وجهات نظر مُقدَّمة من المسؤولين

جودة البيانات، وأداء الذكاء الاصطناعي، والثقة

Data & Trust Alliance

IBM

في ما يتعلق بالذكاء الاصطناعي، **تشير جودة البيانات** بشكل عام إلى دقة واكتمال واتساق مجموعة معينة من البيانات، وحسن توقيتها، وتميزها، وصلاحيتها، فضلاً عن ملاءمة البيانات للغرض الذي تُستخدم من أجله، ولكن سيعتمد المعنى العملي لعبارة "بيانات عالية الجودة" في أي سياق معين على احتياجات المنظمة وحالة الاستخدام المحددة.

في اقتصاد اليوم، تعمل البيانات على توجيه كل قرار تتخذه الشركات، بدءًا من تطوير المنتجات، ثم إدارة سلسلة التوريد، وحتى إصدار فواتير العملاء. ويمكن أن يؤدي الفشل في فهم وإدارة جودة بياناتك إلى مخاطر قانونية ومالية كبيرة، ومخاطر متعلقة بالسمعة، وفي النهاية، سيحد ذلك من قدرة المؤسسة على الابتكار والازدهار.

"البيانات هي غذاء الذكاء الاصطناعي. إذا كنت لا تعرف جودة البيانات، فلن تعرف جودة نتائج الذكاء الاصطناعي التي تحصل عليها."

—**جوان ستونير،** زميلة في مجال البيانات والذكاء الاصطناعي في Mastercard

أصبحت هذه المخاوف ملحّة لا سيما مع تسابق قادة الأعمال لاعتماد الذكاء الاصطناعي بسبب ما يتيحه من الأتمتة والابتكار. تعد نماذج الذكاء الاصطناعي وأدوات العمل التي تقدمها فعالة وآمنة، مثل البيانات التي يجري تدريبها عليها، والتي تستمر في تعلمها منها أثناء تفاعلها مع العملاء، ومصادر المعلومات الأخرى. تقول جوان ستونير، زميلة في مجال البيانات والذكاء الاصطناعي في Mastercard: "البيانات هي غذاء الذكاء الاصطناعي." "إذا كنت لا تعرف جودة البيانات، فلن تعرف جودة نتائج الذكاء الاصطناعي التي تحصل عليها."

بسيطة، الخوارزميات المدربة على البيانات غير الصالحة تؤدي إلى نتائج سيئة للأعمال. نظرًا لأن معظم المؤسسات اليوم تعتمد على البيانات، الخاصة بها إلى جانب البيانات التي تصل إليها من جهات خارجية، يحتاج قادة الأعمال إلى اتباع نهج مدروس وقائم على القيمة لتحقيق جودة البيانات. هناك ثلاث خطوات بالغة الأهمية في هذه العملية.

الخطوة الأولى
تحديد معنى جودة البيانات لشركتك

يعد تعريف "الجودة" أمرًا وثيق الصلة بالسياق بالنسبة لأي منظمة؛ لأنه يعتمد على الأهداف التجارية لكل شركة وحالات استخدامها. تقول إيوانا مازار، نائبة رئيس استراتيجية بيانات المؤسسة في UPS: "ما يهم في مختلف الشركات هو مجالات التركيز المختلفة والجوانب المختلفة لجودة البيانات."

بغض النظر عن الاختلافات السياقية، هناك اعتبارات أساسية مهمة لكل منظمة عندما يتعلق الأمر بالبيانات المستخدمة في تطبيقات الذكاء الاصطناعي. يحدد المنتدى الاقتصادي العالمي ستة عناصر أساسية لجودة البيانات ينبغي تقييمها عند التدريب وفحص الذكاء الاصطناعي، سواء تم بناؤه داخليًا أو شراؤه من أحد البائعين:

- **الدقة:** هل يمكنك تأكيد وإثبات أن البيانات التي تستخدمها تمثل موضوعات فعلية في العالم؟
- **الاكتمال:** هل يمكنك التأكد من أن البيانات شاملة وأن جميع قيمها أو حقولها كاملة؟

- **الاتساق:** هل يمكنك ضمان أن البيانات المخزنة في أماكن متعددة في الشبكات والتطبيقات، مستقرة ومتسقة في تنسيقها والقيم التي تمثلها؟

