مارك كوكلبير ج

أخلاقيات الذكاء الاصطناعي

ترجمة هبة عبد العزيز غانم

سلسلة المعارف الأساسية

أخلاقيات الذكاء الاصطناعي

تأليف
مارك كوكلبيرج

ترجمة
هبة عبد العزيز غانم

مراجعة
هبة عبد المولى أحمد

هنـــداوي

أخلاقيات الذكاء الاصطناعي

مارك كوكلبيرج

AI Ethics

Mark Coeckelbergh

الناشر مؤسسة هنداوي
المشهرة برقم ١٠٥٨٥٩٧٠ بتاريخ ٢٦ / ١ / ٢٠١٧

يورك هاوس، شييت ستريت، ونسدور، SL4 1DD، المملكة المتحدة
تليفون: ٨٣٢٥٢٢ ١٧٥٣ (٠) ٤٤ +
البريد الإلكتروني: hindawi@hindawi.org
الموقع الإلكتروني: https://www.hindawi.org

تصميم الغلاف: ولاء الشاهد

الترقيم الدولي: ٨ ٣٦٨٥ ٥٢٧٣ ١ ٩٧٨

صدر الكتاب الأصلي باللغة الإنجليزية عام ٢٠٢٠.
صدرت هذه الترجمة عن مؤسسة هنداوي عام ٢٠٢٤.

المحتويات

إلى أرنو

تمهيد السلسلة

تُقدِّم «سلسلة المعارف الأساسية» التي تَنشرها مؤسسة «إم آي تي بريس» كُتبًا موجزةً بلُغة جَزلة سهلة الفهم، وشكلٍ أنيق، وحجمٍ صغير يُلائم الجيب، تُناقِش الموضوعات التي تُثير الاهتمام في الوقت الحالي. ولَمَّا كانت كُتب هذه السلسلة من تأليف مُفكِّرين بارزين، فإنها تُقدِّم آراء الخبراء بشأن موضوعاتٍ تتنوَّع بين المجالات الثقافية والتاريخية، إضافةً إلى العِلمية والتقنية.

في ظِل ما يَشيع في هذا العصر من إشباعٍ لَحظي للمعلومات، أضحى لدى الجميع القدرةُ على الوصول إلى الآراء والأفكار والشروح السطحية بسرعةٍ وسهولة، وأصبح من الصعوبة بمكانٍ أن يَحظى المرء بالمعرفة الأساسية التي تُيسِّر فَهمًا صادقًا للعالَم؛ وما تفعله كتب هذه السلسلة هو أنها تُحقِّق ذلك الغرض. وكلُّ كتاب من هذه الكتب المُختصَرة يُقدِّم للقارئ وسيلةً مُيسَّرة للوصول إلى الأفكار المُعقَّدة، من خلال تبسيط المواد المُتخصِّصة لغير المُختصِّين، وشَرْح الموضوعات المهمة بأبسط طريقةٍ مُمكنة.

بروس تيدور
أستاذ الهندسة البيولوجية وعلوم الكمبيوتر
«معهد ماساتشوستس للتكنولوجيا»

شكر وتقدير

لا يعتمد هذا الكتاب على عملي الخاص في موضوع أخلاقيات الذكاء الاصطناعي فحسب، بل يعكس المعرفة والخبرة في هذا المجال بأكملِه. وسيكون من المُستحيل إدراج جميع الأشخاص الذين ناقشتُهم وتعلَّمتُ منهم على مدار السنوات الماضية، لكن المجتمعات ذات الصلة والسريعة النمو التي أعرفها تضم باحثين في مجال الذكاء الاصطناعي مثل جوانا بريسون ولوك ستيلز، وزملائي الفلاسفة في مجال التكنولوجيا مثل شانون فالور ولوتشيانو فلوريدي، وأكاديميين يسعَون إلى الابتكار المسئول في هولندا والمملكة المتحدة، مثل بيرند ستال في جامعة دي مونتفورت، وبعض الأشخاص الذين الْتَقيتُ بهم في فيينا، مثل روبرت ترابل، وسارة سبيكرمان، وولفجانج (بيل) برايس، وزملائي الأعضاء في الهيئات الاستشارية ذات التوجُّهات السياسية، فريق الخبراء الرفيع المستوى المَعنيِّ بالذكاء الاصطناعي (المفوضية الأوروبية) والمجلس النمساوي للروبوتات والذكاء الاصطناعي، ومن ضِمنهم على سبيل المثال لا الحصر راجا شاتيلا، وفيرجينيا ديج نوم، وجيروين فان دين هوفن، وسابين كوسيجي، وماتياس شوتز. أودُّ أيضًا أن أشكر بحرارة زاكاري ستورمز للمساعدة في التدقيق اللغوي للكتاب وتنسيقه، ولينا ستاركل وإيزابيل والتر على دعمهما في البحث عن الأدبيات.

الفصل الأول

أيتها المرآة على الحائط

الضجة والمخاوف التي يُثيرها الذكاء الاصطناعي: أيتها المرآة على الحائط: مَن الأذكى في العالَم؟

عندما أُعلنت النتائج، اغرورقت عينا اللاعب لي سيدول بالدموع. حقَّق «ألفا جو»، وهو برنامج ذكاء اصطناعي طوَّرَتْه شركة «ديب مايند» التابعة إلى جوجل، فوزًا ٤-١ في لعبة «جو» (لعبة «جو» هي لعبة استراتيجية قديمة ظهرت في الصين ويُشارك فيها لاعبان اثنان). تاريخ الحدث: مارس ٢٠١٦. قبل عقدَين من الزمان، خسر لاعب الشطرنج جاري كاسباروف الحاصل على لقب «جراند ماستر» (الأستاذ الكبير) أمام الآلة «ديب بلو»، والآن فاز برنامج كمبيوتر على بطل العالم لثماني عشرة مرة؛ لي سيدول، في لعبة مُعقَّدة كان يُنظَر إليها على أنها لعبة لا يمكن أن يلعبها إلا البشر، باستخدام حدسِهم وتفكيرهم الاستراتيجي. الأدهى من ذلك أن الكمبيوتر لم يفُز باتباع القواعد المُعطاة له من قِبَل المُبرمجين، وإنما عن طريق تعلُّم الآلة القائم على الملايين من مباريات «جو» السابقة وعلى اللعب ضدَّ نفسه. في مثل هذه الحالة، يُعِد المبرمجون مجموعات البيانات ويُنشِئون الخوارزميات، ولكن لا يُمكنهم معرفة التحرُّكات التي سيأتي بها البرنامج. فالذكاء الاصطناعي يتعلَّم من تلقاء نفسه. وبعد عددٍ من التحرُّكات غير المعتادة والمفاجئة، اضطُرَّ بطل العالم لي إلى الانسحاب (Borowiec 2016).

إنه إنجاز رائع حقَّقَه الذكاء الاصطناعي. ولكنه، مع ذلك، يُثير المخاوف في قلوبنا. إننا مُعجبون بجمال الحركات، ولكننا أيضًا حزانى، وربما حتى خائفون. نأمُل في أن تساعدنا أنظمة الذكاء الاصطناعي الأكثر ذكاءً في إحداث ثورة في الرعاية الصحية أو في إيجاد حلولٍ

لجميع أنواع المشكلات المجتمعية، ولكن يُراودنا القلق من أن تسيطر الآلات على زمام أمورنا. فهل تستطيع الآلات أن تتفوَّق علينا وتتحكَّم فينا؟ هل لا يزال الذكاء الاصطناعي مجرد أداة، أم إنه سيُصبح رويدًا رويدًا سيدنا لا محالة؟ تُذكِّرنا هذه المخاوف بكلمات «هال» كمبيوتر الذكاء الاصطناعي في فيلم الخيال العلمي الذي أخرجه ستانلي كوبريك: «٢٠٠١: ملحمة الفضاء» (٢٠٠١: سبيس أوديسي)، حين قال ردًّا على الأمر البشري «افتح أبواب المركبة الصغيرة»: «أخشى أنني لا أستطيع أن أفعل ذلك يا ديف.» وإذا لم يكُن هناك خوف، فقد يكون هناك شعور بالحزن أو خيبة الأمل. لقد أطاح داروين وفرويد بإيماننا بتميُّزنا، وبإحساسنا بالتفوُّق، وأطاحا بأوهام السيطرة التي يعيش فيها البشر؛ والآن جاء دور الذكاء الاصطناعي ليُوجِّه ضربةً أخرى إلى صورة البشر عن ذواتهم. إذا كانت الآلة تستطيع القيام بذلك، فماذا تبقَّى لنا؟ ماذا نحن؟ هل نحن مجرَّد آلات؟ هل نحن آلات رديئة، بها الكثير من العيوب والأخطاء؟ وماذا سيحدُث لنا؟ هل سنُصبح عبيدًا للآلات؟ أو ما هو أسوأ، مجرد مصدر للطاقة، كما في فيلم «المصفوفة» (ذا ماتريكس)؟

التأثير الحقيقي والواسع النطاق للذكاء الاصطناعي

ولكن إنجازات الذكاء الاصطناعي لا تقتصر على الألعاب أو عالَم الخيال العلمي. فالذكاء الاصطناعي يحدث الآن وهو مُتوغِّل في كل ما حولنا، وغالبًا ما يكون مُضمَّنًا على نحو غير مرئي في أدواتنا اليومية وبكونه جزءًا من الأنظمة التكنولوجية المعقَّدة (Boddington 2017). ونظرًا إلى النمو الهائل لقدرة الكمبيوتر، وإتاحة البيانات (الضخمة) بسبب وسائل التواصل الاجتماعي والاستخدام الهائل لمليارات الهواتف الذكية، وشبكات المحمول السريعة، أحرَزَ الذكاء الاصطناعي، وخاصة تعلُّم الآلة، تقدُّمًا كبيرًا. وقد مكَّنَ هذا الخوارزميات من تولِّي العديد من أنشطتنا، بما في ذلك التخطيط والكلام والتعرُّف على الوجوه واتخاذ القرار. يمتلك الذكاء الاصطناعي تطبيقاتٍ في العديد من المجالات، بما في ذلك النقل والتسويق والرعاية الصحية والتمويل والتأمين والأمن والجيش والعلوم والتعليم والعمل المكتبي والمساعدة الشخصية (مثل جوجل دوبلكس[1] والترفيه والفنون (مثل استرجاع الموسيقى وتأليفها) والزراعة، وبالطبع التصنيع.

تتمُّ عمليات إنشاء الذكاء الاصطناعي واستخدامه لدى شركات تكنولوجيا المعلومات والإنترنت. على سبيل المثال، لطالما استخدمت جوجل الذكاءَ الاصطناعي في مُحرِّك البحث الخاص بها. كما يستخدم فيسبوك الذكاء الاصطناعي في الإعلانات المستهدفة وإشارات

الصور. كذلك تستخدم مايكروسوفت وأبِل الذكاء الاصطناعي في تشغيل مساعدَيهما الرقميين. لكن الذكاء الاصطناعي لا يقتصر على قطاع تكنولوجيا المعلومات بمعناه الضيِّق. فهناك، على سبيل المثال، الكثير من الخُطط الملموسة، والتجارب في مجال السيارات الذاتية القيادة. فهذه التقنية تعتمد أيضًا على الذكاء الاصطناعي. كما تستخدِم الطائرات دون طيار الذكاءَ الاصطناعي، مثلها مثل الأسلحة الذاتية التشغيل التي يمكن أن تقتُل دون تدخُّلٍ بشري. بل إن الذكاء الاصطناعي قد استُخدِم بالفعل في اتخاذ القرار في المحاكم. ففي الولايات المتحدة، على سبيل المثال، استُخدم نظام «كومباس» للتنبُّؤ بالذين يُحتمَل أن يُعاودوا ارتكاب الجرائم. يدخل الذكاء الاصطناعي أيضًا في المجالات التي نعتبِرها عمومًا أكثر شخصية أو حميمية. على سبيل المثال، يمكن للآلات الآن قراءة وجوهنا، ليس فقط للتعرُّف علينا، ولكن أيضًا لقراءة انفعالاتنا واسترداد جميع المعلومات المرتبطة بنا.

> الذكاء الاصطناعي يحدث الآن وهو مُتوغِّل في كلِّ ما حولنا، وغالبًا ما يكون مُضمَّنًا على نحوٍ غير مرئي في أدواتنا اليومية.

الحاجة إلى مناقشة المشكلات الأخلاقية والمجتمعية

يمكن أن يكون للذكاء الاصطناعي العديد من الفوائد. ويمكن استخدامه لتحسين الخدمات العامة والتجارية. على سبيل المثال، يُعد التعرُّف على الصور شيئًا مفيدًا في الطب؛ إذ ربما يساعد في تشخيص أمراضٍ مثل السرطان ومرض ألزهايمر. ولكن مثل هذه التطبيقات اليومية للذكاء الاصطناعي تُظهِر أيضًا كيف تُثير التقنيات الجديدة تخوُّفات أخلاقية. واسمحوا لي أن أُقدِّم بعض الأمثلة على أسئلةٍ حول أخلاقيات الذكاء الاصطناعي.

هل يجب أن تحتوي السيارات الذاتية القيادة على قيودٍ أخلاقية مضمَّنة؟ وإذا كان الأمر كذلك، فما نوع هذه القيود وكيف ينبغي تحديدها؟ على سبيل المثال، إذا واجهت سيارة ذاتية القيادة موقفًا يتعيَّن عليها فيه الاختيار بين أن تصطدم بطفلٍ أو تصطدم بجدارٍ لإنقاذ حياة الطفل، ولكن مع احتمال قتل راكِبها، فماذا تختار؟ وهل ينبغي ترخيص الأسلحة الفتَّاكة الذاتية التشغيل من الأساس؟ كم عدد القرارات التي نُريد تفويضها إلى الذكاء الاصطناعي، وما القَدْر الذي نُفوِّضه منها؟ ومَن سيكون المسئول عندما يحدُث خطأ ما؟ في إحدى القضايا، وضَعَ القضاة ثقتهم في خوارزمية «كومباس» أكثر من ثِقتهم في

الاتفاقات التي توصَّل إليها الدفاع والادعاء.[2] فهل سنعتمد كثيرًا على الذكاء الاصطناعي؟ تُعد خوارزمية «كومباس» أيضًا مُثيرة للجدل إلى حدٍّ كبير؛ نظرًا إلى أن الأبحاث أظهرت أن الأشخاص الذين تنبَّأت الخوارزمية بأنهم سيُعيدون ارتكاب الجرائم ولكنهم لم يفعلوا كانت النسبة الكبرى منهم مِن السود (Fry 2018). وبالتالي يمكن للذكاء الاصطناعي أن يُعزِّز التحيُّز والتمييز غير العادل. ويمكن أن تنشأ مشكلات مُماثلة مع الخوارزميات التي تُوصي بقراراتٍ بشأن طلبات الرهن العقاري وطلبات التقدُّم للوظائف. أو فلنُفكر فيما يُسمى بالشرطة التنبُّؤية: تُستخدَم الخوارزميات للتنبؤ بالمكان المُحتمَل لارتكاب الجرائم (على سبيل المثال، أي منطقة في المدينة) ومَن قد يرتكِبها، ولكن قد تكون النتيجة أن تُستهدَف مجموعات اجتماعية واقتصادية أو عِرقية مُعيَّنة للمراقبة الشرطية بدرجةٍ أكبر من غيرهم من المجموعات. وقد جَرَت الاستعانة بالفعل بالشرطة التنبُّؤية في الولايات المتحدة، وكما يُظهر تقرير حديث لمنظمة «ألجوريذم ووتش» (٢٠١٩)، فقد استُعين بها أيضًا في أوروبا.[3] وغالبًا ما تُستخدَم تقنية التعرُّف على الوجوه القائمة على الذكاء الاصطناعي لأغراض المُراقبة، ومِن ثَم يمكن أن تُشكِّل انتهاكًا لخصوصية الأفراد. كما يُمكنها بشكلٍ أو بآخَر التنبؤ بالميول الجنسية لدى الأفراد. الأمر لا يتطلَّب أي معلومات من هاتفك أو أي بيانات بيومترية (بيانات المقاييس الحيوية). وتقوم الآلة بعملها عن بُعد. ومِن ثَم فإننا باستخدام الكاميرات الموجودة في الشوارع والأماكن العامة الأخرى، يمكن التعرُّف علينا و«قراءتنا»، بما في ذلك التعرف على حالتنا المزاجية. وعن طريق تحليل بياناتنا، يمكن التنبؤ بصحَّتنا العقلية والجسدية؛ دون عِلمنا بذلك. ويمكن لأصحاب العمل استخدام التكنولوجيا لمُراقبة أدائنا. ويمكن للخوارزميات النشطة على وسائل التواصل الاجتماعي أن تنشر خطاب الكراهية أو المعلومات الخطأ؛ على سبيل المثال، يمكن أن تظهر الروبوتات السياسية في هيئة أشخاصٍ حقيقيِّين وتنشُر محتوًى سياسيًّا. إحدى الحالات المعروفة هي برنامج الدردشة الآلي من مايكروسوفت لعام ٢٠١٦ المُسمى «تاي» المُصمَّم لإجراء محادثات مَرِحة على تويتر، ولكن عندما أصبح أكثر ذكاءً، بدأ في نشر تغريدات تحمِل دلالاتٍ عنصرية. يمكن لبعض خوارزميات الذكاء الاصطناعي إنشاء خطابات فيديو كاذبة، مثل الفيديو الذي جرى إنشاؤه ليُشبه بشكلٍ مُضلِّل خطابًا لباراك أوباما.[4]

غالبًا ما تكون النوايا طيبة. ولكن هذه المشكلات الأخلاقية عادةً ما تكون نتائج غير مقصودة للتكنولوجيا: فمعظم هذه التأثيرات، مثل التحيُّز أو خطاب الكراهية، لم يقصدها مطورو التكنولوجيا أو مُستخدموها. علاوةً على ذلك، هناك سؤال مهم يجب طرحه دائمًا:

من أجل مَن يتم التحسين؟ من أجل الحكومة أم من أجل المواطنين؟ من أجل الشرطة أم من أجل مَن تستهدفهم الشرطة؟ من أجل بائع التجزئة أم من أجل الزبون؟ من أجل القضاة أم من أجل المُتهمين؟ كما تظهر الأسئلة المتعلقة بالسلطة والهيمنة، كالحال على سبيل المثال عندما يقتصر تشكيل التكنولوجيا على عددٍ قليل من الشركات الضخمة (Nemitz 2018). فمَن الذي يُشكل مُستقبل الذكاء الاصطناعي؟

يُلقي هذا السؤال الضوء على الأهمية الاجتماعية والسياسية للذكاء الاصطناعي. تتعلَّق أخلاقيَّات الذكاء الاصطناعي بالتغيُّر التكنولوجي وتأثيره على حياة الأفراد، ولكنها تتعلق أيضًا بالتحولات التي تحدُث في المجتمع وفي الاقتصاد. وتدلُّ قضايا التحيُّز والتمييز بالفعل على أن الذكاء الاصطناعي مُرتبِط بالمجتمع. ولكنه يُغيِّر أيضًا الاقتصاد، وبالتالي ربما يُغيِّر الهيكل الاجتماعي لمجتمعاتنا. ووفقًا لمكافي وبرينجولفسون (٢٠١٤)، فقد دخلنا عصر الآلة الثاني، الذي لا تكون فيه الآلات مُكملة للبشر فحسب، كما في الثورة الصناعية، ولكنها أيضًا بدائل للبشر. ونظرًا إلى أن المِهن والأعمال من جميع الأنواع ستتأثر بالذكاء الاصطناعي، فمن المتوقَّع أن يتغيَّر مجتمعنا تغيُّرًا جذريًّا مع دخول التقنيات التي وَصفت في يومٍ من الأيام في روايات الخيال العلمي حيِّز العالم الحقيقي (McAfee and Brynjolfsson 2017). فما هو مستقبل العمل؟ وما نوع الحياة التي سنعيشها نحن عندما يتولى الذكاء الاصطناعي القيام بالوظائف؟ ومَن «نحن»؟ ومَن الذي سيستفيد من هذا التحوُّل ومن سيخسر؟

هذا الكتاب

استنادًا إلى الإنجازات المُذهلة التي تم تحقيقها، فهناك الكثير من الضجة المُثارة حول الذكاء الاصطناعي. ويُستخدَم الذكاء الاصطناعي بالفعل في مجموعةٍ واسعة من مجالات المعرفة والممارسات البشرية. وقد أثارت الأولى تكهُّناتٍ جامحة حول مستقبل التكنولوجيا، كما أثارت مناقشاتٍ فلسفيةً مهمَّة حول معنى أن تكون إنسانًا. بينما خلقت الثانية إحساسًا بالإلحاح من جانب الأخلاقيين وصانعي السياسات لضمان أن تُفيدنا هذه التكنولوجيا بدلًا من أن تخلق أمام الأفراد والمجتمعات تحديات لا يُمكنهم التغلُّب عليها. وتُعد هذه المخاوف الأخيرة أكثر عمليةً وإلحاحًا.

تتعلق أخلاقيات الذكاء الاصطناعي بالتغيُّر التكنولوجي وتأثيرِه على حياة الأفراد، ولكنها تتعلق أيضًا بالتحوُّلات التي تحدث في المجتمع وفي الاقتصاد.

يتناول هذا الكتاب، الذي كتبَه فيلسوف أكاديمي لدَيه أيضًا خبرة في تقديم المشورة من أجل وضع السياسات، كِلا الجانبَين؛ فهو يتعامل مع الأخلاقيات على هذه المستويات كافة. ويهدف إلى إعطاء القارئ نظرةً عامة جيدة على المشكلات الأخلاقية التي يُثيرها الذكاء الاصطناعي، بدءًا من السرديات المؤثرة حول مستقبل الذكاء الاصطناعي والأسئلة الفلسفية حول طبيعة الإنسان ومُستقبله، وانطلاقًا إلى القضايا الأخلاقية المُتعلقة بالمسئولية والتحيُّز وكيفية التعامُل مع المسائل العملية الواقعية التي أثارتها التكنولوجيا عن طريق وضع السياسات؛ لا سيما إذا كان ذلك قبل فوات الأوان.

لكن ماذا سيحدُث إذا «فات الأوان»؟ بعض السيناريوهات متشائمة ومتفائلة في الوقت نفسه. اسمحوا لي أن أبدأ ببعض الأحلام والكوابيس حول مستقبل التكنولوجيا، والسرديات المؤثرة التي تبدو، ولو للوهلة الأولى على الأقل، ذات صِلةٍ بتقييم الفوائد والمخاطر المُحتمَلة للذكاء الاصطناعي.

الفصل الثاني

الذكاء الفائق والوحوش ونهاية العالم بالذكاء الاصطناعي

الذكاء الفائق وتجاوز الإنسانية

أدَّت الضجة المُحيطة بالذكاء الاصطناعي إلى ظهور جميع أنواع التكهُّنات حول مستقبل الذكاء الاصطناعي ومستقبل ما سيكون عليه الإنسان. إن إحدى الأفكار الشائعة، والتي تتكرَّر كثيرًا في وسائل الإعلام وفي النقاشات العامة حول الذكاء الاصطناعي، بل ينشرها أيضًا خبراء التكنولوجيا المؤثِّرون الذين يُطوِّرون تقنية الذكاء الاصطناعي مثل إيلون ماسك وراي كورزوايل، هي فكرة الذكاء الفائق، وبشكلٍ أكثر عمومية، فكرة أن الآلات ستُسيطر علينا، وتستعبِدنا وليس العكس. بالنسبة إلى البعض، هذا حلم؛ وبالنسبة إلى الكثيرين، هذا كابوس. وهناك مَن يرَون أنه حلم وكابوس في الوقت نفسه.

فكرة الذكاء الفائق هي أن الآلات ستتفوَّق على الذكاء البشري. وهي غالبًا ما ترتبط بفكرة انفجار الذكاء الاصطناعي والتفرُّد التكنولوجي. ووفقًا لنيك بوستروم (٢٠١٤)، سنقع في مأزقٍ يُماثل ذلك الذي وقعت فيه الغوريلا، التي يعتمد مصيرها اليوم علينا بشكلٍ كامل. إنه يرى طريقَين على الأقل لبلوغ الذكاء الفائق وما يُسمَّى أحيانًا بانفجار الذكاء الاصطناعي. أحدهما أن الذكاء الاصطناعي سوف يُطوِّر تحسينًا ذاتيًّا تكراريًّا؛ إذ يستطيع الذكاء الاصطناعي تصميم نسخةٍ مُحسَّنة من نفسه، والتي بدورها تُصمِّم نسخةً أكثر ذكاءً من نفسها، وهكذا دواليك. أما الطريق الآخَر فهو محاكاة الدماغ بالكامل أو تحميله: دماغ بيولوجي يُمكِن مسحه ضوئيًّا وصُنع نموذج له، ثم إعادة إنتاجِه في مكوناتٍ برمجية ذكيَّة ومِن خلالها. يتم بعد ذلك توصيل هذه المُحاكاة للدماغ البيولوجي

بجسم إنسان آلي. وستؤدي مثل هذه التطوُّرات إلى انفجارٍ في الذكاء غير البشري. حتى إن ماكس تجمارك (٢٠١٧) يتخيل أن فريقًا ما يُمكِنه إنشاء ذكاء اصطناعي يُصبح في منتهى القوة بحيث يستطيع إدارة الكوكب. ويكتب يوفال هراري عن عالَمٍ لم يعُد فيه البشر يسيطرون، ولكنهم يعبدون البيانات ويثقون في قدرة الخوارزميات على اتخاذ قراراتهم. وبعد انهيار كلِّ أوهام الإنسانيين والمؤسسات الليبرالية، لن يبقى للبشر إلا أن يحلموا بالاندماج في تدفُّق البيانات. يسير الذكاء الاصطناعي في مساره الخاص، «الذهاب إلى حيث لم يذهب أي إنسانٍ من قبل؛ وإلى حيث لا يمكن لأي إنسانٍ أن يتبعه» (Harari 2015, 393).

ترتبط فكرة انفجار الذكاء الاصطناعي ارتباطًا وثيقًا بفكرة «التفرُّد التكنولوجي»: لحظة في تاريخ البشرية سيُحدِث فيها التقدُّم التكنولوجي الهائل تغييرًا دراماتيكيًّا بحيث لا نعود نستوعِب ما يحدث و«تنتهي الشئون الإنسانية كما نفهمها اليوم» (Shanahan 2015, xv). في عام ١٩٦٥، تكهَّنَ عالم الرياضيات البريطاني إيرفينج جون جود بآلة فائقة الذكاء تُصمِّم آلاتٍ أفضل؛ وفي التسعينيات، رأى مؤلف الخيال العلمي وعالم الكمبيوتر فيرنور فينج أن هذا سيعني نهاية عصر الإنسان. وقد اقترح رائد علم الكمبيوتر جون فون نيومان بالفعل الفكرة في خمسينيات القرن العشرين. وتبنَّى راي كورزوايل (٢٠٠٥) مصطلح «التفرُّد» وتوقَّع أن الذكاء الاصطناعي، جنبًا إلى جنبٍ مع أجهزة الكمبيوتر وعلم الوراثة وتكنولوجيا النانو وعلم الروبوتات، سيؤدي إلى نقطةٍ يكون فيها ذكاءُ الآلة أقوى من كلِّ الذكاء البشري مُجتمعًا، ويندمج عندها الذكاء البشري وذكاء الآلة في النهاية. وسوف يتجاوز البشر حدود أجسامهم البيولوجية. وكما جاء في عنوان كتابه: «التفرُّد قريب». وهو يعتقد أن هذا سيحدث حوالي عام ٢٠٤٥.

ليس لهذه القصة بالضرورة نهاية سعيدة: ففي رأي بوستروم وتجمارك وآخرين، ثمَّة «مخاطر وجودية» مُرتبطة بالذكاء الفائق. وقد تكون نتيجة هذه التطوُّرات أن الذكاء الاصطناعي الفائق سوف يُسيطر ويتولَّى زمام الأمور ويُهدِّد حياة الإنسان الذكية. وسواء أكان هذا الكيان واعيًا أم لا، وبصورة أعم مهما كانت حالته أو كيفية نشوئه، فإن القلق هنا يتعلَّق بما سيفعله هذا الكيان (أو ما لا يفعله). قد لا يهتمُّ الذكاء الاصطناعي بأهدافنا البشرية. ونظرًا لعدم امتلاكِه جسدًا بيولوجيًّا، فإنه لن يفهم حتى المعاناة البشرية. ويُقدم بوستروم تجربةً فكرية لذكاءٍ اصطناعي يُحدَّد له هدف مُعيَّن وهو تصنيع مشابك الورق بأكبر كمٍّ مُمكِن، فما كان منه إلا أن حوَّل كوكب الأرض والبشر الذين يعيشون عليه إلى

موارد لإنتاج مشابك الورق. إذَن التحدِّي الذي يُواجهنا اليوم هو التأكُّد من أننا نبني ذكاءً اصطناعيًّا لا يُثير بطريقةٍ ما مشكلة السيطرة هذه؛ بمعنى أنه يفعل ما نريد ويأخذ حقوقنا في الاعتبار. على سبيل المثال، هل يجب أن نحدَّ بطريقةٍ ما من قدرات الذكاء الاصطناعي؟ وكيف يُمكننا احتواء الذكاء الاصطناعي؟[1]

ثمَّة أفكار أخرى مترابطة وذات صلة؛ ألا وهي الأفكار المتعلِّقة بتجاوز الإنسانية. في ضوء الذكاء الفائق والإحباط من الضعف البشري و«الأخطاء»، يجادل أنصار تجاوز الإنسانية مثل بوستروم بأننا بحاجةٍ إلى تعزيز الإنسان: جعله أكثر ذكاءً، وأقل عُرضةً للمرض، وأطوَل عمرًا، وربما حتى خالدًا، مما يؤدي إلى ما يُسمِّيه هاراري «الإنسان الإله»: ترقية البشر إلى آلهة. وكما قال فرانسيس بيكون في «دحض الفلسفات»: البشر «آلهة فانية» (Bacon 1964, 106). لماذا لا نُحاول تحقيق الخلود؟ ولكن حتى لو لم نستطع تحقيق ذلك، فإن الآلة البشرية، وفقًا لمُناصري تجاوُز الإنسانية، بحاجةٍ إلى ترقية. فنحن إذا لم نفعل ذلك، فسيُخاطر البشر بأن يظلوا «الجزء المُتخلف غير الكفء بشكل متزايد» من الذكاء الاصطناعي (Armstrong 2014, 23). إن البيولوجيا البشرية بحاجةٍ إلى إعادة تصميم، ولذا يتساءل بعض مؤيدي تجاوز الإنسانية، لماذا لا نتخلَّص تمامًا من الأجزاء البيولوجية ونُصمِّم كائناتٍ ذكية غير عضوية؟

على الرغم من أن معظم الفلاسفة والعلماء الذين يُروِّجون لهذه الأفكار يحرصون على تمييز آرائهم عن الخيال العلمي والدين، فإن العديد من الباحثين يُفسِّرون أفكارهم بهذه المصطلحات بالضبط. بادئ ذي بدء، ليس من الواضح مدى ارتباط أفكارهم بالتطوُّرات التكنولوجية الحالية وعلوم الذكاء الاصطناعي، وما إذا كان هناك فرصة حقيقية للوصول إلى الذكاء الفائق في المُستقبل القريب، هذا إن أمكن الوصول إليه من الأساس. إذ يرفض البعض تمامًا إمكانية الوصول إليه (انظر الفصل التالي)، وحتى هؤلاء الذين على استعدادٍ لقبول إمكانية الوصول إليه من حيث المبدأ، مثل العالِمة مارجريت بودن، فإنهم لا يعتقدون أنه من المُرجَّح الوصول إليه عمليًّا. إن فكرة الذكاء الفائق تفترِض أننا سنُطوِّر «الذكاء الاصطناعي العام»، أو الذكاء الذي يكافئ الذكاء البشري أو يتفوَّق عليه، وهناك العديد من العقبات التي يجب التغلُّب عليها قبل تحقيق ذلك. وترى بودن (٢٠١٦) أن الذكاء الاصطناعي ليس واعدًا كما يتوقَّع الكثيرون. وفي تقريرٍ صادر عن البيت الأبيض عام ٢٠١٦، تم التأكيد على اتفاق خبراء القطاع الخاص على أن الذكاء الاصطناعي العام لن يتحقَّق على الأقل قبل عقود. كما يرفض العديد من

الباحثين في مجال الذكاء الاصطناعي الرؤى المُظلمة المتشائمة التي يُروِّج لها بوستروم وآخرون، ويحضُّون على استخدام الذكاء الاصطناعي بشكلٍ إيجابي، كمساعدٍ أو زميل. ولكن المسألة لا تتعلق بما سيحدث فعليًّا في المستقبل. بل يوجَد شيء آخَر يُثير القلق وهو أن هذه المناقشة حول تأثيرات الذكاء الاصطناعي في المستقبل (البعيد) تُشتِّت الانتباه عن المخاطر الحقيقية والموجودة حاليًّا للأنظمة التي تم نشرها فعليًّا (Crawford and Calo 2016). يبدو أن هناك خطرًا حقيقيًّا أنه في المستقبل القريب، لن تكون الأنظمة ذكيةً بما فيه الكفاية وأننا سنفهم آثارها الأخلاقية والاجتماعية بشكلٍ غير كافٍ، ومع ذلك سنستخدِمها على نطاقٍ واسع. كما أن التركيز المُفرِط على الذكاء، بوصفه سِمةً رئيسية للإنسانية، وهدفًا نهائيًّا وحيدًا، هو أيضًا أمر مشكوك فيه (Boddington 2017).

مع ذلك، تستمر الأفكار مثل الذكاء الفائق في التأثير على المناقشة العامة. ومن المُحتمل أن تؤثِّر أيضًا على تطوُّر التكنولوجيا. على سبيل المثال، لا يُعتبر راي كورزوايل من دُعاة المستقبلية فحسب. بل إنه يشغل منصب مدير الهندسة في شركة جوجل منذ عام ٢٠١٢. كما يبدو أن إيلون ماسك، الرئيس التنفيذي لشركة تيسلا وشركة سبيس إكس، وهو شخصية عامة معروفة جدًّا، يؤيد سيناريوهات الذكاء الفائق والمخاطر الوجودية (سيناريوهات الهلاك؟) التي وضعها بوستروم وكورزوايل. وقد حذَّر مرارًا من خطورة الذكاء الاصطناعي، واعتبَرَه تهديدًا وجوديًّا وزعم أننا لا يُمكننا التحكُّم في الشيطان (Dowd 2017). ويعتقد ماسك أن البشر سينقرضون على الأرجح، ما لم يُدمَج الذكاء البشري والذكاء الآلي أو نتمكَّن من الهروب إلى المريخ.

ربما تكون هذه الأفكار مؤثرة للغاية لأنها تمسُّ مخاوف وآمالًا عميقةً تتعلَّق بالبشر والآلات داخل وعينا الجمعي. وسواء قَبِلنا هذه الأفكار المُحدَّدة أو رفَضْناها، فإن هناك صِلاتٍ واضحة بالسرديات الخيالية في الثقافة البشرية والتاريخ التي تُحاول أن تفهم الإنسان وعلاقته بالآلات. ويجدُر بنا أن نُوضِّح هذه السرديات لكي نفهم بعض هذه الأفكار على نحوٍ أفضل ونضعها في سياقها الصحيح. وبشكلٍ عام، فإنه من المُهم أن ندمج بحث السرديات في أخلاقيات الذكاء الاصطناعي، على سبيل المثال، لكي نفهم الأسباب التي تجعل بعض السرديات مُنتشرة، ومَن أنشأها، ومَن الذي يستفيد منها (Royal Society 2018). كما يمكن أن يُساعدنا في إنشاء سرديَّات جديدة حول مستقبل الذكاء الاصطناعي.

وحش فرانكنشتاين الجديد

مِن السبل التي يُمكننا اتخاذها لتجاوُز الضجة المثارة أن نفكِّر في بعض السرديات ذات الصلة من تاريخ الثقافة البشرية التي تُشكل المناقشة العامة الحالية حول الذكاء الاصطناعي. فليست هذه هي المرة الأولى التي يتساءل فيها الناس عن مُستقبل البشرية ومُستقبل التكنولوجيا. ومهما كانت بعض الأفكار المتعلقة بالذكاء الاصطناعي تبدو غريبة، فإننا يُمكننا استكشاف صِلتها بأفكار وسرديات أكثر شهرة توجَد في وعينا الجمعي، أو بشكلٍ أدق، في الوعي الجماعي للغرب.

أولًا، هناك تاريخ طويل للتفكير في البشر والآلات أو المخلوقات الاصطناعية في الثقافات الغربية وغير الغربية على حدٍّ سواء. يُمكن العثور على فكرة إنشاء كائنات حية من مادة غير حية في قصص الخلق في الثقافات السومرية والصينية واليهودية والمسيحية والإسلامية. فقد كانت لدى الإغريق فكرة إنشاء بشَر اصطناعيين، وخاصة النساء الاصطناعيات. على سبيل المثال، في الإلياذة، يُقال إن هيفايستوس يقوم على خدمته خَدَم مصنوعون من الذهب يُشبهون النساء. وفي أسطورة بيجماليون الشهيرة، يقع النحَّات في حُب تمثال امرأة صنَعَه من العاج. ويتمنَّى أن تدبَّ فيه الروح ويُصبح امرأة حقيقية، فتُحقِّق له الإلهة أفروديت أُمنيته: فتصبح شفتاها دافئتَين وجسدُها ناعمًا. ويُمكننا بسهولة هنا ملاحظة الصِّلة بين ذلك وبين الروبوتات الجنسية المعاصرة.

هذه السرديَّات لا تأتي فقط من الأساطير: ففي كتابه «الأوتوماتا»، قدَّم عالِم الرياضيات والمهندس الإغريقي هيرون السكندري (ولد عام ١٠) أداة اكتشفَت في البحر، وهي آلية «أنتيكيثيرا»، التي تُحدد أنها كمبيوتر تناظُري إغريقي يعتمد على آلية مُعقَّدة من التروس والمُسنَّنات. ولكن القصص الخيالية التي تجعل الآلات تُشبِه البشر تسلُب ألبابنا بشكلٍ خاص. فلنأخذ، على سبيل المثال، أسطورة الجوليم: وحش مصنوع من الطين صنَعَه حاخام في القرن السادس عشر، ثم فقَدَ السيطرة عليه. هنا نواجِهُ نسخة مُبكِّرة من مشكلة التحكُّم. ويمكن تفسير أسطورة بروميثيوس بهذه الطريقة أيضًا؛ إذ يسرق النار من الآلهة ويُعطيها إلى البشر، لكنه يُعاقَب بعد ذلك. وعقوبته الأبدية هي أن يُربط بصخرةٍ بينما يأكل النسر كبِدَه كلَّ يوم. وقد كان الدرس القديم من هذه الأسطورة هو التحذير من الغطرسة: فهذه القدرات ليست مُقدَّرة للبشر.

ومع ذلك، في رواية ماري شيلي «فرانكنشتاين» — التي تحمل العنوان الفرعي الدال «بروميثيوس الحديث» — يُصبح إنشاء حياة ذكية من مادة غير حيَّة مشروعًا

علميًّا حديثًا. حيث ينشئ العالم فيكتور فرانكنشتاين كائنًا شبيهًا بالإنسان من أجزاء الجثث، لكنه يفقد السيطرة عليه. ومع أن الحاخام استطاع أن يُسيطر على الجوليم في النهاية، فإن الأمر ليس كذلك في هذه الحالة. ويمكن اعتبار فرانكنشتاين رواية رومانسية تُحذِّر من التكنولوجيا الحديثة، ولكنها تستند إلى العلم في زمنِها. على سبيل المثال، يلعب استخدام الكهرباء — وهي تقنية جديدة جدًّا في ذلك الوقت — دورًا مهمًّا؛ إذ تُستخدَم لإحياء الجثة. كما أنها تُشير إلى المغناطيسية وعِلم التشريح. في ذلك الوقت، كان المفكِّرون والكتَّاب يناقشون طبيعة الحياة وأصلها. ما قوة الحياة؟ لقد تأثرت ماري شيلي بعلوم عصرها.[2] وتُظهر القصة كيف كان الرومانسيون في القرن التاسع عشر مفتونين في كثيرٍ من الأحيان بالعلم، فضلًا عن أملهم في أن يُحرِّرنا الشِّعر والأدب من الجوانب الأكثر ظُلمةً في الحداثة (Coeckelbergh 2017). يجب ألَّا نعتبر هذه الرواية بالضرورة ضد العلم والتكنولوجيا؛ إذ يبدو أن الرسالة الرئيسية التي تحرص على توصيلها هي أن العلماء ينبغي أن يتحملوا مسئولية اختراعاتهم. يهرب الوحش، ولكنه يفعل ذلك لأن صانعه يرفضه. يجب أن نتذكَّر هذا الدرس فيما يتعلَّق بأخلاقيات الذكاء الاصطناعي. ومع ذلك، تؤكِّد الرواية بوضوح خطر التكنولوجيا التي تخرج عن السيطرة، وعلى وجه الخصوص خطر البشر الاصطناعيين الذين يُصيبهم الجنون. تعود هذه المخاوف للظهور على السطح في القلق المُعاصر من أن يخرج الذكاء الاصطناعي عن السيطرة.

> في رواية ماري شيلي «فرانكنشتاين» — التي تحمل العنوان الفرعي الدال «بروميثيوس الحديث» — يُصبح إنشاء حياة ذكية من مادة غير حية مشروعًا علميًّا حديثًا.

وعلاوةً على ذلك، كما هو الحال في رواية «فرانكنشتاين» وأسطورة «الجوليم»، تظهر سردية المنافسة: فالمخلوقات الاصطناعية تتنافس مع الإنسان. وتستمرُّ هذه السردية في تشكيل خيالنا العلمي حول الذكاء الاصطناعي، ولكنها أيضًا تؤثِّر على تفكيرنا المُعاصر في التكنولوجيا مثل الذكاء الاصطناعي والروبوتات. فلنأخُذ مسرحية «روبوتات روسوم العالمية» التي كتبت عام ١٩٢٠ مثالًا، وهي تتناول قصة الروبوتات العبيد التي تتمرَّد على سيدها وتثور عليه، أو فيلم «٢٠٠١: سبيس أوديسي» (٢٠٠١: أوديسة الفضاء) الذي أنتج عام ١٩٦٨ والذي ذكرناه من قبلُ، ويتحدَّث عن ذكاء اصطناعي يبدأ في قتل طاقم المركبة الفضائية لتحقيق مهمَّته، أو فيلم «إكس ماكينا» الذي أنتج عام ٢٠١٥

ويروي قصة روبوت الذكاء الاصطناعي «أفا» التي تنقلب على صانعها. كما يندرج تحت سردية الآلات التي تتمرَّد علينا مجموعة أفلام «المُدمِّر» (ترمينيتور). وقد وصف كاتب الخيال العلمي أيزاك أسيموف هذا الخوف بـ «عقدة فرانكنشتاين»: الخوف من الروبوتات. ويرتبط هذا أيضًا بالذكاء الاصطناعي اليوم. وهو أمر يتعيَّن على العلماء والمُستثمرين التعامُل معه. فبعضهم يُحاربون هذا الخوف؛ وبعضهم يساعد في خلقه والحفاظ عليه. وقد أشرتُ بالفعل إلى مثال «ماسك». وثمَّة مثال آخَر على شخصية مؤثرة ساهمت في نشر الخوف من الذكاء الاصطناعي وهو عالم الفيزياء ستيفن هوكينج، الذي صرَّح في عام ٢٠١٧ بأن خلق الذكاء الاصطناعي يمكن أن يكون أسوأ حدَثٍ في تاريخ حضارتنا (Kharpal 2017). إن «عقدة فرانكنشتاين» منتشرة وعميقة الجذور في الثقافة والحضارة الغربية.

التسامي ونهاية العالم بسبب الذكاء الاصطناعي

ثمة مقدمات لأفكار مثل «تجاوز الإنسانية» و«التفرُّد التكنولوجي» في تاريخ التفكير الديني والفلسفي الغربي أو على الأقل توجَد أفكار مشابهة لها، ولا سيما في التقاليد اليهودية المسيحية وفي الفكر الأفلاطوني. وعلى عكس ما يعتقده الكثيرون، فإن الدين والتكنولوجيا كانا دائمًا مُترابطَين في تاريخ الثقافة الغربية. ودعوني أحصر نقاشي هنا في التسامي ونهاية العالم.

في الدين اللاهوتي، يقصد بالتَّسامي أن الإله «فوق» العالم المادي والجسدي ومُستقل عنه، وهي فكرة مُناقضة لفكرة أنه موجود في العالَم وأنه جزء منه (الحلولية). في التقليد اليهودي المسيحي الأحادي اللاهوتي، يُرى الله على أنه يتسامى فوق خلقه. ويُمكن في الوقت نفسه أيضًا أن يُرى على أنه مُتغلغل في كل مخلوقاتِه وفي كل الكائنات (أي إنه يحلُّ فيها)، وعلى سبيل المثال، في اللاهوت الكاثوليكي، يُفهم الله كما يتجلَّى من خلال ابنه (المسيح) والروح القدس. ويبدو أن سرديات الذكاء الاصطناعي التي تتجلى فيها «عقدة فرانكنشتاين» تؤكد فكرة التسامي بمعنى أن هناك انفصالًا أو فجوة بين الخالق والمخلوق (بين الإنسان الإله والذكاء الاصطناعي)، دون إعطاء الكثير من الأمل في إمكانية تجاوز هذه الفجوة.

على عكس ما يعتقده الكثيرون، فإن الدين والتكنولوجيا كانا دائمًا مُترابطَين في تاريخ الثقافة الغربية.

التسامي يمكن أيضًا أن يُشير إلى تجاوز الحدود، أو تخطِّي شيءٍ ما. في التاريخ الديني والفلسفي الغربي، اتخذت هذه الفكرة في كثيرٍ من الأحيان شكلَ السمو فوق العالم المادي والجسدي وتجاوُز حدوده. على سبيل المثال، في منطقة البحر المتوسِّط في القرن الثاني الميلادي، كانت الغنوصية تنظر إلى المادَّة جميعها باعتبارها شرًّا، وتهدف إلى تحرير الشعلة الإلهية من الجسد البشري. وفي وقتٍ أسبَق، رأى أفلاطون الجسد سجنًا للروح. وعلى عكس الجسد، كان ينظر إلى الروح على أنها خالدة. وفي الميتافيزيقا الخاصة به، ميَّز أفلاطون بين الأشكال، التي هي أبدية، والأشياء الموجودة في العالم، التي تتغير؛ فالأولى تتسامى فوق الأخيرة وتتجاوزها. وهناك أفكار في مبدأ تجاوز الإنسانية تُذكِّرنا بهذا. فهي تُحافظ على هدف التسامي بمعنى تجاوز القيود البشرية، وليس هذا فحسب، بل إن الطرق الخاصة التي يُفترض أن يحدث بها هذا التسامي تستحضر أفلاطون والغنوصية: لتحقيق الخلود، يجب التسامي فوق الجسد البيولوجي عن طريق تحميل أدواتٍ اصطناعية وتطويرها. بشكلٍ أكثر عمومية، عندما يَستخدِم الذكاء الاصطناعي والعلوم والتكنولوجيا ذات الصلة الرياضيات لاستخلاص أشكالٍ أكثر نقاءً من العالَم المادي الفوضوي، يمكن تفسير ذلك على أنه برنامج أفلاطوني يتحقَّق بواسطة وسائل تكنولوجية. ومن هنا يتبيَّن أن خوارزمية الذكاء الاصطناعي هي آلة أفلاطونية تستخلِص شكلًا (أو نموذجًا) من عالَم الظواهر (البيانات).

التسامي يمكن أيضًا أن يعني تجاوز الحالة الإنسانية. في التقليد المسيحي، يمكن أن يأخذ هذا شكل محاولة رأب الفجوة بين الله والبشر من خلال تحويل البشر إلى آلهة، ربما عن طريق استعادة تشابُههم مع الآلهة وكمالهم الأصلي (Noble 1997). ولكن سَعْي مؤيدي تجاوز الإنسانية للخلود ليس جديدًا، بل يمكن تتبُّعه إلى العصور القديمة. إذ يُمكننا أن نجده في الميثولوجيا الميزوبوتامية (الأساطير التي تأتي من منطقة ما بين النهرَين): تحكي لنا قصة «ملحمة جلجامش»، وهي واحدة من أقدم القصص المكتوبة عن البشرية، عن ملك أوروك (جلجامش)، الذي يبحث عن الخلود بعد وفاة صديقه إنكيدو. ولكنه يفشل في العثور عليه: ومع ذلك، ينجح في الحصول على نبتةٍ يُقال إنها تُعيد الشباب، ولكن تسرقها أفعى، وفي النهاية، يتعيَّن عليه أن يتعلَّم الدرس بأن عليه مواجهة

حقيقة موتِه هو شخصيًّا؛ إذ إن سعيه إلى الخلود بلا جدوى. على مرِّ التاريخ، كان الناس يبحثون عن إكسير الحياة. واليوم، تبحث العلوم عن علاجاتٍ مضادَّة للشيخوخة. ومِن هذا المنطلق، فإن سعي مؤيدي مبدأ تجاوز الإنسانية إلى الخلود أو إلى إطالة العمر ليس جديدًا أو غريبًا؛ بل هو واحد من أقدم أحلام البشرية وأهداف العلم المُعاصر. وفي أيدي مؤيدي تجاوز الإنسانية، يُصبح الذكاء الاصطناعي هو أداة التجاوز التي تَعِدنا بالخلود.

من المفاهيم القديمة الأخرى التي تساعدنا على وضع أفكار تجاوز الإنسانية في سياقها، ولا سيما فكرة التفرُّد التكنولوجي، مفهوم نهاية العالم (أبوكاليبس) والأخروية. ومصطلح «أبوكاليبس» عند الإغريق القدماء، الذي يلعب أيضًا دورًا في الفكر اليهودي والمسيحي، يُشير إلى كشف الحجاب. وفي الوقت الحاضر، يُشير هذا المصطلح غالبًا إلى نوع معيَّن من الكشف: وهو كشف سيناريو نهاية الزمان أو نهاية العالم. وفي السياقات الدينية، نجد مصطلح «الأخروية»: وهو جزء من علم اللاهوت يتعلَّق بالأحداث النهائية للتاريخ والمصير النهائي للبشرية. وتنطوي معظم الأفكار الأخروية وتلك التي تتعلَّق بنهاية العالم على تخريب أو تدمير جِذري وغالبًا عنيف للعالم، والاتجاه نحو مستوى أعلى من الواقع والكينونة والوعي. ويُذكرنا ذلك أيضًا بالطوائف والجماعات المتطرفة المتشائمة التي كانت وما تزال تتنبَّأ بالكوارث ونهاية العالم. ورغم أن مؤيدي تجاوز الإنسانية في العادة ليس لهم علاقة بمثل هذه الطوائف والممارسات الدينية، فإن فكرة التفرد التكنولوجي تُشبه إلى حدٍّ ما سرديات نهاية العالَم والأخروية والتنبؤ بالكوارث، وهذا أمر واضح.

