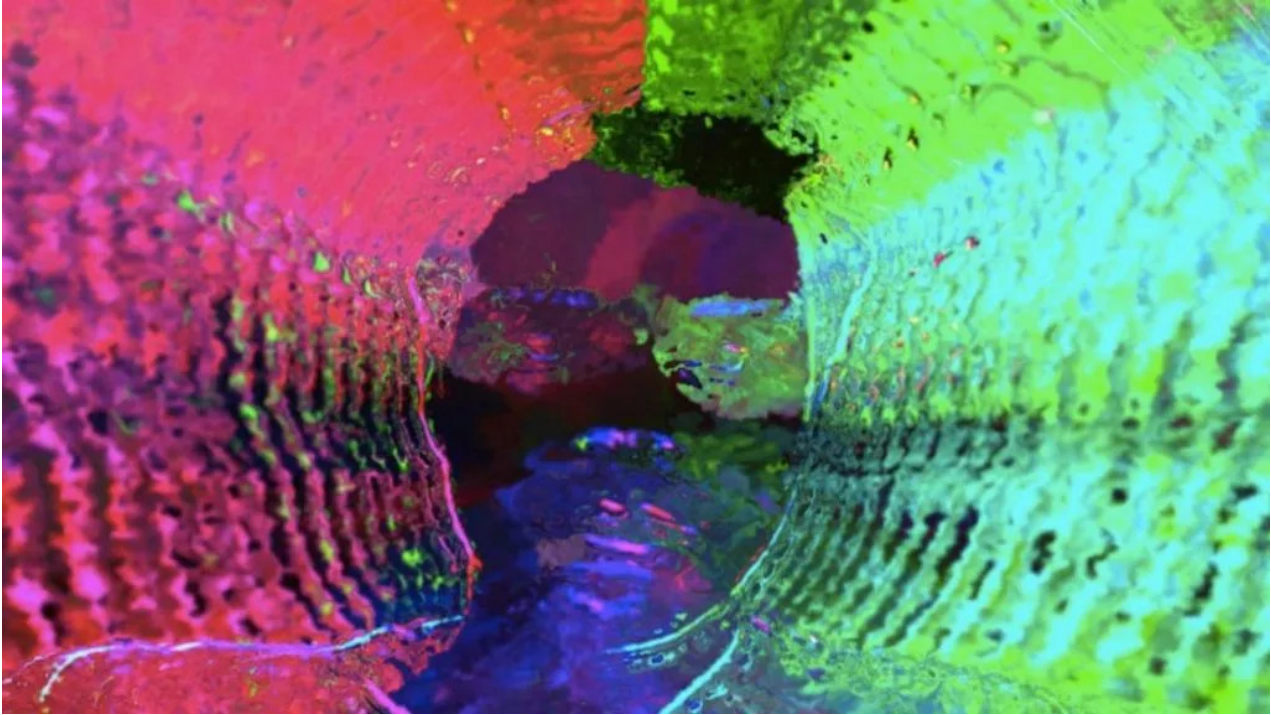


ما هو الذكاء الاصطناعي القابل للتفسير؟ وهل نحن في حاجة إليه؟



ترجمة وتحرير: نون بوست

تُوقش موضوع الذكاء الاصطناعي مرارا وتكرارا خلال السنوات الأخيرة، حيث تباينت الآراء حوله. ولكن قبل مناقشة مدى دقة الذكاء الاصطناعي، لا بدّ من طرح السؤال التالي: إذا كان الذكاء الاصطناعي يحاول نمذجة كل من تفكيرنا وعملية صنع القرار، فيجب حينها أن نكون قادرين على شرح كيفية اتخاذنا للقرارات! أليس كذلك؟

في الحقيقة، هناك تحوّل على مستوى التعلّم الآلي تتسارع وتببطأ وتيرته منذ الخمسينات. ففي الماضي القريب، كان التعلّم الآلي يعدّ المجال اللافت للنظر والأكثر دراسة، حيث يهدف إلى نمذجة كل من طريقة أخذ القرار، والسلوك، وردود الفعل. وعموماً، أدت النتائج الناجحة التي تم الحصول عليها في مجال التعلّم الآلي إلى نمو سريع في تطبيق الذكاء الاصطناعي. ويُعدّ العمل المتقدّم بأن يكون عبارة عن أنظمة مستقلة قادرة على إدراك الذات، والتعلّم، وصنع القرار، والحركة.

