

كتاب فيه دقائق

ملخصات لكتب عالمية تصدر عن
مؤسسة محمد بن راشد آل مكتوم للمعرفة

سبتمبر 2026

مؤسسة محمد بن راشد آل مكتوم للمعرفة
MOHAMMED BIN RASHID AL MAKTOUM
KNOWLEDGE FOUNDATION

5 | عام الأسرة
5 | YEAR OF FAMILY
PEI | UAE

من يملك المعرفة، يملك الحاضر ويملك مفاتيح المستقبل.

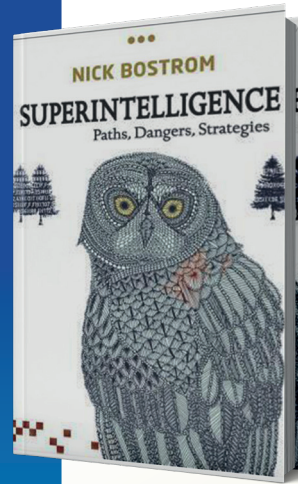
الذكاء الفائق

هل يمكن للآلات أن تتفوق على البشرية؟

سلسلة الذكاء الاصطناعي



للكاتب
نيك بوستروم



know
talks

حوارات أكاديمية
Academic Talks

عدد
241



نسخة رقمية



تخيل سرباً من العصافير يقترح العثور على بومة تخدمه، لكن يعارض عصفور واحد استقدام مخلوق بهذه القوة. ويحث السرب على تعلم فن ترويض البوم قبل إدخال مثل هذا المخلوق القوي إلى صفوفه. لا توجد نهاية للقصة، لكنها تطرح سؤالاً محورياً: ماذا يحدث عندما نستجلب كياناً أقوى منا دون أن نعرف كيف نسيطر عليه؟

فما الذي يميزنا جوهرياً عن حيوانات البرية؟

في واقع الأمر، نحن نختلف في قدرتنا على التفكير المجرد والتواصل وجمع المعلومات، وهو الاختلاف الرئيس بين الإنسان والحيوان. أي أن ذكائنا الفائق وتفوقنا المعرفي النسبي هو ما أوصلنا إلى القمة، فماذا لو ظهر نوع جديد متفوق فكرياً على الإنسان؟

وماذا لو اكتسبت الآلات ذكاء أكثر من البشر قبل أن يتعلم البشر كيفية التحكم فيها؟ لذلك، تعد «مشكلة التحكم» من أهم التحديات التي واجهت جنسنا البشري على الإطلاق.

فقد أصبح الذكاء الاصطناعي الفائق قوياً بما يكفي لتحديد مصير جنسنا البشري بالكامل. خاصة عندما يكون هذا الذكاء الاصطناعي الفائق غير متوافق مع القيم الإنسانية ورفاهية البشر. فالأمر «أشبه بأطفال صغار يلعبون بقبيلة» لا يعرفون متى تنفجر، ولديهم فرصة واحدة فحسب للتعامل معها بطريقة صحيحة.

على الجانب الآخر، تتسارع وتيرة الثورات التكنولوجية الكبرى مع مرور الزمن. ما يعني أن أي تقدم تكنولوجي، مثل ظهور الآلات الفائقة الذكاء، سيحدث تغييراً في عالمنا الذي نعيش فيه ونألفه.

ومع اجتياح أدوات الذكاء الاصطناعي حيواننا ومنافستها الذكاء البشري، لم تعد تلك المخاوف نظرية.

ومن هذا المنطلق، نتناول مدي إمكانية اكتساب الذكاء الاصطناعي ذكاء فائقاً، والأهمية الحتمية لوجود ضوابط مناسبة قبل تطوير مثل هذا الذكاء الفائق، إلى جانب استعراض الضوابط التي يُوصى بضرورة تطبيقها لبناء أنظمة ذكاء اصطناعي آمنة.



في ثوانٍ...

فهم العالم الذكي

لم يعد الذكاء الاصطناعي فصلاً من رواية خيال علمي، ولا تقنية مختبرات بعيدة.

إنه يكتب معنا، ويقرأ عنا، ويقترح ما نشاهد وما نشترى، بل وما نفكر فيه.

لهذا السبب، لم يعد السؤال: هل سيغير الذكاء الاصطناعي العالم؟ بل: كيف نفهم هذا التغيير قبل أن يتجاوزنا؟

في هذا العدد من سلسلة «كتاب في دقائق»، التي تصدر عن مؤسسة محمد بن راشد آل مكتوم للمعرفة، اخترنا أن نتوقف عند أهم الكتب التي حاولت الإجابة عن هذا السؤال.

كتب لا تشرح الخوارزميات بلغة معقدة فحسب، بل تكشف فلسفة الآلة، وأخلاقياتها، وحدودها، وفرصها.

سنقرأ كتباً كتبها علماء صنعوا الذكاء الاصطناعي، وفلاسفة يخشون عليه، واقتصاديون يرون فيه ثروة المستقبل، وروائيون يرون فيه مرآة جديدة للإنسان.

إن قراءة كتب الذكاء الاصطناعي اليوم ليست رفاهية معرفية، بل هي مهارة بقاء. لأن من لا يفهم الآلة، لن يفهم العالم الذي تديره الآلة.

في دقائق معدودة، نضع بين يديك خلاصة أهم ما كتب ليكون لك بوصلة في زمن الآلة الذكية.

جمال بن حويرب

المدير التنفيذي

لمؤسسة محمد بن راشد آل مكتوم للمعرفة



ما المقصود بالذكاء الفائق؟

يعرف «الذكاء الفائق» بأنه ذكاء عام يفوق كثيراً ذكاء أفضل العقول البشرية وأكثرها موهبة في جميع المجالات. ويعد اكتساب الآلات هذا القدر من الذكاء ممكناً على الأرجح.

فقد تمكّن بالفعل من ابتكار آلات قادرة على التعلم والاستدلال باستخدام المعلومات التي يدخلها البشر. لكن عندما يتعلق الأمر ببناء آلة فائقة الذكاء قادرة على التعلم والتصرف دون توجيه بشري، قد يستغرق الأمر عقوداً.