بالتالي، بينما يستند تطوير الذكاء الاصطناعي إلى علمٍ من المُفترَض أنه لا خيالي ولا ديني، وبينما ينأى مؤيدو تجاوز الإنسانية بأنفسهم عادةً عن الدين ويرفضون أيَّ اقتراحٍ بأن أعمالهم تستنِد إلى الخيال، إلا أن الخيال العلمي والأفكار الدينية والفلسفية القديمة تلعب بالضرورة دورًا مُهمًّا عندما نناقش مُستقبل الذكاء الاصطناعي من هذا المنطلق.

كيفية تجاوز سرديات المنافسة وتجاوز الضجَّة المُثارة حول الذكاء الاصطناعي

يمكن للمرء أن يتساءل الآن: هل هناك سبُل للنجاة؟ هل يُمكننا تجاوز سرديات المنافسة وإيجاد طرقٍ أكثر رسوخًا لفهم مستقبل الذكاء الاصطناعي والتكنولوجيا المُماثلة؟ أم إن التفكير الغربي حول الذكاء الاصطناعي محكوم عليه بالبقاء في سجن هذه المخاوف

العصرية وجذورها القديمة؟ هل يُمكننا تجاوز الضجة المثارة حول الذكاء الاصطناعي؟ أم ستظلُّ المناقشة مُنصَبَّة على الذكاء الفائق؟ أعتقد أن لدَينا سبلًا للنجاة.

رغم أن مؤيدي تجاوز الإنسانية في العادة ليس لهم علاقة بمثل هذه الطوائف والمُمارسات الدينية، فإن فكرة التفرُّد التكنولوجي تُشبه إلى حدٍّ ما سرديات نهاية العالم والأخروية والتنبؤ بالكوارث.

أولًا، يمكننا تجاوز الثقافة الغربية للعثور على أنواع مختلفة من السرديَّات غير المَبنية على «عقدة فرانكنشتاين» فيما يخص التكنولوجيا وطرق التفكير غير الأفلاطونية. على سبيل المثال، في اليابان حيث تتأثر ثقافة التكنولوجيا بديانات الطبيعة أكثر من الغرب، وتحديدًا بديانة الشنتو، وحيث صوَّرت الثقافة الشعبية الآلات كمُساعدين، نجد موقفًا أكثر ودًّا تجاه الروبوتات والذكاء الاصطناعي. هنا، لا نجد عقدة فرانكنشتاين. وتنطوي طريقة التفكير التي يُطلق عليها أحيانًا «الأرواحية» على أن الذكاء الاصطناعي يمكن أيضًا من حيث المبدأ أن يمتلك روحًا أو نفسًا، ويمكن أن يُعتبَر مقدسًا. وهذا يعني عدم وجود سردية تنافُسية؛ وعدم وجود رغبة أفلاطونية في تجاوز المادية والدفاع المُستمر عن الإنسان بوصفه كائنًا يسمو فوق الآلة ويتجاوزها، أو يختلف عنها اختلافًا جوهريًّا. في حدود معرفتي، لا تشتمِل الثقافة الشرقية على أفكار حول نهاية الزمان. وعلى عكس الديانات التوحيدية، تحمل ديانات الطبيعة فهمًا دوريًّا للزمن. وبالتالي، يمكن أن يساعد النظر إلى ما هو أبعد من الثقافة الغربية (أو في واقع الأمر إلى الماضي القديم للغرب، حيث نجد أيضًا ديانات طبيعة) في التقييم النقدي للسرديات السائدة حول مستقبل الذكاء الاصطناعي.

ثانيًا: لتجاوز الضجة المُثارة حول الذكاء الاصطناعي وتجنُّب حصر مناقشة أخلاقيات الذكاء الاصطناعي في أحلام المستقبل البعيد وكوابيسه، يُمكننا (١) استخدام الفلسفة والعلم لفحص ومناقشة الافتراضات المتعلِّقة بالذكاء الاصطناعي والإنسان الذي يلعب دورًا في هذه السيناريوهات والمناقشات (مثل: هل الذكاء العام مُمكن؟ ما الفارق بين الإنسان والآلة؟ ما العلاقة بين الإنسان والتكنولوجيا؟ ما الوضع الأخلاقي للذكاء الاصطناعي؟)؛ و(٢) النظر بتفصيلٍ أكثر إلى ماهية الذكاء الاصطناعي الموجود وما يفعله اليوم في التطبيقات المختلفة؛ و(٣) مناقشة المشكلات الأخلاقية والاجتماعية الأكثر واقعيةً وإلحاحًا التي يُثيرها الذكاء الاصطناعي كما يُطبق اليوم؛ و(٤) التفكير في سياسة

الذكاء الاصطناعي للمستقبل القريب؛ و(٥) طرح تساؤل عما إذا كان التركيز على الذكاء الاصطناعي في الخطاب الجماهيري الحالي مُفيدًا في ضوء المشكلات الأخرى التي تُواجهنا، وما إذا كان تركيزنا ينبغي أن ينصبَّ على الذكاء الاصطناعي وحدَه. وسوف نتبع هذه المسارات في الفصول القادمة من الكتاب.

الفصل الثالث

كل ما له علاقة بالبشر

هل الذكاء الاصطناعي العام مُمكن؟
هل هناك فروق جوهرية بين الإنسان والآلة؟

تفترض رؤية أنصار تجاوز الإنسانية للمُستقبل التكنولوجي أن الذكاء الاصطناعي العام (أو الذكاء الاصطناعي القوي) ممكن، ولكن هل هو كذلك؟ بعبارةٍ أخرى، هل يُمكننا إنشاء آلات تتمتَّع بقدرات معرفية تُشبه تلك الخاصة بالبشر؟ إذا كانت الإجابة لا، فإن رؤية الذكاء الفائق بالكامل تُصبح غير ذات صِلة بأخلاقيات الذكاء الاصطناعي. فإذا كان مِن المُستحيل أن تتمتَّع الآلات بالذكاء البشري العام، فإننا غير مُضطرين إلى أن نقلق بشأن الذكاء الفائق. بشكلٍ عام، يبدو أن تقييمنا للذكاء الاصطناعي يعتمِد على فهمنا لماهية الذكاء الاصطناعي في الوقت الحالي وما يُمكن أن يصبح عليه في المستقبل، كما يعتمد على رؤيتنا للفروق بين الإنسان والآلة. على الأقل منذ منتصف القرن العشرين، ناقش الفلاسفة والعلماء ما تستطيع أجهزة الكمبيوتر أن تقوم به وما يُمكن أن تُصبح عليه، والفروق بين الإنسان والآلة الذكية. دعونا نُلقي نظرةً على بعض هذه النقاشات، التي تتناول ماهية الإنسان وما يجب أن يكون عليه، بقدْر ما تتناول ماهية الذكاء الاصطناعي وما يجب أن يكون عليه.

هل يمكن لأجهزة الكمبيوتر أن تتمتَّع بالذكاء والوعي والإبداع؟ هل يُمكنها فهم الأشياء وإدراك المعاني؟ هناك تاريخ من النقد والشك في إمكانية وجود ذكاء اصطناعي مُشابِهٍ لذكاء الإنسان. في عام ١٩٧٢، نشر هيوبرت دريفوس، فيلسوف ذو خلفية في علم الظواهر، كتابًا بعنوان «ما لا تستطيع أجهزة الكمبيوتر فعله».[1] منذ الستينيَّات، كان دريفوس يُظهر انتقادًا شديدًا للأساس الفلسفي للذكاء الاصطناعي وشكَّك في وعوده: وقال إن برنامج الذكاء الاصطناعي البحثي محكوم عليه بالفشل. وقبل أن ينتقل إلى

بيركلي، كان يعمل في معهد ماساتشوستس للتكنولوجيا، وهو مكان مُهم لتطوير الذكاء الاصطناعي، والذي كان يعتمد أساسًا في ذلك الوقت على المُعالجة الرمزية. رأى دريفوس أن الدماغ ليس جهاز كمبيوتر وأن العقل لا يعمل عن طريق المُعالجة الرمزية. إن لدَينا خلفية لا واعية من المعرفة المشتركة القائمة على الخبرة وما يمكن أن يُطلِق عليه هايدجر «كينونتنا في العالم»، وهذه المعرفة ضِمنية ولا يمكن تشكيلها. وتعتمد خبرة الإنسان، حسب رأي دريفوس، على المُمارسة بدلًا من المعرفة. ولا يستطيع الذكاء الاصطناعي الْتِقاط هذا المعنى والمعرفة الضمنية؛ وإذا كان هذا هو هدف الذكاء الاصطناعي، فهذا محضُ أساطير. فالبشر وحدَهم قادرون على رؤية ما هو ذو صلة لأنهم، بوصفِهم كائنات مُتجسِّدة ووجودية، يشاركون في العالم وقادرون على الاستجابة لمتطلبات الوضع.

هناك تاريخ من النقد والشك في إمكانية وجود ذكاء اصطناعي مُشابِهٍ لذكاء الإنسان.

في ذلك الوقت، واجهَ دريفوس الكثير من المعارضة، ولكن في وقتٍ لاحق، لم يعُد الكثيرون من باحثي الذكاء الاصطناعي يعِدُون بتحقيق الذكاء الاصطناعي العام أو يتوقَّعون تحقيقه. وانتقلت أبحاث الذكاء الاصطناعي من الاعتماد على مُعالجة الرموز إلى نماذج جديدة، ومنها تعلُّم الآلة القائم على الإحصاء. وفي حين كانت هناك فجوة هائلة في وقت دريفوس بين عِلم الظواهر والذكاء الاصطناعي، فإن العديد من باحثي الذكاء الاصطناعي اليوم يعتنِقون مناهج العلوم المعرفية المتجسِّدة والموجودة، التي تدَّعي أنها أقرب إلى علم الظواهر.

ومع ذلك، فإن اعتراضات دريفوس لا تزال صائبةً وتُظهر كيف يمكن أن تتعارَض وجهات نظر الإنسان غالبًا مع الآراء العلمية، خاصة — ولكن ليس حصريًّا — فيما يُسمَّى بالفلسفة القارية. يُشدِّد الفلاسفة القاريون عادةً على أن البشر والعقول البشرية مختلفة اختلافًا جوهريًّا عن الآلات، ويُركِّزون على التجربة الإنسانية الواعية والوجود الإنساني، الذي لا يمكن ولا ينبغي اختزاله في أوصاف شكلية أو تفسيرات علمية. من جهة أخرى، يؤيد بعض الفلاسفة — غالبًا من منطلق التقليد التحليلي للفلسفة — رؤية للإنسان تدعم الباحثين في مجال الذكاء الاصطناعي الذين يعتقدون أن الدماغ والعقل البشري يُشبهان ويعملان حقًّا مثل نماذج الكمبيوتر الخاصة بهم. ومن أمثلة هؤلاء الفلاسفة

بول تشيرشلاند ودانييل دنيت. يعتقد تشيرشلاند أن العلم، وخاصة عِلم الأحياء التطوُّري وعلم الأعصاب، والذكاء الاصطناعي يُمكنهما تفسير الوعي البشري تفسيرًا كاملًا. ويعتقد أن الدماغ عبارة عن شبكةٍ عصبية مُتكرِّرة. وينكر وجود أفكار أو تجارب غير مادِّية فيما يُطلَق عليه المادية الإقصائية. فما نُسمِّيه أفكارًا وتجارب ما هو إلا حالات للدماغ. وينكر دنيت أيضًا وجود أي شيءٍ بخلاف ما يحدُث في الجسم: ويرى أننا «نحن أنفسنا نوع من الروبوتات» (Dennett 1997). وإذا كان الإنسان في الأساس آلة واعية، فإن مثل هذه الآلات مُمكنة، وليس فقط من حيث المبدأ ولكن في الواقع. يُمكننا أن نحاول صنعها. ومن الأهمية بمكانٍ أن كلًّا من الفلاسفة القاريين والتحليليين يُعارضان الثنائية الديكارتية التي تفصل بين العقل والجسم، ولكن لأسبابٍ مختلفة: فالفلاسفة القاريون يعتقدون أن وجود الإنسان يتعلَّق بكونه في العالَم الذي لا يُفصل فيه العقل عن الجسم، أما الفلاسفة القاريون فيعتقدون لأسبابٍ مادية أن العقل ليس شيئًا مُستقلًّا عن الجسم.

ولكن ليس جميع الفلاسفة التحليليِّين يرَون أن الذكاء الاصطناعي العام أو القوي مُمكن. من وجهة نظر الفيلسوف فيتجنشتاين (في وقتٍ لاحق)، يمكن للشخص أن يُجادل بأنه في حين يمكن لمجموعةٍ من القواعد أن تصف ظاهرةً معرفية، فإن ذلك لا يعني بالضرورة أن لدَينا فعليًّا قواعد في رءوسنا (Arkoudas and Bringsjord 2014). كما هو الحال مع انتقاد دريفوس، يُثير هذا مشكلة لنوعٍ واحد من أنواع الذكاء الاصطناعي، وهو الذكاء الاصطناعي الرمزي، إذا افترض أن هذه هي الطريقة التي يُفكِّر بها البشر. ثمَّة انتقاد فلسفي آخَر للذكاء الاصطناعي يأتي من جون سيرل، الذي يُعارض فكرة أن برامج الكمبيوتر يمكن أن تكون لديها حالات معرفية حقيقية أو فهم للمعنى (Searle 1980). وفيما يلي التجربة الفكرية التي يُقدِّمها، والتي تُعرَف باسم حجَّة الغرفة الصينية: يُحبَس سيرل في غرفة ويُعطى كتابات صينية ولكنه لا يعرف الصينية. ومع ذلك، يستطيع الرد على الأسئلة التي يطرحها أشخاصٌ خارج الغرفة يتحدثون بالصينية لأنه يستخدم كُتيِّب القواعد الذي يُمكِّنه من إنتاج الإجابات الصحيحة (مُخرجات) استنادًا إلى المستندات (المدخلات) التي يتلقَّاها. وهو يستطيع القيام بذلك بنجاحٍ دون فهم اللغة الصينية. وبالمثل، يُجادل سيرل، يُمكن لبرامج الكمبيوتر إنتاج مُخرجَات استنادًا إلى مدخلات بالاستعانة بالقواعد التي تُزوَّد بها، ولكنها لا تفهم شيئًا. بمصطلحات فلسفية أكثر تخصُّصًا: لا تمتلك برامج الكمبيوتر قصدية، ولا يمكن خلق فهم حقيقي بواسطة الحوسبة الشكلية. أو كما يقول بودن (٢٠١٦)، الفكرة هي أن المعنى يأتي من البشر.

على الرغم من أن برامج الكمبيوتر الحالية للذكاء الاصطناعي غالبًا ما تختلف عن تلك التي انتقدها دريفوس وسيرل، فإن النقاش لا يزال مُستمرًّا. يعتقد العديد من الفلاسفة أن هناك فروقًا حاسمة بين طريقة تفكير البشر وأجهزة الكمبيوتر. على سبيل المثال، يمكن للمرء اليوم أن يُجادل بأننا كائنات قادرة على خَلْق المعنى، وواعية ومُتجسِّدة وحية، ولا يمكن تفسير طبيعتنا وعقولنا ومعرفتنا بالمقارنة بالآلات. ومع ذلك، عليك أن تلاحظ أنه حتى العلماء والفلاسفة الذين يعتقدون أن هناك الكثير من التشابُه بين البشر والآلات من حيث المبدأ، وأن الذكاء الاصطناعي العام مُمكن نظريًّا، يرفضون في كثيرٍ من الأحيان رؤية بوستروم للذكاء الفائق وأفكار مُماثلة تَعتبِر أن الذكاء الاصطناعي المُشابه لذكاء الإنسان قد أصبح قاب قوسين أو أدنى من التحقُّق. فبودن ودنيت كلاهما يعتقدان أن الذكاء الاصطناعي العام صعب جدًّا تحقيقه عمليًّا، وبالتالي ليس شيئًا يجب القلق بشأنه في الوقت الحالي.

> نحن كائنات قادرة على خَلْق المعنى، وواعية ومتجسدة وحية، ولا يمكن تفسير طبيعتنا وعقولنا ومعرفتنا بالمقارنة بالآلات.

وبناءً عليه يمكننا القول إن هناك، في خلفية النقاش حول الذكاء الاصطناعي، تبايُن عميق في الآراء حول طبيعة الإنسان والذكاء البشري والعقل والفهم والوعي والإبداع والمعنى والمعرفة البشرية والعلوم، وهكذا. فإذا كان ثمة «معركة» من الأساس، فهي معركة تتعلَّق بالإنسان بقدْر ما تتعلَّق بالذكاء الاصطناعي.

الحداثة و(ما بعد) الإنسانية وما بعد الظاهرية

من وجهة نظرٍ أوسَع في العلوم الإنسانية، من المهم أن نضع هذه النقاشات حول الذكاء الاصطناعي والإنسان في سياقٍ أوسَع للوقوف على ماهيتها وما تنطوي عليه. فهذه النقاشات لا تتعلَّق بالتكنولوجيا والإنسان فحسب، ولكنها تعكس انقساماتٍ عميقة في الحداثة. دعوني أمرُّ مرور الكرام على ثلاثة انقسامات تُساهم بشكلٍ غير مباشر في تشكيل المناقشات الأخلاقية حول الذكاء الاصطناعي. الانقسام الأول هو انقسام ظهَرَ في مُستهَلِّ عصر الحداثة بين حركتَي التنوير والرومانسية. أما الآخَران فهما تطوُّرات حديثة

نسبيًّا: الأول بين الإنسانية وتجاوز الإنسانية، ويبقى حبيس توترات الحداثة، والثاني بين الإنسانية وما بعد الإنسانية، والذي يُحاول تخطِّي الحداثة.

إحدى وسائل فهم النقاش حول الذكاء الاصطناعي والإنسان هي أن نضع في الاعتبار التوتُّر القائم بين التنوير والرومانسية في الحداثة. في القرنَين الثامن عشر والتاسع عشر، تحدى العلماء والمُفكرون التنويريون الآراء الدينية التقليدية وزعموا أن العقل والشك والعلم تُظهِر لنا ماهية الإنسان والعالم الحقيقية، على عكس المُعتقدات المُسلَّم بها غير المُبررة بالحجج أو غير المدعومة بالأدلة. وكانوا متفائلين حيال ما يمكن أن يقدِّمه العلم لصالح الإنسانية. ردًّا على ذلك، قال الرومانسيون إن العقل المجرَّد والعلم الحديث قد أفقدا العالَم سِحره وأننا في حاجة إلى إعادة الغموض والسحر اللذَين يُريد العلم القضاء عليهما. عند النظر إلى النقاش حول الذكاء الاصطناعي، يبدو لنا أننا لم نبتعِد كثيرًا عن ذلك. على سبيل المثال، يستهدف عمل دنييت حول الوعي وعمل بودين حول الإبداع تقديم تفسيراتٍ لكل شيء، أو كما يقول دنييت «فك السحر». فهذان الفيلسوفان مُتفائلان بأن العلم يُمكنه كشف غموض الوعي والإبداع وغيرهما. إنهما يُعارضان كلَّ مَن يقاوم جهود فك سحر الإنسان، مثل الفلاسفة القاريين الذين يسيرون في ركب ما بعد الحداثة ويُشدِّدون على غموض معنى أن تكون إنسانًا؛ بعبارة أخرى: الرومانسيين الجُدد. يبدو أن سؤال «هل نفكُّ السحر أم نحتفظ بغموض الإنسان؟» هو السؤال الرئيسي في المناقشات التي تتناول الذكاء الاصطناعي العام ومُستقبله.

أما التوتر الثاني فهو بين مؤيدي الإنسانية ومؤيدي تجاوز الإنسانية. ما هو «الإنسان»، وماذا يجب أن يكون؟ هل من المُهم الدفاع عن الإنسان كما هو، أم يتعيَّن علينا تعديل تصوُّرنا له؟ يحتفي دُعاة الإنسانية بالإنسان كما هو. ومن الناحية الأخلاقية، يُشدِّدون على القيمة الجوهرية والمتفوِّقة للبشر. ويُمكننا العثور على أفكار دعاة الإنسانية في النقاش الدائر عن الذكاء الاصطناعي في الحجج التي تُدافع عن حقوق الإنسان وكرامته كأساسٍ لأخلاقيات الذكاء الاصطناعي، أو في الحجة المؤيدة لأن يكون البشر وقِيَمُهم في قلب وفي مركز مسألة تطوير الذكاء الاصطناعي ومُستقبله. هنا غالبًا ما تتَّفق الإنسانية مع التفكير التنويري. ولكن يُمكن أن تأخذ أيضًا أشكالًا أكثر تحفظًا أو رومانسية. كذلك يُمكننا أن نعثر على الإنسانية في مقاومة مشروع دُعاة تجاوز الإنسانية. فبينما يعتقد دُعاة تجاوز الإنسانية أنَّ علينا المُضِي قدمًا نحو نوع جديدٍ من الإنسان يتم تحسينُه بواسطة العلم والتكنولوجيا، يدافع الإنسانيون عن الإنسان كما هو، ويشددون على قيمته وكرامته، التي يُقال إنها مهدَّدة من قِبل علوم دعاة تجاوز الإنسانية وفلسفتهم.

ردود الفعل الدفاعية تجاه التكنولوجيا الجديدة لها تاريخها الخاص. ففي العلوم الاجتماعية والإنسانية، كثيرًا ما تُنتقَد التكنولوجيا باعتبارها تهديدًا للإنسانية والمجتمع. على سبيل المثال، كان كثيرٌ من فلاسفة القرن العشرين شديدي التشاؤم حيال العلم، وحذروا من سيطرة التكنولوجيا على المجتمع. ولكن الصراع الآن لا يتعلَّق فقط بحياة الإنسان والمجتمع، بل يتعلَّق بالإنسان نفسه: هل نحن بصدد تحسينه وتطويره أم لا؟ هذا هو السؤال. فمن جهة، يُصبح الإنسان نفسه مشروعًا علميًّا تكنولوجيًّا، قابلًا للتحسين والتطوير. وبمجرَّد أن يُفَك سحر الإنسان — من خلال داروين وعلم الأعصاب والذكاء الاصطناعي — يُمكننا أن نبدأ في تحسينه. ويمكن للذكاء الاصطناعي أن يُساعدنا في تحسين الإنسان. ومن جهة أخرى، يجب علينا أن نحتضِن الإنسان كما هو. وربما يقول البعض: دائمًا ما يفوتنا أن نُدرك ماهية الإنسان. فنحن لا نستطيع أن نفهمه فهمًا تامًّا بواسطة العلم.

تستمر هذه التوتُّرات في تقسيم العقول والقلوب في هذا النقاش. فهل يُمكننا تخطِّيها؟ عمليًّا، يمكن للمرء أن يتخلَّى عن هدف إنشاء ذكاءٍ اصطناعي شبيه بالإنسان. ولكن حتى في هذه الحالة، تظلُّ هناك خلافات بشأن وضع «آلات الذكاء الاصطناعي كنماذج للبشر» المُستخدَم في علم الذكاء الاصطناعي. هل تُعلِّمنا حقًّا شيئًا عن كيفية تفكير البشر؟ أم إنها تُعلِّمنا فقط شيئًا عن نوعٍ معيَّن من التفكير، على سبيل المثال تفكير يمكن صياغته بواسطة الرياضيات، أو تفكير يهدف إلى السيطرة والتلاعُب؟ إلى أي مدًى يُمكننا حقًّا التعلُّم من هذه التقنيات عن الإنسان؟ هل البشرية أكبر مما يستطيع العلم أن يُدرك؟ حتى في المناقشات الأكثر اعتدالًا، تظهر الصراعات بشأن الحداثة.

للخروج من هذا المأزق، يُمكن للمرء اتباع نهجِ دارسي العلوم الاجتماعية والإنسانية الذين استكشفوا طرقًا «غير حديثة» للتفكير خلال الخمسين عامًا الماضية. أوضح كتَّابٌ أمثال برونو لاتور وتيم إنجولد أنه يمكننا العثور على طرق أقل ميلًا للمُقارنة بين ثنائيات وأكثر ميلًا للُّجوء إلى اللاحداثة عند التعامُل مع العالَم من أجل تجاوز الخلاف ما بين التنوير والرومانسية. يُمكننا عندئذٍ أن نُحاول اجتياز الفجوة الحديثة بين البشر وغير البشر ليس من خلال العِلم الحديث أو من خلال تجاوز الإنسانية، التي ترى من وجهة نظرها أن البشر والآلات ليسا في صراعٍ أساسي، ولكن من خلال الفكر ما بعد الإنساني من وجهة النظر (ما بعد) الإنسانية. وهذا يؤدي إلى التوتر الثالث: بين الإنسانية وما بعد الإنسانية. يُشكِّك مؤيدو ما بعدَ الإنسانية، الذين يُعارضون الإنسانيين المُتهمين بالعُنف

مع غير البشر، مثل الحيوانات، تحت مُسمَّى القيمة الفائقة للإنسان، يُشكِّكون في مركزية الإنسان في الأنظمة الأنطولوجية والأخلاقية الحديثة. فهم يرَون أن غير البشر مُهمُّون أيضًا، وأننا يجب ألا نخاف من عبور الحدود بين البشر وغير البشر. وهذا اتجاه مُثير للاستكشاف لأنه يأخُذنا خارج سردية المنافسة بين البشر والآلات.

يُقدم مناصرو ما بعد الإنسانية، من أمثال دونا هاراواي، رؤيةً تصوِّر أن العيش مع الآلات، بل ربما الاندماج معها، لم يعُد يُرى كتهديدٍ أو ككابوس، كما كان يرى من قبل دعاة الإنسانية، أو كحلمٍ يتحقَّق لمناصري تجاوز الإنسانية، ولكنه وسيلة يُمكن من خلالها عبور الحدود الأنطولوجية والسياسية بين البشر وغير البشر. ومن ثَم يمكن أن يكون الذكاء الاصطناعي جزءًا ليس من مشروع دُعاة تجاوز الإنسانية، ولكن من مشروع دُعاة ما بعد الإنسانية المُهم، الذي يدخل من جانب العلوم الإنسانية والفنون بدلًا من العلم. يتم عبور الحدود ليس باسم العلم والتقدُّم العالمي، كما قد يرغب بعض مناصري تجاوز الإنسانية التنويريين في القول، ولكن باسم سياسة مناصري ما بعد الإنسانية وأيديولوجية عبور الحدود. ويمكن لما بعد الإنسانية أيضًا أن تُقدِّم شيئًا آخَر يتعلَّق بالذكاء الاصطناعي: يمكنها أن تحُثَّنا على الاعتراف بأنه «ليس ثمة حاجة لأن يكون غير البشر مُماثِلين لنا ويجب عدم جعلهم مُماثِلين لنا». يبدو أن الذكاء الاصطناعي يمكنه، بالاستناد إلى آراء ما بعد الإنسانية، أن يُحرِّر نفسه من عبء تقليد الإنسان أو إعادة بنائه ويمكنه استكشاف أشكالٍ مختلفة من الوجود والذكاء والإبداع، وما إلى ذلك. ليس هناك حاجة لأن يُصنَّع الذكاء الاصطناعي على صورتنا. فالتقدُّم هنا يعني تجاوز الإنسان وقبول غير البشر لكي نتعلَّم منهم. وعلاوةً على ذلك، يمكن أن يتفق كلٌّ من دعاة تجاوز الإنسانية وما بعد الإنسانية على أنه بدلًا من التنافُس مع الذكاء الاصطناعي لأداء مهمة معيَّنة، يُمكننا أيضًا تحديد هدفٍ مشترك، يتم التوصُّل إليه من خلال التعاون وحشد أفضل ما يمكن أن يقدِّمه البشر والذكاء الاصطناعي من أجل التوجُّه نحو تحقيق ذلك الهدف المشترك.

وسيلة أخرى لتجاوز سردية المنافسة — وهي وسيلة تقترب في بعض الأحيان من مفاهيم ما بعد الإنسانية — هي نهج في فلسفة التكنولوجيا يُسمَّى ما بعد الظاهرية. يستنِد دريفوس إلى علم الظواهر أو الظاهرية، ولا سيما أعمال هايدجر. ولكن الأفكار ما بعد الظاهرية، التي بدأها الفيلسوف دون إيده، تتجاوز فلسفة التكنولوجيا الظاهرية التي ابتكرها هايدجر بالتركيز على كيفية تفاعل البشر مع تقنيات بِعَينها ولا سيما

المصنوعات المادية. يُركِّز هذا النهج، الذي يتعاون في كثيرٍ من الأحيان مع دراسات العلوم والتكنولوجيا، على البُعد المادي للذكاء الاصطناعي. قد يُنظَر إلى الذكاء الاصطناعي في بعض الأحيان على أنه ذو طابعٍ مُجرَّد أو شكلي، غير مُتصل بمصنوعاتٍ مادية وبِنيات أساسية مُحدَّدة. ولكن جميع الشكليات والتجريدات والعمليات الرمزية المذكورة سابقًا تعتمِد على أدواتٍ مادية وبِنيات أساسية مادية. على سبيل المثال، كما سنرى في الفصل التالي، يعتمِد الذكاء الاصطناعي الحالي بشكلٍ كبير على الشبكات وإنتاج كميَّات ضخمة من البيانات باستخدام الأجهزة الإلكترونية. تلك الشبكات والأجهزة ليست مجرد أشياء «افتراضية» ولكن يتعيَّن إنتاجها وصيانتها بشكلٍ مادي. وعلاوةً على ذلك، يتحدَّث ما بعد الظاهِريِّين، مثل بيتر بول فيربيك، عكس التقسيم الحديث بين الموضوع والمحمول، عن التشكيل المُتبادل بين البشر والتكنولوجيا، أو على الأحرى التشكيل المُتبادل بين الموضوع والمحمول. وبدلًا من رؤية التكنولوجيا كتهديد، يؤكِّدون أن البشر مَيَّالون إلى التكنولوجيا (بمعنى أنهم كانوا دائمًا يستخدمون التكنولوجيا؛ أي إنها جزء من وجودنا وليست شيئًا خارجيًّا يُهدِّد هذا الوجود)، وأن التكنولوجيا تُساعد البشر على التعامُل مع العالم. بالنسبة إلى الذكاء الاصطناعي، يبدو أن هذه الرؤية تَعني أن المعركة الإنسانية للدفاع عن الإنسان ضد التكنولوجيا هي معركة مُضلِلَة. وبدلًا من ذلك، وفقًا لهذا النهج، كان الإنسان دائمًا ميَّالًا إلى التكنولوجيا، ولهذا علينا أن نسأل كيف يُساعد الذكاء الاصطناعي البشر في التعامُل مع العالم ونحاول تشكيل هذه المُساعدات بشكلٍ تفاعُلي بينما لا يزال بإمكاننا: إننا نستطيع مناقشة الأخلاقيات في مرحلة تطوير الذكاء الاصطناعي، بل يتعيَّن علينا ذلك، بدلًا من أن نشكو فيما بعدُ من المشكلات التي يُسبِّبها.

> يبدو أن الذكاء الاصطناعي يُمكنه، بالاستناد إلى آراء ما بعد الإنسانية، أن يُحرِّر نفسه من عبء تقليد الإنسان أو إعادة بنائه ويُمكنه استكشاف أشكالٍ مختلفة من الوجود والذكاء والإبداع، وما إلى ذلك.

ومع ذلك، ربما يشعُر المرء بالقلق من أنَّ رُؤى مُناصري ما بعد الإنسانية وما بعد الظاهرية ليست ناقدةً بما فيه الكفاية؛ لأنها شديدة التفاؤل وشديدة البُعد عن الممارسة العلمية والهيكلية، وبالتالي فهي ليست حسَّاسة بما فيه الكفاية تجاه الأخطار الحقيقية والعواقب الأخلاقية والمُجتمعية للذكاء الاصطناعي. إن عبور الحدود التي لم يسبق عبورها

لا يكون بالضرورة من دون مشكلات، وفي الممارسة العملية قد لا تفيد أفكار ما بعد الإنسانية وما بعد الظاهرية في حمايتنا من التسلُّط والاستغلال الذي قد نُعاني منه جرَّاء استخدام تقنيات كالذكاء الاصطناعي. يُمكن للمرء أيضًا أن يُدافع عن رؤية أكثر تقليدية للإنسان أو يُطالب بنوعٍ جديد من الإنسانية، بدلًا من أن يدعم ما بعدَ الإنسانية. وهكذا يستمر النقاش.

الفصل الرابع

أهي حقًّا مجرد آلات؟

التشكيك في المكانة الأخلاقية للذكاء الاصطناعي: الوكالة الأخلاقية واكتساب المكانة الأخلاقية

إحدى القضايا التي أُثيرَت في الفصل السابق تتعلَّق بما إذا كان غير البشر مُهمِّين أيضًا. يعتقد الكثيرون اليوم أن الحيوانات مُهمَّة من الناحية الأخلاقية. ولكن لم يكُن الأمر كذلك دائمًا. على ما يبدو، كنَّا مُخطئين في الماضي بشأن الحيوانات. فإذا كان الكثيرون اليوم يعتقدون أن الآلات المدعومة بالذكاء الاصطناعي مجرد آلات، فهل يرتكِبون خطأً مُماثلًا؟ هل تستحِقُّ الآلات المدعومة بالذكاء الاصطناعي الفائقة الذكاء مكانةً أخلاقية؟ هل ينبغي أن نُعطيها حقوقًا؟ أم إنه من الخطورة بمكانٍ أن نُفكِّر حتى في مسألة ما إذا كانت الآلات يُمكن أن تحظى بمكانةٍ أخلاقية؟

إحدى الطرق لمناقشة ما هو الذكاء الاصطناعي وما يمكن أن يُصبح عليه هي السؤال عن المكانة الأخلاقية للذكاء الاصطناعي. ونحن هنا نتطرَّق إلى أسئلةٍ فلسفية مُتعلِّقة بالذكاء الاصطناعي، ليس عبر الميتافيزيقا أو الإبستمولوجيا أو تاريخ الأفكار، ولكن عبر فلسفة الأخلاق. يمكن أن يُشير مصطلح «المكانة الأخلاقية» (ويُسمى أحيانًا «الأهمية الأخلاقية») إلى نوعَين من الأسئلة. الأول يتعلَّق بما يُمكِن للذكاء الاصطناعي القيام به من الناحية الأخلاقية؛ بعبارةٍ أخرى، ما إذا كان يمكن أن يتمتَّع بما يُطلق عليه الفلاسفة «الوكالة الأخلاقية»، وإذا كان الأمر كذلك، فهل يتمتَّع بالوكالة الأخلاقية الكامِلة؟ ماذا يعني هذا؟ يبدو أن أفعال الذكاء الاصطناعي اليوم لها بالفعل عواقِب أخلاقية. سيتَّفق معظم الناس على أن لدى الذكاء الاصطناعي شكلًا «ضعيفًا» من أشكال الوكالة الأخلاقية بهذا المعنى، والذي يُشبِه، على سبيل المثال، مُعظم السيارات اليوم؛ إذ

يمكن أن يكون للأخيرة أيضًا عواقِب أخلاقية. ولكن إذا سلَّمنا بأن الذكاء الاصطناعي يزداد ذكاءً واستقلالًا، فهل يمكن أن يتمتَّع بشكلٍ أقوى من أشكال الوكالة الأخلاقية؟ هل يجِب أن يتم منحُه أو سيتطوَّر لدَيه بعض القدرة على التفكير الأخلاقي والقدرة على إصدار الأحكام واتخاذ القرارات؟ على سبيل المثال: هل يُمكن وهل يجب أن نعتبر السيارات الذاتية القيادة التي تستخدِم الذكاء الاصطناعي ذات وكالة أخلاقية؟ هذه الأسئلة تتعلَّق بأخلاقيات الذكاء الاصطناعي، بمعنى أنها تتطرَّق إلى ماهية القدرات الأخلاقية التي يمكن أو ينبغي أن يتمتَّع بها الذكاء الاصطناعي؟ ولكن الأسئلة المُتعلقة بـ «المكانة الأخلاقية» يمكن أيضًا أن تُشير إلى كيف ينبغي أن نُعامل الذكاء الاصطناعي. هل الذكاء الاصطناعي «مجرد آلة»، أم أنه يستحق شكلًا من أشكال الاحترام الأخلاقي؟ هل يجب علينا مُعاملته بطريقةٍ مختلفة عن الطريقة التي نتعامل بها مثلًا مع آلة التحميص أو المغسلة؟ هل يجب أن نمنح حقوقًا لكيانٍ صناعي ذكي للغاية، إذا تم تطوير مثل هذا الكيان يومًا ما، حتى لو لم يكُن بشريًّا؟ هذا ما يُطلِق عليه الفلاسفة السؤال المتعلق بـ «اكتساب المكانة الأخلاقية». هذا السؤال يتعلق بأخلاقيات الذكاء الاصطناعي بذاته، ولكنه يتعلَّق بأخلاقياتنا تجاهه. هنا يكون الذكاء الاصطناعي موضعَ اهتمامٍ من الناحية الأخلاقية، بدلًا من كونه وكيلًا أخلاقيًّا مُحتملًا في حدِّ ذاته.

> هل الذكاء الاصطناعي «مجرد آلة»؟ هل يجب علينا معاملته بطريقةٍ مختلفة عن الطريقة التي نتعامل بها مثلًا مع آلة التحميص أو المغسلة؟

الوكالة الأخلاقية

لنبدأ بالتحدُّث عن سؤال الوكالة الأخلاقية. إذا كان الذكاء الاصطناعي يُمكن أن يُصبح أكثر ذكاءً مما هو عليه اليوم، فيُمكننا أن نفترض أنه يستطيع أن يُطوِّر قدرته على التفكير الأخلاقي وأنه يستطيع أن يتعلَّم كيف يتَّخِذ البشر القرارات بشأن القضايا الأخلاقية. ولكن هل سيكون هذا كافيًا لكي يحظى بالوكالة الأخلاقية الكاملة؛ أي الوكالة الأخلاقية التي يتمتَّع بها الإنسان؟ هذا السؤال ليس خيالًا علميًّا بالكامل. فإذا كنا نعتمِد اليوم على الخوارزميات في اتخاذ بعض قراراتنا، على سبيل المثال في السيارات أو المحاكم، فيبدو أنه سيكون من المُهمِّ أن تكون تلك القرارات سليمةً من الناحية الأخلاقية. ولكن ليس

من الواضح ما إذا كانت الآلات يمكن أن تتمتَّع بنفس القدرات الأخلاقية التي يتمتَّع بها البشر. إنها تتمتَّع بالوكالة الأخلاقية بمعنى أنها تقوم بأفعالٍ في العالَم، وهذه الأفعال لها عواقب أخلاقية. على سبيل المثال، قد تتسبَّب سيارة ذاتية القيادة في حادث، أو قد يوصي الذكاء الاصطناعي بسجن شخص معيَّن. هذه السلوكيات والخيارات ليست حيادية من الناحية الأخلاقية؛ إذ إن لها عواقب أخلاقية واضحة على الأشخاص ذوي الصلة. ولكن للتعامُل مع هذه المشكلة، هل يجب منح الوكالة الأخلاقية للذكاء الاصطناعي؟ وهل يمكن أن يتمتَّع بوكالة أخلاقية كاملة؟

هناك مواقف فلسفية مُتنوِّعة حيال هذه الأسئلة. يقول بعض الأشخاص إن الآلات لا يمكن أن تتمتَّع أبدًا بالوكالة الأخلاقية. ويرى هؤلاء أن الآلات ليس لدَيها القدرات اللازمة للوكالة الأخلاقية، مثل الحالات العقلية أو الانفعالات أو الإرادة الحرة. ولذلك هناك خطورة في أن نفترِض أنها تستطيع اتخاذ قراراتٍ سليمة أخلاقيًّا وأن نعتمِد عليها في اتخاذ مثل هذه القرارات اعتمادًا كاملًا. على سبيل المثال، قالت ديبورا جونسون (٢٠٠٦) إن أنظمة الكمبيوتر لا تتمتَّع بوكالةٍ أخلاقية خاصة بها: إنها من إنتاج البشر وتُستخدَم من قِبَلهم، والبشر وحدَهم لديهم الحرية والقدرة على التصرُّف واتخاذ القرارات من الناحية الأخلاقية. وبالطريقة نفسها، يمكن للمرء أن يقول إن الذكاء الاصطناعي من إنتاج البشر، وبالتالي يجب أن يكون اتخاذ القرارات الأخلاقية في الممارسات التكنولوجية من اختصاص البشر. على النقيض من ذلك، هناك أولئك الذين يعتقدون أن الآلات يمكن أن تتمتَّع بوكالةٍ أخلاقية كاملة تمامًا مثل البشر. ويزعم الباحثون مثل مايكل وسوزان أندرسون، على سبيل المثال، أنه من حيث المبدأ يمكن، بل يجب، أن تُمنح الآلات نوعًا من الأخلاق البشرية (Anderson and Anderson 2011). ويُمكننا تزويد الذكاء الاصطناعي بالمبادئ، وربما تكون الآلات حتى أفضل من البشر في الوصول إلى القرارات الأخلاقية نظرًا لأنها أكثر عقلانية ولا تنجرف وراء عواطفها. وقد جادل البعض، لدحض هذه الفكرة، بأن القواعد الأخلاقية كثيرًا ما تتضارب (على سبيل المثال، انظر إلى قصص الروبوتات لأسيموف، حيث تتسبَّب القوانين الأخلاقية للروبوتات دائمًا في مشكلات للبشر والروبوتات)، وأن مشروع إنشاء «آلات أخلاقية» من خلال تغذِيَتها بالقواعد يستند إلى افتراضاتٍ خاطئة بخصوص طبيعة الأخلاق. فالأخلاق لا يمكن اختزالها في اتِّباع القواعد، كما أنها ليست مسألة عواطف بشرية فحسب؛ ولكن هذه العواطف قد تكون ضرورية للغاية للحُكم الأخلاقي. فإذا كان الذكاء الاصطناعي العام مُمكنًا على الإطلاق، فإننا لا نُريد نوعًا من

«الذكاء الاصطناعي المريض نفسيًّا» أي الذي يتمتَّع بالعقلانية الكاملة ولكنَّه لا يهتمُّ باهتمامات الإنسان لأنه يفتقر إلى المشاعر (Coeckelbergh 2010).

لهذه الأسباب، يمكن أن نرفض فكرة تمتُّع الذكاء الاصطناعي بوكالة أخلاقية كاملة رفضًا تامًّا، أو يمكن أن نتَّخِذ موقفًا وسطًا: يجب أن نمنح الذكاء الاصطناعي نوعًا من القواعد الأخلاقية، ولكن ليس كل القواعد الأخلاقية. يستخدم ويندِيل فالاخ وكولين ألين مُصطلح «القواعد الأخلاقية الوظيفية» (٢٠٠٩، ٣٩). تحتاج أنظمة الذكاء الاصطناعي إلى بعض القدرة على تقييم العواقب الأخلاقية لأفعالِها. والمنطق وراء هذا القرار واضح في حالة السيارات ذاتية القيادة: ستتورَّط السيارة على الأرجح في مواقف تتطلَّب اتخاذ خيارٍ أخلاقي ولكن لا يوجَد وقت للاستعانة بالبشر لاتخاذ القرار أو انتظار التدخُّل البشري. وفي بعض الأحيان، تكون هذه الخيارات عبارة عن معضلة. يتحدَّث الفلاسفة عن معضلة عربة الترام، وهي تجربة فكرية تتعلق بسَير عربة ترام على مسار سكك حديدية ويجب عليك الاختيار بين عدم فِعل أي شيء، الأمر الذي سيؤدِّي إلى مَوت خمسة أشخاص مُقيَّدين بالمسار، أو سحب الرافعة وإرسال العربة إلى مسارٍ آخَر، حيث يكون هناك شخصٌ واحد مقيَّد به ولكنه شخص تعرِفه. ما هو الشيء السليم أخلاقيًّا الذي يتوجَّب عليك القيام به؟ بالمثل، يقول أنصار هذا النهج إن السيارة الذاتية القيادة قد تُضطَرُّ إلى اتخاذ خيار أخلاقي، على سبيل المثال، بين قتْل المشاة العابرين على الطريق والاصطدام بحائط، مما يؤدي إلى موت السائق. ما الخيار الذي يجِب أن تتَّخِذه السيارة؟ يبدو أنه سيتعين علينا اتخاذ هذه القرارات الأخلاقية (مُسبقًا) والتأكُّد من تغذية السيارات بها من قِبَل المُطوِّرين. أو ربما نحتاج إلى بناء سيارات مزوَّدة بالذكاء الاصطناعي تتعلَّم من اختيارات البشر. ومع ذلك، قد يُثار سؤال عما إذا كان إعطاء الذكاء الاصطناعي قواعد هو وسيلة جيدة لتمثيل الأخلاق البشرية، هذا إن كان من المُمكن «تمثيل» الأخلاق من الأساس، وإذا كانت مُعضلة عربة الترام تبين شيئًا جوهريًّا في الحياة والتجربة الأخلاقية. أو، من منظورٍ مختلف تمامًا، يمكن للمرء أن يتساءل عما إذا كان البشر في الواقع قادِرين على اتخاذ قرارات أخلاقية بكفاءة. ولماذا نُقلد أخلاق البشر من الأساس؟ إن مُناصري تجاوز الإنسانية، على سبيل المثال، يرَون أن الذكاء الاصطناعي سوف يتمتع بأخلاقٍ فائقة لأنه سيكون أكثر ذكاءً منَّا.

هذا التشكيك في التركيز على الإنسان يُوجِّهنا إلى موقفٍ آخَر، لا يتطلَّب وكالةً أخلاقية كاملة ويُحاول ترك الموقف الأخلاقي المُتمحوِر حول الإنسان. وقد دافع لوتشيانو فلوريدي

وجيه دبليو ساندرز (٢٠٠٤) عن أخلاقٍ لا عقل لها وغير مُستندة إلى خصائص يمتلكها البشر. ويُمكننا جعل الوكالة الأخلاقية تعتمِد على التمتُّع بمستوًى كافٍ من التفاعل والاستقلال والقدرة على التكيُّف وكذلك القدرة على القيام بتصرُّفات ذات طابع أخلاقي. ووفقًا لهذه المعايير، فإن كلب البحث والإنقاذ يتمتَّع بالوكالة الأخلاقية، ولكن كذلك روبوت الذكاء الاصطناعي الذي يتولَّى تصفية الرسائل البريدية غير المرغوب فيها. وبالمِثل، يمكن تطبيق معايير لا تتمحوَر حول الإنسان لمنح الروبوتات الوكالة الأخلاقية، كما اقترح جون سالينز (٢٠٠٦): إذا كان الذكاء الاصطناعي مُستقلًّا عن المُبرمجين ويمكننا تفسير سلوكه بأن نعزو إليه القصد الأخلاقي (مثل قصد فعل الخير أو الشر)، وإذا نمَّ سلوكه عن فَهم مسئوليته تجاه وكلاء أخلاقِيِّين آخرين، فإن هذا الذكاء الاصطناعي يتمتَّع بالوكالة الأخلاقية. ومن ثَم، فإن هذه الآراء لا تتطلَّب الوكالة الأخلاقية الكاملة إذا كان ذلك يعني الوكالة الأخلاقية البشرية، ولكنها تُعرِّف الوكالة الأخلاقية بطريقةٍ تكون من حيث المبدأ مُستقلةً عن الوكالة الأخلاقية الكاملة للبشر والقُدرات البشرية المطلوبة لذلك. ومع ذلك، هل ستكون مثل هذه الوكالة الأخلاقية الاصطناعية كافيةً إذا حُكِمَ عليها وفقًا للمعايير الأخلاقية البشرية؟ عمليًّا، يكمن القلق، على سبيل المثال، في أن السيارات ذاتية القيادة قد لا تُطبق القواعد الأخلاقية الكافية. أما من حيث المبدأ، فيكمن القلق في أننا نبتعِد كثيرًا عن الأخلاق البشرية هنا. ويعتقد الكثيرون أن الوكالة الأخلاقية مُرتبطة ويجب أن تكون مرتبطة بالإنسانية والشخصية. وهؤلاء لا يميلون إلى اعتناق أفكار مؤيِّدي ما بعد الإنسانية أو مؤيدي تجاوز الإنسانية.

اكتساب المكانة الأخلاقية

ثمة موضوع آخَر مُثير للجدَل ويتعلَّق باكتساب الذكاء الاصطناعي لمكانةٍ أخلاقية. تخيَّل أن لدَينا ذكاءً اصطناعيًّا فائقًا. هل من المَقبول أخلاقيًّا إيقاف تشغيلِه، أو «قتله»؟ وإذا ما نظرنا عن كثَب إلى الذكاء الاصطناعي الحالي: هل من المقبول ركل كلبٍ آلي مُزود بالذكاء الاصطناعي؟[1] إذا كانت الآلات المدعومة بالذكاء الاصطناعي ستكون جزءًا من الحياة اليومية، كما يتوقَّع العديد من الباحثين، فإن مثل هذه الحالات ستظهر بالضرورة وتُثير مسألة كيف يجِب على البشر التصرُّف تجاه هذه الكيانات الاصطناعية. ومع ذلك، ليس علَينا أن ننظُر إلى المُستقبل البعيد أو إلى الخيال العلمي. فقد أظهرت الأبحاث أن الناس في الوقت الحالي يتعاطفون مع الروبوتات ويتردَّدون في «قتلها» أو «تعذيبها»

(Suzuki et al. 2015; Darling, Nandy, and Breazeal 2015)، حتى إذا لم تكُن مزوَّدة بالذكاء الاصطناعي. ويبدو أن البشر لا يحتاجون من الكيانات الاصطناعية سوى القليل جدًّا من أجل إضفاء الإنسانية أو الشخصية عليهم والتعاطُف معهم. فإذا أصبحت هذه الكيانات الآن مزودةً بالذكاء الاصطناعي، مما يجعلها أشبَه بالإنسان (أو بالحيوان)، يبدو أن هذا يجعل مسألة إكساب المكانة الأخلاقية أكثر إلحاحًا. على سبيل المثال، ماذا ينبغي أن يكون ردُّ فعلنا تجاه الأشخاص الذين يتعاطفون مع الذكاء الاصطناعي؟ هل هُم مُخطئون؟

ربما يكون قول إن الآلات المدعومة بالذكاء الاصطناعي هي مجرد آلات وإن الأشخاص الذين يتعاطفون معها ببساطةٍ مُخطئون في تقديرهم للأمور وفي عواطفهم وتجربتِهم الأخلاقية هو الأقرب إلى البديهة. إذ يبدو لنا، عند النظرة الأولى، أننا لا ندين بشيءٍ إلى الآلات. فهي أشياء، وليست أشخاصًا. ويُفكر الكثير من الباحِثين في مجال الذكاء الاصطناعي بهذا المنطق. على سبيل المثال، ترى جوانا برايسون أن الروبوتات هي أدوات ومُمتلكات وأنه ليس لدَينا أي التزاماتٍ تجاهها (Bryson 2010). قد يتفق الذين يتبنَّون هذا الموقف بشدَّة على أنه إذا كان لدى الآلات المدعومة بالذكاء الاصطناعي القُدرة على الوعي، ولدَيها حالات عقلية، وما إلى ذلك، فإننا مُطالبون بأن نمنحها مكانةً أخلاقية. ولكنهم سيقولون إن هذا الشرط لا يتحقَّق اليوم. وكما رأينا في الفصول السابقة، قد يقول البعض إنه لن يتحقَّق أبدًا؛ ويقول آخَرون إنه يمكن تحقيقُه من حيث المبدأ، ولكن هذا لن يحدث في المستقبل القريب. ولكن النتيجة المُترتِّبة على السؤال المُتعلق بالمكانة الأخلاقية هي أنه في الوقت الحالي وفي المستقبل القريب، يُفترَض أن نتعامَل مع الآلات المدعومة بالذكاء الاصطناعي كأشياء، إلا إذا ثبت خلاف ذلك.

