ترجمة مقال

الذكاء الاصطناعي وعَزْو المسؤولية والتسويغ العِلاقي لمطلب تفسير الأفعال والقرارات الآليَّة*

تأليف: مارك كوكِلبِرج**

ترجمة: طارق عثمان***

الـمُلخَّص

تُناقِش هذه الدراسة إشكالية "عَزْو المسؤولية" التي يطرحها استعمال تقنيات الذكاء الاصطناعي؛ إذ يُفترَض أنْ يكون البشر وحدهم الفاعلين المسؤولين، لكنَّ هذا الافتراض يطرح علينا إشكالات كثيرة سنُناقِشها انطلاقاً مِن الشرطين الأرسطيين للمسؤولية: السيطرة، والمعرفة. وفي ما يختصُّ بشرط السيطرة، فإنَّنا سنضيف إلى مشكلة الأيادي الكثيرة الشهيرة مشكلة أُخرى نُسمّيها مشكلة الأشياء الكثيرة، وكذلك سنُشدِّد على البُعْد الزمني، ثمَّ نُركِّز على شرط المعرفة الذي يلفت انتباهنا إلى مسألتين مُهِمَّتين، هما: الشفافية، والتفسيرية. ولكنْ، في مُقابِل النقاشات الشائعة، سنُحاجِج بأنَّ إشكالية "معرفة الفاعل المسؤول" مُرتبِطة بالطرف الآخر في علاقة المسؤولية، وهو الطرف المفعول به أو المسؤول عنه؛ أي الطرف الذي يقع عليه فعل الطرف الأوَّل، ويحقُّ له أنْ يطلب منه تفسيراً لأفعاله التي فعلها به، ولقراراته التي اتَّخذها في شأنه، باستعمال الذكاء الاصطناعي. وبالسَّيْر على هذا النهج العِلاقي، تُقدِّم لنا المسؤولية، بوصفها مُساءَلة، تسويغاً إضافياً مُهِمّاً، بل أساسياً، لمطلب تفسير الأفعال والقرارات الآليَّة، وهو تسويغ ليس مبنياً على الفاعلية، وإنَّما يقوم على المفعولية الأخلاقية.

الكلمات المفتاحية: الذكاء الاصطناعي، عَزْو المسؤولية، الـمُساءَلة، الفاعلية الأخلاقية، المفعولية الأخلاقية، الشفافية، التفسيرية.

---

* تُرجمت هذه المقالة عن الأصل الإنجليزي:

Coeckelbergh, Mark (2020). Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability, Science and Engineering Ethics, 26(4), 2051–2068.

** فيلسوف بلجيكي مُتخصِّص في فلسفة التقنية، أستاذ فلسفة التقنية في قسم الفلسفة بجامعة فيينا، ومعظم أعماله تتركَّز في مجال الذكاء الاصطناعي والأخلاق الرقمية والسلطة الخوارزمية.

*** باحث مصري ومُترجِم مستقل، بكالوريوس صيدلة، جامعة القاهرة. البريد الإلكتروني: dr.tareq.osman@gmail.com، https://orcid.org/0009-0002-2256-9794

تم تسلُّم الترجمة بتاريخ 9/ 3/ 2025م، وقُبِل للنشر بتاريخ 25/ 5/ 2025م.

للاقتباس: كوكِلبِرج، مارك (2020). "الذكاء الاصطناعي وعزو المسؤولية والتسويغ العِلاقي لمطلب تفسير الأفعال والقرارات الآلية"، ترجمة: طارق عثمان، **مجلة الفكر الإسلامي المعاصر (إسلامية المعرفة سابقاً)**، مجلد 32، العدد 111، 261–192. DOI: 10.35632/citj.v31i111.17273

كافة الحقوق محفوظة للمعهد العالمي للفكر الإسلامي © 2026

مُقدِّمة

أضحت أخلاق الذكاء الاصطناعي موضوعاً رائجاً في النقاشات العلمية والنقاشات العامية عن مستقبل التقنية، إثر التقدُّم والنجاح الكبير الذي أحرزه الذكاء الاصطناعي عامة، وتطبيقات التعلُّم الآلي بوجه خاص في الآونة الأخيرة. وقد تراوحت الإسهامات العلمية بين النظر في مسألة الذكاء الاصطناعي الخارق (Bostrom, 2014) ومناقشة العديد من التقنيات والظواهر -ستتاح في المستقبل القريب على الأرجح مقارنةً بالذكاء الاصطناعي الخارق- مثل السيّارات الذاتية القيادة (Helveke & Nida-Rümelin, 2015; Nyholm & Smids, 2016). أمّا النقاشات العامية فإنَّها تدور -من باب المثال- حول مستقبل العمل البشري، والخصوصية الفردية. ومن الأمثلة على ذلك في قطاع النقل، النقاشات القريبة العهد التي دارت حول أتمتة الطائرات،[1] عَقِب تحطُّم طائرة بوينج 737 ماكس بإندونيسيا، في شهر تشرين الأوَّل (أكتوبر) عام 2018م، بسبب خلل في البرمجة، أخفق الطيّارون في معالجته في الوقت المناسب.[2]

ومن الموضوعات الـمُهِمَّة التي تتطلَّب مزيداً من النقاش في هذا الصدد، المسؤولية. فنظراً إلى أنَّ الذكاء الاصطناعي قد مَكَّن المجتمع من أتمتة كثير من المهام آليّاً على نطاق أوسع من قبلُ؛ فإنَّه يحقُّ لنا طرح السؤالين الآتيين: مَنِ المسؤول عن منافع استعمال هذه التقنية ومضارِّها؟ وإذا كان من الواجب تناول هذه المشكلة على نحوٍ استباقي في نطاق كلٍّ من التقنية والسياسة، فما معنى تطوير ذكاء اصطناعي مسؤول؟ إنَّنا لا ننظر إلى ذلك بوصفه مسألة فلسفية شائقة فحسب، بل نَعُدُّه مسألة علمية خالصة في حاجة مُلِحَّة إلى مزيد من الدراسة. ولهذه المشكلة أيضاً صِلَة بأخلاق تقنيات الأتمتة (الأُخرى)، مثل الإنساليات (الروبوتات)، وإنْ كانت بعض المشكلات تختصُّ ببعض أشكال الذكاء الاصطناعي دون غيرها، مثل مشكلة الصندوق الأسود.

---

[1] لفظ (أتمتة) يعني التشغيل الآلي، وهو ترجمة للكلمة الإنجليزية (automation) التي تعني إنجاز المهام التي يؤدّيها البشر عادةً باستخدام تقنيات آليَّة ذكية، بعيداً عن أيِّ تدخُّل بشري (الـمُترجِم).

[2] https://www.theguardian.com/world/2019/mar/20/lion-air-pilots-were-looking-at-handbook-when-plane-crashed.

تُركِّز هذه الدراسة على مسألة عَزْو المسؤولية في تقنيات الذكاء الاصطناعي التي تُستخدَم في أتمتة أفعال البشر والقرارات التي يَتَّخِذونها بوجه عام، عِلْماً بأنَّ هذه المسألة تلفت انتباهنا إلى مشكلة لا تتعلَّق فقط بالطرف الفاعل المسؤول، وإنَّما تختصُّ بالطرف المفعول به أيضاً؛ أيِ الطرف الذي تقع عليه آثار أفعال الطرف الأوَّل وقراراته.

بدايةً، لا بُدَّ من التمييز بين شرطين أرسطيين للمسؤولية؛ أوَّلهما يتعلَّق بالسيطرة والحاجة إلى تحديد الفاعل المسؤول، وثانيهما يُعْنى بفحص معرفة (عِلْم) الفاعل المسؤول. فلنفترض أوَّلاً أنَّ البشر هم المسؤولون اليوم، حتّى لو اكتسبت تقنيات الذكاء الاصطناعي قَدْراً أكبرَ من الفاعلية؛ إذ يتعذَّر على غيرهم أنْ يكونوا هم المسؤولين. صحيحٌ أنَّ تقنيات الذكاء الاصطناعي تَتَّصِف بالفاعلية، لكنَّها لا تستوفي الشروط الـمُعتمَدة للفاعلية والمسؤولية الأخلاقيتين. وبالرغم من ذلك، يعترض عَزْو المسؤولية وتوزيعها مُعوِّقات عِدَّة، ليس مَردُّها فقط إلى مشكلة الأيادي الكثيرة، وإنَّما تُعْزى أيضاً إلى ما نُسمّيه مشكلة الأشياء الكثيرة. ومن الـمُهِمِّ كذلك مراعاة البُعْد الزمني عند مناقشة مسألة الأسباب. بعد ذلك، يتعيَّن علينا تناول مشكلات المعرفة الـمُتعلِّقة ببعض تطبيقات التعلُّم الآلي، التي تُفْضي إلى جملة من العواقب على المسؤولية. وبوجه عام، تُناقَش هذه المشكلات من جهة الشفافية، ومن جهة تفسيرية الأفعال والقرارات الآليَّة.[3] ولكنْ، في مُقابِل أغلب النقاشات العلمية والنقاشات العامية التي تعرض لهذه المشكلات، لا بُدَّ لنا من تناول مسألة التفسيرية بصورة أُخرى؛ إذ لا ينبغي لنا إغفال الطرف الذي تقع عليه مَغَبَّة هذه الأفعال أو القرارات الآليَّة؛ أيْ أولئك الذين يُحمَّل الفاعلون الأخلاقيون المسؤولية عنهم. فمن منظور عِلاقي، لا يوجد في علاقة المسؤولية فاعلون أخلاقيون فقط، وإنَّما يوجد مفعولون أخلاقيون أيضاً.

[3] سنكتفي في هذه الدراسة بلفظ (تفسيرية) في مُقابِل مصطلح (explainability)، عوضاً عن "تفسيرية الأفعال والقرارات الآليّة"، على سبيل الاختصار، وليُفهَم من السياق أنَّ المقصود هو تفسير الأفعال والقرارات الصادرة عن تطبيقات الذكاء الاصطناعي والتعلُّم الآلي تحديداً، بإظهار عِلَلها أو الأسباب الكامنة وراءها (الـمُترجِم).

ثمَّ سنُحاجِج بأنَّ تسويغ مطلب التفسيرية ليس قائماً على شرط المعرفة فحسب (أيْ أنْ يكون الفاعل الأخلاقي على عِلْم بما يفعله)، بل يجب إقامته على أساس المطلوب من هذا الفاعل أخلاقياً؛ وهو تعليل فعله أو قراره لأولئك الذين يُسأَل عنهم؛ أيِ الذين تقع عليهم آثار هذا الفعل أو القرار. وفي الختام، سنتطرَّق إلى المضامين الأُخرى لمصطلح "المسؤولية العِلاقية"، وهي: المسؤولية الجماعية، والوجه الاجتماعي للتفسير والتعليل.

بادئ ذي بَدْء سنسلك نهجاً أرسطياً في تناول مسألة عَزْو المسؤولية، بالرغم من عِلْمنا أنَّه – كغيره من نُهُجٍ تناولها الـمُعتمَدة أو الشائعة– نهج محدود ومُقيَّد. فعلى سبيل المثال، يفترض هذا النهج أنَّ التقنيات ليست سوى أدوات ووسائل، وينبني على ذلك أنَّ البشر وحدهم هم الفاعلون المسؤولون. ويميل هذا النهج كذلك إلى التركيز على المسؤولية الفردية، ولا سيّما على الطرف الفاعل المسؤول، مُتجاهِلاً الطرف الآخر الذي تقع عليه أفعال الطرف المسؤول. ولهذا سنأتي على ذكر بعض الافتراضات التي يتبنّاها هذا النهج (مثل: عَدِّ التقنية مُجرَّد أداة فقط، وقَصْر المسؤولية على البشر) من دون تفصيل القول فيها؛ نظراً إلى ضيق الـمَقام، ودراستنا إيّاها بالتفصيل في موضع آخر، لكنَّنا سنأتي على نقد اثنين منها (الفردية، والفاعل المسؤول)، مُدرِكين –في الوقت نفسه– أنَّ المنظور العِلاقي يتيح لنا سلوك نهج أكثر تشدُّداً وموضوعيةً من الناحية الاجتماعية، وأكثر مراعاةً للطرف الذي تقع عليه أفعال الطرف المسؤول، مقارنةً بالنُهُج الـمُعتمَدة في هذا الجانب.

