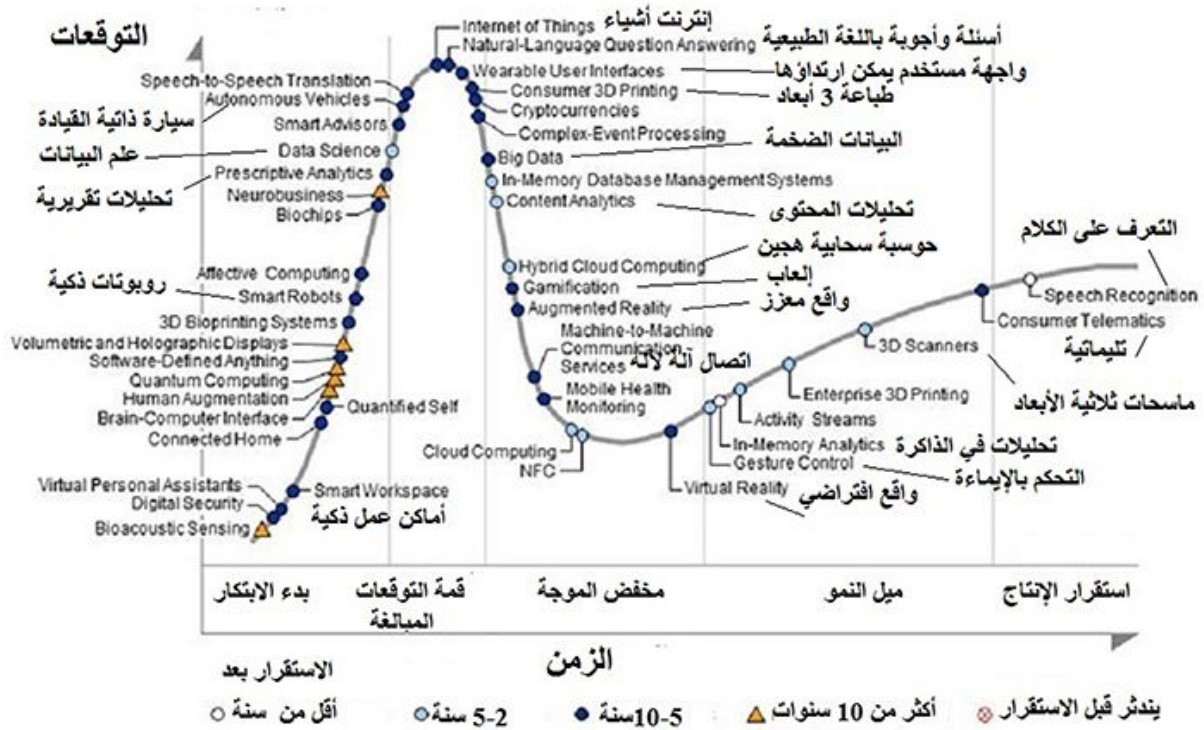


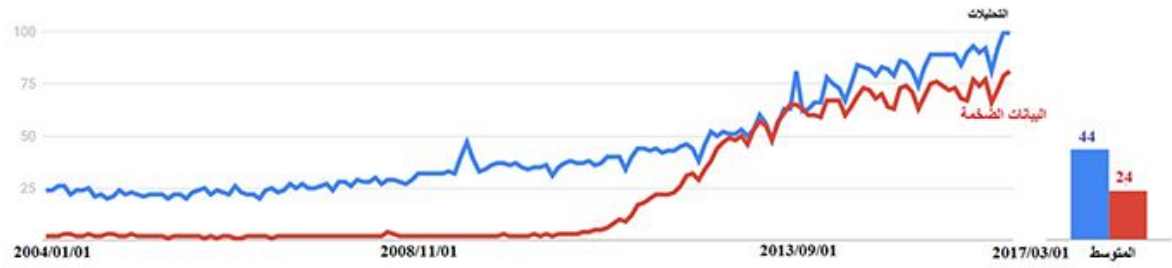
ظهر في السنوات الأخيرة مصطلح البيانات الضخمة (1) لوصف نموذج جديد لتطبيقات البيانات. وتميل موجة التقانات الجديدة إلى الظهور مع كثير من الضجيج، لكن يمكن أن نستغرق بعض الوقت لنقول ما هو جديد ومختلف. ويبين الشكل رقم (1) منحني موجة التوقعات التقانية عامي 2011 م و2014م، وفيها تجاوزت تقانة البيانات الضخمة قمة التوقعات خلال هذه المدة، ويبين الشكل رقم (2) تنامي اتجاهات الاهتمام بهذه التقانة والتقانات الأخرى ذات الصلة خلال السنوات الخمس الأخيرة.

ويوفر نموذج البيانات الضخمة حلاً فاعلاً وقابلاً للتعامل مع تزايد حجم البيانات لاكتشاف الأنماط والمؤشرات والدلائل أو غيرها من المعلومات والمعرفة القادرة على جعل البيانات الضخمة قابلة للإدارة، ومربحة، وأداة فاعلة للتنمية.

الشكل رقم (1): منحني موجة التوقعات التقانية



الشكل رقم (2): اتجاهات الاهتمام بالبيانات الضخمة والتحليلات المنطقية خلال المدة (2004-2017م)



خصائص البيانات الضخمة

بينما يجري تعريف البيانات الضخمة في عددٍ لا يُحصى من الطرائق، فإن أساس خصائص نموذج البيانات الضخمة أنها كبيرة جداً (الحجم)، وتصل بسرعة عالية (السرعة)، وتتغير كثيراً (التغير)، وتحتوي على كثير من الموضوعات (المصادقية)، وهي أيضاً متنوعة كثيراً (التنوع) لدرجة لا يمكن معالجتها داخل هيكل الحوسبة باستخدام النهج والتقنيات التقليدية. ولدى التقانات التي يجري تقديمها لدعم هذا النموذج مجموعة واسعة من الواجهات، وهو ما يجعل من الصعب بناء الأدوات والتطبيقات التي تدمج البيانات الضخمة من مصادرها المتعددة.

