

أخلاقيات الذكاء الاصطناعي: الحاجة إلى مبادئ توجيهية لضبط التداعيات وتعزيز الموثوقية "المبادئ التوجيهية للهيئة السعودية سدايا SADAIA نموذجاً"

Artificial Intelligence Ethics: The Need for Guidelines to Control onsequences and Promote Trust-SADAIA guidelines as a model-

أحمد بن سعدة^{1*}، عبد القادر بن سعدة²

¹جامعة عمار ثليجي الأغواط (الجزائر)، a.bensada@lagh-univ.dz

²جامعة العربي التبسي تبسة (الجزائر)، abdelkaderbensaade@gmail.com

تاريخ النشر: 2025/06/30

تاريخ القبول: 2025/06/23

تاريخ الاستلام: 2025/03/13

ملخص:

تهدف هذه الدراسة إلى استكشاف الجوانب الأخلاقية لمفهوم الذكاء الاصطناعي، مع التركيز على تداعياته الأخلاقية التي أوجبت وضع مبادئ توجيهية تهدف إلى تعزيز الثقة في هذه التقنية واستخداماتها، وكذلك تقليل المخاطر المحتملة في حال حدوثها، هذه التداعيات دفعت بالعديد من المنظمات العالمية والمؤسسات البحثية على غرار الهيئة السعودية للبيانات والذكاء الاصطناعي سدايا SADAIA إلى تحديد أهم المبادئ والممارسات التي تضمن مواجهة هذه التحديات وضبط استخدامات الذكاء الاصطناعي. خلصت الدراسة إلى أن الاهتمام بأخلاقيات الذكاء الاصطناعي أصبح أمراً ضرورياً بسبب الأضرار الفردية والمجتمعية الناتجة عن سوء استخدام أو تصميم هذه التقنية، مما دفع العديد من التنظيمات الحكومية وغير الحكومية إلى اقرار مجموعة من المبادئ العملية التي تركز على التنمية المستدامة والرفاهية المجتمعية، وحماية القيم الإنسانية والعدالة، وتكريس الشفافية وقابلية التفسير، إضافة إلى توفير الأمان وضمان المسؤولية وحق المساءلة، كل هذا دعماً للممارسات الأخلاقية المقبولة وإنتاج تطبيقات ذكاء اصطناعي تتسم بالقيم بالأخلاقيات.

الكلمات المفتاحية : ذكاء اصطناعي؛ أخلاقيات؛ أخلاقيات ذكاء الاصطناعي؛ مبادئ توجيهية؛ سدايا.

تصنيف JEL: O31؛ O32؛ C88.

Abstract:

This study aims to explore the ethical aspects of the concept of artificial intelligence, focusing on its ethical implications, which have necessitated the development of guidelines aimed at enhancing confidence in this technology and its uses, as well as mitigating potential risks should they occur. These implications have prompted numerous international organizations and research institutions, such as the Saudi Data and Artificial Intelligence Authority (SADAIA), to identify the most important principles and practices that ensure these challenges are addressed and the uses of artificial intelligence are regulated.

The study concluded that attention to AI ethics has become imperative due to the individual and societal harms resulting from the misuse or misdesign of this technology. This has prompted many governmental and non-governmental organizations to adopt a set of practical principles that focus on sustainable development and societal well-being, protecting human values and justice, promoting transparency and interpretability, as well as providing safety and ensuring responsibility and accountability. All of this supports acceptable ethical practices and the production of ethical AI applications.

Keywords: Artificial intelligence; ethics; AI ethics; guidelines; SADAIA.

Jel Classification Codes : O31; O32; C88.

*المؤلف المرسل.

مقدمة:

يعد الحديث عن أخلاقيات الذكاء الاصطناعي من أكثر المواضيع جدية وحيوية في مجال التطور التكنولوجي، فظهور الآلات الذكية التي قد تضاهي أو تتفوق في كثير من الأحيان على القدرات البشرية يفرض ضرورة الاستعداد لمستقبل الذكاء الاصطناعي غير القابل للتنبؤ، ونظراً للنمو المتسارع الذي تشهده الممارسات والتقنيات المتعلقة بالذكاء الاصطناعي وتنوع استخدام هذه التقنيات في تحليل البيانات الضخمة من خلال إنشاء أنظمة تعمل على نماذج وخوارزميات ذات تطور عالٍ، فقد زادت احتمالية ظهور تهديدات ومخاوف أخلاقية تصاحب هذا التطور، لذا سارعت العديد من الجهات الحكومية والغير حكومية بتطوير وتبني العديد من المبادئ التوجيهية الأخلاقية وهياكل حوكمة واضحة لنشر الذكاء الاصطناعي في بيئة تتسم بالأمان والموثوقية، كما تضمن تضييق هامش المخاطر في حال حدوثها.

● إشكالية الدراسة:

من أجل الامام والاحاطة بأكبر قدر ممكن من جوانب موضوع الدراسة، وفق منهجية علمية تقود إلى الحصول على النتائج المطلوبة يمكن من طرح الإشكالية التالية:

- "إلى أي مدى تساهم المبادئ التوجيهية في الحد من التداعيات الأخلاقية للذكاء الاصطناعي وتعزيز الأمان والموثوقية؟".
- وهذا ما يدفعنا لطرح الأسئلة الفرعية التالية:
- ما المقصود بأخلاقيات الذكاء الاصطناعي؟
- ما أهم المساهمات في وضع المبادئ التوجيهية للقضايا الأخلاقية في الذكاء الاصطناعي؟
- ما هي المبادئ التوجيهية التي أقرتها الهيئة السعودية سدايا SADAIA لمواكبة القضايا الأخلاقية للذكاء الاصطناعي؟

● فرضيات الدراسة:

- توجد علاقة إيجابية بين تبني مبادئ أخلاقيات الذكاء الاصطناعي ومستويات الأمان والموثوقية في استخداماته.
- يؤدي غياب الأطر الأخلاقية الواضحة إلى زيادة المخاطر المرتبطة باستخدام تقنيات الذكاء الاصطناعي.
- يساهم وجود هيئات تنظيمية متخصصة (مثل الهيئة السعودية سدايا SADAIA) في تعزيز الحوكمة الرشيدة لتقنيات الذكاء الاصطناعي.
- الوعي المجتمعي والسياسات الحكومية يعتبران عاملين حاسمين في تطوير أطر أخلاقية فعالة للذكاء الاصطناعي.

● أهمية الدراسة:

تستمد الدراسة أهميتها من كونها خطوة إضافية على مسار البحث في مجال الذكاء الاصطناعي وتأكيداً على ضرورة أن تكون مخرجات هذا الأخير عادلة وآمنة وموثوقة، كما أن هذه الدراسة جاءت كضرورة فرضتها التحديات والقضايا الأخلاقية التي يمكن أن تؤدي إلى انخفاض مستوى الثقة في أنظمة الذكاء الاصطناعي، زيادة على ذلك ضرورة التعريف بالجهود التي تبذلها مختلف الحكومات والمؤسسات في وضع أطر قانونية وتنظيمية لتوجيه وتطوير استخدامات الذكاء الاصطناعي بطريقة مسؤولة وأخلاقية.

● أهداف الدراسة:

هدفت الدراسة إلى إعطاء إطار نظري موثوق حول أخلاقيات الذكاء الاصطناعي، ثم دعم هذا الإطار بمجموعة من المساهمات العملية من طرف الحكومات والمؤسسات في وضع أسس ومبادئ توجيهية لممارسات الذكاء الاصطناعي على خلفية التصاعد الكبير في التحديات الأخلاقية لهذا المفهوم، وفي إطار تحقيق هذا الهدف تسعى الدراسة إلى تحقيق الأهداف الفرعية التالية:

- التعريف بالذكاء الاصطناعي؛
- تحليل القضايا الأخلاقية للذكاء الاصطناعي ومحاولة تصنيفها وفق عدة أبعاد؛

- التطرق لأهم الاسهامات العملية حول المبادئ التوجيهية لأخلاقية الذكاء الاصطناعي؛
- التعرف على جهود الهيئة السعودية سدايا SADAIA في اقرار مبادئ توجيهية للقضايا الأخلاقية للذكاء الاصطناعي.

● منهج الدراسة:

تعتمد هذه الدراسة على المنهج الوصفي التحليلي، باعتباره الأنسب لمعالجة الإشكالية المطروحة المتعلقة بأخلاقيات الذكاء الاصطناعي والمبادئ التوجيهية ذات الصلة، ويقوم هذا المنهج على جمع وتحليل المعطيات النظرية المرتبطة بالذكاء الاصطناعي ومخاطره، وأبعاده الأخلاقية، وتصنيفات أخلاقياته، من خلال تحليل الأدبيات العلمية والوثائق المرجعية، إضافة إلى مقاربات المؤسسات الدولية والوطنية التي وضعت مبادئ توجيهية لأخلاقيات الذكاء الاصطناعي.

كما تم التركيز بشكل خاص على تحليل نموذج الهيئة السعودية للبيانات والذكاء الاصطناعي (SDAIA)، عبر التطرق مضامين المبادئ الأخلاقية التي تقترحها، ودراسة مدى تكاملها مع التوجهات العالمية في هذا المجال، وبهذا، تُمكن الدراسة من الوقوف على الفجوات النظرية والتطبيقية في تأطير القضايا الأخلاقية لتقنيات الذكاء الاصطناعي.

