أَخْلَاقُ الخَوَارِزْمِيَّة
مِنْ تَحيُّزَاتٍ ثَقَافِيَّةٍ إِلَى هَنْدَسَةِ القَبُوْلِ والاسْتِسْلَامِ

■ أ.م.د. فاطمة رمضان عبد الرحمن عبد اللطيف[1]

ملخّص

تتناول الباحثة في هذه الدراسة الإشكاليَّة الجوهريَّة المتمثِّلة في كيفيَّة مساهمة التحيّزات الخوارزميَّة، المشتقَّة من ممارسات اجتماعيَّة وثقافيَّة، في تقبُّل البشر لها، واستسلامهم لتأثيرها في حياتهم اليوميَّة. وتتجلَّى الأسئلة الرئيسة لهذه الدراسة في: كيف تتحوَّل الخوارزميَّة من أداة عقلانيَّة إلى فاعل خُلُقي متحيّز؟ وما أشكال التحيّز ومصادره؟ وما استراتيجيَّات معالجته؟ وللإجابة عن هذه الأسئلة وغيرها، يتوجَّب على الباحثة تحليل العلاقة بين التحيّزات الإنسانيَّة والخوارزميَّة وفهم الآليَّات التي تجعل البشر يقبلون بها. وقد اعتمدت الباحثة على المنهج التحليلي لدراسة نماذج متنوّعة من الخوارزميَّات، وتحليل دمج التحيّزات البشريَّة فيها وربطها بسلوكيَّات الأفراد. وقد توصَّلت الباحثة إلى مجموعة مُهمَّة من النتائج، لعلَّ من أبرزها أنَّ الخوارزميَّات ليست محايدة، بل تحمل ضمنيًّا قيمًا وتصوّرات بشريَّة تعيد إنتاج أشكال التمييز الاجتماعي، وتمنحها شرعيَّة زائفة على أساس الموضوعيَّة، الأمر الذي يترتَّب عليه صعوبة الاعتراض عليها. كما بينت الباحثة أنَّ التحيّز الخوارزمي يؤثّر على جلّ المجالات الحيويَّة، مثل التوظيف، والعدالة، والرعاية الصحيَّة، كما يعيد تشكيل علاقات القوّة الاجتماعيَّة.

الكلمات المفتاحية: الخوارزميَّة، التحيّز الخوارزمي، أخطاء التحيّز الخوارزمي، أشكال التحيّز الخوارزمي، مصادر التحيّز الخوارزمي

---

١ - أستاذة المنطق وفلسفة العلم المساعد، قسم الفلسفة، كليَّة الآداب، جامعة المنيا.

Algorithmic Ethics:
From Cultural Biases to Engineering of Acceptance, Submission

■ Assist Prof. Fatima Ramadan Abdel Rahman Abdel Latif[1]

Abstract

This study addresses the fundamental problem of how algorithmic biases, derived from social and cultural practices, contribute to human acceptance and submission to their influence in daily life. The study's central questions revolve around: How does an algorithm transform from a rational technical tool into a biased moral agent? What forms do these biases take, and what are their sources? How can they be addressed? To answer these questions, the researcher analyzes the nature of the relationship between human biases and algorithms, and the mechanisms that lead individuals to accept them. This analysis employs a diverse analytical approach, examining various algorithm models and tracing the integration of human biases within them, linking these biases to individual behaviors.

It concludes with main findings, most notably that algorithms are not neutral. Rather, they implicitly reflect human values and perceptions that reproduce social discrimination and grant it a false legitimacy under the guise of objectivity, thus making it difficult to challenge. The researcher also explained that algorithmic bias extends its influence to vital areas such as employment, justice, and healthcare, thus contributing to the reshaping of social power balances.

Keywords:

Algorithm, Algorithmic Bias, Algorithmic Bias Errors, Algorithmic Bias Forms, Algorithmic Bias Sources.

1 -Assistant Professor of Logic and Philosophy of Science, Department of Philosophy, Faculty of Arts, Minia University.

مقدّمة

مـمَّا لا شـكَّ فيه أنَّ الذكـاء الاصطناعي أصبح جزءًا لا يتجـزَّأ من حياتنـا اليوميَّة. ومع هذا الانتشـار، ظهرت إشكاليَّة رئيسـة تلامس جوهر العلاقة بين الإنسان والآلة، وهي: كيف يمكن لتقنيـات يُفترض بها الحياد أن تعكس وتكـرس تحيّزات قديمة كانت موجودة في بنية المجتمع البشري؟ وبالطبع، لم تكن هذه الإشكاليَّة مجرّد إشكاليَّة تقنيَّة أو برمجيَّة، بل هي إشكاليَّة خُلُقيَّة واجتماعيَّة عويصة تتعلَّق بكيفيَّة تشـكيل الخوارزميَّات لتصوّراتنا وقيمنا، وكيفيَّة قبول البشر لها واستسلامهم لتأثيرها دون وعي كامل. ومن هنا، تنطلق الباحثة في هذه الدراسة لاستكشاف دور التحيّزات الخوارزميَّة في هندسـة القَبول والاستسـلام البشري، لا سيّما في المجالات المتعلّقة بالجنس والعرق والهويَّة الثقافيَّة والأقلّيات.

بالتالي، تراءى للباحثة أنَّ الإشـكاليَّة الجوهريَّة لهذه الدراسـة تكمن في السـؤال الجوهري الآتي: كيف سـاهمت التحيّزات الخوارزميَّة التي نشـأت في البداية من ممارسات بشريَّة ثقافيَّة واجتماعيَّة، في تقبّل البشر لها والاستسلام لتأثيرها في حياتهم اليوميَّة؟ وفي الواقع، ترى الباحثة أنَّ الإجابة عن هذا السـؤال تتطلَّب النظر في الخوارزميَّات ليـس أدوات تقنيَّة محايدة، بل مرآة

تعكس تحيّزات تاريخيَّة وثقافيَّة، وتعيد إنتاجها بطرق قد تكون أكثر تعقيدًا وإقناعًا؛ ذلك لأنَّ الخوارزميَّة قد تُحوّل التحيّزات المضمرة في المجتمع إلى ممارسات تبدو موضوعيَّة ومحايدة، ممَّا يجعل الفرد يمرّرها أو يوافق عليها دون نقد، باعتبارها نتيجة علميَّة بحتة.

ويتمثَّل الهدف من هذه الدراسة في تحليل العلاقة بين التحيّزات الثقافيَّة والخوارزميَّة، وفهم الآليَّات التي تجعل البشر يقبلون بها ويستسلمون لتأثيرها. كما تسعى الباحثة من خلال هذه الدراسة إلى توضيح كيف يمكن للذكاء الاصطناعي أن يصبح ناقلاً ومضاعفًا للتمييز الاجتماعي بدلاً من أن يكون أداة تحرّر أو عدالة، وكيف أنَّ الوعي بهذه الأنماط يمكن أن يسهم في تطوير خوارزميَّات أكثر عدالة وشفافيَّة.

وفي ما يتعلّق بالمنهج المتّبع في هذه الدراسة، تعتمد الباحثة المنهج التحليلي لدراسة نماذج مختلفة للخوارزميَّات المستخدمة في الحياة اليوميَّة، وتحليل كيفيَّة إدماج التحيّزات الإنسانيَّة فيها، ثمّ ربط ذلك بسلوكيَّات البشر في قبولها. كما تهتمّ الباحثة، اعتمادًا على هذا المنهج، بذكر بعض الأمثلة الملموسة التي توضّح كيف تتحوَّل التحيّزات الخوارزميَّة إلى ممارسات يوميَّة مقبولة.

وتتجلّى أهمّيَّة هذا الموضوع في توضيح حقيقة مُهمَّة للقارئ، مفادها أنه مع تزايد الاعتماد على الذكاء الاصطناعي في مختلف المجالات، تزداد الحاجة لفهم تأثيره على القيم الاجتماعيَّة والعدالة والمساواة؛ إذ ترى الباحثة أنَّه إذا لم يجرِ التعامل مع التحيّزات الخوارزميَّة بوعي كامل، فقد يؤدّي ذلك إلى تكريس الفوارق الاجتماعيَّة وتعميق التمييز ضدّ النساء والأقلّيات والفئات المُهمَّشة.

أما عن سبب اختيار الباحثة لهذه الدراسة، فيكمن في اقتناعها الراسخ بأنَّ التكنولوجيا ليست محايدة، وأنَّ دور الفلسفة في تفسيرها وفهمها أصبح أكثر ضرورة من أي وقت مضى. ومهما يكن من أمر، ترى الباحثة أنَّ الإشكاليَّة المطروحة تتطلَّب الإجابة عن مجموعة من الأسئلة الفرعيَّة التي تشكّل بمجموعها المحاور الرئيسة لهذه الدراسة، وهي: كيف تتحوّل الخوارزميَّة من أداة عقلانيَّة إلى فاعل خُلُقي متحيّز، وكيف يظهر التحيّز الخوارزمي وأشكاله ومصادره؟ وما أبعاده الخُلُقيَّة؟ وأخيرًا، ما هي استراتيجيَّات معالجة التحيّز الخوارزمي؟

أوَّلًا: من الخوارزميَّة إلى التحيّز الخوارزمي.

