

عشر قواعد بسيطة لبحوث مسؤولية في البيانات الضخمة

ماثيو زوك، سولون باروكاس، دانا بويد، كيت كروفورد، إميلي كيلر، سيتا بينيا غانغادهاران، أليسا غودمان، راشيل هولاندر، باربرا أ. كونيف، جاكوب ميتكالف، أرفيند نارايانان، ألون درا نيلسون، فرانك باسكوالي



المخلص:

يقدم هذا المقال، الصادر عن أعضاء في مجلس البيانات الضخمة والأخلاق والمجتمع الذي مولته المؤسسة الوطنية للعلوم الأمريكية، عشر قواعد للتعامل مع المسائل الأخلاقية المعقدة التي تثيرها بحوث البيانات الضخمة، انطلاقاً من أن كل بحث في الظواهر الاجتماعية والطبية والنفسية والاقتصادية يتعامل مع بشر، وأن على الباحثين مسؤولية تقليل الضرر المحتمل. تتناول القواعد الخمس الأولى سبل الحد من الضرر: الإقرار بأن البيانات تمثل أناساً، وفهم الخصوصية بوصفها سياقاً لا ثنائية، والحماية من إعادة تحديد الهوية، والمشاركة الأخلاقية للبيانات، ومراعاة حدود البيانات. وتتناول القواعد الخمس الأخرى بناء أفضل الممارسات: مناقشة الخيارات الأخلاقية الصعبة، ووضع مدونات سلوك، وتصميم البيانات والأنظمة لتكون قابلة للتدقيق، والنظر في العواقب الأوسع، ومعرفة متى يجوز الخروج على القواعد. ويخلص المقال إلى أن البحث المسؤول لا يعني منع البحث، بل ضمان سلامته ودقته وتعظيم نفعه وتقليل ضرره.

الكلمات المفتاحية: البيانات الضخمة، أخلاقيات البحث، الخصوصية، إعادة تحديد الهوية، مشاركة البيانات

مقدمة:

شهد استخدام أساليب بحوث البيانات الضخمة (big data) نمواً هائلاً في السنوات الخمس الماضية في الأوساط الأكاديمية وفي قطاع الصناعة على حد سواء. ومع ازدياد حجم مجموعات البيانات المتاحة وتعقيدها، ازدادت كذلك الأسئلة الأخلاقية التي تثيرها بحوث البيانات الضخمة. وتعد هذه الأسئلة أكثر إلحاحاً كلما تجاوزت البيانات والأجندات البحثية ما هو معتاد في العلوم الحاسوبية والطبيعية، لتتناول على نحو أكثر مباشرة جوانب حساسة من السلوك البشري والتفاعل والصحة. فأدوات بحوث البيانات الضخمة تتغلغل على نحو متزايد في حياتنا اليومية، ومن

ذلك التقيب في السجلات الطبية الرقمية بحثًا عن رؤى علمية واقتصادية، ورسم خرائط العلاقات عبر وسائل التواصل الاجتماعي، والتقاط كلام الأفراد وأفعالهم عبر أجهزة الاستشعار، وتتبع الحركة في المكان، وتشكيل سياسات الشرطة والأمن عبر «الشرطة التنبؤية»، وغير ذلك كثير.

غير أن الإمكانيات المفيدة للبيانات الضخمة في العلم والصناعة تقابلها تحديات جديدة تواجه الباحثين، وكثيرًا ما تقع خارج نطاق تدريبهم وما اعتادوه. فعلماء الاجتماع باتوا يتعاملون اليوم مع هياكل البيانات والحوسبة السحابية، في حين يتعين على علماء الحاسوب التعامل مع بروتوكولات حماية الأشخاص موضوع البحث ومجالس المراجعة المؤسسية (institutional review boards - IRBs). ومع أن الصلة بين المعطى الفردي والأشخاص الحقيقيين قد تبدو مجردة إلى حد بعيد، فإن نطاق كثير من أشكال البيانات الضخمة وحجمها وتعقيدها يخلق منظومة غنية يندمج فيها المشاركون من البشر ومجتمعاتهم اندماجًا عميقًا، ويكونون فيها عرضة للضرر. وهذا التعقيد يتحدى أي مجموعة معيارية من القواعد، ويجعل وضع إرشادات عامة أمرًا عسيرًا.

ومع ذلك، فإن الحاجة إلى توجيه في مجال بحوث البيانات الضخمة المسؤولة واضحة، ويقدم هذا المقال مجموعة من «عشر قواعد بسيطة» للتعامل مع المسائل الأخلاقية المعقدة التي ستتشأ لا محالة. وعلى غرار سلسلة القواعد المستمرة في مجلة (PLOS Computational Biology)، تنطوي التوصيات التي نعرضها على دقائق أكثر مما توحى به كلمتا «بسيطة» و«قواعد». وترتبط هذه الدقائق حتمًا بالمقدمة التي تنطلق منها مقالتنا: إن كل بحث في البيانات الضخمة يتناول ظواهر اجتماعية أو طبية أو نفسية أو اقتصادية يتعامل مع بشر بوصفهم مشاركين في البحث، وعلى الباحثين مسؤولية أخلاقية في تقليل الضرر المحتمل إلى أدنى حد.

إن تنوع مصادر البيانات وموضوعات البحث والمقاربات المنهجية في البيانات الضخمة ينفي إمكان وضع قائمة تحقق واحدة تناسب الجميع؛ ولذلك جاءت هذه القواعد أقل تحديدًا مما قد يأمله بعضهم. بل إننا نحث الباحثين على إدراك المشاركين من البشر والأنظمة المعقدة التي تنطوي عليها بياناتهم، وعلى جعل معالجة الأسئلة الأخلاقية جزءًا من سير عملهم المعتاد. ولتحقيق هذه الغاية، بنينا القواعد الخمس الأولى حول كيفية تقليل احتمال الضرر الناجم عن ممارسات بحوث البيانات الضخمة؛ أما القواعد الخمس الأخرى فتركز على السبل التي يستطيع بها الباحثون الإسهام في بناء أفضل الممارسات الملائمة لمقارباتهم التخصصية والمنهجية. وفي صميم هذه القواعد، ندعو باحثي البيانات الضخمة الذين يرون أن بياناتهم منفصلة عن القدرة على إلحاق الضرر إلى إعادة النظر في افتراضاتهم. وتبين الأمثلة الواردة في هذه المقالة مدى تكرار الحالات التي أفضت بيانات تبدو بريئة ومجهلة الهوية إلى أسئلة أخلاقية غير متوقعة وآثار ضارة.