● الدراسات السابقة:

- دراسة (جمال بوغام، 2024) بعنوان: "الذكاء الاصطناعي وأخلاقياته"، تهدف هذه الدراسة إلى تحليل قضية الذكاء الاصطناعي والاجابة عن التساؤل حول الحدود الأخلاقية ومحاولة مناقشة المفارقة بين قيمة الذكاء الاصطناعي المعرفية والتقنية والمخاوف والتداعيات التي تخلفها تطبيقاته، اعتمد الباحث على المنهج الوصفي التحليلي وتوصل إلى أن الذكاء الاصطناعي يعد فجرا جديدا في مجال التكنولوجيا، ولكن لا ينبغي التهاون بمخاطره التي تشكل تحديا في العديد من الجوانب، كما أن القضايا الأخلاقية تنصدر هذه المخاطر وتطرح التساؤل حول كيفية برجمة القيم في الذكاء الاصطناعي، الأمر الذي يفرض على حكومات ومؤسسات التشريع والتنظيم التحكم في استخدام الذكاء الاصطناعي وتطبيق قوانين حماية البيانات وتعزيز أمان للمستخدمين.

- دراسة (2023,Changwu Huang et al) بعنوان: "An Overview of Artificial Intelligence Ethics" مقال مجلة Ieee Transactions Onartificial Intelligence، المجلد (04) العدد (04)،هدفت هذه الدراسة إلى إلقاء نظرة شاملة على مجال الذكاء الاصطناعي من خلال تلخيص وتحليل المخاطر الأخلاقية التي أثارها الذكاء الاصطناعي، والتطرق للمبادئ التوجيهية الأخلاقية والمبادئ الصادرة عن المنظمات المختلفة، إضافة إلى مناقشة أساليب معالجة القضايا الأخلاقية في الذكاء الاصطناعي، وطرق تقييم أخلاقياته، اعتمد الباحثون على المنهج الوصفي التحليلي للتأصيل المفاهيمي لموضوع الدراسة ومراجعة وتصنيفمجموعة من المبادئ التوجيهية من مصادر مختلفة، وخلصت الدراسة إلى أن محاولة معالجة القضايا الأخلاقية في الذكاء الاصطناعي وجعله يتصرف بشكل أخلاقي تتسم بالصعوبة والتعقيد، كما أن هذه المهمة تتطلب جهدا مشتركا من علماء الذكاء الاصطناعي والمهندسين والفلاسفة والمستخدمين وصانعي السياسات الحكومية.

- دراسة (2022,Kimon Kieslich et al) بعنوان: "Artificial intelligence ethics by design. Evaluating public perception on the importance of ethical design principles of artificial intelligence"، هدفت هذه الورقة البحثية إلىمناقشة مبادئ أخلاقيات الذكاء الاصطناعي واجراء مقارنة بينها، إضافة إلى تقديم إجابات أولية حول الأهمية النسبية لهذه المبادئ الأخلاقية بالنظر إلى حالة استخدام الذكاء الاصطناعي في الكشف عن الاحتيال الضريبي، اعتمد الباحثون على المنهج الوصفي التحليلي لمسح آراء (1099) من المستخدمين الألمان، لتتوصل الدراسة إلى المبحوثين يقيمون المبادئ الأخلاقية على أنها بنفس القدر من الأهمية، كما أن جزءا ملحوظا من الشعب الألماني لا يطالب بتطبيق الذكاء الاصطناعي الأخلاقي، إذا أن هذا النقص في المطالبة بالمتطلبات الأخلاقية يعرقل تنفيذ المبادئ الأخلاقية ويفرض ضرورة زيادة الاهتمامبالذكاء الاصطناعي الأخلاقي.

- دراسة (2020, Keng SIAU & Weiyu Wang) بعنوان: "Artificial intelligence (AI) ethics: Ethics of AI and ethical AI"، هدفت هذه الورقة البحثية إلى معالجة إحدى القضايا الحرجة والمتعلقة بكيفية معالجة التحديات الأخلاقية والمعنوية المرتبطة بالذكاء الاصطناعي، اعتمد الباحثان على المنهج الوصفي التحليلي وانتهت الدراسة إلى أن أخلاقيات الذكاء الاصطناعي تدرس المبادئ الأخلاقية والقواعد والمبادئ التوجيهية والسياسات واللوائح المتعلقة بالذكاء الاصطناعي، كما أن الذكاء الاصطناعي الأخلاقي هو الذكاء الاصطناعي الذي يؤدي ويتصرف بشكل أخلاقي، لذا يجب على المرء أن يدرك ويفهم القضايا الأخلاقية والمعنوية المحتملة التي قد تسببها الذكاء الاصطناعي لصياغة المبادئ الأخلاقية والقواعد والمبادئ التوجيهية والسياسات واللوائح اللازمة للذكاء الاصطناعي (أي أخلاقيات الذكاء الاصطناعي).

● مناقشة الدراسات السابقة:

- بعد استعراض الدراسات السابقة يظهر جليا بأن ممارسات الذكاء الاصطناعي بإمكانها أن تخلف أخطارا وتداعيات على العلاقات الإنسانية، هذه الحقيقة أدت إلى تزايد الاهتمام حول ضرورة البحث عن قواعد أخلاقية عملية تؤطر الممارسات العملية للذكاء الاصطناعي في جميع المجالات، كما أن هذه الدراسات كان هدفها تجميع وتوحيد جهود الحكومات والمؤسسات التي بادرت إلى تأسيس مبادئ توجيهية حول القضايا الأخلاقية للذكاء الاصطناعي، وهو ذات الهدف الذي قامت من أجله هذه الدراسة، من خلال التأصيل النظري لهذا المفهوم والتطرق إلى العديد من الاسهامات التي اهتمت بتطوير المبادئ التوجيهية التي تضمن النشر الآمن للذكاء الاصطناعي والتحكم في تداعياته. وفيما يمكن تمييز الدراسة الحالية عن الدراسات السابقة المقترحة من الأوجه التالية:
- تلتقي الدراسات السابقة مع الدراسة الحالية في التطرق لموضوع الذكاء الاصطناعي من جانب التداعيات الأخلاقية التي تصاحب الانتشار السريع لتقنياته وممارساته؛
 - تتفق الدراسات السابقة التي تناولت ريادة الأعمال الاستراتيجية مع دراستنا في ضرورة التأسيس لمبادئ توجيهية تجسد القيم الأخلاقية في استعمالات الذكاء الاصطناعي وترفع من درجة الثقة والأمان في تقنياته؛
 - النصيب الأكبر من الدراسات السابقة حول أخلاقيات الذكاء الاصطناعي كان باللغة الإنجليزية وما كان منها باللغة العربية فهو يتناول هذا المفهوم من جانب التأصيل النظري أكثر من التوسع في المبادئ التوجيهية للقضايا الأخلاقية، في حين حاولت هذه الدراسة العربية الموازنة بين المطلبين معا.
 - تختلف الدراسة الحالية عن الدراسات السابقة في الحدود المكانية، حيث تتناول هذه الدراسة اسهامات الهيئة السعودية سدايا SADAIA في وضع مبادئ توجيهية للقضايا الأخلاقية للذكاء الاصطناعي، في حين كانت الدراسات السابقة تتناول مبادئ توجيهية على مستوى دول أخرى على غرار أستراليا وألمانيا.

المحور الأول: إطار مفاهيمي حول أخلاقيات الذكاء الاصطناعي:

يتناول هذا المحور مدخلا تأسيسيا لفهم الذكاء الاصطناعي من حيث المفهوم والنشأة والتطور، كما يستعرض أبرز مخاطره وتحدياته في السياقات المعاصرة. ويناقش المحور الأبعاد الأخلاقية المرتبطة باستخدام الذكاء الاصطناعي، مع تصنيف القضايا الأخلاقية التي يثيرها، بالإضافة إلى عرض الأدوات والمعايير المستخدمة في تقييم مدى التزام أنظمة الذكاء الاصطناعي بالمبادئ الأخلاقية.

أولا: مدخل للذكاء الاصطناعي:

بالرجوع إلى تاريخ مصطلح "الذكاء الاصطناعي" فقد تمت صياغة المصطلح في عام 1955 من قبل مجموعة من الباحثين - John McCarthy, Marvin L. Minsky, Nathaniel Rochester and Claude E. Shannon - والذين نظموا ورشة عمل صيفية شهيرة لمدة شهرين في كلية Dartmouth حول "دراسة الذكاء الاصطناعي"، حيث يُعترف بهذا الحدث على نطاق واسع باعتباره البداية لدراسة الذكاء الاصطناعي. (Stewart & Nyholm, 2021, p. 02)

أما عن التعريف العلمي للمصطلح فيعرف قاموس أكسفورد الحي الذكاء الاصطناعي Artificial Intelligence على أنه "نظرية تطوير أنظمة الكمبيوتر القادرة على أداء المهام التي تتطلب عادةً ذكاءً بشرياً، مثل الإدراك البصري والتعرف على الكلام واتخاذ القرار والترجمة بين اللغات" (Alicia & Albert, 2021, p. 13). كما يعرف Philip Jansen الذكاء الاصطناعي على أنه "علم وهندسة الآلات ذات القدرات التي تعتبر ذكية وفقاً لمعايير الذكاء البشري" (كوكليبرج، 2024، صفحة 51).

