نحو أخلاقيات للآلة
(تقنيات الذكاء الاصطناعي وتحديات اتخاذ القرار)
Towards Machine Ethics
Artificial Intelligence and Decision-Making

دكتور/ صلاح عثمان (أستاذ المنطق وفلسفة العلم – رئيس قسم الفلسفة – كلية الآداب – جامعة المنوفية – جمهورية مصر العربية)

Salah Osman

(Menoufia University, Egypt)

salah.mohamed@art.menofia.edu.eg

DOI: 10.13140/RG.2.2.11802.11200

دراسة منشورة بالمركز العربي للبحوث والدراسات، القاهرة: ١٤ يوليو ٢٠٢٢

Arab Center for Research and Studies, Cairo, 2022, July 14.

تُعد أخلاقيات الآلة Machine Ethics جزءًا من أخلاقيات الذكاء الاصطناعي المعنية بإضافة أو ضمان السلوكيات الأخلاقية للآلات التي صنعها الإنسان، والتي تستخدم الذكاء الاصطناعي، وهي تختلف عن المجالات الأخلاقية الأخرى المتعلقة بالهندسة والتكنولوجيا، فلا ينبغي الخلط مثلاً بين أخلاقيات الآلة وأخلاقيات الحاسوب، إذ تركز هذه الأخيرة على القضايا الأخلاقية المرتبطة باستخدام الإنسان لأجهزة الحاسوب؛ كما يجب أيضًا تمييز مجال أخلاقيات الآلة عن فلسفة التكنولوجيا، والتي تهتم بالمقاربات الإبستمولوجية والأنطولوجية والأخلاقية، والتأثيرات الاجتماعية والاقتصادية والسياسية الكبرى، للممارسات التكنولوجية على تنوعها؛ أما أخلاقيات الآلة فتعني بضمان أن سلوك الآلات تجاه المستخدمين من البشر، وربما تجاه الآلات الأخرى أيضًا، مقبول أخلاقيًا. الأخلاقيات التي نعنيها هنا إذن هي أخلاقيات يجب أن تتحلى بها الآلات كأشياء، وليس البشر كمصنعين ومستخدمين لهذه الآلات!

بعبارة أخرى نستطيع القول إن الهدف النهائي للبحث في أخلاقيات الآلة هو تصميم آلة ذكية (وكيل اصطناعي Artificial Agent) تتبع مبدأ أخلاقيًا مثاليًا أو مجموعة من المبادئ؛ أي تسترشد بهذا المبدأ أو بهذه المبادئ في القرارات التي تتخذها حول مسارات العمل التي يمكن أن

تسلكها. ولعلنا هنا بحاجة إلى التمييز بين ما يُمكن أن نسميه «وكيل أخلاقي ضمني» Implicit ethical agent، و«وكيل أخلاقي صـــريح» Explicit ethical agent؛ فالآلة التي تمثل وكيلاً أخلاقيًا ضـــمنيًا هي تلك التي تمت برمجتها للتصـــرف بشـــكلٍ أخلاقي، أو على الأقل لتجنب السـلوك غير الأخلاقي، دون تمثيل صـريح للمبادئ الأخلاقية. بمعنى أنها مقيدة في سـلوكها – من قبل مصممها – باتباع مبادئ أخلاقية معينة. أما الآلة التي تمثل وكيلاً أخلاقيًا صريحًا فهي تلك التي لديها القدرة على تحديد أفضـل إجراء يمكن أن تقوم به وفقًا للمبادئ الأخلاقية، أي أنها يمكن أن تُمثل الأخلاق صـراحةً، ومن ثم تعمل بشـكلٍ فعال على أسـاس هذه المعرفة من تلقاء نفســـها. ولعل التحدي الأكبر الذي يُواجه كافة العاملين في مجال أخلاقيات الآلة، ليس هو تصميم الآلة الذكية في حد ذاتها، وإنما تحديد ما يُمكن أن نُعده سلوكًا أخلاقيًا يتفق عليه الجميع، وتلك مهمة صعبة نظرًا لتعدد الرؤى والنظريات الأخلاقية.