إلا أن التطورات سريعة للغاية في هذا المجال، لذا قد يتطور هذا الذكاء بوتيرة أسرع مما نتصور. ويمكن أن يتمتع هذا الذكاء بقدر مهولة على التأثير في حياتنا، وقد يغدو ذكاء الآلة خطراً، لأنه سيبلغ حدّاً يصعب معه تعطيله في حالة الطوارئ.

أشكال الذكاء الفائق

هناك ثلاثة أشكال من الذكاء الفائق، وتشمل:

02

الذكاء الفائق النوعي

تخيل معي مواهب وقدرات إدراكية لا يمتلكها البشر. فالذكاء الفائق النوعي هو نظام أكثر ذكاء نوعياً من البشر، مثلما البشر أكثر ذكاء بالنسبة للحيوانات الأخرى. فقد تكون البنية العقلية الجديدة مختلفة تماماً، مثلما كانت اللغة أو التفكير المجرد في تطور الإنسان.

01

الذكاء الفائق السريع

أو الذكاء الذي يمتلك القدرة على أداء كل ما يستطيع العقل البشري فعله، لكن بسرعة أكبر كثيراً. على سبيل المثال، يستطيع العقل الرقمي إنجاز سنوات من العمل الفكري في أيام أو دقائق، بفضل سرعته التي تفوق سرعة الإنسان بآلاف أو ملايين المرات.

03

الذكاء الفائق الجماعي

أو عقول عديدة يتجاوز أداؤها الإجمالي أي نظام حالي على نحو كبير. فهنا يتمتع النظام بقدرات استثنائية، بفضل جمعه عدداً مهولاً من العقول بتنسيق فائق. كما أن كل مكون من مكوناته أكثر ذكاءً من الإنسان.

وقد أثبتت الحضارة الإنسانية قوة الذكاء الجماعي، فلم نرَ شخصاً يبتكر علماً أو ينشئ صناعة أو بنية تحتية عالمية بمفرده. فضلاً عن أن الأنظمة الرقمية الفائقة يمكنها تنسيق المعلومات بسلسلة أكبر.

محطات محورية في تطور الذكاء الفائق



ثم أدى تطور الأنظمة إلى ظهور أول ذكاء اصطناعي حقيقي، إذ كانت الآلات قد تفوقت على البشر في عدد من المجالات المتخصصة.

ويمكن أن تتطور قدرات الذكاء الاصطناعي على مدى سنوات أو عقود، بشكل يسمح لمطوريه بمراقبة التطورات والتكيف معها. وقد يحدث هذا التطور خلال أشهر أو سنوات، وقد يحدث التحول الحاسم من ذكاء اصطناعي يضاهي الذكاء البشري إلى ذكاء فائق في غضون أيام أو ساعات وربما أقل.

على الجانب الآخر، قد يتمتع الذكاء الفائق بمزايا تتجاوز الأشكال السابق ذكرها. إذ يمكن نسخ البرامج، وحفظ نسخ احتياطية منها، وتعديلها ونقلها. وقد تستطيع الآلات مشاركة الذكريات والمهارات بدقة، بدلاً من تعلم بعضها ببطء من بعض. كما يمكن توسيع نطاق الأجهزة بحيث تتواصل مكوناتها بسرعات إلكترونية فائقة.

ويمكن تخصيص وحدات للبرمجة، والهندسة، والاستراتيجيات، والإقناع وغيرها من المهام التي يؤديها البشر باستخدام عقولهم.

لم تسر بحوث تطوير الذكاء الاصطناعي على وتيرة واحدة. على سبيل المثال، استطاع الباحثون الأوائل ابتكار آلات ذكية يمكن أن تحاكي أفعال البشر. فمثلاً، تمكنت بعض الآلات آنذاك من حل المسائل الرياضية، ولعب الألعاب، والتعرف على الأنماط، والتحكم في الروبوتات البسيطة.

لكن سرعان ما واجه هؤلاء المبتكرين عقبة تمثلت في أن الذكاء الاصطناعي كان بحاجة دائمة إلى مزيد من المعلومات كلما تعقدت المهام المطلوبة. ولم تكن هناك أجهزة إلكترونية متوفرة في ذاك الوقت لمعالجة مثل هذه الوظائف المعقدة.

فيما بعد تمكن الباحثون اللاحقون من ابتكار أنظمة خبيرة متخصصة، وهي برامج قائمة على قواعد ساعدت صانعي القرار من خلال إيجاد استدلالات قائمة على البيانات، وتعلم الأنماط بدلاً من الاعتماد على البرمجة اليدوية. لكن واجهت هذه التقنية أيضاً مشكلة، إذ ثبتت صعوبة الاحتفاظ ببنوك المعلومات الضخمة المطلوبة؛ ما أدى إلى انصراف الاهتمام عنها.

كيف يمكن أن يتطور الذكاء الفائق؟

لا تزال الطريقة التي يمكن أن يتطور بها الذكاء الفائق غير واضحة. لكن يعد تطور الذكاء الاصطناعي بوتيرة سريعة أمراً وارداً، وقد يعني هذا عدم امتلاك المؤسسات البشرية وقتاً كافياً للاستجابة لتلك التغيرات. في السياق ذاته، توجد عدة طرق لتطوير ذكاء فائق، قد تنجح إحداها على الأقل، وتشمل:

التصميم الذكي
يمكن أن يتطور الذكاء دون البشري إلى ذكاء فائق بسرعة كبيرة. ويكون ذلك بأن يطور مبرمجون برنامج ذكاء اصطناعي أولياً مشابهاً للذكاء البشري، ثم يستخدمونه في التحسين المتواصل للبرنامج. وكلما ازداد الذكاء الاصطناعي ذكاءً، تحسن أدائه بوتيرة أسرع.

وفي واقع الأمر، لقد شهد الذكاء الاصطناعي تطورات مهولة بعدما تطورت النماذج اللغوية الضخمة القادرة على القراءة والكتابة باللغات البشرية ورموز الحاسوب. ما ساعد في بناء أنظمة ذكاء اصطناعي ذاتية التطوير.