وجاء في تقرير للصحة العالمية بقلم عالمة Soumya Swaminathan أن مصطلح الذكاء الاصطناعي أو ما يختصر له بـ AI يشير إلى قدرة الخوارزميات المشفرة في التكنولوجيا على التعلم من البيانات حتى تتمكن من أداء مهام آلية دون الحاجة إلى برمجة كل خطوة في العملية بشكل صريح من قبل الإنسان (Swaminathan, 2021)، كما يرى Müller Vincent في دراسته بأن الذكاء الاصطناعي يفهم على أنه أي نوع من أنظمة الحوسبة الاصطناعية التي تُظهر سلوكاً ذكياً، أي سلوكاً معقداً يساعد على تحقيق الأهداف (Müller, 2021, p. 03)، كما يشمل الأنظمة التي تظهر سلوكاً ذكياً من خلال تحليل بيئتها واتخاذ الإجراءات - بدرجة معينة من الاستقلالية- لتحقيق أهداف محددة، بحيث يمكن أن تكون أنظمة الذكاء الاصطناعي قائمة على البرامج البحثية، وتعمل في العالم الافتراضي (مثل المساعدين الصوتيين وبرامج تحليل الصور ومحركات البحث وأنظمة التعرف على الكلام والوجوه) أو يمكن تضمين الذكاء الاصطناعي في الأجهزة (مثل الروبوتات المتقدمة أو السيارات ذاتية القيادة أو الطائرات بدون طيار أو تطبيقات إنترنت الأشياء) (Wynsberghe, 2020, p. 05).

يمكن تعريف الذكاء الاصطناعي على أنه الأنظمة والبرامج التي لها القدرة على محاكاة الذكاء البشري، بما في ذلك القدرة على التفكير والتعلم والتحليل وحل المشكلات، فالذكاء الاصطناعي يعتمد على تقنيات وخوارزميات متنوعة مثل التعلم الآلي، والشبكات العصبية الاصطناعية ومعالجة اللغات الطبيعية، مما يمنحه الاستقلالية في اتخاذ القرارات وفق قواعد البيانات المتاحة له.

ثانياً: مخاطر الذكاء الاصطناعي:

يلخص David Leslie أكثر أشكال مخاطر الذكاء الاصطناعي أهمية في العناصر التالية: (Leslie, 2019, p. 04)

- التحيز والتمييز: يمكن للتقنيات القائمة على البيانات إعادة إنتاج وتعزيز وتضخيم أنماط التهميش وعدم المساواة والتمييز الموجودة في هذه المجتمعات؛
- الحرمان من الاستقلالية الفردية وسبل الانصاف والحقوق: عند الخضوع لقرارات أو تنبؤات أو تصنيفات تنتجها أنظمة الذكاء الاصطناعي، فقد تنشأ مواقف يكون فيها الأفراد غير قادرين على مساءلة الأطراف المسؤولة بشكل مباشر عن هذه النتائج؛
- نتائج غير شفافة أو غير قابلة للتفسير أو غير مبررة: تولد العديد من نماذج التعلم الآلي نتائجاً خوارزمية من خلال العمل على ارتباطات عالية الأبعاد تتجاوز القدرات التفسيرية لتفكير الإنسان؛
- انتهاك الخصوصية: تشكل أنظمة الذكاء الاصطناعي تهديدات للخصوصية نتيجة لعمليات التصميم والتطوير الخاصة بها، وذلك لارتكازها على هيكلية البيانات ومعالجتها، وفي كثير من الأحيان استخدام البيانات الشخصية دون الحصول على الموافقة المناسبة من مصدرها أو يتم التعامل معها بطريقة تضر بأصحابها؛
- نتائج غير موثوقة أو غير آمنة أو رديئة الجودة: تؤدي الإدارة غير المسؤولة للبيانات، وعمليات التصميم والإنتاج المهملة وممارسات النشر المشكوك فيها، إلى توزيع أنظمة الذكاء الاصطناعي التي تنتج نتائج غير موثوقة أو غير آمنة أو رديئة الجودة؛
- عزلة وتفكك التواصل الاجتماعي: تُبنى المجتمعات المنظمة والمتناسكة على علاقات الثقة والتعاطف والتفاهم المتبادل، ومع انتشار تقنيات الذكاء الاصطناعي التي تحقق الاستقلالية بشكل متزايد، نكون في تحد كبير للحفاظ على مثل هذه العلاقات.

ثالثاً: تعريف أخلاقيات الذكاء الاصطناعي:

بداية لا بد من الحديث أولاً عن الرؤى التي يمتلكها البشر حول الأخلاق، وننطلق مما كتبه Velasquez وزملاؤه سنة (1987) عن التجربة التي سأل فيها عالم الاجتماع Raymond Baumhart بعض رجال الأعمال "ماذا تعني الأخلاق بالنسبة لك؟" وكان من بين هذه الأجوبة المعتقدات الثلاثة التالية: (Stefan, 2020, p. 143)

- "إن الأخلاق تتعلق بما تخبرني به مشاعري أنه صواب أو خطأ".
- "إن كونك أخلاقياً يعني القيام بما يتطلبه القانون".
- "تتكون الأخلاق من معايير السلوك التي يقبلها مجتمعنا".

من خلال هذه التجربة نلاحظ مدى الاختلاف في الإجابات المتعلقة بالأخلاق، فمن هذه الأجوبة ما يشير إلى جوانب ذاتية مثل المشاعر والمعتقدات، ومنها ما يشير إلى القواعد والسلوكيات والقوانين، أما من ناحية ارتباطها بالذكاء الاصطناعي فيرى Stefan Trausan-Matu بأن الإجابات من 1 إلى 3 ذات صلة إلى حد ما ببرامج الذكاء الاصطناعي التي تقرر ما إذا كان إجراء معين يتوافق مع بعض قواعد الأخلاق، ومع ذلك يبقى هناك اختلافات جوهرية بين هذه الحالات من حيث التعقيد وإمكانيات التنفيذ.

فالحديث عن أخلاقيات الذكاء الاصطناعي يقود بشكل أساسي إلى بعدين: يتعلق البعد الأول بالسلوك الأخلاقي للأشخاص في تصميم وتصنيع واستخدام أنظمة الذكاء الاصطناعي (أخلاقيات التكنولوجيا)، في حين يتعلق البعد الثاني بسلوك أنظمة الذكاء الاصطناعي (أخلاقيات الآلة)، فالتفكير في البعد الأول يعكس حقيقة أن الآلة المصممة من قبل أشخاص لمحاولة "تقليد" الذكاء البشري يمكن أن يبعث بالفعل عدداً من المشكلات مثل الخداع والتحيزات والأخطاء المعرفية، والتي من الممكن أن تظهر بشكل منهجي إذا تم بناء الذكاء الاصطناعي ليحاكي ويشبه البشر. (Alicia & Albert, 2021, p. 18)، ويختصر David Leslie في مفهوم أخلاقيات الذكاء الاصطناعي في أنها "مجموعة من القيم والمبادئ والتقنيات التي تستخدم معايير مقبولة على نطاق واسع للصواب والخطأ لتوجيه السلوك الأخلاقي في تطوير واستخدام تقنيات الذكاء الاصطناعي". (Leslie, 2019, p. 03)

وجاء في تقرير عن شركة IBM المعد من طرف Francesca Rossi الرائد العالمي في أخلاقيات الذكاء الاصطناعي وآخرون بأن "الذكاء الاصطناعي يوفر رؤية أعمق لحياتنا الشخصية عند التفاعل مع بياناتنا الحساسة" ونظراً لأن البشر معرضون بطبيعتهم للقضايا الأخلاقية، وكونهم مسؤولون عن بناء أنظمة الذكاء الاصطناعي فهناك فرص لتضمين هذه القضايا الأخلاقية البشرية في الأنظمة التي يقومون بإنشائها (Francesca et al, 2022, p. 28)، ويرى Kieslich et al بأن الذكاء الاصطناعي لديه قدرة هائلة لتغيير المجتمع، وبالرغم من أن التنفيذ الواسع النطاق لأنظمة الذكاء الاصطناعي يمكن أن يولد بالتأكيد أرباحاً اقتصادية، إلا أن صناع السياسات والعلماء يسلطون الضوء أيضاً على التحديات الأخلاقية المصاحبة للذكاء الاصطناعي فمعظم العلماء والسياسيين والمطورين يتفقون على أن الذكاء الاصطناعي يحتاج إلى التطوير بطريقة تركز على الإنسان وجديرة بالثقة، من أجل الوصول إلى الذكاء الاصطناعي الذي يعود بالنفع على الصالح العام. (Kieslich et al, 2022, p. 01)، وتناول دليل "أخلاقيات الذكاء الاصطناعي" السعودي أخلاقيات الذكاء الاصطناعي على أنها "مجموعة من القيم والمبادئ والأساليب لتوجيه السلوك الأخلاقي في تطوير تقنيات الذكاء الاصطناعي واستخدامها" (SDAIA, 2024, p. 08).