على الرغم من ذلك، فثمَّة مشكلة واحدة تواجِهنا عند اتخاذ هذا الموقف، وهي أنه لا يفسر ولا يُبرر إحساسنا البديهي الأخلاقي ولا تجاربنا الأخلاقية التي تُخبرنا بأن ثمَّة شيئًا خاطئًا في «إساءة معاملة» الذكاء الاصطناعي، حتى إذا لم تكُن لدَيه خصائص شبيهة بالبشر أو الحيوانات مثل الوعي أو الإحساس. للعثور على مثل هذه التبريرات، يُمكن للمرء اللجوء إلى كانط، الذي اعتبر أنه من الخطأ إطلاق النار على كلب؛ ليس لأن إطلاق النار على كلبٍ ينتهك أي التزاماتٍ تجاه هذا الكلب، ولكن لأن مثل هذا الشخص «يضرُّ بصفات الرحمة والإنسانية في نفسه، والتي يجب أن يُمارسها بناءً على واجباته تجاه البشر» (Kant 1997). أما اليوم فنحن نميل إلى التفكير بطريقةٍ مختلفة تجاه

الكلاب (على الرغم من أن هذا ليس حال الجميع وليس الحال في كل مكان). ولكن يبدو أنه يُمكن تطبيق الحجَّة نفسها على الآلات المدعومة بالذكاء الاصطناعي: يُمكننا أن نقول إننا لا ندين بشيءٍ إلى الآلات المدعومة بالذكاء الاصطناعي، ولكنَّنا مع ذلك ينبغي لنا عدم ركْل أو «تعذيب» آلة مزوَّدة بالذكاء الاصطناعي؛ لأن ذلك يجعلنا غير رحماء تجاه البشر. يُمكن أيضًا استخدام حجَّةِ أخلاقيات الفضيلة، وهي حُجَّة غير مباشرة أيضًا لأنها تتعلَّق بالبشَر وليس بالذكاء الاصطناعي: «إساءة معاملة» الذكاء الاصطناعي خطأ ليس لأن ثمَّة ضررًا سيلحق بالذكاء الاصطناعي، ولكن لأن طابعنا الأخلاقي سيتأذَّى إذا ما فعلنا ذلك. وهذا لا يجعلنا أشخاصًا أفضل. وعلى النقيض من هذا النهج يُمكننا أن نقول إنه في المُستقبل قد تتمتَّع بعض الآلات المزوَّدة بالذكاء الاصطناعي بقيمةٍ جوهرية وتستحقُّ اهتمامنا الأخلاقي، بشرط أن تكون لدَيها خصائص مثل الإحساس. ولا يبدو أن النهج غير المباشر للواجب أو الفضيلة يأخذ هذا الجانب «الآخَر» من العلاقة الأخلاقية على محمل الجد. فهو يُعنى فقط بالبشر. فماذا عن الآلات المزوَّدة بالذكاء الاصطناعي؟ ولكن هل يُمكن للآلات المزوَّدة بالذكاء الاصطناعي أو الروبوتات أن تكون هي «الآخر» كما سأل ديفيد جنكل (٢٠١٨)؟ مرة أخرى، يبدو أن المنطق يقول: لا، الآلات المزوَّدة بالذكاء الاصطناعي ليست لدَيها الخصائص المطلوبة.

> «إساءة معاملة» الذكاء الاصطناعي خطأ؛ ليس لأن ثمة ضررًا سيلحق بالذكاء الاصطناعي، ولكن لأن طابعنا الأخلاقي سيتأذَّى إذا ما فعلنا ذلك.

ثمة نهج مختلف تمامًا يرى أن طريقة تعامُلنا مع مسألة المكانة الأخلاقية هي نفسها تنطوي على إشكالية. يعتمد التفكير الأخلاقي الشائع بشأن المكانة الأخلاقية على ما تمتلِكه الكيانات من خصائص ذات صلةٍ بالأخلاق؛ على سبيل المثال، الوعي أو الإحساس. ولكن كيف نعلَم ما إذا كان لدى الذكاء الاصطناعي فعلًا خصائص مُعينة ذات صلةٍ بالأخلاق أم لا؟ وهل نحن متأكِّدون من ذلك في حالة البشر؟ يقول المُتشكِّكون إننا لسنا متأكِّدين. ومع ذلك، حتى دون هذا اليقين المَعرفي، فإننا لا نزال نُضفي على الإنسان مكانةً أخلاقية على أساس المظهر. ومن المُرجَّح أن يحدُث الشيء نفسه إذا قُدِّر للآلات المزوَّدة بالذكاء الاصطناعي أن تتمتَّع بمظهرٍ وسلوك شبيهَين بالبشر في المستقبل. يبدو أنه بغضِّ

النظر عمَّا يعتبره الفلاسفة من الصواب أخلاقيًّا، سيُضفي البشر، بأية حال، على هذه الآلات مكانة أخلاقية، ويمنحونها حقوقًا، على سبيل المثال. علاوة على ذلك، إذا نظرنا عن كثَب إلى الطريقةِ الَّتي يُضفي بها البشر المكانةَ الأخلاقية «في الواقع»، فإنه يتَّضح على سبيل المثال أن كلًّا من العلاقات الاجتماعية القائمة واللُّغة تلعب دورًا. على سبيل المثال، إذا عامَلْنا قطَّتنا بلُطف، فهذا ليس لأننا ننخرِط في تفكيرٍ أخلاقي بشأن قِطَّتنا، ولكن لأن لدَينا بالفعل نوعًا من العلاقة الاجتماعية معها. إنها بالفعل حيوانٌ أليف ومُرافق لنا قبل أن نقوم بالعمل الفلسفي الذي نُكسِبها بموجبِه مكانة أخلاقية؛ هذا إذا شعرنا من الأساس بحاجةٍ إلى مثل هذه الممارسة. وإذا أطلقنا اسمًا خاصًّا على كلبنا، فإننا — على عكس الحيوانات التي لا تحمِل اسمًا التي نأكُلها — قد منحناه بالفعل مكانةً أخلاقية خاصَّة، بصرْف النظر عن خصائصه الموضوعية. باستخدام مثل هذا النهج العلاقاتي والنقدي وغير المُتزمِّت (Coeckelbergh 2012)، يُمكننا القول إن البشر سوف يمنحون الآلات المزوَّدة بالذكاء الاصطناعي مكانةً أخلاقية بناءً على كيفية تضمينها في حياتنا الاجتماعية وفي لُغتنا وفي ثقافتنا البشرية.

علاوةً على ذلك، نظرًا إلى أن مثل هذه الظروف مُتغيرة تاريخيًّا — فكر مرة أخرى في كيفية مُعاملتنا وتفكيرنا بشأن الحيوانات — ربما تكون هناك حاجة إلى اتخاذ سبُل الحيطة الأخلاقية قبل «تحديد» المكانة الأخلاقية للذكاء الاصطناعي بشكلٍ عام أو لآلة مُعيَّنة مزوَّدة بالذكاء الاصطناعي. ولماذا حتى نتحدَّث عن الذكاء الاصطناعي بشكلٍ عام أو بشكلٍ مجرد؟ يبدو أن هناك شيئًا خاطئًا في الإجراء الأخلاقي لمنح المكانة الأخلاقية: فمن أجل الحُكم على كيانٍ ما، نُخرج هذا الكيان من سياق علاقاته، وقبل أن نحصل على نتيجة إجرائنا الأخلاقي، نتعامَل معه بطريقةٍ رتبوية، سلطوية، مُهيمنة، ككيانٍ نتخذ نحن البشر المُتفوِّقين قرارًا بشأنه. ويبدو أننا قبل حتى أن نفكر في مكانته الأخلاقية، قد وضعناه بالفعل في منزلةٍ مُعيَّنة وربما أيضًا مارَسْنا عليه العنف بمعاملته ككائنٍ نتَّخذ قرارات بشأنه، ونصَّبنا أنفسنا آلهةً محورية قوية عالمة على الأرض يحقُّ لها منح المكانة الأخلاقية للكائناتِ الأخرى. لقد جعلنا أيضًا جميع السياقات والملابسات الاجتماعية غير مَرئية. كما في حالة مُعضلة عربة الترام، لقد اختزلنا الأخلاق في صورة كاريكاتيرية. باستخدام مثل هذا التفكير، يبدو أن فلاسفة الأخلاق يفعلون ما اتُّهِم الفلاسفةُ المؤيدون لدريفوس الباحثين في مجال الذكاء الاصطناعي الرمزي بفِعله: تشكيل وتجريد ثروة من التجربة الأخلاقية والمعرفة الأخلاقية على حساب التخلِّي عما يجعلنا بشرًا، وليس ذلك

فحسب، بل وعلى حساب التضحية بمسألة المكانة الأخلاقية لغير البشر. وبصرف النظر عن المكانة الأخلاقية الفعلية للآلات المزوَّدة بالذكاء الاصطناعي، كما لو كان هذا يمكن تحديده بشكلٍ مُستقل تمامًا عن ذاتية الإنسان، فمن الأهمية بمكانٍ أن نفحص توجُّهنا الأخلاقي ومشروع التفكير الأخلاقي المجرد نفسه، بأسلوب نقدي.

> «إساءة معاملة» الذكاء الاصطناعي خطأ؛ ليس لأن ثمة ضررًا سيلحق بالذكاء الاصطناعي، ولكن لأن طابعنا الأخلاقي سيتأذَّى إذا ما فعلنا ذلك.

نحو قضايا أخلاقية أكثر عملية

كما تُظهر المناقشات في هذا الفصل والفصل السابق، فإن التفكير في الذكاء الاصطناعي يُعلِّمنا أشياء أخرى إلى جانب ما نتعلَّمه بشأن الذكاء الاصطناعي. إنه يُعلمنا أيضًا أشياء عن أنفسنا: عن طريقة تفكيرنا، وطريقة تصرُّفنا في الواقع، والطريقة التي ينبغي أن نتعامَل بها مع غير البشر. فإذا نظرنا إلى الأُسس الفلسفية لأخلاقياتِ الذكاء الاصطناعي، نرى خلافات عميقةً حول طبيعة ومستقبل الإنسانية والعلم والحداثة. إن التشكيك في الذكاء الاصطناعيِّ يكشف اللثام عن عالَم مُظلم من الأسئلة النقدية حول المعرفة البشرية والمُجتمع البشري وطبيعة الأخلاق البشرية.

هذه المناقشات الفلسفية أقل بُعدًا وأقل «أكاديمية» مما قد يعتقِد البعض. وستظلُّ تُعاود الظهور أمامنا عندما نتناول، لاحقًا في هذا الكتاب، المزيدَ من المسائل الأخلاقية والقانونية والسياسية الأكثر عمليةً التي يُثيرها الذكاء الاصطناعي. وسرعان ما ستُواجهنا من جديدٍ بمجرَّد أن نُحاول التطرُّق إلى موضوعاتٍ مثل المسئولية والسيارات ذاتية القيادة، أو شفافية تعلُّم الآلة، أو الذكاء الاصطناعي المُتحيز، أو أخلاقيات الروبوتات الجنسية. إذا كانت أخلاقيات الذكاء الاصطناعي تُريد أن تكون أكثر مِن مجرَّد قائمة بالقضايا، فيجب أن يكون لدَيها ما تقوله حول مثل هذه المسائل.

بعد كلِّ ما قيل، حان الوقت الآن للتحوُّل إلى قضايا أكثر عملية. هذه القضايا لا تتعلَّق بالمُشكلات الفلسفية التي يطرحُها الذكاء الاصطناعي العام المُفترض، أو بالمخاطر المتَّصلة بالذكاء الفائق في المستقبل البعيد، أو بالوحوش المخيفة الأخرى التي يخلقها الخيال العلمي. إنها تتعلق بحقائق الذكاء الاصطناعي القائمة بالفعل، والتي هي أقلُّ

وضوحًا وربما أقل جاذبية، ولكنها لا تزال شديدة الأهمية. إن الذكاء الاصطناعي في الوقت الحالي لا يأخُذ دور وحش فرانكنشتاين أو الروبوتات المُذهلة المزوَّدة بالذكاء الاصطناعي التي تُهدِّد الحضارة، كما أنه أكثر من مجرد تجربةٍ فكرية فلسفية. الذكاء الاصطناعي يتعلَّق بتقنياتٍ سريةٍ غير مرئية ولكنها مُتغلغلة ومنتشرة وقوية ومتزايدة الذكاء، تلك التقنيات التي تُشكِّل بالفعل حياتنا اليوم. ومن ثَمَّ، فإن أخلاقيات الذكاء الاصطناعي تتعلَّق بالتحديات الأخلاقية التي يُثيرها الذكاء الاصطناعي في الوقت الحالي وفي المُستقبل القريب، كما تتعلق بتأثير هذه التحدِّيات على مجتمعاتنا وديمقراطياتنا الهشَّة. إن أخلاقيات الذكاء الاصطناعي تتعلَّق بحياة الناس وبالسياسة. إنها تتعلق بحاجتنا، كأفرادٍ وكمجتمعات، إلى التعامُل مع القضايا الأخلاقية الآن.

الفصل الخامس

التكنولوجيا

قبل مناقشة القضايا الأخلاقية الواقعية المُتعلقة بالذكاء الاصطناعي بمزيدٍ من التفاصيل، لدَينا مهمةٌ أخرى علينا إنجازها لتمهيد الطريق: بعيدًا عن الضجَّة المُثارة حول الذكاء الاصطناعي، علينا أن نفهم هذه التكنولوجيا وتطبيقاتها. فلنُنَحِّ جانبًا الخيالَ العلميَّ لتجاوز الإنسانية والتطلُّعات الفلسفية للذكاء الاصطناعي العام، ولنُلْقِ نظرةً على ماهية تكنولوجيا الذكاء الاصطناعي وكيفية استخدامها اليوم. وبما أن تعريفات الذكاء الاصطناعي وغيرها من المُصطلَحات هي نفسها غير مُتَّفق عليها، فإنني لن أتعمَّق في نقاشاتٍ فلسفية أو سياقات تاريخية. إن هدفي الرئيسي هنا هو أن أُعطي القارئ فكرةً عن التكنولوجيا المعنية وكيفية استخدامها. وسوف أبدأ بالتحدُّث عن الذكاء الاصطناعي بشكلٍ عام؛ أما الفصل التالي، فسيتناول تقنيات تعلُّم الآلة وعلم البيانات وتطبيقاتهما.

ما هو الذكاء الاصطناعي؟

يمكن تعريف الذكاء الاصطناعي بأنه الذكاء الذي تُظهره أو تُحاكيه الرموز البرمجية (الخوارزميات) أو الآلات. ويُثير هذا التعريف سؤالًا حول كيفية تعريف الذكاء. من الناحية الفلسفية، يُعتبَر الذكاء مفهومًا غامضًا. ويمكن القول بأنه ذكاءٌ شبيه بالذكاء البشري. على سبيل المثال، يُعرِّف فيليب جانسن وآخرون الذكاء الاصطناعي بأنه «علم وهندسة الآلات ذات القدرات التي تُعتبر ذكيةً وفقًا لمعايير الذكاء البشري» (٢٠١٨، ٥). وفقًا لهذا التعريف، يتعلق الذكاء الاصطناعي بإنشاء آلات ذكيَّة تُفكر أو تتفاعل مثل البشر. ومع ذلك، يعتقد العديد من الباحثين في مجال الذكاء الاصطناعي أنه ليس هناك داعٍ لأن يكون الذكاء شبيهًا بالذكاء البشري، ويفضلون تعريفًا أكثر حيادًا صِيغَ

بشكلٍ مُستقل عن الذكاء البشري وأهداف الذكاء الاصطناعي العام أو القوي ذات الصلة. ويسردون جميع أنواع الوظائف المعرفية والمهام مثل التعلُّم والإدراك والتخطيط ومُعالجة اللغة الطبيعية والتفكير واتخاذ القرارات وحلِّ المشكلات؛ وغالبًا ما يُعادل ذلك الذكاءَ نفسه. على سبيل المثال، تزعم مارجريت بودين أن الذكاء الاصطناعي «يسعى إلى جعل أجهزة الكمبيوتر تقوم بالأشياء التي يُمكن للعقول البشرية القيام بها». يبدو الأمر في البداية وكأن البشر هم النموذج الوحيد. إلا أنها، تسرد بعد ذلك كل أنواع المهارات النفسية مثل الإدراك والتنبُّؤ والتخطيط، التي تُشكل جزءًا من «الفضاء الغني بقدرات مُعالجة المعلومات المتنوعة» (٢٠١٦، ١). ويمكن أن تكون مُعالجة المعلومات هذه ليست حكرًا على الإنسان. فالذكاء العام، وفقًا لمارجريت بودين، لا يكون بالضرورة بشريًّا. فهناك بعض الحيوانات التي يُمكننا اعتبارها ذكية. ويحلم مؤيدو تجاوز الإنسانية بعقولٍ مُستقبلية لا تكون مضمنة بيولوجيًّا مثلما هو الحال الآن. ومع ذلك، كان هدف تحقيق قدرات شبيهة بقدرات البشر وربما ذكاء عام شبيه بذكاء البشر جزءًا من الذكاء الاصطناعي منذ البداية.

يرتبط تاريخ الذكاء الاصطناعي ارتباطًا وثيقًا بتاريخ علوم الكمبيوتر والتخصُّصات ذات الصِّلة مثل الرياضيات والفلسفة، ومن ثَمَّ فهو يمتدُّ على الأقل إلى العصور الحديثة الباكرة (مثل جوتفريد فيلهلم لايبنيتس ورينيه ديكارت) إن لم يكن إلى العصور القديمة، التي تنتشِر فيها قصص عن حرفيِّين يصنعون كائناتٍ اصطناعية وآلاتٍ ذكية يُمكنها خداع الناس (تذكَّر الشخصيات المُتحركة في اليونان القديمة أو الشخصيات الآلية الشبيهة بالبشر في الصين القديمة). ولكن على العموم يُعتبر الذكاء الاصطناعي قد بدأ في الخمسينيات من القرن العشرين بوصفه تخصُّصًا مستقلًّا، بعد اختراع الكمبيوتر الرقمي القابل للبرمجة في أربعينيَّات القرن العشرين وولادة تخصُّص علم التحكُّم الآلي (السيبرانية)، الذي عرَّفه نوربرت وينر في عام ١٩٤٨ على أنه الدراسة العلمية «للتحكُّم والتواصُل في الحيوان والآلة» (Wiener 1948). وكان نشر ورقة ألان تورينج البحثية لعام ١٩٥٠ بعنوان «الآلات الحاسبة والذكاء» في مجلة «مايند»، والتي قدمت اختبار تورينج الشهير ولكن كانت تتناول بشكلٍ عام سؤال ما إذا كانت الآلات قادرةً على التفكير، وسبقت بالفعل في التكهُّن بالآلات التي يُمكنها التعلُّم وأداء مهام مجرَّدة، كانت لحظة هامة في تاريخ الذكاء الاصطناعي. ومع ذلك، تُعتبر ورشة العمل التي عُقدت في جامعة دارتموث في صيف عام ١٩٥٦ في هانوفر، نيو هامبشاير، بشكلٍ عام هي محل ميلاد الذكاء الاصطناعي المُعاصر. وقد صاغ مُنظمها جون مكارثي فيها مصطلح الذكاء

الاصطناعي، وشاركت فيها أسماءٌ مهمة مثل مارفن مينسكي، وكلود شانون، وألن نيويل، وهيربرت سايمون. وفي حين كان يُنظر إلى علم التحكُّم الآلي على أنه شديد الانشغال بالآلات التناظرية، اهتمَّت ورشة عمل الذكاء الاصطناعي في دارتموث بالآلات الرقمية. كانت الفكرة هي «محاكاة» الذكاء البشري (وليس إعادة خلقِه: فالعملية مختلفة عما يحدث في البشر). وظنَّ الكثير من المشاركين في ورشة العمل هذه أن إنشاء آلةٍ تتمتَّع بنفس ذكاء البشر أمر وشيك الحدوث: توقعوا أنها لن تستغرق في ظهورها أكثر من جيلٍ واحد.

هذا هو هدف «الذكاء الاصطناعي القوي». الذكاء الاصطناعي «القوي» أو «العام» قادر على أداء أي مهام معرفية يمكن للبشر أداؤها، في حين أن الذكاء الاصطناعي «الضعيف» أو «المحدود» يمكن أن يؤدي فقط في مجالاتٍ مُحددة مثل الشطرنج، وتصنيف الصور، وما إلى ذلك. حتى اليوم، لم نُحقِّق الذكاء الاصطناعي العام، وكما رأينا في الفصول السابقة، فإن الشكوك تحُوم حول ما إذا كنَّا سنُحقِّقه على الإطلاق. وعلى الرغم من أن بعض الباحثين والشركات يُحاولون تطوير الذكاء الاصطناعي العام، ولا سيما هؤلاء الذين يؤمِنون بنظرية حاسوبية العقل، فإنه لن يتم تطويره في المُستقبل القريب. ولذا، تُركز الأسئلة الأخلاقية والسياسية في الفصل التالي على الذكاء الاصطناعي الضعيف أو المحدود، الموجود بالفعل حاليًّا والذي من المُرجَّح أن يُصبح أكثر قوةً وانتشارًا في المستقبل القريب.

يمكن تعريف الذكاء الاصطناعي باعتباره علمًا وكذلك باعتباره تكنولوجيا. يمكن أن يكون الهدف من الذكاء الاصطناعي هو تفسير الذكاء والوظائف المعرفية المذكورة تفسيرًا علميًّا أدق. ويُمكن أن يُساعدنا في فهم البشر وغيرهم من الكائنات التي تمتلك ذكاءً طبيعيًّا فهمًا أفضل. وبهذه الطريقة، يكون الذكاء الاصطناعي علمًا وتخصُّصًا يدرس ظاهرة الذكاء بشكلٍ منهجي (Jansen et al. 2018)، وأحيانًا يدرس العقل أو الدماغ. ومن هذا المنطلق، يرتبط الذكاء الاصطناعي بعلومٍ أخرى مثل العلوم المعرفية وعلم النفس وعلم البيانات (انظر القسم اللاحق)، وأحيانًا أيضًا عِلم الأعصاب، الذي يسعى حثيثًا إلى فَهم الذكاء الطبيعي. ولكن قد يكون الهدف من الذكاء الاصطناعي أيضًا هو تطوير تقنياتٍ لأغراضٍ عملية مُختلفة، أو كما يقول بودن «لإنجاز أشياء مُفيدة»: يمكن أن يأخذ شكلَ أدوات، صمَّمها البشر، وتخلق مظهر الذكاء والسلوك الذكي لأغراضٍ عملية. ويمكن للآلات المدعومة بالذكاء الاصطناعي أن تفعل ذلك عن طريق تحليل البيئة (في صورة بيانات) والتصرُّف بدرجةٍ كبيرة من الاستقلالية. في بعض الأحيان، تلتقي الاهتمامات

العلمية-النظرية والأغراض التكنولوجية، على سبيل المثال في عِلم الأعصاب الحوسبي، الذي يستخدِم أدواتٍ من علوم الكمبيوتر لفهم الجهاز العصبي، أو في مشروعاتٍ مُحددة مثل «مشروع الدماغ البشري»[1] الأوروبي، الذي يشمل العلوم العصبية وأيضًا الروبوتات والذكاء الاصطناعي؛ وتجمع بعض مشروعاته ما بين عِلم الأعصاب وتعلُّم الآلة فيما يُعرَف بعلم أعصاب البيانات الضخمة (مثل فو وآخرين ٢٠١٨).

بشكلٍ أعم، يعتمد الذكاء الاصطناعي على العديد من التخصُّصات ويرتبط بها، بما في ذلك الرياضيات (على سبيل المثال، الإحصاء)، والهندسة، واللغويات، والعلوم المعرفية، وعلوم الكمبيوتر، وعلم النفس، وحتى الفلسفة. وكما رأينا، يهتم الفلاسفة والباحثون في مجال الذكاء الاصطناعي على حدٍّ سواء بفهم العقل وظواهر مثل الذكاء والوعي والإدراك والفعل والإبداع. وقد أثَّر الذكاء الاصطناعي على الفلسفة والعكس صحيح. وقد أقرَّ كيث فرانكيش وويليام رامزي بهذا الارتباط بين الذكاء الاصطناعي والفلسفة، وشدَّدا على تعدُّد تخصُّصات الذكاء الاصطناعي، وجمعا الجانبَين العلمي والتكنولوجي في تعريفهما للذكاء الاصطناعي باعتباره «نهجًا مُتعدِّد التخصُّصات لفهم ونمذجة ومُحاكاة الذكاء والعمليات المعرفية عن طريق الاستناد إلى مبادئ وأجهزة حوسبِيَّة ورياضية ومنطقية وميكانيكية وحتى بيولوجية متنوعة» (٢٠١٤، ١). لذلك، يعتبر الذكاء الاصطناعي نظريًّا وعمليًّا، علمًا وتكنولوجيا. ويركز هذا الكتاب على الذكاء الاصطناعي باعتباره تكنولوجيا، على الجانب الأكثر عملية: ليس فقط لأن التركيز داخل الذكاء الاصطناعي قد تحوَّل في هذا الاتجاه، ولكن، على وجه الخصوص، لأن الذكاء الاصطناعي في هذه الصورة له عواقب أخلاقية واجتماعية؛ على الرغم من أن البحث العلمي أيضًا ليس خاليًا تمامًا من العواقب الأخلاقية.

باعتباره تكنولوجيا، يُمكن للذكاء الاصطناعي أن يأخذ أشكالًا مختلفة وعادةً ما يكون جزءًا من نُظم تكنولوجية أكبر: الخوارزميات، والآلات، والروبوتات، وما إلى ذلك. لذلك، في حين قد يتعلَّق الذكاء الاصطناعي بـ «الآلات»، فإن هذا المصطلح لا يُشير إلى الروبوتات وحدَها، ناهيك عن الروبوتات التي تتَّخِذ شكلًا بشريًّا. يُمكن أن يُضمَّن الذكاء الاصطناعي في العديد من أنواع الأنظمة والأجهزة التكنولوجية الأخرى. ويمكن لأنظمة الذكاء الاصطناعي أن تأخُذ شكلَ برنامجٍ يعمل على الويب (مثل الدردشة الآلية ومُحركات البحث وتحليل الصور)، ولكن يُمكن أن يُضمَّن أيضًا الذكاء الاصطناعي في الأجهزة الملموسة مثل الروبوتات أو السيارات أو تطبيقات «إنترنت الأشياء».[2] بالنسبة إلى إنترنت

الأشياء، يُستخدَم أحيانًا مصطلح «الأنظمة الإلكترونية-المادية»، وهي عبارة عن أجهزة تعمل في العالَم المادي وتتفاعل معه. وتُعَد الروبوتات نوعًا من الأنظمة الإلكترونية-المادية، التي تؤثِّر تأثيرًا مباشرًا على العالم (Lin, Abney, and Bekey 2011).

إذا تمَّ تضمين الذكاء الاصطناعي في روبوت، فإنه يُطلق عليه أحيانًا الذكاء الاصطناعي «المتجسِّد». وتعتمد الروبوتات في تأثيرها على العالَم المادي تأثيرًا مباشرًا على مكوِّنات مادية. ولكن كل نظام ذكاءٍ اصطناعي، بما في ذلك البرامج النشطة على الويب، «يفعل» شيئًا ولدَيه أيضًا مكوِّنات مادية مثل الكمبيوتر الذي يعمل عليه، والمُكونات المادية للشبكة والبِنية الأساسية التي يعتمِد عليها، وما إلى ذلك. وهذا يجعل التفرقة ما بين تطبيقات الويب «الافتراضية» والتطبيقات «البرمجية» من ناحية، والتطبيقات المادية أو تطبيقات «الأجهزة» من ناحيةٍ أخرى مسألةً صعبة ومُحيِّرة. إن برامج الذكاء الاصطناعي تحتاج إلى مكوِّنات مادية وبِنية أساسية مادية لكي تعمل، والأنظمة الإلكترونية-المادية لا يمكن اعتبارها ذكاءً اصطناعيًّا إلا إذا تم توصيلها بالبرامج المناسبة. علاوةً على ذلك، من وجهة نظر الظاهرية، قد تندمج المكونات المادية والبرمجية أحيانًا في تجربتنا واستخدامنا للأجهزة: فنحن لا نشعر بأن الروبوت التفاعُلي الذي يأخذ شكلًا بشريًّا ويعمل بواسطة الذكاء الاصطناعي، أو أن جهاز المحادثة بالذكاء الاصطناعي مثل أليكسا، عبارة عن مكونات برمجية أو مكونات مادية، ولكننا نشعر أنهما جهاز تكنولوجي واحد (وأحيانًا نشعر أنهما شِبه أشخاص، مثل دُمية «هالو باربي»).

من المُرجَّح أن يكون للذكاء الاصطناعي تأثير كبير على علم الروبوتات، وذلك على سبيل المثال من خلال التقدُّم في معالجة اللغة الطبيعية والتواصُل الشبيه بتواصل الإنسان. في كثيرٍ من الأحيان يُطلَق على هذه الروبوتات اسم «الروبوتات الاجتماعية»؛ لأنها مُصمَّمة بهدف المشاركة في الحياة الاجتماعية اليومية للبشر، على سبيل المثال، كرفاقٍ أو مساعِدين، من خلال التفاعُل مع البشر بطريقةٍ طبيعية. ومن ثَمَّ، يمكن أن يُعزز الذكاء الاصطناعي مزيدًا من التطورات في الروبوتات الاجتماعية.

ومع ذلك، بغض النظر عن المظهر والسلوك الكلي للنظام وتأثيره على البيئة المُحيطة به، وهو ما يُعتبر مهمًّا جدًّا من الناحية الظاهرية والأخلاقية، فإن أساس «الذكاء» في الذكاء الاصطناعي هو برنامج: «خوارزمية» أو مجموعة من الخوارزميات. والخوارزمية هي مجموعة وتسلسُل من التعليمات، مثل الوصفة، تُخبر الكمبيوتر أو الهاتف الذكي أو الآلة أو الروبوت أو أي شيءٍ آخَر يتم تضمينها فيه بما يجب أن يفعل. وهي تؤدي إلى

مُخرجات مُعيَّنة بناءً على المعلومات المتاحة (المدخلات). وتُطبَّق الخوارزمية لحلِّ مشكلةٍ ما. ولكي نفهم أخلاقيات الذكاء الاصطناعي، علينا أولًا أن نفهم كيفية عمل خوارزميات الذكاء الاصطناعي وما تقوم به. وسوف أتحدَّث أكثر عن هذا الموضوع هنا وفي الفصل القادم.

المناهج والمجالات الفرعية المُختلفة

هناك أنواع مختلفة من الذكاء الاصطناعي. يمكن القول أيضًا إن هناك مناهج أو نماذج بحثٍ مختلفة. كما رأينا في انتقاد دريفوس، غالبًا ما كان الذكاء الاصطناعي على مدار التاريخ ذكاءً اصطناعيًّا رمزيًّا. وكان هذا هو النموذج السائد حتى أواخر الثمانينيات. ويعتمِد الذكاء الاصطناعي الرمزي على التمثيلات الرمزية للمهامِّ المعرفية العالية المُستوى مثل التفكير التجريدي واتخاذ القرارات. على سبيل المثال، قد يتَّخِذ قرارًا استنادًا إلى الهيكل الشجري لاتخاذ القرار؛ وهو عبارة عن نموذجٍ للقرارات وعواقِبها المُمكنة، ويُمثَّل غالبًا بشكلٍ رسومي يُشبه المُخطط الانسيابي. وتحتوي الخوارزمية التي تفعل ذلك على عباراتٍ شرطية: قواعد لاتخاذ القرار على صورة if … then …، بحيث يلي if الشرط ويلي then النتيجة. وهذه العملية حاسمة وغير عشوائية. وبالاستناد إلى قاعدة بياناتٍ تُمثِّل المعرفة الخبيرة البشرية، يُمكن لِمثل هذا الذكاء الاصطناعي اتخاذ القرار، مُعتمدًا على كمٍّ هائل من المعلومات، والتصرُّف كنظامٍ خبير. ويستطيع أن يتَّخذ قراراتٍ حكيمة أو يصل إلى توصيات استنادًا إلى كتلةٍ ضخمة من المعرفة، قد يكون من الصعب أو من المُستحيل بالنسبة إلى البشر الاطلاع عليها. على سبيل المثال، تُستخدَم هذه الأنظمة الخبيرة في القطاع الطبِّي لتشخيص المرض ووضع خطةِ العلاج. وقد ظلَّت هذه الأنظمة هي الأنجح في مجال الذكاء الاصطناعي لفترةٍ طويلة.

ولا يزال الذكاء الاصطناعي الرمزي مُفيدًا حتى اليوم، ولكن ظهرت أيضًا أنواع جديدة من الذكاء الاصطناعي، يُمكن دمجها أو عدم دمجها مع الذكاء الاصطناعي الرمزي، وهي قادرة على التعلُّم ذاتيًّا من البيانات، على عكس الأنظمة الخبيرة. ويتم ذلك من خلال استخدام نهجٍ مختلف تمامًا. ويعتمد نموذج البحث «التشابُكي»، الذي تم تطويره في الثمانينيات من القرن العشرين كبديلٍ لما أُطلِق عليه اسم «الذكاء الاصطناعي القديم» ويعرف اختصارًا بـ GOFAI، وتكنولوجيا «الشبكات العصبية» على فكرة أننا بدلًا

من تمثيل الوظائف المعرفية العُليا، يجب علينا بناء شبكات مُترابطة بالاستناد إلى وحداتٍ بسيطة. ويدعي مؤيدو هذا النهج أن هذا يُشبه الطريقة التي يعمل بها الدماغ البشري؛ إذ ينشأ الإدراك من تفاعُلات بين وحداتِ المعالجة البسيطة المُسمَّاة «الخلايا العصبية» (ومع ذلك، فهي لا تُشبه الخلايا العصبية البيولوجية). ويُستخدَم العديد من الخلايا العصبية المُترابِطة. يُستخدَم هذا النهج وهذه التكنولوجيا كثيرًا في «تعلُّم الآلة» (انظر الفصل التالي)، والذي يُطلق عليه بعد ذلك «التعلُّم العميق» إذا كانت الشبكات العصبية تتكوَّن من عدة طبقاتٍ من الخلايا العصبية. وتُعتبَر بعض الأنظمة هجينة؛ على سبيل المثال، يُعتبَر «ألفا جو» الذي طوَّرَته شركة «ديب مايند» نظامًا هجينًا. وقد أدَّى التعلُّم العميق إلى حدوث تطوُّر في مجالات مثل رؤية الآلة ومُعالجة اللغة الطبيعية. ويمكن أن يكون تعلُّم الآلة الذي يَستخدِم شبكة مُحايدة بمنزلة «صندوق أسود»؛ بمعنى أنه في حين أن المُبرمِجين يعرفون تصميم الشبكة، فإنه ليس واضحًا للآخرين ماذا يحدث بالضبط في طبقاتها الوسيطة (بين المدخلات والمخرجات) وبالتالي كيف تتَّخِذ قرارًا. وهذا عكس ما يحدُث في الهيكل الشجري لاتخاذ القرار، الذي يكون واضحًا وقابلًا للتفسير، ومن ثَم يمكن فحصُه وتقييمه من قِبل البشر.

ثمَّة نموذج مُهم آخَر في مجال الذكاء الاصطناعي وهو ذلك الذي يَستخدِم مناهج أكثر تجسيديةً وأكثر اعتمادًا على المواقف، مركزًا على التفاعُل والمهام الحركية بدلًا مما نُطلق عليه المهام المعرفية العُليا. والروبوتات التي صنعها باحثون في مجال الذكاء الاصطناعي مثل رودني بروكس من «إم آي تي» لا تحلُّ المشكلات باستخدام تمثيلاتٍ رمزية ولكن عن طريق التفاعُل مع البيئة المُحيطة. على سبيل المثال، صُمِّمَ الروبوت «كوج» الشبيه بالبشر، الذي تمَّ تطويره في التسعينيات من القرن العشرين، بحيث يتعلَّم من خلال التفاعل مع العالم، كما يفعل الأطفال. وعلاوةً على ذلك، يعتقد بعض الأشخاص أن العقل يمكن أن ينشأ فقط من الحياة؛ وبالتالي، لإنشاء الذكاء الاصطناعي، يجب أن نُحاول إنشاء حياةٍ اصطناعية. ويتبع بعض المهندسين نهجًا أقلَّ مِيتافيزيقية وأكثر عملية؛ إذ يأخذون الأحياء نموذجًا لتطوير تطبيقاتٍ تكنولوجية عملية. وهناك أيضًا آلات تطوُّرية مزودة بالذكاء الاصطناعي تستطيع أن تتطوَّر. ويمكن لبعض البرامج، باستخدام ما يُسمَّى بخوارزميات الوراثة، تغيير نفسها.

هذا التنوُّع في مناهج الذكاء الاصطناعي ووظائفه يشير إلى أن الذكاء الاصطناعي اليوم له العديد من المجالات الفرعية: تعلُّم الآلة، ورؤية الكمبيوتر، ومعالجة اللغة

الطبيعية، والأنظمة الخبيرة، والحوسبة التطوُّرية، وهلمَّ جرًّا. وغالبًا ما يكون التركيز اليوم على تعلُّم الآلة، ولكن هذا ليس سوى مجالٍ واحد من مجالات الذكاء الاصطناعي، حتى وإن كانت هذه المجالات الأخرى مُتصلةً غالبًا بتعلُّم الآلة. وقد تم تحقيق تطورات هائلة مؤخرًا في رؤية الكمبيوتر ومعالجة اللغة الطبيعية وتحليل البيانات الضخمة عن طريق تعلُّم الآلة. على سبيل المثال، يمكن استخدام تعلُّم الآلة لمعالجة اللغة الطبيعية استنادًا إلى تحليل الكلام والمصادر المكتوبة مثل النصوص الموجودة على الإنترنت. وقد أثمر هذا العمل عن إنشاء أجهزة المحادثة الحديثة. مثال آخَر هو التعرُّف على الوجوه استنادًا إلى رؤية الكمبيوتر والتعلُّم العميق، ويمكن استخدامه، على سبيل المثال، في مجال المراقبة.

التطبيقات والتأثير

يمكن تطبيق تكنولوجيا الذكاء الاصطناعي في مجالاتٍ مختلفة (لها تطبيقات متنوعة)، تتراوح ما بين التصنيع والزراعة والنقل، والرعاية الصحية والتمويل والتسويق والجنس والترفيه والتعليم ووسائل التواصُل الاجتماعي. في مجال البيع بالتجزئة والتسويق، تُستخدَم أنظمة التوصية للتأثير في قرارات الشراء ولتقديم إعلاناتٍ مستهدفة. أما في مجال وسائل التواصُل الاجتماعي، يمكن أن يشغَل الذكاء الاصطناعي الروبوتات: وهي عبارة عن حساباتِ مُستخدمين تظهر على أنها أشخاصٌ حقيقيون ولكنها في الواقع برامج. ويُمكن لِمثل هذه الروبوتات أن تنشر محتوًى سياسيًّا أو تُجري دردشةً مع مُستخدِمين من البشر. وفي مجال الرعاية الصحية، يُستخدَم الذكاء الاصطناعي لتحليل بياناتٍ من ملايين المرضى. وما زالت الأنظمة الخبيرة تُستخدَم أيضًا في هذا المجال. في مجال التمويل، يُستخدَم الذكاء الاصطناعي لتحليل مجموعاتٍ ضخمة من البيانات لتحليل السوق وأتْمَتَةِ التعامُلات المالية. وغالبًا ما يتم تضمين نوع من الذكاء الاصطناعي في الروبوتات المُصمَّمة لتكون مرافقًا للإنسان. والطيار الآلي والسيارات ذاتية القيادة تستخدم الذكاء الاصطناعي. ويمكن لأصحاب العمل استخدام الذكاء الاصطناعي لمراقبة الموظفين. كما أن ألعاب الفيديو تحتوي على شخصياتٍ مدعومة بالذكاء الاصطناعي. وتستطيع الآلات المزوَّدة بالذكاء الاصطناعي تأليف المُوسيقى أو كتابة مقالات الأخبار. كما تستطيع تقليدَ أصوات الأشخاص وحتى إنشاء مقاطع فيديو مُزيفة لخطابات.

نظرًا إلى تنوُّع تطبيقات الذكاء الاصطناعي، من المُرجَّح أن يكون له تأثير واسع النطاق، سواء اليوم أو في المُستقبل القريب. فإذا فكَّرْنا مثلًا في الشرطة التنبُّؤية وإمكانية التعرُّف على الكلام، اللذَين يخلقان إمكانيات جديدة للأمان والمراقبة، ووسائل النقل بين الأفراد والسيارات ذاتية القيادة التي يُمكن أن تُحدِث تحوُّلًا في مدنٍ بأكملها، والتداول الخوارزمي العالي التردُّد الذي يُشكِّل بالفعل الأسواق المالية، أو التطبيقات التشخيصية في القطاع الطبي التي تؤثر في اتخاذ القرارات السليمة. يجب أيضًا ألا ننسى العلوم كأحد المجالات الرئيسية التي تأثَّرَت إلى حدٍّ كبير بالذكاء الاصطناعي: عن طريق تحليل مجموعاتٍ ضخمة من البيانات، يمكن للذكاء الاصطناعي مساعدة العلماء في اكتشاف ارتباطات لم يكونوا ليُدركوها لولاه. وهذا ينطبق على العلوم الطبيعية مثل الفيزياء، ولكن أيضًا على العلوم الاجتماعية والعلوم الإنسانية. ومن المُؤكَّد أن يؤثر الذكاء الاصطناعي في مجال العلوم الإنسانية الرقمية الناشئ، على سبيل المثال، عن طريق تعليمنا المزيد عن البشر وعن المُجتمعات البشرية.

يؤثر الذكاء الاصطناعي أيضًا على العلاقات الاجتماعية، كما أن له تأثيرًا اجتماعيًّا واقتصاديًّا وبيئيًّا أوسع (Jansen et al. 2018). ومن المُرجَّح أن يشكل الذكاء الاصطناعي التفاعلات البشرية ويؤثر على الخصوصية. ويُقال إنه قد يزيد من التحيُّز والتمييز. ومن المتوقَّع أن يؤدي إلى فقدان الوظائف وربما إلى إحداث تحوُّلٍ اقتصادي كامل. فمن المُمكن أن يزيد الفجوة بين الأغنياء والفقراء وبين أصحاب النفوذ والمُستضعَفين، معجلًا الظلم والتفاوت الاجتماعي. أما التطبيقات العسكرية، فقد تُغيِّر الطريقة التي يتم بها تنفيذ الحروب، على سبيل المثال، عند استخدام الأسلحة القاتلة ذاتية التشغيل. كذلك يجِب أن نأخُذ في اعتبارنا التأثير البيئي للذكاء الاصطناعي، والذي يشمل زيادة استهلاك الطاقة والتلوُّث. وسوف أُناقش لاحقًا بعض الآثار الأخلاقية والاجتماعية بمزيدٍ من التفصيل، مركزًا على مشكلات الذكاء الاصطناعي ومَخاطره. ولكن يمكن أن يكون للذكاء الاصطناعي أيضًا آثار إيجابية؛ على سبيل المثال، يمكن أن يخلق مُجتمعات جديدة عن طريق وسائل التواصُل الاجتماعي، ويُقلِّل المهام المتكرِّرة والخطيرة عن طريق تكليف الروبوتات بها، ويُحسِّن سلاسل الإمداد، ويُقلِّل استهلاك المياه، وهكذا.

فيما يتعلَّق بالتأثير — إيجابي أو سلبي — يجب ألا نسأل فقط عن طبيعة التأثير ومداه؛ بل أن نسأل أيضًا «مَن» هم المتأثرون وكيف سيتأثَّرون. قد يكون التأثير أكثر إيجابية بالنسبة إلى البعض منه بالنسبة إلى الآخَرين. فهناك العديد من الأطراف المَعنية،

بدءًا من العمال والمرضى والمُستهلكين، إلى الحكومات والمُستثمرين والشركات، وجميعهم قد يتأثرون بطرُق مختلفة. وتنشأ هذه الاختلافات في المكاسب والخسائر من تأثيرات الذكاء الاصطناعي ليس فقط داخل البلدان ولكن أيضًا بين البلدان وأجزاء العالَم. فهل سيعود الذكاء الاصطناعي بالنَّفع على البلدان المتقدِّمة والمُتطورة في المقام الأول؟ وهل من المُمكن أن يكون مفيدًا أيضًا للأشخاص ذوي التعليم المُنخفِض والدخل المُنخفض، على سبيل المثال؟ مَن ستكون لدَيه القدرة على الوصول إلى التكنولوجيا ويكون قادرًا على جنْي فوائدها؟ مَن سيتمكن من تمكين نفسه باستخدام الذكاء الاصطناعي؟ ومَن سيكون مُستبعدًا من هذه الفوائد؟

مَن ستكون لدَيه القدرة على الوصول إلى التكنولوجيا ويكون قادرًا على جني فوائدها؟ مَن سيتمكَّن من تمكين نفسه باستخدام الذكاء الاصطناعي؟ ومَن سيكون مُستبعدًا من هذه الفوائد؟

الذكاء الاصطناعي ليس التكنولوجيا الرقمية الوحيدة التي تُثير مثل هذه الأسئلة. فهناك تقنيات رقمية أخرى خاصة بالمعلومات والاتصالات، وهي أيضًا تؤثر تأثيرًا كبيرًا على حياتنا ومُجتمعاتنا. وكما سنرى، بعض المشكلات الأخلاقية التي يُثيرها الذكاء الاصطناعي ليست حكرًا على الذكاء الاصطناعي وحدَه. على سبيل المثال، هناك مشكلات موازية في تكنولوجيا الأجهزة الذاتية التشغيل. تذكَّر مثلًا الروبوتات الصناعية التي تمَّت برمجتها ولا تُعتبَر ذكاءً اصطناعيًّا، ولكنها لا تزال لها تأثيرات اجتماعية عندما تؤدي إلى البطالة. وبعض مشكلات الذكاء الاصطناعي مُرتبطة بالتقنيات التي يتَّصِل بها الذكاء الاصطناعي، مثل وسائل التواصُل الاجتماعي والإنترنت، التي تُواجهنا بتحدياتٍ جديدة عندما يتم دمجُها مع الذكاء الاصطناعي. على سبيل المثال، عندما تستخدم منصَّات التواصل الاجتماعي مثل «فيسبوك» الذكاء الاصطناعي لتعرِف المزيد عن مُستخدميها، فإن هذا يُثير مخاوف تتعلَّق بالخصوصية.

هذا الاتصال مع التقنيات الأخرى يعني أيضًا أن الذكاء الاصطناعي يكون غير ملحوظٍ في كثير من الأحيان. ويرجع هذا في المقام الأول إلى كونه أصبح بالفعل جزءًا لا يتجزأ من حياتنا اليومية. فالذكاء الاصطناعي كثيرًا ما يُستخدَم في تطبيقات جديدة ومذهلة مثل «ألفا جو». ولكننا يجِب ألا ننسى الذكاء الاصطناعي الذي يشغل بالفعل منصَّات التواصُل الاجتماعي، ومُحركات البحث، وغيرها من الوسائط والتقنيات التي

أضحت جزءًا من تجربتنا اليومية. إن الذكاء الاصطناعي مُتوغِّل في كل شيء. ويمكن أن يكون الفارق بين الذكاء الاصطناعي الفعلي وأشكالٍ أُخرى من التكنولوجيا غامضًا، ممَّا يجعل الذكاء الاصطناعي غير مرئي: إذا تم تضمين أنظمة الذكاء الاصطناعي في التكنولوجيا، فإننا عادةً لا نُلاحظها. وإذا كنا نعرِف بالفعل أنه مُضمَّن، فإنه من الصعب أن نقول ما إذا كان الذكاء الاصطناعي هو الذي يُسبِّب المشكلة أو التأثير، أو إذا كانت التكنولوجيا الأخرى المُتَّصلة به هي المسئولة عن ذلك. بعبارة أخرى، لا يوجد «ذكاء اصطناعي» في حدِّ ذاته: فالذكاء الاصطناعي يعتمد دائمًا على تقنيات أُخرى ويتم تضمينه في مُمارسات وإجراءات علمية وتكنولوجية أوسع. وفي حين أن الذكاء الاصطناعي أيضًا يُثير مشكلاتٍ أخلاقية خاصة به، فإن «أخلاقيات الذكاء الاصطناعي» تحتاج إلى أن تكون مُرتبطة بالأخلاقيات العامة للمعلومات الرقمية وتكنولوجيا الاتصالات، وأخلاقيات الكمبيوتر، وما إلى ذلك.

> يجب ألا ننسى الذكاء الاصطناعي الذي يشغل بالفعل منصَّات التواصل الاجتماعي، ومُحركات البحث، وغيرها من الوسائط والتقنيات التي أضحت جزءًا من تجربتنا اليومية. إن الذكاء الاصطناعي مُتوغِّل في كلِّ شيء.

ثمَّة منطق آخَر يؤكد أنه لا يوجد شيء يُعرف باسم الذكاء الاصطناعي في حدِّ ذاته، وهو أن التكنولوجيا أيضًا دائمًا ما تكون اجتماعيةً وإنسانية: فالذكاء الاصطناعي لا يتعلق فقط بالتكنولوجيا ولكن أيضًا بما يفعله البشر بها، وكيف يستخدمونها، وكيف يُدركونها ويعيشونها، وكيف يُضمِّنونها في بيئاتٍ اجتماعية وتقنية أوسع. وهذا أمر مُهم للأخلاقيات — التي تتعلق أيضًا بقرارات الإنسان — ويعني أيضًا أنه يجب تضمين منظورٍ تاريخي واجتماعي ثقافي. الضجة الإعلامية المثارة حاليًّا حول الذكاء الاصطناعي ليست الضجَّة الأولى التي تُثار حول التقنيات المتقدمة. قبل الذكاء الاصطناعي، كانت «الروبوتات» أو «الآلات» هي الكلمات الرئيسية. كما شهدت تقنيات مُتقدمة أخرى مثل التكنولوجيا النووية، وتكنولوجيا النانو، والإنترنت، والتكنولوجيا الحيوية الكثير من الجدل. ومن المُفيد أن نضع ذلك في اعتبارنا خلال مناقشاتنا حول أخلاقيات الذكاء الاصطناعي؛ إذ ربما يُمكننا أن نستفيد من هذه النقاشات والجدالات. إن استخدام التكنولوجيا وتطويرها

يحدث في سياق اجتماعي. وكما يعلَم الأشخاص المُهتمون بتقييم التكنولوجيا، عندما تكون التكنولوجيا جديدة، يميل الناس إلى أن يُثيروا حولها الكثير من الجدل، ولكن بمجرد أن تُصبح جزءًا من الحياة اليومية، تنخفِض الضجة المُثارة حولها والجدل بشأنها بشكلٍ كبير. ومن المُرجح أن يحدث هذا أيضًا مع الذكاء الاصطناعي. وفي حين أن مثل هذا التوقُّع ليس سببًا وجيهًا لترك مُهمة تقييم الجوانب الأخلاقية والعواقب الاجتماعية للذكاء الاصطناعي، فإنه يُساعدنا في رؤية الذكاء الاصطناعي في سياقه، ومن ثَمَّ يساعدنا في فهمِه على نحوٍ أفضل.