وهذه المقالة ثمرة مشروع امتد عامين بتمويل من المؤسسة الوطنية للعلوم - (National Science Foundation - NSF)، أسس مجلس البيانات الضخمة والأخلاق والمجتمع (Council for Big Data, Ethics, and Society)، وهو مجموعة من 20 باحثًا من طيف واسع من العلوم الاجتماعية والطبيعية والحاسوبية. وقد كلف المجلس بتقديم التوجيه إلى المؤسسة الوطنية للعلوم بشأن أفضل السبل لتشجيع الممارسات الأخلاقية في البحث العلمي والهندسي الذي يستخدم أساليب بحوث البيانات الضخمة وبُناها التحتية [1].

1. أقر بأن البيانات تمثل أناسًا وبأنها قد تلحق الضرر:

من أكثر القواعد جوهرية في بحوث البيانات الضخمة المسؤولية الإدراك الراسخ بأن معظم البيانات تمثل أناسًا أو تؤثر فيهم. فمجرد الانطلاق من افتراض أن كل البيانات تمثل أناسًا إلى أن يثبت العكس يضع صعوبة فصل البيانات عن أفراد بعينهم في صدارة الاهتمام. وهذا المنطق واضح في مجموعات البيانات «الخطرة»، مثل محتوى وسائل التواصل الاجتماعي ذي اللغة التحريضية، لكن حتى البيانات التي تبدو غير ضارة قد تحتوي على معلومات حساسة وخاصة؛ إذ يمكن مثلًا استخراج بيانات عن معدلات ضربات القلب الدقيقة لأشخاص من مقاطع فيديو على يوتيوب [2]. بل إن البيانات التي تبدو لا علاقة لها بالناس قد تؤثر في حياة الأفراد بطرق غير متوقعة، مثل بيانات علوم المحيطات التي تغير ملامح المخاطر التي تواجهها المجتمعات وقيم العقارات، أو سجلات صيغة ملفات الصور القابلة للتبادل (Exchangeable Image Format - EXIF) المرفقة بالصور، التي تحتوي على إحداثيات الموقع وتكشف تنقلات المصور أو حتى موقع منزله.

وقد ينجم الضرر أيضًا عن استخدام مجموعات بيانات تبدو بريئة في تشكيل حياة الأفراد أو وصم جماعات، بآثار تمتد إلى السكان كافة، وكثيرًا ما يحدث ذلك دون إجراءات تتيح التظلم [3,4]. فعلى سبيل المثال، يمكن لخرائط الشبكات الاجتماعية في خدمات مثل تويتر أن تحدد الجدارة الائتمانية [5]، ويمكن لدرجات احتمال العودة إلى الجريمة غير الشفافة أن تشكل قرارات العدالة الجنائية بطريقة متفاوتة عرقياً [6]، كما أن التصنيف القائم على الرموز البريدية أدى إلى تقليل وصول الأمريكيين من أصل أفريقي في مدن الولايات المتحدة إلى خدمة التوصيل في اليوم نفسه من (Amazon Prime) [7]. وتُظهر هذه الحالات البارزة أن البيانات التي تبدو محايدة قد تفضي إلى نتائج تمييزية، فتُفاقم أوجه عدم المساواة الاجتماعية.

وتُظهر حالات أخرى أن مجموعات البيانات «العامة» يمكن تكييفها بسهولة لبحوث شديدة التطفل بدمج بيانات أخرى معها، ومن ذلك استخدام هاغ (Hague) وزملائه [8] سجلات العقارات وتقنيات الترميز الجغرافي لتحديد هوية الفنان المتخفي باسم مستعار بانكسي (Banksy) كما يُزعم [9]. وعلى وجه الخصوص، فإن البيانات التي لا تحكمها ممارسات موافقة فعلية، سواء أكانت محتوى وسائل التواصل الاجتماعي أم بقايا الحمض النووي التي نخلفها وراءنا باستمرار، قد تبدو عامة، لكنها قد تتسبب في انتهاكات غير مقصودة للخصوصية وفي أضرار أخرى [9,10].

انطلق من افتراض أن البيانات تمثل أناسًا (إلى أن يثبت العكس)، واجعل ذلك موجّهًا لتحليلك. لا أحد يُعفى تلقائيًا من المسؤولية الأخلاقية.

2. أدرك أن الخصوصية أكثر من قيمة ثنائية:

تعدّ انتهاكات الخصوصية من الوسائل الرئيسية التي قد تلحق بها بحوث البيانات الضخمة الضرر، ومن المهم إدراك أن الخصوصية سياقية [11] وظرفية [12]، ولا يمكن اختزالها في ثنائية بسيطة بين العام والخاص. فكون شيء ما قد نُشر علنًا لا يعني أن أي استخدام لاحق له سيكون خاليًا من الإشكالات. فالنظر إلى صورة واحدة نشرها فرد على إنستغرام له تبعات أخلاقية تختلف عن النظر في السجل الكامل لكل منشورات شخص ما على وسائل التواصل الاجتماعي. وتتوقف الخصوصية على طبيعة البيانات، والسياق الذي أنشئت فيه وجمعت، وتوقعات المتأثرين بها وأعرافهم. واعلم أن موقفك من الاستخدام المقبول ومن الخصوصية قد لا يتوافق مع موقف أصحاب البيانات التي تستخدمها، إذ تختلف تفضيلات الخصوصية بين المجتمعات وداخلها.