وعموماً يمكن القول بأن أخلاقيات الذكاء الاصطناعي هي مجموعة المبادئ والمعايير التي تحكم سلوك الذكاء الاصطناعي وتوجهه نحو القيم الإنسانية، فأخلاقيات الذكاء الاصطناعي تساعد على مواجهة تحديات تطوير الذكاء الاصطناعي وضمان استخدامه بطرق تفيد المجتمع، كما أن هذه الأخلاقيات تشمل العديد من القيم مثل العدالة والشفافية والمساءلة وحماية الخصوصية والأمن من الخداع والتمييز وجميع التأثيرات المجتمعية المحتملة.

رابعا: أبعاد أخلاقيات الذكاء الاصطناعي:

لأخلاقيات الذكاء الاصطناعي أربعة أبعاد رئيسية يجب مراعاتها منذ المراحل الأولى لتبني الحلول القائمة على الذكاء الاصطناعي أو تطويرها أو استخدامها، بهدف معرفة أين تنشأ القضايا والتحديات المختلفة المتعلقة بالأخلاقيات وكيفية التعامل معها أو تجنبها مستقبلا:

الشكل رقم (01): أبعاد أخلاقيات الذكاء الاصطناعي



المصدر: (SDAIA, 2024, p. 09)

خامسا: تصنيفات القضايا الأخلاقية للذكاء الاصطناعي:

في دراسة لـ (Changwu Huang et al, 2023) تم مناقشة أهم القضايا التي لها علاقة بأخلاقيات الذكاء الاصطناعي، وبالرجوع إلى العديد من الاسهامات العلمية كالتقارير الحكومية والمنشورات الأكاديمية قام الباحثون بتصنيف القضايا الأخلاقية على أساس العديد من الأوجه المختلفة، ونقوم بدورنا بتلخيص أهم هذه الأوجه في الشكل التالي والتعليق عليها:

الشكل رقم (02) : تصنيفات القضايا الأخلاقية للذكاء الاصطناعي



المصدر: من اعداد الباحثين بالاعتماد على المصادر السابقة

1. التصنيف على أساس ميزات الذكاء الاصطناعي والعوامل البشرية والتأثير الاجتماعي:

في الذكاء الاصطناعي تتم مناقشة القضايا الأخلاقية من هذا المنظور بشكل رئيسي في ثلاث فئات:

- القضايا الأخلاقية الناجمة عن ميزات الذكاء الاصطناعي: مثل الافتقار إلى الشفافية الناجم عن غموض التعلم الآلي والذي يجعل الخوارزميات أو النماذج غامضة للمستخدمين وحتى المطورين ومشكلة تسرب المعلومات (الشخصية) أو العبث بها أيضا من

القضايا الأخلاقية الخطيرة التي تهدد أمن وخصوصية بيانات الأفراد والمنظمات، إضافة إلى مشكلة مقدار الحيز الذي يمكن أن يعطى لنظام الذكاء الاصطناعي في الاستقلالية وعدم تدخل الانسان، وقصدية الذكاء الاصطناعي في تحديد ما هو ضار وما هو مفيد، والمسؤولية تجاه الوفاء بالقواعد الاجتماعية.

- **القضايا الأخلاقية الناجمة عن العوامل البشرية:** مثل مشكلة المساءلة عن المسؤول عند فشل نظام الذكاء الاصطناعي في بعض المهام وهي مشكلة ترجع أساسا إلى العوامل البشرية المشاركة في تصميم ونشر الذكاء الاصطناعي، إضافة إلى مدى قدرة العامل البشري على انشاء معايير أخلاقية شاملة وغير متحيزة تضمن أخلاقية الذكاء الاصطناعي، كما أنه من الضروري على كل من له يد في تصميم وبرمجة وتطبيق نظام الذكاء الاصطناعي أن يكون على دراية بقوانين حقوق الإنسان وضمان عدم انتهاكها.
- **التأثير الاجتماعي للقضايا الأخلاقية للذكاء الاصطناعي:** مثل مشكل الأتمتة التي تقوم على استبدال العمال بالأنظمة الآلية والروبوتات والتي تمثل قلقا كبيرا للمجتمع، إضافة إلى مشكلة مدى توفر التقنيات الناشئة لكل أطراف المجتمع وسهولة الوصول إليها، كما أن إهمال ضرورة الحفاظ على الديمقراطية وضمان الحقوق المدنية وعدم التحيز يعد أيضا من المشاكل الكبيرة ذات البعد الاجتماعي للذكاء الاصطناعي.

2. التصنيف على أساس نقاط ضعف الذكاء الاصطناعي والبشر:

أما من منظور نقاط الضعف فيتم مناقشة القضايا الأخلاقية من جانبين رئيسيين هما:

- **القضايا الأخلاقية من نقاط الضعف في الذكاء الاصطناعي:** والتي من بينها حاجة الذكاء الاصطناعي لأحجام كبيرة من البيانات قد تدفع بالمستخدمين إلى التعدي عن البيانات الشخصية للآخرين، كما أن مشكلة العمل بالخوارزميات الخاطئة يقود حتما إلى قصورها وتقدم نتائج وتنبؤات سيئة للمستخدم، بالإضافة إلى أن التعلم العميق الذي يعد صندوق أسود في الذكاء الاصطناعي يطرح مشكلة عدم قابلية التفسير والثقة في نتائجه.
- **القضايا الأخلاقية من نقاط الضعف البشرية:** إن المشكلة الأكبر في هذا الجانب هي إساءة الأشخاص لاستخدام تقنيات الذكاء الاصطناعي التي لها قدرة كبير على تجسيد وفبركة وقائع وحقائق غير موجودة، كما أن قضية الصداقة التي يمكن ان تنشأ بينا لانسان والروبوت قد تكون لها العديد من الآثار السلبية على الأفراد.

3. التصنيف على أساس الخوارزمية والبيانات والتطبيق:

من هذا المنظور تم تصنيف القضايا الأخلاقية وفق ثلاثة أبعاد كما يلي:

- **القضايا الأخلاقية المتعلقة بالخوارزميات:** والتي يقع في مقدمتها مشكل أمن الخوارزميات ومدى قدرة المصممين على الحفاظ عليها وعدم تسربها أو اكتشاف الثغرات في تطبيقها والذي يؤدي إلى أضرار مباشرة على الأفراد، إضافة إلى عدم قابلية هذه الخوارزمية للتفسير يحد من قدرة الانسان على معرفة سبب العديد من القرارات التي يصدرها الذكاء الاصطناعي أو التنبؤ بها.
- **القضايا الأخلاقية ذات الصلة بالبيانات:** من بين أكبر المشاكل في هذا الجانب نجد مشكلة حماية الخصوصية والحفاظ على البيانات الشخصية للمستخدمين، ففي الكثير من الحالات يتم إلغاء تحديد هوية المعلومات الشخصية بطريقة ما وبالتالي فلن يتم اعتبارها معلومات شخصية أو حساسة وغير محمية ويسهل عملية نقلها والتصرف بها وهذا ما يعد من بين أكبر القضايا الأخلاقية للذكاء الاصطناعي.

- القضايا الأخلاقية المتعلقة بالتطبيق: حيث يمكن أن تنشأ العديد من المشاكل المتعلقة بتطبيق الذكاء الاصطناعي كالتمييز الخوارزمي الذي قد يدعم العنصرية أو يفرق بين الذكور والإناث في عمليات معينة، إضافة إلى إساءة استخدام الخوارزمية في مواقف معينة أو اتخاذ قرارات محددة.

سادساً: أدوات تقييم الذكاء الاصطناعي الأخلاقي:

باعتبار الذكاء الاصطناعي مجموعة واسعة من التقنيات التي لها أيضاً مجموعة من الآثار القانونية والأخلاقية، كما أنه لا يوجد حل واحد يناسب جميع القضايا الأخلاقية الناشئة، ومع ذلك فهناك أدوات يمكن استخدامها لتقييم المخاطر وضمان الامتثال والرقابة حيث يمكن اختيار أنسب الأدوات لكل ظرف على حدة، ونوضح هذه الأدوات في الجدول التالي:

الجدول رقم (01): أدوات تقييم الذكاء الاصطناعي الأخلاقي

1. تقييم الأثر: تقييمات قابلة للتدقيق للتأثيرات المباشرة وغير المباشرة المحتملة للذكاء الاصطناعي، والتي تعالج التأثيرات السلبية المحتملة على الأفراد والمجموعات والمجموعات إلى جانب إجراءات التخفيف.	2. المراجعة الداخلية أو الخارجية: استخدام متخصصين لمراجعة الذكاء الاصطناعي و/أو استخدام أنظمة الذكاء الاصطناعي للتأكد من أنها تلتزم بالمبادئ الأخلاقية والسياسات والتشريعات.	3. تقييم المخاطر: استخدام تقييمات المخاطر لتصنيف مستوى المخاطر المرتبطة بتطوير و/أو استخدام الذكاء الاصطناعي.
4. إرشادات أفضل الممارسات: تطوير مبادئ أفضل الممارسات المتاحة للمساعدة في توجيه المطورين ومستخدمي الذكاء الاصطناعي إلى الممارسات القياسية الذهبية.	5. معايير الصناعة: توفير الأدلة التعليمية وبرامج التدريب وربما الشهادات للمساعدة في تنفيذ المعايير الأخلاقية في استخدام الذكاء الاصطناعي وتطويره.	6. التعاون: برامج تعمل على تعزيز وتحفيز التعاون بين الصناعة والأوساط الأكاديمية في تطوير الذكاء الاصطناعي "الأخلاقي" من حيث التصميم.
7. آليات المراقبة والتحسين: المراقبة المنتظمة للذكاء الاصطناعي للتأكد من دقته وإنصافه وملاءمته للمهمة المطروحة، وينبغي أن يشمل هذا أيضاً النظر في ما إذا كانت الأهداف الأصلية للخوارزمية لا تزال ذات صلة.	8. آليات الاستئناف: سبل الاستئناف عندما يؤثر قرار آلي أو استخدام خوارزمية سلباً على أحد أفراد الجمهور.	9. الاستشارة: استخدام الاستشارة العامة أو المتخصصة لإعطاء الفرصة لأصحاب المصلحة الرئيسيين لمناقشة القضايا الأخلاقية المتعلقة بالذكاء الاصطناعي.