لا شكَّ في أنَّ لمناقشة المفاهيم والإشكالات القانونية نفعاً كبيراً في تناول المسائل الفلسفية والمسائل العلمية الـمُتعلِّقة بالمسؤولية عن الذكاء الاصطناعي. فعلى سبيل المثال، قدَّم تِرنر في هذا الصدد عرضاً مفيداً لإشكالات المسؤولية الـمُتعلِّقة بعدد من الصكوك القانونية في قانون الأحوال الشخصية والقانون الجنائي (Turner, 2018). وهذا موضوع شائق ومُهِمٌّ جداً، بيد أنَّنا لن نتطرَّق إليه في هذه الدراسة.

فنحن سنضرب على مدار الدراسة أمثلة من قطاع النقل والمواصلات في الـمَقام الأوَّل، وتحديداً السيّارات والطائرات الذاتية القيادة، عِلْماً بأنَّ ما سنطرحه هنا ينطبق على جميع استعمالات

الذكاء الاصطناعي خارج هذا القطاع، بل ينطبق على جميع أشكال تقنيات الأتمتة وتطبيقات الذكاء الاصطناعي؛ فليس مَرامنا من هذه الدراسة تسليط الضوء على حالات مُعيَّنة، وإنَّما الإسهام في بناء إطار نظري للتفكير في المسؤولية الـمُرتبِطة باستعمال الذكاء الاصطناعي وغيره من تقنيات الأتمتة الـمُتقدِّمة، إلى جانب التركيز على تقنيات التعلُّم الآلي تحديداً.

أوَّلاً: إشكالية "عَزْو المسؤولية": مَنِ الفاعل المسؤول؟

من الـمُهِمِّ أنْ نكون قادرين على عَزْو المسؤولية حينما يقع الذكاء الاصطناعي في خطأ ما. فمثلاً، في حالة السيّارات والطائرات الذاتية القيادة باستعمال الذكاء الاصطناعي، لا بُدَّ من طرح السؤال الآتي: إذا تسبَّب نظام التشغيل الآلي في إحدى السيّارات أو الطائرات الذاتية القيادة في حادثة، فمَنِ المسؤول؟ وإلى جانب هذا السؤال عن المسؤولية التي تلي وقوع الحادث، يجب طرح السؤال الاستباقي الآتي: كيف يُمكِن تشكيل التقنية ومحيطها الاجتماعي على نحوٍ يضمن استعمال الذكاء الاصطناعي بمسؤولية؟ فعلى سبيل المثال، كيف يُمكِن تطوير سيّارات وطائرات، وتضمينها أنظمة من الذكاء الاصطناعي تساعد على منع وقوع الحوادث؟ سنُركِّز هنا على مناقشة مسألة عَزْو المسؤولية الـمُرتبِطة باستعمال الذكاء الاصطناعي، بيد أنَّ لهذا النقاش تَبِعات مُتعلِّقة بتطوير الذكاء الاصطناعي على نحوٍ مسؤول.

وبوجه عام، حينما يفعل البشر ويُقرِّرون، فإنَّ هذه الفاعلية ترتبط بالمسؤولية؛ إذ تُؤثِّر أفعالهم وقراراتهم في العالَم وفي الآخرين، ومن ثَمَّ يتحمَّلون المسؤولية عنها. غير أنَّ تعيين الشخص الذي يجب عَزْو المسؤولية إليه ليس أمراً سهلاً في جميع الأحوال؛ إذ قد لا يَتَّضِح -على وجه التحديد- مَنِ الذي تسبَّب في مآلات الأحداث (الضارَّة أو النافعة)، وحتّى لو تبيَّن الشخص المسؤول عن هذه الأحداث، فرُبَّما يكون مُكرَهاً عليها، أو لم يكن على عِلْم بحقيقة ما فعله وقتئذٍ. وهذا يقودنا إلى جملة من الأسئلة، أهمُّها: كيف يُمكِن للأفراد أو للمجتمع عَزْو المسؤولية على نحوٍ صائب؟ ومتى يكون ذلك؟ وإلى مَنْ تُعْزى هذه المسؤولية؟ وما مقتضى ذلك على عَزْو المسؤولية في حالة الذكاء الاصطناعي؟

إنَّ من الطرائق المفيدة لتناول مسألة عَزْو المسؤولية البَدْءَ بشروط المسؤولية. فمنذ عهد أرسطو، كان يوجد شرطان مأثوران -على أقل تقدير- لعَزْو المسؤولية عن فعل ما، وهما: شرط السيطرة، وشرط المعرفة (Fischer & Ravizza, 1998; Rudy-Hiller, 2018). ومن ثَمَّ، يكون المرء مسؤولاً عن فعل ما إذا قام به حقّاً (أيْ كان هو الـمُتسبِّبَ في الفعل (مُحدِثه)، ومُسيطِراً عليه بقَدْرٍ كافٍ)، وكان على عِلْم أو دراية بما فعله. وفي هذا السياق، حاجج أرسطو في كتاب "الأخلاق إلى نيقوماخوس" بأنَّ عَزْو المسؤولية يستلزم صدور الفعل عن الفاعل، وألّا يكون هذا الفاعل جاهلاً بما يفعله (Aristotle, 1984: 1109b30-1111b5).

ولكنْ، إلى أيِّ حَدٍّ ينطبق هذان الشرطان (الفاعلية/ السيطرة، والمعرفة) على حالة استعمال تقنيات الذكاء الاصطناعي؟

فلننظر بدايةً في الشرط الأوَّل، وهو شرط السيطرة، ولنبحث في تفاصيله وأحواله بشيء من التروّي والتدبُّر. فقد أظهرت الدراسات الحديثة التي تناولت هذا الموضوع كثيراً من التشديد على مسألة السيطرة الفردية، مُؤكِّدةً أنَّ الفاعل لا يكون مسؤولاً عن فعل أو قرار إلّا إذا كان مُسيطِراً على هذا الفعل أو القرار (Fischer & Ravizza, 1998; McKenna, 2008). وبوجه عام، فإنَّ للنِّيَّة والحرية صِلَة وثيقة بهذا الشرط؛ إذ شدَّد أرسطو -كما ذكرنا آنفاً- على أنَّ المسؤولية تستلزم فعل الفاعل لفعله طوعاً وليس كَرْهاً. فمثلاً، إذا تسبَّب سائق ما في حادثة، فإنَّ السؤالين اللذينِ يتعيَّن علينا طرحهما، ويتعلَّقان بشرط السيطرة، هما: هل تسبَّب السائق حقّاً في الحادثة؟ وهل كان فعله طوعياً؟ ثمَّ يأتي السؤال الـمُتعلِّق بشرط المعرفة، وهو: هل كان بوسع هذا السائق أنْ يختار فعلاً آخرَ؟ (أيْ، هل كان لديه بديل لِما وقع؟) وهل كان على عِلْم بما فعله؟ (أيْ، هل كان يُرسِل رسالة نصية طوعاً، لكنَّه كان جاهلاً بعواقب ذلك مثلاً؟).

إنَّ هذه الأسئلة تقوم على افتراض أنَّ الفاعل المسؤول بشري. ولكنْ، ماذا سيحدث إذا كان الفاعل جزءاً لا يتجزَّأ من أنظمة الذكاء الاصطناعي؟ ولا سيّما أنَّ تقنيات الذكاء الاصطناعي

أصبحت تمتاز بقَدْر أكبر من الفاعلية. ومن ثَمَّ، فقد يجد المرء نفسه مدفوعاً إلى طرح السؤال الآتي: هل يُمكِن لهذه التقنيات أنْ تكون فاعلاً مسؤولاً؟ فيجاب عن المطروح بأنَّ مبحث أخلاق الحوسبة والإنساليات يَحْفل بنقاشات مستفيضة عن هذا الموضوع وعن مسألة الفاعلية الأخلاقية ذات الصِّلَة (Sullins, 2006; Johnson, 2006; Sparrow, 2007; Stahl, 2006; Moor, 2006; Wallach & Allen, 2009; Floridi & Sanders, 2004; Coeckelbergh, 2010). والحقيقة أنَّه يُمكِن سَوْق حُجَج مُشابِهة [للحُجَج المطروحة في ذلك المبحث] لتأييد القول بأنَّ تقنيات الذكاء الاصطناعي قد تكون فاعلاً مسؤولاً، أو لدحض هذا القول. كذلك يُمكِن للمرء أنْ يتبنّى، في تناوله إشكالات السيطرة، نهجاً هجيناً يرى فيه المسؤولية مُوزَّعةً عبر شبكة من البشر والآلات (Gunkel, 2018a). فعلى سبيل المثال، حاجج هانسون، مُتأثِّراً ببرونو لاتور، عن عَزْو المسؤولية إلى السيبروج (كائن هجين؛ بشري وآلي) وشبكات الفاعلين (Hanson, 2009).

تُعَدُّ هذه الحلول الـمُقترَحة مسارات شائقة، لكنَّني لن أسلكها هنا؛ إذ إنَّنا سنُسلِّم بما يأتي:

1. تشكيل التقنية الفعل البشري على نحوٍ يتجاوز دورها الأداتي الـمُجرَّد.

2. اتِّصاف الأفعال الصادرة عن تقنيات الذكاء الاصطناعي بأهمية أخلاقية، من دون المغالاة والقول كما قال فربيك (Verbeek, 2006) إنَّ الأخلاق نفسها شأن تقني.[4]

3. احتمال ظهور تقنيات الذكاء الاصطناعي الـمُتقدِّمة بمظهر الفاعل المسؤول (Coeckelbergh, 2009).

وبالرغم من ذلك، سنفترض أنَّ تقنيات الذكاء الاصطناعي لا تستوفي الشروط الـمُعتمَدة للفاعلية الأخلاقية، ومن ثَمَّ شروط المسؤولية، مثل الحرية والوعي؛ ما يعني أنَّه لا يجوز تحميلها المسؤولية. وفي ما يختصُّ بشرطي المسؤولية الأرسطيين، فإنَّنا نرى أنَّه من غير المعقول أنْ نطلب إلى

[4] مقصد فربيك من ذلك هو أنَّ التقنية تُسهِم في تشكيل أفعال البشر، ومن ثَمَّ تُشكِّل أخلاقهم، بتقديمها أجوبة مادية عن السؤال الآتي: كيف أتصرَّف؟ ولهذا قال فربيك إنَّ الـمُهندِسين حين يُصمِّمون البرامج التقنية، فإنَّهم يُمارِسون "الأخلاق بوسائل أُخرى"، ويُرسِّخون الأخلاق في شكل تقنية مادية (الـمُترجِم).

الذكاء الاصطناعي أنْ يفعل شيئاً طواعية، ومن دون جَهالة؛ لأنَّه يفتقر إلى لوازم ذلك؛ إذ لا يُمكِن حقّاً للذكاء الاصطناعي أنْ يفعل ما يفعله بحرية (أيْ يتمتَّع بإرادة حُرَّة)، ولا أنْ يَعْلم (أيْ يكون مُدرِكاً) ما يفعله. ولن ننشغل هنا ببيان إذا كان مَردُّ ذلك إلى أسباب فلسفية أو أنَّه مُجرَّد تواضع (اتِّفاق) اجتماعي يتطلَّب مزيداً من التمحيص النقدي، وإنَّما سنكتفي بالتسليم به على سبيل الافتراض، حتّى نُفسِح المجال أمام غيره من أُطروحات هذه الدراسة.

وإذا صَحَّ هذا الافتراض، فلن يتبقّى أمامنا سوى خيار واحد، هو تحميل البشر مسؤولية ما يفعله الذكاء الاصطناعي. وحتّى لو كان بمقدور (بعض) تقنيات الذكاء الاصطناعي أنْ تفعل أو تُقرِّر (أيْ تتمتَّع بفاعلية عامة)، فإنَّها تفتقر إلى مُؤهِّلات الفاعلية الأخلاقية تحديداً؛ لذا ينبغي أنْ تظلَّ المسؤولية عن أفعالها وقراراتها، وهي أفعال وقرارات مُوكَلة لها من البشر، مُلْقاة على كاهل أولئك البشر الذين يبتكرون هذه التقنيات، ويُكثِرون من استعمالها. وهذا هو الـمُتَّبَع في النُّظُم القانونية الحالية في ما يخصُّ الأشياء غير البشرية على وجه العموم. وحتّى إذا كانت بعض الـمُنظَّمات تحظى بشخصية اعتبارية (قانونية)، فإنَّ "مسؤوليتها" هذه لا تزال تُرَدُّ إلى مسؤولية البشر القائمين عليها. وبناءً على ذلك، فإنَّنا نفترض انتفاء صفة المسؤولية عن الفاعلين التقنيين؛ أيْ إنَّه لا يُمكِن لهؤلاء أنْ يكونوا مسؤولين أو غير مسؤولين. أمّا البشر فقد يكونون مسؤولين، بل يجب تحميلهم مسؤولية ما يفعلونه ويُقرِّرونه؛ نظراً إلى تطويرهم أنظمة الذكاء الاصطناعي وتقنياته الـمُتعدِّدة، ومواظبتهم على استعمال هذه التقنيات.