فعلى سبيل المثال، يستكشف تين (Tene) وبولونيتسكي (Polonetsky) [13] كيف أن تجاوز الأعراف الاجتماعية، ولا سيما في المواقف الجديدة التي تخلقها التقنيات الحديثة، يراه الأفراد أمرًا «مريبًا» حتى حين لا ينتهك لوائح حماية البيانات أو قوانين الخصوصية. فتطبيقات وسائل التواصل الاجتماعي التي تستخدم مواقع المستخدمين لدفع المعلومات إليهم، وتتبع الشركات لنشاط الأفراد على وسائل التواصل الاجتماعي ولاتصالاتهم الخاصة لاكتساب معلومات عن العملاء، والتسويق القائم على أنماط البحث، كلها رآها بعضهم أمرًا «مريبة» أو حتى انتهاكات صريحة للخصوصية. وبالمثل، فإن تداول السجلات الصحية جزء ضروري من تلقي الرعاية الصحية، لكن هذه المشاركة نفسها تثير مخاوف أخلاقية جديدة حين تتجاوز مقدمي الرعاية لتصل إلى المسوقين.

وتتجاوز الخصوصية أيضًا الأفراد لتمتد إلى الجماعات [10]. ولهذا صدى خاص لدى المجتمعات التي كانت تاريخيًا هدفًا لسياسات تمييزية قائمة على البيانات، مثل ممارسة «الخط الأحمر» (redlining) [14, 15]. ومن الأمثلة الأخرى خرائط المجتمعات المحلية، التي أُعدت لتحديد العقارات المثيرة للمشكلات أو لتأكيد حقوق في الأرض، ثم أعاد آخرون استخدامها لتحديد فرص لإعادة التطوير أو الاستغلال [16]. ومن ثم فإن إعادة استخدام مجموعة بيانات تبدو عامة قد تتعارض مع مقاصد الخصوصية الأصلية لمن أنشئوها، وتثير تساؤلات حول ما إذا كان ذلك يمثل بحثًا مسؤولًا في البيانات الضخمة.

ضع بياناتك في سياقها لتتوقع انتهاكات الخصوصية وتقلل الضرر. فإتاحة البيانات أو كونها تبدو عامة لا يضمن انتفاء الضرر، ولا يعني أن منشئي البيانات يوافقون على استخدام الباحثين لبياناتهم.

3. احتراز من إعادة تحديد الهوية في بياناتك:

من الإشكالي افتراض أن البيانات لا يمكن إعادة التعرف على أصحابها. فثمة أمثلة كثيرة على باحثين ذوي نوايا حسنة وأساليب تبدو سليمة أخفقوا في إخفاء هوية البيانات بالقدر الكافي لمنع التعرف لاحقًا على أفراد بعينهم [17]؛ وفي حالات أخرى كانت هذه الجهود سطحية للغاية [18, 19]. وحين تُدمج مجموعات بيانات يُظن أنها مجهولة الهوية مع متغيرات أخرى، فقد يفضي ذلك إلى إعادة تحديد هوية الأشخاص على نحو غير متوقع (reidentification)، أشبه بتفاعل كيميائي ينتج عن إضافة مكّون آخر.

ومع أن إمكان الاستدلال على الهوية من تاريخ الميلاد والجنس والرمز البريدي معروف جيدًا [20]، فثمة عدد من المعايير الأخرى، ولا سيما البيانات الوصفية المرتبطة بالنشاط الرقمي، قد تكون مفيدة بالقدر نفسه أو أكثر في التعرف على الأفراد [21]. ومما يفاجئ كثيرين أن مخططات الشبكات غير الموسومة، وبيانات الموقع والحركة، وملفات الحمض النووي، وسجلات المكالمات المستمدة من بيانات الهواتف المحمولة، بل حتى صور الأقمار الصناعية عالية الدقة للأرض، يمكن استخدامها لإعادة تحديد هوية الأشخاص [22]. غير أن الأهم من تحديد المتغيرات التي تتيح إعادة تحديد الهوية هو إدراك صعوبة التعرف على مواطن الضعف هذه مسبقًا [23]. فالعوامل التي تُستبعد اليوم بوصفها غير ذات صلة أو غير ضارة بطبيعتها، مثل استهلاك البطارية، قد يتبين غدا أنها وسيلة مهمة لتحديد الهوية الشخصية [24]. فعلى سبيل المثال، يمكن لإضافة الموقع المكاني أن تحوّل منشورات وسائل التواصل الاجتماعي إلى وسيلة لتحديد موقع المنزل [25]. كما يمكن للبحث العكسي عن الصور في (Google) أن يربط بطرق غير متوقعة بين أنشطة شخصية كانت منفصلة من قبل، مثل ملفات المواعدة والملفات المهنية [26]. بل إن البيانات المتعلقة بالجماعات، أي «الإحصاءات المجمعة»، قد تكون لها تبعات خطيرة إذا كشفت مثلًا أن مجتمعات بعينها تعاني أكثر من غيرها بكثير من أمراض موصومة أو من سلوك اجتماعي موصوم [27].

حدّد الوسائل المحتملة لإعادة تحديد الهوية في بياناتك، واعمل على تقليلها في نتائج المنشورة إلى أقصى حد ممكن.

4. مارس المشاركة الأخلاقية للبيانات:

في بعض المشروعات، تكون مشاركة البيانات توقعًا لدى المشاركين من البشر المعنيين بها، ومن ثم جزءًا أساسيًا من البحث الأخلاقي. ففي بحوث الأمراض الوراثية النادرة مثلًا، تُشارك العينات البيولوجية أملًا في إيجاد علاجات، مما يجعل النشر شرطًا للمشاركة. وفي مشروعات أخرى، تقدّم مسائل الصالح العام الأوسع، وهي فئة يُقرّ بصعوبة تعريفها، حجبًا مقنعة لمشاركة البيانات، ومن ذلك قاعدة بيانات الأنماط الجينية والأنماط الظاهرية (database of Genotypes and Phenotypes - dbGaP) التي ترعاها المعاهد الوطنية للصحة (NIH)، والتي تتيح بيانات جينومية منزوعة الهوية على نطاق واسع للباحثين، فتُضفي طابعًا ديمقراطيًا على الوصول، أو حجة العدالة التي قدّمها معهد الطب (Institute of Medicine) بشأن قيمة إلزام الباحثين بمشاركة البيانات على مستوى الأفراد المستمدة من التجارب السريرية [28]. ويسهّل طلب موافقة واسعة من المشاركين، بدلًا من موافقة محددة البنية على نحو ضيق، بشأن إدارة البيانات في المراحل اللاحقة لمشاركة البيانات. ويمكن للتصميم الدقيق للبحث وتوجيهات مجالس المراجعة المؤسسية أن يساعد في توضيح إجراءات الموافقة. ومع ذلك، ننبّه إلى أنه حتى حين يُحصل على موافقة واسعة مقدّمًا، ينبغي للباحثين مراعاة المصلحة الفضلى للمشاركة البشري، والنظر استباقيًا في احتمال انتهاكات الخصوصية ومسائل إعادة تحديد الهوية. ويكتسب ذلك أهمية خاصة في بيانات الحمض النووي البشري، التي يمكن التعرف على صاحبها على نحو فريد.