بداية المصطلح والتطور

كان أول ظهور لمصطلح (البيانات الضخمة) في بدايات عام 2000م، لكن زادت التوقعات حديثاً انتشار المصطلح من ضمن أوائل الاتجاهات التقنية المقبلة؛ إذ توقعت أهميتها البازغة وانتشارها مراكز الأبحاث التقنية؛ مثل: جارتتر، وماكينزي، وشركة إي بي إم العملاق التقني. ولم يخل الأمر من اهتمام دوائر سياسية كثيرة بهذا الموضوع التقني؛ مثل: إدارة الرئيس الأمريكي السابق أوباما، والمفوضية الأوروبية التي عدت البيانات الضخمة رصيماً أساسياً للاقتصاد، والمجتمع الأوروبي على غرار الموارد البشرية والمالية والطبيعية الكلاسيكية. وركزت دوائر علمية أبحاثها في هذا المجال؛ مثل: مؤسسة العلوم الوطنية الأمريكية، ومجلس الأبحاث الهندسية والعلوم الطبيعية الكندي، ومعهد مهندسي الكهرباء والإلكترونيات الأمريكي، وبرنامج الأبحاث والابتكار الأوروبي، ومجلة الطبيعة، ومجلة العلوم، وقطاع الأعمال والاقتصاد (مثل منتدى دافوس الاقتصادي). وكذلك انتشر الحديث عن البيانات الضخمة في وسائل الإعلام؛ مثل: نيويورك تايمز، وويل ستريت جورنال، والإيكونوميست، وهذا العدد من مجلة الفيسل العلمية.

في عام 2007م، كانت البشرية قادرة على تخزين 290 إكسا بايت من البيانات المضغوطة، وتستطيع التواصل تقريباً عبر 2×10^{21} زيتاً بايت)، وتنفيذ 6.4 إكسا بايت من التعليمات في الثانية على أجهزة الحاسب للأغراض العامة، ونمت القدرة الحاسوبية العامة الغرض بمعدل سنوي قدره 58%. وارتفعت قدرة العالم على القيام بالاتصالات الثنائية الاتجاه بنسبة 28% سنوياً، تلتها زيادة في المعلومات المخزنة عالمياً بنسبة 23%. وشهدت قدرة البشرية على انتشار المعلومات الأحادي الاتجاه من خلال قنوات البث نمواً سنوياً متواضعاً نسبياً (6%)، وهيمنت القناة الرقمية على الاتصالات السلكية واللاسلكية منذ عام 1990م حتى بلغت 99.9% عام 2007م، وكانت أغلبية الذاكرة التقنية لدينا في هيئة رقمية منذ أوائل عام 2000م حتى بلغت 94% عام 2007م (2). وستتجاوز كمية البيانات الرقمية

المنتجة خلال السنوات الثماني المقبلة 40 زيتا بايت، وهو ما يعادل 5200 جيجا بايت من البيانات لكل رجل وامرأة وطفل على وجه الأرض (الجدول رقم: 1، والشكلان رقما: 3، و4). ولوضع الأمور في نصابها 40 زيتا بايت (3) هو 40 تريليون جيجابايت. وتشير التقديرات إلى أن هذه الكمية تبلغ 57 أضعاف عدد كل حبات الرمال على جميع الشواطئ على وجه الأرض. ومن المتوقع أن تتضاعف جميع البيانات كل عامين حتى عام 2020م. ومن الآن حتى عام 2020م لن ينتج البشر معظم البيانات، بل تنتجها الأجهزة التي ستحدث معاً عبر شبكات البيانات، ويشمل ذلك على سبيل المثال: أجهزة الاستشعار، والأجهزة الذكية التي تتواصل مع الأجهزة الأخرى (تواصلًا مباشرًا، ومن آلة إلى آلة في إنترنت الأشياء والمدن الذكية والسيارات الذاتية القيادة). لكن حتى الآن، لم يتم اكتشاف سوى جزء صغير جداً من قيمة البيانات التي تم إنتاجها من خلال استخدام (تحليلات البيانات). ويقدر بحلول عام 2020م أن 33% من جميع البيانات ستحتوي على معلومات قد تكون ذات قيمة إذا جرى تحليلها.

الجدول رقم (1): وحدات أحجام البيانات

الجدول رقم (1): وحدات أحجام البيانات

الحجم	عدد البايت	القيمة الأسية العشرية
يوتّا بايت	1000.000.000.000.000.000.000.000	
زيتّا بايت	1000.000.000.000.000.000.000.000	
إكسا بايت	1000.000.000.000.000.000	
بيتا بايت	1000.000.000.000.000	
ثيرا بايت	1000.000.000.000	
جيجا بايت	1000.000.000	
ميغا بايت	1000.000	
كيلو بايت	1000	

