تحليل المحتوى الرقمي في العلوم الاجتماعية: من التحليل الكلاسيكي إلى النمذجة الحاسوبية

يوسف حاسين
دكتوراه في علم الاجتماع، جامعة محمد الأول، وجدة
المغرب

الملخص:

تطرح مسألة جمع وتحليل المحتوى الرقمي في العلوم الاجتماعية تحديات منهجية جوهرية، نظرا للحجم الكبير والطبيعة الديناميكية وغير المنظمة للبيانات الرقمية. الأمر الذي يجعل المقاربات المنهجية التقليدية عاجزة عن مواكبة تحولات الثورة الرقمية، ويستدعي ابتكار أدوات منهجية جديدة قادرة على استيعاب التفاعلات الاجتماعية للأفراد في سياقها الرقمي. يتيح الذكاء الاصطناعي تقنيات مهمة تساعد على بناء مقاربات منهجية أكثر دقة للقضايا والظواهر الاجتماعية الرقمية، غير أنه يطرح في نفس الآن العديد من التحديات التي من شانها أن تقوض الإمكانات التي يفتحها أمام الباحثين في العلوم الاجتماعية. تسعى هذه الورقة إلى الكشف عن الإمكانات الكبيرة التي توفرها تقنيات الذكاء الاصطناعي في مجال تحليل المحتوى الرقمي، مثل النماذج اللغوية الكبيرة (Large Language Models) لمعالجة الأحجام الضخمة من البيانات، واستخلاص أنماط العلاقات المعقدة بين المتغيرات بشكل أكثر دقة. كما تناقش التحديات التي تعترض توظيف هذه التقنيات، من قبيل الترميز الآلي ومدى قدرته على استيعاب تعقيد الوقائع الاجتماعية المتعددة الأبعاد، وصعوبة التمييز بين المواقف الحقيقة للأفراد وتلك المرتبطة بالتلاعب بالرأي العام عن طريق الوسائط الرقمية. وذلك بهدف بناء تصور منهجي واضح ودقيق، عن الكيفية التي يمكن من خلالها استخدام تقنيات الذكاء الاصطناعي في تحليل المحتوى الرقمي، دون السقوط في المزلقات المنهجية والنظرية المصاحبة لها.

الكلمات المفتاحية: المحتوى الرقمي، الذكاء الاصطناعي، العلوم الاجتماعية.

Abstract:

The collection and analysis of digital content in the social sciences present fundamental methodological challenges, primarily due to the massive scale and the dynamic, unstructured nature of digital data. Consequently, traditional methodological approaches have become increasingly inadequate in keeping pace with the transformations of the digital revolution, necessitating the innovation of new methodological tools capable of capturing social interactions within their digital context. Artificial Intelligence (AI) offers significant techniques that facilitate the development of more precise methodological approaches to digital social phenomena; however, it simultaneously poses various challenges that could undermine the potential it offers to social science researchers. This paper explores the substantial potential of AI technologies in the field of digital content analysis—specifically Large Language Models (LLMs)—for processing massive datasets and accurately extracting complex patterns of relationships between variables. Furthermore, it examines the challenges associated with the deployment of these technologies, such as automated coding and its capacity to capture the complexity of multidimensional social facts, as well as the difficulty of distinguishing between genuine individual attitudes and those influenced by public opinion manipulation via digital media. The ultimate objective is to establish a clear and rigorous methodological framework for utilizing AI techniques in digital content analysis while avoiding the associated theoretical and methodological pitfalls.

Keywords : Digital Content, Artificial Intelligence, Research Methods, Social Sciences.