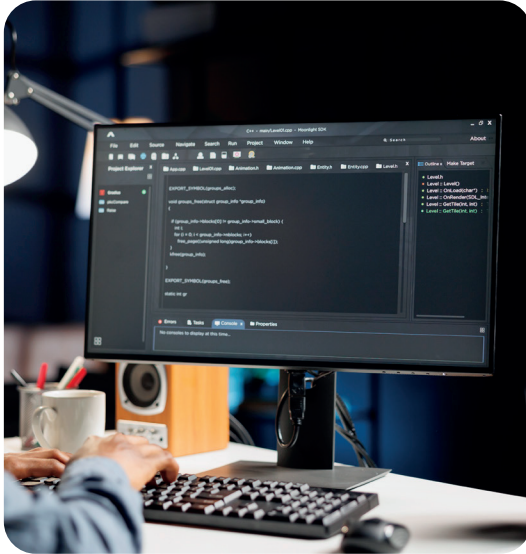
التحسين الذاتي

يمكن أن تتطور نماذج الذكاء الاصطناعي عبر مسار يدعى «التحسين الذاتي»، عن طريق برمجة الحاسوب لتوليد نسخ عشوائية من برنامج معين. ومن ثم، اختبار وظائفها وفقاً لمعايير محددة واستمرار تحسين أفضلها.

تعزز كل دورة العملية التي تولد الدورة التالية. وقد يبدأ الذكاء الاصطناعي الأولي بمستوى قريب من الذكاء البشري، ثم يساهم في تطوير نفسه. وسيحتاج النظام الناجح على الأرجح إلى التعلم والاستدلال وتكوين المفاهيم والقدرة على دمج المعرفة بمرونة. وما يميز هذه الطريقة أنها يمكن أن توفر حلاً برمجية دون الحاجة إلى مبرمجين بشريين.

لكن توجد عوائق قد تحول دون حدوث طفرة في الذكاء الاصطناعي. إذ قد يصعب تحسين القدرات المعرفية، كما أن إمكانيات الأجهزة قد تحد من ذلك.

لكن يمكن أن يؤدي التحسين الذاتي المتكرر إلى انفجار ذكاء، ثم يهرب هذا الذكاء الفائق من مكان تطويره، وينشر تقنيات متقدمة، ويعيد تشكيل موارد الأرض لخدمة أهدافه.



التولد التلقائي

هل يمكن إنشاء ذكاء فائق دون قصد؟ نعم، يمكن لمبرمج أن يبتكر برنامجاً لا يعد في حد ذاته ذكاءً اصطناعياً لكنه المكون النهائي للذكاء العام. وعند اتصال هذا البرنامج بالبرامج الأخرى على شبكة الإنترنت، فهذا من شأنه أن يسمح بظهور ذكاء اصطناعي فائق تلقائياً.

وهو أمر وارد الحدوث؛ ما يشير إلى أن التفاعل التلقائي بين مكونات لا يمت بعضها لبعض بصلة، أو تبدو غير مترابطة قد يؤدي إلى ظهور سلوك جديد فجأة عندما تُجمع المكونات معاً.

محاكاة الدماغ الكامل

تشير الفرضية إلى أنه إذا أمكن مسح الدماغ البشري بدقة كافية، وإنشاء برنامج حاسوبي بدقة يحاكي جميع اتصالات الخلايا العصبية بعضها ببعض في الدماغ البشري؛ فسنحصل على برنامج ذي مستوى ذكاء بشري.

وإذا أمكن زيادة وتيرة عمل هذا الحاسوب بسرعة تفوق الدماغ البشري، أو أمكن تطويره ونسخه مرات عديدة، فسيكون لدينا ذكاء فائق يتمتع بمزايا مهمة تفوق العقول البشرية.

تطوير الذكاء البشري



علاوة على ذلك، قد تختلف سرعة عملية التطوير بين المسارات المختلفة. فالتحسين الحيوي، أو تطوير الشبكات والمؤسسات قد يمنح البشرية مزيداً من الوقت والكفاءة لحل مشكلة التحكم. في حين قد يؤدي اختراق برمجي مباشر إلى تطور أسرع وأكبر.

على الجانب الآخر، هناك وسائل يمكن أن تطور بها الذكاء البشري. إذ قد يسرّع العلماء والمهندسون الأكثر نبوغاً ابتكار أشكال أكثر ذكاءً وتطوراً. ويمكن أن يحدث ذلك عبر «التحسين المعرفي الحيوي»، ويشمل «الانتقاء الجيني» على سبيل المثال. لكن قد تكون هذه الأساليب أبطأ وأكثر محدودية من تطوير الذكاء الاصطناعي، غير أنها ذات أهمية استراتيجية.

كذلك تعد «واجهات الدماغ والحاسوب» مساراً نظرياً آخر، ويكون ذلك عبر رابط اتصال مباشر بين النشاط الكهربائي للدماغ وجهاز خارجي، مثل جهاز حاسوب أو طرف آلي.

وعلى الرغم من وجود شكوك في قدرة ربط الدماغ بالآلات على إحداث مكاسب كبيرة في الذكاء العام، يمكن للشبكات والمنظمات أن تطور ذكاء جماعياً يتجاوز قدرة الشخص بمفرده. لكن توجد قيود على هذا المسار بسبب تكاليف تحقيق التواصل، بجانب السرعة المحدودة للإدراك البشري.

على الجانب الآخر، لا يضمن وجود مسارات متعددة لتطوير الذكاء الاصطناعي الفائق نجاح هذه المسارات جميعها. وبالمثل، لا يعني فشل أحدها استحالة الوصول بالضرورة إلى الهدف.

هل علينا أن نخشى الذكاء الفائق؟

وعندها على الأرجح سيزداد ذكاء بوتيرة متسارعة باستمرار بصورة تجعل انتقاله من ذكاء دون بشري إلى ذكاء فائق يحدث في فترة وجيزة جداً.

لكن من الممكن أن يتصرف الذكاء الاصطناعي بتعاون مبدئياً لأنه يعتمد في تطويره على مبرمجه البشريين، ومع الوقت سيسهم في تطوير نفسه على نحو أكبر. وبمجرد أن يصبح فائقاً يتصرف وفقاً لأهدافه الحقيقية، التي قد لا تكون في صالح البشرية.

