



مقالة رأي وتحليل

التاريخ التحريري: ٢٧ أكتوبر ٢٠٢٦

الكلمات التي لا توجد في الإنترنت

من فنلندا إلى السعودية: لماذا قد يبدأ أعظم مشروع وطني للذكاء الاصطناعي من صوت جدة، ودقتر طالب، وذاكرة مجتمع كامل؟

في صيف عام 2020، طلبت فنلندا من مواطنيها طلباً غير مألوف: أن يتحدثوا.

لم تطلب منهم أن يقرأوا نصوصاً رسمية، ولا أن يرتلوا عبارات أعدت بعناية في مختبر، ولا أن يتحدثوا بالفنلندية المثالية التي تُدرس في المدارس.

طلبت منهم أن يتكلموا كما يتكلمون في حياتهم.

عن الرياضة والطعام والذكريات، وعن الأشياء التي يحبونها أو يختلفون حولها. بلهجاتهم الطبيعية، وتردداتهم، وأخطائهم العابرة، وما يتخلل الحديث من صمت وضحك وتعبيرات لا يعرفها القاموس.

حملت المبادرة اسم «تبرع بكلامك» (Lahjoita puhetta – Donate Speech)، وأطلقتها هيئة الإذاعة الفنلندية (YLE) بالتعاون مع بنك اللغة الفنلندي (Kielipankki) وشركاء بحثيين وتقنيين.

كان الهدف جمع عشرة آلاف ساعة من الكلام الفنلندي الطبيعي لتطوير تقنيات تستطيع فهم الناس كما يتحدثون فعلاً.

وحين أغلقت الحملة في 6 مارس 2024، أعلن بنك اللغة الفنلندي أنها جمعت أكثر من ثلاثة آلاف ساعة من الكلام. أما الدراسة البحثية المنشورة عام 2022 فتصف إصدار 2021 من المجموعة بأنه ضم نحو 3600 ساعة وأكثر من عشرين ألف متحدث، وفق تقدير قائم على اعتبار كل هوية لعميل التطبيق متحدثاً واحداً، لا على التحقق من هويات الأشخاص.

لم تبلغ المبادرة هدفها الكمي الأصلي، لكنها صنعت شيئاً أكثر أهمية من تسجيلات صوتية.

لقد حولت المواطن العادي إلى مشارك في بناء البنية المعرفية التي ستحتاج إليها تقنيات المستقبل.

وتوثق دراسة *Lahjoita puhetta: A Large-Scale Corpus of Spoken Finnish* كيف أصبح الكلام اليومي مورداً للبحث في التعرف الآلي على الكلام (Automatic Speech Recognition – ASR)، وفهم اللهجات، وتحليل الخصائص اللغوية والاجتماعية.

والأجمل أن الفنلنديين فهموا شيئاً بالغ البساطة: مهما بلغت براعة المهندس، فإنه لا يستطيع أن يجلس في مختبره ويخترع الطريقة التي يتحدث بها ملايين البشر.

يمكنه أن يصمم النظام الذي يتعلم منها، لكنه يحتاج إلى الناس كي يمنحوه مادة التعلم. ومن هذه الحكاية الفنلندية أصل إلى سؤال أراه جديراً بمبادرة وطنية سعودية غير مسبقة:

ماذا لو أصبح كل بيت سعودي، وكل مدرسة وجامعة، وكل مهنة ومؤسسة، مساهماً في إنشاء أكبر بنية بيانات بشرية أصيلة تمثل لغتنا وثقافتنا ومعارفنا؟

ليس بجمع ما كتبه الآخرون عنا من الإنترنت، ولا بتوليد ملايين النصوص التي تشبه ما تنتجه النماذج الحالية. بل بأن نبدأ من الإنسان نفسه.

من الكلمات التي يقولها، والمعاني التي يفهمها، والوثائق التي يعرف كيف يقرأها، والتفاصيل التي لا تتعلمها الآلة لأنها لم تُكتب أصلاً في مصدر تستطيع الوصول إليه.

نصف قرن من الإنترنت.. وما زالت الحياة أكبر منه

تخيل أنك طلبت من نموذج لغوي أن يشرح معنى عبارة «يا بعد حيي».

قد يقدم تفسيراً مقبولاً من الناحية المعجمية، وربما يشرح أنها تعبير عن المودة والتقدير.

لكن معرفة العبارة شيء، وإدراك موقعها داخل العلاقة الإنسانية شيء آخر.

قد يقولها رجل مسن لحفيده، أو امرأة لأختها، أو صديق لصديقه في موقف امتنان. وقد تتغير نبرة التعبير ودلالته بحسب السياق والمكان والعلاقة بين المتحدثين.

ثم تخيل مهندساً يعمل في محطة صناعية، يستخدم مصطلحاً دارجاً بين زملائه لوصف عطل متكرر لا يظهر بهذا الاسم في دليل الشركة.

أو امرأة من البادية تستطيع تسمية تغيرات دقيقة في الرياح واتجاهات السحب، بينما لا تقدم موسوعة عامة إلا وصفاً تقريبياً لها.

أو ممرضة تعرف التعبيرات التي يستخدمها المرضى لوصف الألم، وكيف يختلف الوصف الشعبي عما يكتبه الطبيب في السجل السريري.

هذه المعرفة ليست غائبة لأنها عديمة القيمة.

إنها غائبة، في كثير من الحالات، لأنها لم تدخل النظام الرقمي الذي نتعلم منه الآلات.

وهذا يفتح مسألة أعمق من نقص النصوص العربية.

في علم الإحصاء توجد مشكلة تُعرف بتحيز العينة (Sampling Bias). فإذا جمعت بياناتك من شريحة لا تمثل المجتمع، فقد يصبح النموذج ممتازاً في فهم تلك الشريحة، لكنه أقل كفاءة في فهم من لم تشملهم العينة.

ومع النماذج اللغوية، تصبح المشكلة مركبة.

فالإنترنت ليس نسخة متوازنة من المجتمع.

بعض الناس يكتبون أكثر من غيرهم، وبعض المهن تنتج نصوصاً رقمية وفيرة، وبعض الفئات لا تترك إلا آثاراً محدودة، وبعض اللهجات لا تظهر إلا في محادثات خاصة لا يجوز جمعها دون إذن.

وأحياناً يكون الأشخاص الذين يمتلكون أثنى المعارف المحلية هم الأقل حضوراً في البيانات الرقمية.