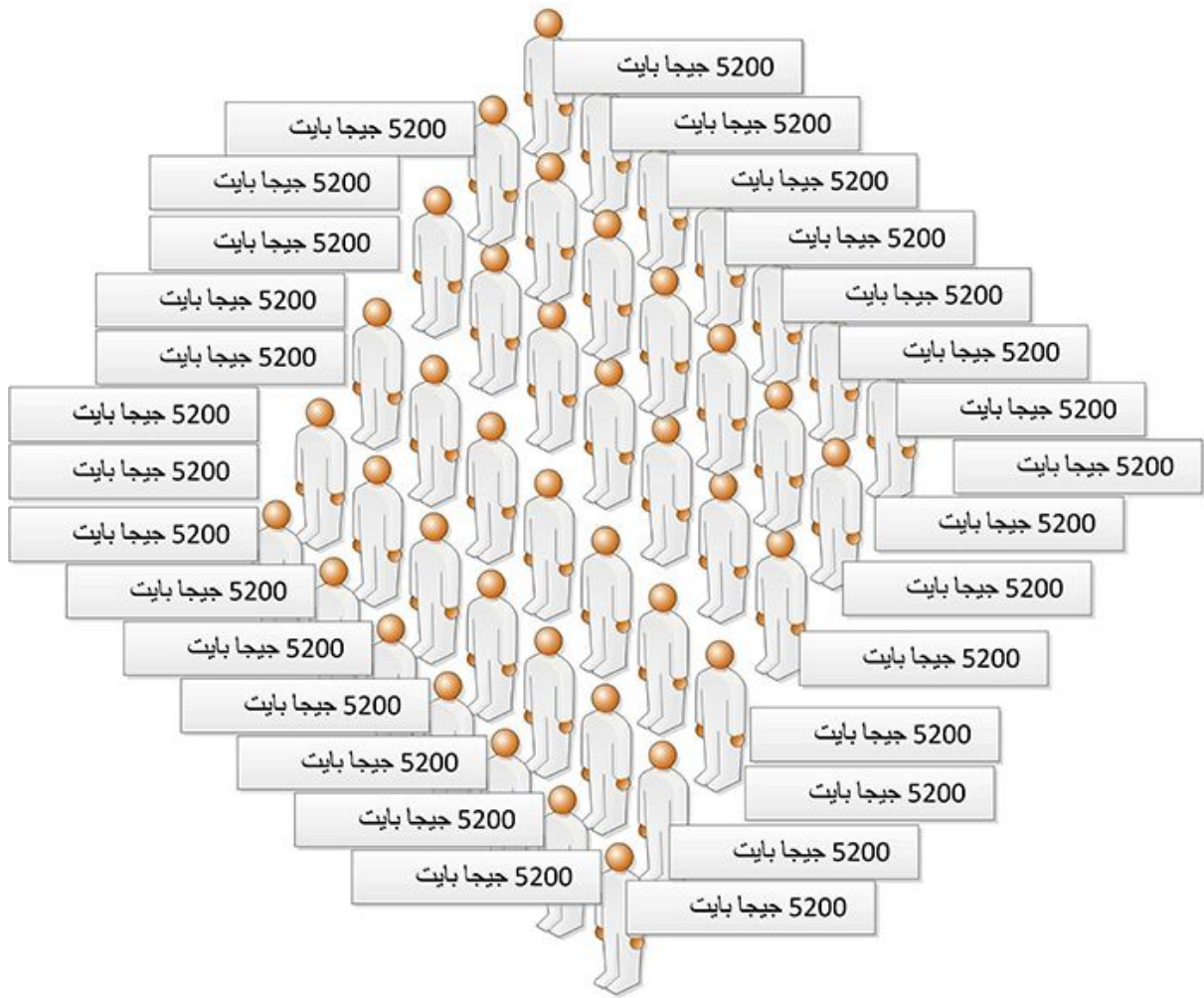
الشكل رقم (3): مقارنة أحجام البيانات لجينوم الإنسان، ودماع فأر، ودماع إنسان، وحاسوب عملاق، وكمية البيانات المنتجة عام 2017م



كوادريليون بايت = بيتا بايت

الشكل رقم (4): كمية البيانات الرقمية المنتجة 40 زيتا بايت

وهي تعادل 5200 جيجا بايت من البيانات لكل رجل وامرأة وطفل على وجه الأرض



يشمل العالم الرقمي كل شيء؛ من الصور ومقاطع المرئيات على الهواتف النقالة التي يتم تحميلها على اليوتيوب إلى الأفلام الرقمية ومحتويات أجهزة التلفاز العالية الوضوح. ويشمل طبيعته الحال- مزيداً من البيانات للشركات التقليدية؛ مثل: البيانات المصرفية وأجهزة الصراف الآلي، وتسجيلات كاميرات الأمن في المطارات والأحداث الكبرى مثل دورات الألعاب الأولمبية، فضلاً عن بيانات الاصطدام دون الذري التي سجلتها منظمة الأبحاث النووية الأوروبية (سيرن) في مصادم (هادران الكبير). وباستخدام (تحليلات البيانات) يمكن أن نكتشف أنماط الاستخدام في وسائل الإعلام الاجتماعية، والعلاقات المتبادلة في البيانات العلمية والطبية من الدراسات المنفصلة وتقاطع المعلومات الطبية مع البيانات الاجتماعية، فضلاً عن وجوه بلايين الناس في لقطات كاميرات الأمن. إضافةً إلى ذلك، فإن البيانات التي تمّ التققيب فيها لا بد من (وسمها)(4) مع (البيانات الوصفية)(5) لإعطائها سياقاً، ويشمل ذلك على سبيل المثال: إضافة التاريخ إلى مرئية أو معلومات تحديد الموقع الجغرافي لصور أو مرئيات الهاتف الذكي، أو الحصول على معلومات مفيدة من مخازن البيانات الضخمة مثل الاتجاهات الشرائية للمستهلك.

مكّن النّقدّم في خدمات تقانة المعلومات والاتصالات والإنترنت الناس من إنشاء بيانات ضخمة ونقلها وتخزينها مع تزايد وتيرة هذا النمو الهائل باستمرار، وأصبحت قواعد البيانات كبيرة جداً ومعقّدة، ويجري توالدها بسرعة، لدرجة أنه ثبت عدم كفاية الأساليب التقليدية لمعالجة البيانات وفشلت ملاحقة هذا النمو المتعدد.