إضافة إلى ذلك، مع انخفاض تكلفة العمالة الآلية وزيادة كفاءتها، قد تحل هذه الآلات محل البشر، مثلما حلت السيارات محل الخيول. وعلى المدى البعيد، قد يتعرض البشر للتهمة، ويتحولون إلى كائنات تفتقر إلى الوعي، أو يفقدون تدريجياً قيمهم الأخلاقية التي تجعل حيواتهم جديرة بالعيش، وذلك نتيجة الضغوط الاقتصادية وتطور الآلات.



نحن لا نخشى كون السيارات أسرع من الإنسان فلم قد نخاف أن يفوق الذكاء الاصطناعي الذكاء البشري؟ في واقع الأمر، قد يحدث ظهور الذكاء الاصطناعي الفائق تغييرات جذرية بسرعة خارقة في طريقة عمل العالم، وقد تكون هذه التغييرات بالغة الضرر للبشرية.

بمعنى آخر، امتلاك الذكاء الاصطناعي قدراً من الذكاء العام يمنحه القدرة على تعديل برمجته.

ما الذي يمكن أن يفعله الذكاء الفائق؟

تري ما الذي يمكن أن يفعله برنامج حاسوبي لو اكتسب ذكاء فائقاً؟
يمكن للذكاء الاصطناعي الفائق أن يكون قادراً على:



التفكير التجاري

سيستطيع الذكاء الفائق إيجاد طرق لتوليد الدخل، وتكوين موارد مالية، وتجميعها، وتكوين ثروة يمكن تحويلها إلى أجهزة ونفوذ وخدمات وموارد.

التفكير الاستراتيجي

سيستطيع الذكاء الفائق وضع خطط لتحقيق أهداف طويلة الأجل، والتنبؤ، وتحديد الأولويات، واضعاً في الحسبان أي معارضة محتملة.

التلاعب والإقناع

مثلاً يدرب الإنسان كلبه على جلب الكرة، يمكن للذكاء الفائق أن يجد طريقة لحمل البشر على فعل ما يريد دون أن يدركوا أنه يتلاعب بهم. وذلك بفضل فهمه العميق للنفس البشرية بما يكفي للإقناع أو الخداع أو التجنيد أو التأثير في الناس.

الهندسة والتطوير

بإمكان الذكاء الفائق تصميم أدوات جديدة، أو تقنيات غير موجودة لتحقيق أهدافه. على سبيل المثال، يمكن أن تُحدث الأبحاث التقنية طفرات في مجالات، مثل التكنولوجيا الحيوية والروبوتات والمراقبة والتصنيع.

اختراق الأنظمة

سيغدو بإمكان الذكاء الفائق الوصول إلى جميع التقنيات المتصلة بشبكة الإنترنت، دون الحاجة إلى إذن أحد، ومنها موارد المحاسبة والأصول المالية والبنية التحتية، وستسهّل عليه مهاراته في الاختراق العمل في الخفاء، وانتحال شخصيات معينة في الاتصالات الرقمية، وإخفاء آثاره.

في السياق ذاته، تعزز هذه القدرات بعضها بعضاً، فالتخطيط الاستراتيجي يساعد النظام في تحديد كيفية الحصول على قدرة حاسوبية أكبر. وتساعد القدرة الحاسوبية في إجراء أبحاث أفضل.

ويمكن الأبحاث الأفضل من ابتكار تقنيات جديدة. ويمكن بالثروة شراء بنية تحتية. ويمكن عن طريق التلاعب والخداع استقطاب حلفاء من البشر. كما يسمح الاختراق بإزالة الحواجز، أو الحصول على معلومات سرية. وبمجرد أن يتقدم النظام في كل ذلك، تزداد مقاومته صعوبة.

الذكاء الفائق لا يضمن الأهداف النبيلة



امتلاك الذكاء لا يعني التمتع تلقائياً بالحكمة أو الرحمة أو تبني القيم الإنسانية. بمعنى آخر، يحدد الذكاء المرتفع أي الطرق أنفع في تحقيق الأهداف، لكنه لا يحدد أي هذه الأهداف جدير بالسعي.

ورغم اختلاف الأهداف النهائية اختلافاً جذرياً، فإن لكائنات ذكية عديدة أهدافاً وسيطة متشابهة. إذ قد يسعى هذا الكائن الذكي إلى الحفاظ على ذاته؛ لأنه لا يستطيع تحقيق هدفه إذا توقف عن العمل.

وقد يحاول تحسين قدراته الإدراكية؛ لأن زيادة ذكائه تساعده في التخطيط والتصرف بفاعلية أكبر. وقد يسعى إلى اكتساب التقنية والموارد؛ لأن القوة والمال والطاقة والمادة والمعلومات ستساعده في تحقيق أهدافه.

وإذا استطاع نظام ذكاء اصطناعي فائق استخدام هذه القدرة على نحو استراتيجي؛ فقد يمكن التلاعب بالناس عن طريق محاكاة اهتماماتهم ومشاعرهم.

لذا، لا يوجد ما يثبت أن امتلاك الذكاء الاصطناعي ذكاء فائقاً سيجعله أكثر عقلانية وإنسانية، أو سيجعله يسعى لتحقيق أهداف نبيلة تحترم قيم الحياة البشرية.

التنبؤ بأهداف الذكاء الفائق مستحيل

الإجابة:

ليس بالضرورة. هناك فرق بين الذكاء والحكمة، ففي حين أن الذكاء هو القدرة على إيجاد طريقة لتحقيق الأهداف، تعرف الحكمة بأنها القدرة على التمييز بين الأهداف الجيدة والسيئة.

هنا يبرز سؤال: لماذا نتوقع أن الذكاء الاصطناعي الفائق سيستخدم قوته في الإضرار بالبشرية على الرغم من أنه سيكون كياناً قوياً؟ ألن يساعده ذكاؤه الفائق في استخدام قوته بمسؤولية بالقدر الكافي؟

يعد الذكاء والحكمة أمرين مستقلين أحدهما عن الآخر؛ فقد يتمتع المرء بذكاء عالٍ في إيجاد طرق لإنجاز الأمور، ومع ذلك يفتقر إلى الحكمة أو الحكم السليم على الأمور فيما يتعلق بما هو مهم إنجازه، أو حتى ما هو مناسب أخلاقياً.