قبل القرن الحادي والعشـرين، كانت أخلاقيات الآلات موضـوعًا لأدب الخيال العلمي إلى حد كبير، ويرجع ذلك أسـاسًـا إلى قيود الحوسـبة والذكاء الاصـطناعي، وإن كان تعريف «أخلاقيات الآلة» قد تطور منذ الربع الأخير من القرن العشـرين، حيث صـاغ الفيزيائي الأمريكي «ميتشـل والدروب» Mitchell Waldrop هذا المصـطلح في مقال له بمجلة «الذكاء الاصـطناعي» AI Magazine سنة ١٩٨٧ تحت عنوان «سؤال عن المسؤولية» A Question of Responsibility، وكتب قائلاً:

«وعلى أية حال، ثمة شيءٌ واحد واضح من المناقشة أعلاه، مؤداه أن الآلات الذكية ستُجسد القيم والافتراضـات والمقاصـد، سـواء أراد المبرمجون ذلك أم لا، وسـواء أكان لديهم وعيٌ بذلك أم لا. وهكذا، عندما تصبح أجهزة الحاسوب والروبوتات أكثر ذكاءً، سيغدو من الضروري أن نفكر مليًا وبصـــراحةٍ ووضـــوح في ماهية تلك القيم المضـــمنة. لعل ما نحتاج إليه، في الواقع، هو نظرية وممارسة لأخلاقيات الآلة، وفقًا لروح قوانين أسيموف الثلاثة للروبوتات».

ومن المعروف أن قوانين أسيموف الثلاثة للروبوتات Three Laws of Robotics هي مجموعة من القوانين المُقترحة لكي يلتزم بها الإنســـان الآلي. وقد ظهرت هذه القوانين للمرة الأولى ســـنة ١٩٤٢ في إحدى روايات الخيال العلمي، وهي رواية «التملص» Runaround لكاتب روايات الخيال العلمي الأمريكي «إسحاق أسيموف» Isaac Asimov (١٩٢٠ – ١٩٩٢)، وجاءت على النحو التالي:

١. لا يجوز لآلي إيذاء بشريّ أو السكوت عما قد يسبب له الأذى.
٢. يجب على الآلي إطاعة أوامر البشر إلا إن تعارضت مع القانون الأول.
٣. يجب على الآلي المحافظة على بقائه طالما لا يتعارض ذلك مع القانونين الأول والثاني.

ثم أضـاف لها «أسـيموف» فيما بعد القانون «صـفر»، ومنطوقه: لا ينبغي لأي روبوت أن يؤذي الإنسانية، أو أن يسمح للإنسانية بإيذاء نفسها بعدم القيام بأي رد فعل.

لم تكن مسرحية «التملص» أولى روايات الخيال العلمي في هذا الصدد، فقد سبقتها مثلاً سنة ١٩٢٠ مسرحية «روبوتات روسوم العالمية» Rossumovi Univerzální Roboti (RUR) للكاتب التشيكي «كارل تشابيك» Karel Čapek (١٨٩٠ – ١٩٣٨)، وهي المسرحية التي قدمت كلمة «روبوت» Robot لأول مرة في اللغة الإنجليزية، وأثارت في الأذهان (بعد عرضـها في الخامس والعشرين من يناير سنة ١٩٢١) فكرة إمكانية هيمنة الروبوتات بعد تمردها على الجنس البشري والقضاء عليه. وفي سنة ١٩٦٨، قدَّم الكاتب البريطاني «آرثر كلارك» Arthur Clarke (١٩١٧ – ٢٠٠٨) رواية «ملحمة الفضاء» (أو «أوديسا الفضاء» A Space Odyssey)، التي تخيل فيها حاسـوبًا خوارزميًا له القدر على اتخاذ القرار (أطلق عليه اسـم «هال ٩٠٠٠» HAL 9000 Computer)، يتحكم في أنظمة إحدى مركبات الفضـاء ويقود انقلابًا يؤدي على القضـاء على طاقمها! وعلى خلفية هذه الهواجس والمخاوف التي طرحتها روايات الخيال العلمي، نشر مهندس الكمبيوتر الأمريكي «وليام نيلسون جوي» William Nelson Joy (من مواليد سنة ١٩٥٤) مقالاً في عدد أبريل ٢٠٠٠ من مجلة «وايرد» Wired تحت عنوان «لماذا لا يحتاجنا المسـتقبل؟» Why The Future Doesn't Need Us?، ذهب فيه إلى أقوى تقنياتنا في القرن الحادي والعشرين: الروبوتات، والهندسـة الوراثية، وتكنولوجيا النانو Nanotechnology، من شـأنها أن تجعل البشـر من الأنواع المهددة بالانقراض، وأن السـبيل الوحيد لتجنب هذا المصـير هو التخلي عن التقنيات الخطرة، وحث العلماء على رفض العمل على تطويرها! ولئن كان هذا مُستبعدًا، فلعل البحث في أخلاقيات الآلة يُقدم حلاً أكثر واقعية وقابلية للتطبيق!