آفاق استخدام البيانات الضخمة

هناك اتفاق واسع النطاق بين قادة التجارة والأكاديميين والحكومات حول إمكانيات البيانات الضخمة في إثارة الابتكار، وتنمية التجارة، والعمل محركاً للتقدم والتنمية، وتوفّر موارد البيانات الضخمة القدرة على الإجابة عن أسئلة كانت الإجابة عنها من قبل بعيدة المنال، ومن ذلك:

– الصحة:

يمكن لتحليلات البيانات الضخمة أن تقلّل من تكاليف العناية الصحية، وتحسّن الخدمات الصحية والوقاية من الأمراض؛ فالإجابة عن سؤال (كيف يمكن مواجهة وباء محتمل شديد الأثر من خلال الكشف في وقت مبكر بما يكفي للتدخل في الوقت المناسب؟) ستؤدي إلى انطلاق ثورة مفيدة في العناية الصحية الإلكترونية، منها على سبيل المثال: استخدام بيانات الهاتف المحمول لتتبع المرضى في مراكز العلاج، وتحليل البيانات لتحديد أفضل علاج لكل حالة معينة، وتحديد أنماط الأعراض الجانبية للأدوية، واختيار الطبيب والممرّض المناسبين لحالة المرض والمريض، وخفض التكاليف مع الجودة العالية في الوقت نفسه، ومساندة الابتكار والبحوث والتطوير في الأدوية والعلاج واستخدام الحوسبة الإدراكية. ومما لا شك فيه أن البيانات المتراكمة في السجلات الصحية هي مصدر غني لتحسين الصحة العامة؛ لكي تقهم المجتمعات المحلية احتياجاتها الخاصة، ومن أجل اكتشاف أوجه عدم المساواة والتفاوت في الرعاية داخل المجتمع كله (الشكل رقم: 5).

الشكل رقم (5): نظام عناية صحية باستخدام البيانات الضخمة والتحليلات



تشير التقديرات إلى أنه يمكن لصناعة الرعاية الصحية توفير مليارات الدولارات باستخدام تحليلات البيانات الضخمة للتنبؤ ثم استخراج كنز من المعلومات من سجلات الصحة الإلكترونية، ومطالبات التأمين، والوصفات الطبية، والدراسات السريرية، والتقارير الحكومية، ونتائج المختبرات. ويمكن استخدام تحليلات البيانات الضخمة في مراجعة البيانات السريرية؛ حتى تكون قرارات العلاج قائمة على أفضل البيانات المتاحة بدلاً من تقدير الأطباء فقط. ويمكن تخفيض وقت الانتظار الطويل في المستشفيات لغرفة بمجرد حساب التنبؤ متى يمكن أن تصبح الأسرة فارغة، ويمكن احتواء تفشي الإنفلونزا إذا قامت السلطات الصحية بتتبع أرقام ومواقع أولئك الذين أصابهم المرض في البيانات الضخمة. وسيحصل المرضى على خدمة صحية عالية الكفاءة، ومثال ذلك ما تقوم به مؤسسة كايزر الصحية في كاليفورنيا (6)، التي تخدم تسعة ملايين عضو، ويعمل بها 17 ألف طبيب، ولها 611 فرعاً، و37 مستشفى. وفي مجال الرعاية الصحية الوقائية، لدى كايزر أكثر السجلات الصحية الإلكترونية في الولايات المتحدة الأمريكية، واحتل تخزين البيانات لديها نحو 30 بيتا بايت، وتبلغ هذه السجلات الضعف تقريباً كل سنتين. وعلى الرغم من بدء إنشاء السجلات الإلكترونية منذ عقد من الزمان، إلا أنها وضعت أخيراً أدوات تحليلية لاستخراج معنى من المعلومات، وتحسين الرعاية الصحية؛ فعلى سبيل المثال: تقوم المؤسسة بعمليات البحث في الرسوم البيانية الطبية الإلكترونية لنحو 1500 فرد أو نحو ذلك من الأفراد الذين زاروا أو اتصلوا بالمؤسسة، ويمكنها أن تنتج تقارير يومية للمرضى الذين تتطلب رعايتهم متابعة مستمرة؛ مثل: اختبارات الدم، أو التحصين. ويبحث البرنامج أيضاً عن الثغرات في الرعاية باستخدام المعرفة المستخرجة من الأدلة السريرية التي توفر أفضل ممارسات الرعاية؛ مثل: كيفية مراقبة الدم، ومستويات السكر لمرضى السكر، وبدلاً من رؤية 20 مريضاً فقط في اليوم أصبح الأطباء قادرين على رؤية نحو 75-100 مريض.

هذه الأنواع من أدوات دعم الطبيب والطب، التي تستخدم جنباً إلى جنب مع السجلات الطبية الإلكترونية، هي مستقبل الرعاية الصحية، وتستخدم كايزر أيضاً تحليلات الصحة التنبؤية لتحسين الإجراءات في المستشفيات؛ لأن الرعاية تُقاس بالدقيقة، ولها إجراءات حاسمة لفعل الشيء الصحيح في الوقت الصحيح، كما تقوم بتطوير برامج لمنع إخفاق المرضى في المستشفى، والتنبؤ بطول الإقامة بالمستشفى، وخلق نظم الإنذار المبكر لتحديد المضاعفات بعد إجراء العمليات، والحد من إعادة قبول مرضى سبقت لهم الإصابة بالحالة نفسها.

وفي الوقت نفسه، يقوم علماء الحاسب في جامعة كاليفورنيا باستخدام البيانات الضخمة لعلاج أفضل للمرضى الذين يعانون مرض باركنسون، واضطراب الجهاز العصبي التقدمي، وأنشأ الفريق خوارزمية تقوم بتحليلات منطقية للبيانات من أجهزة الاستشعار التي تتبع حركات المريض، بما في ذلك أجهزة الاستشعار الثلاثية الأبعاد، مماثلةً لتلك المستخدمة في نظام ألعاب كينكت من مايكروسوفت (7)، والهاتف الذكي، ومتحسسات الجسم المدمجة في أردية إلكترونية. وتقوم أجهزة الاستشعار برصد تقدّم المرض والعلاج بفاعلية في الوقت الحقيقي؛ فإذا انخفض مدى الحركة أو المرونة يتم تنبيه مقدّمي الرعاية لذلك، وقد يصفون أدوية مختلفة، أو يحاولون مع المريض عمل تمارين عضلات أخرى (الشكل رقم 6).