- **حسن التوقيت:** هل يمكنك قياس التأخير بين وقت إنشاء البيانات ووقت استخدامها، والتأكد من أن التأخير لا يؤثر على دقة البيانات؟

- **التفرُّد:** هل يمكنك تحديد أي تكرار أو تداخلات في مجموعات البيانات التي تستخدمها لتدريب نماذج الذكاء الاصطناعي لديك؟

- **الصلاحية:** هل يمكنك ضمان تسجيل بياناتك بالنسق السليم أو الصياغة الصحيحة المناسبة للغرض المقصود أو حالة الاستخدام، بما في ذلك البيانات الوصفية، مثل أنواع البيانات الصحيحة ونطاقاتها وأنماطها؟

الوصفية، يمكن للشركات المساعدة في الكشف عن البيانات غير المهمة أو المكررة والتخلص منها، بالإضافة إلى البيانات التي ربما جرى التلاعب بها أو "تسليحها" أو "تسميمها".

الملاءمة للغرض، وتتعلق بما إذا كانت البيانات تمثل المجموعات المناسبة من الناس أو السوق أو العوامل الملائمة لحالة استخدام أدوات الذكاء الاصطناعي التي يجري تدريبها. إذا كانت البيانات التي تختارها لتدريب الذكاء الاصطناعي لديك ليست مناسبة لغرضك، فسوف يفشل الذكاء الاصطناعي. على سبيل المثال، إذا جرى استخدام خوارزمية الخدمات المالية المدربة على بيانات السوق الأوروبية لتحليل الأسواق في أفريقيا، فلن يعمل النموذج بفعالية.

والأهم من ذلك أن مجموعة البيانات القديمة قد تعكس تحيزات سابقة، فالذكاء الاصطناعي المدرب على تلك البيانات سوف يؤدي إلى استمرار هذه التحيزات. يقول ماناف ميسرا، كبير مسؤولي البيانات

"في الشركات المختلفة، ما يهم هو مجالات التركيز المختلفة والجوانب المختلفة لجودة البيانات."

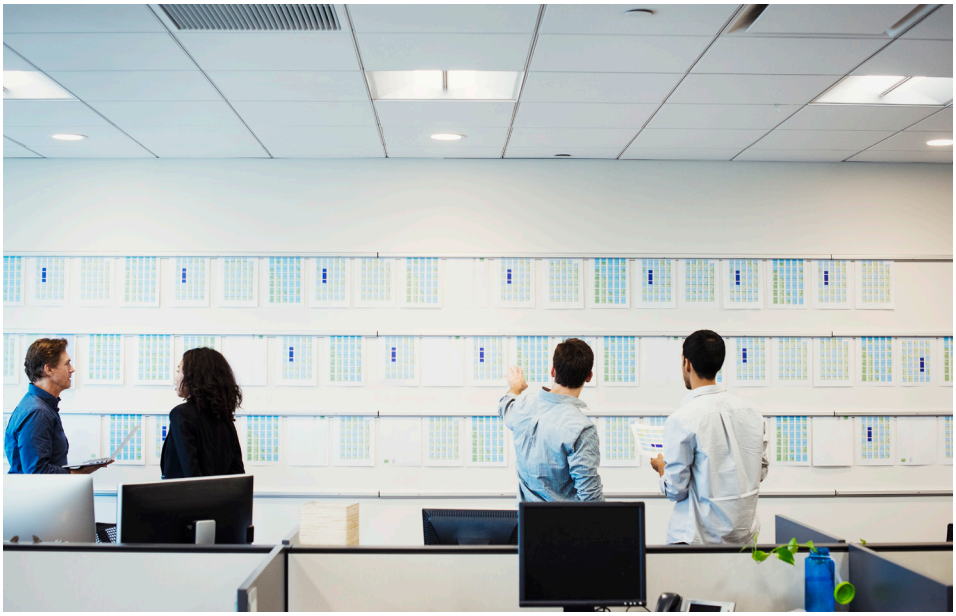
—**إيوانا مازاري**، نائبة رئيس استراتيجية بيانات المؤسسة في UPS

في حين أن هناك جوانب أخرى لجودة البيانات التي قد تكون مهمة بالنسبة لحالات الاستخدام الفردية للمنظمات، فإن هذه الاعتبارات الستة توفر صورة عامة لما يعتبر "بيانات جيدة".