غير أنَّ تبنّي هذا الافتراض لا يَحلُّ إشكالية "عَزْو المسؤولية"؛ لأنَّ القول إنَّ البشر وحدهم هم الفاعلون المسؤولون يثير تساؤلات عِدَّة، أهمُّها: هل يُمكِن للبشر أنْ يتحمَّلوا دائماً مسؤولية استعمال الذكاء الاصطناعي؟ وأيُّ أصناف البشر يتحمَّلون هذه المسؤولية تحديداً؟ دعونا نعرض بعض إشكالات "عَزْو المسؤولية" الـمُتعلِّقة بشرط السيطرة/ الفاعلية:

أ. الاختلاف على مَنْ تقع عليه مسؤولية الفعل

قد يَرِدُ إلى الذهن سؤال مفاده: ماذا سيحدث إذا فقد البشر السيطرة الكاملة على تقنيات الذكاء الاصطناعي؟ فيجاب عن المطروح بما سمّاه ماتيس (Matthias, 2004) فجوة المسؤولية التي تتعلَّق بآلات تتعلَّم وتفعل ما يُطلَب منها على نحوٍ مستقل، فكيف يُمكِن تحميل البشر المسؤولية حين يفقدون زمام السيطرة الكافية أو السيطرة الكاملة؟ فمثلاً، إذا سلَّمنا الآلات السيطرة على المعاملات المالية، كما في حالة التداول الخوارزمي الفائق السرعة، فكيف يُمكِن للبشر أنْ يكونوا مسؤولين عن هذه المعاملات؟ وإذا كانت القيادة الذاتية تعني عجز البشر عن التدخُّل بسرعة في الوقت المناسب، فمَنْ يتحمَّل المسؤولية عندما تقع حادثة ما؟

والحقيقة أنَّ الأصل في هذه الإشكالية يعود غالباً إلى عامل الوقت، وهو ما يظهر جَلِيّاً في الصراعات والحروب؛ إذ يمتاز نظام الدفاع الآلي عن غيره من الأنظمة بقدرته الفائقة على التصدّي سريعاً لأيِّ هجوم خلال وقت قصير جداً. أمّا في حالة تدخُّل العنصر البشري في هذه العملية فلن يكون بمقدور هذا العنصر الرَّدُّ سريعاً حين يُشَنُّ هجوم مُباغِت بعد إنذار خاطف (أو من دون سابق إنذار). ولهذا يشيع اليوم استعمال منظومات الدفاع الآليَّة أثناء المعارك والحروب (Horowitz & Scharre, 2015). والشيء نفسه ينطبق على أنظمة الدفاع المضادة للصواريخ؛ إذ إنَّها تعتمد في عملها على مبدأ التشغيل الآلي الذي يمتاز بالسرعة الفائقة مقارنةً بتلك الأنظمة التي تُطلَق يدوياً بأمر من القيادة العسكرية؛ إذ يكون قرار إطلاق الأخيرة بطيئاً بعض الشيء. وبالمثل، تُستعمَل برامج حوسبة دفاعية آليَّة للرَّدِّ على الهجمات السيبرانية؛ إذ يُمكِن لهذه البرامج أنْ تتدخَّل على نحوٍ سريع ومستقل عن البشر؛ لحماية الأنظمة الـمُستهدَفة قبل وقوع أيِّ ضرر.

ولكنْ، كيف يُمكِن للبشر أنْ يتحمَّلوا مسؤولية هذه الأفعال والقرارات وقتئذٍ بالرغم من عدم قدرتهم على إلغاء أوامر التشغيل الآليَّة؟ لا شكَّ في أنَّ المرء يقف حائراً أمام هذا الوضع الشائك؛ فلا يدري كيف يُمكِن التعامل معه، ولا يدري إنْ كان لازماً استعمال هذا الضرب من تقنيات الذكاء

الاصطناعي، ولا يدري إنْ كانت الضرورة تقتضي ابتكار تقنيات ذكية تتيح للعنصر البشري التدخُّل لحظة عملية التشغيل الآليَّة. ولكنْ، كيف سيكون الحال لو أنَّ البشر لم يَعُدْ لهم مَوطِئ قَدَم في دائرة اتِّخاذ القرار؟ وماذا سيحدث إذا لم يَعُدْ لمطلب الرقابة البشرية على اتِّخاذ القرار أيُّ معنى، بحيث يتعذَّر على الإنسان التدخُّل فيه بأيِّ حال؟ من الـمُلاحَظ أنَّ استعمال تقنيات الأتمتة يسير في هذا الاتجاه داخل كثير من النطاقات. ففي الطائرات التجارية الكبيرة مثلاً، يُستعمَل الطيّار الآلي في أغلب الأوقات، ولكنْ عند وقوع أيِّ خَطْب، لا يستطيع الطيّارون البشر التدخُّل في الوقت المناسب كما حصل في حادثة سقوط طائرة البوينج 737 ماكس. فهل تُعَدُّ أنظمة الأتمتة والذكاء الاصطناعي هذه مقبولة أخلاقياً أم يجب وضع حدود لتقنيات الأتمتة؟

ب. تعدُّد الأطراف المشاركة في الفعل

بوجه عام، من الصعب تحديد الفرد المسؤول عن الفعل، ولا يُمكِن الجزم إنْ كانت المسؤولية تقع على كاهل فرد واحد فقط أم لا، بل يستحيل تحميل أيِّ فرد بعينه المسؤولية. فقد يوجد أكثر من شخص مسؤول؛ نظراً إلى كثرة الفاعلين الـمُساهِمين في الفعل. ولهذا يشار أحياناً إلى هذه الإشكالية في عَزْو المسؤولية وتوزيعها بما يُسمّى مشكلة الأيادي الكثيرة (van de Poel et al., 2012)، وهي مشكلة تطرحها تطبيقات الذكاء الاصطناعي أيضاً. ويَحْضرنا في هذا الـمَقام حادثة سيّارة الأجرة (أوبر) الذاتية القيادة، التي وقعت في شهر آذار (مارس) عام 2018م في ولاية أريزونا الأمريكية، وتسبَّبت في مقتل سيِّدة راجلة، فمَنِ المسؤول هنا؟ رُبَّما يكون المسؤول المباشر عن هذه الحادثة الأليمة مُطوِّر برنامج التشغيل، أو الشركة التي صنعت السيّارة، أو شركة الأجرة (أوبر) التي استخدمتها، أو الشخص الذي كان يستعملها، أو السيِّدة الراجلة نفسها، بل حتّى الجهة الحكومية الـمَعْنِيَّة (ولاية أريزونا). والظاهر أنَّ برنامج التشغيل قد تعطَّل في هذه الحادثة، شأنه في ذلك شأن تعطُّل برامج التشغيل في حوادث تصادم الطائرات، فتعامل البرنامج مع السيِّدة بوصفها شيئاً غير موجود على أرض الواقع[5]؛ ما قد يُلْقي باللائمة على كلِّ مَنْ شارك في تصميم برنامج التشغيل، وعمل على تطويره.

[5] https://www.extremetech.com/cars/268915-fatal-arizona-crash-ubercar-saw-woman-called-it-a-false-positive.

وفي معرض الرَّدِّ على ذلك، طرح بعض الكُتّاب مفهوم "المسؤولية الـمُوزَّعة"؛ إذ قال تاديو وفلوريدي في ذلك: "غالباً ما تكون آثار القرارات والأفعال المبنية على الذكاء الاصطناعي ناجمة عمّا لا يُحْصى مِن التفاعلات بين كثير من الفاعلين، ومن بينهم الـمُصمِّمون، والـمُطوِّرون، والـمُستخدِمون، وبرنامج التشغيل، ومُكوِّنات الجهاز المادية... إنَّ توزيع الفاعلية يقتضي توزيع المسؤولية" (Taddeo & Floridi, 2018: 751). ورُبَّما يبدو ذلك حَلّاً نظرياً جيِّداً لمشكلة الأيادي الكثيرة، ولكنَّ التسليم بتوزُّع المسؤولية في حالة الذكاء الاصطناعي لا يَحُلُّ هذه المشكلة العملية؛ كيف نعزو المسؤولية؟ ذلك أنَّ المرء قد لا يعرف جميع الـمُساهِمين والأطراف المشاركة في عملية التفاعل. يضاف إلى ذلك وجود العديد من الصعوبات التي أبرزها:

– مشاركة كثير من الأطراف في الفعل، من دون تحميلهم جميعاً القَدْر نفسه من المسؤولية. وهذا ما لمسناه في برنامج المحادثة "تاي" في موقع التواصل الاجتماعي "تويتر" (إكس حالياً)، وهو برنامج أنتجته شركة مايكروسوفت عام 2016م، وأثار كثيراً من الجدل؛ لإصداره تعليقات مُتعصِّبة عِرْقياً ومُناهِضة للنساء في محاوراته مع الـمُستخدِمين؛ ما أوجب إغلاقه. ففي هذه الحالة، يوجد كثير من الأطراف الفاعلة: الـمُبتكِرون، والـمُصمِّمون، وشركة مايكروسوفت، والـمُستخدِمون. وبالرغم من ذلك، فقد عزا سواريز–جونـزالو وزملاؤه –في تحليلهم لهذه الحالة– المسؤولية في الـمَقام الأوَّل إلى الـمُصمِّمين والـمُبتكِرين الذين قرَّروا استخدام البرنامج في موقع "تويتر"، لا إلى مُستخدِمي هذا الموقع الذين تحاوروا مع البرنامج (Suárez-Gonzalo et al., 2019). والحقيقة أنَّنا لا نجد مُسوِّغاً كافياً يجعل الـمُستخدِمين أقل تحمُّلاً للمسؤولية في هذه الحالة، ولكنَّ الأمر الذي لا مِراء فيه هنا هو أنَّ توزُّع المسؤولية لا يُحتِّم دائماً تحميلها لجميع الأطراف بالتساوي.

– تعذُّر عَزْو المسؤولية؛ لأنَّ طرفاً أو أكثرَ قد يُقلِّل من شأن مساهمته؛ لتعزيز مصلحته، والتهرُّب من المسؤولية. ومن ثَمَّ، فقد أصاب سواريز–جونـزالو وزملاؤه في التشديد على الطبيعة الاجتماعية لِـما حدث، وفي تحديد بعض المصالح الـمُؤثِّرة، ودور وسائل الإعلام في حالة برنامج المحادثة الذي أشرنا إليه آنفاً؛ إذ رأوا أنَّ الخطاب الإعلامي في هذه الحالة قد خدم مصالح شركة

مايكروسوفت (Suáres-Gonzalo et al., 2019: 8). وبالمثل، فإنَّ قطاع النقل يضمُّ كثيراً من الأطراف الفاعلة، ورُبَّما يكون لكلِّ طرف منها مصالح مختلفة تجعله يُقلِّل من شأن دوره في حادثة طيران ما أو حادث سيّارة ذاتية القيادة مثلاً، وهو ما يجعل عَزْو المسؤولية أمراً عسيراً من الناحية العملية.

– تحديد الأطراف المشاركة في الفعل، وتعرُّف زمن حدوث هذا الفعل. ولهذا تسعى التحقيقات في حوادث الطيران –مثلاً– إلى تحديد ما حدث، وزمن حدوثه، والـمُتسبِّب فيه على وجه التحديد. وهذا يأخذنا إلى الإشكالية التالية.