بالتحديد بعد التسعينات، أصبح مفهوم التعلّم العميق يعتمد على الماضي، لكن حظيت الشبكات العصبية المتكررة والشبكات العصبونية الالتفافية والتعلّم المعزّز والشبكات الجدليّة بنجاح ملحوظ. وعلى الرغم من الحصول على نتائج إيجابية، إلا أنه ليس من الكافي تفسير أو شرح قرارات وسلوك هذه الأنظمة للمستخدمين البشر.

مجال الذكاء الاصطناعي القابل للتفسير

إن نماذج التعلّم العميق المصممة مع مئات الملايين من الشبكات العصبية الاصطناعية، ليست معصومة عن الخطأ. فمن الممكن أن تفقد هذه النماذج مصداقيّتها بسرعة خاصة عندما يتم ببساطة تضليلها كما هو الحال في ”هجوم البكسل الوحيد“. بعد ذلك، يصبح من المحتمّ طرح سؤال حول مدى نجاحها من عدمه!



كلب
(سفينة)



سيارة
(كلب)



سيارة
(طائرة)



قط
(كلب)



كلب
(سفينة)



غزال
(كلب)



ضفدع
(كلب)



ضفدع
(شاحنة)



كلب
(قط)



ضفدع
(شاحنة)



حصان
(قط)



سفينة
(شاحنة)



حصان
(سيارة)



كلب
(حصان)



سفينة
(شاحنة)

صرّحت وزارة الدفاع بأن الأنظمة الأكثر ذكاء واستقلالية وتكافلية تواجه جملة من التحديات. وفي هذا السياق، قالت وزارة الدفاع إن "الذكاء الاصطناعي القابل للتفسير، وبالتحديد التعلم الآلي القابل للتفسير، سيكون ضروريا إذا ما أراد المحاربون في المستقبل أن يفهموا ويثقوا في الجيل الناشئ من الآلات التي تعمل وفقا للذكاء الاصطناعي". وعلى العموم، يتنامى التعقيد الذي يتسم به هذا النوع من التطبيقات المتقدمة في ظل النجاحات ليصبح التفسير، مهمة صعبة. فحتى خلال بعض المؤتمرات، نجد جلسات مخصصة فقط لمناقشة هذا الموضوع.

الأسباب الكامنة وراء الأنظمة الجديدة للآلة/ للتعلم العميق

إنه يهدف إلى شرح الأسباب الكامنة وراء الأنظمة الجديدة للآلة/ للتعلم العميق لتحديد مواطن القوة والضعف وفهم كيفية التصرف في المستقبل. ولتحقيق هذا الهدف، تتمثل الاستراتيجية في تطوير تقنيات تعلم اصطناعية إما جديدة أو معدلة من شأنها أن تُنتج نماذج تكون قابلة لمزيد التعريف.

في الواقع، تهدف هذه النماذج إلى دمجها مع أحدث تقنيات التفاعل الإنساني الحاسوبي، حيث من الممكن أن تحوّل النماذج إلى مربعات حوار مفهومة ومفيدة للمستخدم النهائي. ومع ثلاثة توقعات أساسية، يعدّ من المنشود مقارنة النظام:

- اشرح الغرض من كيفية تأثر الأطراف التي تقوم بتصميم واستخدام النظام.

- اشرح كيفية استخدام مصادر البيانات والنتائج.

”يعدّ الذكاء الاصطناعي القابل للتفسير من بين البرامج القليلة الحالية التي أطلقتها وكالة مشاريع البحوث المتطورة الدفاعية ”داربا“

– اشرح كيف تؤدي المُدخلات في إحدى نماذج الذكاء الاصطناعي إلى المُخرجات.

”يعدّ الذكاء الاصطناعي القابل للتفسير من بين البرامج القليلة الحالية التي أطلقتها وكالة مشاريع البحوث المتطورة الدفاعية ”داربا“ التي يُتوقع أن تُمكن أنظمة الموجة الثالثة من الذكاء الاصطناعي، التي تفهم الآلات فيها السياق والبيئة التي تعمل فيها، مع مرور الوقت من تبني نماذج توضيحية أساسية من شأنها أن تسمح لها بتوصيف ظواهر العالم الحقيقي“.

