

## أهمية البيانات الضخمة للذكاء الاصطناعي



ترجمة وتحرير: نون بوست

في الوقت الراهن، لا توجد أية أصول في الاقتصاد الرقمي أهم من البيانات. فقد أصبحت واسعة الانتشار إلى حد اعتبارها "النفط الجديد". وفي أحد عناوين مجلة الإيكونومست، وصفت البيانات على أنها "المورد الأكثر قيمة في العالم".

أضحت البيانات تحظى اليوم بتقدير كبير بسبب الدور الأساسي الذي تلعبه في توفير حلول التعلم الآلي والذكاء الاصطناعي. ويتطلب تدريب نظام الذكاء الاصطناعي على العمل بفعالية، بداية من محرك توصيات نتفليكس إلى سيارات غوغل ذاتية القيادة، مجموعات هائلة من البيانات.

نتيجة لذلك، بات هاجس البيانات يكبر شيئاً فشيئاً. ووفقاً للحكمة السائدة، يصنع الذكاء الاصطناعي الجيد من خلال الكثير من البيانات. وتتنافس الشركات في هذا المجال، بداية من شركة "آي بي إم" وصولاً إلى "جنرال إلكتريك"، على إعادة تصنيف أنفسهم على أنهم "شركات بيانات". من جهته، لا يخفي صندوق رؤية سوفت بانك، أكبر مستثمر تكنولوجي في العالم، حقيقة أن تركيزه يصب على أصول البيانات عند بحثه عن الشركات الناشئة التي تستحق الدعم. وعلى حد تعبير المدير التنفيذي لشركة سوفت بانك، ماسايوشي سون، "من يتحكم بالبيانات هو الذي سيحكم العالم".

تعمل شركات التكنولوجيا الرائدة، بداية من المنافسين القدامى في السوق على غرار شركة إنفيديا وحتى الشركات الناشئة مثل إبلويد أنتويشن، على تطوير طرق لتصنيع بيانات عالية الدقة رقمياً ودون تكلفة

نظرا لأن عالم الأعمال والتكنولوجيا يتجهان بشكل متزايد نحو البيانات باعتبارها ذات أهمية كبرى، لا يوجه سوى القليل من الاهتمام إلى حقيقة مهمة تتمثل في أنه من المحتمل ألا يكون مستقبل الذكاء الاصطناعي مرتبطا كثيرا بالبيانات الضخمة. وفي مجال الذكاء الاصطناعي، تبذل جهود مختلفة لتطوير أشكال محسنة من الذكاء الاصطناعي لا تتطلب مجموعات ضخمة من البيانات المصنفة. وفي الواقع، ستعيد هذه التقنيات تشكيل فهمنا للذكاء الاصطناعي وإدخال تغييرات كبيرة على مجال الأعمال. علاوة على ذلك، سيستفيد أصحاب الأعمال كثيرا من هذا الأمر.

### البيانات الاصطناعية

من أجل تدريب نماذج التعلم العميق، ينبغي على العاملين في هذا المجال جمع الآلاف أو الملايين أو حتى المليارات من نقاط البيانات. بعد ذلك، ينبغي عليهم تسمية كل نقطة بيانات، وهي عملية مكلفة وتحدث بشكل يدوي. فماذا لو لم يكن الباحثون بحاجة إلى جمع وتسمية البيانات من الواقع، ويمكنهم بدلا من ذلك، إنشاء مجموعة البيانات التي يحتاجون إليها من لغة برمجة السكراتش؟

في الحقيقة، تعمل شركات التكنولوجيا الرائدة، بداية من المنافسين القدامى في السوق على غرار شركة إنفيديا وحتى الشركات الناشئة مثل إبلويد أنتويشن، على تطوير طرق لتصنيع بيانات عالية الدقة رقميا ودون تكلفة. كما يمكن تصميم مجموعات من البيانات التي تم إنشاؤها بشكل مصطنع وفقا لاحتياجات الباحثين الدقيقة ويمكن أن تشمل مليارات السيناريوهات البديلة. وفي هذا الإطار، قال مايك سكولونيس، المسؤول عن تقنية المحاكاة في نفيديا: ”يعد الخروج وتغيير الإضاءة في العالم الحقيقي أمرا مكلفا للغاية، كما لا يمكنك تغيير الإضاءة في مشهد خارجي“. لكن في ظل البيانات الاصطناعية، أصبح يمكنك القيام بذلك.