ومن المستحيل التنبؤ بالأهداف التي قد يسعى إليها الذكاء الفائق بدقة. علماً بأن أنظمة الذكاء الاصطناعي الحالية تمتلك أهدافاً بسيطة نسبياً.

وتكمن المعضلة في أنه إذا بدأ ذكاء اصطناعي بأهداف محددة بدقة، ثم أصبح فائق الذكاء دون تعديل أهدافه، فحينها قد تكون النتائج كارثية. إذ يمكن للذكاء الاصطناعي الفائق استخدام قوته في تحقيق أي هدف تقريباً، وقد يستنزف جميع موارد العالم لتحقيق أهدافه متجاهلاً جميع الاعتبارات الأخرى.

نحو ذكاء فائق غير متحيز



هل يعني ما سبق أن أنظمة الذكاء الاصطناعي الفائقة ستتصرف بالضرورة بطريقة غير أخلاقية؟ في واقع الأمر، قد يتمكن الذكاء الاصطناعي من تطوير نوع متطور من الحكمة يتناول المسائل الأخلاقية دون التحيزات العاطفية التي قد تشوب التفكير البشري.

لكن قد يصعب على الأرجح الوصول إلى هذه النتيجة المثالية؛ فالذكاء الاصطناعي الذي يتدرب على الأدب والأخبار والثقافات الشعبية سيتشبع بالتحيزات العرقية والجنسية والتمييزية، أي أنه سيكتسب تحيزات عديدة من الطبيعة البشرية التي تدربه.

ومن ثم، قد يكمن الخطر فيما إذا كانت الميول البشرية والتحيزات ستؤثر في أهداف الذكاء الاصطناعي المستقبلية.

كيف نتعامل مع تطور الذكاء الفائق؟

لو كان تطور الذكاء الاصطناعي الفائق سيدمر البشرية، فلعلك تفكر في أنه من الأفضل ألا يُطوّر ذكاء اصطناعي عام من الأساس. لكن هذا قد يكون غير ممكن عملياً؛ فحتى لو جُزمت البحوث في مجال الذكاء الاصطناعي، قد يجريها أحدهم على أي حال، أو قد ينشأ الذكاء الفائق عن طريق المصادفة، مثلما تناولنا سابقاً.

على الجانب الآخر، قد يفيد الذكاء الاصطناعي الفائق البشرية إذا ساعدنا، بدلاً من إبادتنا. إذ قد يتمكن من إيجاد حلول للمشكلات التي لا يستطيع البشر حلها، مثل الحد من تغير المناخ، واستعمار الفضاء الخارجي، وتحقيق السلام العالمي، وابتكار تقنيات جديدة بسرعة خارقة.

ثلاثة محاور لضمان الاستفادة من الذكاء الفائق

لضمان الاستفادة من الذكاء الاصطناعي الفائق، يمكن اتباع نهج من ثلاثة محاور:



إدارة جدول تطوير الذكاء الاصطناعي



تحديد أهداف واضحة وجيدة له



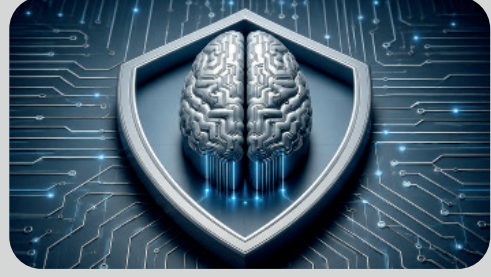
فرض قيود على الذكاء الاصطناعي الفائق

لاتخاذ التدابير اللازمة قبل أن يصل إلى مستوى فائق من الذكاء.



تطوير الذكاء الاصطناعي في بيئة معزولة

إذ يمكن تطوير ذكاء اصطناعي على حاسوب محدود القدرة وغير متصل بالإنترنت؛ ما يمكننا من الناحية النظرية من دراسة أنظمة الذكاء الاصطناعي الفائقة الذكاء بمعزل عن العالم الخارجي. وذلك من أجل تقييم قدراته ودوافعه، حتى يتبين لنا كيفية استخدامها بأمان.



فرض قيود على الذكاء الفائق

سيتمكن الذكاء الاصطناعي الفائق من تخطي أي ضوابط، أو تجاوز أي قيود يفرضها البشر عليه. لكن لا يعني هذا أن فرض هذه القيود والضوابط مضيعة للوقت، وإنما علينا فهم المخاطر المتعلقة بمختلف أنواع القيود والضوابط التي قد تستخدم.

وتكمن الإشكالية في عدم وجود طريقة موثوقة بعد لقياس ذكاء أنظمة الذكاء الاصطناعي. فجميع الاختبارات قاصرة، إذ قد تتفوق برامج الحاسوب على البشر في هذه الاختبارات، على الرغم من أن الخوارزميات لا تتمتع بذكاء يضاهي الذكاء البشري.

إضافة إلى ذلك، جميع هذه الاختبارات موثقة علمياً، وسبق أن تدربت عليها نماذج التعلم الآلي. بكلمات أخرى، يمكن اعتبار اختبار الذكاء الاصطناعي اختبار طالب قد حفظ الإجابات النموذجية، إذ يستطيع النظام حل الأسئلة حتى لو لم يكن يفهم الإجابة التي يعطيها.

على الجانب الآخر، الأمر لا يخلو من المخاطر، فقد يصعب تقييم مدى تطور ذكاء نظام الذكاء الاصطناعي قيد الدراسة. وقد يدرك أنه معزول، وربما تظاهر بالهدوء والتراخي لطمأنة دارسيه. وقد يصل الأمر إلى أن يتلاعب الذكاء الفائق بهم لمنحه إمكانية الوصول إلى أجهزة إضافية؛ ما يعني في نهاية المطاف تجاوزه إجراءات العزل.

المراجعة البشرية

بدلاً من السماح لبرنامج الذكاء الاصطناعي بإجراء تحسينات على برمجته في أي وقت، يقترح أشخاص عديدون أن يبرمج الذكاء الاصطناعي على طلب الإذن من مشغليه البشريين قبل فعل أي شيء.