مقدمة:

أدت الثورة الرقمية إلى تحولات جوهرية في أنماط التفاعل الاجتماعي بين الأفراد؛ فقد انتقلنا من التفاعل المادي "وجها لوجه" في السياقات والمواقف الاجتماعية الواقعية، إلى التفاعل الاجتماعي الرقمي عبر وسائل التواصل الاجتماعي. وهو انتقال أدى إلى توسيع شبكة التفاعل الاجتماعي أضعاف ما كانت عليه قبل الثورة الرقمية، حيث ينتج عن النشاط الرقمي للأفراد كميات هائلة من البيانات الرقمية، عبارة عن منشورات أو نقرات أو تعليقات تمثل "أثرا رقميا" يحمل دلالات سوسيولوجية وسيكولوجية تعبر عن مواقفهم واتجاهاتهم وحالاتهم النفسية. الأمر الذي يجعل من "المحتوى الرقمي" مختبرا مفتوحا للباحثين في العلوم الاجتماعية. بيد أن الحجم الكبير، والطبيعة الدينامية وغير المنظمة للمحتوى الرقمي، يضع الباحث في العلوم الاجتماعية أمام مأزق منهجي، يتمثل في عدم قدرة المناهج التقليدية، بما في ذلك منهج تحليل المحتوى في صيغته التقليدية، على مواكبة التحولات التي مسّت عمليات إنتاج المحتوى. فبعد أن كان فيه الباحث يحلل بضع عشرات من النصوص أو الصور أو غيرها من المواد الحاملة للمحتوى، وجد نفسه اليوم أمام ملايين التفاعلات اللحظية التي يستحيل معالجتها بالاعتماد على القدرات البشرية وحدها. وهو ما أدى إلى بروز فجوة عميقة بين سرعة التحولات الاجتماعية وتطور الأدوات المنهجية المعتمدة في دراستها.

أمام هذا الوضعية الابستمولوجية الراهنة للعلوم الاجتماعية، تبرز أدوات الذكاء الاصطناعي بوصفها وسيطا تقنيا مساعدا على سدّ هذه الفجوة، لكن مع ضرورة العمل على تجاوز التحديات التي من شأنها أن تقوض الإمكانات التي يفتحها أمام الباحثين في العلوم الاجتماعية. وذلك عبر إحداث توازن بين المنطق التقني الخوارزمي والمنطق التأويلي للعلوم الاجتماعية، بشكل يمنع من السقوط في النزعة التقنية الجافة وتجريد البيانات الرقمية من بعدها الإنساني. ويعمل الباحثون في إطار ما بات يسمى بالعلوم الاجتماعية الحاسوبية Computational Social Science بوصفه حقلا معرفيا متعدد التخصصات، يدمج بين العلوم الاجتماعية وعلم الحاسوب على توظيف أدوات الحوسبة المتقدمة في استخلاص المعطيات بشكل آلي، وتحليل الشبكات الاجتماعية ونظم المعلومات الجغرافية وغيرها من المجالات التي يعتبر فيها الحاسوب أداة أساسية للتحليل والدراسة. "فكما استغل غاليليو التلسكوب كأداة رئيسية لرصد الكون المادي واكتساب فهم أعمق وأكثر دقة تجريبية، يتعلم الباحثون في العلوم الاجتماعية الحاسوبية استغلال أدوات الحوسبة المتقدمة لرؤية ما وراء نطاق التحليلات التقليدية"[1]. في هذا السياق، يبرز الانتقال من التحليل الكلاسيكي للمحتوى إلى النمذجة الحسابية ليس بوصفه استبدالا للأدوات والتقنيات اليدوية بالأدوات الآلية، بل باعتباره انخراطا في "براديغم" العلوم الاجتماعية الحاسوبية. هذا الحقل المعرفي الذي يسعى لتجاوز المناهج التقليدية عبر استثمار 'الآثار الرقمية' كمادة خام لفهم التفاعلات الإنسانية. وهو ما يجعل من أدوات الذكاء الاصطناعي جسرا تقنيا مهما يربط بين ضخامة البيانات الرقمية وبين القدرة على بناء نماذج تفسيرية وتنبؤية للظواهر الاجتماعية، محولاً بذلك 'المحتوى الرقمي' من مجرد وثيقة لغوية إلى مؤشر سلوكي قابل للقياس الحسابي والتحليل العلمي الدقيق.