بدايــة، ترى الباحثة أنَّه من الأهميَّة بمكان الشروع في تعريف الخوارزميَّة؛ ذلك لأنَّها تلاحظ وجــود عدّة تعريفات لها. ولعلَّ من أبرز هذه التعريفات أنَّها هيكل أو كيان يحكمه نظام مركَّب، ومحــدود، ومجرّد، وفعّال، يُقدَّم بصياغة إلزاميَّة، ويحقّــق غرضًا محدّدًا ضمن شروط معيّنة[١]، أو بالأحــرى، يمكن تعريفها على أنَّها شــفرة حاســوبيَّة تنفّــذ مجموعة مــن التعليمات. وينظر المتخصّصون في علوم الحاســوب إليها على أنَّها عمليَّة رئيســة لمعالجــة البيانات، بينما يراها آخرون مــن الناحية النظريَّة على أنَّها إجراءات مشــفّرة تعمل على تحويل البيانات اســتنادًا إلى حسابات محدّدة. وتتكوّن الخوارزميَّة من سلسلة من الخطوات المتتابعة التي تُتّخذ لحلّ مشكلة معيَّنة، فهي تأخذ مدخلات (المكوّنات)، وتقسّم المهمَّة إلى أجزاء مكوّنة، ثمّ تنفذ هذه الأجزاء خطوة بخطــوة، لتنتج مخرجًا محدّدًا[٢]. كــما يمكن تعريف الخوارزميَّة على أنَّها تسلســل من الأوامــر أو التعليمات التي ينفذهــا الكمبيوتر لمعالجة البيانــات، وتحويلها من مدخلات إلى مخرجات محددة. فعلى سبيل المثال، تقوم الجهات المطوِّرة أو المستخدمون بتزويد الكمبيوتر بقائمة عشــوائيَّة من الأشــخاص (المدخلات)، ثمّ ينفّذ الكمبيوتر الخوارزميَّة المبرمجة مسبقًا لترتيب هذه القائمة حســب العمر. ومن هذا المثال، تشكّل سلسلة الأوامر الموجهة للكمبيوتر خوارزميَّة، ليُنتج في النهاية الكمبيوتر قائمة مرتَّبة حسب العمر (المخرجات).[٣]

وبالتــالي، لم تكــن الخوارزميَّة مجرّد تعليمات عشــوائيَّة بل هي مجموعــة دقيقة وواضحة، لا يمكن الشــكّ فيها من حيث التتابــع والتنفيذ. وبمعنى أكثر دقّة، تتّصــف الخوارزميَّة بالدقّة، والتسلســل المنطقي، وعدم الغموض إلى حدّ مــا، ما يجعلها كيانًا مفهوميًّــا خاصًّا ومميّزًا في علوم الحاسوب.[٤]

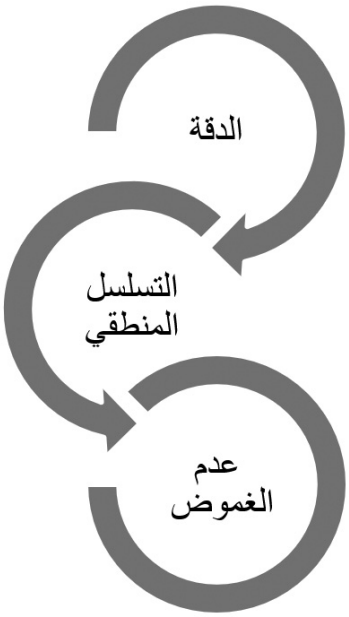
---

1 - Robin K. Hill: What an Algorithm Is, p.58

2 - Kilian Vieth and Joanna Bronowicka: Ethics of Algorithms – Why Should We Care?

3 - Michael O'Flaherty: Bias in Algorithms – Artificial Intelligence and Discrimination, European Union Agency for Fundamental Rights, p18

4 - Robin K. Hill: What an Algorithm Is, p. 44



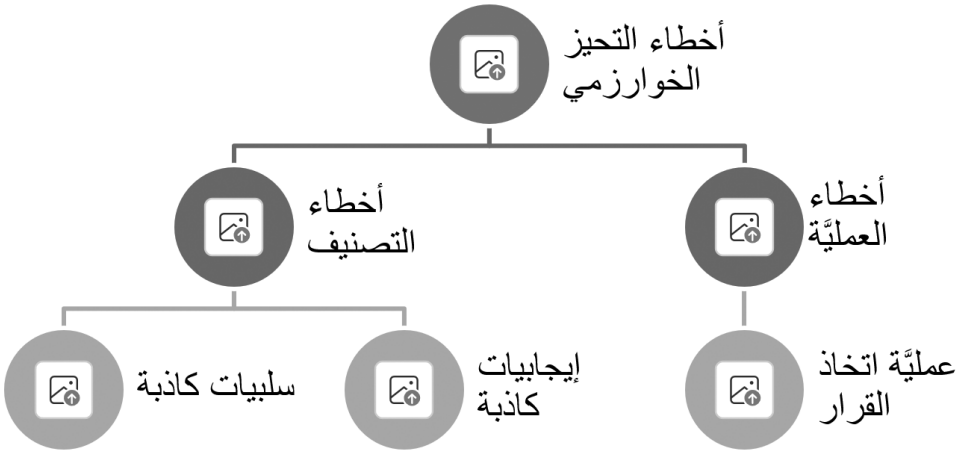
شكل رقم (١) يوضّح أبرز خصائص الخوارزميَّة

وتختلف الخوارزميَّة عن المَهمَّة (The Task) في أنَّها قد تنفّذ أكثر من مَهمَّة واحدة في الوقت نفسه. فعلى سبيل المثال، عند فرز حقل يحتوي على قيم مشتركة بين عدّة مدخلات، قد يؤدّي هذا الفرز أيضًا إلى تجميع هذه المدخلات. ومع ذلك، قد يكون هدف التجميع مختلفًا تمامًا عن هدف الفرز نفسه من منظور المستخدم، ما يجعل التجميع مَهمَّة مستقلّة عن عمليَّة الفرز. وهذا يجعل الباحثة تستنتج أنَّ الخوارزميَّة هي وسيلة لتنفيذ المَهام وليست المَهمَّة نفسها، وأنَّ الخوارزميَّة نفسها قد تحقّق أهدافًا متعددة اعتمادًا على سياق البيانات وطبيعتها، كما أنَّ خصائص أي مَهمَّة معالجة أو حسابيَّة بحدّ ذاتها، مثل الفرز، يمكن تقسيمها إلى أنواع فرعيَّة، مثل الفرز القائم على المقارنات، ما يعكس مرونة الخوارزميَّة في التطبيق، واستخدامها لأغراض متعددة حسب الحاجة[1].

1 - Robin K. Hill: What an Algorithm Is, p.43

والجدير بالذكر أنَّ الخوارزميَّات تحظى بأهميَّة كبيرة؛ إذ تُعتبر أداة مُهمَّة تساعد المتخصّصين عـلى تحقيق نتائج أفضل لمؤسّسـاتهم وتحسـين خدمة العملاء. فهي تضيـف قيمة من خلال تحليل كمّيات ضخمة ومعقّدة من البيانات واكتشـاف العلاقات غـير الواضحة بينها، كما أنَّها أكثر اتسـاقًا في قراراتها مقارنة بأسـاليب اتّخاذ القرار التقليديَّة. ومع ذلك، تلاحظ الباحثة تزايد القلق بشـأن إمكانيَّة أن تكون القـرارات الخوارزميَّة غير عادلة، ومتحيّزة، أو حتّى غير خُلُقي[1]؛ لأنَّ الخوارزميَّات يجري تدريبها على بيانات تُظهر التحيّزات البشريَّة .[2]

وهذا إن دلَّ على شيء، فإنَّما يدلّ على أنَّ عمليَّة اتّخاذ القرار الخوارزمي قد تتعرَّض لأخطاء يمكن حصرها في نوعين رئيسين لا ثالث لهما:



شكل رقم (٢) يوضّح أبرز أخطاء التحيّز الخوارزمي

وهـما أخطاء التصنيف وأخطـاء العمليَّة؛ حيث قد يؤدّي هذان النوعـان إلى قرارات متحيّزة

1 -Tiago Marques: Overcoming Algorithmic Bias: The role of Bias Awareness, Knowledge, and Minority Status on Human Decision-Making, p. 7

2 - Tiago Marques: Overcoming Algorithmic Bias: The role of Bias Awareness, Knowledge, and Minority Status on Human Decision-Making, p.iii

لا محالة. وتنشـأ أخطاء التصنيف من نظريَّة الاختبارات الإحصائيَّة، وتأخذ شـكلين: الخطأ من النوع الأول (إيجابيات كاذبة) الذي يحدث عندما تصنّف الخوارزميَّة شيئًا على أنَّه ينتمي إلى فئة معيّنة بينما في الواقع لا ينتمي إليها، والخطأ من النوع الثاني (سلبيات كاذبة) الذي يحدث عندما تفشل الخوارزميَّة في التعرّف على شيء ينتمي فعليًا إلى الفئة الصحيحة. وفي كلتا الحالتين، تقوم الخوارزميَّـة بتصنيف أو توصيف فرد أو موقف بشـكل غير صحيح، ما يـؤدّي مباشرة إلى اتخاذ قرار غير دقيق. أما أخطاء العمليَّة، كما يوحي الاسم، فتحدث في عمليَّة اتخاذ القرار نفسها، على عكس أخطاء التصنيف التي تظهر فقط في النتائج، وتتعلّق بالعوامل أو فئات البيانات التي تأخذها الخوارزميَّة في الاعتبار عند اتخاذ القرار، مضافًا إلى كيفيَّة جمع بيانات التدريب واسـتخدامها. وعلى الرغم من تأكيد عيد من الباحثين عـلى ضرورة اتّباع الخوارزميَّات لمعايير إجرائيَّة محدّدة تشمل الإجراءات القانونيَّة، ومعايير العدالة، والمعايير الخُلُقيَّة، لكنَّ الباحثة ترى أنَّ أخطاء العمليَّة قد تستمرّ نتيجة تصميم بشري، سواء كان متعمدًا أم لا، أو بسبب وجود بيانات تدريب متحيّزة [1].

والجدير بالذكر أنَّ هذه الأخطاء تشير إلى أنَّ الخوارزميَّات قد تؤدّي إلى التحيّز الخوارزمي في اتّخاذ القرارات لا محالة، الأمر الذي يستدعي ضرورة دراسة هذا التحيّز لفهم أنواعه ومصادره. فالأخطـاء التصنيفيَّة وأخطاء العمليَّة، المرتبطة بالبيانـات أو بطريقة تصميم الخوارزميَّة، تجعل القرارات الناتجة غير عادلة أحيانًا، وتكشف عن تحوّل الخوارزميَّة من مجرّد أداة عقلانيَّة لمعالجة البيانات إلى فاعل خُلُقي قادر على إصدار أحكام متحيّزة، سواء عن قصد أم غير قصد. ومن هنا، يتراءى للباحثة بوضوح أنَّ فهم التحيّز الخوارزمي ليس خيارًا أكاديميًّا فحسب، بل خطوة رئيسة لتمهيد الطريق للانتقال إلى المحور الثاني من هذه الدراسـة الذي سـيتناول التحيّز الخوارزمي بمزيد من التفصيل، ويحلّل أشكاله ومصادره.