ت. ارتباط مسألة عَزْو المسؤولية وتوزيعها بالبُعْد الزمني

لا بُدَّ من تحديد المسؤول عن الفعل، سواء أكان فرداً واحداً أم مجموعةً من الأفراد. وكذلك تحديد زمن وقوع الفعل ومكانه. وهذا ينطبق على استعمال العديد من التقنيات الذكية، ويشمل عدداً من العمليات الـمُحدَّدة، كما في حالة حدوث مشكلة ما في برمجة الطيّار الآلي؛ إذ يتعيَّن تحديد المسؤول عن هذه المشكلة، وتعرُّف وقت حدوثها في مقصورة الطائرة. والحقيقة أنَّ مسألة عَزْو المسؤولية قد ترتبط ارتباطاً وثيقاً بتطوير تقنيات الذكاء الاصطناعي، عِلْماً بأنَّ استعمال هذه التقنيات وتطويرها رَهْن بسلسلة سببية طويلة من الفاعلية البشرية. ونجد ذلك مُتحقِّقاً في مجال الذكاء الاصطناعي خاصة؛ إذ يوجد لكلِّ برنامج مُعقَّد تاريخ طويل حافل بالتجارب والتعديلات والتحسينات، وهو ما يتطلَّب مشاركة كثير من الـمُتخصِّصين والخبراء في مختلف مراحل تطوير البرنامج، ولا سيّما أنَّ كثيراً من البرامج قد تخضع للاستخدام –أثناء عملية تطويرها– في أكثر من شركة أو مُنظَّمة، بل قد تُستخدَم في أكثر من قسم داخل الشركة أو الـمُنظَّمة نفسها.

وهذا الحال قد ينطبق أيضاً على تقنيات التعلُّم الآلي؛ إذ تستغرق عملية إنتاج قواعد البيانات وانتقائها ومعالجتها وقتاً طويلاً، وهي عملية يشارك فيها كثير من مُهندِسي البرمجة وغيرهم من

الفنيين والتقنيين، وتتمُّ في أماكن وأوقات مختلفة؛ إذ نجد مِنْ هؤلاء مَنْ يُقدِّم البيانات (طوعاً أو كَرْهاً، عن عِلْم أو عن جَهْل)، ومَنْ يجمعها ويُعالِجها، ومَنْ يبيعها، ومَنْ يُحلِّلها، وهَلُمَّ جَرّاً. وفي السياق نفسه، قد تعمل بعض الشركات والمؤسسات التقنية على تطوير برامج ذكاء صناعي خاصة بمجال تطبيقي مُحدَّد، لكنَّ هذه البرامج تُستخدَم لاحقاً في مجالات تطبيقية أُخرى تختلف اختلافاً كاملاً عن المجال الأوَّل. فمثلاً، قد تُستخدَم برامج تعرُّف الوجوه الـمُخصَّصة للأغراض الطبية في أعمال الرقابة وعالَم الجريمة؛ ما يجعلها تأخذ طابعاً استخبارياً شرطياً.

وتجدر الإشارة إلى أنَّ تقنيات الذكاء الاصطناعي الـمُستخدَمة تتطلَّب صيانة دورية مستمرة؛ لكيلا تتعطَّل الأنظمة الـمُعتمِدة عليها أو يتراجع أداؤها، كما في حال أنظمة قيادة الطائرات. ومن ثَمَّ، يتعيَّن على الخبراء والـمُهندِسين والفنيين المشاركة في فحص هذه الأنظمة؛ حتّى يتسنّى عَزْو المسؤولية على نحوٍ صائب. والحقيقة أنَّ عملية الفحص والمتابعة والصيانة قد تكون مُعقَّدة وعسيرة جداً في حالة الذكاء الاصطناعي إذا لم يتوافر سجل يُوثِّق تاريخ التقنية الـمُستخدَمة، أو استحال تعقُّبه؛ إذ سنجد أنفسنا أمام مشكلة في عَزْو المسؤولية. ولهذا تُعَدُّ عملية التعقُّب والمتابعة (Dignum et al., 2018; Mittelstadt et al., 2016) وسيلة مُهِمَّة لإعمال عَزْو المسؤولية، والوفاء بمطلب التفسيرية، وتنفيذ كليهما (European Commission AI HLEG, 2019).

واللافت في موضوع الذكاء الاصطناعي أنَّه لا يقتصر فقط على وجود أيادٍ بشرية كثيرة، وإنَّما يمتدُّ ليشمل ما يُسمّى الأشياء الكثيرة؛ أي الكثير من التقنيات المختلفة. فإلى جانب مختلف البرامج التي تُسهِم في عملية تطوير الذكاء الاصطناعي، توجد أشياء أُخرى عديدة؛ أيْ مصنوعات تقنية مادية. ولهذه الأشياء أهميتها؛ لأنَّها تُسهِم سببياً في الفعل التقني، ومن ثَمَّ تتمتَّع بقَدْر من الفاعلية، لا سيَّما أنَّها تعمل بالتناغم مع كثير من الـمُكوِّنات الأُخرى. فمثلاً، يُمكِن لِمجَسٍّ (جهاز استشعار) مُتعطِّل يرتبط ببرنامج تشغيل إحدى الطائرات أنْ يُسهِم سببياً في تحطُّمها. ولهذا، فمن الـمُهِمِّ تعرُّف بِنية النظام التقني كاملاً، وتحديد المسؤول عن تطوير أجزائه واستعمالها وصيانتها (مثل الـمِجَسِّ)،

واستكناه طبيعة التفاعلات بين هذه الأجزاء. فقد يقال –مثلاً– إنَّ تعطُّل أحد الـمِجَسّات هو الذي تسبَّب في تحطُّم طائرة البوينج 737 ماكس المذكورة آنفاً.[6]

وبالمثل، فإنَّ البُعْد الزمني مُهِمٌّ جداً؛ لأنَّ عملية تطوير برامج التشغيل تستغرق غالباً وقتاً طويلاً، ولأنَّ إنتاج الـمُكوِّنات المادية واستعمالها يشمل كثيراً من المصنوعات، ومن ثَمَّ يحتاج إلى كثير من الـمُطوِّرين والـمُنتِجين وغيرهم. وهذا ما نَلْحظه على الطائرات والسيّارات الحديثة؛ إذ إنَّها تتألَّف من مُكوِّنات كثيرة، تَنْظمها تفاعلات وعمليات عِدَّة. وتأسيساً على ذلك كلِّه، فإنَّ من الـمُهِمِّ تبيُّن كلِّ هذه العمليات والعلاقات والتفاعلات البِنيوية والزمنية؛ حتّى نتمكَّن من عَزْو المسؤولية بصورة صحيحة بعد وقوع الحوادث، وكذلك تطوير التقنيات ومتابعتها وصيانتها؛ تجنُّباً للحوادث والأعطال. ولا نقصد هنا العمليات والعلاقات والتفاعلات بين البشر فحسب، بل العمليات والعلاقات والتفاعلات بين البشر والأشياء من جهة، وبين الأشياء نفسها من جهة أُخرى. وهذا يعني أنَّ مسؤولية التقنية ليست فقط نتاج التفاعل بين البشر وحدهم، وإنَّما هي نتاج تفاعل البشر مع الأشياء، وتفاعل الأشياء بعضها مع بعض.

ومقتضى ذلك أنَّه لا يَتَّضِح دائماً إذا كان الذكاء الاصطناعي وحده هو سبب المشكلة أم لا؛ فقد يُعْزى سببها إلى أحد مُكوِّنات النظام التقني غير الذكية أثناء تفاعل هذا الـمُكوِّن مع الذكاء الاصطناعي، مثل حدوث عطل مُفاجِئ في أحد الـمِجَسّات. وقد يكون سبب المشكلة وجود خلل في التفاعلات (البرمجية أو المادية) بين أجزاء النظام. ومهما يكن من أمر، فليس واضحاً أين تنتهي مسؤولية الذكاء الاصطناعي، وأين تبدأ مسؤولية التقنيات الأُخرى، بل إنَّ كُنْه الذكاء الاصطناعي نفسه ليس واضحاً؛ إذ يُستعمَل مصطلح "الذكاء الاصطناعي" للإشارة إلى الخوارزميات الذكية على وجه العموم. ولكنْ، توجد نقاشات مستمرة عن تعريفه في نطاق العِلْم والفلسفة ووضع السياسات. وبصرف النظر عن هذه النقاشات التي تتناول كُنْه الذكاء الاصطناعي، فإنَّ ما يهمُّنا في

[6] http://syner giain sight s.in/morni ng-brief /lion-air-610-crash -fault y-senso rs-not-fixed .

سياق عَزْو المسؤولية هو أنْ نتبيَّن -على وجه التحديد- مُكوِّنات النظام التقني، وطبيعة ترابطها، وتفاعل بعضها مع بعض، ودورها في حدوث المشكلات، وعلاقتها بالفاعلين من البشر.

ختاماً، فإنَّ الإشكالية الشائكة اليوم تتمثَّل في أنَّ الـمُستخدِم النهائي للذكاء الاصطناعي يُوظِّفه في تسيير شؤونه على نحوٍ طوعي وحُرٍّ، من دون أنْ يعرف كُنْه الذكاء الاصطناعي، بل إنَّه قد لا يعرف -حقيقةً- أنَّه يستعمله من الأساس. وهذا يقودنا إلى الإشكاليات الأُخرى الـمُتعلِّقة بشرط المعرفة.

ثانياً: إشكاليات المعرفة: الشفافية والتفسيرية

يتمثَّل شرط المسؤولية الثاني لأرسطو في المعرفة (Fischer & Ravizza, 1998: 13)؛ أيْ لا بُدَّ للمرء أنْ يكون على عِلْم بما يفعل، أو كما قال أرسطو: "لا بُدَّ ألّا يكون المرء جاهلاً." وعلى وجه التحديد، فقد ميَّز أرسطو بين عدد من أوجه هذه الجهالة؛ إذ قال: "وعليه، قد يكون الإنسان جاهلاً بماهيَّته، وبما يفعله، وبماهيَّة مَنْ (أو ما) يقع عليه فعله، وأحياناً بوسيلة (أداة مثلاً) فعله، وبغاية ما يفعله (للحماية مثلاً)، وبطريقته في فعله (بلطف أم بعنف مثلاً)" (Aristotle, 1111a3-5). وقد حَمَل رودي-هيلر شرط انعدام الجهالة هذا على معنى الدراية (الإدراك)، وفرَّق بين عدد من ضروبها، مُقتفِياً في ذلك أثر الفلاسفة (التحليليين) المعاصرين: الدراية بالفعل، والدراية بمضمون الفعل الأخلاقي، والدراية بعواقب الفعل، و(بحسب بعضهم) الدراية ببدائل الفعل (Rudy-Hiller, 2018). لكنَّ المثير للانتباه أنَّ أرسطو قد ذكر ضمن أوجه الجهالة وجهاً يتعلَّق بالأداة حين قال: "وأحياناً بوسيلة فعله." ومن الـمُلاحَظ أنَّ الفلاسفة يغفلون غالباً عن هذا الوجه، ولكنَّ ذلك لا ينبغي أنْ يشمل فلاسفة التقنية؛ إذ يتعيَّن عليهم مراعاة هذا الوجه، وأخذه في الاعتبار؛ لأنَّ معرفة المرء طبيعة التقنية التي يستعملها تُعَدُّ أمراً مُهِمّاً لعَزْو المسؤولية.

وبوجه عام، يُمكِن للنقاشات التي تتناول إشكاليات المعرفة الـمُتعلِّقة بالذكاء الاصطناعي أنْ تهتدي بهذه التحليلات الفلسفية للجهالة. ولكنْ، كيف يُمكِن لمُستخدِمي الذكاء الاصطناعي وأولئك الذين يُطوِّرونه أنْ يعرفوا ما يفعلونه؟

لقد جَرَتِ العادة أنْ يعرف الـمُبرمِجون ومُستخدِمو تقنيات الذكاء الاصطناعي ما يريدونه من الذكاء الاصطناعي، أو يعرفون -بالأحرى- ما يريدون من الذكاء الاصطناعي أنْ يفعله لهم؛ إنَّهم يعرفون الهدف، ويُدرِكون جيِّداً العواقب المقصودة (الناتجة من الفعل) أو الغاية -بحسب اصطلاح أرسطو- لكنَّهم لا يعرفون -بالضرورة- العواقب غير المقصودة لِما يفعلونه، ولا يعرفون كذلك المضمون الأخلاقي لهذا الفعل. فهم قد لا يعرفون -مثلاً- أنَّه يوجد تحيُّز في قاعدة البيانات التي يستعملونها أو في الخوارزميات، وقد لا يعرفون تحيُّزاتهم الشخصية التي تسري في تصميمهم للخوارزميات، والتي تعمل على توجيهها. كذلك قد يجهلون وجود أيِّ مضمون أخلاقي لِما يفعلونه. والشيء نفسه ينطبق على الخبير البشري؛ فقد يكون مُتحيِّزاً أيضاً، وقد تكون الخوارزميات أقل تحيُّزاً منه كما قال سنشتين (Sunstein, 2018)، ويُمكِن للخوارزميات أنْ تساعد كذلك على رصد التفريق الـمُجحِف كما قال كلينبرج وزملاؤه (Kleinberg et al., 2019)، لكنَّ ما يعنينا هنا هو حالات الجهالة الـمُوشَّحة بالتحيُّز التي تزيدها التقنية أو تعمل على ديمومتها.