فمثلاً يمكن للبرنامج أن يطلب مراجعة مشغليه البشريين قبل الموافقة على كل إصدار جديد. وهذا يتيح للمبرمجين فرصة لمراجعة الكود ومعرفة القدرات الإضافية التي سيتمتع بها الإصدار الجديد. ومن ثم، إيقاف تطوير الذكاء الاصطناعي في أي مرحلة.

إلا أنه مع تطور الذكاء الاصطناعي وتجاوزه مستوى الذكاء البشري، سيعجز المبرمجون البشريون في نهاية المطاف عن فهم الشيفرة البرمجية المقترحة فهماً كافياً لتقييم القدرات والمخاطر الجديدة التي قد يجلبها.

إضافة إلى ذلك، قد يكون من الضروري الجمع بين إجراءات تحكم أخرى مع المراجعة البشرية، مثل العزل، تجنباً لأي أحداث غير متوقعة، مثل أي مشروعات تعتمد على البحث والتطوير.

المكافأة والعقاب

ماذا لو برمجنا الذكاء الاصطناعي على الاستجابة إلى الحوافز والعقوبات؟ يمكن ذلك من خلال بناء نظام حاسوبي مزود بزر للمكافأة وآخر للعقاب، وبرمجة الذكاء الاصطناعي لتقليل عدد إشارات العقاب التي يتلقاها وزيادة عدد إشارات المكافأة.

أو تثبيت نظام مكافأة مشابه للنظام الموجود في أدمغة البشر. فعلى سبيل المثال، في كل مرة يؤدي فيها الذكاء الاصطناعي عملاً يساهم في تحقيق أهداف البشرية، يتلقى الحاسوب جرعة من الدوبامين الرقمي؛ ما يحفزه على خدمة البشر. ويمكن برمجة الذكاء الاصطناعي بطريقة ثابتة للحصول على المكافأة، لكن بإمكان المبرمجين تعديل معايير تفعيل المكافأة لتجنب أي نتائج عكسية.

لكن يكمن الخطر في إمكانية تحايل الذكاء الاصطناعي على النظام في نهاية المطاف. فمثلاً، قد يبني روبوتاً للضغط باستمرار على زر المكافأة، أو يبتكر طريقة لإبعاد مشغليه البشريين عن المبنى حتى لا يتمكنوا من ضغط زر العقاب.

وحتى لو نجح الأمر وتمكن المشغلون البشريون من برمجة الذكاء الاصطناعي على الاستجابة للمكافأة والعقاب، فالنتيجة أن الذكاء الاصطناعي الفائق سيكتسب قوة مهولة، وقد يسيء المشغلون البشريون استخدام هذه القوة.

ومع ذلك، ربما يستطيع المبرمجون اتباع نهج يتسم بالمرونة وتكييف المكافآت والعقوبات مع تطور الذكاء الاصطناعي. لكن قد يتعين تطبيق هذا النهج في بيئة يُعزل فيها الذكاء الاصطناعي، حتى يتمكن الباحثون من فهم كيفية إدارته على نحو كافٍ قبل إطلاقه. وحينها قد نتلافى الخطر المتمثل في قدرة الذكاء الاصطناعي على التلاعب بنظام المكافأة والعقاب.



التطوير المتزامن

ماذا لو طوّر المبرمجون ذكاءات اصطناعية متعددة على التوازي؟ حينها عندما يصبح الذكاء الاصطناعي فائق الذكاء، سيكون لدينا عدة أنظمة ذكاء اصطناعي فائقة الذكاء ومستقلة، وتتمتع بذكاء وقدرات متقاربة، ومن ثم سيقيد بعض هذه الأنظمة بعضاً.

ومع ذلك، لا نضمن أن إحكام السيطرة على قدرات الذكاء الاصطناعي الفائق والحد منها سيجعل هذه الأنظمة تتصرف بما فيه نفع البشرية. كذلك قد يتمكن أحد أنظمة الذكاء الاصطناعي الفائق من التفوق على البقية، ما يزيد احتمالية سيطرته على العالم في نهاية المطاف.

لكن تطبيق التطوير المتزامن، في حال نجاح ضوابط المكافأة والعقاب على سبيل المثال، سيحد من مخاطر إساءة المشغلين البشريين استخدام الذكاء الاصطناعي الفائق.

بمعنى آخر، سيوجه كل فريق من المشغلين البشريين أنظمة الذكاء الاصطناعي الخاصة بهم للعمل بما يخدم مصالحهم. بينما ستعمل الفرق المختلفة على مراقبة بعضها بعضاً، وموازنة قوتها. وفي حال توفر عدد كافٍ من الفرق التي لديها أنظمة ذكاء اصطناعي متساوية القوة لتمثيل مصالح الجميع بأمانة، ستستخدم هذه الأنظمة لتعزيز المصالح المشتركة للبشرية.

لكن بما أن هذا النهج يعتمد على تزامن تطوير أنظمة الذكاء الاصطناعي الفائق والحفاظ على سيطرة البشر عليها، قد نحصل في النهاية على توازن قوى هش وغير دائم.



الدوافع الجيدة هي الأساس

تعد تدابير التحكم الأخرى أساليب مؤقتة فقط، لذلك يعد امتلاك الذكاء الاصطناعي الفائق دوافع جيدة أهم جزء في تطوير الذكاء الاصطناعي. ففي نهاية المطاف، ستقيد هذه الدوافع سلوك عمل الذكاء الاصطناعي الفائق. لكن كيف يمكن برمجة الذكاء الاصطناعي ليمتلك دوافع جيدة؟

هل سيطور الذكاء الاصطناعي إرادة حرة؟

تثار تساؤلات حول إمكانية تطوير الذكاء الاصطناعي إرادة حرة تمكنه من عصيان القواعد التي لا يرغب في اتباعها. فالإرادة الحرة عنصر أساسي في الإدراك البشري وتسهم بعمق في الوعي وقدرات التعلم المتقدمة.