الفصل السادس

لا تنسَ (علم) البيانات

تعلُّم الآلة

بما أن العديد من الأسئلة الأخلاقية حول الذكاء الاصطناعي تتعلق بتقنيات تعتمِد كليًّا أو جزئيًّا على تعلُّم الآلة وعلم البيانات ذي الصِّلة، فإنه يجدُر بنا أن نُلقي الضوء على هذه التقنية والعلم.

يُشير «تعلُّم الآلة» إلى البرامج التي يُمكنها «التعلُّم». والمصطلح مُثير للجدل: فالبعض يقولون إن ما تقوم به ليس تعلُّمًا حقيقيًّا لأنها لا تتمتَّع بإدراكٍ حقيقي؛ والتعلُّم مقصور على البشَر فحسب. على أي حال، يحمل تعلُّم الآلة الحديث «تشابهًا ضئيلًا أو مُنعدمًا مع ما قد يحدُث في عقول البشر» (Boden 2016, 46). وهو يعتمد على الإحصاءات؛ إذ إنه عملية إحصائية. ويُمكن استخدامه لمهامَّ متنوعة، ولكن المهمة الأساسية غالبًا ما تكون هي التعرُّف على الأنماط. ويُمكن للخوارزميات التعرُّف على الأنماط أو القواعد الموجودة في البيانات واستخدام تلك الأنماط أو القواعد لتفسير البيانات وتوقُّع البيانات المُستقبلية.

يحدُث ذلك ذاتيًّا؛ بمعنى أنه يحدُث دون تعليماتٍ وقواعد مباشرة يُعطيها المبرمج. وعلى عكس الأنظمة الخبيرة التي تعتمِد على خبراء بشريين في المجال يشرحون القواعد للمُبرمِجين الذين يتولَّون بعد ذلك برمجة هذه القواعد، تبحث خوارزمية تعلُّم الآلة عن قواعد أو أنماطٍ لم يُحدِّدها المبرمج. كل ما عليك هو تحديد الهدف أو المهمة فقط. وسوف يستطيع البرنامج أن يُكيِّف سلوكَه بما يتوافق مع مُتطلبات المهمة. على سبيل المثال، يمكن لتعلُّم الآلة المساعدة في التمييز بين البريد الإلكتروني العشوائي غير المرغوب فيه والبريد المُهم من خلال فحص عددٍ كبير من الرسائل وتعلُّم ما يُعتبَر عشوائيًّا. مثال آخَر:

لإنشاء خوارزمية تتعرَّف على صور القطط، لا يُقدِّم المبرمجون للكمبيوتر مجموعةً من القواعد تُعرِّف فيها ما هي القطط، ولكنهم يُتيحون للخوارزمية إنشاء نموذجٍ خاصٍّ بها لصور القطط. وتُحسِّن الخوارزمية من أدائها ذاتيًّا لتحقيق أعلى دقَّة تنبؤ بالاستناد إلى مجموعةٍ من صور القطط وغير القطط. وبالتالي، تهدف إلى تعلُّم ما هي صور القطط. ويُقدِّم البشر تقارير، ولكنهم لا يُغذُّونها بتعليماتٍ أو قواعد مُحددة.

كان العلماء في السابق يُنشئون نظرياتٍ لتفسير البيانات والتنبُّؤ بها؛ في حين يُنشئ الكمبيوتر في تعلُّم الآلة نماذج خاصة به تتناسب مع البيانات. إذَن فنقطة البداية هي البيانات، وليس النظريات. ومن هذا المُنطلق، لم تعُد البيانات «سلبية» بل «نشطة»: «فالبيانات نفسها هي التي تُحدِّد ما يجب القيام به بعد ذلك» (Alpaydin 2016, 11). يُدرِّب الباحثون الخوارزمية باستخدام مجموعات البيانات الموجودة (على سبيل المثال، رسائل البريد الإلكتروني القديمة)، وعندئذٍ تستطيع الخوارزمية التنبُّؤ بالنتائج من البيانات الجديدة (على سبيل المثال، البريد الإلكتروني الوارد الجديد) (CDT 2018). يُشار أحيانًا إلى التعرُّف على الأنماط في كمياتٍ كبيرة من المعلومات (البيانات الضخمة) باسم «التنقيب عن البيانات»، تشبيهًا له باستخراج المعادن القَيِّمة من الأرض. ومع ذلك، فإن المصطلح مُضلِّل لأن الهدف هو استخراج أنماطٍ من البيانات، وتحليل البيانات، وليس استخراج البيانات نفسها.

يمكن أن يكون تعلُّم الآلة «مُوجَّهًا»، مما يَعني أن الخوارزمية تركِّز على متغيِّر مُعيَّن يُعرَف باسم هدف التنبؤ. على سبيل المثال، إذا كان الهدف هو تقسيم الأشخاص إلى فئتَين (على سبيل المثال، خطورة أمنية عالية أو منخفضة)، فإن المُتغيرات التي تتنبأ بهاتَين الفئتَين معروفة بالفعل، وبالتالي تتعلَّم الخوارزمية التنبُّؤ بالانتماء إلى إحدى الفئتَين (الخطورة الأمنية العالية أو الخطورة الأمنية المنخفضة). يُدرِّب المبرمج النظام عن طريق توفير أمثلة وغيرها، على سبيل المثال، صور للأشخاص الذين يُشكِّلون خطورة أمنية عالية وأمثلة للأشخاص الذين لا يُشكلون خطورة أمنية. يكون الهدف أن يتعلَّم النظام التنبُّؤ بمَن ينتمي إلى كل فئة، أي مَن يُشكل خطورةً أمنية عالية ومَن لا يشكل بناءً على البيانات الجديدة. إذا أُعطِي النظام ما يكفي من الأمثلة، فإنه سيكون قادرًا على التعميم من هذه الأمثلة ومعرفة كيفية تصنيف البيانات الجديدة، مثل صورةٍ جديدة لراكبٍ يمرُّ عَبْر أمن المطار. أما تعلُّم الآلة «غير المُوجَّه» فيعني عدم تقديم هذا النوع من التدريب، وأن الفئات غير معروفة: ومن ثَم تُنشئ الخوارزميات فئاتٍ خاصَّة بها. على سبيل المثال،

يُنشئ الذكاء الاصطناعي فئاتٍ أمنيةً خاصةً به استنادًا إلى المُتغيرات التي يُحددها؛ لا التي يُقدمها إليه المبرمج. وربما يعثر الذكاء الاصطناعي على أنماطٍ لم يُحدِّدها خبراء المجال (في هذا السياق: الخبراء الأمنيون). ويمكن أن تبدو الفئات التي أنشأها الذكاء الاصطناعي من منظور البشر عشوائية للغاية. وربما لا يكون لها معنى. ولكنها موجودة من الناحية الإحصائية. وفي بعض الأحيان يكون لها معنى، وفي هذه الحالة يمكن لهذه الطريقة أن تُعطينا معرفةً جديدة حول الفئات في العالم الواقعي. أما التعلُّم «المُعزَّز»، فإنه يتطلب تقييمًا للمُخرجات إن كانت جيدة أم سيئة. وهذا يُشبه فكرة الثواب والعقاب. فالبرنامج لا يُخبَر أيُّ الإجراءات يجب أن يُتخَذ، ولكنه «يتعلم» من خلال عملية تكرارية أي الإجراءات التي تؤدي إلى الثواب. ففي المثال الأمني السابق، يتلقى النظام تقريرًا (أو بيانات) من الخبراء الأمنِيِّين بحيث «يعرف» ما إذا كان قد قام بعملٍ جيد عندما يجري تنبؤًا معينًا. فإذا لم يُسبب الشخص الذي تنبأ النظام بأنه ذو خطورة أمنية منخفضة أيَّ مشكلاتٍ أمنية، فإن النظام يتلقى تقريرًا بأن مخرجاته كانت جيدة ومن ثَم «يتعلم» منه. يجب ملاحظة أن هناك دائمًا نسبةً من الخطأ: فالنظام ليس دقيقًا بنسبة ١٠٠ في المائة. يجب أيضًا ملاحظة أن المُصطلَحَين الفنِّيَّين «موجَّه» و«غير موجَّه» لا علاقة لهما بمدى التدخُّل البشري في استخدام التكنولوجيا: ففي حين أن الخوارزمية تتمتَّع ببعض الاستقلالية، فإن البشر في جميع أنواع تعلُّم الآلة يتدخَّلون بطرقٍ مختلفة.

هذا صحيح أيضًا فيما يخصُّ البيانات في مجال الذكاء الاصطناعي، بما في ذلك ما يُسمَّى بـ «البيانات الضخمة». اكتسب تعلُّم الآلة القائم على البيانات الضخمة الكثير من الاهتمام بسبب توفر كميات كبيرة من البيانات وزيادة قدرة الكمبيوتر (الأرخص). يتحدَّث بعض الباحثين عن «زلزال البيانات» (Alpaydin 2016, x). نحن جميعًا نُنتِج بيانات من خلال أنشطتنا الرقمية، مثلما يحدُث على سبيل المثال عندما نستخدِم وسائل التواصل الاجتماعي أو عندما نشتري منتجاتٍ عبر الإنترنت. هذه البيانات مهمة بالنسبة إلى الجهات التجارية وأيضًا بالنسبة إلى الحكومات والعلماء. لقد صار جمع البيانات وتخزينها ومعالجتها أسهل بكثير على المؤسسات (Kelleher and Tierney 2018). وليس ذلك بسبب تعلُّم الآلة فقط: فالبيئة الرقمية الأوسع وتقنيات الوسائط الرقمية الأخرى تلعب دورًا مُهمًّا في هذا الصدد. إذ تيسر التطبيقات عبر الإنترنت ووسائل التواصُل الاجتماعي جمع البيانات من الأفراد. كما أن تخزين البيانات أصبح أقلَّ تكلفة، وأصبحت

أجهزة الكمبيوتر ذات إمكانياتٍ أكبر. كل هذا كان مُهمًّا لتطوير الذكاء الاصطناعي بشكلٍ عام، وعلم البيانات بشكل خاص.

علم البيانات

نستنتِج مما سبق أن تعلُّم الآلة يرتبط بـ «علم البيانات». إذ يهدف علم البيانات إلى استخراج أنماطٍ مفيدة وذات معنًى من مجموعات البيانات، وفي الوقت الحالي هذه المجموعات كبيرة جدًّا. يستطيع تعلُّم الآلة تحليل هذه المجموعات الكبيرة من البيانات آليًّا. ويعتمِد تعلُّم الآلة وعلم البيانات على الإحصاءات، أو على الانتقال من الملاحظات الفردية إلى توصيفاتٍ عامة. فعلماء الإحصاء يهتمُّون بالعثور على ارتباطاتٍ في البيانات من خلال التحليل الإحصائي. وتبحث عمليات إنشاء النماذج الإحصائية عن العلاقات الرياضية بين المدخلات والمخرجات. وهذا هو ما تساعد فيه خوارزميات تعلُّم الآلة.

> نحن جميعًا نُنتج بيانات من خلال أنشطتنا الرقمية، كما يحدث على سبيل المثال عندما نستخدِم وسائل التواصُل الاجتماعي أو عندما نشتري مُنتجات عبر الإنترنت.

ولكن علم البيانات ينطوي على أكثر من مجرد تحليل البيانات بواسطة تعلُّم الآلة. إذ يجب جمع البيانات وإعدادها قبل تحليلها، وبعد ذلك يجب تفسير نتائج التحليل. وينطوي علم البيانات على تحدِّيات مثل كيفية الحصول على البيانات وتنقيتها (على سبيل المثال، من وسائل التواصل الاجتماعي والويب)، وكيفية الوصول إلى كميةٍ كافية من البيانات، وكيفية جمع مجموعات البيانات معًا، وكيفية إعادة هيكلة مجموعات البيانات، وكيفية اختيار مجموعات البيانات ذات الصلة، وأي نوع من البيانات يتم استخدامه. لذلك لا يزال البشَر يلعبون دورًا مهمًّا في جميع المراحل وفيما يتعلق بجميع هذه الجوانب، بما في ذلك صياغة المشكلة، والحصول على البيانات، وإعداد البيانات (مجموعة البيانات التي تتدرَّب عليها الخوارزمية ومجموعة البيانات التي ستُطبق عليها)، وإنشاء خوارزمية التعلُّم أو اختيارها، وتفسير النتائج، واتخاذ قرار حول الإجراء الذي يجب اتخاذه (Kelleher and Tierney 2018).

تظهر التحدِّيات العلمية في كل مرحلة من هذه العملية، وعلى الرغم من أن البرامج قد تكون سهلة الاستخدام، فإن مواجهة هذه التحديات تتطلَّب وجود المعرفة البشرية الخبيرة المُتخصِّصة. وعادةً ما يكون التعاون بين البشر أمرًا ضروريًّا أيضًا، على سبيل المثال، بين علماء البيانات والمهندسين. ومن الوارد حدوث أخطاء طوال الوقت، لذا فإن الاختيار البشري والمعرفة البشرية والتفسير البشري أمر حاسم الأهمية. فالبشر مهمُّون في هذا السياق لتفسير الأمور على نحوٍ معقول وتوجيه التكنولوجيا نحو البحث عن عوامل وعلاقات مختلفة. والذكاء الاصطناعي، من وجهة نظر بودن (٢٠١٦)، يفتقر إلى فهمنا للصِّلات والعلاقات. ويمكننا أن نُضيف أنه يفتقر أيضًا إلى الفهم والتجربة والحساسية والحكمة. وهذه حجة جيدة تدعم نظريًّا ومبدئيًّا ضرورة مشاركتنا نحن البشر في الأمر. ولكن ثمة حجة عملية أيضًا تدعم عدم خروج البشر من المشهد؛ وهي أن البشر يشاركون بالفعل عمليًّا في الأمر. فدون المبرمجين وعلماء البيانات، لن تستطيع التكنولوجيا القيام بوظيفتها ببساطة. علاوةً على ذلك، كثيرًا ما يتم دمج الخبرة البشرية مع الذكاء الاصطناعي، على سبيل المثال، عندما يستخدم الطبيب استراتيجية علاج سرطان يوصي بها الذكاءُ الاصطناعي، ولكنه في الوقت نفسه يعتمد على تجاربه وحدسه كخبير. فإذا أُلغي التدخل البشري، يمكن أن تسوء الأمور أو تفقد معناها أو ببساطة تُصبح غير منطقية.

ولنضرب مثلًا بالمشكلة المعروفة التالية من الإحصاء، والتي تؤثر بدَورها على استخدام تعلُّم الآلة: الارتباطات لا تعني بالضرورة علاقاتٍ سببية. يُقدم تايلر فيجين في كتابه «الارتباطات الزائفة» (٢٠١٥) بعض الأمثلة الجيدة على ذلك. في الإحصاء، الارتباط الزائف هو الارتباط الذي تكون فيه المُتغيرات غير مرتبطة فيما بينها بعلاقاتٍ سببية ولكنها قد تبدو كذلك؛ ويكون الارتباط ناجمًا عن وجود عاملٍ ثالث غير مرئي. من بين الأمثلة التي يُقدِّمها فيجين الارتباط بين معدل الطلاق في ولاية مين ومعدل استهلاك السمن النباتي للفرد الواحد، أو الارتباط بين معدل استهلاك جبن الموتزاريلا للفرد الواحد والحصول على دكتوراه في الهندسة المَدنية.[1] ربما يعثر الذكاء الاصطناعي على مثل هذه الارتباطات، ولكن يجب أن يتدخَّل البشر لتقرير الارتباطات التي تستحقُّ مزيدًا من الدراسة من أجل العثور على علاقاتٍ سببية.

فضلًا عن ذلك، في المرحلة التي يتم فيها جمع البيانات وتصميم أو إنشاء مجموعة البيانات، نجري اختياراتٍ فيما يخصُّ كيفية التجريد عن الواقع (Kelleher and Tierney 2018). والتجريد عن الواقع لا يكون مُحايدًا أبدًا، والتجريد نفسه ليس واقعًا؛ وإنما هو

تمثيل للواقع. وهذا يَعني أنه يُمكننا مناقشة مدى جودة هذا التمثيل وملاءمته، فيما يتعلق بغرَض مُعين. قارن هذا بأية خريطة: الخريطة نفسها ليست هي الإقليم، وقد اختار البشَر طريقة تصميم الخريطة لغرَضٍ مُعين (على سبيل المثال، خريطة لملاحة السيارات مقابل خريطة طوبوغرافية للتنزُّه سيرًا على الأقدام). في تعلُّم الآلة، يعمل التجريد باستخدام الأساليب الإحصائية على إنشاء نموذج للواقع؛ إنه ليس الواقع الفعلي. كما يتضمَّن ذلك اختيارات: اختيارات بشأن الخوارزمية نفسها التي تُوفِّر العملية الإحصائية التي تأخذنا من البيانات إلى النمط/القاعدة، ولكن أيضًا اختيارات بشأن تصميم مجموعة البيانات التي تتدرَّب عليها الخوارزمية. يعني هذا الجانب الاختياري، ومن ثَم الجانب البشري، في تعلُّم الآلة أنه يُمكننا أن نطرح أسئلةً نقدية حول الاختيارات التي تُتَّخَذ، بل يجب علينا أن نفعل ذلك. على سبيل المثال، هل مجموعة البيانات التي سيتم التدريب عليها تُمثل السكان تمثيلًا جيدًا؟ هل هناك أي تحيُّزات في البيانات؟ كما سنرى في الفصل القادم، هذه الاختيارات والقضايا ليست مجرد أسئلة فنية ولكن لها أيضًا جانب أخلاقي شديد الأهمية.

التطبيقات

لتعلُّم الآلة وعلم البيانات تطبيقاتٌ عديدة، ذَكرتُ بعضها بالفعل تحت العنوان الأعم المُتمثل في الذكاء الاصطناعي. هذه التقنيات يُمكن استخدامها للتعرُّف على الوجوه (بل للتعرُّف على الانفعالات بناءً على تحليل الوجوه)، أو تقديم اقتراحات بحث، أو قيادة السيارة، أو إجراء توقُّعات شخصية، أو التنبُّؤ بمَن سيعاود ارتكاب الجريمة، أو التوصية بموسيقى مُعينة للاستماع إليها. وتستخدَم في مجال المبيعات والتسويق، للتوصية بمنتجات وخدمات. على سبيل المثال، عندما تشتري شيئًا على موقع أمازون، سيجمع الموقع بياناتٍ عنك ثم يُقدم توصيات على أساس نموذج إحصائي يستند إلى بياناتٍ من جميع العملاء. استخدمت شركة وولمارت في متاجرها تقنية التعرُّف على الوجوه للتصدي للسرقة؛ وقد تستخدم في المُستقبل التقنية نفسها لتحديد ما إذا كان المُتسوقون سعداء أم مُحبَطين. كما أن للتقنيات تطبيقات مختلفة في مجال التمويل. تعاونت وكالة إكسبريان للمرجعية الائتمانية مع الذكاء الاصطناعي المدعوم بتعلُّم الآلة لتحليل البيانات المُتعلقة بالمُعاملات والقضايا المنظورة في المحاكم من أجل التوصية بما إذا كان يجب تقديم قرضٍ لمُقدِّم طلب لرهن عقاري. وتستخدم أمريكان إكسبريس تعلُّم

الآلة لتوقع المعاملات الاحتيالية. وفي مجال النقل، يُستخدَم الذكاء الاصطناعي والبيانات الضخمة لإنشاء سيارات ذاتية القيادة. على سبيل المثال، تستخدِم شركة بي إم دبليو نوعًا من تقنية التعرُّف على الصور لتحليل البيانات الواردة من أجهزة الاستشعار والكاميرات في السيارة. وفي مجال الرعاية الصحية، يمكن أن يُساعد الذكاء الاصطناعي المدعوم بتعلُّم الآلة في تشخيص السرطان (على سبيل المثال، في تحليل صور الأشعة لتشخيص مرض السرطان) أو اكتشاف الأمراض المُعدية. على سبيل المثال، أجرى نظام الذكاء الاصطناعي لشركة ديب مايند تحليلًا لمليون صورة من صور أشعة العيون وبيانات المرضى، مُدربًا نفسه على تشخيص أعراض حالات العيون المرضية المُتدهورة. وقد تجاوز نظام واتسون الذي أنشأته شركة آي بي إم مُمارسة لعبة «جيوباردي» ويستخدم لتقديم توصياتٍ بشأن علاج السرطان. كما تُزوِّد أجهزة الرياضة والصحة التي يمكن ارتداؤها تطبيقات تعلُّم الآلة بالبيانات. وفي مجال الصحافة، يمكن لتعلُّم الآلة كتابة تقارير إخبارية. على سبيل المثال، في المملكة المتحدة، تستخدِم وكالة أنباء «بريس أسوسييشن» الروبوتات في كتابة تقارير الأخبار المحلية. ويدخل الذكاء الاصطناعي أيضًا إلى المنزل والمجال الشخصي، على سبيل المثال، في شكل روبوتات تتولَّى جمع البيانات وأجهزة تفاعُلية مساعدة متَّصلة بمعالجة اللغة الطبيعية. تتحدَّث دُمية «هالو باربي» إلى الأطفال باستخدام مُعالجة اللغة الطبيعية التي تُحلل المحادثات المسجلة. فكلُّ ما يقوله الأطفال يتم تسجيلُه وتخزينه وتحليله في وحدات الخدمة الخاصة بـ «توي توك». ثم يُرسل ردًّا إلى الجهاز: وتجيب دمية «هالو باربي» على أساس ما «تعلمته» عن مُستخدمها. ويستخدِم فيسبوك تقنيات التعلُّم العميق والشبكات العصبية لهيكلة وتحليل البيانات الآتية مما يقرُب من مليارَي مستخدم للمنصَّة يُنتجون بياناتٍ غير مُهيكلة. وهذا يساعد الشركة في تقديم إعلانات مُستهدفة. ويحلِّل إنستجرام صور ٨٠٠ مليون مُستخدِم بهدف بيع الإعلانات إلى الشركات. ويستخدم نتفليكس محركات التوصية التي تُحلِّل بيانات العملاء، لكي يُحوِّل نفسه من موزع إلى منتج محتوى: فإذا كنتَ تستطيع التنبُّؤ بما يرغب الناس في مشاهدته، فيُمكنك إنتاجه بنفسك وتحقيق ربح منه. بل إن علم البيانات استُخدِم في مجال الطهي. على سبيل المثال، بناءً على تحليل نحو ١٠٠٠٠ وصفة، يُنشئ نظام شيف واتسون الذي أنتجته شركة آي بي إم وصفاته الخاصة التي تقترح توليفات جديدة للمكونات.[2] ويمكن أيضًا استخدام الذكاء الاصطناعي المدعوم بتعلُّم الآلة في التعليم، والتوظيف، والعدالة الجنائية، والأمن

(على سبيل المثال، الشرطة التنبؤية)، واسترجاع الموسيقى، والأعمال المكتبية، والزراعة، والأسلحة العسكرية، وما إلى ذلك.

في الماضي، كانت الإحصاء من المجالات غير الجذابة. أما اليوم، فبعد أن أصبحت جزءًا من عِلم البيانات وفي شكلٍ يُدمَج فيه الذكاء الاصطناعي مع البيانات الضخمة، أصبحت الإحصاء شديدة الجاذبية. إنها السحر الجديد. إنها المجال الذي تُفضِّله وسائل الإعلام. كما أنها تُعتبر مجالَ أعمال ضخمًا. فالبعض يتحدَّثون عن نوعٍ جديد من التنقيب عن الذهب؛ والتوقعات هائلة. علاوةً على ذلك، فهذا النوع من الذكاء الاصطناعي ليس خيالًا علميًّا أو محضَ نبوءة، كما تُبين الأمثلة التي ضربناها أن ما يُسمَّى بالذكاء الاصطناعي المحدود أو الضعيف موجود بالفعل وواسع الانتشار. وفيما يتعلق بتأثيره المُحتمَل، فليس هناك ما يُمكننا أن نصِفه بأنه محدود أو ضعيف. لذلك، فإنه من الضروري جدًّا أن نُحلِّل ونُناقش العديد من القضايا الأخلاقية التي أثارتها تقنيات تعلُّم الآلة وغيرها من تقنيات الذكاء الاصطناعي وتطبيقاتها. وهذا هو موضوع الفصول القادمة.

في الماضي، كانت الإحصاء من المجالات غير الجذَّابة. أما اليوم، فبعد أن أصبحت جزءًا من علم البيانات وفي شكلٍ يُدمج فيه الذكاء الاصطناعي مع البيانات الضخمة، أصبحت الإحصاء شديدة الجاذبية. إنها السحر الجديد.

الفصل السابع

الخصوصية وغيرها من القضايا

إن العديد من المشكلات الأخلاقية المُتعلقة بالذكاء الاصطناعي معروفة من مجال أخلاقيات الروبوتات والأتمتة أو، بشكلٍ أعم، من مجال أخلاقيات التكنولوجيا الرقمية وتكنولوجيا الاتصالات. ولكن هذا في حدِّ ذاته لا يُقلِّل من أهميتها. وعلاوةً على ذلك، فإن هذه القضايا — بسبب التكنولوجيا وطريقة ارتباطها بتقنياتٍ أخرى — تكتسب بُعدًا جديدًا وتُصبح أكثر أهمية وإلحاحًا.

الخصوصية وحماية البيانات

فلنُفكر، على سبيل المثال، في مسألة الخصوصية وحماية البيانات. ينطوي الذكاء الاصطناعي، ولا سيما تطبيقات تعلُّم الآلة التي تتعامَل مع البيانات الضخمة، غالبًا على جمع المعلومات الشخصية واستخدامها. ويُمكن أيضًا استخدام الذكاء الاصطناعي للمُراقبة، في الشارع وأيضًا في مكان العمل وفي كل مكان، وذلك من خلال الهواتف الذكية ووسائل التواصُل الاجتماعي. وفي كثيرٍ من الأحيان، لا يعلم الناس حتى أن البيانات تُجمَع، أو أن البيانات التي قدموها في سياقٍ ما تُستخدَم بواسطة أطراف أخرى في سياقٍ آخر. كما أن البيانات الضخمة غالبًا ما تَعني أن (مجموعات) البيانات التي تحصل عليها المنظمات المختلفة يتم دمجها معًا.

يتطلَّب الاستخدام الأخلاقي للذكاء الاصطناعي جمع البيانات ومعالجتها ومشاركتها بطريقةٍ تحترم خصوصية الأفراد وحقَّهم في معرفة ما يحدث لبياناتهم، والوصول إلى بياناتهم، والاعتراض على جمع بياناتِهم أو على مُعالجتها، ومعرفة أن بياناتهم تُجمع وتُعالج وأنهم بعدئذٍ يخضعون لقراراتٍ يتخذها الذكاء الاصطناعي (في حالة حدوث ذلك

بالفعل). وتُثار العديد من هذه القضايا أيضًا في سياقات تكنولوجيا المعلومات وتكنولوجيا الاتصالات الأخرى، وكما سنرى فيما بعدُ في هذا الفصل، تعتبر الشفافية شرطًا مُهمًّا أيضًا في تلك الحالات (انظر لاحقًا في هذا الفصل). كما تُثار قضايا حماية البيانات في أخلاقيات البحث، على سبيل المثال، في أخلاقيات جمع البيانات لأبحاث العلوم الاجتماعية.

ومع ذلك، عند النظر إلى السياقات التي يُستخدَم فيها الذكاء الاصطناعي اليوم، تُصبح قضايا الخصوصية وحماية البيانات أكثر تعقيدًا. فإن احترام هذه القِيَم والحقوق يكون سهلًا إلى حدٍّ ما عند إجراء استبيانٍ كعالِم اجتماع: إذ يمكن للباحث إبلاغ المشاركين في الاستبيان وطلب موافقتهم بشكلٍ صريح، ومِن ثَم سيكون من المعروف نسبيًّا ما سيحدث للبيانات. ولكن البيئة التي يُستخدَم فيها الذكاء الاصطناعي وعلم البيانات اليوم عادةً ما تكون مختلفةً تمامًا. فلنتناول مثلًا وسائل التواصل الاجتماعي: على الرغم من معلومات الخصوصية والتطبيقات التي تطلُب من المُستخدِمين الموافقة، فإن المُستخدِمين لا يعرفون بوضوح ما يحدُث لبياناتهم أو حتى أي بيانات يتم جمعها؛ وإذا كانوا يرغبون في استخدام التطبيق والاستمتاع بفوائده، فعليهم أن يوافقوا. وفي كثيرٍ من الأحيان، لا يعلم المُستخدمون حتى أن الذكاء الاصطناعي يُشغِّل التطبيق الذي يستخدمونه. وغالبًا ما تُنقَل البيانات المُعطاة في سياقٍ ما إلى نطاقٍ آخَر واستخدامها لأغراض مختلفة (إعادة استخدام البيانات في أغراضٍ أخرى)، على سبيل المثال، عندما تبيع الشركات بياناتها إلى شركات أخرى أو تنقل البيانات بين أجزاءٍ مختلفة من نفس الشركة دون عِلم المستخدِمين بهذا.

التلاعُب والاستغلال والمُستخدِمين المُستهدفين

تُشير هذه الظاهرة الأخيرة أيضًا إلى احتمالية التلاعُب بالمُستخدمين واستغلالهم. يُستخدَم الذكاء الاصطناعي للتحكم فيما نشتريه، وفي الأخبار التي نُتابعها، وفي الآراء التي نثِق بها، وغير ذلك. وقد أشار الباحثون في النظرية النقدية إلى السياق الرأسمالي الذي يحدث فيه استخدام وسائل التواصل الاجتماعي. على سبيل المثال، يمكن القول إن مُستخدمي وسائل التواصل الاجتماعي يؤدُّون «عملًا رقميًّا» مجانيًّا (Fuchs 2014) من خلال إنتاج البيانات لصالح الشركات. ويُمكن أن يشمل هذا الشكل من أشكال الاستغلال أيضًا الذكاء الاصطناعي. فبوصفنا مُستخدمين لوسائل التواصل الاجتماعي، نحن نتعرض لخطر أن نصبح القوة العاملة المُستغلَّة غير المأجورة، التي تنتج البيانات لصالح الذكاء الاصطناعي

الذي يُحلل بياناتنا بعد ذلك لصالح الشركات التي تستخدم البيانات، والتي عادةً ما تتضمَّن أطرافًا أخرى أيضًا. وهذا يُذكِّرنا أيضًا بتحذير هيربرت ماركوزه في ستينيات القرن العشرين بأنه حتى في المجتمعات المُسمَّاة مجتمعات «حرة»، و«غير شمولية»، هناك أشكال خاصة من السيطرة، وخاصة استغلال المُستهلكين (Marcuse 1991). يكمن الخطر هنا في أن الذكاء الاصطناعي قد يؤدي حتى في الديمقراطيات الحديثة إلى أشكالٍ جديدة من التلاعُب والمراقبة والاستبداد، ليس بالضرورة في شكل سياسات استبدادية ولكن بطريقة أكثر خفاءً وفعالية: من خلال تغيير الاقتصاد بطريقة تُحوِّلنا جميعًا — في استخدامنا للهواتف الذكية والتفاعلات الرقمية الأخرى — إلى ما يُشبه الأبقار التي يتم حلبها للحصول على بياناتها. ولكن يمكن أيضًا استخدام الذكاء الاصطناعي للتلاعُب في السياسة بشكلٍ أكثر مباشرة، على سبيل المثال، من خلال تحليل بيانات وسائل التواصُل الاجتماعي لدعم حملات سياسية مُعينة (كما في الحالة الشهيرة لشركة كامبريدج أناليتيكا، التي استخدمت بيانات مُستخدمي فيسبوك — دون موافقتهم — لأغراض سياسية في انتخابات الرئاسة الأمريكية عام ٢٠١٦)، أو عن طريق استخدام روبوتات لنشر رسائل سياسية على وسائل التواصُل الاجتماعي استنادًا إلى تحليل بيانات الأفراد من حيث تفضيلاتهم السياسية للتأثير على عمليات التصويت. كما أن البعض يُساورهم القلق من أن يُحوِّل الذكاء الاصطناعي، من خلال تولِّيه المهام المعرفية نيابةً عن البشر، قد يُحوِّل مُستخدِميه إلى أطفال على المستوى العقلي عن طريق «تقليل قدرتهم على التفكير بمحض أنفسهم أو اتخاذ قراراتهم الخاصة بما يجب فعله» (Shanahan 2015, 170). علاوةً على ذلك، لا يكمن خطر الاستغلال في جانب المستخدم فحسب: فالذكاء الاصطناعي يعتمد على أجهزة صنعها أشخاص، وقد ينطوي إنشاء هذه الأجهزة على استغلال هؤلاء الأشخاص. وقد يدخل الاستغلال أيضًا في تدريب الخوارزميات وإنتاج البيانات التي تُستخدَم لصالح الذكاء الاصطناعي وعن طريقه. إن الذكاء الاصطناعي ربما يجعل الحياة أيسر بالنسبة إلى مُستخدِميه، ولكن ليس بالضرورة بالنسبة إلى أولئك الذين يُنقِّبون عن المعادن، أو بالنسبة إلى مَن يتعاملون مع المُخلفات الإلكترونية، أو إلى مَن يُدربون الذكاء الاصطناعي. على سبيل المثال، لا يقتصر ما يقوم به تطبيق «أليكسا» الذي طوَّرته أمازون إكو على إنشاء مُستخدِمين يؤدُّون عملًا مجانيًّا ويُصبحون مصادر للبيانات ويُباعون كمنتجات؛ بل هناك عالَم من العمل البشري يكمُن خلف الكواليس: فعُمَّال التنقيب عن المعادن، والعمال على السفن، والعمال الذين يُصنفون مجموعات البيانات، كل هؤلاء في خدمة تجميع رءوس الأموال وتراكُمها لدى عدد قليل جدًّا من الأشخاص (Schwab 2018).

قد يؤدي الذكاء الاصطناعي إلى أشكالٍ جديدة من التلاعُب والمراقبة والاستبداد، ليس بالضرورة في شكل سياساتٍ استبدادية ولكن بطريقةٍ أكثر خفاءً وفعالية.

بعض مُستخدمي الذكاء الاصطناعي أكثر تعرُّضًا للخطر من غيرهم. ونظريات الخصوصية والاستغلال غالبًا ما تفترض أن المستخدِم شخص بالِغ سليم الجسم، صغير السن نسبيًّا، في كامل قواه العقلية. لكنَّ العالَم الحقيقي مليء بالأطفال وكبار السن والأشخاص الذين لا يتمتعون بقوًى عقلية «طبيعية» أو «كاملة»، وغيرهم. مثل هؤلاء المُستخدمين الضعفاء أكثر عُرضةً للخطر. ويمكن انتهاك خصوصيتهم أو التلاعب بهم بسهولة، ويوفِّر الذكاء الاصطناعي فرصًا جديدة لهذه الانتهاكات وعمليات التلاعب. فكِّر مثلًا في الأطفال الصغار الذين يتحدثون مع دميةٍ متصلة بنظام تكنولوجي مدعوم بالذكاء الاصطناعي: على الأرجح، هؤلاء الأطفال لا يعلمون شيئًا عن الذكاء الاصطناعي المُستخدَم أو عن جمع بياناتهم، فما بالك بما يُفعَل بمعلوماتهم الشخصية. إن روبوت الدردشة أو الدمية الذكية المدعومة بالذكاء الاصطناعي لا تستطيع فقط أن تجمع الكثير من المعلومات الشخصية عن الطفل وأبويه بهذه الطريقة، بل يُمكنها أيضًا التلاعُب بالطفل باستخدام واجهة اللغة والصوت. ومع تحوُّل الذكاء الاصطناعي إلى جزءٍ من «إنترنت الألعاب» (Druga and Williams 2017) وإنترنت الأشياء (الأخرى)، تُصبح هذه مشكلة أخلاقية وسياسية. إن شبح الشمولية والاستبداد يُعاود الظهور مجددًا: ليس في قصص الخيال العلمي المُتشائمة أو في كوابيس ما بعد الحروب القديمة، ولكن في التكنولوجيا الاستهلاكية الموجودة بالفعل في الأسواق.

الأخبار الكاذبة، وخطر الشمولية، وتأثيرها على العلاقات الشخصية

يمكن أن يُستخدَم الذكاء الاصطناعي أيضًا في إنتاج خطاب الكراهية والمعلومات الزائفة، أو في إنشاء روبوتات تبدو كأشخاصٍ ولكنها في الواقع مجرد برامج مدعومة بالذكاء الاصطناعي. وقد سبق وأشرت بالفعل إلى روبوت الدردشة «تاي» وخطاب أوباما الزائف. قد يؤدي ذلك إلى عالَمٍ لا يمكن فيه التمييز بوضوح بين ما هو حقيقي وما هو زائف، عالم تتداخَل فيه الحقائق مع الخيال. وسواء كان يجِب تسميتها «ما بعد الحقيقة» أم لا (McIntyre 2018)، تساهم هذه التطبيقات للذكاء الاصطناعي بشكلٍ واضح في

المشكلة. بالطبع، كان يُوجَد تلاعب ومعلومات كاذبة قبل ظهور الذكاء الاصطناعي. فالأفلام، على سبيل المثال، كانت دائمًا تخلُق أوهامًا، والصحف كانت تنشر الدعاية الكاذبة. ولكن بعد ظهور الذكاء الاصطناعي، جنبًا إلى جنبٍ مع إمكانيات وبيئة الإنترنت ووسائل التواصل الاجتماعي الرقمية، يبدو أن المشكلة تزداد تعقيدًا وحِدَّة. ويبدو أن هناك المزيد من الفُرَص للتلاعُب، مما يعرض التفكير النقدي للخطر. وكل هذا يُذكرنا مرة أخرى بخطورة الشمولية، التي تستفيد من الْتباس الحقيقة وتنتج أخبارًا زائفة لأغراض أيديولوجية.

ومع ذلك، حتى في اليوتوبيا الليبرالية قد لا تكون الحياة غاية في الإشراق والبهاء. إذ إن المعلومات الكاذبة تنخر في جدار الثقة ومن ثَم تفسد النسيج الاجتماعي. ويُمكن أن يؤدي الاستخدام المُفرط للتكنولوجيا إلى تقليل التواصُل، أو على الأقل التواصل الهادف، بين الأفراد. في عام ٢٠١١، قدمت شيري تيركل ادعاءً يتعلق بالتكنولوجيا مثل أجهزة الكمبيوتر والروبوتات: لقد انتهى بنا الأمر إلى توقُّع المزيد من التكنولوجيا، والقليل من أنفسنا. ويمكن أيضًا استخدام هذه الحجَّة فيما يتعلق بالذكاء الاصطناعي: تكمُن المشكلة في أن الذكاء الاصطناعي، في شكل وسائل التواصل الاجتماعي أو في شكل «الرفاق» الرقميين، يُعطينا وهْم الرفقة ولكنه يُزعزع استقرار العلاقات الحقيقية مع الأصدقاء والأحباء والعائلات. وعلى الرغم من أن هذه المشكلة كانت موجودةً بالفعل قبل الذكاء الاصطناعي وتزداد تفاقمًا مع ظهور كل وسيطٍ جديد من الوسائط (قراءة الصحف أو مشاهدة التلفيزيون بدلًا من التحدُّث وإدارة حوار)، فإنه يمكن القول إن التكنولوجيا الآن، في وجود الذكاء الاصطناعي وتطبيقه، قد أصبحت أفضل بكثيرٍ في خلق وهْم الرفقة، وأن هذا يزيد من خطر الوحدة أو تدهور العلاقات الشخصية.

السلامة والأمان

هناك أيضًا مخاطر أوضح. فالذكاء الاصطناعي، لا سيَّما في حال تضمينه في أنظمة الأجهزة التي تعمل في العالَم الفعلي، يحتاج أيضًا إلى أن يكون آمنًا. ولنضرب مثلًا على ذلك بالروبوتات الصناعية: يفترض ألا تُلحِق هذه الروبوتات الأذى بالعمال. ومع ذلك، تحدث أحيانًا حوادث في المصانع. ويمكن للروبوتات أن تقتُل، حتى لو كان ذلك نادرًا نسبيًّا. ومع ذلك، في الروبوتات التي تعتمِد على الذكاء الاصطناعي، تُصبح مشكلة السلامة أكثر تحدِّيًا: فهذه الروبوتات قد تتمكَّن من العمل جنبًا إلى جنبٍ مع البشر، وقد تتمكَّن

من تجنُّب إلحاق الأذى بالبشر «على نحوٍ ذكي». ولكن ماذا يعني ذلك بالضبط؟ هل يجب أن تتحرك ببطءٍ أكبر عندما تكون قريبةً من البشر، مما يُبطئ العملية، أم أنه من المقبول التحرك بسرعةٍ عالية من أجل إنجاز العمل بكفاءة وسرعة؟ هناك دائمًا احتمالات لحدوث خطأ من نوع ما. فهل يجب أن تنطوي أخلاقيات السلامة على الوصول إلى حلول وسط؟ تُثير الروبوتات المدعومة بالذكاء الاصطناعي في بيئة المنزل أو في الأماكن العامة أيضًا قضايا تتعلَّق بالسلامة. على سبيل المثال، هل يجب على الروبوت دائمًا تجنُّب الاصطدام بالبشر أم أنه من المقبول أحيانًا أن يُعرقل الروبوت شخصًا من أجل الوصول إلى هدفه؟ هذه ليست مسائل تقنية بحتة ولكن لها جانب أخلاقي: إنها مسألة حياة بشرية وقِيَم مثل الحرية والكفاءة. كما أنها تُثير مشكلاتٍ تتعلَّق بالمسئولية (سنتحدث عن هذا بتفصيل أكثر لاحقًا).

ثمَّة مشكلة أخرى كانت موجودة بالفعل قبل ظهور الذكاء الاصطناعي في المشهد، ولكنها تستحقُّ تجديد اهتمامنا بها؛ ألا وهي مشكلة الأمان. في عالم مُتصل بالشبكات، يمكن اختراق أي جهاز إلكتروني أو برنامج واختراقه والتلاعُب به من قبل أشخاص لديهم نوايا خبيثة. فكلُّنا نعلم بشأن فيروسات الكمبيوتر، على سبيل المثال، التي يمكن أن تُخرِّب جهاز الكمبيوتر الخاص بك. ولكن عند تزويد أجهزتنا وبرامجنا بالذكاء الاصطناعي، يمكن أن تزيد إمكانياتها وقدراتها، وعندما تحظى بوكالةٍ أخلاقية أكبر ويكون لهذا عواقب مادية في العالم الفعلي، تُصبح مشكلة الأمان أكبر بكثير. على سبيل المثال، إذا اختُرِقَت سيارتك الذاتية القيادة التي تعمل بالذكاء الاصطناعي، فسوف تُعاني مما هو أكثر من مجرد «مشكلة في الكمبيوتر» أو «مشكلة في البرنامج»؛ قد تلقى حتفك. وإذا اختُرِق برنامج إحدى البَنى التحتية المهمة (مثل الإنترنت، أو المياه، أو الطاقة ... إلخ) أو جهاز عسكري ذي قدراتٍ مدمرة، فمن المرجَّح أن يتعرض المجتمع بأكمله إلى اضطرابٍ كبير وسوف يتعرض الكثير من الأشخاص للضرر. في التطبيقات العسكرية، يشكِّل استخدام الأسلحة الفتاكة الذاتية التشغيل خطورة أمنية واضحة، لا سيَّما على المُستهدَفين بالطبع بهذه الأسلحة (وعادة ما لا يكونون من الغرب) ولكنه يشكِّل خطورة أيضًا على أولئك الذين ينشرونها: إذ يمكن دائمًا اختراقها وتحويلها ضدهم. علاوةً على ذلك، قد يؤدي سباق التسلُّح الذي يشمل هذه الأسلحة إلى حربٍ عالمية جديدة. ولا يلزمنا أن ننظر بعيدًا في المُستقبل: فإذا كانت الطائرات دون طيار (غير المُزودة بالذكاء الاصطناعي) يُمكنها بالفعل حاليًّا السيطرة على مطارٍ كبير في لندن، فإنه ليس من الصعب تخيُّل مدى

هشاشة منشآت بِنيتنا الأساسية اليومية وكيف يمكن للاستخدام المؤذي أو لاختراق الذكاء الاصطناعي أن يُسبب اضطراباتٍ جسيمة وعملياتٍ تدميريةً هائلة. لاحظ أيضًا أنه، على عكس التكنولوجيا النووية على سبيل المثال، فإن استخدام تكنولوجيا الذكاء الاصطناعي الحالية لا يتطلَّب معدَّات باهظة الثمن أو تدريبًا طويلًا؛ ومِن ثَم فالعائق أمام استخدام الذكاء الاصطناعي لأغراضٍ خبيثة مُنخفض نسبيًّا.

تُذكرنا أيضًا المشكلات العادية المتعلقة بالأمان مع السيارات ومنشآت البنية التحتية مثل المطارات بأنه على الرغم من أن بعض الأشخاص أكثر عرضةً للخطر من غيرهم، فإننا «جميعًا» مُعرَّضون للخطر في ظلِّ تقنياتٍ مثل الذكاء الاصطناعي لأننا، مع زيادة تمتُّع هذه التقنيات بالوكالة وزيادة تفويضنا لها لتأدية المزيد من المهام، نُصبح جميعًا أكثر اعتمادًا عليهم. وهناك احتمالٌ دائم أن تسير الأمور على غير ما نروم. ومن ثَم، يُمكننا القول إن المخاطر التكنولوجية الجديدة ليست مجرد مخاطر تكنولوجية، وإنما تتجاوز ذلك لتُصبح مخاطر تهدِّد وجودنا بصفتنا بشرًا (Coeckelbergh 2013). يمكن رؤية المشكلات الأخلاقية المطروحة هنا على أنها مخاطر إنسانية: فالمخاطر التكنولوجية تُهدد وجودنا كبشرٍ في نهاية المطاف. وبقدْر ما نعتمد على الذكاء الاصطناعي، وبقدْر ما يكون الذكاء الاصطناعي أكثر من مجرد أداةٍ نستخدمها؛ فإنه يُصبح جزءًا من هويتنا ومن المخاطر التي تحيق بنا في العالم.

> في عالمٍ مُتصل بالشبكات، يمكن اختراق أي جهاز إلكتروني أو برنامج واختراقه والتلاعُب به من قبل أشخاص لديهم نوايا خبيثة.

كذلك يُثير تمتُّع الذكاء الاصطناعي بالوكالة الأخلاقية، لا سيما إذا كانت تحلُّ محلَّ الوكالة الأخلاقية البشرية، مشكلة أخلاقية أخرى تزداد أهميةً مع مرور الوقت: ألا وهي المسئولية. وهذا هو موضوع الفصل القادم.

الفصل الثامن

لامسئوليةُ الآلات والقرارات غير المُبررة

كيف يمكن أن نسند المسئولية الأخلاقية وما الكيفية الواجبة لذلك؟

عند استخدام الذكاء الاصطناعي لاتخاذ قرارات وللقيام بأشياء بالنيابة عنَّا، فإننا نواجِهُ مشكلة مشتركة في جميع تقنيات الأتمتة، غير أن هذه المشكلة تزداد أهميةً عندما يُمكِّننا الذكاء الاصطناعي من تفويض المزيد والمزيد من القرارات إلى الآلات أكثر بكثيرٍ مما كنا نفعل في الماضي: وهذه المشكلة هي إسناد المسئولية.[1] إذا مُنح الذكاء الاصطناعي وكالة أكبر وأخذ على عاتقه ما كان يتولَّاه البشر في الماضي، فكيف نُسند المسئولية الأخلاقية عن أفعاله؟ مَن المسئول عن الأضرار والفوائد التي تنشأ عن التكنولوجيا عندما يفوض البشر الوكالة والقرارات إلى الذكاء الاصطناعي؟ وفيما يخصُّ المخاطر تحديدًا: مَن المسئول عند حدوث خطأ ما؟

عندما يقوم البشَر بأداء مهامَّ واتخاذ قرارات، فنحن عادةً ما نربط الوكالة بالمسئولية الأخلاقية. فأنت مسئول عما تفعله وعن القرارات التي تتَّخذها. وإذا كان لدَيك تأثير على العالم وعلى الآخرين، فأنت مسئول عن عواقب أفعالك. وفقًا لأرسطو، هذا هو الشرط الأول للمسئولية الأخلاقية، المعروف باسم الشرط التحكُّمي: في الأخلاقيات النيقوماخية، يقول أرسطو إن الفعل يجب أن ينشأ من الفاعل. ولهذا الرأي أيضًا جانب تقييمي: إذا كان لديك وكالة وإذا كنتَ قادرًا على اتخاذ قرارات، فينبغي أن تتحمَّل المسئولية عن أفعالك. وما نريد تجنُّبه من الناحية الأخلاقية هو أن يُوجَد شخص يتمتع بالوكالة والقدرة ولكنه لا يتحمل المسئولية. أضاف أرسطو أيضًا شرطًا آخَر فيما يخص المسئولية الأخلاقية: أنت مسئول إذا كنت تعلم ما تفعله. وهذا شرط إدراكي: يجب أن تكون واعيًا بما تفعل وعلى

دراية بعواقبه المحتملة. وما نحتاج إلى تجنُّبه هنا هو شخص تصدُر عنه أفعال لا يدري ماهيتها، وهو ما قد يؤدي في النهاية إلى عواقب وخيمة.

