصدر التحيز كمعيار أساسي، ويضع التحيزات الخوارزمية في ثلاث فئات: تحيز البيانات، وتحيز تصميم الخوارزمية، والتحيز البشري. ومع ذلك، يحث الباحثون والممارسون في مجال الذكاء الاصطناعي على البحث عن الخيار الأخير، لأن التحيز البشري يكمن وراء التحيزين الآخرين ويتفوق عليهما(Shashkina Victoria, 2024)، وهو ما يمكن أن يؤدي بدوره إلى مواقف خطيرة، والمهم أن ندرك أن التحيز يمكن أن يحدث من أحد المصادر الرئيسية وهي:

1.3.3. التحيزاتف بالبيانات

يعتمد أداء نظام الذكاء الاصطناعي بشكل كبير على البيانات الأساسية المستخدمة في التدريب النموذجي، فأحد الطرق المؤدية إلى التحيز الخوارزمي هو من خلال الانحرافات في التدريب أو إدخال البيانات المقدمة إلى الخوارزمية، يتم تدريب الخوارزميات أو تعلمها لاستخدامات أو مهام معينة وبالتالي تؤدي إلى استجابات متحيزة لتلك المهام (London & Danks, 2017).

فعندما يتم جمع البيانات المستخدمة لتدريب نماذج التعلم الآلي من مصادر متحيزة أو عندما تكون البيانات غير كاملة، أو تفتقد الى معلومات مهمة (غير تمثيلية)، أو التمثيل المفرط أو تحتوي على أخطاء (Kate Crawford & Ryan Calo, 2016)، مما يعني أن النموذج النهائي سيعكس التحيزات في تلك البيانات وكيفية جمعها(Lev Craig, 2023) .

يمكن أن ينشأ التحيز في البيانات من مجموعة متنوعة من المصادر وهي:

- **التحيز التاريخي في البيانات**: تتشكل بسبب عوامل تاريخية. وترجع في المقام الأول إلى أن البيانات التاريخية المستخدمة لتدريب هذه الخوارزميات غالبا ما تعكس التحيزات المتراكمة في صنع القرار البشري، والتي يمكن أن تؤدي إلى إعادة إنتاجها وتضخيمها في النماذج الحاسوبية. وعلاوة على ذلك، قد تتعزز التحيزات وتستمر دون علم المستخدم (Nicol Turner Lee, Paul Resnick, & Genie Barton, 2019).
- **التحيز المسبق**: تتعلم هذه الخوارزميات بشكل أساسي ما تم تدريسه من قبل البشر. وبالتالي يمكن أن تحمل جميع التحيزات الواعية أو اللاواعية من الحكم البشري.
- **تحيز العينة**: تحدث هذه مشكلة عندما لا تكون البيانات المستخدمة لتدريب النموذج كبيرة بما يكفي، أو غير ممثلة بدرجة كافية، أو غير مكتملة جدا بحيث لا يمكنها تدريب النظام بشكل كاف
- **تحيز القياس**: يحدث عندما تعكس البيانات الأساسية مشاكل تتعلق بكيفية تقييمها أو قياسها أو جمعها أو تخزينها او بسبب عدم اكتمال البيانات وغالبا ما يكون هذا سهوا أو نقصا في الإعداد، مما يؤدي إلى عدم تضمين مجموعة البيانات التي ينبغي أخذهما في الاعتبارمما يؤدي إلى نتائج مشوهة.
- **تحيز الوكيل**: يعني أن مجرد استبعاد السمات المحمية للأفراد في البيانات قد لا يلغي أو حتى يخفف من التأثير المحتمل للتحيز على مجموعات معينة.

❖ **التحيز الناشئ** : يشير هذا إلى التحيز الذي يظهر بعد فترة من استخدام الخوارزمية، نتيجة لتغيير المعرفة المجتمعية أو السكان أو القيم الثقافية، ونتيجة لذلك، قد لا تعكس البيانات المختارة والمستخدمة داخل الخوارزمية ظهور المعرفة الجديدة أو القيم المجتمعية.

❖ **التحيز التأكيدي**: ويحدث عندما يعتمد الذكاء الاصطناعي كثيرا على المعتقدات أو الاتجاهات الموجودة مسبقا في البيانات، مما يضاعف التحيزات الموجودة، ويصبح غير قادر على تحديد أنماط أو اتجاهات جديدة .

2.3.3. **التحيزات في التصميم الخوارزمي**

يمكن أن يؤدي إنشاء خوارزميات تنتج أخطاء بشكل متكرر، أو نتائج غير عادلة، أو حتى تضخيم التحيز الكامن في البيانات المعيبة، بسبب أخطاء البرمجية، مثل قيام المطور بترجيح العوامل بشكل غير عادل في اتخاذ القرار الخوارزمي بناءا على تحيزاته الواعية أو اللاواعية، وتشمل التحيزات في التصميم الخوارزمي ما يلي:

❖ **التحيز الموضوعي:** يمكن أن ينشأ من هدف الخوارزمية نفسها. على سبيل المثال، قد تهدف الخوارزمية المستخدمة لتقييم مخاطر التخلف عن سداد الائتمان لمقترضي الرهن العقاري إلى تعظيم معدل النجاح الإجمالي للتنبؤات. ومع ذلك، قد يعني الإفراط أو النقص في تسجيل مجموعات معينة أن الخوارزمية أكثر دقة لبعض المجموعات منغيرها مما يؤثر على نتائجها على الرغم من زيادة الدقة الإجمالية عبر جميع المجموعات.

❖ **تحيز الترجيح:** يعكس تحيز الترجيح حقيقة أن الأوزان غالبا ما يتم تطبيقها على الميزات الموجودة في الخوارزمية، وإذا لم يتم تطبيقها بشكل صحيح، فيمكن أن تتأثر النتائج.

❖ **تحيزالتقييم:** ينشأ تحيز التقييم من الأساليب المستخدمة لتقييم الخوارزميات. من الممارسات الشائعة تقسيم البيانات إلى مجموعات بيانات التدريب والاختبار للتحقق من الأداء الخوارزمي، ومع ذلك، يمكن أن تؤدي البيانات المقسمة بشكل غير لائق إلى تحيز العينة الموصوفة سابقا مما يؤدي إلى إنشاء عينة فرعية من البيانات التي انتقلت من عدم وجود تحيز، إلى تقديم التحيز.

❖ **تحيز التجانس خارج المجموعة** : ينشأ هذا التحيز نتيجة قيام المطورين بإنشاء خوارزميات أقل قدرة على التمييز بين الأفراد الذين لا يشكلون جزءا من مجموعة الأغلبية في بيانات التدريب، مما يؤدي إلى التحيز العنصري وسوء التصنيف والإجابات غير الصحيحة .

3.3.3. **التحيز في الاستخدام البشري**

غالبا ما يتم تفسير النتائج الخوارزمية واستخدامها من قبل البشر، والبشر غير معصومين من الخطأ، ونتيجة لذلك يمكن أن يتسرب التحيز الشخصي دون أن يدرك الممارسون ذلك، يمكن أن يؤثر هذا على مجموعة البيانات أو سلوك النموذج، وتشمل هذه التحيزات تحيز القرار وتحيز المعتقد وتحيز التفسير، وفقا لتقرير المعهد الوطني للمعايير والتكنولوجيا (NIST)، أشار ان مصدر التحيز هذا هو أكثر شيوعا، وأن " العوامل المؤسسية والمجتمعية البشرية والنظامية تشكل مصادر مهمة للتحيز في مجال الذكاء الاصطناعي، وتشمل هذه العوامل تأثير السياقات الاجتماعية والاقتصادية والثقافية على كيفية تصميم أنظمة الذكاء الاصطناعي ونشرها واستخدامها.