غير أن هذه الأنواع من المشروعات، التي تحكم فيها الموافقة المستنيرة والحق في الانسحاب قواعد الاستخدام والمشاركة حكمًا جيدًا، باتت على نحو متزايد الاستثناء لا القاعدة في البيانات الضخمة. ففي مجتمعنا الرقمي، تلاحقنا سحب من البيانات مؤلفة من آثار الحياة اليومية، كمعاملات بطاقات الائتمان، ونتائج الفحوص الطبية، وصور الدوائر التلفزيونية المغلقة (CCTV) ومقاطعها المصورة، وتطبيقات الهواتف الذكية، وما إلى ذلك، وهي تُجمع بموجب شروط خدمة إلزامية لا بموجب تصميم بحثي مسؤول يشرف عليه مسؤولو الامتثال في الجامعات. ومع أننا قد نتمنى تطبيق معايير الموافقة المستنيرة والحق في الانسحاب، فإن مصادر البيانات الضخمة غير الرسمية هذه تجمعها جهات غير الباحث، كشركات البرمجيات الخاصة والهيئات الحكومية وشركات الاتصالات. ولا تكون هذه البيانات متاحة للباحثين إلا بعد إنشائها، مما يجعل الحصول على موافقة مستنيرة مسبقًا أمرًا مستحيلًا، كما أن الاتصال بالمشاركين من البشر بأثر رجعي لطلب الإذن كثيرًا ما يحظره مالك البيانات، أو يستحيل القيام به على نطاق واسع.

لا شك في أن الباحثين داخل شركات البرمجيات والمؤسسات الحكومية التي تجمع هذه البيانات يتحمّلون مسؤولية خاصة في معالجة الشروط التي تُجمع بموجبها البيانات؛ لكن ذلك لا يعفي المستخدم النهائي للبيانات المشاركة. وباختصار، يقع عبء الاستخدام الأخلاقي (انظر القواعد 1 إلى 3) والمشاركة الأخلاقية على عاتق الباحث، لأن شروط الخدمة التي أُنتجت بموجبها بيانات الأشخاص موضوع البحث قد تكون في الغالب فضفاضة للغاية ولا توفر إلا حماية ضئيلة من انتهاكات الخصوصية. وفي هذه الظروف، يتعيّن على الباحثين الموازنة بين متطلبات الجهات الممولة بشأن مشاركة البيانات [29] ومسؤولياتهم تجاه البشر الذين تمثّلهم البيانات التي حصلوا عليها. وينبغي للباحث إبلاغ الجهات الممولة بالمخاوف الأخلاقية المحتملة قبل بدء البحث، والاحتراز من إعادة تحديد الهوية قبل المشاركة.

شارك البيانات وفق ما تحدده بروتوكولات البحث، لكن عاج استباقياً المخاوف من الضرر المحتمل الناجم عن البيانات الضخمة المجموعة بطرق غير رسمية.

5. ضع في الحسبان مواطن قوة بياناتك وحدودها؛ فالضخامة لا تعني تلقائياً الأفضلية:

لإجراء بحوث في البيانات الضخمة تكون دقيقة ومسؤولة معاً، من المهم وضع مجموعات البيانات في سياقها الصحيح، بما في ذلك تضارب المصالح. ويؤثر السياق أيضاً في كل مرحلة من مراحل البحث: من اقتناء البيانات، إلى تنظيفها، إلى تفسير النتائج، ونشرها. وفي مرحلة اقتناء البيانات، من الضروري فهم مصدر البيانات والقواعد واللوائح التي جمعت وفقها. ويكتسي ذلك أهمية خاصة في البحوث التي تُجرى في بيئات تنظيمية متساهلة نسبياً، قد يتعارض فيها استخدام الإجابات عن أسئلة البحث مع توقعات من قَدَمُوا البيانات. ومن المقاربات الممكنة الاستفادة من الأعراف الأخلاقية المتبعة في تتبع مصدر القطع الأثرية وأصلها، وكثيراً ما يجري ذلك بالتعاون والتشارك مع المجتمعات التي جاءت منها (مثل علماء الآثار الذين يعملون في مجتمعات السكان الأصليين لتحديد مآل الثقافة المادية). وعلى نحو مماثل، يستخدم علماء الحاسوب تقنيات تتبع سلاسل البيانات (data lineage) لتتبع تطور مجموعة البيانات، وكثيراً ما يستخدمونها لتعقب الأخطاء في البيانات.

ويؤثر الوعي بسياق البيانات الأساس لتوضيح متى تكون بياناتك وتحليلاتك ناجعة ومتى لا تكون كذلك. ومع أن ثمة إغراء بتفسير النتائج المستندة إلى البيانات الضخمة بوصفها نتيجة واضحة، فإن من الخطوات الأساسية في البحث العلمي التعبير الواضح عما تمثله البيانات أو المؤشر وعما لا تمثله. هل تبقى نتائجك قاطعة بالقدر نفسه إذا تحول تفسيرك لمنشور على وسائل التواصل الاجتماعي من تسجيل لواقعة إلى أداء لهوية اجتماعية؟ ونظراً إلى الطبيعة الفوضوية، شبه العضوية، لكثير من مجموعات البيانات المستمدة من الأفعال الاجتماعية، فمن الجوهري أن يكون الباحثون متنبهين إلى المعاني المتعددة المحتملة للبيانات.