هناك بُعدان إضافيان أساسيان لضمان جودة البيانات الخاصة للذكاء الاصطناعي:

مصدر البيانات، وتعرِّفه رابطة الشبكة العالمية على أنه "معلومات عن الكيانات والأنشطة والأشخاص المشاركين في إنتاج جزء من بيانات أو شيء معين، ويمكن استخدامها لإعداد تقييم عن جودتها أو موثوقيتها أو جدارتها بالثقة". يشمل المصدر تفاصيل عن كيفية وصول مؤسستك إلى البيانات وأي حقوق قانونية ذات صلة، وهو أمر مهم لا سيما مع بيانات الجهات الخارجية. يتم تسجيل المصدر جزئيًا في البيانات الوصفية التي توضح تفاصيل عن أصل حزمة البيانات وأي تغييرات جرت عليها، بالإضافة إلى الطوابع الزمنية. ومن خلال أتمتة إنشاء البيانات والتحليلات في بنك Regions: "يتمثل جزء مهم من مسؤولية البيانات في ضمان حصولنا على مجموعة كاملة ووفيرة من البيانات التي تعكس بدقة جميع المجتمعات التي نخدمها". "عندما ندرّب الأدوات والنماذج الاصطناعية برؤية شاملة حقًا، فإننا نلتزم بالاستخدام الأخلاقي والمسؤول للبيانات، وتظهر النتائج في صورة ملاءمة أكبر وشمولية مالية للجميع."

تقول دينا مندلسون، مسؤولة الخصوصية ومديرة الامتثال في شركة Transcarent، وهي شركة ناشئة في الرعاية الصحية الرقمية: "تمثل مسائل المصدر والملاءمة مصدر قلق دائم يؤرقنا". تقول مندلسون: "ينبغي علينا عدم المبالغة في الحماس تجاه مجموعة البيانات الكبيرة للغاية التي يمكن أن تحل ظاهريًا جميع مشكلاتنا." وتضيف قائلة: "نحن بحاجة إلى معرفة مصدر البيانات، ومن يتحكم في البيانات، وما هي الحقوق الفردية أو حقوق الشركات المطبقة، وما إذا كانت ذات صلة بغرضنا من الاستخدام وستوصلنا إلى النتيجة المرجوة."



يعد مصدر البيانات وملاءمتها للغرض أمرًا مهمًا لا سيما في سياق الجيل الجديد من الذكاء الاصطناعي التوليدي المتاح للجمهور، والذي يجري تدريبه على مجموعة هائلة من البيانات من الإنترنت. تقول ستونير: "إن امتلاك القدرة على فهم مصدر بياناتك ونسبها سيصبح أكثر أهمية بالنسبة للشركات التي ترغب في استخدام الذكاء الاصطناعي التوليدي لأغراض تجارية."

يقول ديفيد كوكس، نائب رئيس قسم نماذج الذكاء الاصطناعي في IBM Research®، إن المصدر على وجه الخصوص هو العنصر الأساسي. يقول كوكس: "يجب على كل قائد أعمال أن يتساءل من أين تأتي بيانات الذكاء الاصطناعي الخاصة به، وما إذا كان قد تم تنظيفها وإعدادها للتدريب". ويضيف قائلاً: "إذا حدثت مشكلة غدًا في مجموعة معينة من البيانات، فستحتاج المؤسسات إلى الالتزام بحذف البيانات وإعادة تدريب النموذج".

إن كبار مقدمي خدمات تخزين البيانات وأنظمة الإدارة والتحليل، بما في ذلك AWS، وAzure، وGoogle Cloud، وIBM، بالإضافة إلى العديد من الشركات المتخصصة، يقدمون أدوات لإدارة مصدر البيانات.

الخطوة الثانية:
إنشاء معايير وحوكمة على مستوى المؤسسة

يُعد إنشاء معايير متسقة لجودة البيانات أمرًا ضروريًا للاستخدام الأخلاقي والمسؤول للذكاء الاصطناعي. ومع ذلك لا توجد معايير عالمية لجودة البيانات في الأعمال التجارية؛ لأن جودة البيانات تعني لكل شركة أمرًا مختلفًا. لذلك من المهم للمنظمات أن تحدد معايير جودة البيانات الخاصة بها، ويجب أن يُنفذ ذلك بدقة.