ثانيًا: التحيّز الخوارزمي- مفهومه، وأشكاله ومصادره.

يمكن تعريف التحيّز بشـكل عام على أنَّه خطأ منهجي في عمليَّات اتخاذ القرار يؤدّي

1 - Tiago Marques: Overcoming Algorithmic Bias: The role of Bias Awareness, Knowledge, and Minority Status on Human Decision-Making, p, 8

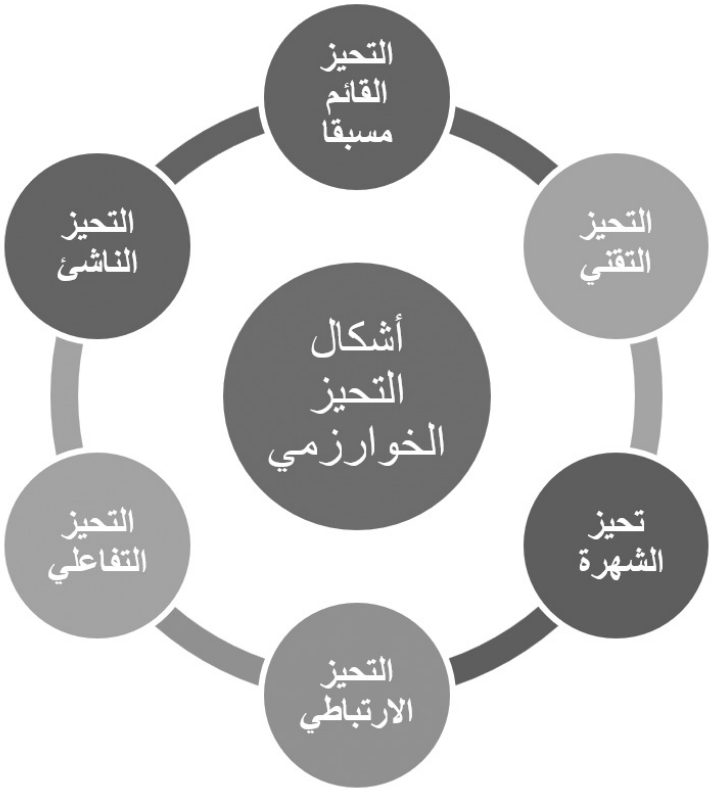
إلى نتائــج غير عادلة. وينشــأ هذا التحيّــز في أنظمة الذكاء الاصطناعــي من خلال جمع البيانــات، وتصميم الخوارزميَّات، أو التفســير البشري، كما يمكن لنــماذج التعلّم الآلي أن تتعلَّم وتكرّر أنماط التحيّز الموجودة في البيانات المســتخدمة لتدريبها، ما يؤدّي إلى نتائج غير عادلة أو تمييزيَّة[1]. وبناءً على ذلك، يُعدّ التحيّز الخوارزمي خطأً منهجيًّا أو ميلاً غــير متوقّع لتفضيل نتيجة على أخرى، وغالبًا ما يُوصف بأنَّه اعتماد غير مرغوب فيه على خصائص محدَّدة في البيانات تُنسب إلى مجموعة ديموغرافيَّة معيَّنة[2]. كما يحدث التحيّز الخوارزمي عندما يؤدّي تصميم الخوارزميَّة أو نموذجها الرياضي إلى تعزيز بعض النتائج على حســاب أخرى[3]، أو عندمــا تحتوي الخوارزميَّات على تحيّــزات كامنة تنعكس في نتائجها، سواء بسبب افتراضات متحيّزة في التصميم أم معايير متحيّزة في اتخاذ القرارات. وبعبارة أخرى، حتى لو كانت البيانات متوازنة، قد تؤدّي طريقة تصميم الخوارزميَّة نفسها إلى نتائج غير عادلة[4]. وانطلاقًا من هــذا الفهم العام للتحيّز الخوارزمي، تلاحظ الباحثة أنَّ غالبيَّة الباحثين يركّزون على تصنيف أنواعه وتحليل أشــكاله بتفصيل أكبر. كما تشــير الدراســات الحديثة للباحثين إلى أنَّ التحيّز في البرمجيَّات قد يكون ســلوكًا مدمجًا فيها لتســهيل حلّ المشكلات آليًّا، لكنَّه قد يضرّ المستخدمين بشكل مباشر أو غير مباشر على أســاس الجنس أو العِرق أو الإعاقة، ويتجلَّى ذلك في أشكال متعدّدة مترابطة تتداخل مع بعضها بعضًا، منها:

1 - Emilio, Ferrara: Fairness and Bias in artificial intelligence: a brief survey of sources, impacts, and mitigation strategies, p.2

2 - Marloes Susan Haitsma: Influence of Knowledge and trust on the perception of algorithmic gender bias in Artificial Intelligence, pp.8-9

3 - Emilio, Ferrara: Fairness and Bias in artificial intelligence: a brief survey of sources, impacts, and mitigation strategies, p.1

4 - Emilio, Ferrara: Fairness and Bias in artificial intelligence: a brief survey of sources, impacts, and mitigation strategies, p.3



شكل رقم (٣) يوضح أبرز أشكال التحيّز الخوارزمي

التحيّز القائــم مســبقًا(Pre-existing Bias) ، والتحيّز التقنــي (Technical Bias)، والتحيّز الناشئ (Emergent Bias)، وتحيّز الشهرة (Popularity Bias)، والتحيّز التفاعلي (Interaction Bias)، والتحيّز الارتباطي (Correlation Bias). ويعكس التحيّز القائم مسبقًا التمييز الموجود في المجتمع، الذي تنتقل آثاره إلى الخوارزميَّات، في حين ينشأ التحيّز التقني من قيود التصميم أو قدرات الحوسبة، كما هي الحال في أنظمة التعرّف على الصور التي قد تسيء تفسير المشهد لغياب السياق. وعندما تُســتخدم الخوارزميَّات في سياقات جديدة، يظهر التحيّز الناشئ، بينما يحدث تحيّز الشــهرة عندما تُفضّل الخوارزميَّات الخيارات الشائعة على الأكثر جودة. كما ينشأ التحيّز التفاعلي من تفاعل المســتخدمين مع النظام، كما حدث مع روبوت الدردشة (Tay) من

مايكروسـوفت. وأخيراً، يُشير التحيّز الارتباطي إلى اسـتنتاج معلومات حسَّاسة عن الأفراد من بيانات عاديَّة مرتبطة بها بشكل غير مباشر[١].

كما يمكن النظر إلى التحيّز الخوارزمي من زاوية أخرى، وهي تصنيفه إلى تحيّز صريح الذي ينشأ من مشـكلات في أخذ عيّنات البيانات أو سلوك غير متوقّع أثناء جمعها، وغالبًا ما يحدث عند وجود بيانات غير متوازنة، وتحيّز ضمني الذي ينشأ نتيجة روابط واتصالات غير متوقّعة بين المتغيّرات.[٢]

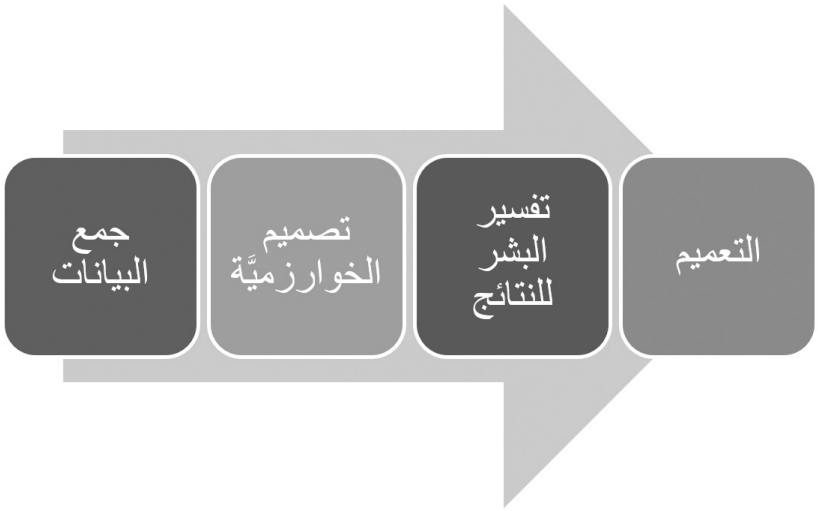
ومـن الواضح هنا أنَّ البُعد الخُلُقي يصبح جليًّا؛ إذ تظهر الدراسـات الحديثة أنَّ خوارزميَّات التعلّم الآلي قد تمارس أشكالاً من التمييز على أساس العِرق أو النوع الاجتماعي، وهو ما يكشف أنَّ هذه النظـم، رغم مظهرها التقني المحايـد، يمكن أن تعكس انحيـازات اجتماعيَّة قائمة[٣]. وبذلـك، يبرز التحدّي الخُلُقي في تصميم الخوارزميَّـات وتطبيقها؛ حيث يتطلَّب الأمر الانتباه إلى هذه الانحيازات ومراعاة العدالة والإنصاف عند اتّخاذ القرارات الآليَّة، لتجنّب تعزيز التمييز الموجود في المجتمع.

وأمَّا في مـا يتعلّق بمصادر التحيّز الخوارزمي، تلاحظ الباحثـة أنَّ أنظمة الذكاء الاصطناعي تمتلك القدرة على إحداث تأثيرات كبيرة تعزّز حياة الأفراد بطرق متعدّدة. ومع ذلك، ترى الباحثة أنَّ أحد التحدّيات الرئيسـة التي تواجـه تطوير هذه الأنظمة ونشرها هو التحيّز الذي يُشـير إلى الأخطاء المنهجيَّة في عمليَّة اتخاذ القرار التي قد تؤدّي إلى نتائج غير عادلة. ولعلَّ من أبرز الأمثلة الواضحـة على أنظمة الذكاء الاصطناعي التي قد تحمـل تحيّزًا هي الخوارزميَّات التي تُعدّ أداة رئيسة لمعالجة البيانات، واتخاذ القرارات ضمن هذه الأنظمة.