فعلى سبيل المثال، هل الأفضل أن يفسر منشور على فيسبوك أو صورة على إنستغرام بوصفه تأكيداً لظاهرة ما أو رفضاً لها، أم ملاحظة بسيطة، أم محاولة لتحسين المكانة داخل شبكة من الأصدقاء؟ ومع أن أيّاً من هذه التفسيرات قد يكون صحيحاً، فإن غياب السياق يزيد من صعوبة تسويغ اختيار فهم بعينه دون آخر. ويعزز التأمل في المعاني المتعددة المحتملة للبيانات وضوح فرضيات البحث، كما يجعل الباحثين واعين بالاستخدامات المحتملة الأخرى لبياناتهم. ومرة أخرى، فإن فعل التفسير عملية بشرية، ولأن أحكام من يستخدمون بياناتك (أو يعيدون استخدامها) قد تختلف عن أحكامك، فمن الضروري توضيح مواطن قوة البيانات وأوجه قصورها معاً.

وثق مصدر بياناتك وتطورها. لا تبالغ في تقدير وضوحها؛ وأقر بفوضويتها وتعدد معانيها.

6. ناقش الخيارات الأخلاقية الصعبة:

تخضع البحوث التي تشمل مشاركين من البشر في المؤسسات الممولة اتحادياً لمجالس المراجعة المؤسسية المكلفة بمنع الضرر عبر إجراءات راسخة، وهي مألوفة لكثير من الباحثين. غير أن مجالس المراجعة المؤسسية ليست الحكم الوحيد في الأخلاق؛ فكثير من المسائل الأخلاقية المتعلقة بالبيانات الضخمة يقع خارج نطاق ولايتها. وتحديداً، لأن باحثي البيانات الضخمة كثيراً ما يواجهون مواقف غريبة على مجالس المراجعة المؤسسية أو خارجة عن ولايتها، فإننا نؤكد أهمية مناقشة هذه المسائل ضمن مجموعات من الأقران.

وبدلاً من النظر إلى غياب الحلول القاطعة وبروتوكولات الحوكمة بوصفه خللاً، فالأنسب أن يفهم بوصفه ميزة ينبغي للباحثين تبنيها في عملهم، فالنقاش والجدل في المسائل الأخلاقية جزء أساسي من التطوير المهني، داخل التخصصات وفيما بينها، لأنه قادر على تكوين مجتمع ناضج من الممارسين المسؤولين. كما أن إدخال هذه النقاشات في المقررات الدراسية والتدريب يمكن أن يؤهل محكمين مؤهلين على نحو خاص لطرح هذه الأسئلة الأخلاقية، وللحث على إدراك الحاجة إلى هذه الحوارات.

ومن الشروط المسبقة لأي قواعد أو لوائح أخلاقية رسمية القدرة على إجراء مثل هذه النقاشات المفتوحة. وكما تكتب عالمة الاجتماع الرقمي والباحثة في الأخلاق أنيت ماركهام (Annette Markham) [30]: «يمكننا أن نجعل [أخلاقيات البيانات] موضوعاً أيسر طرّاً إذا تناولنا الأخلاق بوصفها تتعلق بخيارات نتخذها عند منعطفات حاسمة؛ خيارات سيكون لها أثر لا محالة». ونظرًا إلى طبيعة البيانات الضخمة، فإن جمع الباحثين في المجالات التقنية والعلمية والاجتماعية والإنسانية معًا في المشروعات يتيح لهذا النقاش أن يبرز بوصفه مصدر قوة، لأنه، إن أحسن إجراؤه، يوفر وسيلة لفهم المسائل الأخلاقية من منظورات متعددة ولكسر العزلة بين التخصصات [31]. وثمة عدد من النماذج الجيدة لبحوث الأخلاق متعددة التخصصات، مثل الدورات التدريبية التي يقدمها مركز بحوث العلم والعدالة (Science and Justice research center) في جامعة كاليفورنيا، سانتا كروز [32]، ومناهج «القيم في التصميم» (Values in Design) [33]. كما يمكن لخدمات الاستشارات في أخلاقيات البحث، المتاحة في بعض الجامعات بفضل برنامج جائزة العلوم السريرية والانتقالية (Clinical and Translational Science Award - CTSA) التابع للمعاهد الوطنية للصحة (NIH)، أن تكون موارد للباحثين [34].

وتوقّر بعض الحالات الأخلاقية الأشهر في مجال «البيانات الضخمة»، مثل دراسة فيسبوك عن العدوى العاطفية [35]، ساحات مثمرة للغاية للنقاشات بين التخصصات. لماذا قد ترى مجموعة من الباحثين في هذه الدراسة مقارنة غير ضارة نسبيًا، في حين ترى مجموعات أخرى فيها أوجه قصور أخلاقية كبيرة؟ وأين يختلف الباحثون في رسم الحد الفاصل بين البحث المسؤول وغير المسؤول، ولماذا؟ إن فهم الطرق المختلفة التي يناقش بها الناس هذه التحديات والعمليات يوفر ضابطاً مهماً للباحثين، ولا سيما إذا كانوا من تخصصات لا تركز على الاهتمامات المتعلقة بالأشخاص موضوع البحث.

علاوة على ذلك، فإن الظهور الواسع الذي يحيط بهذه الأحداث يعني أنها تمثل (للأفضل أو للأسوأ) الرؤية «العامة» لبحوث البيانات الضخمة، وأن الانخراط عضوًا فاعلاً في هذا الحوار يضمن للباحثين إسماع رؤاهم بدلاً من أن يكونوا مجرد متلقين لقرارات السياسات. وسعيًا إلى دفع هذه النقاشات قُدماً، أصدر مجلس البيانات الضخمة والأخلاق والمجتمع عددًا من دراسات الحالة التي تركز تحديدًا على بحوث البيانات الضخمة، وورقة بيضاء تتضمن توصيات لبدء هذه الحوارات المهمة.

أشرك زملاءك وطلابك في الحوار حول الممارسة الأخلاقية في بحوث البيانات الضخمة.