كيف تبدو "الدقة"؟ وضع برناردو تافاريس، كبير مسؤولي التقنية والبيانات في شركة Kenvue (المعروفة سابقًا باسم Johnson & Johnson Consumer Health)، مقياسًا لمؤشر جودة البيانات (DQI)، ويُعبَّر عنه كنسبة مئوية. يقول تافاريس: "في حالة جودة بيانات المنتج، يجري قياس DQI كنسبة مئوية من السجلات المتوفرة والدقيقة والمتصلة في جميع أنظمتنا". ويضيف قائلاً: "نحن نفعل ذلك من خلال تطبيق Business Rules التلقائية. واستغرق الأمر منا أكثر من عام لتحديد هذا المقياس". بدأ فريقه بفرقة عاملة ركزت على المنتجات الأكثر ربحية، وأوصلت البيانات إلى نسبة 95% من مؤشر DQI.

واليوم، يمكن للشركة مقارنة المنتجات، بدءًا من Listerine وانتهاءً بـ Neutrogena عن طريق النسب المئوية لمؤشر DQI. يقول تافاريس: "إنه رقم نتتبعه على مستوى الإدارة العليا، ومن السهل تذكره والإبلاغ به." بعد أن أظهر فريق تافاريس فعالية مؤشر DQI للأقسام التي تمثل قيمة عالية للعمل، تمكّن من الحصول على التمويل لتصميم أدوات مفتوحة المصدر لإنشاء "السجل الذهبي" وتنقيته، والبدء في توسيع نطاق مبادرتهم. يقول تافاريس: "إننا نعمل ببطء ولكن بثقة على ابتكار فكرة مؤشر DQI لجميع كياناتنا الأساسية".

معايير جودة البيانات هي الأساس لإدارة البيانات. غالبًا ما يكون كبير مسؤولي البيانات (CDO) في المنظمة مسؤولاً عن وضع تلك المعايير، وتنفيذ إطار عمل الإدارة لتطبيقها، بالإضافة إلى العمل مع محللي وعلماء البيانات، ومطوري البرامج، والمحامين، وفرق الامتثال، وأي شخص يستخدم أدوات الذكاء الاصطناعي في المؤسسة.

"لا تحاول حل العديد من التحديات والمشكلات في وقت واحد. اختر الجانب الأهم والأكثر خطورة، وابدأ بسؤال رئيسي لا تستطيع الإجابة عليه؛ لأن بياناتك ليست جيدة."

- **برناردو تافاريس**، رئيس قسم التقنية ومسؤول البيانات في Kenvue

وفقًا لتقرير حديث لشركة Gartner، فإن زيادة الاستثمار في البيانات والتحليلات، والطلبات المفروضة على فرق البيانات في المنظمة "تعكس الثقة المتزايدة في إمكانيات كبار مسؤولي البيانات والتحليلات، والاعتراف بأن دور مكتب البيانات لا غنى عنه في الأعمال، ولكن يؤدي ذلك إلى مزيد من العمل مع تزايد الضغط على البيانات والتحليلات لتحقيق نتائج ملموسة للأعمال."

إذا لم يكن لدى منظمتك منصب كبير مسؤولي البيانات، فلا تدَع ذلك يمنعك من التصرف، فكما تقول ستونير "من المحتمل أن لديك شخص ما في شركتك كان يراقب جودة البيانات لفترة من الوقت". وتنصح الرؤساء التنفيذيين "بالاهتمام بأي مناصب أو وظائف حالية تتمتع بالخبرة. يمكن أن يكون رئيسًا تنفيذيًا للمعلومات أو شخصًا في قسم تقنية المعلومات لديك، أو شخصًا متخصصًا في مجال ذكاء الأعمال".

وبعيدًا عن عمليات إدارة جودة البيانات، فإن تأثير الإمكانيات الجديدة الذي ينتشر بصورة متزايدة سيتطلب تعزيز ثقافة تركّز على البيانات والذكاء الاصطناعي.