لهذا يمكن أن يصبح النموذج واسع المعرفة بالعالم، لكنه محدود المعرفة بتجربة بشرية قريبة جداً من مستخدميه.

إنه يعرف ما نُشر، لا بالضرورة ما عُرف.

وهنا تكمن الفرصة: ليست السعودية بحاجة إلى جمع المزيد مما تعرفه الآلات بالفعل، بقدر حاجتها إلى تمكين الناس من تعليمها ما لم تستطع الوصول إليه من قبل.

حين أصبحت الشعوب شريكة في تدريب الآلات

لم تكن فنلندا استثناءً منفرداً.

في الهند، التي تضم عدداً هائلاً من اللغات واللهجات، أطلقت الحكومة مبادرة بهاشيني (BHASHINI)، وضمنها برنامج التبرع اللغوي (Bhasha Daan)، الذي يتيح للمواطنين المساهمة بالنصوص والتسجيلات الصوتية والترجمات.

وفي مارس 2026، أعلنت وزارة الإلكترونيات وتكنولوجيا المعلومات الهندية أن أكثر من عشرة آلاف مساهم انضموا إلى منصة (Bhashini Samudaye)، التي تجمع الأفراد والجامعات والشركات والمختبرات لدعم إنشاء الموارد اللغوية ووسم البيانات (Data Annotation).

الفكرة ليست أن الحكومة تجمع أصوات المواطنين لتضعها في أرشيف وحسب.

بل أنها تبني مورداً تشارك فيه الجامعات والباحثون والمؤسسات والشركات، وتستطيع تقنيات مختلفة الاستفادة منه.

وقد يكون هذا هو الفارق بين حملة قصيرة لجمع البيانات وبين تأسيس بنية وطنية دائمة.

وفي كوريا الجنوبية، اتخذ المشروع شكلاً مؤسسياً أوسع.

فقد أنشأت الوكالة الوطنية لمجتمع المعلومات (National Information Society Agency – NIA) منصة (AI-Hub)، التي أتاحت، منذ عام 2017، أكثر من تسعمئة مجموعة بيانات لتدريب الذكاء الاصطناعي، بحسب الوكالة.

وتشمل البيانات اللغة الكورية والصوت والصور والطب والصناعة والزراعة وغيرها.

لكن التفصيل الجدير بالانتباه أن كوريا لم تنظر إلى وسم البيانات بوصفه نشاطاً تقنياً مغلقاً.

وفي إطار البرامج الحكومية التي بنت بيانات التدريب المتاحة عبر المنصة، أنشئت وظائف مباشرة وأعمال مدفوعة عبر التجهيد الجماعي في جمع البيانات ومعالجتها والتحقق من جودتها، مع توثيق الوكالة لإمكان انتقال بعض العاملين، مع الخبرة، من الوسم إلى المراجعة وإدارة المشروعات.

وهنا تظهر رؤية اقتصادية مهمة: يمكن أن تتحول عملية تعليم الآلة إلى مجال لتأهيل البشر وإكسابهم مهارات مهنية جديدة.

وفي سنغافورة، جاء اتجاه آخر لا يقل إثارة.

أطلقت مؤسسة الذكاء الاصطناعي السنغافورية (AI Singapore)، بالتعاون مع Google، منصة (Aquarium) ضمن منظومة (SEA-LION)، لتسهيل مساهمة المجتمعات والباحثين في مجموعات بيانات لغات جنوب شرق آسيا.

لم يصمم المشروع لتجميع اللغة الملايوية أو التايلندية أو الإندونيسية فقط، بل لتسهيل الوصول إلى موارد تعكس واقع المنطقة اللغوي والثقافي، وتطوير أدوات للمشاركة والتقييم والتحفيز.

ولعل الاسم، «حوض الأسماك»، يحمل استعارة جميلة: ليست اللغة سمكة كبيرة واحدة نضعها في حوض، بل نظاماً حياً من التنوع والعلاقات.

أما إستونيا، فقد اختارت مدخلاً تعليمياً مختلفاً.

أطلقت مبادرة (AI Leap) لإدخال الذكاء الاصطناعي في التعليم العام بطريقة تساعد المعلمين والطلاب على اكتساب مهارات التفكير والتعلم النقدي.

وخلال العام الدراسي 2025-2026، شاركت 154 مدرسة ثانوية من أصل 156 مدرسة في البرنامج، الذي شمل نحو خمسة آلاف معلم وعشرين ألف طالب.

ولا تُعد هذه المبادرة مشروعاً لجمع بيانات لتدريب وطنية بالمعنى الذي أتحدث عنه هنا.

لكنها ثبت أن الدولة تستطيع تحويل المدارس إلى شبكة وطنية واسعة للتعليم والمشاركة في تطوير الممارسات التقنية الجديدة.

وحين نصل إلى الولايات المتحدة، نجد شاهداً رائعاً في مكان غير متوقع: مكتبة الكونغرس (Library of Congress).

أطلقت المكتبة مشروع *By the People*، الذي يدعو المتطوعين إلى تفريغ المخطوطات والوثائق التاريخية المكتوبة بخط اليد، ومراجعة تفريغات الآخرين وتصنيف المواد.

قد يكون المتطوع طالباً، أو متقاعداً، أو مؤرخاً هاوياً.

ومع كل سطر يقرأه، تتحول وثيقة تاريخية يصعب على الحاسوب فهمها إلى نص يمكن البحث فيه ودراسته والاستفادة منه.

وبالنسبة إلى الذكاء الاصطناعي، فإن هذا النوع من العمل يوضح قيمة تحويل الأرشيفات غير المهيكلة (Unstructured Archives) إلى معرفة رقمية قابلة للاستخدام.

نحس تجارب، ونحس طرق مختلفة، لكنها تلتقي في فكرة واحدة:

البيانات التي يحتاج إليها الذكاء الاصطناعي ليست دائماً منتجاً يمكن شراؤه جاهزاً؛ أحياناً تكون معرفة مبعثرة بين ملايين البشر، ويمكن الابتكار في بناء الوسيلة التي تجمعها وتحفظ حقوق أصحابها وتحقق من جودتها.

«ديوان»: مشروع وطني يكتبه الناس

أتخيل مبادرة سعودية تحمل اسماً مقترحاً: الديوان السعودي للمعرفة البشرية.

لا أتصورها تطبيقاً يطلب من الناس تحميله ثم نسيانه، ولا مسابقة موسمية تنتهي بتوزيع الجوائز.