إذا ما انطلقنا من الممارسة الطبية، بعد فحص بيانات المريض، يجب أن يفهم الطبيب ويشرح للمريض أنه أعلمه بخطر الإصابة بنوبة قلبية بناء على التوصية التي قدّمها نظام دعم القرار. في هذه المرحلة، يجب أولاً تحديد البيانات التي تم تقييمها، لأن هذه الخطوة تعد من المقاييس الضرورية. ومن المهم أيضاً تحديد البيانات التي تحتاج إليها وما يجب القيام به لإجراء التقييم المناسب.

الجانب السيكولوجي للتفسير

دعونا نلقي نظرة على النقطة التي نرفض فيها استخدام تكنولوجيا الذكاء الاصطناعي في مجال التعليم لأننا لا نستطيع شرح كيفية اتخاذ هذه التقنية للقرارات. ولا يستطيع أغلب الأشخاص شرح الطريقة التي مكنتهم من اتخاذ القرار الصحيح. دعونا نتخيل كيف يمكن للشخص التوصل إلى اتخاذ قرار على مستوى هذا النموذج. فعندما نتحدث عن تركيبنا البيولوجية على المستوى الكيميائي والفيزيائي، فإننا نتحدث عن الإشارات الكهربائية التي تنتقل من خلية دماغية إلى خلية دماغية أخرى. إذا لم تكن راضياً عن هذا التفسير، فأخبرني كيف قررت طلب فنجان من القهوة.

تعدّ التفسيرات والشفافية والوضوح استعراضات تحليلية، كما أن إجراء التحاليل دون اختبار يشبه اقتناء تذكرة قطار باتجاه واحد، وهو الأمر الذي يبعث شعوراً بالأمان

عندما تكون رفقة أصدقائك في المقهى، قد يطلب أحد أصدقائك قهوة مثلجة، في حين يطلب صديقك الآخر قهوة ساخنة، أما الآخر فيطلب كوباً من الشاي. قد يجعلك هذا الأمر تطرح أسئلة عديدة؛ مثل السبب وراء اختيار هؤلاء الأشخاص شرب قهوة مثلجة أو قهوة ساخنة؟ أو كيف يمكن تفسير طريقة عمل المواد الكيميائية والشبكات العصبية في الدماغ؟ ومن المؤكد أنك ترغب في إيجاد تفسيرات لكيفية بدأ الإنسان في فهم عملية اتخاذه للقرارات.

إن كل ما عليك فعله هو التركيز على المعلومات التي تتلقاها وترسلها، ثم أروي قصة مضحكة. وهناك نهج مماثل فيما يتعلق بالمسائل التحليلية والمسائل المهمة. وتعدّ التفسيرات والشفافية والوضوح استعراضات تحليلية، كما أن إجراء التحاليل دون اختبار يشبه اقتناء تذكرة قطار باتجاه واحد، وهو الأمر الذي يبعث شعوراً بالأمان. وفي ظروف مثالية، قد ترغب في وجود نظام قادر على تقديم أفضل أداء ممكن، والحصول على أفضل تفسير مقنع، لكن الحياة هي التي تجبرنا على الاختيار.

الأداء مقابل التفسير

التفسير: تتمكن من فهم المعنى، في حين أن تقنية الذكاء الاصطناعي لا تعمل بشكل جيد.