4.2.آثار التحيز الخوارزمي

بالرغم من الفوائد العديدة لخوارزميات للذكاء الاصطناع (AI) ، ولكنه تجلب معها مخاطر وتحديات محتملة، أحد المخاوف الرئيسية هي الآثار السلبية للتحيز في الذكاء الاصطناعي التي يمكن أن تؤدي إلى قرارات مضللة وعواقب سلبية على الأفراد والمنظمات والمجتمع اهمها(Nima Kordzadeh & Maryam Ghasemaghae, 2022):

❖ **الاثار الفردية:** تشمل العواقب الفردية حرمان وصول الافراد إلى الخدمات الأساسية مثل القروض والائتمان والمزايا الأخرى ودفعا ولأسعار أعلى من المعتاد وعدم المساواة المهنية التي تؤثر على الأقليات أو فرص العمل أو حتى إلى افتراضات واتهامات واعتقالات وإدانات غير مشروعة مما يؤدي إلى التمييز ضد الفئات المهمشة وبالتالي يؤثر على نظرة الناس لأنفسهم وللآخرين وعلى فرصهم وتفاعلاتهم، ويمكن أيضا أن يكون مصدرا لرسائل تسويقية محرجة أو خاطئة.

❖ **الآثار التنظيمية:** تتضمن التأثيرات التنظيمية انتهاك سياسات تكافؤ الفرص، وخلق مناخ غير أخلاقي، ومعدل دوران مرتفع للموظفين، ومعدل مرتفع لفقدان العملاء بسبب التمييز الخوارزمي وعدم الرضا، يمكن أن ينتهي الأمر بالمنظمات إلى زيادة أو نقص المعروض من المواد الخام أو المخزون بسبب ضعف توقعات طلب العملاء، كما يمكن أن تؤدي الأخطاء إلى انخفاض مستويات الثقة في التعلم الآلي ومقاومة اعتماد الذكاء الاصطناعي.

❖ **التأثيرات على المستوى المجتمعي:** تشمل التأثيرات على مستوى المجتمع فجوة الثروة المتزايدة بين المجموعات المحرومة تاريخيا وغيرها ويمكن أن يؤدي الاستخدام الواسع النطاق لأنظمة الذكاء الاصطناعي المتحيزة إلى ترسيخ وإدامة الصور النمطية والتمييز بين الجنسين أو حتى تضخيم الروايات التمييزية على أساس لون البشرة أو العرق أو المظهر الجسدي أو أولئك الذين ينتمون إلى خلفيات اجتماعية واقتصادية أقل(Joy Buolamwini & Timnit Gebru, 2018). وعرقلة الجهود الرامية إلى تحقيق المساواة والشمولية(Latanya Sweeney, 2013). و تعرض الأشخاص في تلك المجموعات والمجتمع ككل للأذى دون أن يدركوا ذلك. كما ان المنظمات تعاني من الضرر الذي يلحق بعلامتها التجارية وسمعتها، وفي الوقت نفسه، الأضرار المالية اللاحقة.

5.3. اتجاهات واستراتيجيات في معالجة التحيز الخوارزمي

يتخذ قادة التكنولوجيا في جميع أنحاء العالم خطوات للحد من التحيز في مجال الذكاء الاصطناعي يتطلب التخفيف من التحيز الخوارزمي وحماية العدالة ضرورة اتباع نهج استباقي متعدد الأوجه، حيث يمكن لمبادئ "العدالة المقصودة" أن تساعد في توجيه طريقة تطوير أنظمة الذكاء الاصطناعي التي من غير المرجح أن تتسبب بالتحيز أو تشجع عليه، كما يتطلب حل مشكلة التحيز في الذكاء الاصطناعي التعاون بين الجهات الفاعلة في صناعة التكنولوجيا، وصناع السياسات، وعلماء الاجتماع، ولا يزال أمام صناعة التكنولوجيا طريق طويل لتقطعه قبل أن تتمكن من القضاء على تحيز الذكاء الاصطناعي (Shashkina Victoria, 2024).

1.5.3. الاطر التنظيمية لمعالجة والحد من التحيز الخوارزمي

إستجابة للمخاوف المتزايدة المحيطة بالتحيز الخوارزمي، تدرك الحكومات والمنظمات بشكل متزايد الحاجة إلى مبادئ توجيهية تحكم استخدام الخوارزميات، واتخذ المنظمون والمشرعون بعض الخطوات لمكافحة قضايا التحيز القائمة على التحليلات، على سبيل المثال، تفرض اللائحة العامة لحماية البيانات (GDPR) للاتحاد الأوروبي أحكاما تتناول الشفافية الخوارزمية والحق في التفسير، وتمكين الأفراد من فهم كيفية تأثير الخوارزميات على حياتهم.

كما ان قانون مستقبل الذكاء الاصطناعي في الولايات المتحدة لعام 2017 يفرض قيودا على معالجة البيانات وممارسات الأعمال التي تعمل بالذكاء الاصطناعي، على أمل تعزيز المساءلة الخوارزمية وتقليل التحيز الخوارزمي (Robert J. Domanski, 2019).تهدف هذه الجهود التنظيمية إلى التخفيف من المخاطر المرتبطة بالتحيز الخوارزمي وتعزيز ممارسات الذكاء الاصطناعي الأخلاقية(Nicol Turner Lee، Detecting racial bias in algorithms and machine learning، 2018). كما تهدفمعايير IEEE 7003لاعتبارات التحيز الخوارزمي، والتي تم تقديمها كجزء من المبادرة العالميةIEEE حول أخلاقيات الأنظمة المستقلة والذكية، إلى

تعزيز المساواة والإنصاف وبالمثل، فإن مدونة الأخلاقيات التي تم تطويرها داخل التحالف السويسري للخدمات كثيفة البيانات تحث الشركات على اتخاذ قرارات أكثر عدلا تعتمد على البيانات(Nick Barney & Ronald Schmelzer, 2023).

2.5.3 حوكمة الذكاء الاصطناعي للحد من التحيزات الخوارزمية

يبدأ تحديد ومعالجة التحيز في الذكاء الاصطناعي بحوكمة الذكاء الاصطناعي، أو القدرة على توجيه وإدارة ومراقبة أنشطة الذكاء الاصطناعي في المؤسسة كما تخلق حوكمة الذكاء الاصطناعي مجموعة من السياسات والممارسات والأطر لتوجيه التطوير والاستخدام المسؤول لتقنيات الذكاء الاصطناعي، عند تنفيذها بشكل جيد .ففي عام 2019، أصدر الاتحاد الأوروبي المبادئ التوجيهية الأخلاقية للذكاء الاصطناعي الجدير بالثقة من قبل فريق من الخبراء رفيع مستوى المعني بالذكاء الاصطناعي، و وفقا للمبادئ التوجيهية ، يجب أن يكون الذكاء الاصطناعي(European commission , 2019):

- قانونيا ويتوافق مع جميع القوانين واللوائح المعمول بها.
- أخلاقيا، بما يضمن الالتزام بالمبادئ والقيم الأخلاقية.
- متينا وقويا، سواء من المنظور التقني أو الاجتماعي.