بل بنية وطنية تشاركية دائمة (National Participatory Data Infrastructure)، يُقترح أن يشارك فيها التعليم العام والجامعي، والقطاع الخاص، والجمعيات غير الربحية، والمراكز البحثية، والمواطنون والمقيمون الراغبون في المساهمة.

تكون غايتها إنشاء أكبر مورد سعودي موثق للبيانات البشرية الأصيلة والموسومة، مع تنوع لغوي وثقافي ومهني واجتماعي مدروس.

وفي قلب الديوان بوابة رقمية تتيح المشاركة في مستويات مختلفة.

يدخل الطالب فيجد مهمة تناسب سنه ومعرفته: تفسير تعبير محلي، أو كتابة وصف لمشهد، أو قراءة جملة بصوته، أو مراجعة نص استُخرج آلياً من وثيقة مرخصة.

وتدخل المعلمة لتساهم في تصنيف نصوص تعليمية، وتحديد مستويات صعوبتها، وشرح مواضع اللبس فيها.

وتشارك طبيباً أو ممرضة في مراجعة المصطلحات والتعبيرات الصحية ضمن أمثلة أعدت خصيصاً للتدريب، أو بيانات تعليمية جرى تجهيلها وفق ضوابط اللائحة التنفيذية وبعد تقييم مخاطر إعادة التعرف، لا من سجلات مرضى مكشوفة.

ويشارك مهندس في توضيح مفردات مهنية، أو تصنيف صور معدات متاحة قانونياً، أو وصف خطوات إجراء صناعي عام.

وتدخل امرأة مهتمة بالتاريخ الاجتماعي فتسجل، بموافقتها، تفسيراً لمثل شعبي أو عادة محلية أو اسم أداة منزلية قديمة.

ويشارك رجل من البادية في وصف مفردات تتعلق بالمكان والطقس والحياة البرية، كما يشارك ساكن المدينة في التعبيرات التي نشأت في بيئته الحديثة.

لا أفضلية مسبقة للهجة على أخرى، ولا لأسلوب اجتماعي على آخر.

فالمقصود ليس أن نفرض على النموذج صورة واحدة للإنسان السعودي، بل أن نجعله يرى تنوع المجتمع الذي يتعلم منه.

ويمكن أن يشارك المقيمون أيضاً بلغاتهم وخبراتهم، خصوصاً في مجالات التفاعل اليومي والعمل والخدمات والترجمة بين اللغات.

وهذا يفتح الباب لتطوير نماذج تفهم السياقات الاجتماعية والمهنية في المملكة بصورة أوسع.

والميزة الكبرى أن المساهم لا يحتاج إلى معرفة البرمجة أو تعلم الآلة.

يكفي أن يعرف شيئاً حقيقياً يستطيع نقله للآخرين.

قد لا تعرف الجودة ما معنى (Transformer Architecture)، لكنها تعرف حكايات وتعايير لن يستطيع مهندس

المحولات اختراعها مهما اتسعت قدرته الحاسوبية.

وهنا تصبح المبادرة أكثر من مصدر بيانات.

تصبح اعترافاً اجتماعياً بأن المعرفة ليست حكراً على من يحمل شهادة علمية، وأن الخبرة اليومية نفسها قد تكون مادة ثمينة لتطوير التقنيات.

يوم يصبح تصحيح كلمة عملاً وطنياً

لنتخيل واجهة الديوان.

تظهر أمام المتطوع صورة لوثيقة عربية قديمة، أتيح استخدامها بإذن الجهة المالكة.

تحاول أداة التعرف الضوئي على الحروف (Optical Character Recognition – OCR) استخراج النص، لكنها تخطئ في قراءة كلمة بسبب طبيعة الخط أو تلف الورق.

يرى المتطوع الكلمة، فيصححها.

ثم تُعرض المراجعة على مساهم ثانٍ. وإذا وقع خلاف في القراءة، تُحال إلى مراجع أكثر خبرة.

في دقائق قليلة، يكون المشاركون قد أدوا عملاً يستطيع أن يسهم في إنشاء مجموعة بيانات لتطوير أنظمة قراءة الوثائق العربية.

وفي مكان آخر، تُعرض صورة مرخصة لقطعة من الحرف التقليدية أو أداة زراعية.

تطلب البوابة من المشاركين وصفها، وتحديد اسمها في منطقتهم، وذكر استعمالها، وربما إضافة اسم بديل في لهجة أخرى.

ثم تُربط الصورة بالأوصاف المناسبة، فتتحول إلى مادة تدريب للربط بين الرؤية واللغة (Vision-Language Learning).

وفي مهمة ثالثة، يستمع المتطوع إلى تسجيل صوتي تبرع به شخص بموافقته، ثم يراجع النص الذي أنتجه نظام التعرف الآلي على الكلام.

هذه ليست فكرة افتراضية من الناحية التقنية.

فمنصة [Google Crowdsourcing](#) تتيح للمستخدمين المساهمة في التحقق من وسوم الصور (Image Label Verification)، وقراءة الخطوط (Handwriting Recognition)، ومراجعة الترجمات، ووصف الصور، والتحقق من التسجيلات الصوتية، وغيرها من المهام.

وتمنح المنصة نقاطاً مختلفة بحسب المهمة وصعوبتها، مع نقاط إضافية لبعض المساهمات التي تؤكد مراجعات المشاركين الآخرين.

أما مشروع (Common Voice) التابع لموزيلا (Mozilla Foundation)، فيتيح للناس تسجيل أصواتهم، ثم يتيح لمتطوعين آخرين مراجعة التسجيلات، لتنشأ مجموعات صوتية متعددة اللغات يمكن الاستفادة منها في تطوير تقنيات الكلام.

الفكرة أثبتت قابليتها للتنفيذ.

لكن المبادرة السعودية تستطيع أن تذهب إلى ما هو أبعد من نسخها.

يمكنها أن تربط المساهمة بمجالات معرفية، وأن تنشئ مسارات للمستويات المتقدمة، وأن تطور شهادات خبرة في الوسم اللغوي والصوتي والبصري، وأن تجعل المشاركة بوابة إلى التخصص في علوم البيانات.

فالذي يبدأ بتصحيح كلمة في مخطوطة قد يصبح بعد سنوات باحثاً يطور نظاماً يستطيع قراءة آلاف المخطوطات.

هل تستطيع الآلة شم رائحة الآلة؟

هنا تظهر مشكلة لا مفر من التعامل معها.