الشكل رقم (6): رداء إلكتروني ذكي مدمج فيه أقطاب كهربائية واستشعار للحصول على البيانات الصحية وإرسالها لاسلكياً عبر الإنترنت لإجراء التحليلات المنطقية واتخاذ القرار الصائب



ومن ناحية دولية، يقوم برنامج فلونيت (8) لمنظمة الصحة العالمية بجمع بيانات الإنفلونزا المقدّمة من مراكز مراقبة الإنفلونزا الوطنية في العالم وغيرها من مختبرات تتّبع الإنفلونزا الوطنية، ويجري تحميل واستخدام هذه المعلومات لتتّبع حركة الفيروسات عالمياً، وتفسير البيانات الوبائية في الوقت الحقيقي، وعرضها متاحةً للجمهور في مختلف الأشكال، بما في ذلك الجداول، والخرائط، والرسوم البيانية.

– الاقتصاد:

قريباً، ستكون الأشياء من حولنا، وربما حتى ملابسنا أيضاً، مرتبطةً مع شبكة الإنترنت عن طريق المتحسسات الإلكترونية المدمجة، وتشير التقديرات إلى أنه خلال عشر سنوات سيكون هناك 150 مليار جهاز استشعار قياس متّصلة بالشبكات، وسيبلغ عدد سكان الأرض 20 مرة مما هو عليه الحال الآن؛ لذلك فإن كمية البيانات ستتضاعف باطراد كلّ ساعة، وكثير من الشركات تحاول بالفعل تحويل هذه البيانات الضخمة إلى أموال ضخمة.

تأتي الاستخدامات الأكثر شيوعاً للبيانات الضخمة من الشرائح لتتّبع عمليات الأعمال والنتائج، وإنشاء مجموعة واسعة من نماذج التوقعات التنبؤية، وكان لتحليل البيانات الضخمة أثر كبير في كثير من قطاعات الاقتصاد، بما في ذلك التجزئة والتمويل والإعلان والتأمين. ويمكن للبيانات الضخمة الجديدة أن تؤثر في السياسة والبحوث الاقتصادية. ومن منظور السياسات الاقتصادية، تشكّل البيانات الضخمة قدرةً على

استيعاب البيانات ومعالجتها في الوقت الحقيقي، وقيمةً لتحسين فاعلية العمليات في القطاع الخاص والحكومي، وإرشاد سياستهما الاقتصادية. ومن منظور البحوث الاقتصادية، يمكن تصميم البحوث الجديدة ومراقبة النتائج الإضافية للأحداث أو السياسات الاقتصادية، وأتاحت أدوات البيانات الضخمة، التي يجري تطويرها في مجال الإحصاء وعلوم الحاسوب، مثل: التعلّم الإحصائي، وتقنيات استخراج البيانات، كثيراً من التطبيقات المهمة في الاقتصاد؛ فعلى سبيل المثال: لننظر في البيانات التي تمّ جمعها من متاجر البيع بالتجزئة للتوضيح؛ فقبل بضعة عقود ربما جمعت المتاجر بيانات عن المبيعات اليومية، وكان يمكن أن تُعدّ ذات جودة عالية إذا تمّ تقسيم البيانات حسب المنتجات أو فئاتها، بينما في الوقت الحاضر تتيح بيانات الماسح الضوئي تتبّع عمليات الشراء الفردية ومبيعات السلع، والنقاط الوقت الدقيق الذي حدثت فيه، وتاريخ شراء الأفراد، واستخدام بيانات المخزون الإلكتروني لربط المشتريات بمواقع محددة في المخزن على الرفوف أو بمستويات المخزون الحالية. ولا يراقب تجار التجزئة على الإنترنت هذه المعلومات فحسب، بل يمكنهم أيضاً تتبّع سلوك المستهلك بعد البيع، بما في ذلك طلب البحث الأولي، والعناصر التي تمّت مشاهدتها، والتخلص منها، والتوصيات أو العروض الترويجية التي تمّ عرضها، والمراجعات اللاحقة للمنتج أو البائع.

ومن ناحية المبدأ، يمكن ربط هذه البيانات بالعوامل السكانية، والإعلانات، وأنشطة الشبكات الاجتماعية، والإنفاق، أو تاريخ الائتمان، وهو ما أحدث تطوراً موازياً في النشاط التجاري. ومع نقل الشركات عملياتها اليومية إلى أجهزة الحاسب، ثم على شبكة الإنترنت، أصبح من الممكن تجميع مجموعات بيانات غنية عن المبيعات، والتعاقد مع العمال، وشحنات البضائع. وعلى نحو متزايد، هناك أيضاً سجلات إلكترونية لجهود العمل التعاوني، وتقييمات الموظفين، ومقاييس الإنتاجية، ويمكن قول القصة نفسها أيضاً عن القطاع العام من ناحية القدرة على الوصول إلى القرارات الضريبية، وتحليلها، وبرامج التأمين الاجتماعي، والنفقات، والأنشطة التنظيمية.

وهناك مسألة أخرى مهمة أيضاً، وهي تتعلق بتدريب الاقتصاديين على العمل مع مجموعات البيانات الضخمة، ومختلف البرمجيات والأدوات الإحصائية المطلوبة عادةً؛ لإجراء تحليل أفضل لبيانات العملاء، والتسعير عبر دورات حياة المنتج مع المستهلك، أو لتحسين التسويق والإنفاق بالتنبؤ بالمناطق؛ إذ ستكون عروض المنتج أكثر فاعليةً. ومن فرص تحليلات البيانات الضخمة استخدام أمثل للبيانات الواردة من الهواتف المحمولة في الوقت الحقيقي لتعقّب تحركات المتسوقين في المراكز التجارية، وتحليل أنماطهم السلوكية، أو التنبؤ ببنيتهم، وتحديد الأسعار المناسبة، وتوافر تحليلات بيانات مستوى استخدام المنتج، التي تنقلها الرقائق الدقيقة المدمجة في المنتج، جنباً إلى جنب مع بيانات الشبكات الاجتماعية؛ لتحسين منتجات الجيل المقبل، وتطويرها، وإنشاء خدمات ما بعد البيع، وخدمة العروض، ومعرفة تفضيلات العملاء الفردية لاتخاذ قرارات صائبة موضوعية.