توفر هذه الإرشادات الثلاثة سبعة متطلبات أساسية يجب أن تفي بها أنظمة الذكاء الاصطناعي حتى تعتبر جديرة بالثقة والتي تحدد سبعة مبادئ للحوكمة وهي: (1) الوكالة البشرية والإشراف، (2) المتانة التقنية والسلامة، (3) الخصوصية وحوكمة البيانات، (4) الشفافية، (5) التنوع وعدم التمييز والإنصاف، (6) الرفاهة البيئية والمجتمعية، و (7) المساءلة (European commission , 2019).

3.5.3. استراتيجيات لمعالجة تحيز خوارزميات الذكاء الاصطناعي

أن تحقيق القضاء التام على التحيز قد يكون بعيد المنال، ولكن من المهم وضع استراتيجية شاملة تتضمن إجراءات فنية وتشغيلية وتنظيمية، يتضمن ذلك أدوات لتحديد مصادر التحيز المحتملة، وتحسين عمليات جمع البيانات، وتعزيز الشفافية داخل المنظمة، من بين الإستراتيجيات لتقليل التحيز في خورزميات الذكاء الاصطناعي:

- **استراتيجيات لمعالجة التحيز في البيانات:** لمعالجة التحيز في البيانات. يمكن النظر في العديد من الاستراتيجيات من قبل مطوري ومستخدمي الخوارزميات في خطوات ومراحل مختلفة، اهمها: معالجة البيانات مسبقا(Sales Philip, 2021) من خلال تحديد البيانات التمثيلية الدقيقة، جمع بيانات من مجموعة واسعة من المصادر وضمان جودتهامع توثيق ومشاركة كيفية اختيار البيانات وتنقيتها(Nick Barney & Ronald Schmelzer, 2023).
- **استراتيجيات لمعالجة التحيز في التصميم الخوارزمي:** يمكن لنماذج الذكاء الاصطناعي أن تتطور وتتكيف، لذا فإن تدقيق وتحديث الخوارزميات ومراقبة ومراجعة النماذج قيد التشغيل وتقييم احتمالية التحيز ضرورية لاكتشاف التحيزات وتصحيحها عند ظهوره، الذى يقوم مبرمجو الكمبيوتر عند اكتشاف التحيز التصميم الخوارزمي، بفحص مجموعة المخرجات التي تنتجها الخوارزمية للتحقق من النتائج الشاذة. ويمكن أن تكون مقارنة النتائج لمجموعات مختلفة خطوة أولى مفيدة، ويمكن القيام بذلك من خلال محاكاة التنبؤات (الحقيقية والخاطئة) واستخدام خوارزميات المدركة للعدالة قبل تطبيقها على سيناريوهات الحياة الواقعية (Nicol Turner Lee, Paul Resnick, & Genie Barton, 2019).
- **استراتيجيات التخفيف التحيز في الاستخدام البشري:** جميع البشر لديهم القدرة على التحيز، سواء من الاختلاف في الخبرة الحياتية أو التحيز التأكيدي أثناء البحث. ويمكن للأشخاص الذين يستخدمون الذكاء الاصطناعي الاعتراف

بتحيزاتهم لضمان عدم تحيز الذكاء الاصطناعي لديهم، مثل الباحثين الذين يتأكدون من أن أحجام عيناتهم ممثلة (Flori Needle, 2023). وهذا ما يؤكد ضرورة الاعتراف بالتحيز البشري، وتحسين العمليات التي يقودها الإنسان واتباع نهج متعدد التخصصات اي إشراك علماء الأخلاق وعلماء الاجتماع والخبراء من المجالات ذات الصلة في مشاريع الذكاء الاصطناعي لضمان اتباع نهج شامل لتخفيف وتحديد التحيزات بشكل أكثر فعالية.

❖ **استراتيجيات اخرى لمعالجة التحيز الخوارزمي:** يمكن للمطورين استخدام مجموعة متنوعة من مقاييس العدالة، مثل التكافؤ الديموغرافي، أو الاحتمالات المتساوية، أو العدالة الفردية، لتقييم أداء نماذج الذكاء الاصطناعي والتعلم الآلي الخاصة بهم. كما يمكن أن يساعد التقييم المستمر ومراقبة أداء النموذج في تحديد التحيزات وقضايا العدالة عند ظهورها، كما يساهم الاستثمار في أخلاقيات الذكاء الاصطناعي وزيادة اعتماد الذكاء الاصطناعي القابل للتفسير (XAI)(Flori Needle, 2023). ووضع مبادئ توجيهية، التقليل من التحيز الخوارزمي

4. طبيعة التحيز الخوارزمي في الخدمات المصرفية

تتبنى العديد من الشركات وبشكل متزايد تحليلات البيانات وتقنيات البيانات الضخمة والذكاء الاصطناعي (AI) لتحويل وتحسين عملياتها الرئيسية واتخاذ القرارات التنظيمية (Mikalef P, Pappas, I. O, Krogstie, J, & Giannakos, M, 2018) لقد تم استخدام قدرات الذكاء الاصطناعي وخوارزميات التعلم الآلي على نطاق واسع في تطبيقات وقطاعات مختلفة بما في ذلك الخدمات المالية.

في الخدمات المالية، تتنوع التطبيقات والفوائد المحتملة لاستخدام الذكاء الاصطناعي. وأصبحت مؤسسات الخدمات المالية تدرك بشكل متزايد الفوائد الكبيرة التي يمكن أن يحققها استخدام الخوارزميات، ومثل نشر الذكاء الاصطناعي في الخدمات المالية حقبة تحولية تتميز بالكفاءة المعززة والابتكار، وتتراوح الاستخدامات من الخوارزميات لتحسين وتخصيص تجارب العملاء، وصولا إلى تطوير نماذج متطورة تسمح بتوفير الائتمان للعملاء الذين لم يكن من الممكن النظر إليهم في السابق، ومع ذلك، فإن هذا التكامل يبرز أيضا اعتبارات أخلاقية معقدة تتحدى القيم الأساسية للخصوصية والإنصاف والشفافية وتظهر مخاوف التحيز في الذكاء الاصطناعي، الناجم عن مجموعات البيانات المنحرفة أو التصميم الخوارزمي (Naila Iqbal Quresh, Saurabh Suman Choudhuri, Yaramala Nagamani, Raj A Varma, & Rutul Shah, 2024).

والبنوك مثل العديد من القطاعات الأخرى، تتم إعادة اختراعها وإعادة تعريفها بواسطة الذكاء الاصطناعي، مثل الموافقة على القروض أو حدود الائتمان وتقدير درجة الائتمان، لكن العيوب التكنولوجية أكثر شيوعا من العيوب البشرية(Phillip Britt, 2021).

1.4 اهمية الخوارزميات في الخدمات المالية

مع تزايد استخدام البيانات في العقد الماضي، يتم تطبيق الخوارزميات بشكل متزايد بطرق تسترشد بها قرارات العمل وأتمتة العمليات، وعليه هناك ثلاث فئات شاملة لتطبيقها(KPMG, 2021):

❖ **الخوارزميات التي تحسن القرارات:** تستخدم الخوارزميات لاتخاذ قرارات مباشرة بشأن العملاء بناءا على بياناتهم، مما يؤثر بدوره على خيارات العملاء وسلوكهم، ويتم تقييم البيانات المقدمة من عميل ويتم دمجها مع البيانات الموجودةثم تقرر الخوارزمية الأساسية الموافقة أو الرفض أو تقديم خيارات مخصصة لطلب ائتمان بناءا على درجة الائتمان الناتجة. تسمح هذه الأساليب للمقرضين، بالتمييز بين المتقدمين ذوي مخاطر التخلف عن السداد العالية

وأولئك الذين يتمتعون بجدارة ائتمانية ولكنهم يفتقرون إلى تاريخ ائتماني.