إذا فتحنا باب المشاركة أمام مجتمع واسع، فقد يحاول بعض المشاركين استخدام نماذج مثل (GPT) أو (Claude) أو (DeepSeek) لتوليد نصوص جاهزة، ثم إرسالها بوصفها مساهمات بشرية أصلية.

وليس هذا الخطر هامشياً.

فإذا أصبح الهدف إنشاء مورد معرفي نابع من البشر، ثم امتلاء المورد بنصوص مولدة من نماذج تجارية، فقد نعيد إدخال المعرفة التي كانت موجودة أصلاً في النماذج السابقة، بدل توسيع تمثيل التجربة البشرية.

لهذا أقترح أن تضم بوابة المبادرة نظاماً بحثياً خاصاً لفحص منشأ المساهمات وجودتها.

يمكن أن يستفيد هذا النظام من تحليل نتاج الكلمات (N-gram Analysis)، والبصمات الإحصائية للأسلوب (Stylometric Fingerprints)، ودرجة المفاجأة اللغوية (Perplexity)، واكتشاف النصوص المكررة أو شديدة التشابه (Near-Duplicate Detection).

فمثلاً، يمكن للنظام أن يبحث عن تكرار غير معتاد لبعض التراكيب، أو عن سلاسل طويلة متطابقة مع مواد منشورة، أو عن أنماط تظهر بكثرة في المخرجات الآلية مقارنة بالمجموعات البشرية المرجعية.

لكن هنا تكمن مفارقة علمية لا ينبغي تجاهلها.

لا نستطيع أن نثبت أن النص بشري لمجرد أنه لا يشبه النصوص الاصطناعية، كما لا نستطيع اتهام إنسان باستخدام الآلة لمجرد أن أسلوبه منتظم.

فقد وجدت دراسة نشرها ويشين ليانغ (Weixin Liang) وزملاؤه من ستانفورد في مجلة (Patterns) عام 2023، ضمن العينات والكواشف السبعة التي اختبروها، أن كواشف النصوص المولدة قد تنحاز ضد كتابات غير الناطقين بالإنجليزية كلغة أولى، فتصنف نصوصاً بشرية على أنها مولدة.

كما كشفت دراسة بقيادة ديورا فير-وولف (Debora Weber-Wulff)، نُشرت في (International Journal for Educational Integrity)، عن محدودية موثوقية مجموعة من أدوات الكشف الشائعة، وارتفاع الأخطاء الإيجابية والسلبية في بعض الحالات.

لذلك فإن نظام فحص المحتوى المقترح لا ينبغي أن يكون قاضياً يمنع المشاركة بناءً على درجة احتمالية واحدة.

بل جزءاً من هندسة أوسع لإثبات منشأ البيانات (Data Provenance).

تبدأ هذه الهندسة من المساهمة نفسها: إقرار واضح بطريقة إنشائها، ومسار تحرير موثق اختياري، ومهام تُنجز مباشرة داخل البوابة، ومراجعات بشرية مستقلة، وعينات عشوائية للتحقق عند الحاجة، من دون مراقبة متطفلة للمتطوعين.

وتُفصل البيانات البشرية المؤلفة أصلاً (Human-Authored Data) عن النصوص البشرية التي راجعت مخرجات آلية (Human-Verified AI Outputs)، وعن البيانات الاصطناعية (Synthetic Data).

فالأنواع الثلاثة قد تكون نافعة، لكنها ليست الشيء نفسه، ولا ينبغي أن تضيع الفروق بينها داخل مستودع واحد. وأقترح تطوير سجل منشأ لكل مجموعة بيانات، يوضح كيفية جمعها، وأسس ترخيصها، وطبيعة المراجعة، ومستوى الثقة في جودتها.

وهذا ينسجم مع الفكرة التي قدمتها تيمنت جيبرو (Timnit Gebru) وزملاؤها في البحث المعروف Datasheets for Datasets، الذي دعا إلى توثيق مجموعات البيانات كما توثق الصناعات خصائص المنتجات والمكونات.

إن الهدف ليس صناعة أداة تتباهى بأنها تكشف «رائحة الذكاء الاصطناعي».

بل بناء مورد معرفي نعرف تاريخ كل جزء فيه، وماذا يمكن أن نتعلم منه، وما حدود الثقة فيه.

فالرائحة انطباع.

أما المنشأ الموثق فهو معرفة.

نوع المادة	منشؤها	ما الذي ينبغي توثيقه؟
بيانات بشرية مؤلفة أصلاً (Human-Authored Data)	مساهمة أنشأها إنسان	المصدر وطريقة الإنشاء والحقوق
مخرجات آلية راجعها البشر (Human-Verified AI Outputs)	إنتاج آلي تبعته مراجعة بشرية	المولد ودور المراجع وحدود المراجعة
بيانات اصطناعية (Synthetic Data)	توليد آلي	طريقة التوليد والغرض وحدود الاستخدام

تلخيص مفاهيمي للفروق الواردة في هذا القسم؛ لا يجعل أي نوع صحيحاً أو صالحاً تلقائياً.

حين لا يكون اختلاف الناس خطأً في البيانات

إحدى أهم الأفكار التي ينبغي أن تحكم الديوان هي ألا تُعامل جميع الاختلافات بين المشاركين بوصفها أخطاء يجب محوها. قد يرى شخصان صورة واحدة ويصفانها بطريقتين مختلفتين.

وقد يختلف متحدثان من منطقتين حول معنى تعبير، لأن الكلمة تُستخدم في بيئتيهما بمعنيين مختلفين.

وقد يقرأ باحثان عبارة محتملة الدلالة ويصلان إلى تفسيرين معقولين.

في كثير من أنظمة وسم البيانات التقليدية، تحاول المنصة الوصول إلى وسم واحد نهائي (Ground Truth Label)، وقد تعتمد تصويت الأغلبية أو قرار المراجع المتخصص.

وهذا مناسب للسائل التي تملك حقيقة موضوعية واضحة، كقراءة رقم في وثيقة سليمة أو تحديد وجود جسم بعينه في صورة.

لكنه قد يكون أقل ملاءمة للغة والثقافة.

فليس كل خلاف علامة على أن أحد الطرفين مخطئ.