ويضاف إلى ذلك كلِّه أنَّ الـمُبرمِجين ومُستخدِمي تقنيات الذكاء الاصطناعي قد لا يعرفون -على وجه التحديد- العواقب التي ستصيب مَنْ سيتأثَّرون بالخوارزميات. فعلى سبيل المثال، قد يُحرَم شخص ما من الحصول على قرض مالي؛ لأنَّه يعيش فقط في حَيٍّ مُعيَّن، ولكنْ توجد تأثيرات كثيرة أُخرى مُحتمَلة، بعضها أقل مباشرةً وأكثر بُعْداً. فمثلاً، قد لا يعلم مُطوِّر إحدى خوارزميات التداول المالي تَبِعاتها على أولئك الذين سيتضرَّرون من رفع أسعار الغذاء الناتج منها. وكذلك قد يُطوِّر شخص ما خوارزمية لتعرُّف الصور، ثمَّ استخدامها في السياق الأكاديمي، أو في القطاع الصحي، لكنَّها قد تُستعمَل لاحقاً لأغراض المراقبة الشرطية. وحتّى أولئك الذين يُطوِّرون تقنيات المراقبة أو يستعملونها، فإنَّهم قد لا يعرفون دائماً آثارها على وجه التحديد، كأنْ تتسبَّب هذه التقنيات في نشر ثقافة الخوف. إذن، فالتقنيات الذكية ليست فقط مُجرَّد أداة؛ إذ إنَّها تُستخدَم في سياق اجتماعي وسياق عِلاقي، لكنَّ مُطوِّري هذه التقنيات ومُستخدِموها لا يُدرِكون ذلك عادةً.

وهذا يقودنا إلى السؤال الآتي: إلى أيِّ حَدٍّ ينبغي للمُطوِّرين أنْ يكونوا على دراية باستخدامات التقنيات الذكية، ومآلات الاستخدام البديلة الـمُحتمَلة لمبتكراتهم؟ قد يقول المرء: إنَّ هذه الدراية جزء من المسؤولية الواقعة على كاهل الـمُطوِّرين، ولكنْ نظراً إلى القيود المشار إليها؛ فقد لا يكفي القول إنَّهم مسؤولون عن معرفة ذلك ببساطة. ومن ثَمَّ، فإنَّ أكثر الأسئلة إثارة للانتباه في هذا الصدد هي: كيف يُمكِن تمكين الـمُطوِّرين والـمُستخدِمين من حيازة هذه المعرفة؟ وكيف يُمكِن فعل ذلك في مؤسسات التعليم والـمُنظَّمات والشركات المختلفة؟ وكيف يُمكِن التغلُّب على تلك القيود؟

وفوق ذلك كلِّه، فقد يكون مُطوِّرو الذكاء الاصطناعي ومُستخدِموه على جهالة بهذه الأداة. صحيحٌ أنَّهم يعلمون كيف تعمل تقنيات الذكاء الاصطناعي بوجه عام، ومن ثَمَّ يعلمون ما يتوسَّلون به لبلوغ هدفهم أو غايتهم -بحسب اصطلاح أرسطو- غير أنَّ ذلك لا يقتضي أنَّهم يعلمون كلَّ شيء عن هذه التقنيات، وعن سلسلة الأفعال والأشياء الـمُرتبِطة بها. وفي مُقابِل ذلك، نجد أنَّ بعض هؤلاء الأشخاص يعلمون عن الأداة التي يستعملونها أكثر ممّا يعلمه بعضهم الآخر. فعلى سبيل المثال، لا يفهم الـمُهندِسون والمديرون -بالضرورة- ما يفعله الإحصائيون (علماء البيانات) أو ما كانوا يفعلونه قبل تشغيل إحدى التقنيات، والعكس بالعكس.

ويُذكِّرنا هذا المثال مَرَّة أُخرى بأنَّ كثيراً من الأفراد يشاركون في تطوير تقنيات الذكاء الاصطناعي واستخدامها، لكنَّ أرسطو وكثيراً من الفلاسفة المعاصرين لم يُدرِجوا سؤال "مع مَنْ؟" في شروط المسؤولية. فكما هو حال جميع الأفعال التقنية، فإنَّ الأفعال التي تتمُّ بالذكاء الاصطناعي تكون غالباً جماعية، أو تُسهِم فيها أيادٍ كثيرة على مدار الوقت. وهذا أمر مُهِمٌّ بالنسبة إلى مسألة المسؤولية؛ فقد لا يعلم مُستخدِمو الذكاء الاصطناعي، بل قد لا يعلم المديرون أو الـمُنظِّمون شيئاً عن الآخرين الذين أسهموا (أو يُسهِمون) في تطويره واستعماله، ولا عن أفعالهم ومقاصدهم؛ ما يجعل كُلّاً من الاستعمال المسؤول وعَزْو المسؤولية وتوزيعها أمراً عسيراً.

ومن جانب آخر، فإنَّ الشفافية مُهِمَّة أخلاقياً للذكاء الاصطناعي، بل لاستعمال الخوارزميات والبيانات بوجه عام (Mittelstadt et al., 2016). وبالرغم من ذلك، فإنَّ جمهور الدارسين مُتَّفِقون على

الطبيعة الإشكالية لأنظمة الذكاء الاصطناعي المعروفة بالصندوق الأسود؛ أيِ الأنظمة المبنية على التعلُّم الآلي والشبكات العصبونية الإلكترونية، مثل: أنظمة الذكاء الاصطناعي الـمُستخدَمة في السيّارات الذاتية القيادة، والتشخيص الطبي الآلي، وبرامج تعرُّف الصور، وتطبيق "تشات جي بي تي" وما شابه.

ففي أنظمة الذكاء الاصطناعي الكلاسيكية التي تعمل بالرموز، نَلْحظ الوضوح في طريقة خلوص النظام إلى قرار مُعيَّن؛ ذلك أنَّها تعتمد -مثلاً- على شجرة اتِّخاذ قرار مُبرمَجَة من طرف الخبراء الـمُتخصِّصين. أمّا في أنظمة الصندوق الأسود المبنية على التعلُّم الآلي فقد لا تكون طريقة اتِّخاذ النظام لقراره أو توصياته واضحة؛ إذ إنَّها تعتمد على عملية إحصائية يعلم مَنِ ابتكرها كيف تعمل بوجه عام، لكنَّ الـمُطوِّرين والـمُستخدِمين والـمُتأثِّرين بالخوارزميات لا يعرفون كيف يخلص النظام إلى اتِّخاذ قرار مُعيَّن يتعلَّق بشخص ما؛ إنَّهم لا يستطيعون تفسير عملية اتِّخاذ القرار بكامل خطواتها، ولا الكشف عنها.

وهذا الافتقار إلى الشفافية والتفسيرية أمر مُشكِل أخلاقياً؛ لأنَّه -من منظور أرسطي- يجعل مُستخدِمي الذكاء الاصطناعي في جهالة. فمن جانب، لا يعلم مُستخدِمو أنظمة الصندوق الأسود غير الشفّافة ما يفعلونه حين يتوسَّلون بهذه الأنظمة لأداء فعل مُعيَّن أو اتِّخاذ قرار ما. ومن جانب آخر، تتمثَّل جهالتهم في عدم معرفتهم الكافية بالأداة التي يستعملونها (هذا أحد أشكال الجهالة التي نَصَّ عليها أرسطو)، ومن ثَمَّ عدم معرفتهم بما يفعلونه حين يمنحون أحد الأشخاص توصية اعتماداً على هذا الضرب من الذكاء الاصطناعي، أو حين يَتَّخِذون قراراً أو يفعلون فعلاً بوساطته. فعلى سبيل المثال، حين يعمل سائق ما على نقل الرُّكّاب بسيّارة ذاتية القيادة طوال الوقت من دون أنْ يعلم كيف تَتَّخِذ قراراتها، أو حين يستعمل قاضٍ نظام ذكاء اصطناعي لتقدير احتمال عودة مُتَّهَم إلى ارتكاب الجريمة التي يُحاكَم عليها مَرَّة أُخرى، ويُصدِر عليه حُكْماً بناءً على توصية هذا النظام من دون أنْ يعلم كيف خلص إليها النظام؛ فإنَّه يجوز القول: إنَّهما لا يعلمان ما يفعلان بوجه من الوجوه، ويُمكِن عَدُّ ذلك ضرباً مُشكِلاً من الناحية الأخلاقية -على أقل تقدير- من فرط الاعتماد على الذكاء الاصطناعي، إنْ لم يكن محاولة ذميمة للتنصُّل من المسؤولية. غير أنَّ ذَمَّ الأشخاص ولومهم ليس

كافياً -كما ذكرنا آنفاً- للتعامل مع هذه المشكلة؛ إذ يجب اتِّخاذ التدابير اللازمة لتجنيب مُستخدِمي الذكاء الاصطناعي هذا الضرب من الجهالة.

ولكنْ -كما ذكرنا من قبلُ- فإنَّ مُطوِّري البرامج أنفسهم قد لا يعلمون -على وجه التحديد- كيف تخلص أنظمة الصندوق الأسود إلى توصياتها. ولهذا، فإنَّ الجهالة لا تقتصر على مَنْ يتأثَّرون بالذكاء الاصطناعي، الذين قد لا يعلمون أنَّه يفعل شيئاً في خلفية أحد التطبيقات التي يستعملونها، أو حين يطلبون قرضاً من أحد البنوك مثلاً؛ حتّى إنَّهم لا يعلمون كيف يفعل الذكاء الاصطناعي ما يفعله. وبالمثل، فإنَّ الخبراء يعانون حالةً من الجهالة حين يتعلَّق الأمر بأنظمة الذكاء الاصطناعي غير الشفّافة. ومن الـمُلاحَظ في هذا الجانب، وتحديداً في حالة برامج التشغيل عامة، أنَّ بعض الـمُطوِّرين لا يفهمون فهماً كاملاً المقطع البرمجي (الكود) الذي يشتغلون به (قد يكون معظم هذا الكود قد كُتِب منذ زمن طويل على أيدي أشخاص لا يُمكِن تقفّي آثارهم)، وقد لا يعرفون أيضاً الاستخدامات المستقبلية الـمُحتمَلة لهذا المقطع البرمجي، ولا أيّاً من المجالات التي سيُستخدَم فيها. وبعبارة أرسطية، إنَّهم على جهالة "بماهيَّة مَنْ (أو ما) يقع عليه فعله [م]." ولننظر مَرَّة أُخرى في مثال التقنية الطبية التي يُمكِن أنْ تُستعمَل لأغراض عسكرية، في ما يُعرَف بالاستعمال الـمُزدوَج؛ فحتّى لو كان مُطوِّرو الخوارزميات ومُستخدِموها على عِلْم بطريقة عملها، فإنَّهم قد يكونون في عمى مُطبِق عن هذه العواقب البعيدة لأفعالهم.