- **الخوارزميات التي تقلل من الجهد البشري**: يمكن استخدام الخوارزميات لتقليل الجهد البشري المطلوب لمهام محددة تمتاز بالتكرار كبير. وهذا بدوره يمكن أن يؤدي إلى انخفاض التكاليف. على سبيل المثال، تستخدم العديد من مؤسسات الخدمات المالية روبوتات الدردشة الخاصة بالذكاء الاصطناعي للتواصل مع العملاء، تستخدم روبوتات الدردشة هذه معالجة اللغة الطبيعية (" NLP ") ومحاكاة المحادثات البشرية لاستفسارات العملاء.
- **الخوارزميات التي تحل المشكلات المعقدة**: يمكن تطبيق الخوارزميات لحل المشكلات المعقدة التي كانت تعتبر في السابق صعبة للغاية أو معقدة للغاية بالنسبة للبشر. على سبيل المثال، في الخدمات المالية، كان اكتشاف الاحتيال منذ فترة طويلة تحديا،وأظهرت الشركات العاملة في القطاع المالي والتي تستخدم هذه الخوارزميات على نطاق واسع في كل شيء من اكتشاف الاحتيال إلى اتخاذ قرار بتمديد الائتمان. (Nicol Turner Lee, Paul Resnick, & Genie Barton, 2019).

2.4. حالات استخدام الخوارزميات في الخدمات المالية

- **التسعير الآلي**: هذا التحول ساعد المؤسسات على تقليل حجم العمل اليدوي المطلوب من قبل الموظفين، مما أدى إلى تقليل التكاليف. في موازاة ذلك، يستخدم العملاء بشكل متزايد مواقع مقارنة الأسعار لمقارنة مقدمي الخدمات وتحديد الخيارات المناسبة في الوقت الفعلي بناءا على ميزات المنتج والأسعار المقدمة.(KPMG, 2021)
- **الخدمة الذاتية الشخصية للعملاء**:لتلبية هذا الطلب في الخدمات المالية، يستخدم مساعدو الخدمة الذاتية الآليون (مثل روبوتات الدردشة) الخوارزميات لتوليد مشورة مالية مخصصة بناءا على التفاعلات مع العملاء من خلال منصة عبر الإنترنت، في مجالات تشمل دعم العملاء، ووضع الميزانية، وتحديد أهداف الادخار، وتتبع النفقات. تمكن هذه الأدوات مقدمي الخدمات المالية من تقليل تكاليف الدعم التشغيلي مع توفير الخدمات ذات الصلة لقاعدة عملائهم على مدار الساعة طوال أيام الأسبوع.(KPMG, 2021)
- **اكتشاف الاحتيال في الوقت الفعلي**: تحتاج البنوك إلى تحديد المعاملات الاحتيالية المحتملة في الوقت الفعلي لحماية حسابات العملاء. لهذا يمكن لخوارزميات التعلم الآلي تحليل بيانات المعاملات لتحديد الأنماط غير العادية أو الشذوذ الذي قد يشير إلى نشاط احتيالي. ويتم تدريب هذه الخوارزميات على بيانات الاحتيال التاريخية ويمكنها تحديث نماذجها في الوقت الفعلي، وتحديد المعاملات المشبوهة بدقة عالية، مما يسمح للبنوك باتخاذ إجراءات فورية، وبالتالي تعزيز ثقة العملاء وأمنهم(Coursera Staff, 2024)
- **مراقبة وإدارة المخاطر التنبؤية**: تستخدم الخوارزميات أيضا لتطوير نماذج الكشف عن المخاطر داخل مؤسسات الخدمات المالية، وتمكين المقرضين الائتمانيين من التنبؤ بالتخلف المحتمل عن السداد في وقت مبكر، وبالتالي تقدير خطط الدفع المناسبة مع العملاء المتعثرين ماليا لتقليل مخاطر التخلف عن السداد في المستقبل، مما يعود بالنفع على كلا الطرفين(KPMG, 2021).

3.4. أشكال التحيز الخوارزمي في الخدمات المالية

بالنظر إلى لتحيزات خوارزميات الذكاء الاصطناعي في الخدمات المالية، فقد ذكر أن قرارات كانت متحيزة على الرغم من عدم برمجة أي تعصب في النظام(K. Ukanwa & R.T. Rust, 2020) ونتيجة لذلك، أثار استخدام الخورزميات في صنع القرار مخاوف بشأن الخيارات الآلية التي تؤدي إلى نتائج تمييزية(Latanya Sweeney, 2013).وقد يتجلى التمييز في أشكال

مختلفة، مثل التمييز العنصري أو القائم على النوع الاجتماعي، والتحيز الاجتماعي والاقتصادي، وغير ذلك من التفضيلات غير المقصودة، ومن اهم اشكال التحيز الخوارزمي في الخدمات المالية نذكر:

- **التحيز في القرارات الائتمانية :** أدى الاستخدام الواسع النطاق لخوارزميات التعلم الآلي في مجال الائتمان الى تقليل من ذاتية عملية صنع القرار (Ana Cristina Bicharra Garcia, , Marcio Gomes Pinto Garcia, & Roberto Rigobon, 2024)، و باعتبار ان هذه الخوارزميات تعتمد على التاريخ الائتماني المسجل في مجموعات بيانات المؤسسات المالية، فإن العملية غالبا ما تعمل على تعزيز التحيز ضد المجموعات التي يتم تحديدها حسب العرق والجنس والتوجه الجنسي وغيرها من السمات.(Bajracharya, A, Khakurel, U, Harvey, B, & Rawat, D.B, 2023).وباعتبار ان الائتمان أحد أهم المجالات التي قد يكون لتحيز الخوارزميات فيها عواقب وخيمة(Nicole Willing, 2023). مما يجعل من الصعب عليهم الحصول على القروض أو الرهون العقارية (Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Toniann Pitassi, & Rich Zemel, 2012).

تستخدم خوارزميات التعلم الآلي حاليا لمعالجة البيانات وتحديد ما إذا كان الشخص قادرا على سداد القرض. يقوم الشخص الذي يبرمج الخوارزمية بتجميع قاعدة بيانات للأشخاص الذين تقدموا بطلبات قروض في الماضي، بما في ذلك جميع أنواع البيانات (جنس مقدم الطلب وعمره، وما إذا كان قد سدد المبلغ بالكامل أم لا، وما إذا كان قد تأخر في السداد وكم مرة، ومتوسط أجره، ومقدار الضريبة التي دفعها، وفي أي مدينة وحي يعيش) وما إلى ذلك. تطبق الخوارزمية سلسلة من الصيغ الإحصائية على كل هذه البيانات ثم تولد أنماطا تقدر احتمالية سداد عميل محتمل جديد للقرض. وعادة ما تكون الموثوقية هي المعيار الوحيد المستخدم لتطوير هذه الخوارزميات(David Casacuberta, 2017). وقد تتجلى مظاهر التحيز الخوارزمي في قررات الائتمان في ما يلي:

- **التمييز في الوصول:** النتيجة الأكثر أهمية لتحليل طلب القرض هي الموافقة على القرض أو رفضه. ويعني رفض طلب القرض عدم تمكن مقدم الطلب من الوصول إلى خدمة خط الائتمان، ويستند تحليل الائتمان إلى المخاطر المحددة لتخلف مقدمي الطلبات عن السداد والحد الأقصى للمخاطر التي حددتها المؤسسة المالية، ومن الأهمية أيضا تحديد شروط الدفع، مثل سعر الفائدة، والحد الأقصى لعدد الأقساط، ومتطلبات الضمان. وقد يشكل رفض الائتمان، شكلا من أشكال التمييز(Bajracharya, A, Khakurel, U, Harvey, B, & Rawat, D.B, 2023) .