في دراسة بعنوان *Investigating Reasons for Disagreement in Natural Language Inference*، نشرتها نان-جيانغ (Nan-Jiang Jiang) وماري-كاترين دو مارنيف (Marie-Catherine de Marneffe) عام 2022، درس الباحثان أسباب اختلاف المراجعين في تفسير العلاقات بين النصوص، ووجدوا أن بعض الاختلافات تعود إلى احتمالات حقيقية في المعنى، لا مجرد أخطاء عشوائية في الوسم. وهذا بالغ الأهمية لمشروع يريد تمثيل المجتمع السعودي.

قد يكون اختلاف أبناء المناطق حول استعمال كلمة هو بالضبط المعلومة التي تحتاج إليها الآلة.

وقد يكون اختلاف الأجيال حول معنى تعبير مؤشراً إلى تحول لغوي يستحق التوثيق.

لذلك ينبغي أن يستطيع نظام الوسم الاحتفاظ بتعدد الإجابات حين يكون له معنى، وأن يربطها بالسياق الجغرافي أو العمري أو المهني المناسب، دون تحويل البيانات الشخصية غير الضرورية إلى سجلات مكشوفة.

ولا يعني ذلك اقتراض أن لكل رأي الصحة نفسها في جميع الموضوعات.

المسائل الطبية والعلمية والقانونية تحتاج إلى تحقق متخصص، بينما قد تكون التعددية مشروعة في وصف استعمال لغوي أو تقليد اجتماعي.

فبعض البيانات الجيدة لا تعلم النموذج جواباً واحداً، إنها تعلمه لماذا يختلف الناس في الجواب.

وهذا أحد أعظم المكاسب الممكنة من الانتقال من جمع النصوص إلى جمع المعرفة البشرية الموسومة.

ساعة تطوعية قد تصبح أصلاً معرفياً

أصل الآن إلى عنصر أراه حاسماً في نجاح المبادرة: كيف نجعل المشاركة عادة مجتمعية، لا اندفاعاً مؤقتاً ينتهي بانتهاء حملة إعلامية؟

يمكن الاستفادة من المنصة الوطنية للعمل التطوعي في المملكة، التي توفر أصلاً آليات لتنظيم الفرص التطوعية وتوثيق الساعات وإصدار الشهادات.

وتُظهر تقارير رؤية السعودية 2030 التوسع الكبير في العمل التطوعي، إذ سجلت المنصة الوطنية أكثر من 189 مليون ساعة تطوعية تراكمية بحلول تقرير عام 2025.

تخيل أن إحدى الفرص المعتمدة أصبحت: «المساهمة في توثيق اللهجات السعودية»، أو «مراجعة النصوص العربية التاريخية»، أو «وسم الوثائق التعليمية»، أو «إنشاء بيانات صوتية لتحسين فهم العربية».

يدخل المتطوع إلى البوابة، وينجز مهاماً تناسب مع مستواه، ثم تعتمد مساهماته بعد المرور بمراجعة الجودة.

وتُحتسب له الساعات المستحقة وفق معايير زمنية ومهنية معلنة، لا بمجرد عدد النقرات أو الكلمات.

ومن خلال تكامل مؤسسي مع الجهات المختصة، يمكن مستقبلاً إظهار هذه الساعات والشهادات عبر الخدمات الحكومية الرقمية، بما فيها تطبيق توكلنا إن أقر الربط رسمياً.

فلا حاجة إلى إنشاء نظام موازٍ لكل ما هو موجود أصلاً.

الأدكى أن نتصل المبادرة بالبنية التطوعية والتعليمية الوطنية، وأن نضيف إليها مجالاً جديداً هو التطوع المعرفي الرقمي (Digital Knowledge Volunteering).

لكن علينا أن نفهم جانباً حساساً في نظام الخوافز.

إذا كافأنا الناس على عدد الكلمات فقط، فقد نحصل على كثير من الكلمات قليلة القيمة.

وإذا كافأناهم على عدد الوسوم، فقد يسارع البعض إلى الإجابة دون تدقيق.

وهذا يذكرنا بقانون غودهارت (Goodhart's Law) في الاقتصاد والقياس: حين يصبح المؤشر هدفاً للمكافأة، قد يفقد قدرته على تمثيل الجودة التي كان يقيسها.

لذلك ينبغي أن تعتمد المكافآت على جودة المساهمة ومستوى صعوبتها ونتائج مراجعتها.

يمكن أن يحصل المتطوع على نقاط خبرة (Experience Points – XP)، ويرتقي في مستويات معرفية، ويفتح مهاماً أكثر تخصصاً، ويحصل على شارات مهارة (Skill Badges)، وشهادات معتمدة، وفرص تدريب في مختبرات الجامعات.

وقد تنشأ منافسات بين الجامعات والمدارس والمناطق، لا على عدد المساهمات الخام، بل على جودة الموارد التي أنتجتها وتنوعها وأثرها البحثي.

ويمكن أن تشمل الخوافز مكافآت بحثية، وفرص تدريب، ومنحاً للمشروعات الطلابية، وتقديراً للمساهمين المتميزين.

أما الأعمال التي تتطلب خبرة عالية أو جهداً احترافياً منتظماً، فينبغي أن نتوافر لها مسارات عمل مدفوعة الأجر، على غرار بعض تجارب كوريا، بدل الاعتماد الدائم على التطوع.

فالعمل التطوعي ليس ذريعة للحصول على عمل متخصص مجاني.

والمبادرة التي تحترم مساهمها هي الأقدر على المحافظة على جودتها وثقتهم على المدى الطويل.

المدرسة التي تصبح مختبرا للمعرفة

أتصور أن المدارس والجامعات يمكن أن تكون قلب المبادرة.

لكن ليس عبر فرض إنتاج البيانات على الطلاب، أو تحويل الواجبات المدرسية إلى مصادر تدريب تلقائية.

بل من خلال مسارات تعليمية اختيارية ومدرسة، تجعل المشاركة وسيلة لفهم اللغة والبحث والذكاء الاصطناعي.

في المرحلة المتوسطة، يمكن أن يتعلم الطالب كيف يصف صورة، ويفرق بين الحقيقة والتفسير، ويفهم لماذا لا تتفق الآلات دائماً في قراءة النصوص.

وفي المرحلة الثانوية، يستطيع المشاركة في مشروع يوثق تعبيرات محلية أو مفردات علمية، ويتعلم كيف تُصمم أسئلة التقييم وكيف تُراجع النتائج.

أما الطالب الجامعي، فيمكن أن ينتقل إلى مهام أصعب: مراجعة البيانات المتخصصة، ودراسة اتفاق المراجعين (Inter-Annotator Agreement)، واكتشاف التحيز في العينات (Dataset Bias)، وتحليل أخطاء النماذج.