بفضل تقارب البيانات الاصطناعية من بيانات العالم الواقعي من ناحية الدقة، سيعمل ذلك على إضفاء الطابع الديمقراطي على الذكاء الاصطناعي، مما يقوض الميزة التنافسية لأصول البيانات المتعلقة بحقوق الملكية. وفي حال استطاعت الشركة توليد مليارات الأميال من بيانات القيادة الواقعية عبر المحاكاة، فكم ستكون قيمة بضعة ملايين من بيانات القيادة الحقيقية التي استثمرت شركة وايمو في جمعها لمدة عقد من الزمن؟ في عالم يمكن فيه توليد بيانات عند الطلب بتكلفة زهيدة، سيتم قلب ديناميات التنافس في الصناعات.

### التعلم من خلال بضع لقطات

على عكس الذكاء الاصطناعي، لا يحتاج البشر لرؤية الآلاف من الأمثلة من أجل تعلم مفهوم جديد. وفي شأن ذي صلة، أوضحت ورقة بحثية مهمة خاصة بشركة غوغل أنه ”يمكن للطفل تعميم مفهوم ”الزرافة“ من خلال صورة واحدة في كتاب. في المقابل، تحتاج أفضل أنظمة التعلم العميق لدينا إلى مئات أو آلاف الأمثلة“. وحتى يتمكن الذكاء الآلي الاقتراب من الذكاء البشري في قدراته، ينبغي أن يكون قادرا على التعلم والتفكير من خلال بضعة أمثلة على غرار ما يقوم به البشر. وبعد هذا هو الهدف من حقل مهم داخل الذكاء الاصطناعي يعرف باسم ”التعلم من خلال بضع لقطات“.

نظراً لأن نظام البيانات الصغيرة تنتقل من الأوساط الأكاديمية إلى الإنتاج التجاري في السنوات المقبلة، فإنها ستغير الطريقة التي يُطوّر بها الذكاء الاصطناعي بشكل جذري، الشيء الذي يقلل من أهمية البيانات الضخمة في هذه العملية.

في شأن ذي صلة، هناك تقدم مثير في الآونة الأخيرة في التعلم من خلال بضع لقطات، وخاصة في مجال الرؤية الحاسوبية. فضلا عن ذلك، يُطلق على هذه التقنية اسم ”التعلم من خلال لقطات واحدة“ عند استخدام نقطة بيانات واحدة، أو ”التعلم دون لقطات“ في حال عدم استعمال أي بيانات. وطور

الباحثون نماذج من الذكاء الاصطناعي التي، في ظل الظروف المناسبة، يمكن أن تحقق أداءً متطورًا في مهام مثل التعرف على الوجه بناءً على نقطة أو بضع نقاط بيانات.

في المقابل، تبقى هذه التطورات محصورة في العالم الأكاديمي في الوقت الراهن. ونظرًا لأن نظام البيانات الصغيرة تنتقل من الأوساط الأكاديمية إلى الإنتاج التجاري في السنوات المقبلة، فإنها ستغير الطريقة التي يُطوّر بها الذكاء الاصطناعي بشكل جذري، الشيء الذي يقلل من أهمية البيانات الضخمة في هذه العملية.

من جانبه، أوضح أندرو نج، رائد التعلم العميق ورئيس سابق للذكاء الاصطناعي في غوغل وبايدو: "إذا كنت تقوم بفحص مرئي على الهواتف الذكية، لن يكون لديك مليون صورة للهواتف الذكية المخدوشة. فإذا كان بإمكانك إصلاح شيء ما باستخدام 100 أو 10 صور فقط، فسيؤدي ذلك إلى ظهور الكثير من التطبيقات الجديدة".

### التعليم المعزز

يُعتبر نظام التعليم المعزز بصدد تحقيق تقدم مهم في الذكاء الاصطناعي دون الحاجة إلى تناول البيانات الحقيقية. ففي التعليم المعزز، لا يتعلم نموذج الذكاء الاصطناعي من خلال استهلاك بيانات القوى العمياء، ولكن من خلال التجربة والخطأ الموجهين ذاتياً، إذ أن له الحرية في تجربة تصرفات مختلفة في بيئة معينة. لذلك، يطور هذا النموذج سلوكه تدريجياً لأنه يتلقى ردود الفعل حول الإجراءات المفيدة والإجراءات غير المفيدة.