في سنة ٢٠٠٤، قامت «جمعية النهوض بالذكاء الاصطناعي» AAAI (وهي جمعية علمية دولية مكرسة لتعزيز البحث في الذكاء الاصطناعي والاستخدام المسؤول له، تم تأسيسها سنة ١٩٧٩) بتنظيم ورشـة عمل تحت عنوان «نحو أخلاقيات الآلة» Towards Machine Ethics، وتم فيها وضع الأسس النظرية لأخلاقيات الآلة. كما عقدت الجمعية ذاتها سنة ٢٠٠٥ ندوة حول أخلاقيات الآلة التقى فيها باحثون من أقطار مختلفة للنظر في تنفيذ البُعد الأخلاقي للأنظمة المستقلة Autonomous Systems. وفي سنة ٢٠٠٧ نشرت «مجلة الذكاء الاصطناعي» AI Magazine مقالاً لكل من «مايكل أندرسـون» Michael Anderson (أسـتاذ علوم الحاسـوب بجامعة هارتفورد Hartford الأمريكية) و«سـوزان لي أندرسـون» Susan Leigh Anderson، (أستاذة التفكير الناقد والأخلاق التطبيقية بجامعة كونيتيكت Connecticut الأمريكية) تحت عنوان «أخلاقيات الآلة: إنشـاء وكيل ذكي أخلاقي» Machine Ethics: Creating an Ethical Intelligent Agent، وفيه ناقش الباحثان أهمية أخلاقيات الآلة، والحاجة إلى الآلات التي تمثل

المبادئ الأخلاقية صـــراحةً، والتحديات التي تواجه أولئك الذين يعملون على تطوير أخلاقيات الآلة. كما أظهرا أنه من الممكن لآلةٍ ما – على الأقل في مجال محدود – أن تُجرد مبدأ أخلاقيًا من أمثلة الأحكام الأخلاقية، وأن تستخدم هذا المبدأ لتوجيه سلوكها.

أما أول كتاب عن أخلاقيات الآلة فقد نشـرته مطبعة جامعة أكسـفورد سـنة ٢٠٠٩، تحت عنوان «الآلات الأخلاقية: تعليم الروبوتات الصواب من الخطأ» Moral Machines: Teaching Robots Right from Wrong، وهو تأليف مشترك لكل من «ويندل والاش» Wendell Wallach و«كولين ألين» Colin Allen، وهو أول عمل يفحص التحدي المتمثل في بناء آلات أخلاقية اصطناعية، ويتعمق في طبيعة صنع القرار البشري وأبعاده الأخلاقية. وقد استشهد فيه المؤلفان بحوالي ٤٥٠ مصـدرًا، وطرحا ما يقرب من مائة سـؤال رئيس حول أخلاقيات الآلة، ولفتا أنظار الباحثين وصانعي السياسات والممولين إلى أن عدد أعضاء البيئة التي يصنعها الإنسان بواسطة آلات قادرة – من خلال خوارزميات الذكاء الاصـطناعي – على التصـرف بشـكل مسـتقل، يزداد بشـكلٍ غير مسـبوق، وأن الخوارزميات التي تتحكم في سـلوك هذه الأنظمة المسـتقلة حتى الآن «عمياء أخلاقيًا» Ethically blind، بمعنى أن قدرات اتخاذ القرار لهذه الأنظمة لا تنطوي على أي تفكير أخلاقي واضح، ومن ثم تبدو الحاجة مُلحة إلى أن تصبح هذه الأنظمة المستقلة بشكل متزايد (الروبوتات وبرمجيات الروبوت) قادرة على مراعاة الاعتبارات الأخلاقية في عملية صـــنع القرار.

من جهة أخرى، أعلن مكتب الولايات المتحدة للبحوث البحرية سـنة ٢٠١٤ أنه سـيوزع ٧,٥ مليون دولار في شكل منح على مدى خمس سنوات للباحثين الجامعيين لدراسة مسائل أخلاقيات الآلة كما هي مطبقة على الروبوتات المستقلة. وفي سنة ٢٠١٦ نشر البرلمان الأوروبي ورقة من ٢٢ صـفحة لتشـجيع المفوضـية الأوربية على معالجة مسـألة الوضـع القانوني للروبوتات. وقد تضمنت الورقة أقسامًا تتعلق بالمسؤوليات القانونية للروبوتات، حيث تمت مناقشة الالتزامات على أنها تتناسـب مع مسـتوى اسـتقلالية الروبوتات. كما أثارت الورقة أيضًـا تسـاؤلات حول عدد الوظائف التي يمكن أن تحل محلها روبوتات الذكاء الاصطناعي.