وهذا ينطبق أيضاً على تقنيات التعلُّم الآلي وغير ذلك من أشكال الذكاء الاصطناعي الـمُتقدِّمة؛ إذ يعاني مُبتكِرو هذه التقنيات ومُستخدِموها ضرباً من الجهالة، مثل جهلهم بالطريقة التي خلصت بها التقنية إلى توصية ما أو قرار مُعيَّن؛ نظراً إلى افتقارها إلى القَدْر الكافي من الشفافية. صحيحٌ أنَّ ذلك قد لا يُلحِق الضرر بأحد إذا اقتصر الأمر على استخدام الذكاء الاصطناعي في لعبة طاولة مثلاً، غير أنَّه قد يكون مُشكِلاً جداً، بل مُغيِّراً للحياة أو مُهلِكاً لها في أحيان أُخرى. فمثلاً، حين يعجز قاضٍ عن تفسير سبب توصية الذكاء الاصطناعي بإطالة أَمَد سجن أحد الـمُتَّهَمين، أو حين يعجز مصرفي عن تفسير سبب توصية الذكاء الاصطناعي برفض إقراض أحد

الأشخاص، أو حين عجز قائد طائرة البوينج 737 ماكس عن السيطرة عليها بسبب استمرار نظام الطيّار الآلي الـمُتطوِّر تقنياً في الهبوط بالطائرة رغماً عنه، وحين عجزت شركتا الطائرات والرحلات الجوية عن تفسير ما حدث لأهالي الضحايا؛ فهذا كلُّه يضعنا أمام مشكلة عويصة في ما يخصُّ الاستعمال المسؤول للذكاء الاصطناعي وغيره من تقنيات الأتمتة، في ظلِّ غياب كامل للشفافية والتفسيرية. وينطبق هذا القول أيضاً على أنظمة الأتمتة الـمُتقدِّمة عامة، سواء أكانت قائمة على التعلُّم الآلي أم لا.

وبالنظر إلى الأمثلة السابقة، فإنَّ التصرُّف على نحوٍ مسؤول يتطلَّب القدرة على تفسير القرارات للأشخاص، وإجابة السؤال المبدوء بـ"لماذا؟" الذي يطرحه شخص تعرَّض للغَبْن حين صدر بحقِّه قرار مُجحِف أو فُعِل به فعل شنيع.

ثالثاً: شمول مُتحمِّلي عواقب المسؤولية أو المسؤولية بوصفها مُساءَلة: نحو نهج أشد عِلاقية

لا يتمُّ الاهتمام غالباً بتسويغ مطلب التفسيرية. ولكن، لماذا تُعَدُّ التفسيرية مُهِمَّة على وجه التحديد؟ رُبَّما يعود ذلك إلى سببين اثنين -على أقل تقدير- الأوَّل: إنَّها مطلوبة للفاعل المسؤول. فلكي يؤدِّي فعلاً على نحوٍ مسؤول، ويَصِحَّ تحميله المسؤولية عن شيء ما؛ لا بُدَّ أنْ يكون على عِلْم بما يفعله (هذا هو شرط المسؤولية الثاني لأرسطو)؛ أيْ إنَّ التفسيرية تُمكِّن الفاعل من أنْ يكون مسؤولاً. أمّا السبب الثاني فهو مُتعلِّق بالطرف الذي يقع عليه الفعل في علاقة المسؤولية. فلو اقتفينا أثر الذين ينتبهون إلى سؤال: "أمام مَنْ يكون الفاعل مسؤولاً؟" (Duff, 2005)، وسلكنا نهجاً عِلاقياً في تناولنا إشكالات المسؤولية، لتبيَّن لنا عدم وجود فاعل مسؤول فقط؛ أيْ ذاك الذي يفعل، ويُفتَرَض أنْ يفعل على نحوٍ مسؤول، وإنَّما يوجد مفعول به أيضاً؛ أيِ الطرف الذي يتأثَّر بفعل هذا الفاعل (مَنْ يقع عليه فعل الفاعل مباشرةً -بعبارة أرسطية- وغيره من الـمُتأثِّرين بهذا الفعل على

نحوٍ غير مباشر)، ويطلب إلى الفاعل أنْ يفعل على نحوٍ مسؤول؛ أيْ إنَّه ينتظر منه (ويسأله) أنَّ يُعلِّل فعله (ومن ثَمَّ يكون قادراً على تعليله). وهذا يتطلَّب الإجابة عن سؤالين اثنين، هما: مَنِ المسؤول عن شيء ما؟ (عَزْو المسؤولية)، ومَنِ المسؤول أمام مَنْ؟

إنَّ المسؤولية ليست أنْ تفعل شيئاً، وأنْ تعلم فقط ما تفعله، وإنَّما تعني أنْ تُسأَل عنه أيضاً؛ إنَّها مسألة عِلاقية وتخاطبية، بل رُبَّما تكون مسألة حوارية؛ فالطرف الذي يقع عليه فعل الفاعل المسؤول هو الطرف الـمُخاطَب في علاقة المسؤولية، لكنَّ هذا الدور يغيب غالباً عن التحليلات الشائعة للمسؤولية بوجه عام، وعن المسؤولية التقنية خاصة.

وتأسيساً على ذلك، فإنَّ التفسيرية ليست مُجرَّد مطلب من الفاعل نفسه (بوصف ذلك شرطاً أرسطياً لتحميله المسؤولية)، وإنَّما يُمكِن تسويغها بالقول: إنَّ الطرف الذي يقع عليه فعل الفاعل المسؤول يطلب من الأخير (أيْ من الفاعل المسؤول) تقديم تفسير لِما فعله به؛ لذا يجب على الفاعل أنْ يكون قادراً على أنْ يُفسِّر للمفعول به عِلَّة فعله أو قراره أو توصيته. فمثلاً، يحقُّ للمرء أنْ يسأل قاضياً عن عِلَّة قراره مثلما يحقُّ للقاضي أنْ يسأل الـمُتَّهَم عن عِلَّة جريرته. وهذا يُؤكِّد حقيقة أنَّ الأفعال والقرارات البشرية يجب أنْ تكون قابلة للتفسير؛ كي يُمكِن تحميل المسؤولية عليها، سواء أُفعِلت في الماضي أم ما تزال تُفعَل في الحاضر.

ولا بُدَّ من التنويه هنا بأنَّ الطرف المفعول به في علاقة المسؤولية لا يقتصر على البشر –بالضرورة– وإنَّما يشمل الحيوانات أيضاً، بل يقال إنَّه يشمل الآلات كذلك، حتّى إنَّه يوجد نقاش عن إمكانية معاملة الآلات بوصفها مفعولاً به أخلاقياً (Gunkel, 2018b; Bryson, 2016). وبذلك يُمكِن للمرء أنْ يكون مسؤولاً أمام مختلف أنواع الموجودات، ما دام أنَّها (قد) تتأثَّر بأفعاله وقراراته. ومن الـمُلاحَظ في هذا الجانب أنَّ المنظور الأخلاقي الذي يضع الطرف المفعول به موضع القلب منه لا يقتصر –بالضرورة– على مسألة التفسيرية، ولا يُركِّز عليها حصراً. وهذا ما نجده في فلسفة ليفناس الأخلاقية (Levinas, 1969)، التي يمضي فيها المطلب الأخلاقي –على ما يبدو– إلى

ما وراء الكلمات، بصدوره من "وجه" الآخر. غير أنَّ ما يهمُّنا هنا هو تعلُّق المسؤولية بضرب صريح من الـمُساءَلة، وبيان مسألة التفسيرية؛ لذا سيقتصر حديثنا على تحليل علاقة المسؤولية بين أشخاص يُمكِنهم طلب التفسيرات، وتقديمها، وفهمها (يُمكِن أيضاً للأشخاص أنْ يطلبوا تفسيراً بالنيابة عن غير البشر أو عن البشر الآخرين).

وعلى هذا الأساس، لو أنَّ فاعلاً بشرياً اتَّخذ قراراً بناءً على توصية الذكاء الاصطناعي، لكنَّه لم يستطع تفسير سبب هذا القرار، فإنَّ ذلك يُعَدُّ مُشكِلاً من جهة المسؤولية لسببين اثنين، هما: فشل الفاعل البشري في أنْ يكون فاعلاً مسؤولاً عن فعله؛ لعدم عِلْمه بما يفعل، وفشل الفاعل البشري في أنْ يكون فاعلاً مسؤولاً أمام مَنْ يقع عليهم فعله، ومَنْ يحقُّ لهم أنْ يطلبوا منه تفسيراً لفعله أو قراره؛ لأنَّهم مُتأثِّرون به. أمّا إذا لم يكونوا أهلاً لطلب التفسير منه، أو لا يسمح لهم مَقامهم بذلك، فيُمكِنهم تفويض غيرهم بهذا الطلب نيابةً عنهم.

ولهذا يتعيَّن على أخلاق الذكاء الاصطناعي أنْ تُعزِّز تطوير أنظمة ذكاء اصطناعي تُمكِّن الفاعلين (مُطوِّرين ومُستخدِمين) من أنْ يكونوا مسؤولين في أمرين اثنين معاً: تحمُّل مسؤولية ما يفعلونه عن طريق الذكاء الاصطناعي، وإمكانية مساءلتهم مِنْ طرف الـمُتأثِّرين بأفعالهم (أو مِنْ طرف مَنْ يُمثِّلهم). وبذلك، فإنَّ الذي يهمُّنا أخلاقياً عند النظر إلى المسؤولية بوصفها مساءلة، ليس التفسيرية بصفتها سِمَة للأنظمة التقنية مثل الذكاء الاصطناعي، وإنَّما التفسيرية بصفتها مساءلة للفاعل الذي يُطوِّر الذكاء الاصطناعي ويستعمله؛ أيْ إنَّه ينبغي عَدُّ التفسيرية التقنية -ما يُمكِن لنظام الذكاء الاصطناعي أنْ يقوله أو أنْ يجيب عنه- شيئاً في خدمة مطلب أخلاقي أكثر عمومية مُلْقى على كاهل الفاعلين البشر؛ وهو أنْ يكونوا قابلين للمساءلة، وقادرين على التفسير. وهذا يتطلَّب توافر أنظمة تقنية شفّافة تتيح لهم تقديم أجوبة عن الأسئلة التي يُحتمَل أنْ يطرحها عليهم الـمُتأثِّرون بهذه الأنظمة. ومن الناحية الأخلاقية، تتبوَّأ العلاقة بين طرفي المسؤولية -في سياق الذكاء الاصطناعي- موضع القلب؛ إنَّها الغاية. ولذلك يتعيَّن علينا أنْ نُناقِش كيفية تعزيزها بوسائل تقنية وأُخرى غير تقنية.

صحيحٌ أنَّ هذا الحديث الأرسطي عن الوسائل والغايات قد يبدو أداتياً (وظيفياً)، غير أنَّه لا يمنع النظر في العواقب غير المقصودة للتقنية، بل يوجِب النظر فيها. فوظيفة التقنية (الظاهرة) هي أنْ تفعل ما يُفترَض لها أنْ تفعله، مثل الطيّار الآلي الذي وظيفته التحليق بالطائرة. ولكنْ توجد وظيفة أُخرىٰ مَخْفِيَّة لهذه التقنية، غير تلك التي أرادها مُطوِّرو التقنية ومُستخدِموها، وهو ما يثير سؤالاً أخلاقياً مفاده: كيف يُؤثِّر هذا النظام التقني في طريقة تحمُّل الفاعلين البشر للمسؤولية عن الطيران الآلي، وأمام أولئك الذين يحقُّ لهم مساءلتهم، وهم المسافرون في الـمَقام الأوَّل، وذووهم وأصدقاؤهم وأحبّاؤهم، والعاملون علىٰ متن الطائرة، ومديرو شركة الرحلات الجوية، وغيرهم؟ وهكذا يُرْسي هذا النهج العِلاقي -الذي يَعُدُّ المسؤولية مُساءَلة- أرضية صُلْبة لعلاقات المسؤولية التي تُعَدُّ ركيزة أساسية في مناقشة المسؤولية عن الذكاء الاصطناعي.

وبناءً علىٰ ذلك، فإنَّ الابتكار التقني المسؤول -ضمن أشياء أُخرىٰ- يُحتِّم دعم تطوير أنظمة ذكاء اصطناعي تُسهِم في تعزيز هذا النهج من المسؤولية العِلاقية؛ إذ يُمكِن لبعض التدابير التقنية أنْ تُسهِم في جعل الذكاء الاصطناعي أكثر قابلية ومرونة من الناحية التفسيرية. ولهذا يَعْمد بعض الباحثين إلىٰ تطوير تقنيات لفتح صندوق الذكاء الاصطناعي الأسود (Samek et al., 2017)، وإيجاد طرائق تتيح تفسير الذكاء الاصطناعي (Adadi & Berrada, 2018). كذلك يُمكِن اتِّخاذ بعض التدابير القانونية التي تُخوِّل الناس الحقَّ في الحصول علىٰ تفسيرات، لا علىٰ معلومات فقط (مثل: فيمَ يُستعمَل الذكاء الاصطناعي؟ وماذا يفعل الذكاء الاصطناعي؟) كما هو الحال في النظام الأوروبي العام لحماية البيانات. كذلك يقتضي الابتكار التقني المسؤول سؤال الذين قد يتأثَّرون بالتقنية عن ضروب التفسيرات التي يحتاجون إليها، فضلاً عن تخيُّل صور تأثُّرهم الـمُحتمَلة؛ فقد يكون هؤلاء علىٰ جَهْل بالتقنية وعواقبها، بحيث يتعذَّر عليهم التكهُّن بها، في ما يُمثِّل مشكلة أُخرىٰ يجب التصدّي لها.