ولثورة البيانات تأثير في البحوث الاقتصادية من نواحي نطاق النتائج ونوعيتها، والأساليب المستخدمة، وتدريب الاقتصاديين التجريبيين، وسيكون أول تأثير، وهو أوضح ما يكون، في السماح بمقاييس أفضل للآثار والنتائج الاقتصادية. ويمكن أن تساعد البيانات على طرح أنواع جديدة من الأسئلة أكثر دقة وشمولية، وتمكين التصميمات البحثية الجديدة التي تمكّننا من معرفة واقعية بآثار السياسات والأحداث الاقتصادية المختلفة. وليس هناك شكّ يذكر - على الأقلّ في عقولنا - تجاه أن البيانات الضخمة المقبلة ستغير مشهد السياسة والبحوث الاقتصادية، ويعتقد أن تحليل البيانات الضخمة لن يحلّ محلّ الحسّ السليم، والنظرية الاقتصادية، أو الحاجة إلى تصميم بحوث دقيقة، بل سيكملها.

– الأمن:

بسبب التقنية المتاحة الآن؛ مثل: الهواتف الذكية، وكاميرات المراقبة، وأجهزة الاستشعار البيومترية، يمكن كشف علامات؛ مثل: بصمات الأصابع، وملامح الوجه والعيون، ومزيد من البيانات حول الأفراد أكثر مما سبق. ويمكن أن يؤدي تحليل هذه البيانات إلى إنفاذ القانون ومنع الجرائم قبل حدوثها؛ فقد انتقل الآن أسلوب الشرطة الأمريكية من استخدام دبابيس على الخرائط في الزمن السابق إلى وضع علامات إذ وقعت جرائم الشوارع من أجل تخصيص الدوريات الراجلة لإنفاذ القانون، واستخدام برامج الحاسب التي تقوم بتحليل البيانات، وتحديد المناطق التي على الأغلب ستقع فيها الجرائم. وعلى الرغم من أنه يمكن لتلك البرامج الكشف عن النشاط الجنائي إلا أنه لا يزال هناك بعض القصور في منع حدوث الجريمة؛ فالبيانات في كثير من الأحيان تكرر ما يعرفه ضابط الشرطة بالفعل؛ مثل: أن كثيراً من الجرائم تحدث غالباً يوم الجمعة ليلاً عند خروج الناس، أو في الأماكن التي تباع فيها المخدرات غير المشروعة، لكن (تحليل البيانات) ستكون أكثر فائدة عندما يمكن الكشف عن معلومات أكثر تعقيداً، أو تلك التي لا يكون مسؤولو الشرطة قادرين على تحديدها من تلقاء أنفسهم. ويمكن أن تستخدم برامج التحليلات التي يمكنها معالجة البيانات الضخمة التي جُمعت في سجلات من كثير من الأماكن، وأجهزة الاستشعار المنتشرة في كل مكان، في استكمال ممارسات إنفاذ القانون. واستخدمت وكالة الاستخبارات الأمريكية برمجيات تستخرج تنبؤات من البيانات الضخمة من إحدى الشركات المتخصصة (بالانتير)(9) لمنع الهجمات الإرهابية، ويمكنها من خلال دراسة كميات ضخمة من البيانات المتاحة بالفعل بشأن الإرهاب والمشتبه بهم تحليل هذه البيانات مجتمعةً لربط النقاط معاً، والوصول إلى المعلومات بخصوص ما قد يحدث. ويمكن أن تشمل البيانات الضخمة الحمض النووي للمشتبه بهم، ومعلومات الوجه المستمدة من المراقبة الآلية الملحقة بآلات الصراف الآلي، ورخص تأجير السيارات، وأرقام لوحات السيارات التي تم رصدها في مختلف المواقع، وسجلات الهاتف، والأمكنة التي زارها المشتبه بهم. وكشف البرنامج في إحدى الحالات تخطيطاً لشبكات إرهابية تستخدم القنابل في هجماتها على عدة أماكن، وعثر على مشتبه بهم في جريمة قتل وكيل الجمارك الأمريكية. وكانت الشبكات الاجتماعية مفيدة أيضاً؛ لما توفره من كميات هائلة من بيانات عن الجمهور تستفيد منها الشرطة من خلال تمشيط هذه البيانات، ويمكن للبرنامج البحث عن كلمات رئيسة محددة، والسلوكيات التي يمكن أن تشير إلى نشاط غير قانوني؛ فالبرامج ليست فقط لكشف المؤامرات أو ارتكاب السرقات أو صفقات المخدرات، وإنما أيضاً لتحديد أولئك الذين قد يرتكبونها، وزمان حدوث هذه الجرائم ومكانها.

مثال آخر: مشروع بيانات الجرائم في جامعة بنسلفانيا في فيلادلفيا؛ إذ يقوم فريق في قسم علم الجريمة(10) بتطوير خوارزمية لتوقع ضحية جريمة قتل استناداً إلى مجموعة متنوعة من البيانات، بما في ذلك تقارير أقسام الشرطة المحلية. وبدلاً من استهداف القاتل المحتمل، تقوم الشرطة بتحذير الضحايا المحتملين، وتقديم المشورة حول كيفية حماية أنفسهم. وطوّر القسم سابقاً برمجيات للمساعدة على تحديد الإفراج المشروط لأي من السجناء الذين يمكن أن يطلق سراحهم؛ لأنه من غير المحتمل أن يرتكب هذا السجين جريمة مرةً أخرى. وتستند هذه التنبؤات إلى 24 متغيراً من المتغيرات البيانية، بما في ذلك الجنائية والسجلات والأعمار التي ارتكبت فيها الجرائم، وتستخدم نحو 80 % من الإفراج المشروط في الولايات المتحدة الآن أنظمة مماثلة، وتبين أن نسبة الانتكاسة لا تتعدى 15% باستخدام تحليلات البيانات الضخمة.