إذا مُنِح الذكاء الاصطناعي وكالةً أكبر وأخذ على عاتِقِه ما كان يتولَّاه البشر في الماضي، فكيف نُسند المسئولية الأخلاقية عن أفعاله؟

الآن دعونا نرى هل تتحقَّق هذه الشروط عند تفويض القرارات والأعمال إلى الذكاء الاصطناعي. المشكلة الأولى هي أن الذكاء الاصطناعي يمكن أن يتَّخذ قراراتٍ ويؤدي أفعالًا لها عواقب أخلاقية، ولكنه لا يُدرك ما يفعله وغير قادر على التفكير الأخلاقي وبالتالي لا يُمكن اعتباره مسئولًا من الناحية الأخلاقية عما يفعله. يمكن أن تتمتَّع الآلات بالوكالة ولكن ليس بالوكالة الأخلاقية؛ لأنها تفتقِر إلى الوعي والإرادة الحرة والعواطف والقُدرة على تكوين النوايا وما شابَهَ ذلك. على سبيل المثال، وفقًا لرؤية أرسطو، يمكن للبشر فقط أداء الأفعال التطوُّعية والتفكير في أفعالهم. إذا كان هذا صحيحًا، فإن الحلَّ الوحيد هو جعْل البشر مسئولين عما تفعله الآلة. ومن ثَم فإن البشر يُفوِّضون الوكالة إلى الآلة، ولكنهم يحتفِظون بالمسئولية. ونحن نفعل ذلك بالفعل في أنظمتنا القانونية؛ إذ إننا لا نَعتبِر الكلاب أو الأطفال الصغار مسئولين عن أفعالهم، ولكننا نضع المسئولية القانونية على عاتق مَن يتولَّون رعايتهم. وفي مؤسسةٍ ما، قد نُفوِّض مهمةً معينة إلى شخصٍ ما ولكننا نُحمِّل المسئولية للمدير المسئول عن المشروع العام، على الرغم من أن الشخص المُفوض في هذه الحالة يتحمَّل جزءًا من المسئولية.[2] إذَن لماذا لا نسمح للآلة بأداء الأعمال ونحتفظ بالمسئولية على الجانب البشري؟ يبدو أن هذه هي أفضل وسيلة نمضي بها قدمًا، حيث إن الخوارزميات والآلات بلا مسئولية.

ومع ذلك، يواجِه هذا الحلُّ عدة مشكلات في حالة الذكاء الاصطناعي. أولًا، يمكن للنظام المزوَّد بالذكاء الاصطناعي أن يتَّخذ قراراته ويؤدي أفعاله بسرعة كبيرة للغاية، على سبيل المثال، في التداول العالي التردُّد أو في السيارات الذاتية القيادة، مما يحرم الإنسان من الوقت الكافي لاتخاذ القرار النهائي أو التدخُّل في الفعل. فكيف يُمكن للبشر أن يتحمَّلوا المسئولية عن مثل هذه الأفعال والقرارات؟ ثانيًا، لأنظمة الذكاء الاصطناعي تواريخ. عندما يقوم الذكاء الاصطناعي بأشياء في سياق تطبيقٍ مُعين، فربما يُصبح من

غير الواضح مَن أنشأه، ومَن استخدمه أولًا، والكيفية التي يجب بها توزيع المسئولية بين هذه الأطراف المختلفة المعنية. على سبيل المثال، في حالة إنشاء خوارزمية ذكاء اصطناعي في سياق مشروعٍ علمي في الجامعة، ثم تطبيق هذه الخوارزمية للمرة الأولى في المُختبر في الجامعة، ثم في قطاع الرعاية الصحية، وفي وقتٍ لاحق في سياقٍ عسكري. فمَن يتحمَّل المسئولية؟ قد يكون من الصعب تتبُّع جميع البشر المتورِّطين في تاريخ هذه الخوارزمية بالذات، بل في التاريخ السببي الذي أدَّى إلى نتيجةٍ معيَّنة تَحمِل إشكالية أخلاقية. فنحن لا نعرف دائمًا جميع الأشخاص المَعنيِّين في اللحظة التي تُثار فيها مشكلة تتعلق بالمسئولية. فخوارزمية الذكاء الاصطناعي غالبًا ما يكون لها تاريخ طويل يشارك فيه العديد من الأشخاص. وهذا يُفضي بنا إلى مشكلة نمطية في إسناد المسئولية عن الأفعال التكنولوجية؛ إذ غالبًا ما يكون هناك الكثير من الأطراف ويُمكنني أن أضيف، الأشياء.

هناك الكثير من الأطراف بمعنى أن الكثير من الأشخاص يشاركون في الفعل التكنولوجي. في حالة الذكاء الاصطناعي، يبدأ الأمر بالمبرمج، ولكن لدينا أيضًا المستخدم النهائي وآخرون. دعونا نفكر مثلًا في السيارة الذاتية القيادة: هناك المبرمج، ومُستخدِمُ السيارة، وأصحابُ شركة السيارات، والمستخدِمون الآخرون للطريق، وهكذا. في مارس ٢٠١٨، تسبَّبَت سيارة ذاتية القيادة لشركة أوبر في حادثٍ في أريزونا أدَّى إلى وفاة أحد المُشاة. فمَن المسئول عن هذه النتيجة المأساوية؟ يمكن أن يكون المسئولون هم مَن برمجوا السيارة، والأشخاص المسئولين عن تطوير المنتج في الشركة، وشركة أوبر نفسها، ومستخدم السيارة، والشخص السائر، والمشرع (على سبيل المثال، ولاية أريزونا)، وهكذا. إذَن فليس من الواضح على مَن تقع المسئولية. قد يكون الأمر هو أن المسئولية لا يمكن ولا يجب إسنادها إلى شخصٍ واحد؛ وربما تقع على أكثر من شخص. ولكن هذا يعني أنه ليس من الواضح كيفية توزيع المسئولية. فقد تقع المسئولية على بعضهم أكثر من الآخرين.

هناك أيضًا الكثير من الأشياء، بمعنى أن النظام التكنولوجي يتألف من العديد من العناصر المتصلة؛ وعادةً ما يكون هناك العديد من المكونات التي تدخل في النظام. هناك خوارزمية الذكاء الاصطناعي، ولكن هذه الخوارزمية تتفاعل مع أجهزة استشعار، وتستخدم جميع أنواع البيانات، وتتفاعل مع جميع أنواع المكونات المادية والبرمجية. كل هذه الأشياء لها تاريخها ومتصلة بالأشخاص الذين برمجوها أو أنتجوها. وعندما يحدُث خطأ، لا يكون واضحًا لنا بالضرورة ما إذا كان «الذكاء الاصطناعي» هو الذي

سبَّب المشكلة أم مُكوِّن آخَر من مكونات النظام؛ بل إننا لا نعرف بالضرورة أين تنتهي مسئولية الذكاء الاصطناعي وتبدأ مسئولية بقية المكونات التكنولوجية. وهذا يجعل من الصعب إسناد المسئولية وتوزيعها. دعونا نفكر أيضًا في تعلم الآلة وعلم البيانات: كما رأينا، ليس هناك فقط خوارزمية، ولكن أيضًا عملية تشمل مراحل مختلفة مثل جمع البيانات ومعالجتها، وتدريب الخوارزمية، وهكذا؛ وجميع هذه المراحل يدخل فيها عناصر تقنية مختلفة وتتطلَّب قرارات بشرية. مرة أخرى، هناك تاريخ سببي يشترك فيه الكثير من البشر والأجزاء، وهذا يجعل إسناد المسئولية أمرًا صعبًا.

لكي نُحاول التعامل مع هذه القضايا، يمكننا أن نتعلم من الأنظمة القانونية أو نلقي نظرة على كيفية عمل التأمين؛ وسوف أتحدَّث عن بعض المفاهيم القانونية في الفصول المتعلقة بالسياسة. ولكن ثمَّة أسئلة أكثر عمومية تلوح لنا من وراء هذه الأنظمة القانونية وأنظمة التأمين حول وكالة الذكاء الاصطناعي والمسئولية عنه: إلى أي مدى نريد أن نعتمد على تقنية الأتمتة، وهل يُمكننا أن نتحمل المسئولية عما يقوم به الذكاء الاصطناعي، وكيف يُمكننا إسناد المسئوليات وتوزيعها؟ على سبيل المثال، مفهوم الإهمال في القانون يتعلَّق بما إذا كان الشخص قد أدَّى ما عليه من واجب العناية. ولكن ماذا يعني هذا الواجب في حالة الذكاء الاصطناعي، خاصةً أنه من الصعب التنبُّؤ بجميع العواقب الأخلاقية المُحتملة؟

وهذا يقودنا إلى القضية التالية. حتى إذا تمَّ حلُّ مشكلة التحكم، فهناك الشرط الثاني للمسئولية الأخلاقية، والذي يتعلَّق بمشكلة المعرفة. لكي تتحمَّل المسئولية، يجب أن تعرف ما تفعله والنتائج المُترتِّبة على فعلك، وفيما بعد، تعرف ما قمتَ به. وبالإضافة إلى ذلك، هذه المسألة لها جانب سردي: في حالة البشر، نتوقَّع أن يتمكن الشخص من شرح ما قام به أو قرَّرَه. المسئولية إذَن تعني القدرة على الرد والتفسير. فإذا حدث خطأ ما، فنحن نريد ردًّا وتفسيرًا. على سبيل المثال، نطلب من القاضي أن يُفسِّر قراره، أو نسأل الجاني لماذا فعل ما فعلَه. وهذه الشروط تُصبح إشكاليةً للغاية في حالة الذكاء الاصطناعي. أولًا، من حيث المبدأ، لا «يعرف» الذكاء الاصطناعي في الوقت الحاضر ما يفعله، بمعنى أنه ليس واعيًا وبالتالي لا يدرك ما يقوم به ولا يدرك نتائج أفعاله. يمكنه تخزين ما يفعله وتسجيله، ولكنه لا «يعرف ما يقوم به» كما يفعل البشر، الذين يُدركون، بوصفهم كائنات واعية، ما يفعلون ويمكنهم — وفقًا لأرسطو مرة أخرى — التفكير والتأمل في أفعالهم وعواقب تلك الأفعال. وعندما لا تُلبَّى هذه الشروط في حالة البشر، على

سبيل المثال، في حالة الأطفال الصغار جدًّا، فإننا لا نحمِّلهم المسئولية. وكذلك عادةً ما لا نُحمِّل الحيوانات المسئولية أيضًا.[3] وإذا لم يُلبِّ الذكاء الاصطناعي هذه الشروط، فإننا لا نستطيع أن نُحمِّله المسئولية. والحل مرة أخرى هو تحميل المسئولية للبشر عن أعمال الذكاء الاصطناعي، على افتراض أنهم يعرفون ما يقوم به الذكاء الاصطناعي وما يفعلونه باستخدام الذكاء الاصطناعي — وبمراعاة الجانب السردي — وأنهم قادرون على الردِّ عن أفعاله ويُمكنهم تفسير ما قام به الذكاء الاصطناعي.

ومع ذلك، فإن مدى صحة هذا الافتراض ليس أمرًا من السهل تقريره كما قد يبدو للوهلة الأولى. عادةً ما يعرف المبرمجون والمستخدمون ما الذي يرغبون في القيام به باستخدام الذكاء الاصطناعي، أو بدقَّة أكبر: يعرفون ما يريدون من الذكاء الاصطناعي أن يفعله لهم. إنهم يعرفون الهدف النهائي؛ ولهذا السبب يفوِّضون المهمة إلى الذكاء الاصطناعي. وقد يكونون أيضًا على دراية بكيفية عمل التكنولوجيا بشكلٍ عام. ولكن، كما سنرى، هم لا يعرفون بدقَّةٍ دائمًا ما يفعله الذكاء الاصطناعي (في أي لحظة) ولا يُمكنهم دائمًا تفسير ما فعله أو كيف وصل إلى قراره.

الشفافية والقابلية للتفسير

نحن نواجه هنا مشكلة الشفافية والقابلية للتفسير. في بعض أنظمة الذكاء الاصطناعي، تكون الطريقة التي يستخدمها الذكاء الاصطناعي لاتخاذ قراره واضحة. على سبيل المثال، إذا كان الذكاء الاصطناعي يستخدم شجرة اتخاذ القرارات، فإن الطريقة التي يصل بها إلى قراره تكون واضحة. فقد تمَّت برمجتُه بطريقة تُحدِّد القرار، بناءً على مدخلات مُعيَّنة. وبالتالي يمكن للبشر تفسير كيف وصل الذكاء الاصطناعي إلى قراره، ويمكن أن «نطلُب» من الذكاء الاصطناعي أن «يُفسر» قراره. بعد ذلك، يمكن للبشر تحمُّل مسئولية القرار أو، على الأحرى، اتخاذ قرار بناءً على التوصية التي قدَّمَها الذكاء الاصطناعي. ومع ذلك، مع بعض أنظمة الذكاء الاصطناعي الأخرى، ولا سيما تلك التي تستخدم تعلُّم الآلة وخاصة التعلُّم العميق الذي يستخدم الشبكات العصبية، لم يعُد من الممكن للإنسان تقديم هذا التفسير أو اتخاذ قراراتٍ من هذا النوع. حيث لم يعُد واضحًا كيف يصل الذكاءُ الاصطناعي إلى قراره، وبالتالي لا يُمكن للبشَر تفسير القرار بشكلٍ كامل. إنهم يعرفون كيف يعمل النظام الخاص بهم، بشكلٍ عام، ولكن لا يُمكنهم تفسير قرارٍ معيَّن. ولنضرب مثلًا بلعبة الشطرنج المزودة بالتعلُّم العميق: يعرف المبرمجون كيف يعمل الذكاء

الاصطناعي، ولكن الطريقة الدقيقة التي يصل من خلالها الجهاز إلى حركةٍ معيَّنة (أي ما يحدث في طبقات الشبكة العصبية) ليست واضحة ولا يمكن تفسيرها. وهذه مشكلة فيما يخصُّ تحمُّل المسئولية، حيث لا يستطيع البشر الذين يُنشئون الذكاء الاصطناعي أو يستخدمونه تفسير قرار معين، وبالتالي يفشلون في معرفة ما يقوم به الذكاء الاصطناعي ولا يُمكنهم تبرير أفعاله. فمِن ناحية، يعرف البشر ما الذي يقوم به الذكاء الاصطناعي (على سبيل المثال، يعرفون الرموز البرمجية الخاصة بالذكاء الاصطناعي ويعرفون كيف يعمل بشكلٍ عام)، ولكن من ناحية أخرى، هم لا يعرفون (لا يُمكنهم تفسير قرار معيَّن)، وتكون نتيجة ذلك أن البشر الذين يتأثرون بالذكاء الاصطناعي لا يمكن إعطاؤهم معلوماتٍ دقيقة حول ما الذي دفع الآلة إلى الوصول إلى هذا التوقُّع. وبالتالي، على الرغم من أن كل تكنولوجيا الأتمتة تُثير مشكلات فيما يتعلق بالمسئولية، فإننا هنا نواجه مشكلةً تخصُّ بعض أنواع الذكاء الاصطناعي؛ وهي ما يطلَق عليها مشكلة الصندوق الأسود.

علاوةً على ذلك، حتى الافتراض بأن البشر في مثل هذه الحالات يتمتعون بمعرفةٍ حول الذكاء الاصطناعي بشكلٍ عام وحول رموزه البرمجية ليس دائمًا صحيحًا. فعلى الأرجح يعرف المبرمجون الأصليون الرموز البرمجية وكيفية عمل كل شيءٍ (أو على الأقل يعرفون الجزء الذي برمجوه)، ولكن ذلك لا يعني أن المبرِمجين والمُستخدمين اللاحقِين الذين يُغيرون الخوارزمية أو يستخدمونها لتطبيقاتٍ محدَّدة يعرفون تمامًا ما يفعله الذكاء الاصطناعي. على سبيل المثال، قد لا يفهم الشخص الذي يستخدم خوارزمية التداول الذكاء الاصطناعي تمام المعرفة، أو قد لا يعرف مُستخدمو وسائل التواصل الاجتماعي حتى أن الذكاء الاصطناعي يُستخدَم، فما بالك بأن يفهموه. ومن جهة المبرِمجين (الأصليين)، فهم قد لا يعرفون على نحوٍ دقيق الاستخدام «المُستقبلي» للخوارزمية التي يُطورونها أو مختلف مجالات التطبيق التي يُمكن استخدامها فيها، فما بالك بكلِّ التبِعات غير المقصودة للاستخدام المُستقبلي لهذه الخوارزمية. لذلك، حتى بغضِّ النظر عن المشكلة الخاصة بتعلُّم الآلة (التعلم العميق)، هناك مشكلة تتعلَّق بالمعرفة لدرجة أنَّ الكثيرين ممَّن يستخدمونه لا يعرفون ما يفعلون؛ لأنهم لا يعرفون ما الذي يفعله الذكاء الاصطناعي، وما هي تأثيراته، أو حتى أنه مُستخدَم من الأساس. وهذه أيضًا مشكلة فيما يخصُّ جانب المسئولية، وبالتالي فهي مشكلة أخلاقية خطيرة.

في بعض الأحيان، يتم تسليط الضوء على هذه المشكلات في سياق الثقة: فغياب الشفافية يؤدي إلى غياب الثقة في التكنولوجيا وفي الأشخاص الذين يستخدمون هذه

التكنولوجيا. لذلك يسأل بعض الباحثين كيف يُمكننا زيادة الثقة في الذكاء الاصطناعي، ويُحدِّدون الشفافية والقابلية للتفسير كعاملٍ من العوامل التي يمكن أن تزيد من الثقة، فضلًا عن تجنُّب التحيُّز (Winikoff 2018) أو صور الذكاء الاصطناعي المُرعبة («ترمينيتور») (Siau and Wang 2018). وكما سنرى في الفصل القادم، غالبًا ما تهدف سياسات الذكاء الاصطناعي أيضًا إلى بناء الثقة. ومع ذلك، فإن مصطلحات مثل الذكاء الاصطناعي «الجدير بالثقة» مُثيرة للجدل؛ إذ تجعلنا نتساءل هل يجِب أن نحتفظ بمصطلح «الثقة» للحديث عن العلاقات الإنسانية، أم يمكن استخدامه للحديث عن الآلات أيضًا؟ تقول جوانا برايسون (٢٠١٨)، الباحثة في مجال الذكاء الاصطناعي، إن الذكاء الاصطناعي ليس شيئًا يمكن الوثوق به ولكنه مجموعة من تقنيات تطوير البرامج؛ ومن ثَم فهي تعتقد أن مصطلح «الثقة» يجب أن يُحتفَظ به للحديث عن البشر ومؤسساتهم الاجتماعية. وعلاوةً على ذلك، يُثير موضوع الشفافية والقابلية للتفسير تساؤلاتٍ حول نوع المجتمع الذي نرغب في العيش فيه. فهنا لا يكمُن الخطر في مجرد تلاعُب الرأسماليين أو النخب التكنوقراطية وهيمنتهم، مما يخلق مجتمعًا يُعاني من الانقسام إلى حدٍّ كبير. وإنما يتمثل الخطر الأكبر وربما الأعمق الذي يَحيق بنا في أن نعيش في مجتمعٍ عالي التقنية، مجتمع لا تعود فيه حتى هذه النخب قادرةً على معرفة ما تفعله، مجتمع لا يستطيع فيه أحدٌ أن يُفسِّر ما يحدث.

كما سنرى، يقترح صانعو السياسات في بعض الأحيان «الذكاء الاصطناعي القابل للتفسير» و«حق التفسير». إلا إننا لا ندري إن كان من المُمكن أن يكون الذكاء الاصطناعي شفافًا طوال الوقت. يبدو هذا سهلَ التحقيق في الأنظمة الكلاسيكية. ولكن إذا بدا مُستحيلًا من حيث المبدأ شرح كلِّ خطوة في عملية اتخاذ القرار وشرح القرارات المُتعلقة بأفراد مُحدَّدين مع تطبيقات تعلم الآلة المُعاصرة، فلدَينا مشكلة إذَن. هل من الممكن «فتح الصندوق الأسود»؟ قد يكون هذا شيئًا جيدًا، ليس فقط للأخلاق ولكن أيضًا لتحسين النظام (أي، النموذج) والتعلُّم منه. على سبيل المثال، إذا كان النظام أكثر قابلية للتفسير، وإذا كان الذكاء الاصطناعي يَستخدِم ما نعتبره سماتٍ غير ملائمة، عندئذٍ يمكن للبشر اكتشاف هذه المشكلات والمساعدة في القضاء على الارتباطات الزائفة. وإذا كان الذكاء الاصطناعي يُحدد استراتيجيات جديدة لمُمارسة لعبةٍ ويجعل هذه الاستراتيجيات أكثر شفافية للبشر، عندئذٍ يمكن للبشر تعلُّمها من الآلة لتحسين أدائهم في اللعبة. وهذا مُفيد ليس فقط في مجال الألعاب، ولكن أيضًا في مجالات مِثل الرعاية الصحية والعدالة الجنائية

والعلوم. لذلك، يُحاول بعض الباحثين تطوير تقنيات لفتح الصندوق الأسود (Samek, Wiegand, and Müller 2017). ولكن إذا لم يكن ذلك مُمكنًا بعدُ أو كان مُمكنًا بدرجة محدودة، فكيف لنا أن نمضي قدمًا؟ هل تتعلَّق المشكلة الأخلاقية هنا بالاختيار بين الأداء وإمكانية التفسير (Seseri 2018)؟ وإذا كانت تكلفة إنشاء نظام ذي أداءٍ جيد هي نقص في الشفافية، فهل يجِب علينا استخدام مثل هذا النظام، أم لا؟ أم يجب أن نُحاول تجنُّب هذه المشكلة والبحث عن حلول تقنية أخرى، بحيث تكون حتى أنظمة الذكاء الاصطناعي الأكثر تقدمًا قادرةً على تبرير أفعالها للبشر؟ هل يُمكننا تدريب الآلات على القيام بذلك؟

علاوةً على ذلك، حتى إذا كانت الشفافية مرغوبة ومُمكنة، فقد يكون من الصعب تحقيقها عمليًّا. على سبيل المثال، ربما لا تكون الشركات الخاصة على استعدادٍ للكشف عن خوارزمياتها؛ لأنها ترغب في حماية مصالحها التجارية. كذلك قد تحُول قوانين الملكية الفكرية التي تحمي تلك المصالح دون ذلك. وكما سنرى في فصول لاحقة، إذا كان الذكاء الاصطناعي في أيدي الشركات القوية، فإن هذا يُثير السؤال حول مَن يصنع قوانين الذكاء الاصطناعي ومَن يجب أن يصنعها.

ومع ذلك، يجب مراعاة أن الشفافية والقابلية للتفسير من الناحية الأخلاقية لا تتعلَّق بالضرورة بالكشف عن الرموز البرمجية، وهي بالتأكيد لا تقتصر على ذلك فحسب. المسألة تتعلَّق أساسًا بتفسير القرارات للبشر. إنها لا تتعلَّق في المقام الأول بتفسير «كيف يعمل» وإنما تتعلَّق بكيف يُمكنني أنا، بوصفي إنسانًا من المتوقَّع منه أن يكون مسئولًا ويتصرَّف بمسئولية، تفسير قراري. ويُمكن أن تكون كيفية عمل الذكاء الاصطناعي، وكيفية وصوله إلى هذه التوصية، جزءًا من ذلك التفسير. علاوةً على ذلك، فإن الكشف عن الرموز البرمجية بمُفردها لا يعطي بالضرورة معرفةً حول كيفية عمل الذكاء الاصطناعي. فهذه المعرفة تعتمِد على الخلفية التعليمية للإنسان ومهاراته. فإذا كان يفتقر إلى الخبرة التقنية ذات الصلة، فإننا نحتاج إلى نوعٍ آخَر من التفسير. وهذا لا يُذكِّرنا فحسب بمشكلة التعليم ولكنه يؤدِّي بنا إلى سؤالٍ حول نوع التفسير الذي نحتاجُه، ثم ماهية التفسير في حدِّ ذاته.

وهكذا تَطرَح قضية الشفافية والقابلية للتفسير أيضًا أسئلة فلسفية وعلمية مُثيرة للاهتمام، مثل الأسئلة المتعلقة بطبيعة التفسير (Weld and Bansal 2018). ومما يتألف التفسير الجيد؟ وما الفرق بين التفسيرات والأسباب، وهل يمكن للآلات تقديم أي منها؟ وكيف يتَّخذ البشر القرارات في الواقع؟ وكيف يُبرِّرون قراراتهم؟ هناك أبحاث حول هذا

الموضوع في علم النفس المعرفي والعلوم المعرفية، والتي يُمكن استخدامها للتفكير في الذكاء الاصطناعي القابل للتفسير. على سبيل المثال، لا يُقدم الناس عمومًا سلاسل سببيةً كاملة؛ وإنما يختارون تفسيراتٍ ويجيبون عما يعتقدون أنها مُعتقدات الشخص الذي يُفسِّر لهم: التفسيرات اجتماعية (Miller 2018). وربما نتوقع أيضًا أن تكون تفسيرات الآلات مختلفة عن تفسيرات البشر، الذين يُبررون أفعالهم في كثيرٍ من الأحيان بأنها نتيجة للعواطف. ولكن إذا فعلنا ذلك، فهل يعني هذا أننا نعتبر طريقة اتخاذ الآلات للقرارات أفضل من طريقة اتخاذ البشر لها (Dignum et al. 2018)، وإذا كان الأمر كذلك، فهل يجب أن نفعل؟ يتحدث بعض الباحثين عن الاستدلال بدلًا من التفسير. بل إن وينيكوف (٢٠١٨) يطلب «الاستدلال بناءً على القِيَم» من الذكاء الاصطناعي وغيره من الأنظمة المُستقلة، التي يجب أن تكون قادرةً على تمثيل القيم البشرية والاستدلال باستخدام تلك القِيَم. ولكن هل يمكن للآلة أن تقوم بالاستدلال، وكيف يمكن للنظام التكنولوجي «استخدام» القِيَم أو «تمثيلها» من الأساس؟ أي نوع من المعرفة يمتلكها هذا النظام؟ وهل يمتلك معرفة من الأساس؟ وهل يستطيع الفهم من الأساس؟ وكما يسأل بودينجتون (٢٠١٧)، هل يمكن للبشر أن يُعبِّروا بشكلٍ كامل عن قيَمِهم الجوهرية؟

مثل هذه المشكلات مُثيرة للاهتمام من منظور الفلاسفة، ولكنها أيضًا ذات صلةٍ مباشرة بالأخلاقيات، كما أنها واقعية وعملية للغاية. وكما يقول كاستيلفيتشي (٢٠١٦): إن فتح «الصندوق الأسود» مشكلة في العالَم الحقيقي. على سبيل المثال، يجب على البنوك أن تُفسِّر سبب رفض قرضٍ ما؛ ويجب على القضاة تفسير سبب إصدار الأوامر بحبس شخصٍ ما (مرةً أخرى). إن تفسير القرارات ليس فقط جزءًا من طبيعة البشر عندما يتواصلون (Goebel et al. 2018)، بل هو أيضًا مطلب أخلاقي. إن القدرة على التفسير شرط ضروري للسلوك واتخاذ القرارات بشكل مسئول وقابل للمساءلة. ويبدو أنه ضروري لأي مجتمع يرغب في احترام البشر بوصفهم أفرادًا مُستقلين اجتماعيِّين يُحاولون التصرف واتخاذ القرارات بشكلٍ مسئول وفي الوقت نفسه يُطالبون، عن استحقاق، بالحصول على أسبابٍ للقرارات التي تؤثر عليهم وتفسيرات لها. وسواءٌ أكان بإمكان الذكاء الاصطناعي توفير تلك الأسباب والتفسيرات «مباشرةً» أم لا، فإن البشر لا بد أن يكونوا قادرين على الإجابة عند سؤالهم عن الأسباب. إن التحدي الذي يواجِه الباحثين في مجال الذكاء الاصطناعي هو ضمان أنه في حال استخدام الذكاء الاصطناعي لأغراض اتخاذ القرارات من الأساس، فيجب تصميم التكنولوجيا بحيث يتمكن البشر قدْر الإمكان من الإجابة عند سؤالهم عن أسباب اتخاذ تلك القرارات.

الفصل التاسع

التحيز ومعنى الحياة

التحيز

يُعد التحيُّز مشكلةً أخرى من المشكلات ذات الجوانب الأخلاقية والاجتماعية في الوقت نفسه، وهي أيضًا تتعلَّق بالذكاء الاصطناعي القائم على عِلم البيانات بعيدًا عن غيره من تقنيات الأتمتة الأخرى. عندما يتَّخذ الذكاء الاصطناعي — أو على الأحرى، عندما يُوصي باتخاذ — قرارات، قد يُظهر التحيُّز؛ إذ قد تكون القرارات غير منصفةٍ أو غير عادلة تجاه أفرادٍ أو مجموعات بعينها. وعلى الرغم من أن التحيُّز قد يظهر أيضًا عند استخدام الذكاء الاصطناعي التقليدي — على سبيل المثال، نظام خبير يَستخدم شجرة اتخاذ قرارات أو قاعدة بيانات تتسم بالتحيز — فإن قضية التحيز غالبًا ما تكون مرتبطةً بتطبيقات تعلُّم الآلة. وبينما كانت مشكلات التحيز والتمييز موجودة دائمًا في المجتمع، إلا أن القلق يكمُن في أن يؤدي الذكاء الاصطناعي إلى استمرار هذه المشكلات وتفاقُم آثارها.

بينما كانت مشكلات التحيُّز والتمييز موجودةً دائمًا في المجتمع، إلا أن القلق يكمُن في أن يؤدي الذكاء الاصطناعي إلى استمرار هذه المشكلات وتفاقُم آثارها.

غالبًا ما يكون التحيُّز غير مقصود؛ فالمُطوِّرون والمستخدِمون، وغيرهم من أطراف مُعنية مثل إدارة الشركة، لا يتوقعون، في كثيرٍ من الأحيان، آثار التمييز ضدَّ مجموعات أو أفرادٍ مُعينين. ويمكن أن يكون السبب في ذلك هو عدم فهمهم نظام الذكاء الاصطناعي كما ينبغي، أو عدم وعيِهم بشكلٍ كافٍ بمشكلة التحيُّز أو حتى بتحيُّزاتهم الشخصية،

أو بشكلٍ عام عدم تصوُّرهم وعدم تفكيرهم بما فيه الكفاية في العواقب المُحتمَلة غير المقصودة للتكنولوجيا وعدم تواصُلهم مع بعض الأطراف ذات الصلة. يُعَد هذا أمرًا إشكاليًّا نظرًا إلى أن القرارات المُتحيِّزة يمكن أن تكون لها عواقب وخيمة، على سبيل المثال، من حيث الوصول إلى الموارد والتمتُّع بالحريات (CDT 2018)؛ إذ قد لا يحصل الأفراد على وظيفة، أو لا يتمكَّنون من الحصول على ائتمان، أو قد ينتهي بهم الحال في السجن، أو حتى يتعرَّضون للعنف. والمُعاناة لا تقتصِر على الأفراد فحسب؛ إذ قد تتأثر مجتمعات بأسرِها بالقرارات المُتحيزة، على سبيل المثال، عندما تُصنَّف منطقة كاملة في المدينة أو جميع الأشخاص ممَّن لهم خلفية عرقية مُعيَّنة بواسطة الذكاء الاصطناعي على أنهم يشكلون خطورةً أمنية عالية.

ولنعُد مرَّة أخرى إلى مثال خوارزمية كومباس الذي تحدَّثنا عنه في الفصل الأول، تلك الخوارزمية التي تتنبَّأ بمدى احتمالية أن يقوم المُدَّعى عليه بإعادة ارتكاب الجريمة وكان القضاة في فلوريدا يستخدمونها في اتخاذ قراراتهم بشأن إمكانية منح السجين إفراجًا مشروطًا. وفقًا لدراسةٍ أجرَتْها «بروبابليكا»، وهي غرفة إخبارية عبر الإنترنت، كانت النتائج الإيجابية الكاذبة للخوارزمية (المُدَّعى عليهم الذين توقعت الخوارزمية أن يُعيدوا ارتكاب الجرائم ولكنهم في الواقع لم يفعلوا) تميل بشكلٍ مُفرِط إلى الأشخاص من ذوي البشرة السمراء، وكانت النتائج السلبية الكاذبة (المُدَّعى عليهم الذين توقعت الخوارزمية ألا يُعيدوا ارتكاب الجرائم ولكنهم في الواقع فعلوا) تميل بشكلٍ مُفرط إلى الأشخاص ذوي البشرة البيضاء (Fry 2018). ومِن ثَم رأى النقَّاد أن هناك تحيُّزًا ضد المُدَّعى عليهم من ذوي البشرة السمراء. مثال آخَر على ذلك هو أداة «بريدبول»، وهي أداة للتنبُّؤ بالجرائم وقد استُخدِمَت في الولايات المتحدة لتوقُّع احتمالية حدوث جريمةٍ في مناطق مُعيَّنة من المدن وللتوصية بتخصيص موارد الشرطة (على سبيل المثال، أين يجب أن يُجري ضباط الشرطة عمليات التفتيش والتجوال) استنادًا إلى هذه التوقُّعات. وتركزت المخاوف في هذا الصدد في أن يكون النظام مُتحيزًا ضد الأحياء الفقيرة وأحياء المُلوَّنين أو أن تؤدي المراقبة الأمنية المُفرطة إلى كسر الثقة بين الناس في تلك المناطق، مما يُحَوِّل توقع حدوث الجريمة إلى نبوءةٍ تتحقَّق ذاتيًّا (Kelleher and Tierney 2018). ولكن التحيُّز لا يقتصِر على العدالة الجنائية أو المُراقبة الأمنية؛ بل يُمكن أن يعني أيضًا، على سبيل المثال، تعرُّض مُستخدمي خدمات الإنترنت لتحيزاتٍ ضدَّهم إذا صنَّفهم الذكاء الاصطناعي تصنيفًا سيئًا.

قد ينشأ التحيُّز بعدَّة طرُقٍ في جميع مراحل التصميم والاختبار والتطبيق. وإذا ما ركَّزنا على مرحلة التصميم، فسنجد أن التَّحيُّز قد يظهر في اختيار مجموعة البيانات التي سيتم التدريب عليها؛ وفي مجموعة البيانات التي سيتم التدريب عليها نفسها، والتي قد تكون غير مُمثلة أو غير كاملة، وفي الخوارزمية، وفي مجموعة البيانات التي يتم إدخالها إلى الخوارزمية بعد تدريبها، وفي القرارات القائمة على الارتباطات الزائفة (انظر الفصل السابق)، وفي المجموعة التي تُنشئ الخوارزمية، وفي المجتمع الأوسَع. على سبيل المثال، قد لا تكون مجموعة البيانات مُمثِّلة للسكان (كأن تكون مَبنية على رجالٍ أمريكيين بيض) ولكنها تُستخدَم للتنبؤ مع السكان ككلٍّ (الرجال والنساء من خلفيات عرقية مُتنوِّعة). يمكن أيضًا أن يكون التحيُّز مُتعلِّقًا بالاختلافات بين البلدان. فكثير من الشبكات العصبية العميقة المُستخدمة في التعرُّف على الصور تُدرَّب على مجموعة البيانات المُحَدَّدة «إيمدجنت» ImageNet، التي تحتوي على كميةٍ غير متكافئة من البيانات من الولايات المتحدة، في حين أن بلدان مثل الصين والهند، اللَّتين تُمثلان جزءًا أكبر بكثيرٍ من سكان العالم، تُسهمان بنسبة صغيرة فقط (Zou and Schiebinger 2018). وهذا قد يؤدي إلى تحيُّز مجموعة البيانات ثقافيًّا. وبشكلٍ عام، يمكن أن تكون مجموعات البيانات غيرَ كاملةٍ أو ذات جودة رديئة، مما قد يؤدي إلى وجود تحيُّز. كذلك قد يكون التنبُّؤ مَبنيًّا على قدْرٍ ضئيل من البيانات، على سبيل المثال في حالة التنبُّؤ بجرائم القتل: حيث لا يُوجَد هذا الكم الكبير من جرائم القتل، مما يجعل التعميم أمرًا إشكاليًّا. كمثالٍ آخر، يشعر بعض الباحِثين بالقلق إزاء نقص التنوُّع في فِرق تطوير الذكاء الاصطناعي وعلم البيانات؛ حيث يكون معظم علماء الكمبيوتر ومهندسي الكمبيوتر رجالًا بِيضًا من البلدان الغربية تتراوَح أعمارهم ما بين ٢٠ عامًا و٤٠ عامًا، وقد تنعكس تجاربهم الشخصية وآراؤهم، وبالتأكيد تحيُّزاتهم في العملية، وهو ما قد يؤثر سلبًا على الأشخاص الذين لا تنطبق عليهم هذه الأوصاف، مثل النساء، والأشخاص ذوي الإعاقة، وكبار السن، والأشخاص المُلونين، والأشخاص من البلدان النامية.

قد تكون البيانات مُتحيزة أيضًا ضد مجموعاتٍ مُعينة؛ لأن هناك تَحيُّزًا في مُمارسةٍ معينة بشكلٍ خاص أو في المجتمع بشكلٍ عام. على سبيل المثال، ثمة ادعاءات بأن مجال الطب يَستخدِم بشكلٍ رئيسي بياناتٍ من المرضى الذكور، وبالتالي فإنه مُتحيِّز، كذلك هناك التحيُّز ضد الأشخاص المُلوَّنين وهو يُعتبر سائدًا في المجتمع بشكلٍ أوسع. إذا كانت الخوارزمية تستخدِم مثل هذه البيانات، فإن النتائج ستكون أيضًا مُتحيزة. وكما ورد

في مقال مجلة «نيتشر» الافتتاحي عام ٢٠١٦: التحيُّز في المدخلات يؤدي إلى تحيُّز في المُخرجات. وقد تبين أيضًا أن تعلُّم الآلة يمكن أن يكتسِب سماتِ التحيُّز من خلال استخدام البيانات النصيَّة من شبكة الويب العالمية، حيث تعكس هذه البيانات اللغوية الثقافة الإنسانية اليومية، بما فيها من تحيزات (Caliskan, Bryson, and Narayanan 2017). على سبيل المثال، قد تحتوي متون اللغات نفسها على تحيزات جنسية. والمُثير للقلق في هذه الحالة أن الذكاء الاصطناعي ربما يساعد في استمرار هذه التحيُّزات، مما يضرُّ بشكلٍ أكبر الجماعات التي كانت تُعاني من التهميش دائمًا. يمكن أيضًا أن يظهر التحيز إذا كان هناك ارتباط ولكن لا يوجد سبب. على سبيل المثال، في مجال العدالة الجنائية مرة أخرى: قد تستنتج الخوارزمية أنه إذا كان أحد والدَي المُدَّعى عليه قد أُودِع السجن، فإن هذا المُدَّعى عليه من المُرجَّح أن يُودَع السجن أيضًا. حتى لو كان هذا الارتباط قائمًا وحتى لو كان الاستنتاج تنبؤيًّا، يبدو أنه من غير العدل أن يحصل هذا المُدَّعى عليه على عقوبة أشد؛ نظرًا إلى عدم وجود علاقة سببية (House of Commons 2018). وأخيرًا، يمكن أيضًا أن ينشأ التحيُّز بسبب أن صانعي القرارات من البشر يثقون في دقة توصيات الخوارزميات أكثر مما ينبغي (CDT 2018) ويتجاهلون المعلومات الأخرى أو لا يعتمِدون على حُكمهم الشخصي بما فيه الكفاية. على سبيل المثال، قد يعتمد القاضي اعتمادًا كليًّا على الخوارزمية ولا يأخذ في اعتباره العناصر الأخرى. وكما هو الحال دائمًا مع الذكاء الاصطناعي وغيره من تقنيات الأتمتة، تلعب القرارات والتفسيرات البشرية دورًا مهمًّا، وهناك دائمًا خطر الاعتماد الزائد على التكنولوجيا.

ومع ذلك، ليس مِن الواضح ما إذا كان من المُمكن تجنُّب التحيُّز من الأساس، أو حتى ما إذا كان يجب تجنُّبه، فإذا كان من الواجب تجنُّبه، فما التكلفة التي يمكن تحمُّلها في سبيل ذلك. على سبيل المثال، إذا كان تغيير خوارزمية تعلم الآلة لتقليل احتمالات التحيُّز سيكون على حساب جعل توقعاتها أقل دقَّة، فهل يجب علينا تغييرها؟ قد نُضطر إلى الاختيار ما بين فعالية الخوارزمية من ناحية ومكافحة التحيُّز من ناحيةٍ أخرى. هناك أيضًا مشكلة في أنه إذا تمَّ تجاهل سمات مُعينة أو تجاهلها مثل العرق، فإن أنظمة تعلُّم الآلة قد تُحدد ما يعرف بمؤشرات هذه السمات، مما يؤدي أيضًا إلى التحيُّز. على سبيل المثال، في حالة العِرق، قد يكون من المُمكن أن تختار الخوارزمية مُتغيرات أخرى مرتبطة بالعِرق مثل الرمز البريدي. وهل من الممكن وجود خوارزمية خالية تمامًا من التحيز؟ لا يُوجَد توافق بين الفلاسفة أو حتى في المجتمع بشأن العدالة الكاملة أو الإنصاف الكامل.

علاوةً على ذلك، وكما أشرْنا في الفصل السابق، فإن مجموعات البيانات المُستخدمة من قِبل الخوارزميات هي تجريدات عن الواقع وهي نتاج اختيارات بشرية، ومن ثَمَّ فهي لا تكون مُحايدة أبدًا (Kelleher and Tierney 2018). يتوغَّل التحيز في عالَمنا ومُجتمعاتنا؛ وبالتالي، على الرغم من أنه يمكن القيام بالكثير ويجب القيام بالكثير لتقليل التحيز، فإن نماذج الذكاء الاصطناعي لن تخلوَ تمامًا من التحيز (Digital Europe 2018).

علاوةً على ذلك، يبدو بالتأكيد أن الخوارزميات المُستخدمة في اتخاذ القرار دائمًا ما تكون مُتحيزة من منطلق كونها تمييزية؛ إذ إنها مُصممة للتمييز بين مُختلف الاحتمالات. على سبيل المثال، في عملية التوظيف، يُفترَض أن يكون فحص السِّيَر الذاتية ذا طابع مُتحيز وتمييزي تجاه سِمات المرشحين التي تُناسب الوظيفة. ويكمُن السؤال الأخلاقي والسياسي فيما إذا كان هناك تمييز مُعين غير منصف وغير عادل. ولكن مرة أخرى، تختلف وجهات النظر بشأن ما هو منصف وما هو عادل. وهذا يجعل قضية التحيُّز ليست فقط تقنيةً ولكنها أيضًا مرتبطة بالمناقشات السياسية حول الإنصاف والعدالة. على سبيل المثال، هل من العدل مُمارسة التمييز الإيجابي أو التدابير الإيجابية، التي تُحاول محو أثر التحيُّز عن طريق التحيز الإيجابي مع الأفراد أو الجماعات المحرومة؟ هل يجِب أن تكون العدالة عمياء ومحايدة — وبالتالي هل يجب أن تكون الخوارزميات عمياء إزاء العِرق، على سبيل المثال — أم أن العدالة تعني تمييز أولئك المَحرومين بالفعل من أي ميزات، مما يصِل بنا في النهاية إلى نوع من التحيُّز والتمييز (التصحيحي)؟ وهل يجب على السياسة في السياق الديمقراطي أن تُعطي الأولوية لحماية مصالح الأغلبية أم تركز على تعزيز مصالح الأقلية، حتى وإن كانت أقلية محرومة قديمًا أو حاليًّا؟

> هل يجِب أن تكون العدالة عمياء ومُحايدة أم أن العدالة تعني تمييز أولئك المحرومين بالفعل من أي ميزات؟

وهذا يقودنا إلى السؤال حول الإجراءات. حتى إذا اتفقنا على وجود تحيُّز، فهناك طرق مختلفة للتعامُل مع المشكلة. وتشمل هذه الطرق التكنولوجية وكذلك الإجراءات المُجتمعية والسياسية والتعليم. وثمة خلاف حول الإجراءات التي يجِب علينا اتخاذها؛ إذ إنها تعتمد مرة أخرى على مفهومنا للعدالة والإنصاف. على سبيل المثال، تُثير قضية التدابير الإيجابية قضيةً أكثر عمومية حول ما إذا كنَّا يجِب أن نقبل العالم كما هو أم

أننا يجب أن نُشكِّل عالمنا المُستقبلي على نحوٍ فعَّال بطريقةٍ من شأنها تجنُّب استمرار الظلم الذي كان مُستشريًا في الماضي. بعض الناس يرَون أننا يجِب أن نستخدِم مجموعة بياناتٍ تعكس العالم الواقعي. وقد تُمثل البيانات التحيزات الموجودة في المجتمع وقد تُنشئ الخوارزمية نموذجًا من التحيُّزات الموجودة لدى الناس الآن، ولكن هذه ليست مشكلةً يجب أن يقلق بشأنها المُطورون. بينما يرى آخرون أن مثل هذه المجموعة من البيانات موجودة فقط بسبب قرونٍ من التحيز، وأن هذا التحيُّز والتمييز غير عادل وظالِم، وعليه فإنه يجِب تغيير تلك المجموعة من البيانات أو الخوارزمية من أجل تعزيز التدابير الإيجابية. على سبيل المثال، في استجابةٍ إلى نتائج خوارزمية البحث في جوجل التي تبدو مُتحيزةً ضد أساتذة الرياضيات الإناث، يمكن للمرء أن يقول إن هذا يعكس ببساطةٍ حقيقة العالم (وأن هذا هو بالضبط ما يجب أن تفعله خوارزمية البحث)؛ أو يمكن أن نجعل الخوارزمية تُعطي أولويةً لصور أساتذة الرياضيات الإناث من أجل تغيير التصور وربما تغيير العالَم (Fry 2018). ويمكن أيضًا أن نُحاول إنشاء فِرَق تطوير تكون أكثر تنوعًا من حيث الخلفية والرأي والتجربة، وتُمثل بشكلٍ أفضل الفئات التي مِن المحتمَل أن تتأثر بالخوارزمية (House of Commons 2018).

لن يصح الرأي القائل بأنها تعكس الواقع إذا كانت مجموعة البيانات التي سيتم التدريب عليها لا تعكس العالم الواقعي وتحتوي على بياناتٍ قديمة لا تعكس الوضع الحالي. كما أن القرارات المبنية على هذه البيانات تساعد بالفعل في استمرار التمييز الذي كان موجودًا في الماضي بدلًا من الاستعداد للمستقبل. وعلاوةً على ذلك، ثمَّة اعتراض آخَر على الرأي القائل بأنها تعكس الواقع وهو أنه حتى إذا كان النموذج يعكس العالم الواقعي، فإن هذا يمكن أن يؤدي إلى تدابير تمييزية وأضرارٍ أخرى قد تقع على أفرادٍ أو مجموعات بعينها. على سبيل المثال، قد ترفض شركات الائتمان منحَ قروضٍ إلى المُتقدمين على أساس محلِّ الإقامة، أو قد تفرض المواقع الإلكترونية رسومًا أكبر على بعض العملاء مقارنةً بغيرهم استنادًا إلى ملفات العملاء التعريفية التي أنشأها الذكاء الاصطناعي. كذلك يمكن أن تتبع الملفات التعريفية الأفراد عبر النطاقات المختلفة (Kelleher and Tierney 2018). ويمكن أن تربط وظيفة الإكمال التلقائي البسيطة في ظاهرها بِشكلٍ خطأً اسمَك بجريمةٍ ما (الأمر الذي قد يؤدِّي إلى عواقب وخيمة)، حتى إذا كانت خوارزمية البحث الكامنة وراءها تعكس العالَم بشكلٍ صحيح؛ بمعنى أن معظم الناس يُريدون البحث عن اسم المجرم وليس عن اسمك. وثمَّة مثال آخر على التحيُّز، ولكنه ربما ليس

واضحًا بالقدْر نفسه: فنظام استرجاع الموسيقى المُستخدَم في خدمات مثل «سبوتيفاي»، الذي يقدِّم توصياتٍ بناءً على السلوك الحالي (المسارات الموسيقية التي ينقر عليها معظم الناس)، قد يتحيَّز ضد الموسيقى والموسيقيين الذين هم أقلُّ شيوعًا. وحتى إذا كان النظام يعكس العالَمَ الواقعي، فإن هذا يؤدي إلى وضعٍ لا يستطيع فيه بعض الموسيقيين العيشَ من موسيقاهم ويجعل بعض المجتمعات تشعُر بعدم التقدير وعدم الاحترام.

مرة أخرى، في حينِ أن هذه حالات واضحة من التمييز الذي ينطوي على مشكلات، إلا أننا يجِب أن نسأل دائمًا: هل يمكن أن يكون التمييز في حالةٍ مُعينة عادلًا أم لا؟ وإذا كان غير عادل، فما الإجراء الذي سيُتَّخَذ حياله ومَن الذي سيتخِذُه؟ على سبيل المثال، ما الذي يُمكن أن يفعله علماء الكمبيوتر حياله؟ هل يجب أن يجعلوا مجموعات البيانات التي يتم التدريب عليها أكثر تنوعًا، وربما يُنشئون بياناتٍ ومجموعات بيانات «مثالية» كما اقترح إريك هورفيتز من شركة مايكروسوفت (Surur 2017)؟ أم يجب أن تعكس مجموعات البيانات العالَم؟ هل يجب على المطوِّرين تضمين التمييز الإيجابي في خوارزمياتهم، أم يجب عليهم إنشاء خوارزميات «عمياء»؟ إن كيفية التعامُل مع التحيز في الذكاء الاصطناعي ليست مسألة تقنية فحسب؛ بل هي مسألة سياسية وفلسفية. إن المسألة تتعلَّق بنوع المجتمع والعالَم الذي نريده، وإذا كان من الواجِب علينا أن نُحاول تغييره، وإذا كان الأمر كذلك، فما هي الطرق المقبولة والعادلة لتغييره. إنها أيضًا مسألة تتعلَّق بالبشر بقدْر ما تتعلق بالآلات: هل نعتقد أن اتخاذ القرارات البشرية عادل ومنصف، وإذا لم يكن الأمر كذلك، فما دور الذكاء الاصطناعي؟ ربما يُمكن أن يُعلِّمنا الذكاء الاصطناعي شيئًا عن البشر ومجتمعاتهم من خلال الكشف عن تَحيُّزاتنا. وقد تكشف مناقشة أخلاقيات الذكاء الاصطناعي الاختلال الكبير في موازين القوى الاجتماعية والمؤسسية.

وهكذا تصِل المناقشات حول أخلاقيات الذكاء الاصطناعي إلى عُمق قضايا مجتمعية وسياسية حسَّاسة ترتبط بأسئلةٍ فلسفية حول العدالة والإنصاف، وأسئلة فلسفية وعِلمية حول البشر ومجتمعاتهم. واحدة من هذه القضايا هي مُستقبل العمل.

مستقبل العمل ومعنى الحياة

مِن المتوقَّع أن تُحوِّل الأتمتة التي تعتمد على الذكاء الاصطناعي اقتصاداتنا ومجتمعاتنا بشكلٍ جذري، مما يُثير تساؤلاتٍ حول مُستقبل العمل ومعناه، فضلًا عن مستقبل الحياة البشرية ومعناها.