وممّا يزيد من أهمية الإنصات لهؤلاء الذين قد يتأثَّرون بالتقنية، أنَّنا لا نعلم بوضوح -علمياً وفلسفياً- ماهيَّة التفسير الجيِّد، ولا القَدْر المطلوب من التفسير. ولهذا، فإنَّ الإنصات يُمكِّن المرء من

بلوغ هدف الذكاء الاصطناعي الـمُفسِّر على نحوٍ أيسر ممّا هو مُتوقَّع، ومن ثَمَّ لا ينبغي للمرء –على الأرجح– أنْ يطلب من مُطوِّري الذكاء الاصطناعي ومُستخدِميه أكثر ممّا يطلبه من غيرهم؛ إذ بيَّن الباحثون في موضوع التفسير أنَّ الناس لا يطلبون (ولا يُقدِّمون) السلسلة السببية كاملة بوجه عام، وإنَّما ينتقون التفسيرات، ويجيبون عمّا يظنّونه مطلب الآخر؛ فالتفسير مسألة اجتماعية، ولا مِرْية فيه (Miller, 2019). وبالـمُقابِل، فإنَّ الفهم العِلاقي والاجتماعي للتفسير يجعل الذكاء الاصطناعي المسؤول (أيِ الـمُفسِّر) هدفاً عسير المنال؛ لأنَّ التفسير شأن اجتماعي، ومن العسير على الذكاء الاصطناعي أنْ "يفهم" الاجتماعي بما يكفي.

ولكنْ، إذا لم يكن الهدف جعل الآلات تُفسِّر، وطُلِب ذلك من الأشخاص القائمين عليها، الذين يُمكِنهم تفسير الأشياء لغيرهم من البشر، فقد تتاح فرصة للذكاء الاصطناعي الـمُفسِّر تُمكِّنه من النجاح في العمل. وبناءً على ذلك، فإنَّ المسؤولية لا توجِب على المرء أنْ يُفسِّر كلَّ ما يُسهِم في الفعل أو في اتِّخاذ القرار، وإنَّما توجِب عليه أنْ يُقدِّم ما يستطيع من معرفة، وأنْ ينتقي ما قد يحتاج الآخرون إلى معرفته، أو يرغبون في معرفته فقط. أمّا في حالة الذكاء الاصطناعي وغيره من أنظمة الأتمتة الـمُتقدِّمة فلا يُمكِن لذلك أنْ يجري على أيدي الأشخاص إلّا إذا توافرت أنظمة تقنية شفّافة تُمكِّنهم من تفسير الأشياء لغيرهم من الناس، عِلْماً بأنَّ الشفافية الـمُطلَقة هنا غير مطلوبة. كذلك يتعيَّن على هؤلاء الأشخاص أنْ يملكوا من المشيئة والقدرة والعِلْم ما يكفي لتمكينهم من فهم ما قد يطلبه منهم الـمُتأثِّرون بالتقنية، وما قد يسألونهم عنه. والحقيقة أنَّه يُمكِن تيسير ذلك بسؤال الـمُتأثِّرين عن نوعية التفسيرات التي يحتاجون إليها، أو يرغبون فيها حقّاً. ومن ثَمَّ، فإنَّ ما نفترضه هنا هو أنَّ البشر وحدهم قادرون على تفسير أفعالهم وقراراتهم، بل يجب عليهم أنْ يكونوا كذلك. وأنَّ التفسير نفسه شأن اجتماعي وعِلاقي على نحوٍ وثيق.

ولكنْ، هل يحتاج الناس إلى تفسيرات أم إلى تعليلات؟ وهل يُمكِن للتفسيرات أنْ تُحسَب تعليلات؟ ولو أمكن ذلك، فمتى سيتمُّ النظر إليها بوصفها تعليلات؟ يُمكِن أيضاً فهم المسؤولية بصفتها مساءلة من جهة التعليلات، أو تقديم التعليلات على وجه الدِّقَّة. وما ينطبق على التفسيرات

عامة ينطبق على التعليلات في ما يبدو؛ أيْ إنَّ البشر وحدهم قادرون على تقديم التعليلات؛ لذا يقتضي الذكاء الاصطناعي المسؤول أنْ يتولّى البشر هذه المهمة. وعلى هذا الأساس، يتعيَّن على مُطوِّري الذكاء الاصطناعي دعم هذه المهمة البشرية، وذلك بتقديم تعليلات إلى أولئك الذين يطرحون (أو قد يطرحون) أسئلة عن الأفعال التي أُدِّيت والقرارات التي اتُّخِذت بوساطة التقنية. ولو أردنا -بوصفنا مجتمعاً- أنْ نحترم استقلالية البشر وطبيعتهم الاجتماعية، فيجب على مُطوِّري الذكاء الاصطناعي ومُستخدِميه أنْ يجيبوا عن أسئلة الـمُتأثِّرين بالذكاء الاصطناعي، وعن تعليلات الأفعال والقرارات التي وقعت عليهم وتفسيراتها. غير أنَّ تحلّي مُطوِّري الذكاء الاصطناعي ومُستخدِميه بالفاعلية المسؤولة لا يعني فقط تحكُّمهم في ما يفعلونه وعِلْمهم به، وإنَّما يعني أيضاً أنْ يجيبوا، وأنْ يُفسِّروا، وأنْ يُعلِّلوا، وأنْ يتحاوروا. والحقُّ أنَّ مناحي الذكاء الاصطناعي المسؤول والتقنية المسؤولة هذه ليست مسائل هامشية، وإنَّما هي مطالب أخلاقية؛ إذ يجب تعديل التعليم التقني على نحوٍ يُمكِّن مُطوِّري الذكاء الاصطناعي ومُستخدِميه من الإجابة عن السؤال المبدوء بـ"لماذا؟"، ويجب على المجتمع كذلك أنْ يتدبَّر في هذه المسألة: هل يجوز تطوير الأتمتة على نحوٍ كبير جداً بحيث يُمكِن اتِّخاذ القرارات المصيرية بسرعة فائقة كما في حالة الحرب أو حالة التداولات المالية؟ لا شكَّ في أنَّ الجواب أو التعليل أمر مستحيل؛ لأنَّ السؤال المبدوء بـ"لماذا؟" يأتي دائماً بعد فوات الأوان.

وفي الختام، فإنَّ الفهم الأكثر عِلاقية للمسؤولية قد يساعد أيضاً على طرح أسئلة إضافية تتصدّى لمشكلة الأيادي الكثيرة، وتنتقد التشديد على السيطرة الفردية في نظريات المسؤولية الـمُعتمَدة. فعلى سبيل المثال، يُمكِن حَلُّ بعض المشكلات بالإشارة إلى المسؤولية الجماعية إلى جانب توزُّع المسؤولية، عِلْماً بأنَّ المسؤولية الجماعية تعني أمرين اثنين، هما: توزُّع المسؤولية على مجموعة من الأفراد الفاعلين (أيْ إنَّ كلَّ فرد من أفراد المجموعة يتحمَّل المسؤولية)، وتحمُّل فاعل جماعي المسؤولية، مثل: الـمُنظَّمة، والجماعة.

قد يكون الحديث عن الفاعلية الجماعية مفهوماً، لكنَّ الحديث عن المسؤولية الجماعية مُتنازَع فيه بشِدَّة. وهذه مسألة خليقة بأنْ تُفرَد بالدراسة. ويُمكِن للمرء أيضاً أنْ يطرح السؤال الآتي: هل توجد حالات يكون فيها المجتمع كاملاً (أو الـمُواطِنون جميعاً، أو حتّى الثقافة بأكملها) الفاعل والمفعول به في علاقة المسؤولية؟ أيْ إذا كانت خوارزميات الذكاء الاصطناعي مُتحيِّزة، وكان هذا التحيُّز موجوداً في اللغة نفسها (اللغة الإنجليزية مثلاً)، أفلا يعني ذلك أنَّ المسؤولية لا تقع فقط على كاهل مُتحدِّثي هذه اللغة (بوصفهم أفراداً)، وإنَّما تقع أيضاً على كاهل الأُمَّة كلِّها التي تتحدَّث بهذه اللغة وعلى ثقافتها في الماضي والحاضر؟

ولا بُدَّ هنا من استحضار البُعْد الزمني مَرَّة أُخرى؛ ذلك أنَّ تطوُّر لغة مُعيَّنة واستعمالها تاريخياً (يرتبط ذلك بجماعات عِدَّة وتواريخ مختلفة) سيُؤثِّر في مُستخدِميها المعاصرين على نحوٍ مباشر أو غير مباشر، وفي أفعال الذكاء الاصطناعي وعواقبها. وفي هذه الحالة، سيكون من الصعب جداً التصدّي لمسألة كهذه لو قَصَرنا تركيزنا على المسؤولية الفردية. فمثلاً، إذا وجدنا تحيُّزاً ملحوظاً في مُدوَّنة لغوية كاملة، ثمَّ اخترنا برنامج ذكاء اصطناعي يرتكز على التعلُّم الآلي، ودرَّبناه على نصوص موجودة في شبكة الإنترنت، فقد يعيد لنا هذا البرنامج إنتاج التحيُّز التاريخي الجندري مثلاً (Caliskan et al., 2017). ولكي نتصدّى لهذا الضرب من التحيُّز؛ يجب علينا عدم الاكتفاء بالأُطر النظرية التي تُركِّز على الفعل الفردي والمسؤولية الفردية، حتّى لو كان معنى المسؤولية الجماعية غير واضح البتة.

وعلينا أنْ نُلاحِظ أيضاً أنَّ المسؤولية -في هذا السياق الاجتماعي والتاريخي والثقافي- لا يُمكِن غالباً أنْ تكون مُطلَقة، بل لا ينبغي لها أنْ تكون كذلك دائماً؛ فقد يؤدّي الأفراد الذين يستعملون تقنيات عِدَّة (مثل تقنيات الذكاء الاصطناعي) أفعالاً إراديةً، لكنَّ ذلك لا يستمرُّ طويلاً؛ إذ توجد أحوال وأوقات ومؤسسات (وما إلى ذلك) تجعلنا نفقد السيطرة الفردية، بل حتّى السيطرة الجماعية الكاملة، ومن ثَمَّ تُسهِم في تشكيل ما تفعله التقنيات وما تؤول إليه. وإذا كان الحال كذلك، فإنَّ

المسؤولية عن الذكاء الاصطناعي ستكون محدودة نوعاً ما، وسيكون لها بُعْد مأسوي أو تراجيدي (Coeckelbergh, 2011).

وينطوي هذا البُعْد المأسوي على خيار كارثي؛ ففي مُقابِل الحلول التي يطرحها المنظور النفعي ومنظور الواجب الأخلاقي لمعضلات العربة (Trolley dilemmas)، قد يعمل الذكاء الاصطناعي على إيجاد معضلات يستحيل حَلُّها. فمثلاً، قد يؤدّي استعمالنا للسيّارات الذاتية القيادة إلى إيقاعنا في مآزق لا مَخْرج منها، وعلينا عندئذٍ أنْ نتبيَّن هذا البُعْد المأسوي فيها (Sommaggio & Marchiori, 2018)، لكنَّ ذلك لا يعني أنَّه لا يُمكِن (ولا ينبغي) عَزْو المسؤولية في هذه الحالة أبداً، وأنَّه لا يُمكِن (ولا ينبغي) أنْ نفعل شيئاً للتصدّي لهذه المشكلة وأمثالها. فعلى سبيل المثال، إذا وُجِد تحيُّز عِرْقي أو تحيُّز جندري في قاعدة بيانات أو خوارزمية ما، فإنَّه سيكون لهذا التحيُّز وجه جماعي؛ لأنَّه قد يكون نابعاً من اللغة نفسها، وسيكون له وجه مأسوي؛ لأنَّ الأشخاص والمجتمع قد لا يستطيعون أبداً محو هذا التحيُّز الـمُجحِف. غير أنَّه يُمكِن لنا -في سياق استعمال مُعيَّن للذكاء الاصطناعي- إصلاح ذلك -ولو جزئياً على الأقل- عن طريق النظام التقني أو مُستخدِمي الذكاء الاصطناعي. وإذا كان استعمال السيّارات الذاتية القيادة يُفْضي بنا إلى مآزق لا مَخْرج منها، فإنَّ التطوير المسؤول لهذه التقنية يُحتِّم مناقشة هذه المشكلة، ومعالجة بُعْدها المأسوي، وتلمُّس سُبُل التصدّي لها -صراحةً وعلانيةً- مع الـمُستخدِمين جميعاً والمجتمع بأسره، عوضاً عن ترك انطباع مفاده أنَّ التقنية أو النظرية الفلسفية المعيارية يُمكِنها أنْ تُقدِّم الجواب "الصحيح". ولا ينبغي لنا أبداً أنْ نَتَّخِذ من نهج المسؤولية العِلاقي ذريعة للتنصُّل من هذه المسؤولية، وإنَّما يجب علينا فهم أنَّ للفعل المسؤول منحىً تفاعلياً، وتحاورياً، وأنَّه يتطلَّب الشفافية والفهم، بما في ذلك فهم الأبعاد الكارثية للفعل التقني.