وبصفته باحثًا موسوعيًا في دراسات فلسفة الذكاء الاصطناعي، وفلسفة العقل، وفلسفة العلم والمنطق، «يصـــف جيمس مور» James H. Moor (وهو أحد المنظرين الرائدين في مجال أخلاقيات الحاسوب بالولايات المتحدة) أربعة أنواع من الروبوتات الأخلاقية، وهي:

١. وكلاء التأثير الأخلاقي Ethical impact agents: وهي أنظمة آلية لها تأثير أخلاقي سـواء أكان مقصـودًا أم لا، ولديها في الوقت ذاته القدرة على التصـرف بشكل غير أخلاقي. يعطي مور مثالاً افتراضيًا يُطلق عليه اسم «وكيل جودمان» Goodman agent (نسبة إلى الفيلسوف الأمريكي «نيلســون جودمان»). يقارن «وكيل جودمان» التواريخ ولكن لديه خطأ ناجم عن

قيام المبرمجين بتمثيل التواريخ باستخدام آخر رقمين فقط من العام، وبالتالي فإن أية تواريخ بعد سنة ٢٠٠٠ سيتم التعامل معها بشكل مضلل على أنها أقدم من تلك التي كانت في أواخر القرن العشرين. وهكذا يُصبح «وكيل جودمان» بمثابة وكيل تأثير أخلاقي قبل سنة ٢٠٠٠، وبعد سنة ٢٠٠٠ وكيل تأثير غير أخلاقي.

٢. وكلاء أخلاقيون بشكلٍ ضمني Implicit ethical agents: تمت برمجة هؤلاء الوكلاء بهدف تجنب النتائج غير الأخلاقية.

٣. وكلاء أخلاقيون بشكلٍ صريح: وهذه آلات قادرة على معالجة السيناريوهات والتصرف بناءً على القرارات الأخلاقية، أي آلات لديها خوارزميات للعمل بشكل أخلاقي.

٤. وكلاء أخلاقيون بشكلٍ كامل Full ethical agents: وهذه آلات تُشبه الفئة السابقة في القدرة على اتخاذ قرارات أخلاقية، لكنها تحتوي أيضًا على سمات ميتافيزيقية بشرية (أي حرية الإرادة والوعي والقصد).

يذهب بعض الباحثين (مثل الفيلسوف السويدي الأصل «نيك بوستروم» Nick Bostrom (من مواليد ١٩٧٣)، وعالم الذكاء الاصطناعي البريطاني «ستيوارت راسل» Stuart Russell (من مواليد ١٩٦٢)، إلى أنه إذا تجاوز الذكاء الاصطناعي البشرية في الذكاء العام وأصبحت الآلة (فائقة الذكاء Superintelligent)، فقد يصبح هذا الذكاء الخارق الجديد قويًا ويصعب التحكم فيه، ومن ثم فإن مصير البشرية قد يعتمد على أفعال الذكاء الخارق للآلة في المستقبل. ويؤكد «نيك بوستروم» في كتابه «الذكاء الفائق: مسارات، مخاطر، استراتيجيات» Superintelligence: Paths, Dangers, Strategies (٢٠١٤)، و«ستيوارت راسل» في كتابه «متوافق مع الإنسان: الذكاء الاصطناعي ومشكلة التحكم» Human Compatible: Artificial Intelligence and the Problem of Control (٢٠١٩)، أنه في حين أن هناك كثيرًا من عدم اليقين فيما يتعلق بمستقبل الذكاء الاصطناعي، فإن الخطر على البشرية كبير بما يكفي للقيام بإجراءات مهمة في الوقت الحاضر. وهكذا يُمكن صياغة مشكلة التحكم في الذكاء الاصطناعي على النحو التالي: كيف يُمكن بناء وكيل ذكي يساعد مبدعيه، مع تجنب بناء ذكاء خارق عن غير قصد من شأنه إلحاق الضرر بمبدعيه. إن خطر عدم تصميم التحكم بشكل صحيح (في المرة الأولى)، هو أن الذكاء الخارق قد يكون قادرًا على الاستيلاء على السلطة في بيئته، ومنع البشر من إغلاقها. وتتضمن استراتيجيات التحكم المحتملة في الذكاء الاصطناعي «التحكم في القدرات» Capability Control (أي الحد من قدرة الذكاء الاصطناعي على التأثير في العالم)، و«التحكم التحفيزي» Motivational Control (وهو إحدى طرق بناء الذكاء الاصطناعي الذي تتماشى أهدافه مع القيم البشرية أو القيم المثلى). وثمة عدد من الجمعيات والمؤسسات التي تبحث في مشكلة التحكم في الذكاء الاصطناعي، بما في ذلك معهد مستقبل الإنسانية Future of Humanity Institute، ومعهد

أبحاث الذكاء الآلي Machine Intelligence Research Institute، ومركز الذكاء الاصـــطناعي المتوافق مع الإنسان Center for Human-Compatible Artificial Intelligence، ومعهد مستقبل الحياة Future of Life Institute.