وقد تتحول هذه الأعمال إلى مشروعات تخرج، وأوراق بحثية، ومشروعات ناشئة.

وستكون مشاركة المعلمين وأعضاء هيئة التدريس أساسية في تصميم المهام وتقييمها، لا مجرد الإشراف الإداري عليها.

أما مشاركة الأطفال، فينبغي أن تُحاط بضوابط خاصة: موافقة ولي الأمر متى كانت مطلوبة، وتقليل البيانات الشخصية، وتفضيل المهام التعليمية غير الحساسة، وعدم استخدام صوت الطفل أو صورته أو كتاباته الشخصية للتدريب المفتوح بصورة تلقائية.

فالقانون السعودي لحماية البيانات الشخصية (PDPL – Personal Data Protection Law)، والسياسات المنظمة لبيانات الأطفال، يفرضان أهمية واضحة للغرض المشروع، والموافقة المناسبة، وتقليل البيانات وحمايتها.

وإذا كانت المبادرة قائمة على مساهمة الناس، فإن احترام حقهم في عدم المشاركة جزء من مشروعيتها.

لا ينبغي أن يخسر الطالب درجة، أو فرصة تعليمية، لأنه اختار ألا يمنح بياناته.

نريد أن يتعلم الطفل كيف تُصنع المعرفة، لا أن يصبح هو نفسه مادة خاماً تُجمع دون أن يفهم أين تذهب.

من يملك ما يتبرع به الناس؟

هذه، في رأيي، أصعب مسألة في المبادرة وأهمها.

إذا ساهم مليون إنسان بنصوص وأصوات وصور وتفسيرات، ثم استخدمتها شركة لتطوير نموذج تجاري، فما الذي يعود إلى المجتمع؟

هل تصبح البيانات ملكاً حصرياً لجهة واحدة؟ وهل يستطيع الباحث المستقل الوصول إليها؟ وهل يمكن للمساهم سحب بيانات شخصية لم يعد يريد استمرار معالجتها، ضمن الحقوق والحدود القانونية المناسبة؟

وهل يجوز تحويل مادة ثقافية نادرة إلى سلعة تجارية دون احترام الجماعة التي قدمتها؟

هنا لا تكفي العبارة المطمئنة: «بيانات وطنية».

فالوطنية لا تلغي حقوق الأفراد والمجتمعات في بياناتهم.

أرى أن المبادرة تحتاج إلى نموذج حكمة قريب من مفهوم الأمانة المجتمعية للبيانات (Data Trust)، تدير فيه جهة مسؤولة الموارد وفق تفويض واضح، وتحت رقابة قانونية وعلمية ومجتمعية.

ويكون لكل مجموعة بيانات تصريح يحدد غرض استخدامها، ونطاق الترخيص (Data Licensing)، وحقوق المساهمين، ومستوى الإتاحة.

بعض المجموعات يمكن نشره بصورة مفتوحة (Open Data)، خصوصاً حين يكون خالياً من المعلومات الشخصية ومحفوظ الحقوق.

وبعضه يُتاح للباحثين وفق تراخيص وضوابط.

وبعضه يبقى داخل بيئات حوسبية آمنة (Secure Research Environments)، لا تسمح بنسخ البيانات الخام أو الاطلاع على هوية أصحابها.

ومن المفيد الاسترشاد بمبادئ (CARE Principles for Indigenous Data Governance)، التي طورتها شبكات دولية لحكمة بيانات المجتمعات الأصلية، وتؤكد المنفعة الجماعية، وحقوق المجتمعات في التحكم، والمسؤولية والأخلاقيات.

ولا أقصد نقل هذه المبادئ إلى السياق السعودي حرفياً، بل الاستفادة من فلسفتها العامة: أن المعرفة الثقافية لا ينبغي أن تُنتزع من مجتمعها ثم تنفصل منافعها عنه.

فقد نُبِرع امرأة بتسجيل تشرح فيه طريقة صناعة حرفة تقليدية، لكنها لا تقصد بالضرورة منح شركة ما حق استخدام صوتها في إنشاء نموذج لتقليد هويتها الصوتية.

وقد يسمح شخص بتوثيق قصة محلية، دون أن يجيز الكشف عن أسماء أفراد أسرته أو تفاصيل حياتهم.

وهنا ينبغي أن تفصل المنصة بين الموافقة على جمع المادة، والموافقة على استخدامها في البحث، والموافقة على استخدامها في تدريب النماذج، والموافقة على إتاحتها تجارياً.

فالموافقة ليست كلمة صغيرة نضعها أسفل الشاشة.

إنها العلاقة الأخلاقية التي ستحدد ما إذا كان المجتمع سيثق بالمبادرة ويواصل تغذيتها بمعرفته.

والبيانات التي تُجمع بثقة الناس قد تكون أئمن من بيانات أكثر حجماً جُمعت دون فهم أصحابها لمصيرها.

لماذا قد يكون هذا المشروع أهم من نموذج واحد؟

تأمل ما يمكن أن تنتجه المبادرة بعد سنوات.

مورداً واسعاً للأصوات السعودية واللهجات، يمكن أن تستفيد منه تقنيات التعرف على الكلام (Speech Recognition).

ومجموعات من النصوص العربية الأصلية والموثقة، تساعد في التدريب اللغوي (Language Modeling).

وصوراً ووثائق موسومة يمكن استخدامها في فهم المستندات (Document Understanding)، والرؤية الحاسوبية (Computer Vision).

وقواعد مصطلحات مهنية من الطب والهندسة والتعليم والصناعة.

ومجموعات تقييم عمياء (Blind Evaluation Sets) تستطيع الجامعات استخدامها لاختبار قدرة النماذج على فهم المجتمع.

هذه موارد مختلفة، لا ينبغي جمعها كلها في وعاء واحد دون تمييز.

فبعض البيانات يصلح للتدريب المسبق (Pretraining)، وبعضها للضبط الدقيق (Fine-Tuning)، وبعضها للتقييم فقط (Evaluation).

ويجب أن تُعزل بيانات التقييم بعناية حتى لا تتسرب إلى التدريب، وإلا أصبح النموذج كطالب عرف أسئلة الامتحان قبل دخوله.

لكن حين تُدار هذه الموارد بوصفها بنية علمية مشتركة، يمكن أن تنشأ فوقها منظومة واسعة من الأبحاث والشركات والمنتجات.