خاتمة

قدَّمنا في هذه الدراسة عرضاً لبعض الإشكاليات الـمُتعلِّقة بالمسؤولية عن الذكاء الاصطناعي، وحرصنا فيها أنْ نُركِّز على إشكالية عَزْو المسؤولية وإشكالية المسؤولية بوصفها مساءلة. وقد بدأنا

بالحديث عن النهج الشائع في دراسة المسؤولية، الذي يُركِّز على تمتُّع الفاعلين الأخلاقيين -مِمَّنْ يُطوِّرون الذكاء الاصطناعي، ويستعملونه- بالسيطرة والمعرفة، ثمَّ انتقلنا إلى نهج أكثر عِلاقية، لا يُركِّز فقط على درجة معرفة الفاعلين الأخلاقيين بالعواقب غير المقصودة لأفعالهم، ومضمونها الأخلاقي -مثلاً- ودرجة سيطرتهم عليها، وإنَّما يُركِّز أيضاً على أولئك الذين تقع عليهم عواقب هذه الأفعال؛ وهم الذين يحقُّ لهم أنْ يطلبوا تفسيراً للأفعال التي وقعت عليهم، والقرارات التي اتُّخِذت بحقِّهم عن طريق الذكاء الاصطناعي. وقد شدَّدنا -على مدار الدراسة- على البُعْد الزمني (أيْ سلاسل السببية)، وعمليات إنتاج التقنية، وتاريخ المجتمعات التي تُنتَج فيها التقنية، وتُستعمَل.

وهذه الجوانب لا يُمكِن النظر إليها بوصفها تُمثِّل مسألة شائقة فلسفياً فحسب، بل إنَّها تُمثِّل مسألة مُلِحَّة عملياً؛ لأنَّنا نستعمل اليوم أنظمة الذكاء الاصطناعي حقّاً. ولـمّا كان الذكاء الاصطناعي مُتغلغِلاً في حياتنا اليومية، فإنَّه يجوز لنا القول: إنَّ الناس جميعاً يتحمَّلون عواقب استعمال تقنيات الذكاء الاصطناعي. ولو كان الإطار النظري المطروح هنا معقولاً، لساغ للمُتأثِّرين بالذكاء الاصطناعي -بل حُقَّ لهم- أنْ يطلبوا تطوير تقنيات الذكاء الاصطناعي، ووضع ترتيبات اجتماعية تتيح عَزْو المسؤولية وتوزيعها على نحوٍ صائب، وأنْ يطلبوا من الفاعلين الـمَعْنِيّين التصرُّف بمسؤولية أثناء تطويرهم تقنيات الذكاء الاصطناعي واستعمالهم لها.

ووَفقاً للإطار العِلاقي المطروح هنا، فإنَّ التصرُّف بمسؤولية يُؤكِّد أنَّ المجتمع يستحقُّ أنْ يحظى بخبراء ومُشغِّلي ذكاء اصطناعي يعلمون ما يفعلون، ويُسيطِرون عليه، ويُمكِنهم تبليغ ما يفعلونه بالذكاء الاصطناعي وتفسيره وتعليله، سواء أتعلَّق ذلك بالبشر أم بغيرهم؛ فهم مُستعِدّون كلَّ الاستعداد لفعل ذلك؛ حتّى إنَّهم مُلزَمون بكسب قَدْر أكبر من المعرفة بالعواقب غير المقصودة لِـما يفعلونه، ومضمونها الأخلاقي، وكذا طرائق التعامل مع المشكلات الكارثية. وبذلك يصبح استعمال الذكاء الاصطناعي وتقنياته مسؤولاً كلَّ المسؤولية على هذا النحو؛ لكيلا يكون مآله الفشل والإخفاق.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6,* 52138–52160.

Aristotle. (1984). Nicomachean ethics. In J. Barnes (Ed.), *The complete works of aristotle* (Vol. 2, pp.1729–1867). Princeton: Princeton University Press.

Bostrom, N. (2014). *Superintelligence*. Oxford: Oxford University Press.

Bryson, J. (2016). Patiency is not a virtue: AI and the design of ethical systems. In *AAAI spring symposium series*. Ethical and Moral Considerations in Non-Human Agents. Retrieved from 4, Sept 2018, http://www.aaai.org/ocs/index .php/SSS/SSS16 /paper /view/12686 .

Caliskan, A., Bryson, J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356,* 183–186.

Coeckelbergh, M. (2009). Virtual moral agency, virtual moral responsibility. *AI & SOCIETY, 24*(2), 181–189.

Coeckelbergh, M. (2010). Moral appearances: Emotions, robots, and human morality. *Ethics and Information Technology, 12*(3), 235–241.

Coeckelbergh, M. (2011). Moral responsibility, technology, and experiences of the tragic: From Kierkegaard to offshore engineering. *Science and Engineering Ethics, 18*(1), 35–48.

Dignum, V., Baldoni, M, Baroglio, C., Caon, M., Chatila, R., Dennis, L., & Génova, G., et al. (2018). "Ethics by design: Necessity or curse?" Association for the Advancement of Artificial Intelligence. Retrieved from 21, Jan 2019, http://www.aies-confe rence .com/2018/conte nts/paper s/main/AIES_2018_paper 68.pdf.

Duff, R. A. (2005). Who is responsible, for what, to whom? *Ohio State Journal of Criminal Law, 2,* 441–461.

European Commission AI HLEG (High-Level Expert Group on Artificial Intelligence). (2019). *Ethics Guidelines for Trustworthy AI*. Retrieved from 22, Aug 2019, https://ec.europ a.eu/futur ium/en/aiallia nce-consu ltati on/guide lines #Top.

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.

Floridi, L., & Sanders, J. (2004). On the morality of artificial agents. *Minds and Machines, 14*(3), 349–379.

Gunkel, D. J. (2018a). Mind the gap: Responsible robotics and the problem of responsibility. *Ethics and Information Technology*. https://doi.org/10.1007/s1067 6-017-9428-2.

Gunkel, D. J. (2018b). The other question: Can and should robots have rights? *Ethics and Information Technology, 20*(2), 87–99.

Hanson, F. A. (2009). Beyond the skin bag: On the moral responsibility of extended agencies. *Ethics and Information Technology, 11*(1), 91–99.

Helveke, A., & Nida-Rümelin, J. (2015). Responsibility for crashes of autonomous vehicles: An ethical analysis. *Science and Engineering Ethics, 21*(3), 619–630.

Horowitz, M., & Scharre, P. (2015). An introduction to autonomy in weapon systems. CNAS Working Papper. https://www.cnas.org/publi catio ns/repor ts/an-intro ducti on-to-auton omy-in-weapo n-syste ms.

Johnson, D. G. (2006). Computer systems: Moral entities but not moral agents. *Ethics and Information Technology, 8,* 195–204.

Kleinberg, J., Ludwig, J., Mullainathany, S., & Sunstein, C. R. (2019). Discrimination in the age of algorithms. *Journal of Legal Analysis, 10,* 1–62.

Levinas, E. (1969). *Totality and infinity: An essay on exteriority (A. Lingis, Trans.)*. Pittsburgh: Duquesne University.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology, 6*(3), 175–183.

McKenna, M. (2008). Putting the lie on the control condition for moral responsibility. *Philosophical Studies, 139*(1), 29–37.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence, 267,* 1–38.

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society, 3,* 1–21.

Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *Intelligent Systems (IEEE), 21*(4), 18–21.

Nyholm, S., & Smids, Jilles. (2016). The ethics of accident-algorithms for self-driving cars: An applied trolley problem? *Ethical Theory and Moral Practice, 19*(5), 1275–1289.

Rudy-Hiller, F. 2018. The epistemic condition for moral responsibility. Stanford Encyclopedia of Philosophy. Retrieved 26, Aug 2019, https://plato .stanf ord.edu/entri es/moral -respo nsibi lity-epist emic/.

Samek, W., Wiegand, T., & Müller, K.-R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. https://arxiv .org/pdf/1708.08296 .pdf.

Sommaggio, P., & Marchiori, S. (2018). Break the chains: A new way to consider machine's moral problems. *Biolaw Journal, 3,* 241–257.

Sparrow, R. (2007). Killer robots. *Journal of Applied Philosophy, 24*(1), 62–77.

Stahl, B. C. (2006). Responsible computers? A case for ascribing quasi-responsibility to computers independent of personhood or agency. *Ethics and Information Technology, 8,* 205–213.

Suárez-Gonzalo, S., Mas-Manchón, L., & Guerrero-Solé, F. (2019). Tay is you. The attribution of responsibility in the algorithmic culture. *Observatorio, 13*(2), 1–14.

Sullins, J. P. (2006). When is a robot a moral agent? *International Review of Information Ethics, 6*(12), 23–29.

Sunstein, C. R. (2018). Algorithms, correcting biases. Forthcoming, *Social Research*. Available at SSRN: https://ssrn.com/abstr act=33001 71.

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science, 361*(6404), 751–752.

Turner, J. (2018). *Robot rules: Regulating artificial intelligence*. Cham: Palgrave Macmillan.

Van de Poel, I., Nihlén Fahlquist, J., Doorn, N., Zwart, S., & Royakkers, L. (2012). The problem of many hands: Climate change as an example. *Science and Engineering Ethics, 18*(1), 49–67.

Verbeek, P. P. (2006). Materializing morality. *Science, Technology and Human Values, 31*(3), 361–380.

Wallach, W., & Allen, C. (2009). *Moral machines, teaching robots right from wrong*. Oxford: Oxford University Press.

Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability[*]

Mark Coeckelbergh[**]

Tareq Osman[***]

Abstract

This paper discusses the problem of responsibility attribution raised by the use of artificial intelligence (AI) technologies. It is assumed that only humans can be responsible agents; yet this alone already raises many issues, which are discussed starting from two Aristotelian conditions for responsibility. Next to the well-known problem of many hands, the issue of "many things" is identified and the temporal dimension is emphasized when it comes to the control condition. Special attention is given to the epistemic condition, which draws attention to the issues of transparency and explainability. In contrast to standard discussions, however, it is then argued that this knowledge problem regarding agents of responsibility is linked to the other side of the responsibility relation: the addressees or "patients" of responsibility ,who may demand reasons for actions and decisions made by using AI .Inspired by a relational approach, responsibility as answerability thus offers an important additional, if not primary, justification for explainability based, not on agency, but on patiency.

[*] This article was translated from the original English:
Coeckelbergh, Mark (2020). Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability, Science and Engineering Ethics, 26(4), 2051–2068.

[**] A Belgian philosopher specializing in the philosophy of technology, he is a professor of the philosophy of technology in the Department of Philosophy at the University of Vienna. Most of his work focuses on artificial intelligence, digital ethics, and algorithmic power.

[***] Egyptian researcher and freelance translator; B.S. in Pharmacy, Cairo University. Email: dr.tareq.osman@gmail.com, https://orcid.org/0009-0002-2256-9794.

To cite this article: Coeckelbergh, M .(2020) *Artificial intelligence, responsibility attribution, and a relational justification of explainability* .Tareq Osman (trans.), Al-Fikr al-islāmī al-muʿāṣir ,vol. 111.

© 2026 International Institute of Islamic Thought. All rights reserved.