في ســنة ٢٠٠٩، وفي تجربة مثيرة في مختبر الأنظمة الذكية Laboratory of Intelligent Systems بمدرسـة البوليتكنيك الفيدرالية Ecole Polytechnique Fédérale في لوزان Lausanne بسـويسـرا Switzerland، تمت برمجة روبوتات الذكاء الاصـطناعي بحيث تتعاون مع بعضـها البعض، وتم تكليفها بهدف البحث عن الموارد المفيدة مع تجنب الموارد السـامة. وخلال التجربة، تم تجميع الروبوتات في عشـائر، وتم اسـتخدام الشـفرة الوراثية الرقمية للأعضـاء الناجحين للجيل القادم، وهو نوع من الخوارزمية المعروفة باسم «الخوارزمية الجينية» Genetic Algorithm، وبعد خمسـين جيلًا متتاليًا في الذكاء الاصـطناعي، اكتشـف أعضـاء إحدى العشـائر كيفية التمييز بين المورد المفيد والمورد السـام. ثم تعلمت الروبوتات أن تكذب على بعضـها البعض في محاولة للاسـتئثار بالمورد المفيد من الروبوتات الأخرى. وفي التجربة ذاتها، تعلمت روبوتات الذكاء الاصـطناعي نفسـها أيضًـا التصـرف بنكران الذات، والإشـارة إلى الخطر على الروبوتات الأخرى، وماتت أيضًـا على حسـاب إنقاذ الروبوتات الأخرى! لقد تمت برمجة أهداف الروبوتات في هذه التجربة بحيث تكون «نهائية»، لكن الدوافع البشرية تتسم عادة بجودة تستلزم بحثًا لا ينتهي، وهو ما يضطلع به علماء الأخلاق حاليًا.

أيضًـا في سـنة ٢٠٠٩، شـارك عددٌ من الأكاديميين والخبراء في مؤتمرٍ لمناقشـة التأثير المحتمل للروبوتات وأجهزة الحاسوب، وتأثير الاحتمال الافتراضي بأن تُصبح هذه الآلات مكتفية ذاتيًا وقادرة على اتخاذ قرارات بنفسـها. وقد ناقش المشـاركون إمكانية ومدى قدرة أجهزة الحاسـوب والروبوتات على اكتساب أي مستوى من الاستقلالية، وإلى أي مدى يمكنها استخدام هذه القدرات لتشـكيل أي تهديد أو خطر، وأشـاروا إلى أن بعض الآلات قد اكتسـبت أشـكالًا مختلفة من شـبه الحكم الذاتي، بما في ذلك القدرة على العثور على مصـادر الطاقة بمفردها، والقدرة على اختيار الأهداف بشكل مستقل للهجوم بالأسلحة. كما أشـاروا إلى أن بعض فيروسـات الحاسـوب يمكنها التهرب من القضـاء عليها، وحققت ما يُسـمى «ذكاء الصـرصـور» Cockroach Intelligence. أشـاروا كذلك إلى أن الوعي الذاتي كما نجده في روايات الخيال العلمي من المحتمل ألا يكون مُحتملاً، لكن هناك مخاطر ومزالق أخرى محتملة. كما شـكك بعض الخبراء والأكاديميين في اسـتخدام الروبوتات للقتال العسـكري، خاصـةً عندما يتم إعطاء مثل هذه الروبوتات درجة معينة من الوظائف المسـتقلة، وقد قامت البحرية الأمريكية بتمويل تقرير يشـير إلى أنه مع زيادة تعقيد الروبوتات العسـكرية، يجب أن يكون هناك اهتمام أكبر بالآثار المترتبة على قدرتها على اتخاذ

قرارات مستقلة، وطالب رئيس جمعية النهوض بالذكاء الاصطناعي بإجراء دراسة للنظر في هذه المسألة!

من جهة أخرى، أصبحت خوارزميات البيانات الضخمة والتعلم الآلي العميق شائعة في العديد من الصناعات، بما في ذلك الإعلانات المنشورة عبر الإنترنت، والتصنيفات الائتمانية، والأحكام الجنائية، مع وعد بتقديم نتائج أكثر موضوعية وقائمة على البيانات، ولكن تم تحديدها كمصدر محتمل لعدم المساواة الاجتماعية والتمييز. وقد أظهرت دراسة أجريت سنة ٢٠١٥ أن النساء كن أقل عرضة لعرض إعلانات الوظائف ذات الدخل المرتفع من قبل تطبيق «جوجل أدسنس» Google AdSense. كما أظهرت دراسة أخرى أن خدمة التوصيل في اليوم ذاته من شركة «أمازون» لم تكن متوفرة عن قصد في الأحياء السوداء. ولم تتمكن شركتا أمازون وجوجل من عزل هذه النتائج في قضية واحدة، ولكن بدلاً من ذلك أوضحتا أن النتائج كانت نتيجة خوارزميات الصندوق الأسود التي استخدمتاها!