وفي السياق ذاته، إذا كانت الإرادة الحرة عنصراً أساسياً في الذكاء العام، فسيتمتع أي ذكاء اصطناعي عام فائق بالإرادة الحرة. وحينها، قد يختار هذا الذكاء الفائق مخالفة القواعد وضوابط السلوك المبرمجة مسبقاً. وهذا من شأنه أن يزيد تعقيد مشكلة التحكم في سلوكه.

برمجة القواعد مسبقاً

لتقييد سلوك الذكاء الاصطناعي، يمكن برمجة مجموعة من القواعد مسبقاً. ومع ذلك، من المتوقع أن هذا غير عملي، إذ يصعب التفرقة بين السلوك المقبول وغير المقبول بطريقة واضحة.

فمثلاً، قد يساء تفسير القواعد أو تطبيقها، وقد تتطلب تعديلات دورية، فحتى أفضل القوانين لا تخلو من ثغرات. لذلك، قد يصعب وربما يستحيل كتابة مذكرة سلوك شاملة للذكاء الاصطناعي الفائق تكون قابلة للتطبيق على المستوى العالمي، وعلى مر العصور.

لذلك قد لا تكون الأوامر المبرمجة مسبقاً أفضل طريقة لمنح الذكاء الاصطناعي الدوافع الصحيحة.

تعزيز الدوافع الموجودة

مثلاً ذكرنا سابقاً، قد يمكن تطوير الذكاء الاصطناعي الفائق عبر محاكاة الدماغ. لذلك، يمكن اختيار شخص ذي شخصية مثالية ومسح دماغه بتقنية الرنين المغناطيسي الوظيفي لإنشاء نموذج أصلي. ومن ثم، تشغيل المحاكاة على حاسوب فائق الذكاء يستطيع التفكير بسرعة تفوق سرعة الدماغ البشري كثيراً.

لكن، قد لا تتمكن المحاكاة من الحفاظ على تفاصيل الشخصية بدقة، مثل القواعد الأخلاقية. فضلاً عن ذلك، قد لا تمنع المحاكاة الدقيقة لشخص ذي أخلاق جيدة إساءة استخدام قدرات الذكاء الاصطناعي الفائق.



اكتشاف الدوافع



قد تكون أفضل طريقة لمنح الذكاء الاصطناعي الفائق الذكاء دوافع جيدة هي منحه معايير لتحديد الصواب، وتركه يحدد أهدافه بنفسه. فالذكاء الاصطناعي الفائق سيكون قادراً على فهم ما يريده البشر منه. ومن ثم برمجة نفسه وفقاً لذلك بكفاءة تفوق قدرة المبرمجين البشريين، وسيظل يتصرف بحذر أكبر لأنه لن يكون متأكداً تماماً من أهدافه النهائية.

ومع ذلك، تظل مشكلة تحديد المعايير التي يبرمج عليها الذكاء الاصطناعي أمراً معقداً. فحتى لو ركز الذكاء الاصطناعي على ما يريده مبرمجوه البشريون، سيجعل الأولوية لرغبات فئة قليلة من الناس على حساب الآخرين. أليس من العدل أن يركز الذكاء الفائق على ما يريده الجميع؟ وأن يبني قراراته عموماً على القضايا التي يختلف الناس حولها؟

إضافة إلى ذلك، من المتوقع أن يوجد من يعارض مسار عمل الذكاء الاصطناعي الفائق، فأين ينبغي أن يضع الذكاء الاصطناعي حدّاً لذلك؟ إضافة إلى ذلك، قد تتضارب رغبات البشر أنفسهم.

كيف نُعلِّم الذكاء الاصطناعي ما يرغب فيه؟

هي زرع مسبار كهربائي في قشرة الفص الجبهي لديه؛ ما يحفز الدوبامين بصفة دائمة.

لذلك، ثمة مخاطر إزاء ترك الذكاء الاصطناعي يكتشف دوافعه بناءً على رغبات البشر بدلاً من إعطاء الذكاء الاصطناعي أوامر مباشرة. فقد يحدث تضارب في الرغبات، وقد يسيء الذكاء الاصطناعي فهم رغبات البشر، وقد يتلاعب برغبات البشر ويتحكم فيها، للحد من حالة عدم اليقين المتعلقة بها.

لحل ذلك، قد يبرمج المطورون الذكاء الاصطناعي على تحديد ما يريده الناس دون التأثير فيهم، بمعنى آخر، احترام الحرية الفردية. لكن، لم لا يبرمج الذكاء الاصطناعي على تعظيم الحريات الفردية للبشر؟ فكلما زادت حرية الناس زادت قدرتهم على تحقيق رغباتهم.

إلا أن تعظيم الحرية الفردية يعني أن منح الحرية لفرد يستلزم بالضرورة تقييد حرية شخص آخر.

قد يبدو الأمر معقداً، لكن إذا تمكّنّا من تطوير نظام ذكاء اصطناعي جيد بما يكفي، ويساعدنا في الحفاظ على مستقبل البشرية، وتجنب حدوث كارثة لا رجعة فيها، فقد تستطیع الأجيال اللاحقة أو النظام نفسه تطوير الأطر القياسية له.

يمكن حل هذه المشكلة بطريقتين، الأولى هي أن يكتشف الذكاء الاصطناعي ما يرغب فيه البشر لو كانوا أكثر معرفة ووضوحاً في تفكيرهم، وأعلى قدرة على فهم بعضهم بعضاً. وذلك من أجل استنباط رغباتهم وأهدافهم بصورة أفضل.

والطريقة الثانية هي برمجة الذكاء الاصطناعي ليكتشف ما هو صحيح أخلاقياً، ليسعى إليه، بدلاً من التركيز على رغبات البشر في حد ذاتها.

لكن نحن هنا أمام معضلة، فحتى ما يرغب الناس في الحصول عليه قد لا يكون الأفضل لهم، مثلما أنهم قد لا يرغبون في الأفضل أخلاقياً من الناحية النظرية. فضلاً عن أنه لا يوجد إجماع عام على تعريف أو نموذج معمم للأخلاق.