مخاطر وتحديات

من ناحية أخرى، لا تخلو هذه الفوائد والفرص من تحديات ومخاطر لا يمكن إغفالها، مثل:

— الديمقراطية:

من شأن البيانات الضخمة والاضطرابات التقانية البازغة أن تغير طريقة تنظيم المجتمع واتجاهاته؛ ففيما يتعلّق بمشكلة الديمقراطية: العبارة السحرية للبيانات الضخمة تعني (الدفع الضخم)؛ أي أنها مزيج من البيانات الضخمة مع الدفع. ويبدو لدى كثيرين أن ذلك نوعاً من العصا السحرية الرقمية يدفع أحداً ما إلى أن يحكم الجماهير بكفاية من دون الحاجة إلى إشراكهم في العمليات الديمقراطية، فهل يمكن لهذا الشخص أن يتغلب على المصالح المكتسبة للجماهير، ويحسن مسار العالم؟ إذا كان الأمر كذلك فإن المواطنين يمكن أن يحكمهم (حاكم حكيم) مُسلّح ومتمكّن بما لديه من قوة البيانات الضخمة، وقادر على تحقيق النتائج الاقتصادية والاجتماعية المرجوة تقريباً، كما لو كانت لديه عصا سحرية رقمية. لكن نظرة واحدة إلى الأدبيات العلمية ذات الصلة تبين أن محاولات السيطرة على الآراء؛ بمعنى: (تحسينها)، محكوم عليها بالإخفاق بسبب تعقيد المشكلة؛ لأن ديناميات تشكيل الآراء مملوءة بالمفاجآت، ولا أحد يعرف كيف ينبغي استخدام عصا سحرية رقمية.

وبغض النظر عن ذلك، فإن المجرمين والإرهابيين والمتطرفين سيحاولون السيطرة على عصا سحرية رقمية عاجلاً أم آجلاً، وربما حتى من دون أن نلاحظ، وقد تم بالفعل اختراق جميع الشركات والمؤسسات، حتى البنّاجون، والبيت الأبيض، ووكالة الأمن القومي الأمريكية. وتتشأ مشكلة أخرى عندما نفترق إلى الشفافية الكافية، والرقابة الديمقراطية، وتآكل النظام من الداخل، ويمكن أن تتأثر خوارزميات البحث على الويب ونتائج البحث، ويمكن للشركات محاولة جمع مجموعات معينة من الكلمات للحصول على نتائج أكثر ملاءمة لهم، وربما تكون الحكومات قادرة أيضاً على التأثير في النتائج أيضاً؛ فخلال الانتخابات مثلاً— قد يدفعون الناخبين المتردّدين إلى دعمهم، وهو التلاعب الذي يصعب الكشف عنه؛ لذلك فمن يسيطر على هذه التقنية يمكن أن يفوز في الانتخابات عن طريق (دفع) أنفسهم إلى السلطة. ومما يزيد من تفاقم هذه المشكلة أن محرك بحث واحد أو منصة وسائط التواصل الاجتماعي في كثير من البلدان لها حصة سوقية سائدة؛ لذلك يمكن أن تؤثر هذه البيانات بشكل حاسم في اتجاهات الجمهور، وتتدخل في شؤون البلدان عن بُعد.

وعلى الرغم من أن حكم محكمة العدل الأوروبية الصادر في 6 أكتوبر عام 2015م يحدّ من تصدير البيانات الأوروبية غير المحظورة، إلا أنه لم يتم حلّ المشكلة الكامنة في أوروبا، بل حتى في أمكنة أخرى. ومن أجل التلاعب للبقاء من دون أن يلاحظ أحد يجري تفصيل اقتراحات كافية لكل فرد؛ فهذه الطريقة يتم تعزيز الاتجاهات المحلية تدريجياً من خلال التكرار، وهو ما يؤدي إلى حدوث تأثير (فقاعة التصفية)(11) أو (تأثير صدى الغرفة)(12).

في النهاية، كلّ ما قد حصل هو آراء خاصة بالمستخدم تنعكس إليه مرة أخرى، يمكنها أن تعزّز وجهات نظره السياسية والاجتماعية الحالية من دون أن تسمح له بالتفكير بشكل مختلف. ويعود (تأثير الصدى) إلى خوارزميات (التعلم الآلي) من البيانات الضخمة التي تستخدمها شركات مثل: جوجل، وفيسبوك، التي تهدف إلى خدمة محتوى مستخدميها المصمّم خصيصاً لمصالحهم. وتتيح الخوارزميات المخصّصة للمستخدمين إمكانية العثور على المعلومات التي يرغبون فيها من دون التدقيق في صفحات المحتوى غير ذي الصلة، كما تسمح الخوارزميات للشركات بمطابقة المستخدمين بدقة أكبر مع الإعلانات التي تتماشى مع تفضيلاتهم، وهو ما يؤدي إلى تحسين تدفق

الأرباح. والشئ نفسه يمكن أن تقوم به الحكومات، لكنه قد يسبب الاستقطاب الاجتماعي، ويؤدي إلى تشكيل مجموعات منفصلة لم يُدْفَعُ يفهم بعضها بعضاً، وتجد نفسها على نحو متزايد في صراع معاً. وبهذه الطريقة، يمكن للمعلومات الشخصية أن تدمر عن غير قصد التماسك الاجتماعي في الأمة، ويمكن ملاحظة ذلك حالياً في السياسة الأمريكية؛ إذ ينحرف الديمقراطيون والجمهوريون على نحو متزايد لتصبح الحلول الوسطية السياسية بينهم مستحيلة تقريباً، والنتيجة هي تفتت المجتمع، وربما حتى تفككه، وخطورة الأمر على الديمقراطية(13).