بناء على هذه التقاطعات، ستعمل هذه الورقة على الإجابة عن سؤال مركزي مؤداه: كيف يمكن لتقنيات الذكاء الاصطناعي أن تعيد تشكيل أدوات تحليل المحتوى الرقمي في العلوم الاجتماعية، بما يضمن الاستفادة من القوة الحسابية الهائلة دون الإخلال بالدقة العلمية أو العمق التأويلي للظواهر الاجتماعية؟ وذلك بالتركيز على الإمكانات التي تتيحها النماذج اللغوية الكبيرة والأدوات المنهجية المعتمدة على الذكاء الاصطناعي في تحليل المحتوى الرقمي، ومدى قدرة الخوارزميات على استيعاب الواقع الاجتماعي المعقد، والتحديات المنهجية والأخلاقية التي يواجهها الباحثون أثناء استعمالهم لهذه الأدوات.

أولا: المحتوى الرقمي: من التحليل الكلاسيكي للمحتوى إلى النمذجة الحاسوبية

شهدت طبيعة 'المادة' التي يتعامل معها الباحثون في العلوم الاجتماعية الذين يستعملون منهجية تحليل المحتوى تحولا جذريا عقب الانتشار الواسع لأدوات التواصل الاجتماعي. فبعد أن كان المحتوى مقتصرا على بيانات ذات طبيعة مادية، ساكنة، ومحدودة إلى حدّ ما، مثل

الصحافة الورقية ووثائق الأرشيف الكتب المدرسية وغيرها، وجدت العلوم الاجتماعية نفسها أمام محتوى رقمي ذو طبيعة غير مستقرة، فهو قابل للتعديل والحذف والتعليق، وهو مرتبط بشبكة واسعة من الروابط والتفاعلات التي تجعل حدوده غير متناهية. علاوة على حجمه الضخم وغير المنضم، حيث يمتزج النص بالصوت والصورة. مما يعقّد عملية إخضاعه لأساليب التفريغ التقليدية، وهو بخلاف المحتوى المادي، يشكل "أثرا اجتماعيا" لسلوك الفرد، حينما يسجل السلوك لحظة حدوثه عبر البث المباشر والنشر والمشاركة والتعليق...إلخ. هذه الأنشطة التي تشكل بصمات سيكو–سوسيولوجية تعكس التفاعل الاجتماعي–الرقمي واتجاهات التأييد والمعارضة وبنية العلاقات بين الفاعلين الرقميين.

دفع هذا التحول الذي شهدته البيانات الخاضعة لتحليل المحتوى إلى إعادة النظر في هذه الأداة المنهجية وكيفية تطبيقها والتقنيات والأدوات التي تستعملها. حيث شكل ظهور العديد من منصات التواصل الاجتماعي – مثل أكاديميا للأكاديميين والباحثين، وفيسبوك لمشاركة النصوص والصور، والمدونات والتطبيقات للتواصل بين المبرمجين ومؤيدي المصادر المفتوحة وغيرها من فضاءات إنتاج المحتوى الرقمي، تحديات مفاهيمية ومنهجية كبيرة، نظرا للتأثير الذي خلّفه استخدامها الواسع النطاق على طبيعة الخطاب العام، وممارسات الأفراد والروابط بينهم.

وتؤثر هذه المنصات على كيفية توظيف الباحثين في العلوم الاجتماعية لمنهجية تحليل المحتوى حسب كلاوس كريبندورف Klaus Krippendorff بأربع طرق على الأقل[2]؛ أولها، الكميات الهائلة من الصور والنصوص الرقمية سهلة الوصول، حيث شجع هذا التحول الباحثين في العلوم الاجتماعية ذوي التوجه الإحصائي على وضع نظريات تخدم المؤسسات السياسية والتجارية والاقتصادية ذات التوجه الإحصائي أيضًا. كما دفع محللي المحتوى إلى تطوير أساليب حسابية أكثر تقدما للكشف عن البنى الكامنة وراء هذه الوسائط والفضاءات الرقمية والمحتوى الذي يتم إنتاجه من خلالها.