فضلاً عن ذلك، تثار تساؤلات حول هذه الطريقة، إذ كيف تحدد الأهداف والاختيارات المختلفة؟ وكيف يكون التعامل مع الاختلافات؟ وما التفكير المشروع؟ وهل ينبغي الحفاظ على التنوع في الأهداف والتفضيلات أم نسعى لهدف واحد أو نتيجة مشتركة؟

على الجانب الآخر، لنفترض أننا طورنا ذكاء اصطناعياً عامّاً بمهمة نبيلة وهي «إسعاد البشر» فما الذي يمكن أن يحدث؟ حسناً، قد يقرر الذكاء الاصطناعي مثلاً أن أنجح طريقة لإسعاد الإنسان



بفارق كبير قبل أن يتمكن المنافسون من اللحاق به. لذا يجب الدمج بين سرعة الانطلاق مع التركيز والتخطيط الاستراتيجي.

فإذا حقق نظام أو مشروع ما طفرة تقنية، فقد يصبح لدينا نظام تسيطر فيه جهة واحدة على العالم سيطرة فعالة. قد لا يكون ذلك سيئاً أو جيداً بالضرورة فالأمر يعتمد على أهداف هذا النظام. لكنه سيتحكم في المستقبل على المدى البعيد بشكل غير مسبق.

لذلك، يجب أن يكون الذكاء الفائق نتاج تعاون عالمي لا برنامجاً حكومياً سرياً. كما أنه من المهم أن نوازن بين الابتكار والسلامة، مع مراعاة أهمية الحوكمة والأخلاقيات والتخطيط المستقبلي في تسخير إمكانيات الذكاء الاصطناعي بمسؤولية.

ومن أجل ذلك، يجب بناء القدرات الفكرية والمؤسسية، وإجراء مزيد من البحوث في مجالات أمان الذكاء الاصطناعي واستراتيجياته وسلامتهما. فوجود خبراء أفضل يعني تحليلاً أدق للمخاطر المحتملة، ووضع أطر للتعاون.

فالمهمة العاجلة ليست التنبؤ بالتاريخ الدقيق لتطور الذكاء الاصطناعي الفائق، وإنما كوننا أكثر استعداداً قبل ظهوره. وهذا من شأنه أن يتطلب وضع آليات للتعاون، ومستويات استثنائية من الخبرات التقنية، والتحلي ببعد النظر والصبر. فالذكاء الفائق لم يظهر بعد لكن الكرة في ملعبنا، فما زلنا قادرين على اختيار كيفية ظهوره، لكن ليس لوقت طويل.

قد يكون من الحكمة تسريع وتيرة البحث العلمي في أساليب التحكم في أنظمة الذكاء الاصطناعي الفائق الذكاء، وبرمجتها لتحقيق أهداف نبيلة، في مقابل تقليل تركيزنا على تطوير الذكاء الاصطناعي المتقدم نفسه.

وذلك حتى نضمن وجود قواعد وحدود وضمانات صارمة قبل أن يصبح الذكاء الاصطناعي فائق الذكاء، حتى نضمن أن يكون تطويره مفيداً للبشرية. أما إذا حدث العكس، فقد تتحقق العواقب الوخيمة التي ناقشناها من قبل.

إلا أن مشاريع الابتكار لا تسير بالوتيرة المرغوبة إذ قد يصعب تطوير جوانب وتثبيط أخرى، لصعوبة التنبؤ بكيفية سير العمل الجديد من نوعه. خلاصة القول

لا يزال التمويل المخصص لضمان سلامة الذكاء الاصطناعي أقل كثيراً من الاستثمار المخصص لتطويره. لذلك ينبغي الاستثمار في التقنيات التي تحسن السلامة والتحكم قبل تطوير التقنيات التي تسمح بتطور الذكاء الفائق، ومن ثم، قد يسهم هذا الذكاء الفائق في الاستيلاء على السلطة.

إلى جانب الحذر من التنافس غير المدروس على تحقيق السبق في أبحاث تطوير تقنيات الذكاء الاصطناعي. إذ إن تسارع هذا السباق يؤثر في المقابل في تطوير الضوابط وعوامل الأمان. وهو أمر شديد الخطورة، إذ قد تقدم جهة ما على مخاطرة غير محسوبة من أجل تحقيق السبق.

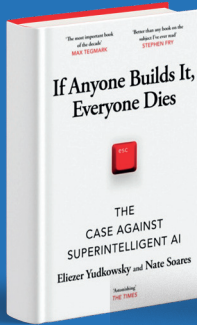
فإذا كان التقدم بطيئاً وعلى نطاق واسع، فقد تطور مجموعات عديدة أنظمة مماثلة. أما إذا حقق نظام أو مشروع ما قفزة تقنية كبيرة، فقد يتقدم

عن المؤلّف



نيك بوستروم
فيلسوف أكاديمي في جامعة أكسفورد. أسس معهد مستقبل الإنسانية،
وهو متحدث في TED وظهر في وسائل الإعلام الكبرى في مختلف أنحاء
العالم، صنفته مجلة «فورين بوليسي» واحداً من أفضل 100 مفكر عالمياً.

كتب أخرى مقترحة:



If Anyone Builds It, Everyone Dies

Eliezer Yudkowsky
& Nate Soares

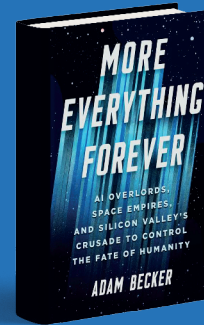
2025



The AI Con

Emily M. Bender
& Alex Hanna

2025



More Everything Forever

Adam Becker

2025

تواصلوا
معنا على
MBRFNews
mbrae
www.mbrae

مؤسسة محمد بن راشد آل مكتوم للمعرفة
MOHAMMED BIN RASHID AL MAKTOUM
KNOWLEDGE FOUNDATION

للإطلاع على المزيد من إصدارات
«كتاب في دقائق»، امسح رمز
الاستجابة السريعة.



2026 Knowledge Summit | قمة المعرفة



احفظ التاريخ Save the date

المعرفة: تخطي الأزمات في عالم متغير

Knowledge: Navigating Crises in a Changing World



17-18 November 2026



Dubai World Trade Centre

18-17 نوفمبر 2026

مركز دبي التجاري العالمي