كما يمكن تدريب خوارزمية تحليل المخاطر باستخدام مجموعة بيانات تعكس معدلات رفض أعلى بشكل غير متناسب للمتقدمات دون أسباب وجيهة مثل الاختلافات في درجات الائتمان أو تاريخ الدفع المتأخر عند مقارنته بالتطبيقات الذكورية، وستفترض الخوارزمية بعد ذلك أن النوع الاجتماعي هو عامل ذو صلة في تقييم مخاطر الائتمان، وبالتالي، فإن الخوارزمية ستتكرر والأسو من ذلك، ستعزز التحيز الجنساني الذي ربما أدى سابقا إلى قرارات قروض متحيزة (Andreas Fuster , Paul Goldsmith-Pinkham, Ansgar Walther,, & Tarun Ramadorai , 2022).

- **التمييز في التسعير:** إن استخدام خوارزميات التسعير لم يعد ظاهرة هامشية بل أصبح ممارسة تجارية سائدة في العديد من الأسواق، ودورا حاسما في مشهد السوق التنافسي وسمة التي تميز اقتصاد المستهلك في الوقت الحاضر وتظل محيرة للغاية من منظور قانوني واقتصادي، ولا تنبع المخاوف المرتبطة بذلك من حداثة التسعير الخوارزمي فحسب. بل إنها تشير أيضا إلى الأنواع الجديدة من المخاطر وهي التمييز الخوارزمي (Mateusz Grochowski, Agnieszka Jabłonowska, Francesca Lagioia , & Giovanni Sartor, 2022).

فالتسعير المستند إلى الخوارزميات هو التعديل السريع للأسعار، بطريقة تربط بها الآلة بين خصائص الأفراد واستعدادهم للدفع استنادا إلى البيانات الداخلية والخارجية في الوقت الفعلي، ولم يعد قائما على تقييمات السوق، ولا يستقر بالضرورة عند النقطة التي يلتقي فيها العرض بالطلب (Mateusz Grochowski, Agnieszka Jabłonowska, Francesca Lagioia , & Giovanni Sartor, 2022)، يمكن أن تؤدي البيانات الضخمة والخوارزميات المتطورة إلى تسعير غير دقيقبسبب البيئات الداخلية والخارجية الديناميكية(Oren Bar-Gill, 2019). وممارسة التمييز في الأسعار اي تحديد أسعار أعلى للمستهلكين الراغبين في دفع المزيد وأسعار أقل للمستهلكين الراغبين في دفع أقل.

- **الاستثمار :**تستخدم خوارزميات الذكاء الاصطناعي أيضا لتطوير استراتيجيات الاستثمار، وقد يؤدي التحيز المتأصل إلى تفضيل صناعات أو مناطق أو فئات سكانية معينة عن غيرها في الحصول على التمويل، مما قد يؤدي إلى تعزيز الفوارق الاقتصادية الراسخة. (Nicole Willing, 2023)
- **نظام الكشف عن الاحتيال:** إن أنظمة الكشف عن الاحتيال في الخدمات المالية كانت تحدد بشكل غير عادل معاملات بعض العملاء كاحتيالية، مما أدى إلى تجميد حساباتهم، وغالبا ما كان المتأثرون هم أفراد من مجتمعات معينة، مما زاد من التوترات الاجتماعية.
- **التسويق المستهدف:** في بعض الحالات، استخدمت البنوك خوارزميات لتحليل البيانات وتوجيه العروض التسويقية وجد أن بعض العملاء، وخاصة ذوي الدخل المنخفض أو من خلفيات عرقية معينة، تم استبعادهم من العروض، مما أثر على فرصهم في الوصول إلى المنتجات المالية(Nicole Willing, 2023)
- **خدمة العملاء:** قد تظهر برامج الدردشة الآلية وتطبيقات خدمة العملاء الأخرى المدعومة بالذكاء الاصطناعي تحيزات في التفاعلات. على سبيل المثال، قد تقدم برامج الدردشة الآلية استجابات أو مستويات مختلفة من المساعدة بناء على المعلومات الديموغرافية للمستخدم، مما يؤدي إلى عدم المساواة في المعاملة.(Nicole Willing, 2023)

وهذا ما أكد عليه قانون تكافؤ فرص الائتمان (ECOA) انه من غير القانوني لمقدمي الائتمان التمييز على أساس الجنس أو العرق أو الأصل القومي أو غيرها من الخصائص المحمية ردا على رفض الائتمان على نطاق واسع للنساء والأشخاص الملونين. كما يحظر على المقرضين جمع هذه البيانات في المقام الأول، حتى لغرض الكشف عن التمييز، بسبب المخاوف من استخدامها بشكل غير صحيح لمزيد من إدامة التحيز في الإقراض الاستهلاكي، وتظهر الأبحاث في الولايات المتحدة أن الاتجاهات التمييزية كان من الصعب اكتشافها في الإقراض الاستهلاكي وأن نقص بيانات السمات الحساسة قد خلق أيضا تحديات في التنفيذ.(Alex Kessler & Jacobo Menajovsky, 2021)

ففي عام 2019، تبين أن الخوارزمية التي تدير بطاقة الائتمان الخاصة بشركة Apple لقرارات الحد الائتماني متحيزة ضد النساء بإعطاء حدود ائتمانية أقل بكثير للنساء من أزواجهن حتى لو كانت الزوجة تتمتع بدرجة ائتمان أعلى (Nima Kordzadeh & Maryam Ghasemaghae, 2022).مما تسبب في رد فعل عنيف من جانب العلاقات العامة ضد الشركة (Ben Dickson, 2020).

وفي تقرير لعام 2021 من The Markup، كان تحيز الذكاء الاصطناعي مسؤولا عن رفض 80٪ من المتقدمين للحصول على الرهن العقاري من السود. وعلى نحو مماثل، كان المقرضون أكثر ميلا إلى رفض المتقدمين من أصل لاتيني بنسبة 40% وأكثر احتمالا بنسبة 50% لرفض المتقدمين من سكان جزر آسيا/المحيط الهادئ، وأكثر احتمالا بنسبة 70% لرفض المتقدمين من الأميركيين الأصليين، وكل هذا بالمقارنة مع المتقدمين البيض المماثلين .(Stephen J. Bigelow, Alexander S.