7. ضع مدونة سلوك لمؤسستك أو مجتمعك البحثي أو قطاعك الصناعي:

إن عملية مناقشة الخيارات الصعبة تُدخل الأخلاق مباشرة في سير العمل البحثي، فتجعل «تزييف الأخلاق» أمرًا غير مقبول بالقدر نفسه الذي يُرفض به تزييف البيانات أو النتائج. واستيعاب هذه النقاشات في صلب العمل، بدلاً من التعامل معها بوصفها أمرًا ثانويًا يأتي لاحقًا أو مشكلة يُعهد بها إلى جهة خارجية، أمر أساسي لنجاح البحث، ولا سيما عند استخدام بيانات الآثار الرقمية (trace data) التي ينتجها الناس. وهذا وثيق الصلة بكل البحوث، بما فيها

بحوث العاملين في قطاع الصناعة ممن لهم وصول متميز إلى تدفقات بيانات الحياة اليومية الرقمية. ولا ينبغي تجنب الاهتمام العام بالاستخدام الأخلاقي لهذه البيانات؛ فهذه المجموعات من البيانات قائمة في نهاية المطاف على بنية تحتية يستخدمها مليارات الناس في عيش حياتهم، وثمة مصلحة عامة قوية في أن يُجرى البحث بمسؤولية.

ومن أفضل السبل لترسيخ ذلك في الممارسة اليومية وضع مدونات سلوك تُستخدم في مؤسستك أو مجتمعك البحثي، وتُدرج في التعليم الرسمي والتدريب المستمر. ويمكن لهذه المدونات أن توفر توجيهًا في التحكيم العلمي للمنشورات وفي النظر في طلبات التمويل. وعمليًا، فإن حالة بارزة لبحث غير أخلاقي تجلب المشكلات لمجال بأكمله، لا للمعنيين بها مباشرة فحسب، علاوة على ذلك، فإن تصميم مدونات السلوك يجعل الباحثين أكثر نجاحًا. فالمسائل التي قد تُتجاهل إلى أن تنفجر، مثل: هل نلتزم بشروط الخدمة أو بتوقعات المستخدمين؟ وهل يرى عامة الناس بحثنا «مريبًا»؟ [13]، يمكن معالجتها بترؤ بدلًا من التدافع لاحتواء الضرر. ويكتسي ذلك أهمية خاصة للشركات الخاصة التي تتعامل مع الجمهور والحريصة على تجنب الاهتمام غير المرغوب فيه.

ومن المزايا الإضافية والأبعد مدى لوضع مدونات السلوك أن من الواضح أن التغيير قادم إلى بحوث البيانات الضخمة. فقد مؤلت المؤسسة الوطنية للعلوم مجلس البيانات الضخمة والأخلاق والمجتمع وسيلةً لاستباق قضية آخذة في التطور، وللاستباق تغييرات تنظيمية مرتقبة في القواعد الاتحادية لحماية الأشخاص موضوع البحث، وهي قيد المراجعة حاليًا [1]. والعمل الفاعل على وضع قواعد لبحوث البيانات الضخمة المسؤولة داخل مجتمع بحثي ما من السبل الأساسية التي يستطيع بها الباحثون الانضمام إلى هذه العملية الجارية.

ضع مدونات ملائمة للسلوك الأخلاقي داخل مجتمعك، واجعل الباحثين في قطاع الصناعة وممثلي المجتمعات المتأثرة مساهمين فاعلين في هذه العملية.

8. صمّم بياناتك وأنظمتك لتكون قابلة للتدقيق:

مع أن مدونات السلوك ستختلف حسب الموضوع والمجتمع البحثي، فإن من عناصرها البالغة الأهمية تصميم البيانات والأنظمة لتكون قابلة للتدقيق (auditability). فعمليات التدقيق الداخلي المسؤولة يمكن دمجها بسهولة في أنظمة التدقيق، كما أنها ترصد العوامل التي قد تسهم في نتائج إشكالية. ويمكن لتطوير عمليات اختبار آلية لتقييم النتائج الإشكالية، وآليات لتدقيق عمل الآخرين في أثناء عمليات المراجعة، أن يساعد في تعزيز البحث ككل. والهدف من قابلية التدقيق هو التوثيق الواضح لتوقيت اتخاذ القرارات، والعودة عند الضرورة إلى مجموعة بيانات سابقة ومعالجة المشكلة من جذورها (مثلًا إذا اختُرقت استراتيجيات إخفاء هوية البيانات).

ويعود التصميم من أجل قابلية التدقيق أيضًا بفوائد مباشرة على الباحثين، إذ يوفر آلية للتحقق المزدوج من العمل، ويُلزم المرء بالإفصاح عن قراراته، فيزيد من قابلية الفهم وقابلية التكرار. فعلى سبيل المثال، كثير من أنواع بيانات وسائل التواصل الاجتماعي وغيرها من بيانات الآثار الرقمية غير مهيكلة، والإجابات حتى عن الأسئلة الأساسية، مثل الروابط الشبكية والموقع والعشوائية، تتوقف على الخطوات المتبعة في جمع البيانات وترتيبها. وتوضّح أنظمة قابلية التدقيق أوجه اختلاف مجموعات البيانات المختلفة (والتحليلات اللاحقة لها) بعضها عن بعض، مما يساعد على الفهم ويُنْتِج بحثًا أفضل.

خطّط لتدقيق ممارساتك في البيانات الضخمة ورخّب به.

9. انظر في العواقب الأوسع لممارسات البيانات والتحليل:

من المهم أيضًا لباحثي البيانات الضخمة المسؤولين أن يفكروا فيما يتجاوز المقاييس التقليدية للنجاح في عالم الأعمال والأوساط الأكاديمية. فعلى سبيل المثال، فإن متطلبات الطاقة للحياة اليومية الرقمية، وهي مصدر رئيسي للبيانات الضخمة في بحوث العلوم الاجتماعية، كبيرة في عصر تغيّر المناخ هذا [36]. فكيف يمكن لبحوث البيانات الضخمة أن تقلّل الأثر البيئي لأعمال تحليلات البيانات؟ وعلى سبيل المثال، هل ينبغي للباحثين أن يأخذوا زمام المبادرة في مطالبة مزودي التخزين السحابي ومراكز معالجة البيانات بالتحوّل إلى مصادر طاقة مستدامة ومتجددة؟ إن باحثي البيانات الضخمة، بوصفهم مستخدمين مهمين للسحابة وذوي حضور علني، يمثلون مجتمعين مجموعة مصالح يمكنها أن تحتشد خلف دعوة كهذه إلى التغيير.