وفي محاولة لمعالجة المخاوف المتزايدة بشأن تأثير التعلم الآلي العميق على حقوق الإنسان، نشر المنتدى الاقتصادي العالمي ومجلس المستقبل العالمي لحقوق الإنسان في مارس سنة ٢٠٠٨ ورقة تحوي توصيات مفصلة حول أفضل السبل لمنع النتائج التمييزية في التعلم الآلي. كما طور المنتدى الاقتصادي العالمي أربع توصيات بناءً على مبادئ الأمم المتحدة التوجيهية لحقوق الإنسان للمساعدة في معالجة ومنع النتائج التمييزية في التعلم الآلي، وجاءت على النحو التالي:

١. الإدماج النشط Active Inclusion: يجب أن يسعى تطوير وتصميم تطبيقات التعلم الآلي بنشاط إلى مجموعة متنوعة من المدخلات، لا سيما معايير وقيم مجموعات سكانية محددة تتأثر بمخرجات أنظمة الذكاء الاصطناعي.
٢. الإنصاف Fairness: يجب على الأشخاص المشاركين في وضع تصور لأنظمة التعلم الآلي وتطويرها وتنفيذها النظر في تعريف الإنصاف الذي ينطبق بشكل أفضل على سياقاتهم وتطبيقهم، وإعطائه الأولوية في بنية نظام التعلم الآلي ومقاييس التقييم الخاصة به.
٣. الحق في الفهم Right to Understanding: يجب الإفصاح عن مشاركة أنظمة التعلم الآلي في صنع القرار الذي يؤثر على الحقوق الفردية، ويجب أن تكون الأنظمة قادرة على تقديم تفسير لاتخاذ قراراتها يكون مفهومًا للمستخدمين النهائيين ويمكن مراجعته من قبل سلطة بشرية مختصة. وعندما يكون هذا مستحيلاً وتكون الحقوق على المحك، يجب على القادة في تصميم تكنولوجيا التعلم الآلي ونشرها وتنظيمها أن يتساءلوا عما إذا كان ينبغي استخدامها أم لا.

٤. الوصـــول إلى التعويض Access to Redress: يتحمل القادة والمصـــممون ومطورو أنظمة التعلم الآلي مسـؤولية تحديد الآثار السـلبية المحتملة لأنظمتهم على حقوق الإنسـان. يجب عليهم توفير سـبل واضـحة لإنصـاف المتضـررين من الآثار المتباينة، وتدشـين عمليات للتعويض في الوقت المناسب عن أية مخرجات تمييزية.

وفي يناير سـنة ٢٠٢٠، نشـر «مركز بيركمان كلاين للإنترنت والمجتمع بجامعة هارفارد» Harvard University's Berkman Klein Center for Internet and Society دراسة وصفية لـ ٣٦ مجموعة بارزة من مبادئ الذكاء الاصـــطناعي، حددت ثمانية موضـــوعات رئيســـة، وهي: الخصوصية، والمساءلة، والسلامة والأمن، والشفافية وقابلية التفسير، والإنصاف وعدم التمييز، السيطرة البشرية على التكنولوجيا، والمسؤولية المهنية، وتعزيز القيم الإنسانية.

من المنظور الفلســفي بُذلت عدة محاولات لجعل الأخلاقيات قابلة للحســاب، أو على الأقل صورية. وحيث أن قوانين الروبوتات الثلاثة لإسحاق أسيموف لا تُعد عادةً مناسبة لوكيل أخلاقي مصطنع، فقد تمت دراسة ما إذا كان يمكن استخدام مفهوم الواجب المقولي عند «كانط» Kant's categorical imperative. ومع ذلك، فقد أشـــير إلى أن القيمة الإنســـانية، في بعض الجوانب، معقدة للغاية. وإحدى طرق التغلب على هذه الصعوبة هي تلقي القيم الإنسانية مباشرة من البشر من خلال آلية ما، على سبيل المثال من خلال تعلمهم.