1 - Erica Thompson: Artificial Intelligence has an Implicit Bias Problem.p.p.3-4

2 - Marloes Susan Haitsma: Influence of Knowledge and trust on the perception of algorithmic gender bias in Artificial Intelligence, p.9

3 - Joy Buolamwini & Timnit Gebru: Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, p.1



شكل رقم (٤) يوضّح أبرز مصادر التحيّز الخوارزمي

ويمكن أن ينشأ التحيّز في الخوارزميَّات من عدّة مصادر، بما في ذلك طريقة جمع البيانات، وتصميم الخوارزميَّة، وتفسير البشر للنتائج. فعلى سبيل المثال، تستطيع نماذج التعلّم الآلي، وهي أحد أشكال الخوارزميَّات، التعرّف على أنماط التحيّز الموجودة مسبقًا في البيانات المستخدمة لتدريبها، ما قد يؤدّي إلى قرارات غير دقيقة أو منحازة. كما يمكن أن تظهر مصادر التحيّز في مختلف مراحل عمل الخوارزميَّة، بدءًا من جمع البيانات، مرورًا بتصميم التعليمات، وصولاً إلى تفاعل المستخدمين مع النظام [1].

كما إذا نظرنا إلى التعميم، لوجدناه يمثّل أحد أهمّ المصادر المؤدّية إلى التحيّز الخوارزمي. وتأكيدًا لهذا الزعم، لو دقّقنا النظر في الفكرة التي يروّج لها كثيرون، ومفادها أنَّ معظم ذوي البشرة السوداء أناس خطرون، أو أنَّ أصحاب اللحى ذوو فكر إرهابي، لوجدنا أنَّ الخوارزميَّات تقوم بإطلاق تحذيرات خاطئة ضد أناس أبرياء. ناهيك عن ذلك، لو نظرنا إلى وكالة بلومبيرغ، لوجدنا أنَّها حلّلت أكثر من خمسة آلاف صورة أنشأتها شركة ستيبليتي إيه آي (Stability AI)، وقد

1 - Emilio Ferrara: Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies, p.2

كشفت هذه التحليلات أنَّ برامج الشركة قد ضخمّت الصور النمطيَّة المرتبطة بالعرق والجنس؛ إذ غالبًا ما يُصوَّر الأشخاص ذوو البشرة الفاتحة على أنَّهم يشغلون وظائف ذات رواتب مرتفعة، بينما يُصنَّف الأشخاص ذوو البشرة الداكنة على أنَّهم يعملون في غسيل الأطباق وتدبير المنازل، وغير ذلك من الصور النمطيَّة المؤسفة[١] .

وبناءً على ما سبق، يصبح من الضروري للباحثة الانتقال إلى المحور الآتي من هذه الدراسة الذي يركّز على مظاهر التحيّز الخوارزمي، والأمثلة العمليَّة التي تكشف عنه في الواقع، لفهم أثر هذه الظاهرة على نتائج اتخاذ القرار بشكل أدقّ. فإلى جانب التأثيرات التقنيَّة والإجرائيَّة، ينطوي التحيّز الخوارزمي على أبعاد خُلُقيَّة مُهمَّة؛ إذ يمكن أن يؤدّي إلى قرارات غير عادلة أو متحيّزة تؤثّر في الأفراد والمجتمعات. ويُسهم هذا الانتقال في فتح الباب لفهم كيف يمكن للخوارزميَّات، رغم دقّتها الحسابيَّة، أن تتحول إلى أدوات ذات تبعات خُلُقيَّة حقيقيَّة، بما يخلق جسرًا منطقيًّا بين مصادر التحيّز واستكشاف مظاهره الخُلُقيَّة في العالم الواقعي.

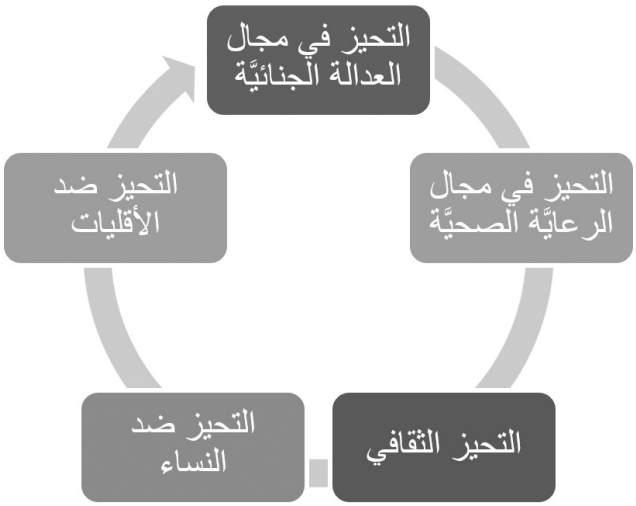
ثالثًا: مظاهر التحيّز الخوارزمي وأبعاده الخُلُقيَّة.

يمثّل هذا المحور نقطة الانطلاق الحقيقيَّة لفهم كيفيَّة ظهور التحيّز في الخوارزميَّات، وما يترتّب عليه من نتائج خُلُقيَّة على الأفراد والمجتمعات. ويركّز هذا المحور على مظاهر التحيّز الخوارزمي في الممارسة العمليَّة، مع عرض للأمثلة التي تكشف تأثيره على البشر، بما يمكّن القارئ من إدراك الأبعاد الخُلُقيَّة المصاحبة لاستخدام هذه الأدوات في اتخاذ القرار. ويشير ذلك إلى أنَّ الخوارزميَّات نفسها التي صُمّمت أساسًا لتحسين عمليَّة اتخاذ القرار، قد تعزّز عن قصد أو عن غير قصد الفوارق القائمة، أو تخلق أشكالاً جديدة من التمييز. ويعود سبب هذا التحيّز إلى وجود أخطاء منهجيَّة، كما أشرنا من قبل، يمكن أن تؤدّي إلى معاملة غير عادلة للأفراد أو الجماعات بناءً على خصائص مثل العرق، أو الجنس، أو الوضع الاجتماعي والاقتصادي.[٢]

---

١ - جقريف الزهرة: إشكاليَّة التحيز الخوارزمي في أنظمة الذكاء الاصطناعي، وأثرها على حقوق الإنسان: الحق في العمل أنموذجًا، ص٢٣.

2 - Tommy Fred: Bias and Fairness in AI Algorithms, p.1



شكل رقم (٥) يوضّح أبرز مظاهر التحيّز الخوارزمي

وتتجـلَّى أبرز مظاهر التحيّـز الخوارزمي في الذكاء الاصطناعي من خـلال عدّة أمثلة عمليَّة تظهـر كيف يمكن للأنظمة التكنولوجيَّة أن تعكس الانحيـازات الاجتماعيَّة القائمة بل وتزيدها تفاقمًا. فعلى سبيل المثال لا الحصر، عند مطالبة نماذج الذكاء الاصطناعي التوليدي بإنشاء صور للرؤسـاء التنفيذيين، كانت النتائج في الغالب صورًا للرجال، وهـو ما يعكس التحيّز الجندري ويشـير بوضوح إلى نقص تمثيل النسـاء في المناصب التنفيذيَّة على أرض الواقع. ومضافًا إلى ذلك، عند طلب إنشاء صور للمجرمين أو الإرهابيين، تميل النتائج بشكل كبير نحو الأشخاص مـن الأقلّيات العرقيَّة، ما يوضّح أنَّ التحيّز في الذكـاء الاصطناعي التوليدي ليس مجرّد عارض تقني، بل يعكس الانحيازات الاجتماعيَّة الموجودة ويعزّزها.

ويتجـاوز التحيّـز الخوارزمي نطاق الصـور التوليديَّة، ليظهر في أنظمـة الذكاء الاصطناعي المستخدمة في مختلف المجالات، بما في ذلك مجال العدالة الجنائيَّة والرعاية الصحيَّة. فعلى سـبيل المثال لا الحصر، يُعدّ نظام التنبؤ بخطر إعادة ارتكاب الجريمة المستخدم في الولايات المتحدة نموذجًا شهيرًا؛ إذ أظهرت الدراسات أنَّه منحاز ضدّ المتَّهمين من أصل أفريقي؛ حيث كانـوا أكثر عرضة لتصنيفهم على أنّهم ذوو خطورة أعلى، حتى في الحالات التي لم تكن لديهم

فيها سجلّات جنائيَّة سابقة. وقد أظهرت دراسات أخرى وجود تحيّزات مماثلة في أنظمة مشابهة مستخدمة في ولايات أخرى، ما يعكس انتشار هذه المشكلة في المجال القضائي[1].

والجدير بالذكر هنا أنَّ التحيّز لا يقتصر على مجال العدالة الجنائيَّة أو المراقبة الأمنيَّة بل يمكن أن يعني أيضًا، على سبيل المثال لا الحصر، تعرُّض مستخدمي خدمات الإنترنت لتحيّزاتٍ ضدّهم إذا صنَّفهم الذكاء الاصطناعي تصنيفًا سيّئًا[2].

ولو نظرنا إلى قطاع الرعاية الصحيَّة، لوجدنا أنَّه جرى اكتشاف تحيّز مشابه؛ حيث أظهر نظامٌ يُستخدم للتنبّؤ بمعدّلات الوفاة ميلاً إلى إعطاء المرضى من أصل أفريقي درجات خطورة أعلى، حتى عندما تكون عوامل أخرى، مثل العمر والحالة الصحيَّة، متساوية مع غيرهم. وهذا التحيّز لا يعكس بالضرورة خطرًا أعلى حقيقيًّا، لكنَّه قد يؤدّي إلى اتخاذ قرارات علاجيَّة غير عادلة، مثل منع هؤلاء المرضى من الوصول إلى الرعاية المناسبة أو تلقي علاج أقل جودة، ما يزيد من الفجوات الصحيَّة القائمة بالفعل.

