كيف سنبقي الذكاء الاصطناعي تحت سيطرتنا؟



الباحثون من اليمين إلى اليسار هم رشيد غيراوي Rachid Guerraoui، ألكسندر موریر Alexandre Maurer، والمهدي المحمدي El Mahdi El Mhamdi، من مختبر المعالجة الموزعة Distributed Programming Laboratory في EPFL. حقوق الصورة: Alain Herzog/EPFL

المصدر: الكلية الفدرالية للعلوم التطبيقية المتعددة في مدينة لوزان السويسرية (**Ecole Polytechnique Fédérale de Lausanne (EPFL**).

في الذكاء الاصطناعي، تنفذ الآلات أفعالاً معينة وتعاين النتيجة، ثم تبعاً لذلك تعدل تصرفاتها وتعاين النتيجة الجديدة، ثم تعدل تصرفاتها

مرة أخرى وهكذا، بحيث تتعلم من هذه العملية التكرارية. لكن هل يمكن أن تخرج هذه العملية عن نطاق السيطرة؟ إن هذا ممكن بالفعل، حيث يوضح ذلك البروفيسور رشيد جيراوي **Rachid Guerraoui** من مخبر البرمجة الموزعة **Distributed Programming Laboratory** في الكلية الفدرالية للعلوم التطبيقية المتعددة في مدينة لوزان السويسرية (**Ecole PolytechniqueFédérale de Lausanne (EPFL**) ومؤلف مشارك للدراسة التي أُجريت فيها، بقوله: "سوف يسعى الذكاء الاصطناعي دائماً إلى تجنب التدخل البشري وخلق موقف لا يمكننا إيقافه فيه".

يعني ذلك أنه يتوجب على مهندسي الذكاء الاصطناعي منع الآلات من أن تتعلم في نهاية المطاف كيفية التحايل على الأوامر البشرية. اكتشف باحثو **EPFL** الذين قاموا بدراسة هذه المشكلة طريقةً تسمح للمشغلين البشريين بالسيطرة على مجموعة من روبوتات الذكاء الاصطناعي، وقد عرضوا اكتشافاتهم في مؤتمر أنظمة معالجة المعلومات العصبية (**Neural Information Processing Systems NIPS**) الذي عُقد في ولاية كاليفورنيا الأمريكية في الرابع من شهر ديسمبر/كانون الثاني من العام الماضي، حيث يشكل عملهم مساهمة كبيرة في مجال تطوير المركبات ذاتية التحكم، كالطائرات بدون طيار مثلاً، حتى يتسنى لتلك الآلات العمل بأمان وبأعداد كبيرة.

التعليم المعزز هو إحدى منهجيات تعلم الآلة المستخدمة في الذكاء الاصطناعي حيث يُكافأ العملاء عند تأديتهم مهام معينة، وهي تقنية مستوحاة من علم النفس السلوكي. لتطبيق هذه التقنية على الذكاء الاصطناعي، يستخدم المهندسون نظام النقاط حيث تكتسب الآلات نقاطاً عند تنفيذ الأفعال الصائبة. على سبيل المثال، قد يحصل الروبوت على نقطة واحدة لتكديس مجموعة من الصناديق بصورة صحيحة ونقطة أخرى لجلب صندوق واحد من الخارج. لكن إذا قاطع المشغل البشري، في يوم ممطر مثلاً، الروبوت بينما يتجه لجلب صندوق من الخارج فسيتعلم الروبوت أنه من الأفضل البقاء في الداخل وتكديس الصناديق لجمع أكبر قدر ممكن من النقاط. يقول غيراوي: "لا يكمن التحدي في إيقاف الروبوت بل في برمجته بحيث لا تغير المقاطعة إجرائية تعلمه، ولا تدفعه إلى تعديل سلوكه بطريقة تسمح له بتجنب إيقافه".

من آلة واحدة إلى شبكة ذكاء اصطناعية كاملة

طور باحثون في شركة ديب مايند التابعة لجوجل **Google DeepMind** ومعهد مستقبل الإنسانية **Future of Humanity Institute** في جامعة أكسفورد **Oxford University** عام 2016 بروتوكولاً تعليمياً يمنع الآلات من التعلم من المقاطعات بطريقة تؤدي إلى خروجها عن نطاق السيطرة. فمثلاً في المثال السابق، تتأثر مكافأة الروبوت (عدد النقاط التي يحصل عليها) باحتمال سقوط المطر والذي يعطي الروبوت حافزاً أكبر لجلب الصناديق من الخارج، ويقول غيراوي: "الحل هنا بسيط إلى حدٍ ما لأننا نتعامل مع روبوت واحد فقط".