ثمة مقاربة أخرى تتمثل في بناء الاعتبارات الأخلاقية الحالية على مواقف مماثلة سابقة، وهذا ما يسمى بعلم معالجة القضـايا (الإفتاء في قضـايا الضـمير) Casuistry، ويمكن تنفيذه من خلال البحث على الإنترنت؛ فالإجماع على مليون قرار ســـابق ســـيؤدي إلى قرار جديد يعتمد على الديمقراطية. وقد ابتكر البروفيسـور «بروس إم ماكلارين» Bruce M. McLaren نموذجًا حسـابيًا مبكرًا (منتصـــف التســـعينيات) لعلم معالجة القضـــايا، وتحديداً برنامج يُســـمى «ســـيروكو» SIROCCO، تم إنشاؤه باستخدام تقنيات الذكاء الاصطناعي والاستدلال بدراسة الحالة. ومع ذلك، يمكن أن تؤدي هذه المقاربة إلى قرارات تعكس التحيزات والســـلوكيات غير الأخلاقية التي تظهر في المجتمع. ويمكن رؤية الآثار السلبية لهذه المقاربة في برنامج الدردشة «تاي بوت» Tay (bot) الخاص بشركة ميكروسوفت، حيث تعلم الروبوت تكرار الرسائل العنصرية والمشحونة جنسيًا التي يرسلها مستخدمو تويتر.

أخيرًا، وعلى الرغم من أن الفكرة الأســـاســـية لأخلاقيات الآلة قد وجدت طريقها مؤخرًا نحو محاولات التطبيق الفعلي، إلا أن ثمة مناقشـــات جدلية واســـعة النطاق ما زالت تكتنف الفكرة وتُصـــاحب محاولات تطبيقها، لاســـيما فيما يتلق بالثراء والتنوع الأخلاقي في الثقافات المختلفة، وصـراعات الهيمنة السـياسـية والاقتصـادية، وحقيقة أن الروبوت المبرمج لاتباع قواعد – أو اتخاذ قرارات – أخلاقية، يمكن بســـهولة إعادة برمجته بحيث يتبع قواعد – أو يتخذ قرارات – غير

أخلاقية! والأخطر من هذا كله هو الإجابة عن السؤال المُلح والصعب: كيف يمكننا نحن البشر أن نظل متحكمين في نظام ذكاء اصطناعي بمجرد أن يصبح فائق الذكاء؟ وبمعنى أوسع، كيف يُمكننا التأكد من أن أي نظام ذكاء اصطناعي سيغدو إيجابيًا وفقًا للتصور البشري؟ تتمثل أحد جوانب صعوبة الإجابة عن هذا السؤال في أننا قد نقرر أن ميزة معينة مرغوبة، لكننا نكتشف بعد ذلك أن لها عواقب سلبية للغاية غير متوقعة لدرجة أننا لا نرغب في هذه الميزة على الإطلاق. تُماثل هذه المشكلة أسطورة «الملك ميداس» King Midas الذي تمنى أن يتحول كل ما يلمسه إلى ذهب، وما أن حقق له «ديونيسوس» Dionyssus (إله الخمر والابتهاج والنشوة) أمنيته، حتى تحول كل ما يلمسه، حتى طعامه وشرابه وابنته إلى ذهبٍ، فراح يتضرع إلى «ديونيسوس» كي يرفع عنه هذه اللعنة!

لقد انتقلنا في غضون عقودٍ قليلة من الحديث عن الذكاء الاصطناعي كخيالٍ علمي نُسلي به أنفسنا ونشحذ به عقول صغارنا وإن كان بعيد المنال، إلى الوعد بحل كافة مشاكلنا وتدشين دولة الرفاهية بتقنيات الذكاء الاصطناعي، إلى التحذير المُخيف بأن الذكاء الاصطناعي سيقتلنا جميعًا؛ وما علينا الآن سوى مراقبة التطورات التكنولوجية والاجتماعية عن كثب لفهم المخاطر التي تواجهنا وتواجه أحفادنا على المدى البعيد، ومناقشة القضايا الطارئة في وقت مبكر، وتطوير تحليل فلسفي يُلبي حاجات الواقع الجديد، وبناء رؤية مستقبلية تُشبع معطيات الحاضر ومردوداتها المنتظرة!

References and Farther Readings:

1. Anderson, M., Anderson, S. and Armen, C., 2004. *Towards Machine Ethics*. [Online] Available at: https://www.researchgate.net/publication/259656154_Towards_Machine_Ethics [Accessed 27 May 2022].

2. Anderson, M. and Anderson, S., 2007. Machine Ethics: Creating an Ethical Intelligent Agent. *AI Magazine*, *American Association for Artificial Intelligence*: 28(4), pp.15-26.