الخطوة الثالثة
ابتكار ثقافة إدارة البيانات على مستوى الشركة

يقول تافاريس: "كان علينا أن نبتكر ثقافة البيانات." ويضيف قائلاً: "بدأ هذا بالإقرار بأن البيانات أصل حقيقي، وإدراك أننا يجب أن يتوفر لدينا أشخاص متحمسون في هذا المجال، والأدوات والهياكل اللازمة لإدارة البيانات من خلال دورات الحياة".

يُعد بناء ثقافة البيانات أمرًا ضروريًا؛ لأن مهمة إدارة البيانات لا تنتهي أبدًا. ستتغير نماذج الذكاء الاصطناعي باستمرار، وكذلك المتطلبات التنظيمية، ولا يمكن تحقيق الإدارة المستمرة بمجرد وضع السياسات. يجب على الإدارة العليا أن تقود جهود التعليم المستمر في مجال جودة البيانات، ويجب أن يقودها المدير التنفيذي للشركة، فمن المهم أن يأتي التوجيه من الإدارة العليا، أي من المدير التنفيذي إلى الموظفين؛ لأنه يجب دمج معايير الجودة في كل وظيفة.

وعلى الرغم من أن بناء ثقافة على مستوى الشركة يعد مشروعًا طويل الأمد، بل ليس له نهاية حقًا، من المحتمل أن يبدأ صغيرًا. يقول تافاريس: "لا تحاول حل العديد من التحديات والمشكلات في وقت واحد." ويضيف قائلاً: "اختر الجانب الأهم والأكثر خطورة، وابدأ بسؤال رئيسي لا تستطيع الإجابة عليه؛ لأن بياناتك ليست جيدة."

بالنسبة لتافاريس، أدى ذلك إلى التركيز على منتجات الشركة التي حققت فيها أصول البيانات أكبر قيمة للشركة.

وحتى ذلك كان يمثل تحديًا. يقول تافاريس: "اجتمع لدينا خبراء وأطراف معنية و70 شخصًا، ولم يتمكنوا من الاتفاق على تعريف مشترك للبيانات الخاصة بالمنتج." ويضيف قائلاً: "إذا كنت تعمل في مجال البحث والتطوير، أو في مجال المبيعات، وما إلى ذلك، فإنك تنظر إلى الأمر بشكل مختلف." لقد توصل الفريق إلى نموذج يحدد العناصر والعلاقات الأساسية التي تمثل حقًا منتج Kenvue. عندئذ بدأوا العمل على جودة البيانات عن طريق "ملء" هذا النموذج ببيانات نظيفة ومتصلة. وبمساعدة الذكاء الاصطناعي والأتمتة وهدف تحقيق جودة البيانات على مستوى الشركة، كان هذا النهج التكراري فعالاً، وجرى تطبيقه لاحقًا على العديد من مجالات "Data360".

وأخيرًا، من الأمور الأساسية لبناء ثقافة بيانات سليمة هي أن تتذكر أن كل جزء من البيانات يمثل إنسانًا: عميلك، وموظفك، وشريكك، ووجوده وأنشطته في العالم. إن التركيز على الأشخاص الموجودين في النظام البنائي لجودة البيانات، وهم الأشخاص الذين ينشئون ويستخدمون ويتمتعون بخدمات أدوات الذكاء الاصطناعي التي تستند إليها البيانات، سيسمح لشركتك بإدارة قراراتها بشكل أفضل فيما يتعلق بالاستخدام المناسب للبيانات في الذكاء الاصطناعي. تقول ستونير: "إذا أعددتَ التصميم مع وضع الإنسان في الاعتبار، فستصبح حواجز الحماية واضحة." ●

أسئلة رئيسية موجهة للإدارة العليا

- ما الذي نطلق عليه **بيانات حساسة؟**
- كيف نقيس حاليًا **سلامة بياناتنا** الخاصة بالذكاء الاصطناعي؟
- هل نقيّم بشكل مستمر **معايير البيانات لدينا** مع تغير احتياجات العمل؟
- ما **التعليم والتدريب** المتوفران لدينا لتوجيه الموظفين فيما يتعلق بالاستخدام المناسب للبيانات؟

تعرّف على المزيد عن تحالف البيانات والثقة

احصل على رؤى من الخبراء عن الذكاء الاصطناعي للأعمال