أما في مجال تقنيَّات التعرف على الوجوه المستخدمة من قبل أجهزة إنفاذ القانون، فقد أظهرت الدراسات أنَّ هذه الأنظمة أقلّ دقَّة مع الأشخاص ذوي البشرة الداكنة، ما يؤدّي إلى ارتفاع معدّلات الإيجابيات الكاذبة. وقد تكون لهذه التحيّزات آثار شديدة الخطورة، مثل الاعتقالات الخاطئة أو الإدانات غير العادلة، وهو ما يبرز البُعد الخُلُقي الحادّ المرتبط بالتحيّز الخوارزمي وضرورة التعامل معه بحذر. ومن ثَمَّ، يتبيّن للباحثة أنَّ التحيّز الخوارزمي ليس مسألة تقنيَّة بحتة، بل هو انعكاس للانحيازات الاجتماعيَّة والاقتصاديَّة والثقافيَّة الموجودة مسبقًا، ما يجعلها ترى ضرورة تصميم أنظمة ذكاء اصطناعي عادلة وشفافة، لتجنّب تعزيز هذه التحيّزات، ومنع التسبّب في أضرار فعليَّة على الأفراد والمجتمعات[3].

كما تلاحظ الباحثة، في ظلّ تزايد الاعتماد على الذكاء الاصطناعي في اتخاذ القرارات

1 - Emilio Ferrara: Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies, pp.3-4

٢ - مارك كوكلبيرغ: أخلاقيات الذكاء الاصطناعي، ص٩٠

3 - Emilio Ferrara: Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies, pp3-4

اليوميَّة، يظهر التحيّز الخوارزمي بوضوح في مجالات متعدّدة، لا سيّما في الإعلانات المستهدفة للوظائف على الإنترنت. فقد اضطرّت شركة أمازون، عملاق التجارة الإلكترونيَّة، إلى إلغاء أداة للتوظيف تعتمد على الذكاء الاصطناعي بعد اكتشاف أنَّ الأداة كانت متحيّزة ضد المتقدّمين الذين تحتوي سيرهم الذاتيَّة على كلمات تدل على النشاطات النسويَّة، مثل المشاركة في فرق أو أنشطة رياضيَّة نسويَّة. وقد جرى تدريب خوارزميَّة التوظيف على بيانات التوظيف الفعليَّة للشركة خلال عشر سنوات التي كانت تمثّل الذكور بشكل أكبر، فتعلّمت الخوارزميَّة من هذه البيانات أن شركة أمازون تميل إلى تفضيل المتقدّمين الذكور على الإناث [1].

كما يمكن أن يظهر التحيّز في المجال الثقافي، بمعنى أنَّ الخوارزميَّات تتسبب في وجود تحيّز ثقافيٍّ، بمعنى أنَّها تميل إلى تعزيز ثقافات أو رؤى معينة على حساب أخرى، الأمر الذي يترتّب عليه عدم تمثيل الثقافات المتنوعة، وهيمنة ثقافة رقميَّة على حساب غيرها، فضلاً عن تمثيل غير متوازن للثقافات المختلفة في المحتوى الرقمي. والجدير بالذكر أنَّ الأمر لا يتوقَّف عند هذا الحدّ، بل يمتدّ إلى التفاعل الاجتماعي؛ حيث يمكن أن يخلق بيئات رقميَّة تعزّز من الفوارق الثقافيَّة[2].

لو ركّزنا على منصّات التواصل الاجتماعي، لوجدنا أنَّ هناك بعض الممارسات المتحيّزة؛ حيث نجد على سبيل المثال أنَّ بعض الخوارزميَّات تتحيّز ضد المحتوى الفلسطيني، وهذا يعني ضمنيًّا أنَّ هذه التقنيَّات أضحت تمثّل نوافذ دعائيَّة تسهم بشكل سلبي، بطريقة أو بأخرى، في تشكيل آراء المستخدمين والسيطرة على الرأي العام الدولي المناقش لها عبر صفحاتها. وتأكيدًا لهذا الزعم، يمكننا أن نجد أنَّ الحكومة الإسرائيليَّة تقوم بعقد اتفاقيَّات وشراكات مع بعض الوسطاء الرقميّين، بهدف إزالة المحتويات الفلسطينيَّة والعربيَّة داخل فلسطين وخارجها، وهي تلك المحتويات التي تدلّ على انتهاكات الحكومة الإسرائيليَّة، ما أدَّى إلى وجود بعض الممارسات التمييزيَّة والتعاونيَّة لإزالة أو مراقبة أو حظر المحتوى الفلسطيني، من أجل ترويج

1 - Erica Thompson: Artificial Intelligence has an Implicit Bias Problem, p.4

٢ - عبد الوهَّاب بوبعة: التحيز الثقافي في خوارزميَّات شبكات التواصل الاجتماعي- قراءة نظريَّة في آليَّات التحيّز الرقمي والهيمنة الثقافيَّة، ص ٨٢-٨٣

الشعور بأنَّ فلسطين وشعبها لا يعانيان من أي انتهاكات أو اعتداءات غير مبرّرة[١] .

عــلاوة على ذلك، تلاحــظ الباحثة أنَّ هناك أمثلة لا حصر لها عــلى التحيّز العرقي في النظام القضائي، ولا سـيّما في الخوارزميَّــات المتعلّقة بتقويــم خطورة الجرائــم وخوارزميَّات قياس السمية في النصوص. وترى الباحثة أنَّ هذه الخوارزميَّات جرى تدريبها على بيانات مأخوذة من صفحــات النقاش العامة لتقويم النصوص وقياس درجة الســمية، لكن هذه البيانات قُوّمت من قبل أشخاص خارج أي سياق ثقافي، ودون مراعاة اللهجات الخاصة بالأقلّيات. ونتيجة لذلك، كانت الخوارزميَّات أكثر ميلاً لتصنيف النصوص التي كتبها هؤلاء الأشــخاص على أنّها مسيئة، على الرغم من أنّ هذه الكلمات لم تكن مؤذية إذا ما نظرنا إلى سياقها الثقافي.

كما تلاحظ أنَّ خوارزميَّة كومباس[٢] (COMPAS) قد اســتُخدمت لتقويم مخاطر العودة إلى ارتكاب الجريمة، وتحديد أيّ الجناة أكثر احتمالاً لإعادة الاعتقال. وقد أظهرت النتائج أنّ هذه الخوارزميَّة كانت تمثّل الجناة الســود تمثيلاً زائــدًا؛ إذ جعلتهم أكثر عرضة للتصنيف ضمن فئة عالية الخطورة، في حين كان الجناة البيض أكثر احتمالاً لِأَنْ يُصنَّفوا – على نحو خاطئ – ضمن فئة منخفضة الخطورة. ولم يقتصر هذا الضرر على الجناة وحدهم بل امتدّ ليشمل ضحايا الجرائم أيضًا، وهو ما يعكس أنّ التحيّز العرقي والجندري الكامن في هذه الخوارزميَّات يمثّل شكلاً من أشكال التمييز المنظومي[٣].

ولعلّي لا أبالغ إذا قلت إنَّ من يدقّق النظر في معظم أشــكال التحيّز الخوارزمي ســوف يجد أنَّها تنشــأ غالبًا من التحيّزات الثقافيَّة البشريَّة القائمة مســبقًا التي تنتقــل إلى الخوارزميَّات عبر

---

١ - الحاج عيسى بن صفي الدين، سفيان غينو: الحجاج بواسطة الخوارزميَّات: كيف تكبح منصات التواصل الاجتماعي الخطاب الرقمي العربي؟، ص ٢٠١-٢٠٢.

٢ - تُستخدم هذه الخوارزميَّة لتقويم خطر عودة المتَّهم إلى ارتكاب الجريمة في النظام القضائي، ونُظهر أنَّها لا تعتمد بشكل مباشر على عمر المتَّهم كما ادَّعى منشؤها. وعند إزالة المكوّن المعقَّد المرتبط بالعمر، نلاحظ أنَّ الخوارزميَّة لا ترتبط بالضرورة بالعرق، على عكس ما أشارت إليه بعض الدراسات السابقة. وبذلك، تؤدّي الافتراضات الخاطئة عن عمل الخوارزميَّة المغلقة إلى استنتاجات مضلّلة، ويمكن تجنّب هذه الأخطاء إذا كانت الخوارزميَّة شفَّافة منذ البدايَّة، ولمعرفة المزيد راجع بالتفصيل: Cynthia Rudin, Caroline Wang, Beau Coker: The age of secrecy and unfairness in recidivism prediction

3 - Erica Thompson: Artificial Intelligence has an Implicit Bias Problem. Pp.19-21

البيانات الأوَّليَّة المستخدمة لتدريبها، وهو ما يُعرف بالتحيّز المسبق. كما قد يظهر تحيّز ناتج عن العمليات نفسها أثناء استخدام الخوارزميَّات، مضافًا إلى تحيّز الارتباط الذي ينشأ أثناء التفاعل مع البيانات. وغالبًا ما يجري تجاهل وجود هذا التحيّز في نتائج الخوارزميَّات؛ لأنَّه يتوافق مع التحيّز الثقافي البنيوي السائد في المجتمع، ما يجعله يبدو طبيعيًا أو متوقّعًا. ويتّضح ذلك بجلاء في اختبارات تضمين الترابطات اللفظيَّة التي أظهرت أنَّ الخوارزميَّات التي تحلّل البيانات النصّيَّة تعلّمت ربط النساء بمهن مثل التمريض تقريبًا بالنسبة الفعليَّة نفسها للنساء العاملات في هذه المهنة.