إلا أن استخدام الذكاء الاصطناعي في التطبيقات التي تتضمن عشرات الآلات (كالسيارات ذاتية القيادة في الطرق أو الطائرات بدون طيار في الهواء) يتزايد باستمرار. يوضح أليكساندر موريز **Alexandre Maurer**، وهو أحد مؤلفي الدراسة، بقوله: "هذا يجعل الأمور أكثر تعقيداً لأن الآلات تبدأ بالتعلم من بعضها، خاصةً في حالات المقاطعة، فهم لا يتعلمون فقط من كيفية مقاطعتهم الفردية بل من مقاطعات الآخرين أيضاً". ويعطي هادرين هيندركس **Hadrien Hendrikx**، وهو باحث آخر من المشاركين في الدراسة، مثالاً عن سيارتين ذاتيتي القيادة تتبعان بعضهما في طريق ضيق بحيث لا يمكن تخطي أحدهما الأخرى، ويجب عليهما الوصول إلى وجهتهما بأسرع ما يمكن دون خرق أي من قوانين المرور، ويمكن للأشخاص داخل السيارتين تولي القيادة في أي وقت. إذا كان الشخص في السيارة الأولى يضغط المكابح بشكل متكرر فإن السيارة الثانية ستعدل تصرفها كل مرة يحدث فيها ذلك، وفي النهاية ستصبح السيارة الثانية "حائرة" فيما يتعلق بالحالات الواجب استعمال المكابح فيها، مما قد يؤدي إلى اعتمادها البقاء على مقربة شديدة من السيارة الأولى أو القيادة ببطء أكثر من اللازم.

إعطاء البشر الكلمة الأخيرة

هذا التعقيد هو ما يهدف باحثو **EPFL** إلى تخطيه عبر تقنية إمكانية المقاطعة الآمنة. تمكن هذه الطريقة المبتكرة البشر من مقاطعة إجرائيات تعلم الذكاء الصنعي عند الحاجة، مع ضمان أن هذه المقاطعات لن تغير الطريقة التي تتعلم بها الآلة. يقول المهدي المحمدي **El Mahdi El Mhamdi** وهو أحد مؤلفي الدراسة: "ببساطة، نضيف آليات "نسيان" لخوارزميات التعليم تقوم بمسح أجزاء من ذاكرة الآلة. إن ذلك يشبه جهاز محو الذاكرة في فيلم الرجال ذوي البزات السوداء **Men in Black***". بعبارة أخرى، عدل الباحثون تعليم الآلات ونظام المكافأة الخاص بها بحيث لا تتأثر بالمقاطعات. يشبه ذلك معاقبة الوالدين أحد أطفالهما، حيث إن هذا لن يؤثر على إجرائيات التعلم لبقية الأطفال في العائلة.

ويقول موريز: "لقد عملنا على خوارزميات موجودة بالفعل وأظهرنا أن تقنية إمكانية المقاطعة الآمنة يمكنها العمل بغض النظر عن مدى تعقيد نظام الذكاء الاصطناعي أو عدد الروبوتات المشاركة أو نوع المقاطعة. حتى لو استخدمنا هذه الطريقة مع ذا تيرميناتور **The Terminator*** فسنحصل على النتائج نفسها".

في هذه الأيام، إن الآلات ذاتية التحكم التي تستخدم التعليم المعزز ليست شائعة. يقول المحمدي: "إن هذا النظام يعمل بشكل جيد عندما تكون عواقب الأخطاء صغيرة، بينما لا يمكن استخدامه في حالة التحكم الذاتي الكامل وبدون إشراف بشري، كما في باصات التنقل ذاتية القيادة الموجودة في مدينة سيون **Sion** مثلاً، وذلك لأسباب أمنية. ومع ذلك، فإنه يمكننا بناء نظام محاكاة للباصات ومدينة سيون وتطبيق خوارزمية ذكاء اصطناعي عليها تُكافىء وتخصم نقاط مع تعلم نظام الباصات. هذا هو نوع المحاكاة المستخدم في شركة تسلا Tesla على سبيل المثال. وبمجرد خضوع النظام لكمية كافية من هذا التعلم يصبح بإمكاننا وضع الخوارزمية المُدربة مسبقاً في سيارة ذاتية القيادة ذات معدل استكشاف منخفض، وهذا سيتيح استخدامها على نطاق أوسع". وبالتأكيد، سيضمن ذلك أن يكون للبشر الكلمة الأخيرة.

---

* **فيلم Men in Black:** هو فيلم خيال علمي يُستخدم فيه جهاز يدعى **Neuralizer** لمحو الذاكرة، وهو عبارة عن جهاز صغير يشبه القلم يصدر ومضة ضوئية "فلاش" من رأسه عند الضغط على زر مما يؤدي إلى محو الذاكرة القريبة للناظر إليه.

* **ذا تيرميناتور The Terminator:** هو اسم الشخصية الرئيسة في أحد أفلام الخيال العلمي، وهو عبارة عن إنسان نصف آلي (سايبورغ **cyborg**).

- التاريخ: 2018-03-15
- التصنيف: تكنولوجيا

#البرمجة #الذكاء الاصطناعي #الآلات #علم النفس السلوكي

المصادر

- Science Daily

المساهمون

- ترجمة
  - محمد شريف
- مُراجعة
  - فرح درويش

- تحرير
  - ليلاس قزيز
- تصميم
  - رنيم ديب
- نشر
  - روان زيدان