(Gillis, & Mary K. Pratt.تشمل بعض الأمثلة البارزة على النتائج السيئة التي سببها التحيز الخوارزمي تطبيقات الائتمان الآلية من Goldman Sachs التي أثارت تحقيقا في التحيز الجنسي(Nick Barney & Ronald Schmelzer, 2023)

4.4. أليات لتجنب التحيز الخوارزمي في الخدمات المالية

هناك بعض الاليات التي يمكن لقطاع الخدمات المالية اتخاذها لتجنب النتائج السلبية الناجمة عن تحيز الخوارزميات، بناءا على المتطلبات التنظيمية الحالية :(Nicole Willing, 2023)

- ❖ توسيع استخدام مجموعة البيانات المتنوعة والتمثيلية ومراجعة شاملة للبيانات المستخدمة في تحليل المعاملات. وهذا يتطلب البحث عن المجموعات غير الممثلة لضمان تضمين بياناتها بشكل كاف لتحسين التوازن.
- ❖ أفضل الممارسات بشكل عام عند استخدام خوارزميات الذكاء الاصطناعي هي فهم مجموعات البيانات المستخدمة لتدريب النماذج واستخدام أدوات تحليل البيانات لتحديد الممارسات السيئة و تحليل تأثيرها على نتائج الخوارزمية. لأن فهم القيود أو التحيزات المحتملة في مجموعات البيانات سيساعد في محاولة التحكم في أي تحيزات في الناتج والمتغيرات المسببة لها.
- ❖ نظام لمراجعة نتائج الخوارزمية بشكل دوري لضمان عدم وجود تحيزات ويتم هذا من خلال تحديث وتعديل الخوارزمية لتقليل الاعتماد على الأنماط السابقة والتركيز على التحليل السياقي وإجراء اختبارات على النموذج للتأكد من عدالة النتائج
- ❖ في الخدمات المالية، بالإضافة إلى القوانين التي تهدف إلى منع التمييز على أساس العرق والجنس والعمر والدين، هناك قوانين محددة للقطاع المالي، مثل قانون تكافؤ الفرص الائتمانية (ECOA) وقانون الإسكان العادل(FHA)، لمنع التحيز تجاه الأقليات(Ana Cristina Bicharra Garcia, , Marcio Gomes Pinto Garcia, & Roberto Rigobon, 2024). تؤسس هذه القوانين عقيدتين قانونيتين وهما:(Solon Barocas & Andrew D Selbst , 2016)المعاملة غير المتساوية والذي يأخذ القرار في الاعتبار صراحة عضوية المجموعة (بشكل مباشر أو غير مباشر)، والتأثير غير المتناسب حيث تؤدي نتائج القرار إلى إلحاق الضرر
- ❖ إن إعطاء الأولوية للشفافية في خوارزميات الذكاء الاصطناعي من خلال تقديم تقارير واضحة للعملاء حول كيفية اتخاذ القرارات ومعلومات حول العوامل التي أثرت على نتائجهم، من شأنه أن يساعد المؤسسات المالية على فهم كيفية اتخاذ القرارات المتعلقة بالامتثال للقواعد التنظيمية وبناء الثقة مع المستهلكين وإدخال خوارزميات تعتمد على الذكاء الاصطناعي التفسيري، مما يساعد في فهم كيفية تأثير المتغيرات المختلفة على النتائج، فتوثيق عمليات اتخاذ القرار في الخوارزمية بشكل واضح يسمح بالتدقيق وتحديد التحيزات المحتملة.

5. خاتمة

خورزميات الذكاء الاصطناعي مثل البشر، بعيدة كل البعد عن الكمال، والشيء المذهل فيهاهو مدى تحيزها للإنسان. ولو كانت لها شخصية وآراء خاصة بها، لربما وقفت في وجه من يغذيها بأمثلة تقطر تعصبا، لهذا غالبا ما ينظر إلى التحيزات الخوارزمية انها مشكلة إنسانية فهي نتاج أدمغة غير كاملة، وليس أنظمة ذكاء اصطناعي محايدة كما يفترض. لكنها لا تعكس التحيزات البشرية فحسب، بل يمكنها تضخيمها على نطاق واسع، بطرق يصعب اكتشافها ومنعها. إلا أن فهم جوهر التحيز الخوارزمي أمر بسيط. بغض النظر عن النية، يجعل النتيجة في النهاية (مهما كان تعريفها) موضع شك وغير موثوقة أو ضارة أو تمييزية لمجموعات معينة من الناس. ومن المهم أيضا أن ندرك انها ليست مشكلة عرضية فهي مشكلة أساسية لن "تصلح نفسها" بأي حال من الأحوال بمرور الوقت.

فالعديد من الحالات التي يتم تصنيفها على أنها "تحيز خوارزمي تنشأ في كل مرحلة من مراحل عملية التطوير والتنفيذ والتطبيق، فإن أحد المتطلبات الأساسية هو فهم العلاقة بين النظام والمعايير الأخلاقية والقانونية المعمول بها في السياقات ذات الصلة.

والمؤسسات المالية مثل العديد من القطاعات الأخرى، تتم إعادة اختراعها وإعادة تعريفها بواسطة الذكاء الاصطناعي، ومثل نشر الذكاء الاصطناعي في الخدمات المالية حقبة تحولية تتميز بالكفاءة المعززة والابتكار وخدمة العملاء وتعزيز الإنتاجية التنظيمية وتطوير منتجات جديدة، لكن العيوب التكنولوجية أكثر شيوعا من العيوب البشرية حيث تحمل القرارات التي تقودها خوارزميات الذكاء الاصطناعي اعتبارات أخلاقية ابرزها التحيز الخوارزمي، مما يؤثر على قدرة الافراد للوصول الى الائتمان وفرص الاستثمار والتي تؤدي إلى إدامة التفاوتات المجتمعية،وتحجب الأساس المنطقي وراء الاستشارات أو القرارات المالية المهمة، مما يقوض الثقة والمساءلة.

❖ **نتائج الدراسة:** من بين النتائج التي توصلت اليها الدراسة ما يلي:

- إن التحيز الخوارزمي هو في الأساس مسألة اجتماعية، وبالتالي، يجب أن تكون الآثار الأخلاقية والعواقب الاجتماعية لهذه الظاهرة في مركز فحصها ومعالجتها، فعواقب وجودها يمكن أن تكون كبيرة وتؤدي إلى تداعيات خطيرة، والى نتائج غير عادلة وإدامة أوجه عدم المساواة الاجتماعية. مما يؤدي إلى ضرر محتمل للأفراد والمجتمعات بشكل غير متناسب.
- إن الأسباب الجذرية للتحيز الخوارزمي في الخدمات المالية متعددة الأوجه ويمكن أن تنبع من مصادر مختلفة. فالعوامل المؤسسية والمجتمعية البشرية والنظامية تشكل مصادر مهمة للتحيز ويتم تجاهلها حاليا، بالإضافة الى الطبيعة السرية لهذه التحيزات، وهذا ما يجعل من الصعب تحديد وتصحيح حالات التحيز.
- إن الوعي التام بمخاطر التحيز الخوارزمي والعمل على الحد منه يعد أولوية ملحة ومسؤولية الجميع حتى وان كانت غير مقصودة ودون نية أو وعي واضح بذلك، فمواجهة هذا التحدي يتطلب من مقدمي الخدمات المالية برامج لتوسيع نطاق تنوع مجموعات البيانات الخاصة بها وتنوع موظفيها لضمان وجهات النظر المتعددة والخبرات المتنوعة لتغذية الأنظمة بنقاط البيانات التي يمكن التعلم منها. وبناء مشهد أكثر شمولا.
- من الشائع بين أنصار الحلول التكنولوجية القول بأن "الخوارزميات ليست جيدة أو سيئة، أو عادلة أو غير عادلة فالعادل أو غير العادل هو الشخص الذي يطبقها، كما أن تحقيق القضاء التام على التحيز قد يكون بعيد المنال إلا أن هناك استراتيجيات يمكننا اتخاذها لتقليل التحيز الخوارزمي في الخدمات المالية