أولًا، هناك مَخاوف من أن يؤدي الذكاء الاصطناعي إلى تدمير الوظائف، الأمر الذي قد يؤدي إلى البطالة الشاملة. وهناك أيضًا سؤال حول نوع الوظائف التي يستطيع الذكاء الاصطناعي تولِّيها: وهل ستقتصِر على وظائف ذوي الياقات الزرقاء (العمالة اليدوية)، كما يُطلَق عليها، أم أن هناك وظائف أخرى يمكن أن يتولَّاها؟ يتنبَّأ تقرير شهير لكلٍّ من بنيديكت فري ومايكل أوزبورن (٢٠١٣) بأن ٤٧ في المائة من جميع الوظائف في الولايات المتحدة يُمكن أتمتَتُها. وتحمل تقارير أخرى أرقامًا أقلَّ إثارةً للجدل، ولكن معظمها يتنبأ بأن فقدان الوظائف سيكون كبيرًا. ويتفق العديد من الكتَّاب على أن الاقتصاد قد تأثر وسيظلُّ يتأثر بشكلٍ كبير (Brynjolfsson and McAffee 2014)، بما في ذلك التغيرات الملحوظة التي حدثت في التوظيف الآن والتي ستحدث في المستقبل. ومن المُتوقَّع أن يؤدي فقدان الوظائف بسبب الذكاء الاصطناعي إلى التأثير على جميع أنواع العامِلين، ليس فقط ذوي الياقات الزرقاء، حيث أصبح الذكاء الاصطناعي قادرًا بشكلٍ متزايد على أداء المهام المَعرفية المُعقدة. إذا كان هذا صحيحًا، فكيف يُمكننا أن نُعِدَّ الأجيال الجديدة لهذا المستقبل؟ ماذا يجب أن يتعلَّموا؟ وماذا يجب أن يفعلوا؟ وماذا لو كان الذكاء الاصطناعي يُفيد بعض الأشخاص أكثر من غيرهم؟

بهذا السؤال الأخير، نعود مرةً أخرى إلى قضايا العدالة والإنصاف، التي شغلت تفكير الفلاسفة السياسيين لعصور. على سبيل المثال، إذا كان الذكاء الاصطناعي سيوسِّع الفجوة بين الأثرياء والفقراء، فهل هذا أمر عادل؟ وإذا لم يكن عادلًا، فما الذي يمكن القيام به حيال ذلك؟ يمكن أيضًا صياغة المشكلة من حيث عدم المساواة (هل سيزيد الذكاء الاصطناعي من عدم المساواة في المجتمعات وفي العالم؟) أو من حيث التعرُّض إلى التأثيرات السلبية: هل سيحظى أصحاب الوظائف والأثرياء والمُتعلَّمون في الدول المتقدمة تكنولوجيًّا بفوائد الذكاء الاصطناعي بينما سيكونُ العاطلون عن العمل والفقراء والأقل تعليمًا في الدول النامية أكثر عرضةً لتأثيراته السلبية (Jansen et al. 2018)؟ وللتعامُل مع قضية أخلاقية وسياسية أخرى أكثر حداثة: ماذا عن العدالة البيئية؟ ما هو تأثير الذكاء الاصطناعي على البيئة وعلاقتنا بالبيئة؟ ماذا يعني «الذكاء الاصطناعي المُستدام»؟ هناك أيضًا سؤال حول ما إذا كانت أخلاقيات الذكاء الاصطناعي وسياساته مُرتبطة بقِيَم البشر ومصالحهم فقط أم لا. (انظر الفصل الثاني عشر.)

من المتوقَّع أن تُحوِّل الأتمتة التي تعتمد على الذكاء الاصطناعي اقتصاداتنا ومُجتمعاتنا بشكلٍ جذري، مما يُثير أسئلة حول مُستقبل العمل ومعناه، فضلًا عن مُستقبل الحياة البشرية ومعناها.

ثمة سؤال آخَر ذو طابع وجودي يتعلَّق بمعنى العمل والحياة البشرية. تفترِض المخاوف من فقدان الوظائف أن العمل هو القيمة الوحيدة والمصدر الوحيد للدخل والمعنى. ولكن إذا كانت الوظائف هي الشيء الوحيد ذو القيمة، فربما علينا عندئذٍ خلق المزيد من الأمراض العقلية، ورفع مُعدل التدخين، وزيادة معدلات السمنة؛ لأن هذه المشكلات هي التي تؤدي إلى خلق وظائف.[1] ونحن لا نُريد ذلك. إذَن فمن الواضح أننا نؤمن بأن هناك قيمًا أخرى أهم من خلق الوظائف في حدِّ ذاته. ولماذا نعتمد على الوظائف لتحقيق الدخل والمعنى؟ يُمكننا تنظيم مجتمعاتنا واقتصاداتنا بطريقةٍ مختلفة. يُمكننا أن نفصل بين العمل والدخل، أو بالأحرى ما نعتبره «عملًا» ودخلًا. فهناك الكثيرون يقومون بالعمل مجَّانًا، على سبيل المثال في المنزل ورعاية الأطفال والمُسنِّين. فلماذا لا يُعتبر هذا «عملًا»؟ ولماذا يكون القيام بذلك النوع من العمل أقلَّ قيمةً وأهمية من غيره من الأعمال؟ ولماذا لا نجعله مصدرًا للدخل؟ علاوةً على ذلك، يعتقد بعض الأشخاص أن الأتمتة يُمكن أن تُتيح لنا المزيد من الرفاهية والراحة. ربما يُمكننا القيام بأشياء أكثر متعةً وإبداعًا، ليس بالضرورة في شكل وظيفة. يُمكننا، بعبارةٍ أخرى، الاعتراض على فكرة أن الحياة ذات المعنى هي فقط حياة تُقضَى في أداء عملٍ مدفوع الأجر ومُنظم مُسبقًا من قِبل الآخرين أو عمل يتم في إطار ما يُطلَق عليه «التوظيف الذاتي». ربما يُمكننا فرض تدابير مُعيَّنة مثل تحديد «دخل أساسي» لنسمح للجميع بفعل ما يرَونه ذا معنًى وقيمة. وبالتالي، ردًّا على مشكلة مُستقبل العمل، يُمكننا أن نُفكر فيما يجعل العمل ذا معنًى، وفي نوع العمل الذي ينبغي للبشر عملُه (أو بالأحرى يُسمَح لهم بعمله)، وفي كيفية إعادة تنظيم مجتمعاتنا واقتصاداتنا بحيث لا يرتبط الدخل بالوظائف والتوظيف.

على الرغم من كلِّ ما قيل، فإن الأفكار اليوتوبية حول المجتمعات المُرفَّهة وغيرها من الجنان ما بعد الصناعية لم تتحقَّق حتى الآن. لقد شهدنا بالفعل عدة موجاتٍ من الأتمتة بدءًا من القرن التاسع عشر حتى الآن، ولكن إلى أي مدًى حرَّرَتنا الآلات وأعتقت رقابنا؟ ربما تولَّت نيابةً عنا بعضَ الأعمال المُضجرة والخطيرة، ولكنها استُخدِمت أيضًا للاستغلال ولم تُغيِّر بشكلٍ جذري الهيكل الهرَمي للمجتمع. وقد استفاد بعض الناس

استفادةً هائلة من الأتمتة، بينما لم يفعل آخرون. وربما تكون الأوهام حول عدم وجود وظائف هي رفاهية محفوظة فقط لأولئك الذين كانوا في جانب المُستفيدين. فضلًا عن ذلك، هل حرَّرَتنا الآلات لنعيش حياةً ذات معنًى أكثر من ذي قبل؟ أم أنها تُهدِّد إمكانية هذه الحياة نفسها؟ هذا نقاش طويل ولا تُوجَد إجابات سهلة عن هذه الأسئلة، ولكن المخاوف التي لدينا تُعد أسبابًا وجيهة لأن نتشكَّك على الأقل في العالم الجديد الجميل الذي رسمَتْه لنا نبوءاتُ الذكاء الاصطناعي.

علاوةً على ذلك، قد لا يكون العمل بالضرورة شقاءً يجب تجنُّبُه أو استغلالًا يجب مقاومته؛ فثمة وجهة نظر أخرى تُشير إلى أن العمل له قيمة في حدِّ ذاته، وأنه يمنح العامل هدفًا ومعنًى، وأن له فوائد مُتنوعة مثل التواصُل الاجتماعي مع الآخرين، والانتماء إلى شيءٍ أكبر، والتمتُّع بالصحة، والحصول على فُرَص لممارسة المسئولية (Boddington 2016). فإذا كان هذا هو الحال، فلربما كان علينا أن نحتفِظ بالعمل للبشر؛ أو على الأقل ببعض أنواع العمل، كالعمل ذي المَغزى الذي يُوفِّر فرصًا لتحقيق هذه الفوائد. أو ربما علينا أن نحتفِظ على الأقل ببعض المهام. وليس على الذكاء الاصطناعي أن يأخُذ على عاتقه وظائف بأكملها، ولكن يُمكن أن يتولَّى بعض المهام ذات القيمة الأقل. ويُمكننا أن نتعاون مع الذكاء الاصطناعي. على سبيل المثال، يُمكننا اختيار عدم تفويض العمل الإبداعي إلى الذكاء الاصطناعي (وهو ما يقترِحه بوستروم) أو يُمكننا اختيار التعاون مع الذكاء الاصطناعي للقيام بأشياء إبداعية. ما يُثير القلق في هذا الصدد هو أنه إذا كانت الآلات ستتولَّى القيام بكلِّ ما نقوم به في حياتنا الآن، فلن يتبقَّى لنا شيء نقوم به، وسنجد حياتنا بلا معنًى. ومع ذلك، فنحن نقول «إذا»؛ ويجب أن نضع في اعتبارنا الشكَّ فيما يمكن أن يقوم به الذكاء الاصطناعي (انظر الفصل الثالث) وحقيقة أن العديد من أنشطتنا ليست «عملًا» ولكنها ذات مغزًى كبير، وبالتالي فإننا سنحتفظ على الأرجح بالكثير لنقوم به. على هذا، يُمكننا أن نقول إن السؤال الآن ليس ماذا سيفعل البشر عندما تتولَّى الآلات القيام بجميع أعمالهم وأنشطتهم، ولكن أي المهام نريد أو نحتاج إلى الاحتفاظ بها للبشر، وما هي الأدوار التي يمكن أن يتولَّاها الذكاء الاصطناعي، إن كان سيتولى أي أدوار، لدعمنا في هذه المهامِّ بطرق أخلاقية ومقبولة اجتماعيًّا.

ختامًا، تدعونا أخلاقيَّات الذكاء الاصطناعي إلى التفكير في ماهية المجتمع الخَيِّر والعادل، وماهية الحياة البشرية ذات المعنى، وماهية الدور الذي تضطلِع به التكنولوجيا والذي يمكن أن تضطلع به فيما يتعلَّق بكلِّ ذلك. ويمكن أن تكون الفلسفة، بما فيها

الفلسفة القديمة، مصدر إلهامٍ للتفكير في تقنيات اليوم والمشكلات التي تجلِبها بالفعل والتي يُحتمَل أن تجلبها من الناحية الأخلاقية والمُجتمعية. فإذا كان الذكاء الاصطناعي يُثير هذه الأسئلة القديمة حول الحياة الجيدة ذات المعنى، فلدَينا مصادر في مختلف التقاليد الفلسفية والدينية يمكن أن تُساعدنا في التعامُل مع هذه الأسئلة. على سبيل المثال، كما اقترحت شانون فالور (٢٠١٦)، فإن تقليد أخلاقيات الفضيلة الذي وضعه أرسطو وكونفوشيوس وفلاسفة قدماء آخرون ربما ما زال يستطيع أن يُساعدنا اليوم للتفكير في معنى ازدهار الإنسان وكيف ينبغي أن يكون في عصر التكنولوجيا. وبعبارةٍ أخرى، قد تُوجَد لدينا بالفعل إجابات عن هذه الأسئلة، ولكن علينا القيام ببعض العمل للتفكير في معنى الحياة الجيدة في سياق التكنولوجيا الحديثة، بما في ذلك الذكاء الاصطناعي.

ومع ذلك، تُواجِه فكرة تطوير «أخلاقيات الذكاء الاصطناعي للحياة الجيدة» وأخلاقيات الذكاء الاصطناعي للعالَم الواقعي بشكلٍ عامٍّ عدة مشكلات. تتمثَّل المشكلة الأولى في السرعة. يفترض نموذج أخلاقيات الفضيلة الذي ورثته الفلسفة الغربية من أرسطو مجتمعًا يتغيَّر ببطءٍ ولا تتغيَّر فيه التكنولوجيا بسرعةٍ كبيرة، ويمتلك فيه الناس وقتًا لتعلُّم الحكمة العملية؛ ولذا، فإنه من غير الواضح كيف يمكن استخدامه للتعامُل مع مجتمعٍ سريع التغيُّر (Boddington 2016) ومع التطوُّر السريع للتقنيات مثل الذكاء الاصطناعي. هل ما زال لدَينا الوقت الكافي للاستجابة ولتطوير الحكمة العملية ونقلها فيما يتعلَّق باستخدام تقنيات مثل الذكاء الاصطناعي؟ هل تأتي الأخلاقيات بعد فوات الأوان؟ عندما تنشُر بومة مِينيرفا جناحَيها (التي ترمز للحكمة عند اليونان)، ربما يكون شكل العالم قد تغيَّر تمامًا ولم يعُد بالإمكان التعرف عليه. فما هو دور مثل هذه الأخلاقيات، وماذا ينبغي أن يكون دورها في سياق التطوُّرات التي تحدث في العالم الواقعي؟

أما المشكلة الثانية، فنظرًا إلى تنوع وتعدد وجهات النظر في هذا الأمر داخل المجتمعات، والاختلافات الثقافية بين المجتمعات، فإن الأسئلة الخاصة بماهية الحياة الجيدة ذات المعنى في ظل وجود التكنولوجيا يمكن الإجابة عنها على نحوٍ مختلف في الأماكن والسياقات المُختلفة، وهي تخضع، من الناحية العملية إلى كل أنواع العمليات السياسية التي قد تنتهي أو لا تنتهي بالتوافق. والاعتراف بهذا التنوُّع والتعدُّد قد يؤدي إلى نهج يميل إلى التعدُّدية. كما يمكن أن يأخذ شكل النسبية. وقد أثارت الفلسفة ونظرية المجتمع في القرن العشرين، خاصةً ما يُعرَف بمدرسة ما بعد الحداثة، الكثير

من الشكوك حول الإجابات التي يُزعَم كونها عالمية في حين أنها نشأت من سياقٍ جغرافي وتاريخي وثقافي مُعين (من «الغرب»، على سبيل المثال) وأنها مرتبطة بمصالح وعلاقات قوة مُعينة. كما أثيرت شكوك حول ما إذا كانت السياسة يجب أن تهدُف إلى التوافُق من الأساس (انظر أعمال شانتال موف، على سبيل المثال، موف ٢٠١٣)؛ وما إذا كان التوافُق مرغوبًا فيه دائمًا، أم أن الصراع الشرس حول مستقبل الذكاء الاصطناعي يمكن أن يكون له بعض الفوائد؟ وعلاوةً على ذلك، هناك مشكلة أخرى تتعلق بالهيمنة: فالتفكير في الأخلاقيات في العالم الحقيقي يعني التفكير ليس فقط فيما يجب القيام به فيما يتعلق بالذكاء الاصطناعي ولكن أيضًا فيمَن سيقرر، ومَن يجب عليه أن يقرر، مستقبل الذكاء الاصطناعي وبالتالي مستقبل مجتمعنا. ودعونا نفكر معًا مرة أخرى في قضايا الحكم الشمولي وهيمنة الشركات الكبيرة. وإذا رفضنا الحكم الشمولي والبلوتوقراطية (حكم الأثرياء)، فماذا يعني اتخاذ قرار ديموقراطي بشأن الذكاء الاصطناعي؟ ما هو نوع المعرفة المتعلق بالذكاء الاصطناعي الذي يحتاجه السياسيون والمواطنون؟ إذا كان هناك فَهم ضعيف للغاية للذكاء الاصطناعي ومشكلاته المحتمَلة، فإننا نواجه خطر التكنوقراطية أو ببساطة عدم وجود سياسة للذكاء الاصطناعي على الإطلاق.

ومع ذلك، كما يُوضح الفصل التالي، يبدو أن واحدة على الأقل من العمليات السياسية المتعلقة بالذكاء الاصطناعي التي ظهرت مؤخرًا جاءت في الوقت المناسب. وتلك هي صنع سياسات خاصة بالذكاء الاصطناعي، وهي عملية استباقية، وتهدف إلى التوافق، وتُظهِر درجةً متزايدة من التقارب، ويبدو أنها تلتزم بنوع من العالمية بلا خجل، وتعتمد على المعرفة الخبيرة، وتزعم — ولو على الأقل شفهيًّا — احترام مبادئ الديمقراطية، وخدمة الصالح العام والمصلحة العامة، ومشاركة جميع الأطراف المَعنية.

الفصل العاشر

السياسات المقترحة

ما يجب القيام به وأسئلة أخرى يتعيَّن على صانعي السياسات الإجابة عنها

نظرًا إلى المشكلات الأخلاقية المُرتبطة بالذكاء الاصطناعي، فإنه من الواضح أن شيئًا ما يجب القيام به. ولذا، تتضمَّن معظم مبادرات السياسات المُتعلقة بالذكاء الاصطناعي أخلاقيات الذكاء الاصطناعي. وجدير بالذكر أن هناك الكثير من المبادرات في هذا المجال في الوقت الحالي. ومع ذلك، ليس من الواضح بالضبط ما يجِب القيام به، وما المسار الذي يجب اتِّخاذه. على سبيل المثال، ليس واضحًا كيفية التعامُل مع مشكلة الشفافية أو التحيُّز، نظرًا إلى التقنيات نفسها، والتحيُّز الذي يُعاني منه المجتمع بالفعل، والآراء المُتباينة حول العدالة والإنصاف. وهناك أيضًا العديد من التدابير المُمكن اتخاذها: إذ يمكن أن تعني السياسة التنظيمَ من خلال إصدار القوانين واللوائح، على سبيل المثال، الأنظمة القانونية، ولكن هناك أيضًا استراتيجيات أخرى قد تكون مُتصلة أو غير متصلة بالأنظمة القانونية، مثل التدابير التكنولوجية، وقواعد الأخلاق، والتعليم. ولا يقتصر التنظيم على القوانين ولكنه يتضمن أيضًا معايير مثل معايير الآيزو. وعلاوة على ذلك، هناك أيضًا أنواع أخرى من الأسئلة التي يتعيَّن الإجابة عنها في السياسات المقترحة؛ فالأمر ليس فقط ما يجِب القيام به، ولكن أيضًا لماذا يجب القيام به، ومتى يجِب القيام به، وما مِقدار ما يجب القيام به، ومَن يجب عليه أن يقوم به، وما هي طبيعة المشكلة ومَداها ودرجة خطورتها وإلحاحها.

أولًا: من المُهم تبرير التدابير المقترحة. على سبيل المثال، قد تستنِد السياسة المُقترحة إلى مبادئ حقوق الإنسان لتبرير اقتراحٍ بالتقليل من اتخاذ القرارات التي تعتمِد على خوارزميات مُتحيزة. ثانيًا: استجابة إلى التطوُّر التكنولوجي، غالبًا ما تأتي السياسة بعد فوات الأوان، عندما تكون التكنولوجيا قد توغَّلت بالفعل في المجتمع ودخلت في كلِّ

شيء. بدلًا من ذلك، يمكن أن نُحاول وضع سياسة قبل أن يكتمِل تطوير التكنولوجيا ويبدأ استخدامها. وفيما يخصُّ الذكاء الاصطناعي، يمكن القول إن هذا ما زال مُمكنًا، إلى حدٍّ ما، على الرغم من أن الكثير من الأنظمة المدعومة بالذكاء الاصطناعي موجودة بالفعل حولنا. والبُعد الزمني مُهم أيضًا فيما يتعلَّق بالنطاق الزمني للسياسة: هل هي مُخصَّصة فقط للسنوات الخمس أو العشر المقبلة، أم تهدف إلى أن تُكوِّن إطار عمل على المدى البعيد؟ هنا علينا أن نختار. على سبيل المثال، يمكن تجاهل التنبؤات على المدى البعيد والتركيز على المستقبل القريب، كما تفعل معظم السياسات المُقترحة، أو يمكن طرح رؤية لمستقبل الإنسانية. ثالثًا: لا يتفق الجميع على أن حلَّ المشكلات يتطلَّب الكثير من التدابير الجديدة. يزعم بعض الأشخاص والمؤسسات أن التشريعات الحالية كافية للتعامُل مع الذكاء الاصطناعي. فإذا كان هذا هو الحال، فإنه يبدو أن المُشرِّعين ليسوا في حاجة إلى القيام بالكثير، في حين أن الذين يُفسرون القانون والذين يُطوِّرون الذكاء الاصطناعي هم مَن يحتاجون إلى العمل الدءوب. ويعتقد آخَرون أنه يجب أن نُعيد التفكير في جوهر المجتمع ومؤسساته، بما في ذلك أنظمتنا القانونية، من أجل التعامُل مع المشكلات الأساسية وإعداد أجيال المُستقبل. رابعًا: يجب أن توضِّح السياسة المقترحة مَن الذي يجب أن يتخذ الإجراءات. وقد لا يقتصِر هذا على جهةٍ واحدة وإنما أكثر من جهة. فكما رأينا، يشترك الكثيرون في أي عملٍ تكنولوجي. ويُثير هذا سؤالًا حول كيفية توزيع المسئولية عن السياسة والتغيير: هل الحكومات أساسًا هي المسئولة عن اتخاذ إجراءات، أم يجب، على سبيل المثال، على الشركات والصناعة اتخاذ إجراءاتٍ خاصة بها لضمان الذكاء الاصطناعي الأخلاقي؟ وفيما يتعلَّق بالشركات، هل يجب مخاطبة الشركات الكبيرة فقط أم أيضًا الشركات الصغيرة والمتوسِّطة الحجم؟ وما هو دور العلماء (المُختصِّين بالكمبيوتر) والمهندسين الأفراد؟ وما هو دور المواطنين؟

خامسًا: تعتمد الإجابة عما يجب القيام به ومقدار ما يجب القيام به، وعن أسئلة أخرى، على كيفية تعريف طبيعة المشكلة نفسها ومَداها ودرجة خطورتها وإلحاحها. على سبيل المثال، هناك اتجاه في سياسات التكنولوجيا (وفي الواقع، في أخلاقيات الذكاء الاصطناعي) لرؤية مشكلاتٍ جديدة في كلِّ مكان. ومع ذلك، كما رأينا في الفصل السابق، فالعديد من المشكلات قد لا تكون حكرًا على التقنيات الجديدة، ولكنها ربما تكون موجودةً منذ وقتٍ طويل. علاوةً على ذلك، كما أظهر النقاش حول التحيُّز، يعتمِد ما نقترح القيام به على كيفية تعريف المشكلة: هل هي مشكلة خاصة بالعدالة، وإذا كانت كذلك، فما هو

نوع العدالة المُهدَّدة؟ سيشكل تعريف المشكلة التدابير التي نقترحها. على سبيل المثال، إذا قدَّمنا تدابير للعمل الإيجابي، فإن هذا يستند إلى تعريفٍ مُعين للمشكلة. وأخيرًا، يلعب أيضًا تعريف الذكاء الاصطناعي دورًا في تحديد السياسة المقترحة ونطاقها، وقد كان هذا التعريف دائمًا مُثيرًا للجدل والنقاشات. على سبيل المثال، هل مِن المُمكن ومن المُستحسَن أن نُميز بوضوح بين الذكاء الاصطناعي والخوارزميات الذكية المُستقلة، أو بين الذكاء الاصطناعي وتقنيات الأتمتة؟ جميع هذه الأسئلة تجعل من صنع السياسات المُتعلقة بالذكاء الاصطناعي أمرًا قد يُثير الجدل بشكلٍ كبير. وبالفعل، نجد العديد من الاختلافات والجدالات، على سبيل المثال حول مدى الحاجة إلى تشريعات جديدة، وحول المبادئ التي يجب الاستناد إليها بالضبط لتبرير التدابير، وحول مسألة ما إذا كان ينبغي تحقيق توازن بين أخلاقيات الذكاء الاصطناعي والاعتبارات الأخرى (مثل تنافسية الشركات والاقتصاد). ومع ذلك، إذا فكَّرنا في وثائق السياسة الفعلية، فسنجد درجةً ملحوظة من التقارُب.

المبادئ الأخلاقية والتبريرات

لقد أدَّى الإحساس الواسع الانتشار بضرورة وأهمية التعامُل مع التحديات الأخلاقية والمُجتمعية التي أثارها الذكاء الاصطناعي إلى سَيل من المبادرات ووثائق السياسات التي لا تُعرِّف فقط بعض المشكلات الأخلاقية المُرتبطة بالذكاء الاصطناعي ولكنها تهدف أيضًا إلى توفير توجيهات مِعيارية للسياسات. وقد اقتُرحت سياسات خاصة بالذكاء الاصطناعي تشتمل على عنصرٍ أخلاقي من قِبل مجموعة متنوعة من الجهات، بما في ذلك الحكومات والهيئات الحكومية مثل اللجان الوطنية للأخلاقيات، وشركات التكنولوجيا مثل جوجل، والمهندسين ومنظماتهم المهنية مثل معهد مهندسي الكهرباء والإلكترونيات، والهيئات الحكومية الدولية مثل الاتحاد الأوروبي، والجهات غير الحكومية وغير الهادفة للربح، والباحثين.

لقد أدَّى الإحساس الواسع الانتشار بضرورة وأهمية التعامُل مع التحديات الأخلاقية والمُجتمعية التي أثارها الذكاء الاصطناعي إلى سَيل من المبادرات ووثائق السياسات.

إذا راجعنا بعض المبادرات والمقترحات الحديثة، يتبيَّن أن معظم الوثائق تبدأ بتبرير السياسة من خلال توضيح المبادئ، ثم تُقدم بعض التوصيات فيما يتعلق بالمشكلات

الأخلاقية المُحددة. وكما سنرى، هذه المشكلات والمبادئ شديدة التشابُه. وفي كثير من الحالات، تعتمِد المبادرات على مبادئ أخلاقية عامة ومبادئ من قانون أخلاقيات المهنة. فدعوني أُراجِع معكم بعض المقترحات.

ترفض معظم المقترحات سيناريو الخيال العلمي الذي تستولي فيه الآلات الفائقة الذَّكاء على زمام الأمور وتتولَّى فيه السيطرة. على سبيل المثال، في فترة رئاسة أوباما، نشرت حكومة الولايات المتحدة تقريرًا بعنوان «الاستعداد لمُستقبل الذكاء الاصطناعي»، تؤكِّد فيه صراحةً على أن المخاوف الطويلة الأمد بشأن الذكاء الاصطناعي الفائق العام «يجب ألا يكون لها تأثير كبير على السياسة الحالية» (المكتب التنفيذي للرئيس ٢٠١٦، ٨). وبدلًا من ذلك، يتناول التقرير المُشكلات الحالية والمُتوقَّعة في المستقبل القريب التي يُثيرها تعلُّم الآلة، مثل التحيُّز ومشكلة أنه حتى المُطورون قد لا يفهمون نظامهم بما فيه الكفاية لتجنُّب مثل هذه العواقب. ويؤكِّد التقرير أن الذكاء الاصطناعي مُفيد للابتكار والنمو الاقتصادي ويُشدِّد على الرقابة الذاتية، ولكنه يقول إن حكومة الولايات المتحدة يُمكنها مراقبة سلامة التطبيقات وعدالتها، وتعديل الأطر القانونية إذا لزِم الأمر.

علاوةً على ذلك، تملك العديد من الدول الأوروبية حاليًّا استراتيجيات للذكاء الاصطناعي تتضمَّن عنصرًا أخلاقيًّا. ويُعد «الذكاء الاصطناعي القابل للتفسير» هدفًا مشتركًا بين العديد من صانعي السياسات. يقول مجلس عموم المملكة المتحدة (٢٠١٨) إن الشفافية وحق التفسير أمور أساسية لنتمكَّن من مساءلة الخوارزميات، ويجب على الصناعات والجهات التشريعية التعامُل مع مسألة اتخاذ القرارات المُتحيزة من قِبل الخوارزميات. كذلك تفحص لجنة مجلس لوردات المملكة المتحدة المختارة المَعنية بالذكاء الاصطناعي التداعِيات الأخلاقية للذكاء الاصطناعي. وفي فرنسا، يقترح تقرير فيلاني العمل نحو تطوير «ذكاء اصطناعي ذي معنًى» لا يؤدي إلى تفاقُم مشكلات الإقصاء، أو يزيد من التفاوت الاجتماعي، أو يؤدي إلى مجتمع تحكُمنا فيه خوارزميات «صناديق سوداء»؛ إذ يجب أن يكون الذكاء الاصطناعي قابلًا للتفسير وصديقًا للبيئة (Villani 2018). كما أنشأت النمسا مؤخرًا مجلسًا استشاريًّا وطنيًّا معنيًّا بالروبوتات والذكاء الاصطناعي،[1] والذي قدَّم توصياتٍ لسياسةٍ تستنِد إلى حقوق الإنسان، والعدالة والإنصاف، والإشراك والتضامُن، والديمقراطية والمشاركة، وعدم التمييز، والمسئولية، وقِيَم أخرى شبيهة. كما تُوصي ورقتها البيضاء بتطوير ذكاء اصطناعي قابلٍ للتفسير وتقول صراحةً إن المسئولية تظلُّ على عاتق البشر؛ ولا يمكن أن يكون الذكاء الاصطناعي

مسئولًا أخلاقيًّا (ACRAI 2018). كذلك، فإن الهيئات والمؤتمرات الدولية نشطة للغاية. فقد نشر المؤتمر الدولي لمُفوضي حماية البيانات والخصوصية إعلانًا بشأن الأخلاقيات وحماية البيانات في الذكاء الاصطناعي، ويتضمَّن مبادئ العدالة، والمساءلة، والشفافية والفهم، والتصميم المسئول، والخصوصية المُتضمنة في التصميم (مفهوم يُطالب بمراعاة الخصوصية في جميع مراحل عملية الهندسة)، وتمكين الأفراد، والحدِّ من التحيز أو التمييز وتخفيف آثارهما (ICDPPC 2018).

يضع بعض صانعي السياسات هدفهم في إطار «الذكاء الاصطناعي الجدير بالثقة». فعلى سبيل المثال، تؤكِّد المُفوضية الأوروبية، التي تُعَد بلا شكٍّ واحدة من أبرز الهيئات العالمية في مجال صُنع سياسات الذكاء الاصطناعي، على أهمية هذا المصطلح. وفي أبريل ٢٠١٨، أنشأت فريقَ خبراء رفيع المستوى معنِيًّا بالذكاء الاصطناعي لوضع مجموعةٍ جديدة من إرشادات الذكاء الاصطناعي؛ وفي ديسمبر ٢٠١٨، أصدر الفريق مُسودة وثيقة عمل تتضمَّن إرشادات أخلاقية تدعو إلى نهجٍ في الذكاء الاصطناعي يتمحور حول الإنسان، وإلى تطوير ذكاءٍ اصطناعي جدير بالثقة، يحترم الحقوق الأساسية والمبادئ الأخلاقية. وكانت الحقوق المذكورة هي كرامة الإنسان، وحرية الفرد، واحترام الديمقراطية، والعدالة، وسيادة القانون، وحقوق المواطن. أما المبادئ الأخلاقية، فهي الإحسان (فعل الخير) وعدم إلحاق الأذى، والاستقلال (الحفاظ على وكالة الإنسان)، والعدالة (أن تكون عادلًا)، والقابلية للتفسير (شفافية التنفيذ). هذه المبادئ مألوفة من مجال أخلاقيات علم الأحياء، ولكن الوثيقة تُضيف إليها القابلية للتفسير، وتتضمَّن تفسيراتٍ تسلط الضوء على المشكلات الأخلاقية الخاصة التي يُثيرها الذكاء الاصطناعي. على سبيل المثال، يُفسِّر مبدأ عدم إلحاق الأذى على المطالبة بأن خوارزميات الذكاء الاصطناعي يجب أن تتجنَّب التمييز، والتلاعُب، والتوجيه السلبي، ويجِب أن تحمي الفئات الضعيفة مثل الأطفال والمهاجرين. أما مبدأ العدالة، فيُفسَّر على أنه يتضمن مطالبة مطوري الذكاء الاصطناعي ومنفِّذيه بضمان احتفاظ الأفراد والمجموعات الأقلية بالتحرُّر من التحيُّز. ويفسر مبدأ القابلية للتفسير على أنه يُطالب بأن تكون أنظمة الذكاء الاصطناعي قابلة للتدقيق و«مفهومة من قِبل البشر على اختلاف مستويات فهمهم وخبرتهم» (European Commission AI HLEG 2018, 10). وتُحدِّد النسخة النهائية، التي صدرت في أبريل ٢٠١٩، بشكلٍ خاصٍّ أن قابلية التفسير لا تتعلَّق فقط بتفسير العملية التقنية ولكن أيضًا بالقرارات البشرية ذات الصِّلة بها (European Commission AI HLEG 2019, 18).

في وقتٍ سابق، أصدرت هيئة استشارية أخرى تابعة إلى الاتحاد الأوروبي، وهي المجموعة الأوروبية المَعنية بالأخلاقيات في العلوم والتقنيات الجديدة بيانًا حول الذكاء الاصطناعي والروبوتات والأنظمة المُستقلة، مقترحةً مبادئ الكرامة الإنسانية، والاستقلال، والمسئولية، والعدالة، والمساواة، والتضامُن، والديمقراطية، وسيادة القانون والمساءلة، والأمان والسلامة، وحماية البيانات والخصوصية، والاستدامة. ويُقال إن مبدأ الكرامة الإنسانية يقتضي إعلام الأفراد بما إذا كانوا يتفاعلون مَع آلة أم معَ إنسان آخر (EGE 2018). كذلك عليك ملاحظة أن الاتحاد الأوروبي لدَيه بالفعل تشريعات قائمة تتعلق بتطوير الذكاء الاصطناعي واستخدامه. وتهدف لائحة حِماية البيانات العامة، التي اعتُمِدت في مايو ٢٠١٨، إلى حماية جميع مواطني الاتحاد الأوروبي وتمكينهم فيما يتعلَّق بخصوصية البيانات. وتتضمَّن مبادئ مثل حق الفرد في نسيان بياناته (يمكن للفرد أن يطلب مسح بياناته الشخصية ووقف معالجة تلك البيانات في المستقبل) والْخصوصية المُتضمَّنة في التصميم. كما تمنح الأفراد المعنيين حق الوصول إلى «معلومات ذات معنى حول المنطق المُضمَّن» في اتخاذ القرارات المؤتمتة ومعلومات حول «العواقب المُتوقَّعة» لِمثل هذه المعالجة (البرلمان الأوروبي ومجلس الاتحاد الأوروبي ٢٠١٦). الاختلاف عن وثائق السياسة هو أن هذه المبادئ المذكورة هنا تُعد مُتطلبات قانونية. إنها بمثابة تشريع مفروض؛ بمعنى أن المؤسَّسات التي تنتهك لائحة حماية البيانات العامة يُمكن تغريمها. ومع ذلك، ثمة تساؤل مطروح عما إذا كانت أحكام لائحة حماية البيانات العامَّة تكافئ الحق الكامل في تفسير القرار (Digital Europe 2018)، وبشكل عام، إذا كانت توفِّر حماية كافية ضد مخاطر اتخاذ القرار المؤتمت (Wachter, Mittelstadt, and Floridi 2017). توفر لائحة حماية البيانات العامَّة الحق في الإعلام باتخاذ القرار المؤتمت ولكن يبدو أنها لا تُطالِب بتفسير الأساس المنطقي لأي قرارٍ بعَينه. وهذه أيضًا مشكلة فيما يتعلَّق باتخاذ القرار في المجال القانوني. وقد طالبت دراسةٌ أجراها مجلس أوروبا، استنادًا إلى عمل لجنةٍ من خبراء حقوق الإنسان، بأن يكون للأفراد الحق في محاكمةٍ عادلة وإجراءات قانونية سليمة بشروط يُمكنهم فهمها (Yeung 2018).

تُعَد المناقشات القانونية ذات أهميةٍ بالطبع في المناقشات المُتعلقة بأخلاقيات الذكاء الاصطناعي وسياسة الذكاء الاصطناعي. وقد ناقش تيرنر (٢٠١٩) المقارنات بالحيوانات (كيف عومِلت وتُعامل في القانون وما إذا كانت تتمتَّع بحقوق) وراجع عددًا من الصكوك القانونية فيما يتعلق بما يمكن أن تعني للذكاء الاصطناعي. على سبيل المثال، عند وقوع

الضرر، فإن مسألة الإهمال تتعلق بما إذا كان شخصٌ ما مُلتزمًا بواجب الرعاية لتجنُّب وقوع ضرر، حتى إذا لم يكن الضرَر الواقع مقصودًا. يمكن أن ينطبق ذلك على مُصمم أو مُدرب الذكاء الاصطناعي. ولكن ما مدى سهولة التنبؤ بعواقب الذكاء الاصطناعي؟ أما القانون الجنائي، فعلى العكس من ذلك، فهو يتطلَّب نِيَّة إيقاع الضرر. ولكن هذا غالبًا ليس الحال مع الذكاء الاصطناعي. من ناحيةٍ أخرى، لا تتعلق المسئولية عن المنتَج بخطأ الأفراد ولكنها تفرِض على الشركة التي أنتجت التكنولوجيا دفع تعويضاتٍ عن الأضرار، بغضِّ النظر عن الخطأ. ويمكن أن يكون هذا أحد الحلول المُمكنة للمسئولية القانونية عن الذكاء الاصطناعي. كذلك تتَّصل قوانين الملكية الفكرية بالذكاء الاصطناعي، مثل حقوق الطبع والنشر وبراءات الاختراع، وقد بدأت مناقشات حول «الشخصية الاعتبارية» للذكاء الاصطناعي، وهو ما يُعد افتراضًا قانونيًّا ولكنه ذريعة تُطبَّق حاليًّا على الشركات ومختلف المُنظمات. فهل يجب أن يُطبَّق أيضًا على الذكاء الاصطناعي؟ في قرارٍ مُثير للجدل في عام ٢٠١٧، اقترح البرلمان الأوروبي أن منح الروبوتات الذاتية التشغيل الأكثر تطورًا منزلة الأشخاص الإلكترونيين هو حلٌّ قانوني مُمكن لقضية المسئولية القانونية؛ وهذه الفكرة لم يتم الاعتراف بها من قِبل المفوضية الأوروبية في استراتيجيتها للذكاء الاصطناعي[2] في عام ٢٠١٨. كذلك اعترض آخرون اعتراضًا حازمًا على فكرة إعطاء حقوقٍ وشخصية للآلات، مُجادِلين، على سبيل المثال، بأنه سيُصبح من الصعب، إن لم يكُن من المُستحيل، محاسبة أي شخصٍ لأن الناس سيسعَون إلى استغلال هذه الفكرة لأغراضٍ ذاتية (Bryson, Diamantis, and Grant 2017). كان هناك أيضًا الحالة الشهيرة لصوفيا، الروبوت الذي منحته السعودية «الجنسية» في عام ٢٠١٧. تُثير مثل هذه الحالة مجددًا مسألة المكانة الأخلاقية للروبوتات والذكاء الاصطناعي (انظر الفصل الرابع).

اقتُرِحَت أيضًا سياسات ذكاء اصطناعي خارج نطاق أمريكا الشمالية وأوروبا. فالصين، على سبيل المثال، لدَيها استراتيجية وطنية للذكاء الاصطناعي. وتُقر خُطتها التنموية بأن الذكاء الاصطناعي هو تكنولوجيا هدَّامة يمكن أن تضرَّ بالاستقرار الاجتماعي، وتؤثر على القانون والأخلاقيات الاجتماعية، وتنتهِك الخصوصية الشخصية، وتخلق مخاطر أمنية؛ ومِن ثَم تُوصي الخطة بتعزيز الوقاية المُستقبلية وتقليل المخاطر المحتمَلة (مجلس الدولة الصيني ٢٠١٧). وتروي بعض الجهات الفاعلة في الغرب سردية منافسة: إنهم يخشَون أن تتجاوزهم الصين أو حتى فكرة أننا نقترِب من اندلاع حربٍ عالمية جديدة. بينما يُحاول آخرون التعلُّم من استراتيجية الصين. وقد يتساءل الباحثون أيضًا

عن كيفية تعامُل الثقافات المختلفة مع الذكاء الاصطناعي بطرُق مختلفة. ويمكن أن يُسهم البحث في مجال الذكاء الاصطناعي نفسه في بناء وجهة نظر مقارنة عابرة للثقافات بشأن أخلاقيات الذكاء الاصطناعي، على سبيل المثال، عندما يُذكرنا بالفروق بين الثقافات الفردية والجماعية فيما يتعلَّق بالمُعضلات الأخلاقية (Awad et al. 2018). ويمكن أن يُثير هذا مشكلاتٍ لأخلاقيات الذكاء الاصطناعي إذا كانت تهدف إلى أن تكون عالمية. ويمكن أيضًا استكشاف كيف تختلف السرديات حول الذكاء الاصطناعي في الصين أو اليابان، على سبيل المثال، عن السرديات الغربية. ومع ذلك، على الرغم من الاختلافات الثقافية، يتبيَّن أن سياسات أخلاقيات الذكاء الاصطناعي مُتشابهة بدرجةٍ كبيرة وملحوظة. فبينما تؤكد خطة الصين أكثر على الاستقرار الاجتماعي والصالح العام الجماعي، إلا أن المخاطر الأخلاقية المُحددة والمبادئ المذكورة ليست مختلفة كثيرًا عن تلك المُقترحة من قِبل الدول الغربية.

على الرغم من الاختلافات الثقافية، يتبيَّن أن سياسات أخلاقيات الذكاء الاصطناعي متشابهة بدرجة كبيرة وملحوظة.

ولكن، كما ذكرنا سابقًا، سياسة أخلاقيات الذكاء الاصطناعي ليست مقصورةً على الحكومات ولجانها وهيئاتها فقط. فقد أخذ الأكاديميون أيضًا زمام المبادرة. على سبيل المثال، اقتُرِحَ إعلان مونتريال بشأن الذكاء الاصطناعي المسئول من قِبل جامعة مونتريال وشمل استشارة المواطنين والخبراء وغيرهم من أصحاب الشأن. ويقول الإعلان إن تطوير الذكاء الاصطناعي يجب أن يُعزِّز رفاه جميع المخلوقات الحية ويُعزِّز استقلال البشر، ويقضي على جميع أنواع التمييز، ويحترم الخصوصية الشخصية، ويَحمينا من الدعاية والتلاعُب، ويُعزِّز النقاش الديمقراطي، ويجعل مختلف الجهات الفاعلة مَسئولين عن مكافحة مخاطر الذكاء الاصطناعي (Université de Montréal 2017). وقد اقترح باحثون آخرون مبادئ الإحسان، وعدم التسبُّب في الأذى، والاستقلال، والعدالة، وقابلية التفسير (Floridi et al. 2018). وتعمل الجامعات مثل كامبريدج وستانفورد على أخلاقيات الذكاء الاصطناعي، غالبًا من وجهة نظر الأخلاق التطبيقية. وكذلك يؤدي الأشخاص العامِلون في مجال الأخلاق المهنية أيضًا عملًا مُفيدًا. على سبيل المثال، قدم مركز ماركولا للأخلاق التطبيقية في جامعة سانتا كلارا مجموعةً من النظريات الأخلاقية كأداةٍ

لُممارسة التكنولوجيا والهندسة، والتي قد تُفيد أيضًا في إثراء أخلاقيات الذكاء الاصطناعي بالمعلومات.[3] كما أبدى فلاسفة التكنولوجيا اهتمامًا كبيرًا بالذكاء الاصطناعي مؤخرًا.

نجد أيضًا مُبادرات بشأن أخلاقيات الذكاء الاصطناعي في عالم الشركات. على سبيل المثال، يدخل في الشراكة بشأن الذكاء الاصطناعي شركات مثل ديب مايند، وآي بي إم، وإنتل، وأمازون، وأبل، وسوني، وفيسبوك.[4] وتدرك العديد من الشركات الحاجة إلى الذكاء الاصطناعي الأخلاقي. على سبيل المثال، نشرت جوجل مبادئ أخلاقيات الذكاء الاصطناعي: تقديم فائدة اجتماعية، وتجنُّب التسبُّب في التحيُّز غير العادل أو تعزيزه، وفرض السلامة، والحفاظ على تحمُّل المسئولية، والحفاظ على تصميم الخصوصية، وتعزيز التميُّز العلمي، وتقييد التطبيقات التي يُحتَمل كونها ضارة أو مُسيئة مثل الأسلحة أو التكنولوجيا التي تنتهِك مبادئ القانون الدولي وحقوق الإنسان.[5] وتتحدَّث شركة مايكروسوفت عن فكرة «الذكاء الاصطناعي من أجل الخير» وتقترح مبادئ العدالة، والموثوقية والسلامة، والخصوصية والأمان، والتضمين، والشفافية، والمساءلة.[6] كما اقترحت شركة أكسنتشر مبادئ عالمية لأخلاقيات البيانات، بما في ذلك احترام الأشخاص الكامنة وراء البيانات، والخصوصية، والتضمين، والشفافية.[7] وعلى الرغم من أن وثائق الشركات تَميل إلى التركيز على الرقابة الذاتية، فإن بعض الشركات تعترف بضرورة اللوائح التنظيمية الخارجية. وقد قال تيم كوك الرئيس التنفيذي لشركة أبل إن اللوائح التنظيمية التكنولوجية، على سبيل المثال، لضمان الخصوصية أمر لا غِنى عنه لأن السوق الحرة التي لا تخضع لرقابة حكومية لا تُفيد في هذه الحالة.[8] ومع ذلك، هناك جدل حول ما إذا كان هذا يتطلب لوائح تنظيمية جديدة. ويدعم البعض مسار اللوائح التنظيمية، بما في ذلك القوانين الجديدة. فقد قدمت ولاية كاليفورنيا بالفعل مشروع قانون يطالب بالكشف عن الروبوتات: فإن استخدام الروبوت بطريقةٍ تُضلِّل الشَّخص الآخر حول هويته الاصطناعية أمر غير قانوني.[9] وتتخذ شركات أخرى موقفًا أكثر تحفظًا. فقد جادلت شركة ديجيتال يوروب (٢٠١٨)، التي تمثل الصناعة الرقمية في أوروبا، بأن الإطار القانوني الحالي مُجهَّز لمعالجة المشكلات المتعلقة بالذكاء الاصطناعي، بما فيها التحيُّز والتمييز، ولكن لبناء الثقة، فإن الشفافية والقابلية للتفسير أمران غاية في الأهمية: يجب أن يفهم الأفراد والشركات متى وكيف تُستخدَم الخوارزميات في اتخاذ القرارات، ونحن بحاجةٍ إلى توفير معلوماتٍ ذات معنى وتيسير عملية تفسير القرارات الخوارزمية.

تلعب الجهات غير الهادفة إلى الربح دورًا أيضًا. على سبيل المثال، تطرح الحملة الدولية لوقف الروبوتات القاتلة العديد من الأسئلة الأخلاقية بشأن التطبيقات العسكرية

للذكاء الاصطناعي[10] ومن جانب دُعاة تجاوز الإنسانية، تُوجَد مبادئ الذكاء الاصطناعي التي اتَّفق عليها المشاركون الأكاديميون والصناعيون في مؤتمر أسيلومار، وهو مؤتمر عقده «معهد مستقبل الحياة» (ماكس تيجمارك وآخرون). وكان الهدف العام هو الحرص على أن يظل الذكاء الاصطناعي مفيدًا، واحترام المبادئ والقيم الأخلاقية مثل السلامة والشفافية والمسئولية، وتوجيه القيم، والخصوصية، والتحكم البشري.[11] هناك أيضًا منظمات مهنية تعمل في مجال سياسات الذكاء الاصطناعي. فقد طرح معهد مهندسي الكهرباء والإلكترونيات، الذي يزعم أنه أكبر منظمة مهنية فنية في العالم، مبادرةً عالمية حول أخلاقيات الأنظمة الذكية والمُستقلة. وبعد مناقشاتٍ بين الخبراء، أثمرت المبادرة عن وثيقةٍ تتضمَّن رؤية لـ «تصميمٍ موجَّه أخلاقيًّا»، تقترح أن يكون تصميم هذه التقنيات وتطويرها وتنفيذها موجهًا بواسطة المبادئ العامة لحقوق الإنسان والرفاه والمساءلة والشفافية والتوعية بشأن سوء الاسْتِخدام. ويمكن أن يكون تضمين الأخلاق في المعايير التكنولوجية العالمية وسيلة فعَّالة للمساهمة في تطوير الذكاء الاصطناعي الأخلاقي.

الحلول التكنولوجية ومسألة الأساليب والتنفيذ

تبين المبادرة العالمية التي طرحها معهد مهندسي الكهرباء والإلكترونيات أنه فيما يتعلق بالتدابير، تُركز بعض وثائق السياسات على الحلول التكنولوجية. على سبيل المثال، كما ذكرنا في الفصل السابق، دعا بعض الباحِثين إلى الذكاء الاصطناعي القابل للتفسير، إلى فتح الصندوق الأسود. وهناك أسباب وجيهة للرغبة في فِعل ذلك؛ إذ إن تفسير المنطق وراء القرار الذي يُتَّخَذ ليس مطلبًا أخلاقيًّا فقط ولكنه أيضًا جانب مُهم من الذكاء البشري (Samek, Wiegand, and Müller 2017). إذَن فالفكرة وراء الذكاء الاصطناعي القابل للتفسير أو الشفَّاف هي أن يكون من السهل فهم أفعال الذكاء الاصطناعي وقراراته. وكما رأينا، فإن هذه الفكرة من الصعب تنفيذها في حالة تعلُّم الآلة الذي يستخدم الشبكات العصبية (Goebel et al. 2018). ولكن يمكن للسياسات بالطبع دعم البحث في هذا الاتجاه.