المصدر: (Dawson et al, 2019, p. 08)

المحور الثاني: المبادئ التوجيهية لأخلاقيات الذكاء الاصطناعي: المبادرات والتحديات

يركز هذا المحور على المبادئ التوجيهية التي تشكل الإطار الأخلاقي لاستخدام وتطوير تقنيات الذكاء الاصطناعي، من خلال استعراض أبرز المبادئ المعتمدة في الأدبيات والممارسات العالمية، إلى جانب تسليط الضوء على الهيئات والمؤسسات التي تعمل على تبني وتفعيل هذه المبادئ. كما يولي المحور اهتماماً خاصاً بمبادئ الهيئة السعودية للبيانات والذكاء الاصطناعي (SDAIA) بوصفها نموذجاً عربياً يسعى إلى موازنة استخدام الذكاء الاصطناعي مع القيم والأطر الأخلاقية المحلية.

أولاً: المبادئ التوجيهية لأخلاقيات الذكاء الاصطناعي

تاريخياً كان كاتب الخيال العلمي الشهير Isaac Asimov أول من قدم قانوناً أخلاقياً لأنظمة الذكاء الاصطناعي، حيث قدم قوانينه الثلاثة للروبوتات في كتابه Runaround (أسيموف 1942)، وقد تم استكمال هذه القوانين الثلاثة لاحقاً بقانون رابع يُسمى القانون الصفري للروبوتات، في كتاب Robots and Empire (أسيموف 1986)، والقوانين الأربعة هي كما يلي: (Stewart & Nyholm, 2021, p. 05)

- لا يجوز للروبوت أن يؤدي إنساناً أو أن يسمح بإيذاء إنسان من خلال التفاعس؛
 - يجب على الروبوت أن يطيع الأوامر الصادرة إليه من البشر إلا إذا كانت هذه الأوامر تتعارض مع القانون الأول؛
 - يجب على الروبوت حماية وجوده طالما أن هذه الحماية لا تتعارض مع القانون الأول أو الثاني؛
 - لا يجوز للروبوت أن يؤدي البشرية أو يسمح للبشرية أن تعاني من الأذى بسبب عدم اتخاذ أي إجراء.
- أما اليوم فالتحديات والمخاوف التي تصاحب انتشار الذكاء الاصطناعي لا بد أن يقابلها اقرار مبادئ توجيهية وأخلاقية وهياكل حوكمة تؤسس لمبدأ الثقة في هذه التقنية واستخداماتها وتضمن تضيق هامش مخاطر الذكاء الاصطناعي في حال حدوثها (بوغانم، 2024، صفحة 23)، وجاء في "تقرير دبي الذكية" أن هناك حاجة ماسة وملحة إلى مناقشة شاملة ومفتوحة حول أخلاقيات الذكاء الاصطناعي، والمبادئ التي يجب على المنظمات التي تستخدم هذه التكنولوجيا تطويرها ومراعاتها (AISHA, 2019, p. 14)، وعلى خلفية الاهتمام المتزايد بالقضايا الأخلاقية المتعلقة بالذكاء الاصطناعي، سارعت العديد من المؤسسات والحكومات تحت إشراف الخبراء في المجال إلى صياغة وثائق السياسات والمبادئ التوجيهية الأخلاقية للذكاء الاصطناعي، هذا التسارع أدى إلى ظهور الكثير من الوثائق التوجيهية لتنوع بذلك المبادئ الأخلاقية للذكاء الاصطناعي وأصبح من الصعب على الباحثين تتبعها واحصائها كلها، ومن جهة أخرى أدى هذا الاهتمام أيضاً إلى توفير العديد من مصادر التمويل والرعاية وإنشاء مراكز بحثية متعددة لأخلاقيات الذكاء الاصطناعي.

إن أحد الموارد الممتازة في هذا السياق هو النظرة العامة التي أجراها (Jobin et al. 2019)، الذين أجروا مراجعة مقارنة كبيرة لـ 84 مجموعة من المبادئ التوجيهية الأخلاقية الصادرة عن المنظمات الوطنية أو الدولية من بلدان مختلفة، وجد Jobin et al تقارباً كبيراً حول خمسة مبادئ رئيسية - الشفافية والعدالة والإنصاف وعدم الإيذاء والمسؤولية والخصوصية - من بين العديد من المبادئ الأخرى (Stewart & Nyholm, 2021, p. 22).

الجدول (02): مختلف المبادئ التوجيهية لأخلاقيات الذكاء الاصطناعي

المبدأ الأخلاقي	عدد وثائق المبادئ التوجيهية الأخلاقية (من 84 وثيقة تم فحصها)	الرموز المضمنة
الشفافية	73	الشفافية، القدرة على التفسير، القدرة على الفهم، التواصل لإفصاح.
العدالة والإنصاف	68	العدالة، الإنصاف، الاتساق، الشمول، المساواة، (عدم التحيز، (عدم التمييز، التنوع، التعددية، إمكانية الوصول لمكانية الرجوع، التعويض، الوصول، التوزيع.
عدم الإيذاء	60	عدم الإيذاء، الأمن، السلامة، الأذى، الحماية، الاحتياط، في النزاهة (الجسدية أو العقلية)، عدم التخريب.
المسؤولية	60	المسؤولية، المساءلة، القدرة على العمل بنزاهة.
الخصوصية	47	الخصوصية، المعلومات الشخصية أو الخاصة.

الإحسان	41	الفوائد، الإحسان، الرفاهية، السلام، الخير الاجتماعي، الصالح العام.
الحرية والاستقلال	34	الحرية، الاستقلال، الموافقة، الاختيار، تقرير المصير، التحرر، التمكين.
الثقة	28	الثقة
الاستدامة	14	الاستدامة، البيئة (الطبيعة)، الطاقة، الموارد (الطاقة).
الكرامة	13	الكرامة
التكافل	6	التضامن، الضمان الاجتماعي، التماسك.

المصدر: (Stewart & Nyholm, 2021, p. 22)

في الجدول أعلاه يشير العمود الأول على اليمين إلى المبدأ الأكثر تداولاً في الوثائق المفحوصة، في حين يدل الرقم الموجود على العمود الثاني إلى عدد وثائق المبادئ التوجيهية الأخلاقية من بين 84 وثيقة مفحوصة تم التي ذكر المبدأ فيها، أما العمود الثالث على اليسار فيشير إلى مجموعة المعايير أو الرموز التي ظهر بها المبدأ المشار إليه في العمود الأول في مختلف الوثائق.

ثالثاً: بعض الهيئات العاملة على تأطير مبادئ أخلاقيات الذكاء الاصطناعي

يرى Keng & Weiyu بأن الذكاء الاصطناعي المتقدم سيؤدي إلى تحقيق مكاسب تجارية غير مسبوق، ولكن في طريق تحقيق ذلك سيتعين على قادة الحكومة والصناعة التعامل مع مجموعة متنوعة من المضاعفات الأخلاقية مثل مشكلات خصوصية البيانات، وتحيز التعلم الآلي، ومخاوف السلامة العامة، بالإضافة إلى مشاكل استبدال الوظائف ومعدل البطالة، وتوجيه استراتيجياتها في تطوير تقنيات الذكاء الاصطناعي واعتمادها واحتضانها، يجب على المنظمات النظر في إنشاء أطر وإرشادات أخلاقيات الذكاء الاصطناعي، حيث بدأت بعض المؤسسات العمل بشأن هذه المسألة ونشرت بعض المبادئ التوجيهية (Keng & Weiyu, 2020, p. 05). ويوضح الجدول أدناه بعض المؤسسات العاملة على الأطر أو المبادئ الأخلاقية لمنظمة الذكاء الاصطناعي وأهدافها ومضمونها:

الجدول (03): أمثلة عن أعمال بعض المؤسسات في مجال أخلاقيات الذكاء الاصطناعي

المضمون	الأهداف	المصادر
معهد مستقبل الحياة (2017)	يؤكد هذا التقرير على "عدم إلحاق الأذى". يتطلب تطوير الذكاء الاصطناعي لإفادة المجتمع، وتعزيز الثقة والتعاون، وتجنب السباقات التنافسية.	السلامة، شفافية الفشل، المسؤولية، حفظ القيم الإنسانية، والخصوصية الشخصية، المنفعة المشتركة، الرخاء المشترك، السيطرة البشرية، عدم التخريب، سباق التسلح بالذكاء الاصطناعي
الرابطة الدولية لمحترفي الخصوصية (IAPP, 2018)	يستكشف الإطار المقترح المخاطر التي تهدد الخصوصية والإنصاف والشفافية والمساواة والعديد من القضايا الأخرى التي يمكن تضخيمها بواسطة البيانات الضخمة والذكاء الاصطناعي. وهي توفر نظرة عامة على كيفية قيام المؤسسات بتشغيل أخلاقيات البيانات وكيفية عكس الاعتبارات الأخلاقية في صنع القرار.	أخلاقيات البيانات، الخصوصية، التحيز، المساواة، الشفافية، حقوق الإنسان
معهد مهندسي الكهرباء والإلكترونيات (IEEE, 2019)	تحديد الممارسات لإنشاء هيكل حوكمة الذكاء الاصطناعي بما في ذلك المعالجة العملية لإدارة البيانات والحوسبة العاطفية والاقتصاد والشؤون القانونية وغيرها من المجالات. إلى جانب ذلك، يتطلب مبدأ IEEE أن يتم تعليم كل من يشارك في تصميم وتطوير الذكاء	حقوق الإنسان، الرفاهية، وكالة البيانات، الفعالية، الشفافية، المساواة، الوعي بإساءة الاستخدام، الكفاءة

	الاصطناعي لتحديد أولويات الاعتبارات الأخلاقية.	
فريق الخبراء رفيع المستوى المعني ب الذكاء الاصطناعي التابع للمفوضية الأوروبية (المفوضية الأوروبية ، 2019)	تم تصميم المبادئ التوجيهية لتوجيه مجتمع الذكاء الاصطناعي في تطوير واستخدام "الذكاء الاصطناعي الجدير بالثقة" (أي الذكاء الاصطناعي القانوني والأخلاقي). وتؤكد المبادئ التوجيهية على أربعة مبادئ: احترام الاستقلالية البشرية، ومنع الضرر، والإنصاف، وقابلية التفسير.	الوكالة البشرية والإشراف، والمتانة التقنية والسلامة، والخصوصية وحوكمة البيانات، والشفافية، والتنوع، والصحة المجتمعية والبيئية، والمساءلة
منظمة الأمم المتحدة للتربية والعلم والثقافة (اليونسكو، 2017)	يهدف المبدأ الأخلاقي المقترح إلى تزويد صانعي القرار بمعايير تتجاوز الاعتبارات الاقتصادية البحتة.	كرامة الإنسان، قيمة الاستقلالية، قيمة الخصوصية، مبدأ "عدم الإضرار"، مبدأ المسؤولية، قيمة الإحسان، قيمة العدالة
إطار أخلاقيات أستراليا (داوسون وآخرون ، 2019)	تسليط الضوء على القضايا الأخلاقية الناشئة أو التي من المحتمل أن تنشأ في أستراليا من تقنيات الذكاء الاصطناعي ويحدد الخطوات الأولية نحو التخفيف منها. المهدف من هذه الوثيقة هو تقديم تقييم عملي للقضايا الرئيسية للمساعدة في تعزيز تطوير الذكاء الاصطناعي الأخلاقي في أستراليا.	الامتثال التنظيمي والقانوني والإنصاف ، وقابلية المنافسة ، وعدم الإضرار ، وحماية الخصوصية والشفافية وقابلية التفسير ، والمساءلة

المصدر: (Keng & Weiye, 2020, p. 07)

ثالثاً: مبادئ الهيئة السعودية للبيانات والذكاء الاصطناعي سدايا SDAIA:

الهيئة السعودية للبيانات والذكاء الاصطناعي سدايا SDAIA أنشئت بموجب الأمر الملكي رقم (أ/471) وتاريخ 29/12/1440هـ، وترتبط مباشرة برئيس مجلس الوزراء، ويرتبط بها تنظيمياً: مكتب إدارة البيانات الوطنية، والمركز الوطني للذكاء الاصطناعي، ومركز المعلومات الوطني، وتتمتع الهيئة بالشخصية الاعتبارية وبالاستقلال الإداري والمالي، ومقرها الرئيس في مدينة الرياض، وتسعى هذه الهيئة إلى تحقيق الأهداف التالية: (نبذة عن سدايا، 2025)

- مواصلة تحديث أجندة البيانات الوطنية والذكاء الاصطناعي بما يتوافق مع المبادئ الرئيسية؛
- تنفيذ أجندة البيانات والذكاء الاصطناعي على الصعيد الوطني؛
- توجيه الجهات التابعة للهيئة بشأن تنفيذ أجندة البيانات والذكاء الاصطناعي؛
- زيادة الوعي العام بشأن إنجازات المملكة العربية السعودية في مجال البيانات والذكاء الاصطناعي؛
- زيادة مساهمة البيانات والذكاء الاصطناعي في تحقيق أهداف رؤية 2030؛
- زيادة استفادة الجهات الحكومية كافة من أجندة البيانات والذكاء الاصطناعي؛
- دعم تحقيق الأهداف الاستراتيجية في الجهات التابعة للهيئة فيما يتعلق بالتشريعات والتشغيل والابتكار؛
- تعزيز صورة المملكة العربية السعودية باعتبارها دولة رائدة عالمياً في مجال البيانات والذكاء الاصطناعي.

وعلى خلفية التطورات السريعة التي تشهدها تقنيات الذكاء الاصطناعي ومواكبة للجهود العالمية في الاستعداد للمخاطر والتداعيات الأخلاقية للاستخدام المسؤول لهذه التقنيات، سارعت الهيئة السعودية للبيانات والذكاء الاصطناعي سدايا SDAIA إلى إنشاء دليل عملي لتبسيط مفهوم أخلاقيات الذكاء الاصطناعي وتوضيح أهميته وتلخيص أهم المبادئ التوجيهية والممارسات التطبيقية

التي ينصح بالأخذ بها عند تطوير أنظمة الذكاء الاصطناعي أو تبنيها، كما يقدم الدليل أيضا مجموعة من الأدوات التي تساعد على تكريس مبادئ أخلاقيات الذكاء الاصطناعي.

يركز الدليل على خمسة مبادئ رئيسية لأخلاقيات الذكاء الاصطناعي، حسب ما تبنته منظمة التنمية الاقتصادية (OECD) عام 2007، واعتمدها مجموعة العشرين بالإضافة إلى 42 دولة أخرى، وأكدت عليها منظمة الأمم المتحدة للتربية والعلم والثقافة (اليونيسكو) عام 2021، وتبنتها 193 دولة، من ضمنها المملكة العربية السعودية. (SDAIA, 2024, p. 12)

الشكل (03): المبادئ التوجيهية لأخلاقيات الذكاء الاصطناعي لـ SDAIA



المصدر: (SDAIA, 2024, p. 12)

- النمو الشامل والتنمية المستدامة والرفاهية: يهدف هذا المبدأ إلى تسخير تقنيات الذكاء الاصطناعي واستخدامها بطريقة مبتكرة وتعود بالنفع على الأفراد والمجتمعات والبيئة وتساهم في النمو والازدهار وتعزيز أهمية التنمية العالمية، ويشمل هذا المبدأ ما يلي: (SDAIA, 2024, p. 14)
 - ✓ النمو الشامل والرفاهية: تقديم المنفعة وتعزيز السلام ورفاهية الإنسان وازدهاره، وخلق الفرص الاجتماعية والاقتصادية.
 - ✓ الاستدامة: تقديم منفعة مستدامة تساهم في تحسين النظام البيئي وحماية حفظ مصادره.
- القيم الإنسانية والعدالة: يركز هذا المبدأ على تسخير تقنيات الذكاء الاصطناعي وتصميمها واستخدامها بطريقة تحترم القانون وحقوق الإنسان بما فيها المساواة والتنوع، بما يضمن للأفراد والمجتمعات العدالة والإنصاف، وتتمحور قيم هذا المبدأ حول: (SDAIA, 2024, p. 14)
 - ✓ الإنسان: أن تكون أنظمة الذكاء الاصطناعي تخدم البشرية وتراعي قضايا حقوق الإنسان.
 - ✓ العدالة: أن يكون نظام الذكاء الاصطناعي طوال دورة حياته عادلا وبما يضمن التنوع العادل والمتكافئ للفوائد والفرص بالإضافة إلى خلو النظام من التحيز والتمييز.

