

الجانب المقلق لاستخدام البيانات المصطنعة

الكاتب



راسل قاسم

د. راسل قاسم

يتزايد الاقتناع بأن البيانات هي نطفة العصر المقبل، وأن من يمتلك البيانات سيكون قادراً على التحكم في الكثير من مفاصل المستقبل.

على الرغم من الكميات الهائلة من البيانات التي ننتجها كل يوم، إلا أنها لم تعد كافية لإشباع نهم أنظمة الذكاء الاصطناعي وتعلم الآلة. من المتوقع أن يبلغ حجم البيانات عام 2025 حوالي 170 زيتابايت، وهذا الرقم مهول فهو مثلاً 10 أضعاف البيانات التي كانت متوفرة في عام 2017. في طريق موزن، يتضخم سوق الذكاء الاصطناعي بشكل مذهش، حيث يُتوقع أن يصل حجم هذا السوق إلى 1.4 تريليون دولار قبل نهاية هذا العقد.

تحتاج برامج الذكاء الاصطناعي وتعلم الآلة إلى أن تتدرب لتتمكن من أداء مهامها بشكل صحيح، وتدريبها يعتمد على توفر بيانات سابقة تتعلم منها لتبدأ بالعمل بشكل مستقل. ونتيجة العدد الكبير من برامج الذكاء الاصطناعي، أصبحت البيانات المتاحة غير كافية، فالاعتماد عليها فقط سيستغرق وقتاً طويلاً، كما أن مستوى الدقة سينخفض، ناهيك عن صعوبة الحصول على بعض البيانات واستخدامها نتيجة أسباب مرتبطة بالخصوصية، أو حقوق الأفراد، أو سرية وحساسية تلك المعلومات.

الحل كان في القيام بتوليد البيانات بشكل آلي، وهي بيانات غير حقيقية تعتمد على مجموعة من البيانات الواقعية، حيث تقوم أنظمة خاصة بتوليد الكميات المرغوبة من البيانات المصطنعة هذه، والتي تستخدمها برامج الذكاء الاصطناعي للتعلم بتكلفة ووقت أقل. وقد أصبح هذا السوق الثانوي قائماً بذاته حيث توجد العديد من الشركات المختصة بتوليد بيانات مصطنعة لشركات وأنظمة الذكاء الاصطناعي.

من جهة أخرى، تشعّبت تطبيقات الذكاء الاصطناعي وتعلم الآلة في القطاعات الاقتصادية والممارسات الإدارية، فنجدها في خدمة المتعاملين من خلال اقتراح المنتجات والخدمات المناسبة لهم، أو أنظمة المحادثة والرد التلقائي على الأسئلة والاستفسارات، وكذلك في مجال التسويق واستهداف الأفراد بشكل مختلف اعتماداً على أنماطهم الاستهلاكية، وصولاً إلى العديد من القطاعات الأخرى كتشخيص الأمراض في القطاع الصحي، وتقييم الملاءات المالية في البنوك، وتحليل المطالبات في شركات التأمين، وتوقع الجرائم وتوجيهاتها في الأجهزة الشرطية، والقيادة الذاتية وتطبيقاتها في ..قطاع النقل، وابتكار أساليب جديدة في التعليم والتدريب، والحديث هنا يطول

للأسف فإن الصورة غير مكتملة الإشراف. فالبيانات المصطنعة كما أنها قدّمت حلاً عملياً لمشكلة تدريب الآلة وتطوير أنظمة الذكاء الاصطناعي، فإنها خلقت تحديات أخرى كالتحيّز والتمييز على أساس الجنس أو العرق، حتى أن التحيزات التي توجد في البيانات الحقيقية يمكن أن تتضخم وتزداد في البيانات المصطنعة، فتتفاقم الممارسات الخاطئة لأن الأنظمة تقوم باستخدام هذه البيانات لبناء تنبؤات واتخاذ قرارات

فمثلاً في تجربة أجريت في جامعة أريزونا في الولايات المتحدة الأمريكية نُشرت نتائجها العام الماضي، تم استخدام صور لأساتذة جامعيين بلغ عددها حوالي 17000 صورة حيث كان 80% منهم ذكوراً و 76% من ذوي البشرة البيضاء، وذلك لتوليد بيانات مصطنعة، وقد كانت النتيجة مجموعة ضخمة من الصور تبدو طبيعية وحقيقية ولكن 93% منها ذكور و99% منها لذوي بشرة بيضاء

هذا الأمر ينسحب على جميع أنواع البيانات المصطنعة التي يتم توليدها وفي مختلف القطاعات، وهذا يعني أنواعاً مختلفة من التحيزات، وبالتالي المزيد من التوقعات والقرارات الخاطئة. وهنا ندق جرس الإنذار إلى وجوب بذل الجهود والاستثمار في ضمان حيادية البيانات المصطنعة، أو الذهاب أبعد من ذلك إلى تصحيح التحيزات الموجودة في البيانات الحقيقية والتخلص منها للوصول إلى ممارسات أكثر عدلاً وشفافية

خبير إداري *