بشكل عام، فإن فكرة تضمين الأخلاق في تصميم التقنيات الجديدة هي فكرة رائعة. ويمكن أن تُساعدنا الأفكار مثل الأخلاقيات المُتضمنة في التصميم أو التصميم الحسَّاس للقيم، التي لها تاريخها الخاص،[12] في تصميم الذكاء الاصطناعي بطريقةٍ تؤدي إلى مزيدٍ من المساءلة والمسئولية والشفافية. على سبيل المثال، يمكن أن تنطوي الأخلاقيات المُتضمَّنة

في التصميم على ضمان التتبُّع في جميع المراحل (Dignum et al. 2018)، مما يُسهم في إمكانية مساءلة الذكاء الاصطناعي. ويمكن تحقيق فكرة التتبع حرفيًّا، بمعنى تسجيل بيانات حول سلوك النظام. وقد طالب وينفيلد وجيروتكا (٢٠١٧) بتنفيذ «صندوق أسود أخلاقي» في الروبوتات والأنظمة المُستقلَّة، ليُسجل ما يفعله الروبوت (البيانات من الأجهزة الاستشعارية ومِن الوضع «الداخلي» للنظام) بطريقةٍ تُشبه الصندوق الأسود المُثبَّت في الطائرات. ويمكن تطبيق هذه الفكرة أيضًا في الذكاء الاصطناعي المُستقل: فعندما يحدث خطأ ما، قد تُساعدنا مثل هذه البيانات في تفسير ما حدث بالضبط. وهذا بدَوره قد يُساعد في التحليل الأخلاقي والقانوني للحالة. وعلاوةً على ذلك، كما يقول الباحثون، وهم مُحقُّون في قولهم، يُمكننا أن نتعلَّم شيئًا من صناعة الطائرات، التي تخضع إلى تنظيمٍ صارم ولدَيها عمليات دقيقة للتحقُّق من السلامة وعمليات مرئية للتحقيق في الحوادث. فهل يُمكن تثبيت بِنية أساسية مُماثلة تضمن التنظيم والسلامة في حالة الذكاء الاصطناعي؟ وللمُقارنة بمجال آخَر من مجالات وسائل النقل، قد اقترحت صناعة السيارات أيضًا شهادةً أو نوعًا من «رخصة القيادة» للمركبات الذاتية التشغيل المدعومة بالذكاء الاصطناعي.[13] يذهب بعض الباحثين إلى أبعَدَ من ذلك ويهدفون إلى إنشاء آلاتٍ أخلاقية، في مُحاولةٍ لتحقيق «أخلاقيات الآلة» بمعنى أن تستطيع الآلات نفسها اتخاذ قراراتٍ أخلاقية. ويُجادل آخرون بأن هذه فكرة خطيرة وأنه يجِب الاحتفاظ بهذه القدرة للبشَر، وأنه من المُستحيل خلق آلات تتمتَّع بالوكالة الأخلاقية الكاملة، ولا حاجة إلى أن تتمتَّع الآلات بالوكالة الأخلاقية الكاملة، ويكفي أن تكون الآلات آمنةً وملتزمة بالقانون (Yampolskiy 2013)، أو قد تُنشَأ أشكال من «الأخلاق الوظيفية» (Wallach and Allen 2009) التي لا تكافئ الوكالة الأخلاقية الكاملة، ولكنها مع ذلك تجعل الآلة مُراعيةً نسبيًّا لقواعد الأخلاق. تُعد هذه المناقشة، التي تتعلَّق مجددًا بموضوع المكانة الأخلاقية، ذات صِلة، على سبيل المثال، في حالة السيارات الذاتية القيادة: وإلى أي مدًى يتعين ويمكن ويُستحسَن تضمين القواعد الأخلاقية في هذه السيارات، وما نوع هذه القواعد الأخلاقية وكيف يُمكن تنفيذها تقنيًّا؟

يمكن أن تُساعدنا الأفكار مثل الأخلاقيات المُتضمنة في التصميم أو التصميم الحسَّاس للقِيَم، في إنشاء الذكاء الاصطناعي بطريقةٍ تؤدي إلى مزيدٍ من المساءلة والمسئولية والشفافية.

يميل صانعو السياسات إلى دعم العديد من هذه الاتجاهات في البحث والابتكار في مجال الذكاء الاصطناعي، مثل الذكاء الاصطناعي القابل للتفسير وبشكلٍ عام، تضمين الأخلاق في التصميم. على سبيل المثال، إلى جانب الأساليب غير التقنية مثل اللوائح التنظيمية، ووضع المعايير، والتعليم، وحوار الأطراف المَعنية وفِرق التصميم الشاملة، ذكر تقرير فريق الخبراء الرفيع المستوى عددًا من الأساليب التقنية ومنها تضمين القواعد الأخلاقية وسيادة القانون في التصميم، وهياكل الذكاء الاصطناعي الجدير بالثقة، والاختبار والتحقق، والتتبع والتدقيق، والتفسير. على سبيل المثال، يُمكن أن تشتمل الأخلاقيات المُضمنة في التصميم على الخصوصية المُضمنة في التصميم. ويُشير التقرير أيضًا إلى بعض الطرق التي يُمكن بها تنفيذ الذكاء الاصطناعي الجدير بالثقة، مثل التتبُّع كطريقةٍ للمساهمة في الشفافية: وفي حالة الذكاء الاصطناعي المُستند إلى قواعد يجِب توضيح كيفية بناء النموذج، وفي حالة الذكاء الاصطناعي المُستنِد إلى التعلُّم يجب توضيح وسيلة تدريب الخوارزمية، بما في ذلك كيفية جمع البيانات واختيارها. ومن المُفترَض أن يضمَن هذا أن يكون نظام الذكاء الاصطناعي قابلًا للتدقيق، ولا سيَّما في المواقف الخطيرة (European Commission AI HLEG 2019).

تُعدُّ مَسألة الأساليب والتنفيذ حاسمةَ الأهمية: حيث إن تحديد عددٍ من المبادئ الأخلاقية شيء، واكتشاف طريقة تنفيذ هذه المبادئ عمليًّا شيءٌ مختلف تمامًا. وحتى المفاهيم مثل الخصوصية المُضمَّنة في التصميم، التي يُفترض أن تكون أقرب إلى عملية التطوير والهندسة، فغالبًا ما تُصاغ بطريقةٍ مجردة وعامة؛ ومِن ثَم فإننا ما زلنا لا ندري بالتحديد ما ينبغي أن نفعله. ويقودنا هذا إلى الفصل التالي لمناقشةٍ موجزة حول بعض التحديات التي تواجِه سياساتِ أخلاقيات الذكاء الاصطناعي.

الفصل الحادي عشر

التحديات التي تُواجه صانعي السياسات

الأخلاقيات الاستباقية: الابتكار المسئول وتضمين القِيَم في التصميم

ربما لا يُدهشنا أن نعرف أن سياسات أخلاقيات الذكاء الاصطناعي تواجِه العديد من التحدِّيات. وقد رأينا أن بعض السياسات المُقترحة تؤيد رؤيةً استباقية لأخلاقيات الذكاء الاصطناعي؛ بمعنى أننا بحاجةٍ إلى مراعاة الأخلاق في المرحلة المُبكرة من تطوير تكنولوجيا الذكاء الاصطناعي. وتكمُن الفكرة في تجنُّب المشكلات الأخلاقية والمُجتمعية التي يخلقها الذكاء الاصطناعي والتي سيكون من الصعب التعامُل معها بمجرد حدوثها. ويتماشى هذا مع أفكار الابتكار المسئول، وتضمين القِيَم في التصميم، وغيرها من الأفكار المُشابهة التي اقتُرِحت على مدار السنوات الأخيرة. وهذا يُحوِّل المشكلة من معالجة الآثار السلبية للتقنيات المُستخدمة على نطاق واسع بالفعل إلى تحمل المسئولية تجاه التقنيات التي يتم تطويرها اليوم.

ومع ذلك، ليس من السهل أن نتوقَّع العواقِب غير المقصودة للتقنيات الجديدة في مرحلة التصميم. إحدى الطرق لتخفيف هذه المشكلة هو بناء سيناريوهات حول العواقب الأخلاقية المُستقبلية. وهناك أساليب مُختلفة لممارسة الأخلاقيات في البحث والابتكار (Reijers et al. 2018)، إحداها ليست فقط دراسة تأثير سرديات الذكاء الاصطناعي الحالية وتقييمها (Royal Society, 2018) ولكن أيضًا خلق سرديات جديدة أكثر واقعيةً حول تطبيقات مُعينة للذكاء الاصطناعي.

النهج المُوجَّه للمُمارسة والنهج التصاعدي: كيف نترجمهما عمليًّا؟

الابتكار المسئول لا يتعلق فقط بتضمين الأخلاقيات في التصميم، ولكنه يتطلَّب أيضًا مراعاة آراء مُختلف الأطراف المَعنية ومصالحهم. وتنطوي الحوكمة الشاملة على إشراك نطاقٍ واسع من الأطراف المَعنية، وإجراء نقاشٍ عام، والتدخل المُجتمعي المُبكر في مرحلة البحث والابتكار (Von Schomberg 2011). وهذا قد يعني، مثلًا، تنظيم مجموعات نقاشٍ مركزة واستخدام تقنيات أخرى لمعرفة رأي الناس في التكنولوجيا.

يتعارض هذا النهج التصاعدي في الابتكار المسئول مع نهج الأخلاقيات التطبيقية الذي يتبعه معظم وثائق السياسات، والذي يميل في الغالب إلى أن يكون نهجًا تنازليًّا ومجردًا. أولًا، يتم إنشاء السياسات غالبًا من قِبل خبراء، دون أن يشارك فيها نطاق واسع من الأطراف المَعنية. ثانيًا، حتى إذا أيَّدت هذه السياسات مبادئ مثل الأخلاقيات المُضمنة في التصميم، فإنها تظلُّ شديدة الغموض فيما يتعلق بما يَعنيه تطبيق هذه المبادئ عمليًّا. ولإنجاح سياسة الذكاء الاصطناعي، يظلُّ التحدي كبيرًا لبناء جسرٍ بين المبادئ الأخلاقية والقانونية المُجردة والعالية المستوى من ناحية، وبين مُمارسات تطوير التكنولوجيا واستخدامها في سياقاتٍ مُعينة، والتقنيات، وأصوات أولئك الذين يشاركون في هذه الممارسات ويعملون في هذه السياقات من ناحية أخرى. يُترك بناء هذا الجسر لَمن تُوجَّه إليهم هذه السياسات المقترحة. فهل يُمكننا القيام بالمزيد في المرحلة الأولى من صنع السياسات، وهل يجِب علينا ذلك؟ نحتاج على الأقل إلى المزيد من العمل على الأساليب والإجراءات والمؤسَّسات التي نحتاجها لجعل أخلاقيات الذكاء الاصطناعي تنجح عمليًّا. ويجب علينا أن نُولي المزيد من الاهتمام للعملية.

الابتكار المسئول لا يتعلق فقط بتضمين الأخلاقيات في التصميم، ولكنه يتطلب أيضًا مراعاة آراء مختلف الأطراف المَعنية ومصالحهم.

فيما يتعلق بالسؤال عمَّن يشارك في وضع أخلاقيات الذكاء الاصطناعي، فإننا نحتاج إلى تطبيق نهج تصاعدي إلى جانب النهج التنازلي، بمعنى الاستماع أكثر إلى الباحثين والمهنيين الذين يتعاملون مع الذكاء الاصطناعي عمليًّا، بل وإلى الأشخاص الذين من المُحتمَل أن يضرَّهم الذكاء الاصطناعي. إذا كنا نؤيد مبدأ الديمقراطية وإذا كان

هذا المفهوم يشمل التضمين والمشاركة في صنع القرار بشأن مُستقبل مجتمعاتنا، فإن سماع صوت الأطراف المَعنية ليس أمرًا اختياريًّا ولكنه إلزامي من الناحيتَين الأخلاقية والسياسية. بينما يشارك بعض صانعي السياسات في نوع من التشاور مع الأطراف المعنية (على سبيل المثال، لدى المفوضية الأوروبية تحالف الذكاء الاصطناعي الخاص بها)،[1] لا يزال من المشكوك فيه ما إذا كانت مثل هذه الجهود تصل حقًّا إلى المُطورين والمستخدمين النهائيين للتكنولوجيا، والأهم من ذلك، إلى أولئك الذين سيتعيَّن عليهم تحمُّل معظم المخاطر والتعايش مع آثارها السلبية. فهل صُنع القرار والسياسات الخاصة بالذكاء الاصطناعي أمرٌ ديمقراطي ينطوي على مشاركة حقًّا؟

إنَّ مفهوم الديمقراطية مُهدَّد أيضًا بحقيقة تركز السلطة في أيدي عددٍ صغير نسبيًّا من الشركات الكبيرة. ويرى بول نيميتز (٢٠١٨) أن تراكُم السلطة الرقمية في أيدي شركات قليلة ينطوي على إشكالية: إذا مارست حفنة من الشركات سُلطتها ليس فقط على الأفراد — من خلال تكوين ملفَّاتٍ تعريفية عنا — ولكن أيضًا على البنية الأساسية للديمقراطية، فإن هذه الشركات، على الرغم من نواياها الحسنة للمساهمة في الذكاء الاصطناعي الأخلاقي، سوف تضع عقباتٍ أمامه. ولذلك، فمِن الضروري وضع لوائح تنظيمية وحدود لحماية المصلحة العامة، ولضمان أن هذه الشركات لن تُشكل القواعد بمفردها. وأشار موراي شاناهان إلى أن «المَيل إلى تركُّز السلطة والثروة والموارد في أيدي عدد قليل يتَّسم بالاستدامة الذاتية» (٢٠١٥، ١٦٦)، مما يجعل من الصعب تحقيق مجتمعٍ أكثر إنصافًا. كما أنه يجعل الأفراد عُرضة لجميع أنواع المخاطر، بما في ذلك الاستغلال وانتهاكات الخصوصية، على سبيل المثال، ما تُسمِّيه دراسة أجراها المجلس الأوروبي «التأثير المُروِّع لإعادة استخدام البيانات» (Yeung 2018, 33).

إذا قارنَّا الوضع مع سياسة البيئة، يُمكن أن نكون مُتشائمين أيضًا بشأن إمكانية أن تتَّخذ البلدان إجراءً فعَّالًا وتعاونيًّا بشأن أخلاقيات الذكاء الاصطناعي. فلنأخُذ، على سبيل المثال، العمليات السياسية المُتعلقة بتغيُّر المناخ في الولايات المتحدة، حيث يتم في بعض الأحيان إنكار مشكلة الاحترار العالمي وتغيُّر المناخ نفسها، وحيث تعمل بعض القوى السياسية ذات النفوذ ضد اتخاذ أي إجراءٍ حيال ذلك، أو النجاح المحدود للغاية لمؤتمرات تغيُّر المناخ الدولية في الاتفاق على سياسة مناخية مشتركة وفعَّالة. وقد يواجِه أولئك الذين يسعون إلى اتخاذ إجراءٍ عالَمي في ظل المشكلات الأخلاقية والمجتمعية التي أثارها الذكاءُ

الاصطناعي صعوبات مُماثلة. فغالبًا ما تتفوق المصالح الأخرى على المصلحة العامة، وهناك ندرة في السياسات الحكومية الدولية الخاصة بالتكنولوجيا الرقمية الجديدة، بما فيها الذكاء الاصطناعي. ومع ذلك، هناك استثناءٌ واحد لذلك وهو الاهتمام العالمي بحظر الأسلحة القاتلة الذاتية التشغيل، التي تحتوي أيضًا على جانب ذكاء اصطناعي. ولكن هذا لا يزال استثناءً، ولا يحظى أيضًا بدعم جميع البلدان (على سبيل المثال، ما زال موضع جدل في الولايات المتحدة).

علاوةً على ذلك، ورغم حسن النية، فإن لكلٍّ من أخلاقيات التصميم والابتكار المسئول قيودهما الخاصة. أولًا، تفترض أساليب مثل التصميم الحساس للقِيَم أنه يُمكننا التعبير عن قِيَمنا، وتفترض جهود بناء الآلات الأخلاقية أننا يمكن أن نُعبِّر بشكلٍ كامل عن أخلاقياتنا. ولكن هذا لا يحدث بالضرورة دائمًا؛ إذ إننا قد لا نستطيع التفكير بوضوحٍ ولا التعبير عن أخلاقياتنا اليومية. ففي بعض الأحيان، نستجيب إلى مشكلات أخلاقية بطريقةٍ مُعينة دون أن نتمكن من تبرير استجابتنا بشكل كامل (Boddington 2017). وكما قال فيتجنشتاين: أخلاقياتنا ليست فقط متجسدة ولكنها مُضمَّنة في شكلٍ من أشكال الحياة. إنها متصلة على نحوٍ عميق بطريقة قيامنا بالأفعال ككائنات متجسدة واجتماعية، وكمجتمعات وثقافات. وهذا يفرض حدودًا على مشروع التعبير الكامل عن الأخلاق والتفكير الأخلاقي. ويمثل أيضًا مشكلة لمشروع تطوير الآلات الأخلاقية، ويتحدى الافتراضات التي تقول إن الأخلاق والديمقراطية يمكن مناقشتهما والتعبير عنهما بالكامل. كما يخلق مشكلة لصانعي السياسات الذين يعتقدون أن أخلاقيات الذكاء الاصطناعي يمكن التعامل معها تمامًا من خلال قائمة من المبادئ أو من خلال أساليب قانونية وتقنية مُحدَّدة. نحن بالتأكيد بحاجةٍ إلى أساليب وإجراءات وعمليات. ولكن كل هذا ليس كافيًا؛ فالأخلاقيات لا تعمل مثل الآلة، وكذلك السياسة والابتكار المسئول.

ثانيًا، يُمكن أن يكون هذان النهجان عائقًا أمام الأخلاقيات عندما يكون من الواجب أخلاقيًّا إيقاف تطوير التكنولوجيا. فغالبًا ما تكون وظيفتهما من الناحية العملية هي تيسير عملية الابتكار، وتعزيز تحقيق الأرباح، وضمان قبول التكنولوجيا. وقد لا يكون هذا بالضرورة سيئًا. ولكن ماذا لو كانت المبادئ الأخلاقية تُشير إلى أنه يجب إيقاف أو تعليق التكنولوجيا، أو تطبيق مُعيَّن من تطبيقاتها؟ اعتبر كروفورد وكالو (٢٠١٦) أن أداتَي التصميم الحسَّاس للقِيَم والابتكار المسئول تعتمدان على افتراض أن التكنولوجيا سيجري تطويرها؛ وتَقلُّ فعاليَّتهما عندما يتعلق الأمر باتخاذ قرار حول ما إذا كان يجب

إنشاء هذه التكنولوجيا من الأساس. على سبيل المثال، في حالة الذكاء الاصطناعي المُتقدِّم مثل تطبيقات تعلُّم الآلة الجديدة، ربما تكون هذه التكنولوجيا لا تزال غير جديرة بالثقة أو لها عيوب أخلاقية خطيرة، وأن بعض تطبيقاتها على الأقل قد يتوجب عدم استخدامها (بعد). وسواء أكان وقف التكنولوجيا هو الحل الأفضل دائمًا أم لا، فإن القضية هي أننا يجِب على الأقل أن نتمتَّع بالحق في طرح السؤال وتقرير ما ينبغي فعله. فإذا كان هذا الحق غائبًا، فسوف يظلُّ الابتكار المسئول ستارًا نُخفي وراءه مواصلة العمل كالمعتاد.

نحو أخلاقيات إيجابية

على الرغم من كلِّ ما قيل، فإن أخلاقيات الذكاء الاصطناعي بشكلٍ عامٍّ لا تتعلَّق بالضرورة بمنع الأشياء (Boddington 2017). هناك عائقٌ آخر يَحُول دون مُمارسة أخلاقيات الذكاء الاصطناعي عمليًّا، وهذا العائق هو أنَّ العديد من الجهات الفاعلة في مجال الذكاء الاصطناعي مثل الشركات والباحثين التقنيِّين لا يزالون يعتبرون الأخلاقيات قيودًا، أو أشياءً سلبية. هذه الفكرة ليست مُضللة بشكلٍ كامل؛ إذ غالبًا ما يجب على الأخلاق أن تُقيِّد، وتَحُد، وتقولَ إن شيئًا ما غير مقبول. وإذا أخذنا أخلاقيات الذكاء الاصطناعي على مَحمل الجد ونفَّذنا توصياتها، فقد نُواجِه بعض التنازُلات، ولا سيَّما على المدى القصير. فقد يكون للأخلاقيات ثمَن لا بد من دفعه؛ سواءٌ على مستوى المال أو الوقت أو الطاقة. ومع ذلك، فمِن خلال تقليل المخاطر، تدعم الأخلاقيات والابتكار المسئول التنميةَ المُستدامة للأعمال التجارية والمجتمع على المدى البعيد. ولا يزال هناك تَحدٍّ في إقناع جميع الجهات الفاعلة في مجال الذكاء الاصطناعي، بمَن فيهم صانعي السياسات، بأن هذا هو الحال فعلًا. لاحظ أيضًا أن السياسة واللوائح التنظيميَّة لا تتعلَّق فقط بحظر الأشياء أو بجعلِها أكثر صعوبةً وتعقيدًا؛ بل يُمكن أن تكون داعمة، من خلال تقديم حوافز، على سبيل المثال.

علاوةً على ذلك، إلى جانب الأخلاقيات السلبية التي تفرض قيودًا، نحن في حاجة إلى توضيح الأخلاقيات الإيجابية وشرحها: لوضع رؤية للحياة الجيدة والمجتمع الجيد. وبينما تلمح بعض المبادئ الأخلاقية المقترحة أعلاه إلى مِثل هذه الرؤية، فلا يزال توجيه المناقشة إلى هذا الاتجاه تحديًّا. كما سبق وذكرنا، لا تتعلَّق المسائل الأخلاقية الخاصة بالذكاء الاصطناعي بالتكنولوجيا فحسب؛ بل تتعلَّق بحياة الإنسان وازدهاره، وتتعلَّق بمُستقبل المجتمع، وربما تتعلق أيضًا بغير البشر، وبالبيئة، وبمُستقبل الكوكب

(انظر الفصل التالي). وهكذا تُعيدنا المناقشات حول أخلاقيات الذكاء الاصطناعي وسياساته من جديدٍ إلى الأسئلة الكبيرة التي يجب أن نطرحها على أنفسنا؛ أفرادًا، ومُجتمعاتٍ، وربما بشرًا. ويمكن للفلاسفة أن يُساعدونا في التفكير في هذه الأسئلة. وبالنسبة إلى صانعي السياسات، يكمُن التحدِّي في تطوير رؤية واسعة للمُستقبل التكنولوجي تتضمَّن أفكارًا حول ما هو مُهم وما هو ذو معنًى وما هو ذو قيمة. على الرغم من أن الديمقراطيات الليبرالية بشكلٍ عامٍّ تتعمَّد تجاهل مثل هذه الأسئلة وتركها للأفراد، ولا تتدخَّل في مثل هذه الموضوعات العميقة مثل ماهية الحياة الجيدة ومِن ثَم فهي «سطحية» (ابتكار سياسي أدَّى إلى تجنُّب بعض أنواع الحروب على الأقل وساهم في الاستقرار والازدهار)، فإنه في ظلِّ التحدِّيات الأخلاقية والسياسية التي تواجِهنا، فإن تجاهل الأسئلة الأخلاقية الأكثر «عمقًا» يُعتبر من قبيل انعدام المسئولية. وينبغي أن تتعلَّق السياسة أيضًا، بما فيها سياسات الذكاء الاصطناعي، بالأخلاقيات الإيجابية.

بشكلٍ عام، لا تتعلَّق أخلاقيات الذكاء الاصطناعي بالضرورة بمنع الأشياء؛ بل نحن في حاجة إلى أخلاقيات إيجابية: لوضع رؤية للحياة الجيدة والمُجتمع الجيد.

ومع ذلك، فالسبيل إلى ذلك من منظور صانعي السياسات، ليس من خلال العمل بشكلٍ فردي وتولِّي دور الملك الفيلسوف كما في فلسفة أفلاطون، ولكن بالعثور على التوازن الصحيح بين التكنوقراطية والديمقراطية التشاركية. الأسئلة التي تُواجهنا هي أسئلة تُهمنا جميعًا؛ وعلينا أن نتشارك جميعًا في الإجابة عنها. لذلك، لا يُمكننا تركها في أيدي فئةٍ قليلة من الأشخاص، سواء أكانوا في الحكومة أم في الشركات الكبيرة. ويُعيدنا هذا إلى الأسئلة حول كيفية إنجاح الابتكار المسئول والمشاركة في سياسات الذكاء الاصطناعي. المشكلة لا تتعلَّق فقط بالسلطة؛ إنها تتعلَّق أيضًا بالخير: الخير للأفراد والخير للمجتمع. إن أفكارنا الحالية حول الحياة الجيدة والمجتمع الجيد — إذا كنا قادِرين على التعبير عنها من الأساس — قد تحتاج إلى نقاشٍ نقدي أعمق بكثير. ودعوني أقترح أنه قد يكون من المُفيد للغرب، على الأقل أن يستكشفوا خيار مُحاولة التعلُّم من أنظمةٍ سياسية أخرى غير غربية وثقافات سياسية أخرى. لا يجوز لسياسة الذكاء الاصطناعي الفعَّالة والمُبررة تجنُّب المشاركة في مثل هذه النقاشات الأخلاقية الفلسفية والسياسية الفلسفية.

تداخُل التخصُّصات وتجاوز التخصُّصات

هناك عوائق أخرى يجب تجاوزها إذا أردْنا جعل أخلاقيات الذكاء الاصطناعي أكثر فعاليةً وأردنا دعم التطوير المسئول للتكنولوجيا، تجنُّبًا لما يُسميه الباحثون التقنيون «شتاءَ» الذكاء الاصطناعي الجديد: إبطاء عملية تطوير الذكاء الاصطناعي والاستثمار فيه. أحد هذه العوائق هو نقص تداخُل التخصُّصات وتجاوز التخصُّصات الكافي. ما زلنا نواجِه فجوة شاسعة في الخلفية والفهم بين المُختصِّين في العلوم الإنسانية والعلوم الاجتماعية من جهة، والمُختصِّين في العلوم الطبيعية والهندسية من جهةٍ أخرى، داخل المُجتمع الأكاديمي وخارجه. حتى الآن، ما زلنا نفتقد الدعم المؤسَّسي لسدِّ الفجوة الواسعة بين هذَين «العالَمَين»، سواء في المجتمع الأكاديمي أو في المجتمع الأوسع. ولكن إذا كنَّا نُريد حقًّا أن نمتلك تكنولوجيا متقدمة أخلاقية مثل الذكاء الاصطناعي الأخلاقي، فيجب علينا أن نُقرِّب بين هؤلاء الأشخاص وبين هذَين العالَمَين، في أقرب وقتٍ ممكن.

ويتطلَّب هذا إحداث تغيير في كيفية إجراء البحث والتطوير — فمثلًا، يجب أن يُشارك فيه ليس فقط الأشخاص التقنيون ورجال الأعمال ولكن أيضًا مُختصُّون في العلوم الإنسانية — وكذلك تغيير كيفية «تعليم» الأشخاص، من الشباب وغيرهم. يجب أن نحرص على أن يُدرك الأشخاص الذين لدَيهم خلفية في العلوم الإنسانية أهمية التفكير في التقنيات الجديدة مثل الذكاء الاصطناعي ويُحاولوا اكتساب بعض المعرفة حول هذه التقنيات وما تقوم به. ومن ناحيةٍ أخرى، يجب جعل العلماء والمهندسين أكثر حساسيةً تجاه الجوانب الأخلاقية والمُجتمعية لتطوير التكنولوجيا واستخدامِها. ومن ثَم عندما يتعلَّمون استخدام الذكاء الاصطناعي، ويُساهمون بعد ذلك في تطوير تكنولوجيا الذكاء الاصطناعي الجديدة، فإنهم لن يرَوا الأخلاقيات موضوعًا هامشيًّا لا يمتُّ بصِلةٍ إلى مُمارساتهم التكنولوجية ولكن يرَونها «جزءًا أساسيًّا» من هذه الممارسات. وعندئذٍ، في الحالة المثالية، ستعني «ممارسة الذكاء الاصطناعي» أو «ممارسة علم البيانات» أن يتمَّ تضمين الأخلاقيات ببساطة بوصفها جزءًا أساسيًّا لا غِنى عنه. على نطاقٍ أوسع، يُمكننا أن نفكر في شكلٍ أكثر تنوُّعًا وشمولية من التعليم أو السرد تتداخل فيه التخصُّصات جذريًّا فيما يتعلق بالأساليب والمناهج، وبالموضوعات، وأيضًا بالوسائط والتقنيات. بعبارةٍ أخرى أوضح، إذا تعلَّمَ المهندسون كيفية العمل باستخدام النصوص وتعلم المُختصون في العلوم الإنسانية كيفية العمل باستخدام أجهزة الكمبيوتر، فسيزداد الأمل في أخلاقيات التكنولوجيا وفي سياسة تصلح للتنفيذ عمليًّا.

مخاطر «شتاء» الذكاء الاصطناعي وخطر الاستخدام اللاواعي للذكاء الاصطناعي

إذا لم يبدأ تنفيذ هذه التوجيهات في السياسة والتعليم على أرض الواقع، وبشكلٍ عام، إذا فشل مشروع الذكاء الاصطناعي الأخلاقي، فإننا لن نُواجِه فقط مخاطر «شتاء» الذكاء الاصطناعي؛ بل إن الخطر الأدهى والأمَرَّ سيكمن في الكارثة الأخلاقية والاجتماعية والاقتصادية التي ستُلَمُّ بنا وسيدفع ثمنها البشر وغير البشر والبيئة. هذا لا يتعلق بالتفرد التكنولوجي، أو بالآلات التي ستدمر العالم، أو بسيناريوهات نهاية العالم الأخرى حول المستقبل البعيد، ولكنه يتعلق بالزيادة البطيئة ولكن المؤكدة في تراكم المخاطر التكنولوجية وما ينجم عنها من تفاقم الضعف البشري والاجتماعي والاقتصادي والبيئي. هذه الزيادة في المخاطر والضعف مرتبطة بالمشكلات الأخلاقية المشار إليها هنا وفي الفصول السابقة، بما فيها الاستخدام الجاهل والمتهور لتقنيات الأتمتة المُتقدِّمة مثل الذكاء الاصطناعي. إن الفجوة في التعليم ربما تزيد من تأثير مخاطر الذكاء الاصطناعي بشكلٍ عام: حتى لو لم تتسبَّب دائمًا في مخاطر جديدة مباشرة، فإنها تُضاعف المخاطر الموجودة بالفعل على نحوٍ استثنائي. حتى الآن، لا يُوجَد ما يُسمى «رخصة قيادة» لاستخدام الذكاء الاصطناعي، ولا يُوجَد تعليم إلزامي لأخلاقيات الذكاء الاصطناعي للباحثين التقنيِّين، ورجال الأعمال، ومسئولي الحكومة وغيرهم من الأشخاص المشاركين في ابتكار الذكاء الاصطناعي واستخدامه وسياساته. هناك الكثير من آلات الذكاء الاصطناعي غير المُروَّضة في أيدي أشخاصٍ لا يعرفون المخاطر والمشكلات الأخلاقية المُرتبطة بها، أو الذين قد تكون لديهم توقعات خطأ بشأن التكنولوجيا. ويكمن الخطر، مرة أخرى، في ممارسة السلطة دون معرفة و(بالتالي) دون مسئولية؛ والأسوأ من ذلك أن يخضع الآخرون إلى هذه السلطة. وإذا كان هناك شرٌّ على الإطلاق، فإنه يُقيم حيثما قالت فيلسوفة القرن العشرين حنة آرنت: في غياب الوعي عن القرارات والعمل اليومي المُمل. وعندما يُفترَض أن الذكاء الاصطناعي غير مُتحيز ويُستخدَم دون فَهم لما يتم القيام به، فإن هذا من شأنه أن يُسهم في تعميق غياب الوعي، ثم في النهاية، في الفساد الأخلاقي للعالم. وتستطيع سياسات التعليم المساعَدة في التخفيف من ذلك وبالتالي المساهمة في جعل الذكاء الاصطناعي جيدًا وذا معنًى.

لا تزال هناك العديد من الأسئلة المُزعجة، وربما المؤلِمة إلى حدٍّ ما، التي غالبًا ما يتم تجاهلها في المناقشات التي تدور حول أخلاقيات الذكاء الاصطناعي وسياساته، ولكنها

تستحقُّ منَّا على الأقل أن نذكُرها هنا، حتى وإن لم نُحللها تحليلًا كاملًا. هل أخلاقيات الذكاء الاصطناعي تتعلَّق فقط بخير البشر وقيمتهم، أم إن علينا أن نراعي أيضًا قِيَم غير البشر وخيرهم ومصالحهم؟ وحتى إذا كانت أخلاقيات الذكاء الاصطناعي تتعلق بشكل رئيسي بالبشر، فهل يمكن أن تكون أخلاقيات الذكاء الاصطناعي ليست بالمسألة الأهم التي يتعيَّن على البشرية الاهتمام بها؟ يقودنا هذا السؤال إلى الفصل الأخير من الكتاب.

الفصل الثاني عشر

تحدِّي تغيُّر المناخ: حول الأولويات وحقبة التأثير البشري

هل يجِب أن تكون أخلاقيات الذكاء الاصطناعي محورها الإنسان؟

على الرغم من أن العديد من المؤلَّفات المُتعلقة بأخلاقيات الذكاء الاصطناعي والسياسات تأتي على ذِكر البيئة أو التنمية المُستدامة، فإنها تؤكِّد على القِيَم الإنسانية وغالبًا ما تتمحوَر حول الإنسان بوضوح. على سبيل المثال، تقول الإرشادات الأخلاقية التي وضعها فريق الخبراء الرفيع المُستوى المعنيِّ بالذكاء الاصطناعي إنه يجب تبنِّي نهجٍ متمحور حول الإنسان للذكاء الاصطناعي «يتمتع فيه الإنسان بمكانةٍ أخلاقية فريدة وراسخة لها أولوية على جميع الأصعدة المدنية والسياسية والاقتصادية والاجتماعية» (European Commission AI HLEG 2019, 10) وقد صاغت الجامعات مثل ستانفورد ومعهد ماساتشوستس للتكنولوجيا سياسات بحثها في سياق الذكاء الاصطناعي المُتمحور حول الإنسان.[1]

غالبًا ما يتم تعريف هذا التمحوُر حول الإنسان فيما يتعلق بالتكنولوجيا بإعطاء الأولوية لخير الإنسان وكرامته على حساب ما قد تتطلَّبه أو تفعله التكنولوجيا. فالتكنولوجيا يجب أن تعود بالفائدة على البشر وأن تخدمهم وليس العكس. ومع ذلك، وكما رأينا في الفصول الأولى، فإن مدى مناسبة هذا التركيز على الإنسان في أخلاقيات الذكاء الاصطناعي ليس واضحًا كما قد يبدو للوهلة الأولى، ولا سيَّما إذا أخذنا في الاعتبار المناهج المؤيدة لتجاوز الإنسانية أو سرديات المنافسة (ما بين الإنسان والتكنولوجيا). وتبين فلسفة التكنولوجيا أن هناك المزيد من الطرق — الأكثر دقةً وتعقيدًا — لتحديد العلاقة بين البشر والتكنولوجيا. علاوةً على ذلك، يُعد النهج المُتمحور حول الإنسان غير

واضح على أقل تقدير، إن لم يكُن مُثيرًا للمشكلات، في ضوء المناقشات الفلسفية حول البيئة والكائنات الحية الأخرى. في فلسفة البيئة وأخلاقياتها، هناك نقاش طويل حول قيمة غير البشر، خاصة الكائنات الحية، وحول كيفية احترام تلك القيمة وهذه الكائنات، وحول المشكلات المُحتملة التي قد تنشأ نتيجة احترام قيمة البشر. وفيما يخصُّ أخلاقيات الذكاء الاصطناعي، فإن هذا يعني أن علينا على الأقل طرح السؤال بشأن تأثير الذكاء الاصطناعي على الكائنات الحية الأخرى والنظر في احتمالية وجود تعارُض بين قِيَم ومصالح البشر وغير البشر.

تحديد الأولويات على النحو الصحيح

يمكن أيضًا القول بوجود مشكلات أخرى أكثر خطورة من تلك التي يُسببها الذكاء الاصطناعي، وأنه من المُهم تحديد أولوياتنا بشكلٍ صحيح. وقد ينشأ هذا الاعتراض من النظر إلى المشكلات العالمية مثل تغيُّر المناخ، التي تُعد وفقًا للبعض المشكلة الأهم التي تحتاج البشرية إلى التصدِّي لها وإيلائها الأولوية نظرًا إلى خطورتها وتأثيرها المُحتمَل على الكوكب كلًّا.

> يُعَد النهج المُتمحور حول الإنسان غير واضحٍ على أقل تقدير، إن لم يكُن مُثيرًا للمشكلات، في ضوء المناقشات الفلسفية حول البيئة والكائنات الحية الأخرى.

بالنظر إلى جدول أعمال الأمم المتحدة للتنمية المُستدامة لعام ٢٠١٥ (الذي يطلق عليه أهداف التنمية المُستدامة)[2] ونظرته العامة إلى القضايا العالمية المتعلقة بما وصفه الأمين العام للأمم المتحدة بان كي-مون «الإنسان والكوكب»، نرى العديد من القضايا العالمية التي تتطلَّب يقظة أخلاقية وسياسية: التفاوت الاجتماعي المُتزايد داخل البلدان وفيما بينها، والحروب والتطرُّف العنيف، والفقر وسوء التغذية، وصعوبة الوصول إلى المياه العذبة، ونقص المؤسسات الفعالة والديمقراطية، وزيادة نِسبة السكان المُتقدِّمين في السن، والأمراض المُعدية والوبائية، ومخاطر الطاقة النووية، ونقص الفرص للأطفال والشباب، وعدم المساواة بين الجنسَين وأشكال التمييز والإقصاء المُختلفة، والأزمات الإنسانية وجميع أنواع انتهاكات حقوق الإنسان، والمُشكلات المتعلقة بالهجرة واللاجئين،

وتغيُّر المناخ والمشكلات البيئية — التي تتعلَّق في بعض الأحيان بتغيُّر المناخ — مثل الكوارث الطبيعية المُتكرِّرة والمُتفاقمة وأشكال تدهور البيئة مثل الجفاف وفقدان التنوع البيولوجي. في ضوء هذه المشكلات الضخمة، هل يجب أن نعتبر الذكاء الاصطناعي أولويتنا الأولى؟ وهل يُشتِّت الذكاء الاصطناعي انتباهنا عن قضايا أكثر أهمية؟

من جهة، يبدو أن التركيز على الذكاء الاصطناعي وغيره من المشكلات التكنولوجية في غير محلِّه عندما يُعاني عدد هائل من البشر ويُعاني العالم بأسره من مشكلاتٍ أخرى كثيرة للغاية. ففي حين أن الناس في أحد أنحاء العالم يُكافحون من أجل الوصول إلى المياه العذبة أو من أجل البقاء على قيد الحياة في بيئاتٍ عنيفة، يقلق آخرون في جزءٍ آخر من العالَم بشأن خصوصيتهم على الإنترنت ويتخيَّلون مُستقبلًا يُحقق فيه الذكاءُ الاصطناعي الذكاءَ الفائق. من الناحية الأخلاقية، يبدو أن شيئًا مُريبًا يحدث، شيئًا يتعلق بالتفاوُت الاجتماعي والظلم العالَميَّين. يجب ألا تغضَّ الأخلاق والسياسات الطرفَ عن مثل هذه المشكلات، التي لا تتعلَّق بالضرورة بالذكاء الاصطناعي على الإطلاق. على سبيل المثال، في البلدان النامية، يُمكن أحيانًا للتكنولوجيا المُنخفضة التكلفة — وليس التكنولوجيا المُتقدمة — المساعدة في حلِّ مُشكلات الناس؛ لأنهم يستطيعون أن يتحمَّلوا تكاليفها ويستطيعون تركيبها وصيانتها.

من جهة أخرى، يمكن أن يُسبب الذكاء الاصطناعي مشكلاتٍ جديدة وأيضًا يعمل على تفاقُم المشكلات القائمة بالفعل في المُجتمعات وفي البيئة. على سبيل المثال، يخشى البعض أن الذكاء الاصطناعي سيوسع الفجوة بين الأغنياء والفقراء، وأنه، مثل العديد من التقنيات الرقمية، سيزيد من استهلاك الطاقة، ويخلق مزيدًا من النفايات. من هذا المنظور، فإن مناقشة أخلاقيات الذكاء الاصطناعي والتعامُل معها ليس تشتيتًا للانتباه ولكنه إحدى الطرق التي يُمكننا من خلالها المساهمة في معالجة مشكلات العالم، بما فيها المشكلات البيئية. ومن ثَم، يُمكننا أن نستخلص أننا بحاجةٍ أيضًا إلى إيلاء الاهتمام للذكاء الاصطناعي: نعم، الفقر والحروب وما إلى ذلك هي مشكلات خطيرة، ولكن الذكاء الاصطناعي يُمكن أيضًا أن يؤدِّي إلى — أو يُساعد على — تفاقم مشكلات خطيرة الآن وفي المُستقبل، ويجب أن يكون في قائمة المشكلات التي تحتاج منا إلى إيجاد الحلول. ومع ذلك، فهذا لا يُجيبنا عن السؤال المتعلق بالأولويات؛ وهو سؤال مُهم على مستوى الأخلاقيات والسياسة على حدٍّ سواء. إن القضية لا تتمثل في وجود إجابات سهلة عن ذلك السؤال؛ بل القضية هي أن هذا السؤال لا يُطرَح حتى في معظم المؤلَّفات الأكاديمية ووثائق السياسات حول الذكاء الاصطناعي.

ففي حين أن الناس في أحد أنحاء العالم يُكافحون من أجل الوصول إلى المياه العذبة أو من أجل البقاء على قيد الحياة في بيئاتٍ عنيفة، يقلق آخرون في جزء آخَر من العالم بشأن خصوصيتهم على الإنترنت.

الذكاء الاصطناعي وتغيُّر المناخ وحقبة التأثير البشري

إحدى أصعب الطرق لطرح السؤال المُتعلق بالأولويات هو التعرُّض لمناقشة مسألة تغيُّر المناخ والموضوعات ذات الصلة مثل حقبة التأثير البشري: «لماذا نقلق بشأن الذكاء الاصطناعي إذا كانت المشكلة الملحَّة هي تغيُّر المناخ وكون مُستقبل الكوكب في خطر؟» أو دعونا نستعير عبارةً من الثقافة السياسية الأمريكية: «إنه المناخ، أيها الغبي!» وسوف أوضِّح هنا هذا التحدِّي وأناقش تداعِياته على التفكير في أخلاقيات الذكاء الاصطناعي.

في حين يرفض بعض المتطرِّفين النتائج العلمية، يُقر العلماء وصانعو السياسات على نطاقٍ واسع بأن تغيُّر المناخ ليس فقط مشكلةً عالمية خطيرة ولكنه أيضًا «أحد أكبر التحدِّيات في عصرنا»، كما هو مذكور في نصِّ أهداف التنمية المُستدامة للأمم المتحدة. وهو ليس مشكلةً مُستقبلية: فدرجة الحرارة العالمية ومستويات البحر ترتفع بالفعل، مما يؤثر على البلدان والمناطق الساحلية المُنخفضة. وقريبًا جدًّا سوف يُضطر المزيد من الناس إلى التعامُل مع عواقب تغيُّر المناخ. ويستنتج الكثيرون من هذا أنه يجب علينا التصرُّف الآن بشكلٍ عاجل للتخفيف من مخاطر تغير المناخ؛ وأنا أقول «التخفيف» لأن العملية ربما قد تجاوزت بالفعل نقطة التوقُّف. إن الفكرة هي أن هذا ليس فقط الوقت المناسب للقيام بشيءٍ ولكن ربما فات الأوان بالفعل لتجنُّب جميع العواقب. وبالمقارنة مع مخاوف مؤيدي تجاوز الإنسانية بشأن الذكاء الفائق، فإن هذه المخاوف مدعومة بشكلٍ أفضل بالأدلة العلمية وحازت دعمًا كبيرًا بين النُّخب المُثقفة في الغرب — التي ضجرت من النزعة الشكية ما بعد الحداثية وسياسات الهوية البيروقراطية — التي ترى الآن سببًا للتركيز على مشكلة يبدو أنها حقيقية للغاية وواقعية للغاية وعالمية للغاية: تغيُّر المناخ يحدث حقًّا ويؤثر على كلِّ شخص وكل شيء في هذا الكوكب. وتدعو حملة جريتا ثونبرج والاعتصامات المناخية، على سبيل المثال، إلى توجيه الاهتمام إلى أزمة المناخ.

«لماذا نقلق بشأن الذكاء الاصطناعي إذا كانت المشكلة الملحَّة هي تغيُّر المناخ وكون مُستقبل الكوكب في خطر؟»

يُستخدَم أحيانًا مفهوم حقبة التأثير البشري لتأطير المشكلة. وهي فكرة طرحها بول كروتزن الباحث في تغيُّر المناخ ويوجين ستورمر عالِم الأحياء، وتنصُّ على أننا نعيش في حقبة جيولوجية زادت فيها قوة البشر على الأرض وعلى نظمها البيئية، مما جعل البشر قوةً جيولوجية. فكِّر في النمو الأُسّي لأعداد البشر والماشية، وفي التوسع العمراني المتزايد، واستنزاف الوقود الأحفوري، والاستخدام الهائل للمياه العذبة، وانقراض الأنواع، وإطلاق المواد السامة، وما إلى ذلك. يعتقد البعض أن حقبة التأثير البشري قد بدأت مع الثورة الزراعية؛ بينما يرى آخرون أنها انطلقت بانطلاق الثورة الصناعية (Crutzen 2006) أو بعد الحرب العالمية الثانية. على أي حال، لقد نشأت قصة جديدة وتاريخ جديد، وربما حتى سردية جديدة. وغالبًا ما يُستخدَم هذا المفهوم في الوقت الحاضر لإثارة القلق بشأن الاحتباس الحراري وتغيُّر المناخ، ولحشد مختلف التخصُّصات (بما في ذلك العلوم الإنسانية) للتفكير في مُستقبل الكوكب.

لا يتبنَّى الجميع هذا المصطلح؛ فهو مصطلح مُثير للجدل حتى بين الجيولوجيين، وقد شكك البعض في تركيزه على أهمية البشر. على سبيل المثال، قد جادلت هاراواي (٢٠١٥) من منظور ما بعد الإنسانية بأن الأنواع الأخرى والعوامل «اللاحيوية» تلعب أيضًا دورًا في البيئة المتحولة. ولكن حتى من دون مفهوم مُثير للجدل مثل حقبة التأثير البشري، فإن تغيُّر المناخ والمشكلات البيئية (الأخرى) ستظلُّ باقية، ويجِب على السياسة التعامُل معها، والأفضل أن يكون ذلك في أقرب وقتٍ ممكن. فماذا يعني هذا بالنسبة إلى سياسة الذكاء الاصطناعي؟

يعتقد العديد من الباحثين أن الذكاء الاصطناعي والبيانات الضخمة يُمكن أن تُساعدنا أيضًا في علاج العديد من مشكلات العالم، بما في ذلك تغيُّر المناخ. وعلى غرار المعلومات الرقمية وتقنيات الاتصالات بشكلٍ عام، يمكن أن يُسهم الذكاء الاصطناعي في التنمية المستدامة وفي التعامُل مع العديد من المشكلات البيئية. ومن المُرجَّح أن يُصبح الذكاء الاصطناعي المُستدام اتجاهًا ناجحًا في البحث والتطوير. ومع ذلك، يمكن أن يجعل الذكاء الاصطناعي الأمور أسوأ فيما يخصُّ البيئة؛ وبالتالي فيما يخصُّنا نحن جميعًا.

ولنتذكَّر مجددًا زيادة استهلاك الطاقة والنفايات. ومن منظور مشكلة حقبة التأثير البشري، فإن المخاطرة تكمُن في أن البشر يمكن أن يستخدموا الذكاء الاصطناعي لإحكام قبضتهم على الأرض، مما سيزيد من حدة المشكلة بدلًا من حلِّها.

هذا يعتبر أمرًا إشكاليًّا بشكلٍ خاص إذا كنا ننظر إلى الذكاء الاصطناعي ليس فقط بوصفه حلًّا ولكن بوصفه الحل الرئيسي. ولنفكر في سيناريو الذكاء الفائق لذكاء اصطناعي يعرف أفضل منا نحن البشر ما هو جيد لنا: ذكاء اصطناعي «حميد» يخدم البشرية من خلال جعل البشر يتصرَّفون لصالحهم ولصالح الكوكب؛ على سبيل المثال، الآلة الإله التي تُعادل تقنيًّا الملك الفيلسوف المذكور في فلسفة أفلاطون. يحل الذكاء الاصطناعي الإله محل الإنسان الإله (Harrari 2015)، ويدير نظام دعم الحياة الخاص بنا ويديرنا. فلحل مشكلات توزيع الموارد، على سبيل المثال، يمكن للذكاء الاصطناعي أن يعمل بوصفه «وحدة خدمة»، يُدير إمكانية وصول البشر إلى الموارد. وستكون قراراته مُستندة إلى تحليله لأنماط البيانات. ويمكن دمج هذا السيناريو مع حلولٍ تكنولوجية مبتكرة مثل الهندسة الجيولوجية. البشر ليسوا الوحيدين الذين يحتاجون إلى الإدارة؛ فالكون كله في حاجة إلى إعادة هندسته. ومن ثَم، يُمكننا استخدام التكنولوجيا لـ «إصلاح» مشكلاتنا ومشكلات الكوكب.

ومع ذلك، فإن هذه السيناريوهات لن تكون فقط مستبدة وتتعدَّى على استقلالية البشر، بل ستساهم أيضًا بشكلٍ أساسي في مشكلة حقبة التأثير البشري نفسها: فالوكالة البشرية المُفرطة، هذه المرة يتم تفويضها من قِبل البشر إلى الآلات، ستُحول الكوكب بأكمله إلى مجرد مَورِد وآلة للبشر. يتم «حل» مشكلة حقبة التأثير البشري من خلال الوصول بها إلى النقيض التكنوقراطي، مما يؤدي إلى عالَمٍ من الآلات يُعامَل فيه البشر أولًا كأطفال يجب رعايتهم وربما في وقتٍ لاحق يتم تجاهلهم تمامًا. وفي هذا النوع من التأثير البشري المُتعلق بالبيانات الضخمة والسيناريو المألوف جدًّا الذي يتم فيه إحلال الآلات محلَّ البشر، نعود مرَّة أخرى إلى سيناريوهات الأحلام والكوابيس.

جنون الفضاء الجديد والإغراء الأفلاطوني

ثمَّة إجابة أخرى على تغيُّر المناخ وحقبة التأثير البشري، والتي هي أيضًا رؤية مُولعة بالتكنولوجيا وربما ترتبط أحيانًا بسرديات تجاوز البشرية، وهي: قد نُدمر هذا الكوكب، ولكن يُمكننا الهرب من الأرض والذهاب إلى الفضاء.

كانت الصورة الأيقونية لعام ٢٠١٨ هي سيارة إيلون ماسك الرياضية طراز تسلا وهي تطفو في الفضاء.[3] ماسك أيضًا لدَيه خُطط لاستعمار المريخ. وهو ليس الشخص الوحيد الذي يُراوده هذا الحلم: فهناك اهتمام مُتزايد بالذهاب إلى الفضاء. وهذا ليس مجرد حلم. إذ تُستثمر أموال طائلة في مشروعات الفضاء. وعلى عكس سباق الفضاء الذي حدث في القرن العشرين، هذه المشروعات يتم دعمُها من قِبل الشركات الخاصة. والمليونيرات المُولعون بالتكنولوجيا ليسوا الوحيدِين المُهتمين بالفضاء، بل إن الفنانين أيضًا شغوفون به بشدة. تُخطط شركة سبيس إكس الخاصة بإيلون ماسك لإرسال فنانين إلى مدار القمر.[4] وتُعد السياحة الفضائية فكرةً أخرى تزداد شيوعًا. فمَن منَّا لا يرغب في الذهاب إلى الفضاء؟ الفضاء مُغرٍ للغاية.