- **الشفافية وقابلية التفسير:** يؤسس هذا المبدأ لبناء الثقة في أنظمة الذكاء الاصطناعي، لذا يجب بناء أنظمة بدرجة عالية من الوضوح والقابلية للتفسير، مع وجود ميزات لتتبع مراحل اتخاذ القرارات المؤتمتة، ولا سيما تلك التي تؤدي إلى آثار ضارة على الأفراد يعني بأن البيانات والخوارزميات والعمليات والغرض من النظام جميعها يجب أن تكون شفافة وقابلة للتفسير للمتأثرين بها بشكل مباشر أو غير مباشر (SDAIA., 2023, p. 20)، حيث يرتبط هذا المبدأ بشرح أنظمة الذكاء الاصطناعي ومخرجاتها إذ تسمح لأصحاب المصلحة بفهم المراحل الرئيسية للتطوير وعمليات اتخاذ القرار، ويتكون من محورين أساسيين: (SDAIA, 2024, p. 21)
 - ✓ **قابلية التفسير:** شرح كيفية وصول أنظمة الذكاء الاصطناعي إلى قراراتها ومخرجاتها بطريقة بسيطة ومفهومة.
 - ✓ **الشفافية:** فهم كيفية تنفيذ كل مرحلة من مراحل دورة حياة نظام الذكاء الاصطناعي لتقديم معلومات تتعلق بحقيقة استخدام الذكاء الاصطناعي في عمليات صنع القرار، المعلومات المتعلقة بالاحتمالات أو المنطق المعتمد في النظام، البيانات التي يستخدمها النظام، والغرض من النظام وكيفية استخدامه.
- **المتانة والأمن والسلامة:** يهتم هذا المبدأ بمعالجة تحديات أمن أنظمة الذكاء الاصطناعي وسلامتها ومواجهة المخاطر الأمنية عليها طوال دورة حياتها، إضافة إلى سلامة مخرجات أنظمة الذكاء الاصطناعي في جميع الظروف المختلفة، ويتضمن المبدأ ركيزتين رئيسيتين هما: (SDAIA, 2024, p. 26)
 - ✓ **المتانة:** إنتاج مخرجات ذات موثوقية عالية في ظل الظروف المختلفة.
 - ✓ **الأمن والسلامة:** والذي يشمل الجانب الرقمي من خلال سلامة البيانات وجودتها وحمايتها واحترام مبادئ الخصوصية والجانب المادي من خلال خلو أنظمة الذكاء الاصطناعي من المخاطر التي تؤثر في السلامة المادية وضمان عدم إلحاق الضرر بالأفراد بطريقة مقصودة أو غير مقصودة.
- **المسؤولية والمساءلة:** يُحمّل مبدأ المساءلة والمسؤولية المصممين والمطورين ومسؤولي ومقيمي أنظمة الذكاء الاصطناعي المسؤولية الأخلاقية عن القرارات والإجراءات التي قد تؤدي إلى مخاطر محتملة وآثار سلبية على الأفراد والمجتمعات، ويجب تطبيق الإشراف البشري والحكومة والإدارة المناسبة عبر دورة حياة نظام الذكاء الاصطناعي بأكملها لضمان جودة آليات مناسبة لتجنب الأضرار وإساءة استخدام هذه التقنية (SDAIA., 2023, p. 23)، فهذا المبدأ يركز على ضمان نزاهة أنظمة الذكاء الاصطناعي وعدالتها وتوضيح المسؤولية والصلاحيات والأطر التنظيمية المتبعة لجميع الإجراءات أو القرارات خلال دورة تطوير الأنظمة أو نشرها واستخدامها، وستضمن هذا المبدأ ما يلي: (SDAIA, 2024, p. 30)
 - ✓ **المسؤولية:** ما يلزم المسؤول من واجبات وما يترتب عليه من تبعات في جميع مراحل دورة تطوير أنظمة الذكاء الاصطناعي.
 - ✓ **المساءلة:** قدرة المسؤول على تفسير قرارات الذكاء الاصطناعي ونتائج الواقعة ضمن مسؤوليته.

الشكل (04): مضمون المبادئ التوجيهية لأخلاقيات الذكاء الاصطناعي لـ SDAIA

النمو الشامل والتنمية المستدامة والرفاهية	القيم الإنسانية والعدالة	الشفافية وقابلية التفسير	المثانة الأمن والسلامة	المسؤولية والمساءلة
الإعتبارات				
- تحقيق الفائدة للأفراد والمجتمعات والبيئة؛ - مراعاة الاستدامة والمسؤولية البيئية وإفادة البشرية بما فيها ذلك الأجيال القادمة عند تطوير أنظمة AI؛ - وضوح آثار نظام AI واستخداماته الإيجابية والسلبية المباشرة وغير المباشرة على الأفراد والمجتمع والبيئة.	- التمثيل الدقيق والشامل للبيانات المغذية لنظام AI للغة المتأثرة؛ - عدم تحيز مخرجات نظام AI؛ - مواعمة نظام AI مع القوانين واحترام الخصوصية وحرية التعبير وحماية البيانات؛ - إتاحة الوصول لنظام AI لمختلف المستخدمين؛ - مراعات اختلاف المهارات المعرفية والثقافية للإنسان عند تطوير AI واستخدامه.	- مراعات تعزيز الشفافية والقابلية للتفسير عند تطوير نظام AI ووضع معايير التقييم؛ - قابلية وسهولة توضيح قرارات نظام AI؛ - إمكانية تتبع العوامل الرئيسية التي أثرت على قرارات نظام AI؛ - ادراك المستخدم مشاركة نظام AI عملية صنع القرار أو عند التفاعل معه.	- ضمان عمل النظام بطريقة آمنة في جميع الظروف؛ - ضمان أمن جميع الأفراد المشغلين وغيرهم؛ - الاعتماد على تدابير أمن وسلامة تتناسب مع حجم المخاطر؛ - ضمان احترام خصوصية البيانات وحمايتها.	- قابلية تحديد المسؤولية؛ - توزيع المساءلة على نتائج نظام AI على المسؤولين في جميع مراحل دورة حياة النظام؛ - أن لا تنسب المسؤولية عن النتائج أو الخسائر إلى النظام بحد ذاته؛ - إتاحة الفرصة للمستخدمين للاعتراض على قرارات ومخرجات AI؛ - لا يصدر AI قرارات بالنيابة عن المستخدم دون علمه؛ - لا يترك أمر قرارات الموت والحياة لنظام AI.
الممارسات التطبيقية				
- اشراك أصحاب المصلحة بصورة مسبقة في إدارة الاستخدام المسؤول لنظام AI؛ - تقييم مصممي ومطوري نظام AI خلو التأثيرات السلبية والاجابية لنظام AI؛ - تحديد آليات لمراقبة آثار نظام AI وتقييمها؛ - الكشف عن الآثار المتوقعة من الحلول القائمة على AI؛ - اعطاء التأثيرين بأنظمة AI الحق في عملية اتخاذ القرار عند تبني هذه الأنظمة وتطويرها.	- وضع آليات متوافقة مع القوانين والقيم والعدالة لأنظمة AI؛ - موازنة المخاطر والفوائد المتوقعة لأنظمة AI؛ - مراعات تنوع الأعمار والخلفيات الثقافية عن تطوير أنظمة AI؛ - وضع القيود والمتطلبات لضمان عدم تحيز وتمييز البيانات؛ - تحديد آليات لتقييم مدى يتسبب نظام AI في تمييزاته؛ - التحقق من جودة البيانات وشمولها بشكل مبكر ومتكرر؛ - مراعات معايير سهولة الوصول؛ - تقييم موثوقية المخرجات؛ - وضع آلية للتغذية الراجعة.	- توثيق المعلومات والمعلومات لفترة تتناسب مع نوع القرار أو مجال النظام؛ - توفير معايير لقياس مدى تناسب الشفافية مع غرض نظام AI؛ - تفسير القرارات؛ - استخدام لغة غير تقنية ومفهومة لأصحاب المصلحة؛ - توفير وسيلة لتتبع القرارات؛ - إيضاح المقصد العام للنظام؛ - مراعات الخصوصية؛ - الإفصاح بالهوية الآلية والتوضيح للمستخدم حقيقة التفاعل مع AI؛ - توفير آلية الاستفسار.	- اختبار قدرة نظام AI على تقديم تنبؤات صحيحة عند اختلاف المدخلات أو الظروف أو تكرار التجارب؛ - وضع منهجية لإدارة المخاطر من خلال: • تحديد المخاطر المحتملة؛ • تقييم المخاطر المحتملة؛ • تحديد طرق الاستجابة؛ • توفير آليات الإبلاغ والتحقق. - وضع الآليات لحماية نظام AI من الثغرات؛ - توثيق العمليات مع الصيانة المستمرة؛ - تحديد صلاحيات الوصول؛ - الامتثال للتشريعات ووضع مقاييس لمبدأ المثانة والأمن.	- مراعات ووضوح السياسات حول قضايا المسؤولية والمساءلة للمصممين والمطورين وغيرهم؛ - توضيح حدود المسؤولية للمستهلك النهائي؛ - وضع آليات للاستئناف وخطط الطوارئ لمعالجة مخرجات النظام الخاطئة أو التخفيف منها؛ - تخصيص لجنة لمعالجة الشكاوي؛ - تحديد آليات التعويض للمستخدمين المتأثرين بقرارات نظام AI ونتائجها الخاطئة.