– الخصوصية:

يُعدّ تعقّب القضايا الشخصية والمراقبة لتحليل زيارات صفحات (الويب) أو الشبكات الاجتماعية ومكالمات الهاتف والبريد الإلكتروني، وتتبع ومراقبة الميول الدينية أو السياسية أو الإرهاب من مخاطر البيانات الضخمة؛ فعلى سبيل المثال: طالب الرئيس الأمريكي السابق أوباما ورئيس الوزراء البريطاني السابق كامرون شركات التقنية (مثل جوجل وفيسبوك) بالتعاون مع الاستخبارات في تتبّع الإرهابيين على شبكات التواصل الاجتماعي والإنترنت(14)، لكن من ناحيته أخرى: أثار هذا الإجراء احتجاج منظمات حقوق الإنسان؛ لما يمثله من تجاوز للخصوصية الشخصية. وتشمل البيانات الضخمة جمع وتحليل البيانات الشخصية عن الأفراد، والمعلومات السكانية، والأنشطة التجارية والحكومية والعسكرية، واستهلاك المياه، والطاقة، والتقارير الوطنية لأغراض مختلفة، والاعتداء على الملكية الفكرية عبر الإنترنت وشبكات التواصل الاجتماعي والبريد الإلكتروني والاشتراكات الحرة في المواقع. وتواجه تقانات وخدمات البيانات الضخمة تحدي حماية خصوصية الأفراد وبياناتهم الحساسة في أثناء دورة المعالجات، وفي الوقت نفسه الاحتفاظ بهذه البيانات مخزنة، وتمهّد قابلية التوسّع السبيل أمام مزيدٍ من التهديدات للأمن والخصوصية. وأكبر عقبة أمام استخدام البيانات الضخمة في تنبؤ النشاط الإجرامي هو أن المبرمجين ومنفذي القانون غير متعاونين.

وهناك تحدّ آخر هو تحديد ما يجب القيام به عندما يشير تحليل البيانات إلى أن شخصاً ما في سبيله إلى ارتكاب جريمة؛ إذ يمكن للدّعاء أن يطلب من القاضي وضع شخص ما تحت الإقامة الجبرية أو السجن إذا كانت هناك الأدلة المادية الكافية، لكن اعتقال شخص بالاعتماد على تحليلات البيانات الضخمة يمكن أن يكون أكثر صعوبة في إقناع القاضي؛ فلعّل البيانات والبرمجيات لا تظهر دائماً الصورة كاملةً، مع أن برامج البيانات الضخمة والتقانات المصاحبة لها تقوم بتوفير البيانات والمعلومات مقدماً، وتساهم في إنفاذ القانون ومنع الجريمة، لكن قبل أن نتمكّن من تحقيق ذلك تحتاج البرامج إلى التحسين، والإجابة عن أسئلة مهمة، مثل: الآثار في الخصوصية الشخصية.

– الاقتصاد:

هناك كثير من التحديات التي تواجه الاقتصاديين الراغبين في الاستفادة من مجموعات البيانات الضخمة والجديدة، وتشمل هذه التحديات سهولة الوصول إلى البيانات الضخمة ودقتها، وتطوير قدرات إدارة البيانات والبرمجة اللازمة للعمل مع مجموعات البيانات الضخمة على نطاق واسع، وأخيراً – وهو الأهم – التفكير في المناهج الإبداعية لتلخيص ووصف واستخراج المعرفة الواردة في هذه البيانات.

– البيانات:

هناك بعض التحديات الجديدة المرتبطة بالبيانات الضخمة، منها:

- الحصول على بيانات الحكومة والقطاع الخاص، فضلاً عن الموارد الحاسوبية اللازمة.
- تمثّل معالجة أحجام من البيانات سريعة الازدياد تحدياً صعباً أيضاً، مع تزايد معدلات تدفق البيانات بسرعة هائلة.
- كيف تتكامل مجموعة أوسع ومختلفة من أشكال البيانات والهياكل والجداول الزمنية والدلالات؟
- عدم التجانس وعدم الاكتمال: يمكن أن تفوّت البيانات التي تمّت معالجتها بتقانات البيانات الضخمة بعض سمات البيانات أو تُدخل ضوضاء في أثناء نقلها حتى بعد تنظيف البيانات وتصحيح الخطأ، ومن المرجح أن يظلّ هناك بعض النقص والأخطاء فيها، ويمكن أن تُدار هذه التحديات خلال تحليل البيانات.
- الحجم: تنامي حجم البيانات أسرع من تنامي التقانات، مثل: قواعد البيانات المتوازنة، وقواعد البيانات داخل الذاكرة، وقواعد البيانات (إس كيو إل)، وخوارزميات التحليلات، ويمثّل حجم البيانات الهائل التحدي الأكثر إلحاحاً لدى الهياكل التقليدية لتقانة المعلومات، إضافةً إلى مكان تخزين هذه البيانات الضخمة.
- التوقيت: يُعدّ معدل الحصول والتوقيت المناسب للعثور على عناصر في وقت محدود في قاعدة بيانات ضخمة تحدياً آخر جديداً في معالجة البيانات، وكذلك الحاجة إلى أنواع من المعايير الجديدة للتمكّن من الاستجابة لطلبات من البيانات تتميز بأوقات محدودة هو تحدّ إضافي.
- الخسارة المالية والسمعة نتيجة خروقات البيانات الضخمة.

أسئلة أساسية

على الرغم من الاتفاق الواسع النطاق على الفرص الكامنة والقيود الحالية التي تطرحها البيانات الضخمة إلا أن هناك عدم توافق مستمراً في الآراء بشأن بعض الأسئلة الأساسية المهمة، وهو ما يسبب الخلط بين المستخدمين المحتملين وإحباط فرص التقدم، ومن هذه الأسئلة:

- ما سمات حلول البيانات الضخمة؟
- كيف تختلف البيانات الضخمة عن بيانات البيانات التقليدية والتطبيقات ذات الصلة؟
- كيف لخصائص البيانات الضخمة أن تتكامل مع البنى التحتية المنتشرة حالياً؟
- ما التحديات المركزية التقانية العلمية والتوحيد القياسي اللازم لمعالجة الحلول القوية للبيانات الضخمة وتسريعها ونشرها؟