- إن اعتماد الذكاء الاصطناعي يعيد تشكيل مستقبل الخدمات المالية بطرق من شأنها أن تزيد من الكفاءة وتزيد من خطر التحيز النظامي في القرارات المتعلقة بالائتمان والاستثمار والتي تؤدي إلى إدامة التفاوتات المجتمعية

❖ **التوصيات:** من خلال ماسبق ذكره نوصي بما يلي:

- من المهم أن تدرك المؤسسات المالية مفاهيم التحيز والعدالة وتحديد أولوياتها لضمان إمكانية تقاسم فوائد هذه التقنيات بشكل عادل وعلى نطاق واسع .وعلى دراية بإمكانية التحيز في المخرجات الخوارزمية وفهم وظائف الخوارزمية وتحديد المصادر المحتملة للتحيز في الذكاء الاصطناعي وقيودها في سياقات مختلفة.
- على المؤسسات المالية، التأكد من أن خوارزميات الذكاء الاصطناعي عادلة ومنصفة وتخدم احتياجات الجميع، ويتحقق هذا من خلال الاهتمام بكيفية اختيار الميزات والأبعاد المراد تضمينها في مجموعة بيانات التدريب،
- ضرورة تعزيز وتضمين الإشراف البشري لتطوير وإدارة خوارزميات الذكاء الاصطناعي وتحسين العمليات التي يقودها لتعرف على التحيزات المحتملة وتعامل معها أثناء التطوير وتقديم وجهات نظر مختلفة حول كيفية تسلل التحيز والتخفيف منها.
- ضرورة أن تكون المؤسسات المالية شديدة اليقظة بشأن التحيز الخوارزمي، لذا يتعين عليها أن تتخذ خطوات استباقية للحد ومعالجة هذه التحيزات في المراحل التطويرية، والسبب الأكثر وضوحا هو أن مجرد فكرة أن خوارزمية الذكاء الاصطناعي متحيزة يمكن أن تقضي على قيمة الكفاءة والإنتاجية التي يقدمها الذكاء الاصطناعي.
- يجب على الهيئات التنظيمية الإصرار على المبادئ الأخلاقية ومعايير التصميم ومدونات قواعد سلوكشركات وبروتوكولات التوثيق، وغيرها من التدابير الوقائية التي تعطي الأولوية للعدالة والمساءلة والشفافية.
- إن التبني الواسع النطاق لخوارزميات الذكاء الاصطناعي في المؤسسات المالية يستدعي تنويع الذكاء الاصطناعي والاستثمار في أخلاقيات الذكاء الاصطناعي وتحديد تنفيذ نهج مسؤول للذكاء الاصطناعي وتجنب التمييز الخوارزمي غير المبرر وغير المقصود وتعزيز العدالة وحماية الحقوق المدنية.

6. قائمة المراجع:

1. Aaron Patzer. (2023, Aug 30). Not All Algorithms Are AI (Part 1): How The Revolution Started. Consulté le 08 13, 2024, sur forbes: https://www.forbes.com/councils/forbestechcouncil/2023/08/30/not-all-algorithms-are-ai-part-1-how-the-revolution-started/
2. Alex Kessler, & Jacobo Menajovsky. (2021, MAY 25). Algorithmic Decisions Cannot Rely on "Blind" Approaches.Consulté le 08 25, 2024, sur centerforfinancialinclusion.org: https://www.centerforfinancialinclusion.org/reducing-bias-in-algorithmic-decisions-cannot-rely-on-blind-approaches/
3. Ana Cristina Bicharra Garcia, , Marcio Gomes Pinto Garcia, & Roberto Rigobon. (2024). Algorithmic discrimination in the credit domain: what do we know about it?. (2024) 39:. AI & SOCIETY, 39, pp. 2059–2098 .
4. Ananny, M, & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society, 20(03), pp. 973–989.
5. Andreas Fuster , Paul Goldsmith-Pinkham, Ansgar Walther,, & Tarun Ramadorai . (2022). "Predictably Unequal? The Effects of Machine Learning on Credit Markets,. Journal of Finance, American Finance Association, , 77(01), pp. 5-47.
6. Arpan Kumar Kar, Shweta Kumari Choudhary, & Vinay Ku. (2022). How can artificial intelligence impact sustainability: A systematic literature review. Journal of Cleaner Production, 376.
7. Bajracharya, A, Khakurel, U, Harvey, B, & Rawat, D.B. (2023). Recent Advances in Algorithmic Biases and Fairness in Financial Services: A Survey. the Future Technologies Conference (FTC) 2022, 01, pp. 1-15. Canada.European commission . (2019, April 08).
8. Ben Dickson. (2020, December 07). How banks use AI to catch criminals and detect bias. Consulté le 09 03, 2024, sur thenextweb: https://thenextweb.com/news/how-banks-use-ai-to-catch-criminals-and-detect-bias

9. Binns R. (2018). What can political philosophy teach us about algorithmic fairness? IEEE Security & Privacy, 16(03), pp. 73-80.
10. Coursera Staff. (2024, mai 24). What Are AI Algorithms?Consulté le 09 03, 2024, sur coursera.org: https://www.coursera.org/articles/ai-algorithms
11. Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Toniann Pitassi, & Rich Zemel. (2012). Fairness through awareness. Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, (pp. 214–226). Cambridge, MA, USA.
12. David Casacuberta. (2017, MARCH 14). Algorithmic Injustice. Consulté le 08 24, 2024, sur cccb.org: https://lab.cccb.org/en/algorithmic-injustice/
13. Ethics Guidelines for Trustworthy AI. Consulté le 08 18, 2024, sur European Union, Digital Single Market: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
14. Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. (McGraw-Hill., Éd.) The Handbook of Social Psychology, 2(4), pp. 357-411.
15. Flori Needle. (2023, May 19). What is AI bias? [+ Data]. Consulté le 08 25, 2024, sur .hubspot: https://blog.hubspot.com/marketing/ai-bias
16. G. Batra, A. Queirolo, & N. Santhanam. (2018, 01 08). Artificial intelligence: The time to act is now. Consulté le 06 19, 2024, sur McKinsey: https://www.mckinsey.com/industries/advanced-electronics/our-insights/artificial-intelligence-the-time-to-act-is-now
17. Harshit Baluja. (2023, October 25). The Complete Guide to AI Algorithms. Consulté le 05 13, 2024
18. Hou Tsung-Yu, Tseng Yu-Chia, Yuan Chien Wen (Tina), & . (2024, Jun). International Journal of Information Management, 76(C).
19. Jake Silberg, & James Manyika. (2019, June 06). Notes from the AI frontier: Tackling bias in AI (and in humans). Consulté le 06 20, 2024, sur McKinsey Global Institute: https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans
20. Joy Buolamwini, & Timnit Gebru. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Dans N. Proceedings of Machine Learning Research (Éd.), 1st Conference on Fairness, Accountability and Transparency, (pp. 77-91). New York, NY, USA.
21. K. Ukanwa, & R.T. Rust. (2020, March 21). Algorithmic Bias in Service. Consulté le 08 17, 2024, sur Marketing Science Institute Working Paper Series Repor: https://marketing.wharton.upenn.edu/wp-content/uploads/2020/11/11.19.2020-Ukanwa-Kalinda-PAPER-AlgorithmicDiscriminationinService2020.pdf
22. Kate Crawford, & Ryan Calo. (2016). There is a blind spot in AI research. Nature , 538, pp. 311–313.
23. Kaya Ismail . (2018, October 26). AI vs. Algorithms: What's the Difference?Consulté le 09 02, 2024, sur https://www.cmswire.com/information-management/ai-vs-algorithms-whats-the-difference/
24. KPMG. (2021, A report prepared for Finastra International . March 2021.). Algorithmic bias and financial services.Consulté le 09 10, 2024, sur Finastra Internationa: https://www.finastra.com/sites/default/files/documents/2021/03/market-insight_algorithmic-bias-financial-service
25. Latanya Sweeney. (2013, may). Discrimination in online ad delivery. communications of the acm , 56(05), pp. 44–54.
26. Lev Craig. (2023, Ap 17). Breaking the cycle of algorithmic bias in AI systems.Consulté le 09 12, 2024, sur 17r 2023.: https://www.techtarget.com/sustainability/feature/Breaking-the-cycle-of-algorithmic-bias-in-AI-systems
27. London, A. J., & Danks, D. (2017). Algorithmic Bias in Autonomous Systems. Dans ijcai.org (Éd.), Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)., (pp. 4691-4697). Australia.
28. Maddalena Favaretto, Eva De Clercq , & Bernice Simoe Elger. (2018). What can political philosophy teach us about algorithmic fairness? Journal of Big Data, 16(03), pp. 73–80.
29. Malik, N, Tripathi, S.N., Kar, A.K, & Gupta, S. (2022). Impact of artificial intelligence on employees working in industry 4.0 led organizations. International Journal of Manpowe, 43(02), pp. 334-354.
30. Mateusz Grochowski, Agnieszka Jabłonowska, Francesca Lagioia , & Giovanni Sartor. (2022). Algorithmic Price Discrimination and Consumer Protection; A Digital Arms Race? Technology and Regulation, pp. 36-47 .
31. Mc Gettigan, T. (2017). . Artificial Intelligence: Is Watson the Real Thing ?. , 13(2),. The IUP Journal of Information Technology, 13(02), pp. 44-69.
32. Mikalef P, Pappas, I. O, Krogstie, J, & Giannakos, M. (2018). Big data analytics capabilities: A systematic literature review and research agenda. .(. Information Systems and e-Business Management, 16(03), pp. 547–578.