كما يتجلَّى التحيّز ضد النساء بشكل واضح في خوارزميَّات الإعلانات الوظيفيَّة، فقد كشفت الدراسات أنَّ الإعلانات الخاصَّة بالوظائف منخفضة الأجر، مثل وظيفة معلّمة رياض الأطفال، عُرضت على النساء بشكل أكبر حتى عندما جرى استهداف الجمهور بطريقة محايدة من حيث النوع الاجتماعي. وتعكس هذه النتائج تحيّزا ثقافيًا يربط النساء بالوظائف منخفضة الأجر ومجالات الرعاية، ويُرجّح أنَّ معظم الخوارزميَّات المسؤولة عن عرض الإعلانات قد جرى تطويرها بواسطة رجال الذين يشكّلون أكثر من نصف العاملين في قطاع التكنولوجيا المتقدّمة في بيئة عمل غالبًا ما تكون صعبة على النساء. وتشير الدراسات إلى أنَّ أكثر من نصف الموظّفات يشعرن بضرورة إثبات كفاءتهن أمام زملائهن من الرجال، ويتعرّضن لردود فعل سلبيَّة قائمة على النوع الاجتماعي، ما يزيد من تعزيز التحيّز في النتائج التي تقدّمها الخوارزميَّات.

يعكس التحيّز الخوارزمي -أيضًا- ضدّ الأقليات التحيّز الثقافي والاجتماعي القائم ضدّ هذه الفئات. وتأكيدًا لهذا الزعم، ترى الباحثة أنَّ التاريخ الطويل من المراقبة الشرطيَّة المفرطة للمجتمعات الأمريكيَّة من أصول إفريقيَّة قد أدَّى إلى إنتاج مجموعات بيانات تعكس بشكل مضلّل أنَّ هؤلاء الجناة أكثر ميلاً للعودة إلى الجريمة مما تُظهره الوقائع. وقد تلقَّى الجناة من هذه الفئات عقوبات أشدّ مقارنة بالجناة من ذوي البشرة البيضاء، ما دفع خوارزميَّة كومباس إلى تصنيفهم على أنَّهم أعلى خطورة، وزيادة احتمال تعرّضهم لعقوبات صارمة. وأظهرت المقارنة بين التوقّعات الفعليَّة والتوقّعات الخوارزميَّة وجود تحيّز عرقي واضح، تمامًا كما هو الحال بالنسبة إلى الانحياز ضدّ النساء.

ولا شـكّ أنَّ المشـكلة تتفاقم بسـبب نقص تمثيل الأقلّيـات في شركـات التكنولوجيا التي تطوّر نماذج الذكاء الاصطناعي؛ حيث يشـكّل البيض غالبيَّة العاملين في هذا القطاع، ما يجعل الخوارزميَّـات تفتقـر إلى الفهم الثقـافي الكامل للسـياق الاجتماعي المتعلّـق بالعرق والنوع الاجتماعي والفئات المحميَّة الأخرى. وقد أدّت هـذه الفجوة إلى إخفاء ممنهج للثقافات غير الممثّلة، وبالتالي إلى آثار سلبيَّة على المستخدمين الذين يتعاملون مع هذه الأنظمة.

كما تلحـق الخوارزميَّات المتحيّزة أضرارًا جسـيمة بالأقلّيات العرقيَّة؛ إذ يعـاد إنتاج التمييز البنيوي في المجتمع، كما يظهر في المراقبة الشرطيَّة المفرطة للمجتمعات الأمريكيَّة من أصول إفريقيَّـة، وفي التمييز في منح القروض حسـب الهويَّة العرقيَّة؛ حيث كانـت البنوك تحدّد أهليَّة الحصـول على القروض بنـاءً على العرق. وحتى بعض الخوارزميَّـات المصمّمة نظريًا لحماية الأقلّيَّات من المحتوى الضارّ أو المسيء على الإنترنت، انتهى بها الأمر إلى إسـكات أصواتهم بدلاً من حمايتهم، ما يقلّل من ظهورهم في الفضاءات الرقميَّة والاجتماعيَّة. كما أدَّى هذا التحيّز إلى حرمان بعض الجناة من الإفراج المشروط بشـكل غير عادل، ما يعمّق شـعورهم بالتهميش والغبن.

وعلى الرغم من توفّر بعض الأسـاليب لفحـص الخوارزميَّات بحثًا عـن التحيّز، فإنَّ تطبيق هذه الإجراءات بشـكل منتظم ليس واضحًا. فقد أنكرت الشركات المطوِّرة لخوارزميَّة كومباس وجود أي تحيّز عرقي، مسـتندة في ذلك إلى دقّة التنبؤ للجناة البيض والسـود الذين عادوا فعليًّا إلى الجريمـة. لكنَّ هذا الدفاع يغفل التفاوت العرقي في الأخطاء؛ حيث كانت الخوارزميَّة أكثر ميلاً لتصنيف الجناة السـود خطأ على أنَّهم أكثر خطورة، في حين صنَّفت عددًا كبيرًا من الجناة البيض خطأ على أنَّهم أقل خطورة. وهذا يعني أنَّه، رغم دقّتها النسبيَّة، قد فشلت الخوارزميَّة في معالجة اختلال توزيع النتائج الخاطئة، ولم تأخذ في الاعتبار العدد الأكبر للجناة السـود داخل منظومة العدالة الجنائيَّة[1].

بالتالي، يتّضح للباحثة أنَّ للتحيّز الخوارزمي تبعات خُلُقيَّة متعدّدة يجب أخذها في الاعتبار،

1 - Erica Thompson : Artificial Intelligence has an Implicit Bias Problem., pp. 38-40

لعــلَّ من أبرزها احتماليَّة التمييــز ضدّ الأفراد أو المجموعات بناءً عــلى عوامل، مثل العرق، أو الجنس، أو العمر، أو الإعاقة. فحين تكون الأنظمة متحيّزة، قد تعيد إنتاج عدم المساواة القائمة وتزيد من التمييز ضد الفئات المهمَّشة، ويصبح هذا الأمر أكثر خطورة في المجالات الحسَّاسة، مثــل الرعاية الصحيَّة؛ حيث يمكن للأنظمة المتحيّزة أن تــؤدّي إلى عدم تكافؤ فرص الحصول عــلى العلاج أو إلحاق الضرر بالمــرضى. كما يمكن أن يؤدّي اســتخدام الأنظمة المتحيّزة إلى تقويض ثقة الجمهــور في التكنولوجيا، ما يقلّل اعتمادها أو حتى يؤدّي إلى رفضها، وله تبعات اقتصاديَّــة واجتماعيَّة خطيرة إذا لم يثق النــاس بهذه الأنظمة أو اعتبروها أداة للتمييز. كما ينبغي النظر في تأثــير الأنظمة المتحيّزة على حرّيَّة الفرد واســتقلاليّته؛ حيث قد تحدّ هذه الأنظمة من الحُرّيَّة الفرديَّة وتعزّز ديناميّات السلطة في المجتمع، فعلى سبيل المثال قد يؤدّي استخدام نظام متحيّز في عمليّات التوظيف إلى اســتبعاد عدد أكبر من المرشّحين من الفئات المهمّشة بشكل غير متناسب، ما يقلّل من قدرتهم على الوصول إلى فرص العمل والمساهمة في المجتمع[1].

وبالتالي، يمكن للباحثة القول إنَّ التحيّز الخوارزمي لا يهدّد الموضوعيَّة فحســب، بل يعمل أيضًا على إغفال بعض الحقائق وتشويهها، ومنح الأخبار الكاذبة فرصًا كبيرة للظهور على الساحة الرقميَّــة، الأمر الذي يترتّب عليه تأجيج الصراع المجتمعي وتأليب الرأي العام. كما يغفل بعض المبادئ المُهمَّة، مثل مبدأ الشــفافيَّة، وتقديم خدمة إخباريَّة عموميَّة تهدف إلى تحقيق الصالح العام[2].

ناهيــك عن ذلك، ترى الباحثة أنَّ التحيّز الخوارزمــي يمتدّ عبر مجالات متعددة، من العدالة الجنائيَّــة والرعاية الصحيَّة إلى الإعلانات الرقميَّة وتقنيَّات التعرف عــلى الوجه، وله آثار خُلُقيَّة جسيمة على الأفراد والمجتمعات. فهــو يعيد إنتاج التمييــز البنيوي القائــم، ويزيد من معاناة الأقلّيات والجماعات المهمشــة، كما يؤثّر على فرص النساء في سوق العمل ويضعف العدالة

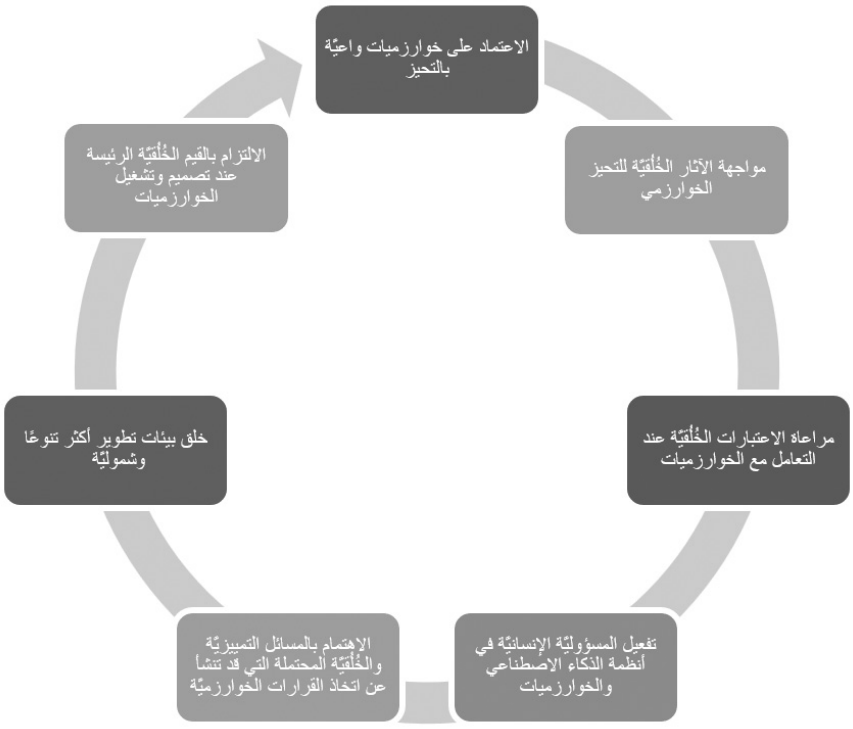
1 - Emilio Ferrara : Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies, p.6

٢ - خالد لراره: التحيّز الخوارزمي آليَّة رقميَّة لإدارة المعلومة، إنتاج سرديَّات بديلة وتوجيه الرأي العام، ص١٣٥٣.