المصدر: من اعداد الباحثين بالاعتماد على بيانات تقرير سدايا SDAIA

خاتمة:

خلُصت هذه الدراسة إلى أن تنامي استخدام تقنيات الذكاء الاصطناعي في مختلف المجالات فرض تحديات غير مسبقة، لا سيما من حيث ضبط هذه التقنيات ضمن أطر أخلاقية تضمن العدالة والشفافية وتحفظ حقوق الأفراد، وقد بينت الدراسة أن الذكاء الاصطناعي ليس مجرد مسألة تقنية أو علمية فحسب، بل هو قضية ذات أبعاد إنسانية، وقانونية، وسياسية، تستدعي إعادة النظر في آليات الحوكمة والتشريع.

من خلال تحليل المبادئ التوجيهية المقترحة في التجارب الدولية، وبالأخص تجربة الهيئة السعودية للبيانات والذكاء الاصطناعي (SDAIA)، اتضح أن بلورة أخلاقيات الذكاء الاصطناعي تتطلب تبني مبادئ واضحة تركز على: احترام الكرامة الإنسانية، وتحقيق العدالة، وضمان الشفافية والمساءلة، وتجنب الانحيازات الخوارزمية، وتعزيز الأمن السيبراني.

كما تؤكد الدراسة على ضرورة التعاون بين الهيئات الحكومية ومراكز الأبحاث والقطاع الخاص من أجل وضع أطر تنظيمية فعالة وملزمة، تتماشى مع الخصائص الثقافية والاجتماعية، وتحاكي المعايير الدولية، فالتحولات السريعة التي يشهدها الذكاء الاصطناعي قد تسبق قدرة المجتمعات على التكيف معها، ما لم يتم تداركها بحوكمة رشيدة واستباقية.

• نتائج الدراسة:

- يفهم الذكاء الاصطناعي على أنه استخدام الآلات للقيام بأشياء تتطلب عادة ذكاء بشريا، في العديد من مجالات الحياة البشرية؛
- إن التأثير السريع والملموس للذكاء الاصطناعي على المجتمع البشري والطرق التي يتفاعل بها مع بعضه البعض، تفرض تحديات أخلاقية واجتماعية وسياسية كبيرة كمشاكل الأمان، الشفافية، التحيز، التمييز، والمساءلة، وهذا ما تتطلب تحليلا أخلاقيا شاملا.
- إن الاهتمام بأخلاقيات الذكاء الاصطناعي جاء كرد فعل على الأضرار الفردية والاجتماعية التي يسببها سوء الاستخدام أو سوء التصميم والآثار السلبية لأنظمة الذكاء الاصطناعي؛
- تهدف الهيئة السعودية سدايا SADAIA إلى تنظيم قطاعات البيانات والذكاء الاصطناعي من خلال وضع سياسات ومعايير وضوابط خاصة بها، وتوضيح كيفية التعامل معها؛
- إن المبادئ الأخلاقية التوجيهية الصادرة عن جهات مختلفة على غرار الهيئة السعودية سدايا SADAIA، تسعى إلى توفير أساليب معالجة القضايا الأخلاقية في الذكاء الاصطناعي، والوفاء بالمبادئ الأخلاقية وتوفير المرجع لتقييم أخلاقيات الذكاء الاصطناعي؛
- تهدف القيم والمبادئ التوجيهية لهيئة سدايا SADAIA إلى تحفيز الممارسات المقبولة أخلاقيا وتحديد الواجبات والالتزامات الأساسية اللازمة لإنتاج تطبيقات الذكاء الاصطناعي الأخلاقية والعادلة والآمنة؛
- إن الحصول على بيئة تكنولوجية تمكن من التصميم الأخلاقي والنشر السليم لأنظمة الذكاء الاصطناعي، تتطلب جهدا جماعيا متعدد التخصصات للحفاظ على ثقافة المسؤولية نحو إنتاج ممارسات سليمة أخلاقيا في كل مرحلة من دورة حياة أنظمة الذكاء الاصطناعي؛

• توصيات الدراسة:

- على ضوء ما توصلت اليه الدراسة من نتائج يمكن عرض جملة التوصيات التالية:
- لا بد أن تكون أخلاقيات الذكاء الاصطناعي من بين أولى الاعتبارات في تطوير تقنيات الذكاء الاصطناعي وليست مجرد فكرة لاحقة، لأن مستقبل البشرية قد يكون مرهونا بالتطوير الصحيح لأخلاقيات الذكاء الاصطناعي؛
- ضرورة دمج اعتبارات الآثار الاجتماعية والأخلاقية لتصميم واستخدام أنظمة الذكاء الاصطناعي في كل مرحلة من مراحل تنفيذ مشاريع الذكاء الاصطناعي؛

- ضرورة بذل جهد تعاوني بين مديري منتجات الذكاء الاصطناعي ومهندسي البيانات والخبراء في المجال لملاءمة تطوير تقنيات الذكاء الاصطناعي مع القيم والمبادئ الأخلاقية التي تحمي وتعزز رفاهية المجتمعات التي تستهدفها هذه التقنيات؛
 - تعزيز دور الهيئات المختصة مثل SDAIA وتوسيع مهامها لتشمل الرقابة الأخلاقية على المنتجات والتطبيقات الذكية؛
 - بناء قدرات وطنية في مجال "الأخلاقيات الرقمية" عبر إدماجها في مناهج التعليم والتكوين.
- وفي الأخير، تفتح هذه الدراسة المجال أمام دراسات مستقبلية أكثر تخصصاً، قد تتناول بالتفصيل تقنيات محددة (مثل الذكاء الاصطناعي التوليدي، أو خوارزميات التعلم الآلي) من زاوية أخلاقية، بما يعزز فهمنا الجماعي للذكاء الاصطناعي كأداة تستوجب الحذر والحكمة.

الإحالات والمراجع:

قائمة المراجع باللغة العربي

1. بوغنام. (2024). جمال بوغنام، الذكاء الاصطناعي وأخلاقياته، مجلة اتواصل، جامعة برج باجي مختار -عناية، الجزائر، المجلد 30، العدد(01).
2. كوكليبرج. (2024). مارك كوكليبرج، ترجمة هبة عبد العزيز غانم، أخلاقيات الذكاء الاصطناعي، مؤسسة هنداوي للنشر.
3. نبذة عن سدايا. (13، 03، 2025). تاريخ الاسترداد: 13.03.2025، الموقع الرسمي للهيئة السعودية للبيانات والذكاء الاصطناعي سدايا SADAIA، على الرابط: <https://sdaia.gov.sa/ar/SDAIA/about/Pages/About.aspx>
4. SDAIA. (2024). أخلاقيات الذكاء الاصطناعي، تقرير الهيئة السعودية للبيانات والذكاء الاصطناعي، السعودية، الإصدار الثاني.
5. SDAIA. (2023). أخلاقيات الذكاء الاصطناعي، تقرير الهيئة السعودية للبيانات والذكاء الاصطناعي، السعودية، الإصدار الأول.

References in english

6. Albert Sabater, Alicia de Manuel. (2021). Artificial Intelligence, Ethics and Society An Overview and Discussion Through the Specialised Literature and Expert Opinions, Observatori d'Ètica en Intel·ligència Artificial de Catalunya (OEIAC), Universitat de Girona.
7. Aisha Bint Butti Bin Bishr.(2019) Smart Dubai Ai Ethics Principles & Guidelines, This document was produced by Smart Dubai in collaboration with Falcon and Associates. Dubai.
8. Changwu Huang, Member, IEEE, Zeqi Zhang, Bifei Mao, and Xin Yao. (2023). An Overview of Artificial Intelligence Ethics, Eee Transactions Onartificial Intelligence, vol 4(no4), pp.799-8189.
9. Dawson D and Schleiger E, Horton J, McLaughlin J, Robinson C, Quezada G, Scowcroft J, and Hajkowicz S. (2019). Artificial Intelligence Australia's Ethics Framework, Commonwealth Scientific and Industrial Research Organisation, CSIRO.
10. Francesca Rossi, Noah Treviño, Almas Ahmed. (2022). Everyday Ethics for Artificial Intelligence, IBM Design Program Office.
11. Keng Siau, Weiyu Wang. (2020). Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI, Published in Journal of Database Management, the School of Computing and Information, vol 31(no 2), 01,15.
12. Kimon Kieslich ,Birte Keller andChristopher Starke. (2022). Artificial intelligence ethics by design. Evaluating public perception on the importance of ethical design principles of artificial intelligence, Big Data & Society, Sage journals.
13. DavidLeslie (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute.
14. Müller Vincent C. (2021). 'Ethics of artificial intelligence', in Anthony Elliott (ed.), The Routledge social science handbook of AI (London: Routledge).
15. Stefan Trauşan-Matu. (2020). Ethics in Artificial Intelligence, International Journal of User-System Interaction, vol 13(no 3), 136,148.
16. John-Stewart Gordon, Sven Nyholm. (2021). Ethics of Artificial Intelligence | Internet Encyclopedia of Philosophy, Internet Encyclopedia of Philosophy.The University of Utrecht, The Netherlands
17. Soumya Swaminathan. (2021). Ethics and governance of artificial intelligence for health: WHO guidance,Geneva, World Health Organization.
18. Aimee van Wynsberghe. (2020). Artificial intelligence: From ethics to policy, STUDY Panel for the Future of Science and Technology EPRS | European Parliamentary Research Service Scientific Foresight Unit (STOA).