إن السعي وراء الاستشهادات أو السمعة أو المال حافز رئيسي لدفع البحث قُدّمًا، لكنه قد يفضي أيضًا إلى نتائج غير مقصودة وغير مرغوب فيها. وفي المقابل، يمكننا أن نسأل: إلى أي مدى يركّز مشروع بحثي ما على تعزيز الصالح العام أو على الفئات المحرومة من الخدمات في المجتمع؟ وهل تُعالج في تدفقات بياناتك أسئلة الإنصاف أو تعزيز القيم العامة الأخرى، أم أن التركيز على البيانات الضخمة يجعلها غير مرئية أو غير ذات صلة بتحليلك [37]؟ وكيف يمكن حماية الموارد العامة الأساسية التي تزداد هشاشة، مثل سجلات السرطان التي تفرضها الولايات؟ وكيف قد يساعد البحث مختلف الفاعلين في عالم الأعمال والسياسة أو يعيقهم؟ ومع أنه ليس من الضروري أن تتناول كل بحوث البيانات الضخمة مسائل اجتماعية وثقافية، فإن من الأهداف الجوهرية للبحث تجاوز فهم العالم إلى النظر في سبل تحسينه.

أدرك أن لإجراء بحوث البيانات الضخمة آثارًا على مستوى المجتمع بأسره.

10. اعرف متى تخرج على هذه القواعد:

القاعدة الأخيرة (المنافية للحدس) هي الدعوة إلى إدراك متى يكون من الملائم الحياد عن هذه القواعد. فعلى سبيل المثال، في أوقات الكوارث الطبيعية أو حالات الطوارئ الصحية العامة، قد يكون من المهم تنحية مسائل الخصوصية الفردية جانبًا مؤقتًا لخدمة صالح عام أكبر. وبالمثل، قد يكون استخدام البيانات الجينية أو غيرها من البيانات البيولوجية المجموعة دون موافقة مستتيرة أمرًا حيويًا في إدارة وباء ناشئ.

علاوة على ذلك، احرص على مراجعة التوقعات التنظيمية والمتطلبات القانونية المرتبطة بحماية الخصوصية في مجموعة بياناتك. فهذا مسار قد يفضي بسهولة إلى تجاوزات، لذا، قبل اتباع هذه القاعدة (للخروج على غيرها)، كن حذرًا من أن تكون «حالة الطوارئ» مجرد ذريعة ملائمة. وأفضل سبيل لضمان ذلك هو بناء الخبرة في الانخراط في النقاشات الصعبة (القاعدة 6)، ووضع مدونات السلوك (القاعدة 7)، وتطوير أنظمة التدقيق (القاعدة 8). وكلما ازداد نضج مجتمع الباحثين فيما يتعلق بإجراءاته وآليات الرقابة والتوازن داخله، ازداد استعداده لتقدير متى يكون الخروج على القواعد مقبولًا. وقد لا تصل في النهاية إلى مجموعة واضحة ونهائية من الممارسات. فكما أن الخصوصية ليست ثنائية (القاعدة 2)، فكذلك البحث المسؤول. وكثيرًا ما تتعلق الأخلاق بإيجاد إجابة جيدة أو أفضل، لا إجابة مثالية، ومن المهم طرح الأسئلة الصعبة (ومحاولة الإجابة عنها). فلا يمكن أن تنشأ ثقافة لبحوث البيانات الضخمة المسؤولة إلا عبر هذا الانخراط.

افهم أن بحوث البيانات الضخمة المسؤولة تتوقف على ما هو أكثر من استيفاء قوائم التحقق.

الخاتمة:

الهدف من هذه المجموعة من القواعد العشر هو مساعدة الباحثين على إنجاز عمل أفضل، وعلى أن يصبحوا في نهاية المطاف أكثر نجاحًا، مع تجنب تعقيدات أكبر، بما فيها انعدام ثقة الجمهور. ولتحقيق ذلك، يتعين على الباحثين التخلي عن عقلية صارمة حين تتعلق المسألة بالتقنيات والمنهجية، وساذجة حين يتعلق الأمر بالأخلاق. فالعبارات من قبيل «البيانات عامة بالفعل» («Data is [sic] already public») [38] تبسيطات غير مسوّغة لمنظومات بيانات أكثر تعقيدًا بكثير، مندمجة في ممارسات اجتماعية أشد تعقيدًا وأكثر ارتهاً بظروفها. فالبيانات تمثل أناسًا، والتمسك بتعريف ساذج صارم يقول بخلاف ذلك [18] سيُلحق الضرر بالجهود البحثية على المدى البعيد، حين تأتي المقاومة من الناس الذين تخضع أفعالهم وأقوالهم للتحليل.

وباختصار، فإن بحوث البيانات الضخمة المسؤولة لا تعني منع البحث، بل التأكد من أن العمل سليم ودقيق، وأنه يعظم النفع ويقلل الضرر إلى أدنى حد. فالمشكلات والخيارات التي يواجهها الباحثون حقيقية ومعقدة وصعبة، ويجب أن تكون استجابتنا كذلك. وعلينا أن نتعامل مع بحوث البيانات الضخمة بالاحترام الذي تستحقه، وأن ندرك أن البحث غير الأخلاقي يقوّض إنتاج المعرفة. فثمة فرص رائعة لفهم المجتمع وعالمنا فهمًا أفضل، لكن هذه الفرص تأتي معها مسؤولية النظر في أخلاقيات خياراتنا في الممارسات والأفعال اليومية لبحثنا. ويقدم مجلس البيانات الضخمة والأخلاق والمجتمع مجموعة أولية من دراسات الحالة والأوراق، بل حتى عشر قواعد بسيطة لتوجيه هذه العملية؛ والآن يقع على عاتقك أن تستخدمها وتحسنها في بحثك.