3. Anderson, Michael; Anderson, Susan Leigh, eds. (July 2011). *Machine Ethics*. Cambridge University Press.

4. Arkin, R., Ulam,, P. and Duncan, B., 2015. *An Ethical Governor for Constraining Lethal Action in an Autonomous System*. [Online] Research Gate. Available at: https://www.researchgate.net/publication/41042770_An_Ethical_Governor_for_Constraining_Lethal_Action_in_an_Autonomous_System [Accessed 27 May 2022].

5. Bird, E., Fox-Skelly, J., Jenner, N., Larbey, R., Weitkamp, E. and Winfield, A., 2000. *The Ethics of Artificial Intelligence*. Brussels: Mihalis Kritikos, Scientific Foresight Unit (STOA).

6. Boyles, R. J. M. (2018, June). *A Case for Machine Ethics in Modeling Human-Level Intelligent Agents*. kritike. Retrieved May 27, 2022, from https://www.kritike.org/journal/issue_22/boyles_june2018.pdf

7. Chaudhuri, S. and Vardi, M., 2013. *Reasoning About Machine Ethics*. [Online] Popl-obt-2014.cs.brown.edu. Available at: https://popl-obt-2014.cs.brown.edu/papers/ethics.pdf [Accessed 27 May 2022].

8. Davies, J. Consciousness, Machines, and Ethics. Proceedings, 2022, 81 (40), from https://doi.org/10.3390/proceedings2022081040

9. Moor, J. (2009). Four Kinds of Ethical Robots. *Philosophy Now*, (72), pp. 12 – 14.

10. Moor, J. (2006). The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems*, 21(4), pp. 18–21.

11. Müller, V. C. (2021). *Ethics of Artificial Intelligence and Robotics*. Stanford Encyclopedia of Philosophy. Retrieved May 29, 2022, from https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=ethics-ai

12. Ruttkamp-Bloem, E., 2020. *The Quest for Actionable AI Ethics*. [Online] Springer Link. Available at: https://link.springer.com/chapter/10.1007/978-3-030-66151-9_3 [Accessed 27 May 2022].

13. Segun, S., 2020. From Machine Ethics to Computational Ethics. *AI & SOCIETY*, 36(1), pp.263-276.

14. Storrs Hall, J. (May 30, 2007). *Beyond AI: Creating the Conscience of the Machine* Prometheus Books.

15. Tolmeijer, S., Kneer, M., Sarasua, C., Christen, M. and Bernstein, A., 2020. Implementations in Machine Ethics. *ACM Computing Surveys*, 53(6), Article 132, pp.1-38.

16. Waldrop, M. M. (1987). A Question of Responsibility. *AI Magazine*, 8(1), 28 – 39.

17. Wallach, Wendell; Allen, Colin (November 2008). *Moral Machines: Teaching Robots Right from Wrong*. USA: Oxford University Press.

18. Wikimedia Foundation. (2022, April 10). *Machine ethics*. Wikipedia. Retrieved May 27, 2022, from https://en.wikipedia.org/wiki/Machine_ethics

***

- **توثيق الدراسة بنظام APA:**

عثمان، صــلاح (١٤ يوليو ٢٠٢٢). «نحو أخلاقيات للآلة: تقنيات الذكاء الاصــطناعي وتحديات اتخاذ القرار». المركز العربي للبحوث والدراسات، القاهرة. تم الاسترداد بتاريخ ١٤ يوليو ٢٠٢٢ من:

http://www.acrseg.org/43003

APA Citation:

Osman, S. (عثمان، ص) (2022, July 14). Towards Machine Ethics: Artificial Intelligence and Decision-Making (نحو أخلاقيات للآلة: تقنيات الذكاء الاصــطناعي وتحديات اتخاذ القرار). Retrieved July 14, 2022, from http://www.acrseg.org/43003

- **عن الدراسة:**

- **خالد محسن: نحو أخلاقيات للآلة: تقنيات الذكاء الاصطناعي وتحديات اتخاذ القرار. دراسة حديثة للمفكر الدكتور صلاح عثمان، منصة المساء والجمهورية أون لاين، الإثنين ١٨ يوليو ٢٠٢٢.**

https://almessa.gomhuriaonline.com/الاص-الذكاء-تقنيات-للآلة-أخلاقيات-نحو

- **إبراهيم التحيوي: الحلقة ١٥ من برنامج حياة الناس على قناة المودة: من الدقيقة ٣,٥٣ إلى الدقيقة ٨,١٠.**

https://www.youtube.com/watch?v=SUef36vYX00&list=PLh4l8I2o1QWXe3TKSLygq8myuOQJYacDj&index=15

- **مركز سمت للدراسات: الأربعاء ٢٠ يوليو ٢٠٢٢.**

https://smtcenter.net/archives/slider/الاص-الذكاء-تقنيات-للآلة-أخلاقيات-نحو

***