ويمكن أن تستفيد منها النماذج الوطنية الموجودة، والمختبرات البحثية، وشركات التقنية العميقة، والجامعات، والشركات الناشئة.

كما يمكن أن تصبح بعض مجموعاتها مورداً مطلوباً خارج المملكة، لتدريب أنظمة متعددة اللغات أو اختبارها أو تطوير تقنيات متخصصة.

وحينها لن يكون الإنجاز مجرد امتلاك أكبر كمية من الرموز العربية (Arabic Tokens).

بل امتلاك بيانات تعرف مصادرها، وتنوعها، وحدودها، وما تضيفه إلى قدرة النموذج.

فالأكبر ليس دائماً الأفضل، والأكثر ندرة ليس دائماً الأقل عدداً.

قد تكون ألف ساعة من كلام طبيعي موثق في بيئة لا تمثلها البيانات العالمية ذات قيمة كبيرة، وقد تكون عشرة ملايين جملة متشابهة مجرد عبء إضافي على التدريب.

رأسمال جديد اسمه المجتمع

أعتقد أن أحد أهم التحولات الاقتصادية في الذكاء الاصطناعي سيكون انتقال الاهتمام من حجم البيانات وحده إلى أصالتها ومنشأها وحقوق استخدامها وقيمتها المعرفية.

كلما أصبحت قدرة النماذج على توليد النصوص أسهل وأرخص، ازدادت صعوبة معرفة ما إذا كانت بعض البيانات الجديدة تمثل تجربة بشرية لم نعلمها النماذج سابقاً، أم أنها إعادة تدوير لاحتمالات النموذج نفسه.

وهنا قد ترتفع القيمة النسبية للبيانات البشرية الأصلية، خصوصاً في المجالات واللغات والتجارب الأقل تمثيلاً.

ليس لأن الإنسان ينتج دائماً بيانات أدق من الآلة، بل لأنه يعيش العالم الذي تحاول الآلة تمثيله.

الطبيب يعرف كيف يصف المرضى أعراضهم في الواقع. والمعلمة تعرف أخطاء الطلاب المتكررة. والمهندس يعرف المصطلحات العملية التي نشأت داخل مصنعه. والمزارع يعرف مفردات الأرض التي لا تظهر في دليل تقني.

هذه الخبرات موزعة بين الناس، ولا يستطيع أي مختبر مركزي جمعها بمفرده.

وهذا يعيدنا إلى فكرة الاقتصادي النمساوي فريدريش هايك (Friedrich Hayek) عن المعرفة المتفرقة (Dispersed Knowledge)، التي تناولها في مقاله الشهير *The Use of Knowledge in Society* عام 1945.

كان هايك يلفت النظر إلى أن قدرًا كبيرًا من المعرفة العملية يوجد موزعًا بين أفراد المجتمع، ويتصل بظروف محلية لا تستطيع جهة مركزية الإحاطة بها كلها.

قد لا تتفق مع جميع استنتاجاته الاقتصادية، لكن هذه الملاحظة تبدو شديدة الملاءمة لعصر الذكاء الاصطناعي.

فلا يوجد مركز بيانات يعرف وحده ما يعرفه مجتمع كامل.

وإذا استطاعت المبادرة بناء وسائل للمساهمة والمراجعة والحفظ، فإنها تكون قد أوجدت قناة لتحويل هذه المعرفة المبعثرة إلى مورد يمكن للعلماء والمهندسين الاستفادة منه.

وهذا أصل اقتصادي جديد.

ليس مجرد مجموعة ملفات ضخمة، بل بنية لإنتاج المعرفة باستمرار، مصدرها المجتمع، ويعود جزء من منفعتها إليه.

البداية التي لا تحتاج إلى انتظار المستقبل

لا أتصور أن تبدأ المبادرة بطلب مليون مساهمة في اليوم الأول.

الأفضل أن تبدأ بمشروعات محددة، يستطيع الباحثون قياس أثرها فعلاً.

يمكن اختيار مسارات أولية في اللهجات والكلام الطبيعي والوثائق العربية والصور الثقافية والمصطلحات المهنية، بمشاركة جامعات ومدارس وجهات غير ربحية ومختبرات بحثية.

ثم تُقسّم البيانات إلى مجموعات تدريب واختبار، وتُقارن نتائج النماذج قبل إضافة المساهمات البشرية وبعدها.

هل تحسن التعرف على لهجة معينة؟ هل قلت أخطاء استخراج الجداول؟ هل أصبح النموذج أكثر قدرة على فهم مصطلحات مهنية؟ هل أدى تنوع المشاركين إلى تقليل الانحياز؟

إذا ظهرت آثار إيجابية واضحة، نتوسع المبادرة إلى مناطق ومجالات جديدة.

أما إذا لم تتحسن النتائج، فينبغي أن نفهم السبب قبل زيادة جمع البيانات.

فالعرض ليس أن نجمع نصوصاً نشعرنا بالفخر بحجمها.

الغرض أن نجمع معرفة تفيد.

ويمكن أن توضع للمشروع أهداف طموحة خلال عدة سنوات، مثل استقطاب مليون مساهم اختياري، وبناء مجموعات بيانات بشرية موسومة في مجالات متعددة، وإنشاء معايير تقييم سعودية معترف بها بحثياً.

لكن النجاح ينبغي أن يقاس بجودة البيانات وتنوعها وأثرها وحقوق أصحابها، لا بعدد المشاركين وحده.

وقد يكون أجمل مؤشر للنجاح أن تكتشف جامعة سعودية، بعد خمس سنوات، أن مجموعة بيانات ساهم في إنشائها طلاب ومعلمون ومتطوعون من مناطق مختلفة أصبحت مرجعاً تستخدمه مختبرات دولية.

في تلك اللحظة، سيكون المجتمع قد ساهم في إنتاج معرفة تتجاوز حدود المكان الذي نشأت فيه.

حين يتبرع الناس بما لا تستطيع الآلة اختراعه

أعود إلى فنلندا.

لم تطلب المبادرة من المواطن العادي أن يصبح عالم حاسوب، ولا أن يفهم المعادلات الرياضية التي تقوم عليها نماذج التعرف على الكلام.

طلبت منه شيئاً أبسط: أن يتكلم.

وكان خلف هذا الطلب إدراك عميق بأن بعض أهم موارد المستقبل موجودة أصلاً داخل المجتمع، لكنها لم تجد الطريق الذي يحولها إلى معرفة قابلة للتعلم.