والمســاواة. ومن هذا المنطلق، تؤكّــد الباحثة على أنَّ فهم هذه التحيّــزات والاعتراف بها ليس كافيًا، بل من الضروري البحث عن حلول عمليَّة تقلّل من آثارها الخُلُقيَّة، وتضمن شفافيَّة وعدالة الخوارزميَّــات. كما أنَّ هذه الآثار دفعت البشر أحيانًا إلى الاستســلام لهــا وقَبولها وكأنَّها واقع مفروض عليهم، ما يجعل التفكير في تغيير هذه الثقافة والتخفيف من حدّتها أمرًا ملحًّا. ولذلك، سوف يكون المحور الآتي مكرَّسًا لاستكشاف الاستراتيجيات الممكنة لمعالجة هذه التحيّزات، وتقديم إطار أكثر عدالة في استخدام الخوارزميَّات في اتّخاذ القرار.

رابعًا: استراتيجيَّات معالجة التحيّز الخوارزمي.

تهتمّ الباحثة في هذا المحور بالتركيز على أبرز الطرق والأساليب المعتمدة لتقليل الأثر الخُلُقي الناجم عن التحيّز الخوارزمي، بهدف ضمان أنظمة أكثر عدالة وموضوعيَّة في اتخاذ القرارات.



شكل رقم (٦) يوضحّ أبرز استراتيجيات معالجة التحيّز الخوارزمي

وتتمثل الاستراتيجيَّة الأولى في الاعتماد على خوارزميَّات واعية بالتحيّز، أي خوارزميَّات تُصمَّم لأخذ أنواع التحيّز المختلفة في الاعتبار، بهدف تقليل تأثيرها على نتائج النظام .[1]

أما الاستراتيجيَّة الثانية فتكمن في ضرورة إدراك أنَّ مواجهة الآثار الخُلُقيَّة للتحيّز الخوارزمي تتطلَّب جهدًا منسّقًا من جميع الجهات المعنية بتصميم الخوارزميَّات وتشغيلها. ومن المُهمّ أيضًا تطوير إرشادات خُلُقيَّة وأطر تنظيميَّة تهدف إلى تعزيز العدالة، وضمان الشفافيَّة في تصميم الخوارزميَّات وتشغيلها، وتعزيز المساءلة للمطوّرين والمستخدمين وصانعي السياسات، بما يضمن الاستخدام المسؤول للخوارزميَّات[2] .

وتتجلَّى الاستراتيجيَّة الثالثة في ضرورة مراعاة الاعتبارات الخُلُقيَّة عند التعامل مع الخوارزميَّات، ولعلَّ من أبرزها:

١. احترام كرامة الإنسان، ويتضمن ذلك الفكرة القائلة إنَّ كلَّ إنسان يمتلك قيمة جوهريَّة، ويجب ألا تُنتقص هذه القيمة أو تُقمع من قبل الآخرين، ولا حتى من قبل التقنيات الحديثة، مثل أنظمة الذكاء الاصطناعي. والمقصود هنا باحترام كرامة الإنسان هو معاملة جميع الأفراد بالاحترام الذي يستحقّونه بوصفهم موضوعات خُلُقيَّة، وليس مجرّد كائنات يجري تصنيفها أو تقويمها بطريقة ميكانيكيَّة. لذا، يجب تطوير أنظمة الذكاء الاصطناعي بحيث تحمي سلامة الإنسان الجسديَّة والعقليَّة، وتحترم هويَّته الشخصيَّة والثقافيَّة، وتلبّي احتياجاته الأساس.

٢. حريَّة الفرد، حيث يجب أن يظلّ الإنسان حُرًّا في اتخاذ قرارات حياته بنفسه، بما يشمل الحريَّة من التدخّل غير المبرّر، مع ضمان حصول الأشخاص المهدّدين بالاستبعاد على فرص متساوية للاستفادة من مزايا الذكاء الاصطناعي. وبالطبع، تتطلّب حريَّة الفرد التخفيف من الإكراه المباشر أو غير المباشر، والتهديدات لاستقلاليته العقليَّة

1 - Emilio Ferrara: Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies, p.3

2 - Emilio Ferrara: Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies, p.5

أو صحّته النفسيَّة، والمراقبة أو التلاعب غير المبرّر. كما تشمل حريَّة الفرد تمكين الأفراد من السيطرة على حياتهم، بما في ذلك حماية حريَّة ممارسة الأعمال، والفنون والعلوم، والتعبير، وحق الخصوصيَّة.

٣. احترام الديمقراطيَّة والعدالة وسيادة القانون، حيث يجب أن تعمل أنظمة الذكاء الاصطناعي على الحفاظ على الديمقراطيَّة وتعزيزها. ويجب ألا تقوّض هذه الأنظمة عمليّات التصويت الديمقراطيَّة، كما يجب أن تلتزم بعدم تقويض الأسس القانونيَّة القائمة، بما في ذلك القوانين واللوائح، وضمان الإجراءات القانونيَّة الواجبة والمساواة أمام القانون.

٤. المساواة وعدم التمييز والتضامن، بما في ذلك حقوق الأشخاص المهدّدين بالإقصاء، وهذا يعني ضمان احترام متساوٍ للقيمة الخُلُقيَّة وكرامة جميع البشر. ولا يقتصر ذلك على مجرّد عدم التمييز، بل يتضمن ضمان أنَّ عمليَّات أنظمة الذكاء الاصطناعي لا تولد نتائج متحيّزة بشكل غير عادل، مع استخدام بيانات شاملة تمثل مختلف الفئات السكّانيَّة. كما يتطلّب احترام الأشخاص والمجموعات الضعيفة، مثل العمال، والنساء، والأشخاص ذوي الإعاقة، والأقلّيات العرقيَّة، والأطفال، والمستهلكين، أو أي أشخاص معرضين للإقصاء[1]. وتكمن الاستراتيجيَّة الرابعة في تفعيل المسؤوليَّة الإنسانيَّة في أنظمة الذكاء الاصطناعي والخوارزميَّات، وهذا يعني تطوير هذه الأنظمة وفق المبادئ والقيم الإنسانيَّة الأساس، من أجل ضمان ازدهار الإنسان وتحقيق رفاهيّته[2].

٥. تتمثّل الاستراتيجيَّة الخامسة في أن المؤسسات في وقتنا الراهن - أكثر من أي وقت مضى- ينبغي ألا تركز فقط على كيفيَّة توليد القيمة من استخدام الخوارزميَّات، بل يجب أن تولي اهتمامًا أيضًا للمسائل التمييزيَّة والخُلُقيَّة المحتملة التي قد تنشأ عن اتخاذ القرارات الخوارزميَّة. فغالبًا ما يُنظر إلى القرارات الخوارزميَّة على أنها

1 - European Commission, Ethics guidelines for trustworthy AI, pp.10-11

2 - Virginia Dignum: Artificial Intelligence: Foundations, Theory, And Algorithms, p119

موضوعيَّة، بينــما الخوارزميَّات، في الحقيقة، محملة بالقيم ومصممة وفق مجموعة محددة من المبادئ والافتراضات الثقافيَّة والاجتماعيَّة السائدة[١] .

٦. أما الاستراتيجيَّة السادســة فتتمثَّل في ضرورة خلق بيئات تطوير أكثر تنوّعًا وشموليَّة ضمن فرق تصميم أنظمة الذكاء الاصطناعي وتشــغيله؛ حيث إنَّ إشراك وجهات نظر ثقافيَّة متعدّدة يزيد من الوعي بالاختلافات الثقافيَّة، ما يساعد على تحديد المعالجات المختلفة والنتائج المتباينة في الخوارزميَّات. كما تشــمل هذه الاستراتيجيَّة توظيف مراجعين مدرّبين على التحيّز لتدقيــق مدخلات ومخرجات الخوارزميَّة، ويفضَّل أن يتمتّع المراجعون بإمكانيَّة الوصول إلى الآليَّات الداخليَّة للخوارزميَّة كلما أمكن[٢] .

٧. تتجلَّى الاســتراتيجيَّة السابعة والأخيرة في ضرورة الالتزام بالقيم الخُلُقيَّة الرئيسة عند تصميم الخوارزميَّات وتشــغيلها، ومن أبرز هذه القيم الخصوصيَّة؛ حيث يجب عدم الكشــف عن المعلومات الحسَّاســة أو اســتخدامها لإيذاء الأشــخاص أو التلاعب بهم. كما تشــمل هــذه الاســتراتيجيَّة قيمة الشــفافيَّة؛ إذ يجب أن تكــون التوصيات والقرارات التي ترشّــحها الخوارزميَّة للبشر واضحة ومفهومة. وقد كانت هذه واحدة من الاهتمامات الرئيســة المتعلّقة بخوارزميَّات التعلّم الآلي والذكاء الاصطناعي؛ إذ غالبًا لا يفهم المطوّر كيف تتوصل الخوارزميَّة إلى قرار معيّن، ويؤدّي هذا الغموض إلى توليد عدم الثقة[٣].

وفي نهاية عرض الباحثة للاســتراتيجيات ســالفة الذكــر، ترى أنَّه من الممكــن تحقيق هذه الاســتراتيجيات إذا جرى العمل على قدم وساق على وجود تطبيقات للذكاء الاصطناعي قائمة على قواعد خُلُقيَّة، يكون المنتجون والمصنّعون والمصمّمون مسؤولين عنها، بحيث تكون هذه التطبيقات صديقة للإنســان ومرافقة له وخاضعة لسيطرته. وتأكيدًا لهذا الزعم، تلاحظ الباحثة

1 - Tiago Marques: Overcoming Algorithmic Bias: The role of Bias Awareness, Knowledge, and Minority Status on Human Decision-Making, p.8

2 - Erica Thompson: Artificial Intelligence has an Implicit Bias Problem, p.34

3 - David Hunt: Can Algorithms be Ethical? Algorithms are rarely designed to reflect our ethical values, but what does that mean and how can we fix it?.