المراجع:

1. Metcalf J, boyd d, Keller E. Perspectives on Big Data, Ethics, and Society. Council for Big Data, Ethics, and Society. 2016. <http://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/>. Accessed 31 May 2016.
2. Wu HY, Rubinstein M, Shih E, Guttag JV, Durand F, Freeman WT. Eulerian video magnification for revealing subtle changes in the world. Eulerian Video Magnification for Revealing Subtle Changes in the World. ACM Transactions on Graphics. 2012;31(4).
3. Crawford K, Schultz J. Big data and due process: Toward a framework to redress predictive privacy harms. BCL Rev. 2014;55: 93-128.
4. Barocas S, Selbst AD. Big data's disparate impact. California Law Review. 2016;104(3): 671-732.
5. Danylo WA, Alisson VB, Alexandre ND, Moacir LM, Jansepetrus BP, Oliveira RF. Identifying relevant users and groups in the context of credit analysis based on data from Twitter. InCloud and Green Computing (CGC), 2013 Third International Conference on 2013 Sep 30 (pp. 587-592). IEEE.
6. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias. Pro Publica. 23 May 2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Accessed 4 September 2016.

7. Ingold D, Spencer S. Amazon Doesn't Consider the Race of Its Customers. Should It? Bloomberg.com 21 April 2016. <http://www.bloomberg.com/graphics/2016-amazon-same-day/>. Accessed 12 June 2016.
8. Hauge MV, Stevenson MD, Rossmo DK, Le Comber SC. Tagging Banksy: using geographic profiling to investigate a modern art mystery. *Journal of Spatial Science*. 2016;61(1):185-90.
9. Metcalf J, Crawford K. Where are Human Subjects in Big Data Research? The Emerging Ethics Divide. *The Emerging Ethics Divide. Big Data and Society*, 2016.
10. Zwitter A. Big data ethics. *Big Data & Society*. 2014;1(2).
11. Nissenbaum H. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press; 2009.
12. Marwick AE, boyd d. Networked privacy: How teenagers negotiate context in social media. *New Media & Society*. 2014;1461444814543995.
13. Tene O, Polonetsky J. Theory of Creepy: Technology, Privacy and Shifting Social Norms, A. *Yale JL & Tech*. 2013;16:59.
14. Massey DS, Denton NA. *American apartheid: Segregation and the making of the underclass*. Harvard University Press; 1993.
15. Davidow B. Redlining for the 21st Century. *The Atlantic*. 5 March 2014. <http://www.theatlantic.com/business/archive/2014/03/redlining-for-the-21st-century/284235/>. Accessed 31 May 2016.
16. Young JC, Gilmore MP. Subaltern empowerment in the Geoweb: Tensions between publicity and privacy. *Antipode*. 2014;46(2):574-91.
17. Barbaro M, Zeller T, Hansell S. A face is exposed for AOL searcher no. 4417749. *New York Times*. 2006 Aug 9;9.
18. Cox J. 70,000 OkCupid Users Just Had Their Data Published. *Motherboard*. 12 May 2016. <http://motherboard.vice.com/read/70000-okcupid-users-just-had-their-data-published>. Accessed 12 June 2016.
19. Pandurangan V. On Taxis and Rainbows: Lessons from NYC's improperly anonymized taxi logs. *Medium*. 2014. <https://medium.com/@vijayp/of-taxis-and-rainbows-f6bc289679a1>. Accessed 10 November 2015.
20. Sweeney L. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 2002;10(05):557-70.

21. Zimmer M. "But the data is already public": on the ethics of research in Facebook. *Ethics and information technology*. 2010;12(4):313-25.
22. Kloumann IM, Kleinberg JM. Community membership identification from small seed sets. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* 2014 Aug 24 (pp. 1366-1375). ACM.
23. Narayanan A, Huey J, Felten EW. A precautionary approach to big data privacy. In *Data protection on the move* 2016 (pp. 357-385). Springer Netherlands.
24. Michalevsky Y, Schulman A, Veerapandian GA, Boneh D, Nakibly G. Powerspy: Location tracking using mobile device power analysis. In *24th USENIX Security Symposium (USENIX Security 15)* 2015 (pp. 785-800).
25. Shelton T, Poorthuis A, Zook M. Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*. 2015;142:198-211.
26. Acquisti A, Gross R, Stutzman F. Face recognition and privacy in the age of augmented reality. *Journal of Privacy and Confidentiality*. 2014;6(2): 1-20.
27. Homer N, Merriman B, Nelson SF. BFAST: an alignment tool for large scale genome resequencing. *PLoS ONE*. 2009;4(11):e7767. pmid:19907642
28. Lo B. Sharing clinical trial data: maximizing benefits, minimizing risk. *Jama*. 2015;313(8): 793-4. pmid:25594500
29. Goodman A, Pepe A, Blocker AW, Borgman CL, Cranmer K, Crosas M, et al. Ten simple rules for the care and feeding of scientific data. *PLoS Comput Biol*.;10(4).
30. Markham A. OKCupid data release fiasco: It's time to rethink ethics education. Medium. Points 18 May 2016. <https://points.datasociety.net/okcupid-data-release-fiasco-ba0388348cd#.g4ofbpnc6>. Accessed 12 June 2016.
31. Ford H. Big Data and Small: Collaborations between ethnographers and data scientists. *Big Data & Society*. 2014;1(2):
32. Science & Justice Research Center (Collaborations Group. Experiments in collaboration: interdisciplinary graduate education in science and justice. *PLoS Biol*. 2013 Jul 30;11(7):e1001619.
33. Knobel C, Bowker GC. Values in design. *Communications of the ACM*. 2011;54(7):26-8.
34. Cho MK, Tobin SL, Greely HT, McCormick J, Boyce A, Magnus D. Research ethics consultation: The Stanford experience. *IRB*. 2008;(6):1-6. pmid:19119757

35. boyd d. Untangling research and practice: What Facebook's "emotional contagion" study teaches us. Research Ethics. 2016;12(1):4-13.
36. Cook G, Dowdall T, Pomerantz D, Wang Y. Clicking clean: how companies are creating the green internet. Greenpeace Inc., Washington, DC. 2014. <http://www.greenpeace.org/usa/wp-content/uploads/legacy/Global/usa/planet3/PDFs/clickingclean.pdf>
37. Zook MA, Graham M. Mapping DigiPlace: geocoded Internet data and the representation of place. Environment and Planning B: Planning and Design. 2007 Jun 1;34(3):466-82.
38. Zimmer M. OkCupid Study Reveals the Perils of Big-Data Science. Wired. 14 May 2016. <https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science/>. Accessed 12 June 2016.