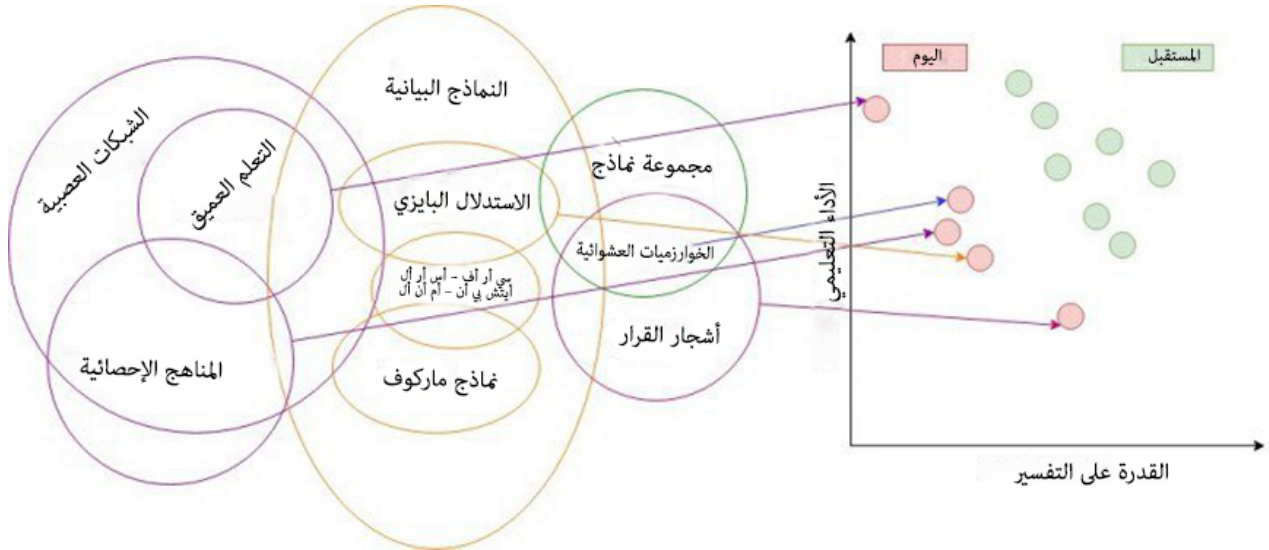
الأداء: لم تتمكن من فهم المعنى، في حين أن تقنية الذكاء الاصطناعي تعمل بشكل جيد.

لن يولي الأكاديميون والباحثون وشركات التكنولوجيا اهتماماً كبيراً بالتفسير لدرجة جعلهم يولون أهمية أكبر للأداء. ومع ذلك، إن الأمر يختلف قليلاً مع الناس والمؤسسات العاملة في هذا القطاع، إذ أن كل ما يرغبون فيه هو الثقة والحصول على تفسيرات.

لا ينبغي أن تتحول تقنية الذكاء الاصطناعي إلى قوة إلهية، حيث نضطر إلى استخدامها دون وجود علاقة سببية

تختلف نماذج الذكاء الاصطناعي المستخدمة من قبل البنوك وشركات التأمين ومقدمي الرعاية الصحية وغيرها من الصناعات المختلفة عن بعضها البعض، ويتمثل السبب في أن النماذج المستخدمة في هذه القطاعات تتطلب وجود لوائح قانونية وشروط أخلاقية مختلفة. في هذه الحالة، نعود إلى نفس النقطة مجدداً؛ إذا كنت ترغب في فهم نظامك وفقاً للشروط التالية، فسيتعين عليك استبداله بنظام أبسط وغير معقد.

أجريت معظم الأبحاث حول هذا الموضوع من قبل وكالة "داربا" وشركة غوغل وشركة ديب مايند وما إلى ذلك. وفي حين أن هذه المؤسسات لا تزال تعمل على تحقيق أهدافها بشكل مكثف، فإن المفهوم من هذه التقارير أنه بغض النظر عن القطاع والشركات التي تستخدم أنظمة الذكاء الاصطناعي، فإن هناك علاقة بين الوضوح والدقة حيث لا يمكن تجنب المقاضلة ولكن يبدو أنها تظل مستمرة لفترة من الوقت.



على أي حال، لا ينبغي أن تتحول تقنية الذكاء الاصطناعي إلى قوة إلهية، حيث نضطر إلى استخدامها دون وجود علاقة سببية. ومن ناحية أخرى، لا ينبغي لنا أن نتجاهل الرؤية التي ستقدمها لنا مستقبلاً. وبشكل أساسي، يجب أن نفكر في إنشاء نماذج مرنة وقابلة للفهم تكون قادرة على أن تعمل معاً في انسجام مع الخبراء الذين لديهم معرفة على المستوى الفني والأكاديمي والآراء من مختلف القطاعات والتخصصات.

المصدر: انترستغ إنجينيرينغ