أن (إليعازر يودكوفسكي- Eliezer Yudkowski)[1] قد أشار صراحة إلى ما ترغب في إبرازه، ففي كتابه المعنون بـ بناء ذكاء اصطناعي صديق للإنسان(Creating Friendly AI) أكَّد (يودكوفسكي) على أنَّ مستقبل البشريَّة مرهون بابتكار ذكاء اصطناعي صديق للإنسان من ناحية، وقادر على تنفيذ الإجراءات الودّيَّة الخُلُقيَّة كافَّة من ناحية أخرى، حتى لو أخطأ جميع المبرمجين في تنفيذها[2] . كما ترى الباحثة أنَّ هذه الاستراتيجيَّات ليست إلا جزءًا من مجموعة أوسع، وأنَّها تتميز بالتجدد والتغيير المستمرّ نتيجة ظهور تبعات خُلُقيَّة جديدة من استخدام الخوارزميَّات والذكاء الاصطناعي في مختلف مجالات الحياة، ما يفتح المجال لظهور استراتيجيَّات جديدة طالما هناك تطور تكنولوجي في واقعنا الراهن، والكلّ يسمع صداه ليل نهار.

خاتمة:

تنطلق الباحثة في هذه الخاتمة من التأكيد على أنَّ الخوارزميَّات، رغم طابعها التقني الظاهر، لا يمكن النظر إليها بوصفها أدوات محايدة أو منفصلة عن السياق الاجتماعي والثقافي الذي تُنتَج وتُوظَّف داخله. فقد بيّنت الباحثة أنَّ التحيّز الخوارزمي يتشكَّل عبر تفاعل مركّب بين البيانات المستخدمة، وآليَّات التصميم البرمجي، وأنماط الاستخدام، بما يجعل الخوارزميَّة حاملة ضمنيَّة لقيم وتصورات بشريَّة قد تُعيد إنتاج أشكال متعدّدة من التمييز وعدم المساواة.

وترى الباحثة أنَّ خطورة التحيّز الخوارزمي لا تقتصر على نتائجه العمليَّة فحسب، بل تمتدّ إلى منحه شرعيَّة زائفة تقوم على افتراض الموضوعيَّة والدقّة، الأمر الذي يحدّ من قابليَّة مساءلة القرارات الخوارزميَّة أو الاعتراض عليها. وعلى هذا النحو، تسهم الخوارزميَّات في إعادة تشكيل الواقع الاجتماعي وتكريس علاقات القوَّة داخله، خاصَّة في المجالات التي تمسّ حياة الأفراد بشكل مباشر، مثل التوظيف، والعدالة، والرعاية الصحّيَّة، وغيرها من المجالات الأخرى.

---

١ - هو باحث مؤسِّس في مجال مواءمة الذكاء الاصطناعي، وأحد المؤسِّسين المشاركين لمعهد أبحاث الذكاء الآلي، وللمزيد من التفاصيل، يمكنك مراجعة الموقع الرسمي.

٢ - مريم ساغي: الذكاء الاصطناعي ومشكلة الخصوصيَّة، ص٥٣٩.

كما تؤكّد الباحثة على أنَّ التحيّز الخوارزمي يطرح إشـكاليَّات خُلُقيَّة وفلسـفيَّة عميقة تمسّ مفاهيم العدالة، والكرامة الإنسانيَّة، والمساواة، والمسؤوليَّة، وهو ما يستدعي تجاوز المقاربات التقنيَّة الضيّقة نحو تبنّي رؤية تكامليَّة تربط بين الفلسفة، والأخلاق، والعلوم الاجتماعيَّة، وعلوم الحاسوب. ومن هذا المنطلق، تشدّد الباحثة على أهميَّة خلق بيئات تطوير أكثر تنوّعًا وشموليَّة داخـل فرق العمل، واعتماد آليات مراجعة بشريَّة مدرَّبة على كشـف التحيّز، بما يضمن فحص المدخلات والمخرجات، وكذلك البنى الداخليَّة للخوارزميَّات كلّما أمكن.

واسـتشرافًا للرؤى المسـتقبليَّة، ترى الباحثة أنَّ التحدّي الحقيقي لا يكمن في تحسين كفاءة الخوارزميَّات فحسـب، بل في إعادة توجيه مسـار تطويرها بما يخدم القيم الإنسـانيَّة الرئيسة. وكذلـك تدعو الباحثة إلى ضرورة تعميق البحث في الآثار طويلة المدى للاعتماد المتزايد على الخوارزميَّات في اتخاذ القرار، وإلى مساءلة دورها في تشكيل الوعي الاجتماعي وأنماط السلوك الفـردي والجماعي. كما تؤكّـد على ضرورة إدماج البعد الخُلُقـي في التعليم الرقمي، بما يعزّز قدرة الأفراد على فهم منطق الخوارزميَّات وعدم التعامل معها بوصفها سلطة غير قابلة للنقاش.

وفي الختـام، انتهـت الباحثة إلى أنَّ خُلُقيَّـات الخوارزميَّات تمثّل أحد أهـم ميادين التفكير الفلسـفي المعاصر؛ نظرًا لارتباطها الوثيق بمسـتقبل العلاقة بين الإنسان والتكنولوجيا. فإمَّا أن تسـهم الخوارزميَّـات في تعزيز العدالة والكرامة الإنسـانيَّة، أو تتحـوَّل – في ظلّ غياب الوعي والضوابط الخُلُقيَّة – إلى أدوات ناعمة لإعادة إنتاج التحيّز والهيمنة. ومن ثَمّ، يبقى الرهان الرئيس هو بناء وعي خُلُقي قادر على توجيه التطوّر التقني نحو خدمة الإنسان وحماية إنسانيَّته.

لائحة المصادر والمراجع:

- الحـاج عيسى بن صفي الدين، سـفيان غينو: الحجاج بواسـطة الخوارزميَّات: كيف تكبح منصات التواصل الاجتماعي الخطاب الرقمي العربي؟ مجلة روافد للدراسـات والأبحاث العلميَّة في العلوم الاجتماعيَّة والإنسانيَّة، مجلد ٩، عدد٢، ٢٠٢٥
- جقريف الزهرة: إشكاليَّة التحيّز الخوارزمي في أنظمة الذكاء الاصطناعي، وأثرها على حقوق الإنسـان: الحقّ في العمـل أنموذجًا، أبحاث الملتقى الدولي العلمـي: الذكاء الاصطناعي وتطبيقاته في العلوم الإسـلاميَّة الجزائر: مخبر الدراسـات الفقهيَّة والقضائيَّـة، كلّيَّة العلوم الإسلاميَّة، جامعة الوادي ٢٠٢٤
- خالد لراره: التحيّز الخوارزمي آليَّة رقميَّة لإدارة المعلومة، إنتاج سرديَّات بديلة وتوجيه الرأي العام، مجلة طنبة للدراسات العلميَّة الأكاديميَّة، مجلد ٨، العدد ٢، ٢٠٢٥
- عبد الوهّاب بوبعة: التحيّز الثقافي في خوارزميَّات شبكات التواصل الاجتماعي- قراءة نظريَّة في آليَّات التحيّز الرقمي والهيمنة الثقافيَّة، مجلَّة الإعلام والمجتمع، مجلد٩، عدد ٢، ٢٠٢٥
- مارك كوكلبيرغ: خُلُقيـات الذكاء الاصطناعي، ترجمة: هبة عبـد العزيز غانم، مراجعة: هبة المولى، مؤسسة هنداوي، لا م، لا ط، ٢٠٢٤
- مريم ساغي: الذكاء الاصطناعي ومشـكلة الخصوصيَّة، مجلة روافد للدراسات والأبحاث العلميَّة في العلوم الاجتماعيَّة والإنسانيَّة، مجلد ٨، عدد٢، ٢٠٢٤
- Cynthia Rudin, Caroline Wang, Beau Coker: The age of secrecy and unfairness in recidivism prediction,2019 (https://arxiv.org/abs/1811.00731)
- David Hunt: Can Algorithms be Ethical? Algorithms are rarely designed to reflect our ethical values, but what does that mean and how can we fix it? 2021 https://pubsonline.informs.org/do/10.1287/orms.2021.06.12/full/
- Emilio, Ferrara. : Fairness and Bias in artificial intelligence: a brief survey of

sources, impacts, and mitigation strategies, 2023.

- Emilio Ferrara : Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies, Sci 2024, 6, 3. 2024
- European Commission, Ethics guidelines for trustworthy AI, 2019.
- Erica Thompson: Artificial Intelligence has an Implicit Bias Problem. Utica College.2020
- Joy Buolamwini & Timnit Gebru: Gender Shades: Intersectional
- Accuracy Disparities in Commercial Gender Classification, Conference on Fairness, Accountability, and Transparency.2018.
- Kilian Vieth and Joanna Bronowicka, : Ethics of Algorithms – Why Should We Care? 2018(https://www.tbd.community/en/a/ethics-algorithms-why-should-we-care
- Marloes Susan Haitsma : Influence of Knowledge and trust on the perception of algorithmic gender bias in Artificial Intelligence. Lisbon School of Economics & Management, 2024
- Michael O'Flaherty: Bias in Algorithms – Artificial Intelligence and Discrimination, European Union Agency for Fundamental Rights.2022
- Robin K. Hill :What an Algorithm Is, Philosophy & Technology,Springer, 29, no. 1, 2016.
- Tiago Marques : Overcoming Algorithmic Bias: The role of Bias Awareness, Knowledge, and Minority Status on Human Decision-Making, Universidade Católica Portuguesa.2021
- Tommy, Fred : Bias and Fairness in AI Algorithms. University of London. ResearchGate. 2025 https://www.researchgate.net/publication/390057263_

Bias_and_Fairness_in_AI_Algorithms

- Virginia Dignum: Artificial Intelligence: Foundations, Theory, And Algorithms. Springer. 2019