33. Ming, Hui Huang, & Roland , T. Rust;. (2021). strategic framework for artificial intelligence in marketing .,. Journal of the Academy of Marketing Science, 49(01), pp. 30-50.
34. Möller, J, Trilling, D, Helberger, N., & van Es, B. (2018). Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. Information, Communication & Society, 21(07), pp. 959-977.
35. Naila Iabal Quresh, Saurabh Suman Choudhuri, Yaramala Nagamani, Raj A Varma, & Rutul Shah. (2024). Ethical Considerations of AI in Financial Services: Privacy, Bias, and Algorithmic Transparency. International Conference on Knowledge Engineering and Communication Systems (ICKECS),, (pp. 1-6). Chikkaballapur, India,.
36. Nick Barney, & Ronald Schmelzer. (2023, Jul 28). 6 ways to reduce different types of bias in machine learning. Consulté le 09 13, 2024, sur techtarget: https://www.techtarget.com/searchenterpriseai/feature/6-ways-to-reduce-different-types-of-bias-in-machine-learning
37. Nicol Turner Lee. (2018). Detecting racial bias in algorithms and machine learning. Journal of Information, Communication & Ethics in Society, 16(03), pp. 252-260.
38. Nicol Turner Lee, Paul Resnick, & Genie Barton. (2019, May 22). Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms.Consulté le 08 05, 24, sur brookings: https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-polic
39. Nicole Willing. (2023, November 23). The Challenges of AI Algorithm Bias in Financial Services. Consulté le 08 17, 2024, sur techopedia: https://www.techopedia.com/ai-algorithm-bias-in-financial-services
40. Nima Kordzadeh, & Maryam Ghasemaghae. (2022). Algorithmic bias: review, synthesis, and future research directions. European Journal of Information Systems, 31(03), pp. 388–409.
41. Oren Bar-Gill. (2019). Algorithmic Price Discrimination When Demand Is a Function of Both Preferences and (Mis) perceptions. The University of Chicago Law Review, 86(02), pp. 1-30.
42. Phillip Britt. (2021, june 01). Tips for battling bias in ai-based personalization. Consulté le 09 09, 2024, sur destinationcrm: https://www.destinationcrm.com/Articles/Editorial/Magazine-Features/Tips-for-Battling-Bias-in-AI-Based-Personalization-147143.aspx
43. Robert J. Domanski. (2019). The AI pandorica: linking ethically-challenged technical outputs to prospective policy approaches. Proceedings of the 20th annual international conference on digital government research, (pp. 409 - 416). Dubai, UAE.
44. Sales Philip. (2021). Algorithms, Artificial Intelligence, and the Law. (J. I. Law., Éd.) 105(01).
45. Shashkina Victoria. (2024, June 04). .What is AI bias really, and how can you combat it?Consulté le 09 04, 2024, sur itrexgroup: https://itrexgroup.com/blog/ai-bias-definition-types-examples-debiasing-strategies/
46. Shin, D, Zaid, B, Biocca, F, & Rasul, A. (2022). In platforms we trust? Unlocking the blackbox of news algorithms through interpretable AI. Journal of Broadcasting and Electronic Media, 66(02), pp. 235–256.
47. Solon Barocas , & Andrew D Selbst . (2016, June). Big Data's Disparate Impact. California Law Review, 104(03), pp. . 671-732.
48. Stephen J. Bigelow, Alexander S. Gillis, & Mary K. Pratt. (s.d.). Machine learning bias (AI bias). Consulté le 08 14, 2024, sur techtarget: https://www.techtarget.com/searchenterpriseai/definition/machine-learning-bias-algorithm-bias-or-AI-bias
49. Tabsharani, F. (2023, May 05). Types of AI algorithms and how they work. . 05 2023. Consulté le 08 03, 2024, sur https://www.techtarget.com/searchenterpriseai/tip/Types-of-AI-algorithms-and-how-they-work
50. Tiago Palma Pagano , Rafael Bessa Loureiro, & and all. (2023). Bias and Unfairness in Machine Learning Models: A Systematic Review Review on Datasets, Tools, Fairness Metrics, and Identification. Big Data Cogan. Comput, 07(15), pp. 1-31.
51. Varsha P. S. (2023). How can we manage biases in artificial intelligence systems – A systematic literature review. International Journal of Information Management Data Insights, 3(1), pp. 1-9.
52. What Is an Algorithm? (2023, Sep 28). Consulté le 07 13, 2024, sur datacamp: https://www.datacamp.com/blog/what-is-an-algorithm
53. W.S. Sarle. (1994). Artificial neural networks and statistical models. Proceedings of the Nineteenth Annual SAS Users Group International Conference, (pp. 1538-1550).
54. زيدان رغداء. (14 04، 2010). مفهوم التحيز عند الدكتور عبد الوهاب المسيري. تاريخ الاسترداد 15 07، 2024، من الملتقى الفكري للإبداع: idSC=16&idC=4&https://almultaka.org/site.php?id=835
55. صديقي، علي. (2011). مفهوم «التحيز» عند عبد الوهاب المسيري. مجلة الكلمة(73).
56. عبد الوهاب المسيري . (24 مايو، 2024). فقه التحيز: من التحيز للنموذج الغربي إلى التحيز لنموذج إسلامي بديل. تاريخ الاسترداد 13 09، 2024، من khotwacenter: https://www.khotwacenter.com