من جهة أخرى، كان أحد أكثر إنجازات الذكاء الاصطناعي انتشاراً على نطاق واسع في السنوات الأخيرة مدعوماً بالتعليم المعزز، ألا وهو هزيمة ديب مايند لأفضل لاعبي لعبة غو العريقة في العالم. في الواقع، تعلم ديب مايند، النموذج الأصلي لألفا غو، اللعبة باستخدام مزيج من البيانات التاريخية والتعليم المعزز، غير أن الإنجاز الرائع جاء مع الألفا غو زيرو الأكثر تطوراً. كما أنه لم يُمنح أية بيانات سابقة بخلاف قواعد اللعبة أو أية معلومات إضافية أخرى. وعلى الرغم من ذلك، يعدّ تعلم لعبة غو أفضل من أي إنسان أو آلة عن طريق اللعب ضد نفسه. والجدير بالذكر أن الألفا غو زيرو هزم الألفا غو الأصلية بنتيجة 100 نظيفة.

إلى جانب ألعاب الطاولة، يجد التعليم المعزز تطبيقات حقيقية في مجال الروبوتات والهندسة الكيميائية والإعلانات، وغير ذلك

من جهته، أوضح فريق ألفا غو زيرو أن "مجموعة البيانات الخبيرة غالباً ما تكون باهظة أو غير موثوق بها أو ببساطة غير متوفرة. وعلى النقيض من ذلك، تُدرّب أنظمة التعليم المعزز انطلاقاً من تجربتها الخاصة، فيُسمح لها بتجاوز القدرات البشرية والعمل في المجالات التي تفتقر إلى الخبرة البشرية".

إلى جانب ألعاب الطاولة، يجد التعليم المعزز تطبيقات حقيقية في مجال الروبوتات والهندسة الكيميائية والإعلانات، وغير ذلك. كما أنه يمثل مقاربة جديدة داخل الذكاء الاصطناعي، أي أنه بدلاً من طلب مجموعات البيانات الضخمة الموجودة مسبقاً، يمكنه توليد بياناته الخاصة، والتعلم من هذه العملية في الأثناء. وبينما يحاول التعليم المعزز إيجاد طريقه إلى التطبيقات التجارية، سيمثل تحدياً آخر لتقليد البيانات الضخمة.

### الاستنتاج

لا شك في أن عالم الذكاء الاصطناعي في حالة تغير مستمر. ومع تقدم هذا المجال بسرعة فائقة، يمكن أن تصبح المنهجيات الحديثة والمتطورة اليوم قديمة في المستقبل. وفي الوقت الراهن، يمثل التعلم العميق نموذج الذكاء الاصطناعي الأكثر هيمنة، والذي يعتمد على ما يصل إلى مليارات من نقاط البيانات

المصنفة لتدريب الشبكات العصبية الاصطناعية على التعرف على الأنماط والتنبؤات. ونظرًا لأن الشبكات العصبية متعطشة للبيانات، فقد أصبح قادة الأعمال والتكنولوجيا مهووسين بتجميع أكبر عدد ممكن من مجموعات البيانات، على أمل أن تكون البيانات هي ميزتهم التنافسية النهائية في عالم يحركه الذكاء الاصطناعي.

في المقابل، يعد التعلم العميق بمثابة محطة واحدة في الطريق الطويل للذكاء الاصطناعي، وليس وجهتها النهائية. كما أن تأسيس استراتيجية عمل طويلة المدى على احتياجات البيانات العصبية الضخمة للشبكات العصبية الحالية يعد دليلاً على الفشل في تقدير التحولات النموذجية المستقبلية في الذكاء الاصطناعي التي نتظرنا. علاوة على ذلك، توضح التطورات الحديثة في العديد من المجالات، مثل البيانات الاصطناعية والتعلم من بضع لقطات والتعليم المعزز، أنه كلما أصبح الذكاء الاصطناعي أكثر ذكاءً في السنوات القادمة، فمن المحتمل أن تتطلب بيانات أقل.

عموماً، ستعيد هذه النماذج الجديدة بناء تاريخ الذكاء الاصطناعي وتحديد الشروط التي ستتنافس عليها الشركات. وبالنسبة لرجال الأعمال والتكنولوجيين الذين يتطلعون إلى التفكير، ستكون هذه فرصة هائلة.

المصدر: فوربس