لا يمثل الذهاب إلى الفضاء مشكلةً في حدِّ ذاته. بل إن له فوائد مُحتملة. على سبيل المثال، يمكن أن تساعد الأبحاث في كيفية البقاء على قيد الحياة في بيئاتٍ أكثر تطرفًا في التعامُل مع المشكلات على الأرض، وفي اختبار التقنيات المُستدامة، واتخاذ منظور كوكبي. ضع في اعتبارك أيضًا أن مشكلة حقبة التأثير البشري يُمكن أن تكون ناجمةً عن أن تكنولوجيا الفضاء منذ سنوات طويلة أتاحت لنا رؤية الأرض من بُعد. وبالنظر إلى صورة سيارة ماسك مرةً أخرى: يعتقد بعض الناس أن السيارة الكهربائية حلٌّ من حلول المشكلات البيئية، دون التشكيك في افتراض أن السيارات هي أفضل وسيلة للنقل ودون التفكير في كيفية إنتاج الكهرباء. على أي حال، هناك أفكار مثيرة للاهتمام.

ولكن أحلام الفضاء تُعد إشكاليةً إذا كانت نتيجتها هي إهمال المشكلات الأرضية، وإذا كانت عرضًا من أعراض الحالة التي شخَّصتها حنة أرنت (١٩٥٨) بالفعل عندما كتبت عن البشر: الكثير من التجريد والاغتراب. أشارت حنة إلى أن العِلم يدعم رغبة دفينة في مغادرة الأرض: حرفيًّا، من خلال تكنولوجيا الفضاء (في عصرها، سبوتنيك) وأيضًا من خلال طُرق رياضية تُجردنا وتَعزلنا مما أصِفه بحياتنا الأرضية الفوضوية المُتجسِّدة والسياسية. ومن هذا المنظور، يمكن تفسير أحلام مؤيدي تجاوز البشرية بالذكاء الفائق وبمُغادرة الأرض على أنها تداعِيات لنوع إشكالي من الاغتراب والهروب. إنها الفكر الأفلاطوني وفكر تجاوز الإنسانية في أوضح صوره؛ إن الفكرة هي التغلُّب ليس فقط على قيود الجسد البشري، ولكن أيضًا على قيود ذلك «النظام الداعم للحياة»: أي الأرض نفسها. فالجسد ليس هو السجن الوحيد، بل الأرض نفسها، ومن ثَم علينا أن نهرُب منها.

بالتالي، فإحدى مخاطر الذكاء الاصطناعي هي أنه يُمكِّن هذا النوع من التفكير ويُصبح آلة للاغتراب: أداة لمغادرة الأرض وإنكار حالتنا الوجودية الاعتمادية الضعيفة والجسدية والأرضية. بعبارةٍ أخرى: صاروخ. مرة أخرى، لا تُمثل الصواريخ مشكلة في حدِّ ذاتها. إنما المشكلة هي مزج تقنيات مُعينة مع سرديات مُعينة. فعلى الرغم من أن الذكاء الاصطناعي يمكن أن يكون قوة إيجابية بالنسبة إلى حياتنا الشخصية، والمجتمع، والبشرية، فإن مزيجًا من تعزيز الاتجاهات التجريدية والاغترابية في العلوم والتكنولوجيا مع خيالات تجاوز الإنسانية و«تجاوز الأرض» قد يؤدي إلى مستقبلٍ تكنولوجي مؤذٍ للبشر وللكائنات الحية الأخرى على الأرض. إذا هربنا من مشكلاتنا بدلًا من التعامُل معها — كما في مشكلة تغيُّر المناخ، على سبيل المثال — فقد نفوز بالمريخ (حتى الآن) ولكننا سوف نخسر الأرض.

وكالعادة، هناك جانب سياسي آخَر لهذا الموضوع: إذ يمتلك بعض الناس فرصًا ومالًا وقدرةً أكبر على الهروب مقارنةً بالآخرين. المشكلة ليست فقط في أن تكنولوجيا الفضاء والذكاء الاصطناعي لهما تكلفة حقيقية بالنسبة إلى الأرض وأن كلَّ المال المُستثمَر في مشروعات الفضاء لم يُنفَق على مشكلات الأرض الحقيقية مثل الحروب والفقر؛ بل المشكلة هي أن الأثرياء سيكونون قادِرين على الهروب من الأرض التي يُدمِّرونها، في حين يجب على بقيتنا البقاء على كوكبٍ يستحيل العيش فيه بصورة متزايدة (انظر، على سبيل المثال، زيمرمان ٢٠١٥). ومثل الصواريخ والتكنولوجيا الأخرى، يمكن أن يُصبح الذكاء الاصطناعي أداة لـ «بقاء الأكثر ثراءً»، كما أوضح أحد المعلِّقين (Rushkoff 2018). في الوقت الحاضر، يحدث ذلك بالفعل مع تقنيات أخرى: ففي مدن مثل دلهي وبكين، يُعاني معظم الناس من تلوُّث الهواء، بينما يطير الأثرياء إلى مناطق أقل تلوثًا أو يشترون هواءً نقيًّا باستخدام تقنيات تنقية الهواء. ليس الجميع يتنفَّسون الهواء نفسه. والآن، هل سيُساهم الذكاء الاصطناعي في توسيع هذه الفجوات بين الأثرياء والفقراء، مما يؤدي إلى حياة أكثر كربًا وغير صحية للبعض وحياة أفضل للبعض الآخر؟ هل سيَصرِفنا الذكاء الاصطناعي عن المشكلات البيئية؟ يبدو أن فكرة أن الذكاء الاصطناعي ينبغي أن يسعى إلى تحسين الحياة على الأرض، للجميع وليس لفئة معينة، مع الوضع في الاعتبار أن حياتنا تعتمد على كوكب الأرض، تعد متطلبًا أخلاقيًّا. وقد تعيق بعض سرديات الفضاء تحقيق هذا الهدف بدلًا من أن تساعدنا في تحقيقه.

عودة إلى الأرض: نحو ذكاء اصطناعي مستدام

دعوني أعود إلى المشكلة العملية جدًّا للأولويات والمخاطر الحالية والحقيقية المُتعلقة بتغيُّر المناخ. ماذا يجب أن تفعل أخلاقيات الذكاء الاصطناعي وسياساته في ضوء هذه التحديات؟ وعندما تكون هناك خلافات بشأن قيمة حياة الكائنات غير البشرية، فكيف يُمكن حلها؟ سيتفق معظم الناس على أن تسليم السيطرة إلى الذكاء الاصطناعي أو الهروب من الأرض ليست حلولًا جيدة. لكن ما هو الحل الجيد؟ وهل يُوجَد حل؟ إذا ما أجبنا إجابةً نافعة على هذه الأسئلة، فستقودنا بالضرورة إلى الأسئلة الفلسفية المتعلقة بكيفية تعامُلنا بوصفنا بشرًا مع التكنولوجيا ومع بيئتنا. كما تقودنا أيضًا إلى الفصل المتعلق بالتكنولوجيا: ماذا يمكن أن يفعل الذكاء الاصطناعي وعلم البيانات من أجلنا، وماذا يُمكننا أن نتوقَّع من الذكاء الاصطناعي منطقيًّا؟

من الواضح أن الذكاء الاصطناعي يمكن أن يساعدنا في التصدي للمشكلات البيئية. فلنُفكر مثلًا في تغيُّر المناخ. يبدو أن الذكاء الاصطناعي يستطيع على نحوٍ استثنائي أن يساعدنا في مواجهة مثل هذه المشكلات المعقَّدة. إذ يمكن للذكاء الاصطناعي مساعدتنا في دراسة المشكلة، على سبيل المثال، من خلال اكتشاف الأنماط التي لا يُمكننا رؤيتها في البيانات البيئية، نظرًا إلى كثرة هذه البيانات وتعقيدها. كما يمكن أن يساعدنا في الحلول، على سبيل المثال، من خلال مساعدتنا في التعامل مع تعقيد عمليات التنسيق وفي تنفيذ تدابير مثل تقليل انبعاثات المواد الضارة، كما اقترح فلوريدي وآخرون (٢٠١٨). وعلى نطاق أوسع، يمكن أن يساعد الذكاء الاصطناعي من خلال مراقبة ونمذجة الأنظمة البيئية وتمكين حلول مثل الشبكات الذكية للطاقة والزراعة الذكية، كما اقترحت مُدونة المنتدى الاقتصادي العالمي (Herweijer 2018). ويمكن للحكومات وللشركات أيضًا أن تتولَّى الأمر هنا. على سبيل المثال، استخدمت جوجل بالفعل الذكاء الاصطناعي لتقليل استخدام الطاقة في مراكز البيانات.

ومع ذلك، لا يعني هذا بالضرورة «إنقاذ الكوكب». يمكن للذكاء الاصطناعي أيضًا أن يُسبب مشكلات ويجعل الأمور أسوأ. ولنفكر مرةً أخرى في التأثير البيئي السلبي الذي يمكن أن يُخلفه الذكاء الاصطناعي نظرًا إلى الطاقة والبِنى التحتية والمواد التي يعتمِد عليها. ولنفكر ليس فقط في استخدام الذكاء الاصطناعي ولكن أيضًا في إنتاجه: قد تكون الكهرباء مُنتجَة بطرق غير مستدامة، كما أن إنتاج الأجهزة المدعومة بالذكاء الاصطناعي يستهلك الطاقة والمواد الخام وينتج نفايات. أو فلنفكر في «الدفع الذاتي» الذي اقترحه

فلوريدي وآخرون؛ إذ يقترحون أن الذكاء الاصطناعي قد يُساعدنا في التصرف بطرقٍ بيئية جيدة عن طريق مساعدتنا في الالتزام بخيارنا المفروض ذاتيًّا. ولكن هذا الأمر ينطوي على مَخاطره الأخلاقية الخاصة: فليس من الواضح أنه يحترم استقلال البشر وكرامتهم، كما يدعي الكُتَّاب، وقد يسير في اتجاه الذكاء الاصطناعي الحميد الذي يعتني بالبشر لكنه يُدمر حريتهم ويُساهم في مشكلة حقبة التأثير البشري. وهناك على الأقل خطورة فرض أشكالٍ جديدة من السلطة الأبوية والاستبداد. علاوةً على ذلك، قد يتماشى استخدام الذكاء الاصطناعي لمواجهة تغيُّر المناخ مع النظرة العالمية التي تُحوِّل العالم إلى مجرد مُستودع بيانات ومع الرؤية التي تختزل ذكاء الإنسان إلى معالجة البيانات؛ بل ربما نوع أدنى من معالجة البيانات يتطلَّب التحسين بواسطة الآلات. ومن غير المُرجَّح أن تعيد مثل هذه الرؤى تشكيل علاقتنا بالبيئة بطريقة تُخفِّف التحديات مثل تغيُّر المناخ والمشكلات المشار إليها بمصطلح التأثير البشري.

نواجه أيضًا خطر النزعة للحلول التكنولوجية بمعنى أن الاقتراحات لاستخدام الذكاء الاصطناعي لمعالجة المشكلات البيئية يُمكن أن تفترض أن هناك حلًّا نهائيًّا لجميع المشكلات، وأن التكنولوجيا وحدَها يمكن أن تُجيب عن أصعب أسئلتنا، وأننا يمكن أن نحل المشكلات بالكامل عن طريق استخدام الذكاء البشري أو الاصطناعي. ولكن المشكلات البيئية لا يمكن حلُّها عن طريق الذكاء التكنولوجي والعلمي؛ فهي مرتبطة أيضًا بالمشكلات السياسية والاجتماعية التي لا يمكن التصدي لها بالكامل عن طريق التكنولوجيا وحدَها. كما أن المشكلات البيئية دائمًا ما تكون مشكلاتٍ بشرية. والرياضيات وذُريتها التكنولوجية هي أدوات مُفيدة جدًّا، ولكنها محدودة فيما يتعلق بفهم المشكلات البشرية والتعامُل معها. على سبيل المثال، قد تتعارَض القِيَم. ولن يستطيع الذكاء الاصطناعي بالضرورة أن يُساعدنا في الإجابة عن السؤال حول الأولويات، وهو سؤال أخلاقي وسياسي مُهم يجب أن نترك للبشر الإجابة عنه. وتُعلِّمنا العلوم الإنسانية والاجتماعية أن نكون حذِرين جدًّا بشأن الحلول «النهائية».

علاوةً على ذلك، البشر ليسوا الوحيدِين الذين تُواجههم مشكلات؛ فالكائنات غير البشرية أيضًا تواجهها صعوبات، والتي غالبًا ما تُهمَل في المناقشات الخاصة بمستقبل الذكاء الاصطناعي. وأخيرًا، الرأي القائل بأننا يجِب أن نهرب من الأرض، أو الرؤية العالمية التي تقول إن كل شيءٍ عبارة عن بيانات نستطيع نحن البشر التلاعُب بها بمساعدة الآلات، يمكن أن يؤدِّيا في النهاية إلى توسيع الفجوة بين الأغنياء والفقراء وإلى أشكالٍ أوسع نطاقًا

من الاستغلال والانتهاكات للكرامة الإنسانية، بالإضافة إلى تهديد حياة الأجيال القادمة عن طريق المخاطرة بتدمير ظروف الحياة على كوكبنا. إننا نحتاج إلى التفكير العميق في كيفية بناء مجتمعات وبيئات مُستدامة؛ إننا نحتاج إلى التفكير البشري.

الذكاء والحكمة

ومع ذلك، فطريقة تفكير البشر لها جوانب مُتعددة أيضًا. والذكاء الاصطناعي مرتبط بنوعٍ واحد من أنواع التفكير البشري والذكاء البشري: النوع المعرفي الأكثر تجريدًا. هذا النوع من التفكير قد أثبت نجاحًا كبيرًا، ولكنه له قيوده وهو ليس النوع الوحيد من التفكير الذي يُمكن أو يجب علينا مُمارسته. والإجابة عن الأسئلة الأخلاقية والسياسية المُتعلقة بكيفية العيش، وكيفية التعامُل مع بيئتنا، وكيفية التعامُل بشكلٍ أفضل مع الكائنات الحية غير البشرية تتطلَّب ما هو أكثر من الذكاء البشري التجريدي (على سبيل المثال، الحُجج، والنظريات، والنماذج) أو التعرُّف على الأنماط بواسطة الذكاء الاصطناعي. نحتاج إلى أشخاصٍ أذكياء وآلات ذكية، ولكننا أيضًا بحاجةٍ إلى الحدس والخبرة التي لا يمكن وصفُها بوضوح كامل، ونحتاج إلى التحلِّي بالحكمة العملية والفضيلة استجابةً إلى المشكلات والمواقف المادية ومن أجل تحديد أولوياتنا. قد تستنير هذه الحكمة بالعمليات المعرفية التجريدية وبتحليل البيانات، ولكنها تستند أيضًا إلى التجارب المُتجسِّدة الخاصة بالعلاقات والمواقف التي نمرُّ بها في العالم، وإلى التعامُل مع أشخاص آخرين، ومع المادية، ومع بيئتنا الطبيعية. ومن المُحتمل أن يعتمد نجاحنا في التصدي للمشكلات الكبيرة التي تُواجهنا في عصرنا على مزيج من الذكاء التجريدي — البشري والاصطناعي — والحكمة العملية الملموسة التي تم تطويرها على أساس التجارب والمُمارسات البشرية الملموسة والخاصة بالمواقف، بما في ذلك تجاربنا مع التكنولوجيا. وأيًّا كان الاتجاه الذي سيسير فيه تطوير الذكاء الاصطناعي، فإن البشر وحدَهم هم مَن يُواجِهون تحدِّي تطوير هذا النوع الأخير من المعرفة والتعلم. وعلى البشر أن يتصدَّوا له. فالذكاء الاصطناعي قادر على التعرُّف على الأنماط، ولكن الحكمة لا يمكن تفويضها إلى الآلات.

مسرد المصطلحات

الابتكار المسئول: نهج يميل إلى جعل الابتكار أكثر أخلاقية ومسئولية على الصعيد المجتمعي، وينطوي عادةً على تضمين الأخلاق في التصميم ومراعاة آراء أصحاب الشأن ومصالحهم.

الأخلاقيات الإيجابية: الأخلاقيات المرتبطة بالطريقة التي ينبغي أن نعيش بها (معًا)، وتستند إلى رؤية للحياة الجيدة والمجتمع الجيد. وتتناقض مع الأخلاقيات السلبية، التي تضع قيودًا وتحدد ما ينبغي ألا نفعله.

الأخلاقيات المُضمَّنة في التصميم: نهج لأخلاقيات التكنولوجيا وعنصر أساسي في «الابتكار المسئول» الذي يهدف إلى دمج الأخلاقيات في مرحلة تصميم التكنولوجيا وتطويرها. وفي بعض الأحيان، نُسميها «تضمين القيم في التصميم». ومن المصطلحات المشابهة لهذا المصطلح «التصميم الحسَّاس للقِيَم» و«التصميم المُتماشي مع الأخلاق».

تجاوز الإنسانية: الاعتقاد بأن البشر يجِب أن يُعززوا أنفسهم من خلال التقنيات المُتقدمة، وبهذه الطريقة يتجاوزون حالتهم الإنسانية؛ بمعنى أن الإنسانية يجب أن تنتقِل إلى مرحلة جديدة. وهذه أيضًا حركة دولية.

التحيز: التمييز ضد أو لصالح أفراد بأعيُنهم أو مجموعات بعينها. في سياق الأخلاقيات والسياسة، يُثار السؤال حول ما إذا كان تَحيُّز معين ظالمًا أو غير عادل.

تعلُّم الآلة: آلة أو برنامج يُمكنه أن يتعلم تلقائيًّا: ليس بالطريقة التي يتعلَّم بها البشر، ولكن بناءً على عملية حسابية وإحصائية. يمكن لخوارزميات التعلُّم، من خلال تغذيتها بالبيانات، تحديد الأنماط أو القواعد في البيانات وإجراء توقعات للبيانات المستقبلية.

التعلُّم العميق: شكل من أشكال «تعلم الآلة» يستخدم الشبكات العصبية المكونة من عدة طبقات من «الخلايا العصبية»: وحدات مُعالجة بسيطة مترابطة فيما بينها وتتفاعل.

التفرُّد التكنولوجي: الفكرة التي تقول بأنه ستحين لحظة في تاريخ الإنسان عندما يجلب انفجار في الذكاء الآلي تغييرًا جذريًّا في حضارتنا يجعلنا لا نفهم بعدها ما يحدث.

حقبة التأثير البشري (الأنثروبوسين): الحقبة الجيولوجية الحالية المزعومة التي زادت فيها قوة البشر وتأثيرهم على الأرض ونظمها البيئية، مما جعل البشر قوة جيولوجية.

الذكاء الاصطناعي: الذكاء الذي تُظهره أو تُحاكيه الوسائل التكنولوجية. غالبًا ما يُفترض أن معنى «الذكاء» في هذا التعريف يستند إلى مقاييس الذكاء البشري، ويُقصَد به القدرات والسلوكيات الذكية التي يُظهرها البشر. ويمكن أيضًا أن يُشير المصطلح إلى العلم أو إلى التقنيات، مثل خوارزميات التعلم.

الذكاء الاصطناعي الجدير بالثقة: الذكاء الاصطناعي الذي يمكن للإنسان الوثوق فيه. يمكن أن تُشير شروط هذه الثقة إلى مبادئ أخلاقية (أخرى) مثل الكرامة الإنسانية واحترام حقوق الإنسان، وما إلى ذلك، و/أو إلى العوامل الاجتماعية والتقنية التي تؤثر فيما إذا كان الناس يرغبون في استخدام التكنولوجيا. استخدام مصطلح «الثقة» فيما يتعلق بالتكنولوجيا مُثير للجدل.

الذكاء الاصطناعي الرمزي: الذكاء الاصطناعي الذي يعتمد على التمثيلات الرمزية للمهام المعرفية العليا، مثل التفكير المجرد واتخاذ القرارات. ويمكن أن يستخدم شجرة اتخاذ القرار ويأخذ شكل نظام خبير يتطلب مدخلات من خبراء المجال.

الذكاء الاصطناعي العام: الذكاء المُشابه لذكاء البشر، ويمكن تطبيقه على نطاقٍ واسع بالمقارنة مع الذكاء الاصطناعي المحدود، الذي يمكن تطبيقه على مشكلةٍ أو مُهمة مُعينة فقط. ويُطلق عليه أيضًا الذكاء الاصطناعي «القوي» في مقابل الذكاء الاصطناعي «الضعيف».

الذكاء الاصطناعي القابل للتفسير: الذكاء الاصطناعي الذي يمكن أن يشرح للبشر تصرفاته أو قراراته أو توصياته، أو يمكن أن يوفر معلومات كافية حول كيفية الوصول إلى نتيجته.

الذكاء الاصطناعي المستدام: الذكاء الاصطناعي الذي يُمَكِّن ويساهم في طريقة عيش مستدامة للبشرية ولا يدمر النظم البيئية على الأرض التي يعتمد عليها البشر (وأيضًا العديد من غير البشر).

الذكاء الفائق: الفكرة التي تقول بأن الآلات سوف تتفوَّق على ذكاء الإنسان. ويرتبط الذكاء الفائق أحيانًا بفكرة «انفجار الذكاء الاصطناعي» الذي يُسبِّبه تصميم الآلات الذكية لآلات أكثر ذكاءً.

علم البيانات: علم متعدد التخصصات يستخدم الإحصاءات والخوارزميات وغيرها من الأساليب لاستخراج أنماطٍ مفيدة وذات معنًى من مجموعات البيانات؛ المعروفة أحيانًا باسم «البيانات الضخمة». في الوقت الحالي، يُستخدَم تعلُّم الآلة في هذا المضمار. وبجانب تحليل البيانات، يهتم علم البيانات أيضًا باستخراج البيانات وإعدادها وتفسيرها.

القابلية للتفسير: القدرة على التفسير أو قابلية التفسير. في سياق الأخلاقيات، فإنه يُشير إلى القدرة على الشرح للآخرين لماذا قمتَ بشيء مُعين أو لماذا اتخذت قرارًا بعينِه؛ وهذا جزء مما يَعنيه أن تكون مسئولًا.

ما بعد الإنسانية: مجموعة من المُعتقدات التي تُشكك في الإنسانية، وخصوصًا المكانة المحورية للإنسان، وتوسع دائرة الاهتمام الأخلاقي لتشمل غير البشر.

المسئولية الأخلاقية: يمكن استخدامها كمرادفٍ لمعنى أن يتحلى المرء بالأخلاق، ومن ثَم فإنها تشير إلى تحقيق نتائج جيدة أخلاقيًّا، والالتزام بالمبادئ الأخلاقية، والتمتع بالفضيلة، واستحقاق الثناء، وما إلى ذلك؛ حسب النظرية المعيارية المُفترضة. يمكن للمرء أيضًا أن يتساءل عن الشروط التي بموجبها يمكن إسناد المسئولية إليه. تُعد شروط إسناد المسئولية الأخلاقية هي الوكالة الأخلاقية والمعرفة. وتؤكد نُهُج العلاقات أن المرء يكون دائمًا مسئولًا أمام الآخرين.

المكانة الأخلاقية: المنزلة الأخلاقية التي يتمتَّع بها كيان ما؛ أي كيف ينبغي التعامُل مع هذا الكيان.

الوكالة الأخلاقية: القُدرة على الفعل والتفكير والحُكم واتخاذ القرار الأخلاقي، بدلًا من مجرد وجود عواقب أخلاقية.

ملاحظات

الفصل الأول: أيتها المرآة على الحائط

(1) See https://www.youtube.com/watch?v=D5VN56jQMWM.

(2) See the case of Paul Zilly as told by Fry (2018, 71–72). More details in Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias," ProPublica, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

(3) For example, in 2016 a local police zone in Belgium started using predictive policing software to predict burglaries and vehicle theft (Algorithm Watch 2019, 44).

(4) BuzzFeedVideo, "You Won't Believe What Obama Says in this Video!" https://www.youtube.com/watch?v=cQ54GDm1eL0&fbclid=IwAR1oD0AlopEZa00XHo3WNcey_qNnNqTsvHN_aZsNb0d2t9cmsDbm9oCfX8A.

الفصل الثاني: الذكاء الفائق والوحوش ونهاية العالم بالذكاء الاصطناعي

(1) Some talk of taming or domesticating AI, although the analogy with wild animals is problematic, if only because in contrast to the "wild"

AI some imagine, animals are limited by their natural faculties and can be trained and developed only up to some point (Turner 2019).

(2) It is often suggested that Mary Shelley must have been influenced by her parents, who discussed politics, philosophy, and literature, but also science, and by her partner Percy Bysshe Shelley, who was an amateur scientist especially interested in electricity.

الفصل الثالث: كل ما له علاقة بالبشر

(1) Dreyfus was influenced by Edmund Husserl, Martin Heidegger, and Maurice Merleau-Ponty.

الفصل الرابع: أهي حقًّا مجرد آلات؟

(1) A real-world case of this was the robot dog Spot who was kicked by its developers to test it, something that met with surprisingly empathetic responses: https://www.youtube.com/watch?v=aR5Z6AoMh6U.

الفصل الخامس: التكنولوجيا

(1) See https://www.humanbrainproject.eu/en/.

(2) See, for example, the European Commission's AI High Level Expert Group's (2018) definition of AI.

الفصل السادس: لا تنسَ (علم) البيانات

(1) See http://tylervigen.com/spurious-correlations.

(2) Concrete examples such as Facebook, Walmart, American Express, Hello Barbie, and BMW are drawn from Marr (2018).

الفصل الثامن: لامسئوليةُ الآلات والقرارات غير المُبررة

(1) One could ask, however, if decisions made by AIs really count as decisions, and if so, if there is a difference in the kind of decisions we delegate or should delegate to AIs. In this sense, the problem regarding responsibility of or for AI raises the very question of what a decision is. The problem also connects with issues about delegation: we delegate decisions to machines. But what does this delegation entail in terms of responsibility?

(2) Indeed, this case is more complicated since one could argue that the delegate is then still responsible for that particular task—at least to some extent—and it may not be clear how the responsibility is distributed in such cases.

(3) Note that this was and is not always the case; as Turner (2019) reminds us, there are cases of animals being punished.

الفصل التاسع: التحيز ومعنى الحياة

(1) Thanks to Bill Price for the thought experiment.

الفصل العاشر: السياسات المقترحة

(1) See: https://www.acrai.at/en/.

(2) The resolution can be found here: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html?redirect#title1.

(3) See: https://www.scu.edu/ethics-in-technology-practice/conceptual-frameworks/.

(4) See: https://www.partnershiponai.org/.

(5) See: https://www.blog.google/technology/ai/ai-principles/.

(6) See: https://www.microsoft.com/en-us/ai/our-approach-to-ai.

(7) See: https://www.accenture.com/t20160629T012639Z_w_/us-en/_acnmedia/PDF-24/Accenture-Universal-Principles-Data-Ethics.pdf.

(8) See: https://www.businessinsider.de/apple-ceo-tim-cook-on-privacy-the-free-market-is-not-working-regulations-2018-11?r=US&IR=T.

(9) See: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001.

(10) See: https://www.stopkillerrobots.org/.

(11) See: https://futureoflife.org/ai-principles/.

(12) Consider people such as Batya Friedman and Helen Nissenbaum in the United States, and later Jeroen van den Hoven and others in the Netherlands, who have been championing the ethical design of technology for some time.

(13) See: https://www.tuev-sued.de/company/press/press-archive/tuv-sud-and-dfki-to-develop-tuv-for-artificial-intelligence.

الفصل الحادي عشر: التحديات التي تُواجه صانعي السياسات

(1) See: https://ec.europa.eu/digital-single-market/en/european-ai-alliance.

الفصل الثاني عشر: تحدِّي تغيُّر المناخ: حول الأولويات وحقبة التأثير البشري

(1) See: https://hai.stanford.edu/ and https://hcai.mit.edu.

(2) See: https://sustainabledevelopment.un.org/post2015/transformingourworld.

(3) See: https://www.theguardian.com/science/2018/feb/07/space-oddity-elon-musk-spacex-car-mars-falcon-heavy.

(4) See: https://cosmosmagazine.com/space/why-we-need-to-send-artists-into-space.

قراءات إضافية

Alpaydin, Ethem, 2016, *Machine Learning*, Cambridge, MA: MIT Press.

Arendt, Hannah, 1958, *The Human Condition*, Chicago: Chicago University Press.

Aristotle, 2002, *Nichomachean Ethics*, Translated by Christopher Rowe, with commentary by Sarah Broadie, Oxford: Oxford University Press.

Boddington, Paula, 2017, *Towards a Code of Ethics for Artificial Intelligence*, Cham: Springer.

Boden, Margaret A., 2016, *AI: Its Nature and Future*, Oxford: Oxford University Press.

Bostrom, Nick. 2014, *Superintelligence*, Oxford: Oxford University Press.

Brynjolfsson, Erik, and Andrew McAfee, 2014, *The Second Machine Age*, New York: W. W. Norton.

Coeckelbergh, Mark, 2012, *Growing Moral Relations: Critique of Moral Status Ascription*, New York: Palgrave Macmillan.

Crutzen, Paul J., 2006, "The 'Anthropocene,'" In *Earth System Science in the Anthropocene*, edited by Eckart Ehlers and Thomas Krafft, 13–18. Cham: Springer.

Dignum, Virginia, Matteo Baldoni, Cristina Baroglio, Maruiyio Caon, Raja Chatila, Louise Dennis, Gonzalo Génova, et al. 2018, “Ethics by Design: Necessity or Curse?” Association for the Advancement of Artificial Intelligence. http://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_68.pdf.

Dreyfus, Hubert L., 1972, *What Computers Can’t Do*, New York: Harper & Row.

Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena, 2018, “AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations.” *Minds and Machines* 28, no. 4: 689–707.

Frankish, Keith, and William M. Ramsey, eds. 2014. *The Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press.

European Commission AI HLEG (High-Level Expert Group on Artificial Intelligence). 2019. “Ethics Guidelines for Trustworthy AI.” April 8, 2019. Brussels: European Commission. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top.

Fry, Hannah. 2018. *Hello World: Being Human in the Age of Algorithms*. New York and London: W. W. Norton.

Fuchs, Christian. 2014. *Digital Labour and Karl Marx*. New York: Routledge.

Gunkel, David. 2012. *The Machine Question*. Cambridge, MA: MIT Press.

Harari, Yuval Noah. 2015. *Homo Deus: A Brief History of Tomorrow*. London: Hervill Secker.

Haraway, Donna. 1991. “A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century.” In *Simians,*

Cyborgs and Women: The Reinvention of Nature, 149–181. New York: Routledge.

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. 2017. "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems," Version 2. IEEE, 2017. http://standards.Ieee.org/develop/indconn/ec/autonomous_systems.html.

Kelleher, John D. and Brendan Tierney. 2018. *Data Science*. Cambridge, MA: MIT Press.

Nemitz, Paul Friedrich, 2018. "Constitutional Democracy and Technology in the Age of Artificial Intelligence." *Philosophical Transactions of the Royal Society A* 376, no. 2133. https://doi.org/10.1098/rsta.2018.0089.

Noble, David F. 1997. *The Religion of Technology*. New York: Penguin Books.

Reijers, Wessel, David Wright, Philip Brey, Karsten Weber, Rowena Rodrigues, Declan O'Sullivan, and Bert Gordijn. 2018. "Methods for Practising Ethics in Research and Innovation: A Literature Review, Critical Analysis and Recommendation." *Science and Engineering Ethics* 24, no. 5: 1437–1481.

Shelley, Mary. 2017. *Frankenstein*. Annotated edition. Edited by David H. Guston, Ed Finn, and Jason Scott Robert. Cambridge, MA: MIT Press.

Turkle, Sherry. 2011. *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books.

Wallach, Wendell, and Colin Allen. 2009. *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press.

المراجع

Accessnow. 2018. “Mapping Regulatory Proposals for Artificial Intelligence in Europe.” https://www.accessnow.org/cms/assets/uploads/2018/11/mapping_regulatory_proposals_for_AI_in_EU.pdf.

ACRAI (Austria Council on Robotics and Artificial Intelligence). 2018. “Die Zukunft Österreichs mit Robotik und Künstlicher Intelligenz positive gestalten: White paper des Österreichischen Rats für Robotik und Künstliche Intelligenz.”

“Algorithm and Blues.” 2016. *Nature* 537:449.

AlgorithmWatch. 2019. “Automating Society: Taking Stock of Automated Decision Making in the EU.” A report by AlgorithmWatch in cooperation with Bertelsmann Stiftung. January 2019. Berlin: AW AlgorithmWatch GmbH. http://www.algorithmwatch.org/automating-society.

Alpaydin, Ethem. 2016. *Machine Learning*. Cambridge, MA: MIT Press.

Anderson, Michael and Susan Anderson. 2011. “General Introduction.” In *Machine Ethics*, edited by Michael Anderson and Susan Anderson, 1–4. Cambridge: Cambridge University Press.

Arendt, Hannah. 1958. *The Human Condition*. Chicago: Chicago University Press.

Arkoudas, Konstantine, and Selmer Bringsjord. 2014. “Philosophical Foundations.” In *The Cambridge Handbook of Artificial Intelligence*, edited by Keith Frankish and William M. Ramsey. Cambridge: Cambridge University Press.

Armstrong, Stuart. 2014. *Smarter Than Us: The Rise of Machine Intelligence*. Berkeley: Machine Intelligence Research Institute.

Awad, Edmond, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. 2018. “The Moral Machine Experiment.” *Nature* 563:59–64.

Bacon, Francis. 1964. “The Refutation of Philosophies.” In *The Philosophy of Francis Bacon*, edited by Benjamin Farrington, 103–132. Chicago: University of Chicago Press.

Boddington, Paula. 2016. “The Distinctiveness of AI Ethics, and Implications for Ethical Codes.” Paper presented at the workshop Ethics for Artificial Intelligence, July 9, 2016, IJCAI-16, New York. https://www.cs.ox.ac.uk/efai/2016/11/02/the-distinctiveness-of-ai-ethics-and-implications-for-ethical-codes/.

Boddington, Paula. 2017. *Towards a Code of Ethics for Artificial Intelligence*. Cham: Springer.

Boden, Margaret A. 2016. *AI: Its Nature and Future*. Oxford: Oxford University Press.

Borowiec, Steven. 2016. “AlphaGo Seals 4–1 Victory Over Go Grandmaster Lee Sedol.” *Guardian*, March 15. https://www.theguardian.com/technology/2016/mar/15/googles-alphago-seals-4-1-victory-over-grandmaster-lee-sedol.

Bostrom, Nick. 2014. *Superintelligence*. Oxford: Oxford University Press.

Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age*. New York: W. W. Norton.

Bryson, Joanna. 2010. “Robots Should Be Slaves.” In *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, edited by Yorick Wilks, 63–74. Amsterdam: John Benjamins.

Bryson, Joanna. 2018. “AI & Global Governance: No One Should Trust AI.” United Nations University Centre for Policy Research. *AI & Global Governance*, November 13, 2018. https://cpr.unu.edu/ai-global-governance-no-one-should-trust-ai.html.

Bryson, Joanna, Mihailis E. Diamantis, and Thomas D. Grant. 2017. “Of, For, and By the People: The Legal Lacuna of Synthetic Persons.” *Artificial Intelligence & Law* 25, no. 3: 273–291.

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. 2017. “Semantics Derived Automatically from Language Corpora Contain Human-like Biases.” *Science* 356:183–186.

Castelvecchi, Davide. 2016. “Can We Open the Black Box of AI?” *Nature* 538, no. 7623: 21–23.

CDT (Centre for Democracy & Technology) 2018. “Digital Decisions.” https://cdt.org/issue/privacy-data/digital-decisions/.

Coeckelbergh, Mark. 2010. “Moral Appearances: Emotions, Robots, and Human Morality.” *Ethics and Information Technology* 12, no. 3: 235–241.

Coeckelbergh, Mark. 2011. “You, Robot: On the Linguistic Construction of Artificial Others.” *AI & Society* 26, no. 1: 61–69.

Coeckelbergh, Mark. 2012. *Growing Moral Relations: Critique of Moral Status Ascription*. New York: Palgrave Macmillan.

Coeckelbergh, Mark. 2013. *Human Being @ Risk: Enhancement, Technology, and the Evaluation of Vulnerability Transformations*. Cham: Springer.

Coeckelbergh, Mark. 2017. *New Romantic Cyborgs*. Cambridge, MA: MIT Press.

Crawford, Kate, and Ryan Calo. 2016. “There Is a Blind Spot in AI Research.” *Nature* 538:311–313.

Crutzen, Paul J. 2006. “The ‘Anthropocene.’” In *Earth System Science in the Anthropocene* edited by Eckart Ehlers and Thomas Krafft, 13–18. Cham: Springer.

Darling, Kate, Palash Nandy, and Cynthia Breazeal. 2015. “Empathic Concern and the Effect of Stories in Human–Robot Interaction.” In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO–MAN)*, 770–775. New York: IEEE.

Dennett, Daniel C. 1997. “Consciousness in Human and Robot Minds. In *Cognition, Computation, and Consciousness*, edited by Masao Ito, Yasushi Miyashita, and Edmund T. Rolls, 17–29. New York: Oxford University Press.

Digital Europe. 2018. “Recommendations on AI Policy: Towards a Sustainable and Innovation–friendly Approach.” Digitaleurope.org, November 7, 2018.

Dignum, Virginia, Matteo Baldoni, Cristina Baroglio, Maruiyio Caon, Raja Chatila, Louise Dennis, Gonzalo Génova, et al. 2018. “Ethics by Design: Necessity or Curse?” Association for the Advancement of Artificial Intelligence. http://www.aies–conference.com/2018/contents/papers/main/AIES_2018_paper_68.pdf.

Dowd, Maureen. 2017. “Elon Musk’s Billion–Dollar Crusade to Stop the A.I. Apocalypse.” *Vanity Fair*, March 26, 2017. https://www.vanityfair.com/news/2017/03/elon–musk–billion–dollar–crusade–to–stop–ai–space–x.

Dreyfus, Hubert L. 1972. *What Computers Can’t Do.* New York: HarperCollins.

Druga, Stefania and Randi Williams. 2017. “Kids, AI Devices, and Intelligent Toys.” MIT Media Lab, June 6, 2017. https://www.media.mit.edu/posts/kids-ai-devices/f.

European Commission. 2018. “Ethics and Data Protection.” http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-data-protection_en.pdf.

European Commission Directorate-General of Employment, Social Affairs and Inclusion. 2018. “Employment and Social Developments in Europe 2018.” Luxembourg: Publications Office of the European Union. http://ec.europa.eu/social/main.jsp?catId=738&langId=en&pubId=8110.

European Commission AI HLEG (High-Level Expert Group on Artificial Intelligence). 2018. “Draft Ethics Guidelines for Trustworthy AI: Working Document for Stakeholders.” Working document, December 18, 2018. Brussels: European Commission. https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai.

European Commission AI HLEG (High-Level Expert Group on Artificial Intelligence). 2019. “Ethics Guidelines for Trustworthy AI.” April 8, 2019. Brussels: European Commission. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top.

EGE (European Group on Ethics in Science and New Technologies). 2018. “Statement on Artificial Intelligence, Robotics and ‘Autonomous’ Systems.” Brussels: European Commission.

European Parliament and the Council of the European Union. 2016. “General Data Protection Regulation (GDPR).” https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679.

Executive Office of the President, National Science and Technology Council Committee on Technology. 2016. “Preparing for the Future of Artificial Intelligence.” Washington, DC: Office of Science and Technology Policy (OSTP).

Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena. 2018. "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." *Minds and Machines* 28, no. 4: 689–707.

Floridi, Luciano, and J. W. Sanders. 2004. "On the Morality of Artificial Agents." *Minds and Machines* 14, no. 3: 349–379.

Ford, Martin. 2015. *Rise of the Robots: Technology and the Threat of a Jobless Future*. New York: Basic Books.

Frankish, Keith, and William M. Ramsey. 2014. "Introduction." In *The Cambridge Handbook of Artificial Intelligence*, edited by Keith Frankish and William M. Ramsey, 1–14. Cambridge: Cambridge University Press.

Frey, Carl Benedikt, and Michael A. Osborne. 2013. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" Working paper, Oxford Martin Programme on Technology and Employment, University of Oxford.

Fry, Hannah. 2018. *Hello World: Being Human in the Age of Algorithms*. New York: W. W. Norton.

Fuchs, Christian. 2014. *Digital Labour and Karl Marx*. New York: Routledge.

Goebel, Randy, Ajay Chander, Katharina Holzinger, Freddy Lecue, Zeynep Akata, Simone Stumpf, Peter Kieseberg, and Andreas Holzinger. 2018. "Explainable AI: The New 42?" Paper presented at the CD-MAKE 2018, Hamburg, Germany, August 2018.

Gunkel, David. 2012. *The Machine Question*. Cambridge, MA: MIT Press.

Gunkel, David. 2018. "The Other Question: Can and Should Robots Have Rights?" *Ethics and Information Technology* 20:87–99.

Harari, Yuval Noah. 2015. *Homo Deus: A Brief History of Tomorrow*. London: Hervill Secker.

Haraway, Donna. 1991. “A Cyborg Manifesto: Science, Technology, and Socialist–Feminism in the Late Twentieth Century.” In *Simians, Cyborgs and Women: The Reinvention of Nature*, 149–181. New York: Routledge.

Haraway, Donna. 2015. “Anthropocene, Capitalocene, Plantationocene, Chthulucene: Making Kin.” *Environmental Humanities* 6:159–165.

Herweijer, Celine. 2018. “8 Ways AI Can Help Save the Planet.” *World Economic Forum*, January 24, 2018. https://www.weforum.org/agenda/2018/01/8–ways–ai–can–help–save–the–planet/.

House of Commons. 2018. “Algorithms in Decision–Making.” Fourth Report of Session 2017–19, HC351. May 23, 2018.

ICDPPC (International Conference of Data Protection and Privacy Commissioners). 2018. “Declaration on Ethics and Data Protection in Artificial Intelligence.” https://icdppc.org/wp–content/uploads/2018/10/20180922_ICDPPC–40th_AI–Declaration_ADOPTED.pdf.

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. 2017. “Ethically Aligned Design: A Vision for Prioritizing Human Well–Being with Autonomous and Intelligent Systems,” Version 2. IEEE. http://standards.Ieee.org/develop/indconn/ec/autonomous_systems.html.

Ihde, Don. 1990. *Technology and the Lifeworld: From Garden to Earth*. Bloomington: Indiana University Press.

Jansen, Philip, Stearns Broadhead, Rowena Rodrigues, David Wright, Philp Brey, Alice Fox, and Ning Wang. 2018. “State–of–the–Art Review.” Draft of the D4.1 deliverable submitted to the European Commission on April 13, 2018. A report for The SIENNA Project, an EU H2020 research and innovation program under grant agreement no. 741716.

Johnson, Deborah G. 2006. “Computer Systems: Moral Entities but not Moral Agents.” *Ethics and Information Technology* 8, no. 4: 195–204.

Kant, Immanuel. 1997. *Lectures on Ethics*. Edited by Peter Heath and J. B. Schneewind. Translated by Peter Heath. Cambridge: Cambridge University Press.

Kelleher, John D., and Brendan Tierney. 2018. *Data Science*. Cambridge, MA: MIT Press.

Kharpal, Arjun. 2017. “Stephen Hawking Says A.I. Could Be ‘Worst Event in the History of Our Civilization.’” CNBC. November 6, 2017. https://www.cnbc.com/2017/11/06/stephen-hawking-ai-could-be-worst-event-in-civilization.html.

Kubrick, Stanley, dir. 1968. *2001: A Space Odyssey*. Beverly Hills, CA: Metro-Goldwyn-Mayer.

Kurzweil, Ray. 2005. *The Singularity Is Near*. New York: Viking.

Leta Jones, Meg. 2018. “Silencing Bad Bots: Global, Legal and Political Questions for Mean Machine Communication.” *Communication Law and Policy* 23, no. 2: 159–195.

Lin, Patrick, Keith Abney, and George Bekey. 2011. “Robot Ethics: Mapping the Issues for a Mechanized World.” *Artificial Intelligence* 175:942–949.

MacIntyre, Lee C. 2018. *Post-Truth*. Cambridge, MA: MIT Press.

Marcuse, Herbert. 1991. *One-Dimensional Man*. Boston: Beacon Press.

Marr, Bernard. 2018. “27 Incredible Examples of AI and Machine Learning in Practice.” *Forbes*, April 30. https://www.forbes.com/sites/bernardmarr/2018/04/30/27-incredible-examples-of-ai-and-machine-learning-in-practice/#6b37edf27502.

McAfee, Andrew, and Erik Brynjolfsson. 2017. *Machine, Platform, Crowd: Harnessing Our Digital Future*. New York: W. W. Norton.

Miller, Tim. 2018. “Explanation in Artificial Intelligence: Insights from the Social Sciences.” *arXiv*, August 15. https://arxiv.org/pdf/1706.07269.pdf.

Mouffe, Chantal. 2013. *Agonistics: Thinking the World Politically*. London: Verso.

Nemitz, Paul Friedrich, 2018. “Constitutional Democracy and Technology in the Age of Artificial Intelligence.” *Philosophical Transactions of the Royal Society A* 376, no. 2133. https://doi.org/10.1098/rsta.2018.0089.

Noble, David F. 1997. *The Religion of Technology*. New York: Penguin Books.

Reijers, Wessel, David Wright, Philip Brey, Karsten Weber, Rowena Rodrigues, Declan O’ Sullivan, and Bert Gordijn. 2018. “Methods for Practising Ethics in Research and Innovation: A Literature Review, Critical Analysis and Recommendation.” *Science and Engineering Ethics* 24, no. 5: 1437–1481.

Royal Society, the. 2018. “Portrayals and Perceptions of AI and Why They Matter.” December 11, 2018. https://royalsociety.org/topics-policy/projects/ai-narratives/.

Rushkoff, Douglas. 2018. “Survival of the Richest.” *Medium*, July 5. https://medium.com/s/futurehuman/survival-of-the-richest-9ef6cddd0cc1.

Samek, Wojciech, Thomas Wiegand, and Klaus-Robert Müller. 2017. “Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models.” https://arxiv.org/pdf/1708.08296.pdf.

Schwab, Katharine. 2018. “The Exploitation, Injustice, and Waste Powering Our AI.” *Fast Company*. September 18, 2018. https://www.fastcompany.com/90237802/the-exploitation-injustice-and-waste-powering-our-ai.

Seseri, Rudina. 2018. “The Problem with ‘Explainable AI.’” *Tech Crunch*. June 14, 2018. https://techcrunch.com/2018/06/14/the-problem-with-explainable-ai/?guccounter=1.

Searle, John. R. 1980. “Minds, Brains, and Programs.” *Behavioral and Brain Sciences* 3, no. 3: 417–457.

Shanahan, Murray. 2015. *The Technological Singularity*. Cambridge, MA: The MIT Press.

Siau, Keng, and Weiyu Wang. 2018. “Building Trust in Artificial Intelligence, Machine Learning, and Robotics.” *Cutter Business Technology Journal* 32, no. 2: 46–53.

State Council of China. 2017. “New Generation Artificial Intelligence Development Plan.” Translated by Flora Sapio, Weiming Chen, and Adrian Lo. https://flia.org/notice-state-council-issuing-new-generation-artificial-intelligence-development-plan/.

Stoica, Ion. 2017. “A Berkeley View of Systems Challenges for AI.” Technical Report No. UCB/EECS-2017-159. http://www2.eecs.berkeley.edu/Pubs/TechRpts/2017/EECS-2017.

Sullins, John. 2006. “When Is a Robot a Moral Agent?” *International Review of Information Ethics* 6: 23–30.

Surur. 2017. “Microsoft Aims to Lie to Their AI to Reduce Sexist Bias.” August 25, 2017. https://mspoweruser.com/microsoft-aims-lie-ai-reduce-sexist-bias/.

Suzuki, Yutaka, Lisa Galli, Ayaka Ikeda, Shoji Itakura, and Michiteru Kitazaki. 2015. “Measuring Empathy for Human and Robot Hand Pain Using Electroencephalography.” *Scientific Reports* 5, article number 15924. https://www.nature.com/articles/srep15924.

Tegmark, Max. 2017. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Allen Lane/Penguin Books.

Turkle, Sherry. 2011. *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books.

Turner, Jacob. 2019. *Robot Rules: Regulating Artificial Intelligence*. Cham: Palgrave Macmillan.

Université de Montréal. 2017. "Montréal Declaration Responsible AI." https://www.montrealdeclaration-responsibleai.com/the-declaration.

Vallor, Shannon. 2016. *Technology and the Virtues*. New York: Oxford University Press.

Vigen, Tyler. 2015. *Spurious Correlations*. New York: Hachette Books.

Villani, Cédric. 2018. *For a Meaningful Artificial Intelligence: Towards a French and European Strategy*. Composition of a parliamentary mission from September 8, 2017, to March 8, 2018, and assigned by the Prime Minister of France, Èdouard Philippe.

Von Schomberg, René, ed. 2011. "Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields." A report from the European Commission Services. Luxembourg: Publications Office of the European Union.

Vu, Mai-Anh T., Tülay Adalı, Demba Ba, György Buzsáki, David Carlson, Katherine Heller, et al. 2018. "A Shared Vision for Machine Learning in Neuroscience." *Journal of Neuroscience* 38, no. 7: 1601–607.

Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. 2017. "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation." *International Data Privacy Law, 2017*. http://dx.doi.org/10.2139/ssrn.2903469.

Wallach, Wendell and Colin Allen. 2009. *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press.

Weld, Daniel S. and Gagan Bansal. 2018. "The Challenge of Crafting Intelligible Intelligence." https://arxiv.org/pdf/1803.04263.pdf.

Winfield, Alan F.T. and Marina Jirotka. 2017. “The Case for an Ethical Black Box.” In *Towards Autonomous Robotic Systems*, edited by Yang Gao, Saber Fallah, Yaochu Jin, and Constantina Lekakou (proceedings of TAROS 2017, Guildford, UK, July 2017), 262–273. Cham: Springer.

Winikoff, Michael. 2018. “Towards Trusting Autonomous Systems.” In *Engineering Multi-Agent Systems*, edited by Amal El Fallah Seghrouchni, Alessandro Ricci, and Son Trao, 3–20. Cham: Springer.

Yampolskiy, Roman V. 2013. “Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach.” In *Philosophy and Theory of Artificial Intelligence* edited by Vincent C. Müller, 289–296. Cham: Springer.

Yeung, Karen. 2018. “A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility within a Human Rights Framework.” A study commissioned for the Council of Europe Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence. MSI-AUT (2018)05.

Zimmerman, Jess. 2015. “What If the Mega-Rich Just Want Rocket Ships to Escape the Earth They Destroy?” *Guardian*, September 16, 2015. https://www.theguardian.com/commentisfree/2015/sep/16/mega-rich-rocket-ships-escape-earth.

Zou, James, and Londa Schiebinger. 2018. “Design AI So That It’s Fair.” *Nature* 559:324–326.


هنـــداوي