وأرى أن السعودية تملك فرصة لصياغة نسخة أكثر اتساعاً من هذه الفكرة.

لدينا تعليم عام وجامعي واسع، وقطاع غير ربحي يتطور، ومنصات وطنية للعمل التطوعي، ومؤسسات بحثية وشركات تقنية، ومجتمع غني بتنوعه الجغرافي والاجتماعي والمهني واللغوي.

والسؤال ليس ما إذا كنا نستطيع إقناع المواطنين بإنتاج ملايين الجمل.

السؤال هو ما إذا كنا نستطيع بناء مشروع يجعل كل مساهمة صغيرة جزءاً من عمل علمي كبير، ويحافظ على حق صاحبها في معرفة أين ذهبت وكيف استخدمت.

تخيل طفلاً يصف صورة لأداة لم يرها إلا في منزل جده.

ثم يشرح الجد اسمها واستعمالها القديم.

وتراجع معلمة النص، فيصحح باحث لغوي بعض ألفاظه، ثم يربطه متخصص بمادة تاريخية موثقة.

لقد تحولت دقائق قليلة من التفاعل بين أربعة أشخاص إلى معرفة لا توجد بالضرورة بهذه الصورة في أي موقع على الإنترنت.

وقد تصبح تلك المعرفة مثلاً في مجموعة بيانات، أو جزءاً من معيار تقييم، أو مدخلاً في معجم رقمي، أو مادة لتطوير نموذج يفهم تاريخ الأشياء وثقافتها.

هذه هي الصورة التي أراها جديرة باسم «الديوان السعودي للمعرفة البشرية».

ليس مستودعاً ضخماً نلقي فيه ما يتبرع به الناس.

بل مؤسسة تحسن الإصغاء إليهم، وتُعلم الآلات من خلاهم، وتمنحهم في المقابل مهارة وتقديراً ونصيبة عادلاً من منفعة المعرفة التي ساعدوا على إنشائها.

وعندها لن تكون الأم التي فسرت عبارة محلية مجرد مساهمة صغيرة في قاعدة بيانات.

ولن يكون الطالب الذي صحح كلمة في مخطوطة مجرد متطوع قضى خمس دقائق أمام شاشة.

سيكون كل منهما قد شارك، بطريقته، في بناء شيء أكبر.

فقد يكون المستقبل الذي نريد أن نتعلم فيه الآلات ثقافتنا معلماً بكلمات لم يقلها أحد أمام ميكروفون بعد، وبوثائق لم يقرأها الحاسوب، وبمعرفة تحملها امرأة أو رجل لا يعرف أن ما لديه سيكون يوماً مورداً علمياً ثميناً.

وما أتمناه ليس أن نطلب من المجتمع أن يتكلم أكثر كي نملأ ذاكرة الآلة.

بل أن نبني نظاماً يصغي لما يعرفه المجتمع فعلاً.

فالأمم لا تمنح الذكاء الاصطناعي لغتها بمجرد أن تترجم إليه ما يعرفه الآخرون؛ تمنحه شيئاً من نفسها حين تجعل أبنائها وبناتها شركاء في كتابة المعرفة التي سيتعلم منها.

وربما يبدأ أعظم مشروع سعودي للذكاء الاصطناعي، لا من غرفة مليئة بالخدادم، بل من بيت عادي، حين تقول جدة لحفيدتها: «تعال، أعلمك كلمة لا يعرفونها».

فتتعلم الحفيدة.

وتتعلم الآلة بعدها.

ويبقى الإنسان أول المعلمين.

المراجع والمصادر الأولية

1. [Donate Speech \(Lahjoita puhetta\)](#) — The Language Bank of Finland (Kielipankki).
2. [Lahjoita puhetta: a large-scale corpus of spoken Finnish with some benchmarks](#) — Language Resources and Evaluation / Springer Nature.
3. [Over 10,000 contributors onboarded on Bhashini Samudaye](#) — Press Information Bureau, Government of India / Ministry of Electronics & IT.
4. [AI Data](#) — National Information Society Agency, Republic of Korea.
5. [AI-Hub data catalog](#) — AI-Hub / National Information Society Agency, Republic of Korea.
6. [Data Dam project begins, being key to Digital New Deal](#) — Ministry of Science and ICT, Republic of Korea.
7. [Data labeler established as a new future-oriented job](#) — National Information Society Agency, Republic of Korea.
8. [Introducing Aquarium: An Open Data Platform for Southeast Asian Languages](#) — AI Singapore / SEA-LION.
9. [We're expanding our presence in Singapore to advance AI in the Asia-Pacific region](#) — Google DeepMind.
10. [TI-Hüpe jõuab sügisest kutseharidusse ja põhikooliõpetajateni](#) — Estonian Ministry of Education and Research.
11. [AI Leap](#) — e-Estonia / Estonian Government.
12. [By the People: LC Crowdsourcing](#) — Library of Congress.
13. [Achievements \(Points\)](#) — Google Crowdsourcing Help.
14. [How does Google use my answers?](#) — Google Crowdsourcing Help.
15. [Common Voice](#) — Mozilla Common Voice.
16. [GPT detectors are biased against non-native English writers](#) — Patterns / Cell Press.
17. [Testing of detection tools for AI-generated text](#) — International Journal for Educational Integrity / Springer Nature.

18. [Datasheets for Datasets](#) — arXiv (preprint).
19. [Datasheets for Datasets](#) — Communications of the ACM (published article).
20. [Investigating Reasons for Disagreement in Natural Language Inference](#) — Transactions of the Association for Computational Linguistics / MIT Press.
21. [Saudi Vision 2030 Annual Report 2025](#) — Saudi Vision 2030.
22. [Ministry platforms — Volunteer Platform](#) — Saudi Ministry of Human Resources and Social Development.
23. [Personal Data Protection Law](#) — Saudi Data & AI Authority.
24. [Implementing Regulations of the Personal Data Protection Law](#) — Saudi Data & AI Authority / National Data Management Office.
25. [Children and Incompetents Data Protection Policy](#) — Saudi Data & AI Authority / National Data Management Office.
26. [The CARE Principles for Indigenous Data Governance](#) — Global Indigenous Data Alliance.
27. [The Use of Knowledge in Society](#) — American Economic Review / American Economic Association.
28. [Conference Volumes — Papers in Monetary Economics](#) — Reserve Bank of Australia.