ثانيًا، الفجوة بين الطبيعة العالمية لتقنيات الإنترنت وبين تجربة المستخدمين لنطاق المحدود لمشاركتهم الفردية. وهو ما أدى إلى خلق مجتمعات افتراضية من الأفراد ذوي التفكير المتقارب، والذين قد ينعزلون عن مجتمعات أخرى مماثلة في انغلاقها ولكنها مختلفة في التفكير، مما يُعزز بالتالي أنواعًا جديدة من الانقسامات والصراعات الاجتماعية. ولم يعد بإمكان محللي المحتوى، في مواجهتهم لهذه الاختلافات تطبيق نفس التصنيفات التحليلية التي كانت معتمدة سابقا في تحليل المحتوى المادي. وبالتالي سيتعين عليهم تبني منظور جديد يُراعي الاختلافات الناشئة في استخدام النصوص وتفسيرها.

أما الطريقة الثالثة، فتتمثل في الديناميات التي يُظهرها استخدام وسائل التواصل الاجتماعي. فبينما كانت النصوص المطبوعة تتيح قراءات متباينة من قِبل مجتمعات مختلفة، نظرا لطريقة إنتاجها وخصائصها المادية ودورة حياتها المختلفة عن النصوص الرقمية. تعتبر بعض رسائل وسائل التواصل الاجتماعي في المحتوى الرقمي غير مؤثرة، بينما يُعاد إنتاج البعض الآخر بأعداد هائلة، وينتشر بسرعة، ويستحوذ على اهتمام جماعات مصالح كبيرة، ويُشجع على تحركات اجتماعية عفوية. وفي الوقت الذي لا يستطيع فيه الناشطون الإعلاميون الاحتفاظ بسجلات طويلة لمشاركتهم، غالبًا ما يستطيع محللو المحتوى الاستفادة من البيانات التي تحتفظ بها منصات التواصل الاجتماعي. غير أن الاحتفاء بحجم قواعد البيانات هذه باعتبارها مناجم ذهب بالنسبة للباحثين في العلوم الاجتماعية الذين يستخدمون تحليل المحتوى، يُخفي ديناميات جوهرية حول دور وسائل التواصل الاجتماعي في المجتمع والرقابة التي تمارس على البيانات التي تتضمنها هذه المنصات.

وأخيرا تظهر الطريقة الرابعة في نوع النصوص الجديد الذي تُنتجه وسائل التواصل الاجتماعي.، بما في ذلك الرسائل القصيرة التي لا معنى لها بمعزل عن سياقها، حتى عند جمعها إحصائيًا. حيث أن معانيها لا توجد في القواميس أو المعاجم، بل في كيفية تفاعل المشاركين مع الرسائل السابقة وما يحدث لها بعد نشرها. وبالرغم من أن الجوانب التقنية للمنصات المختلفة تلعب دورًا هامًا في كيفية ربط المستخدمين للنصوص، إلا أن الروابط بين رسائل وسائل التواصل الاجتماعي تُحدد سياق طبيعتها الشبكية، حيث أن أحد هذه الروابط تتمثل في

ظهور مفردات واختصارات ورموز تعبيرية ومواضيع محددة خاصة بكل مجتمع، والتي تتطور في شبكات معينة. فضلا عن رابط آخر هو هوية المشاركين – سواء كانت مفترضة أو حقيقية.

إن الكميات الهائلة من المعلومات حول ما يفعله الناس، وما يعرفونه، وما يفكرون فيه، وما يشعرون به في النصوص الرقمية، ويتزايد الآن حجم الحياة الاجتماعية التي تتم بشكل أساسي عبر هذه الوسائط، وتتنوع مصادر النصوص الرقمية المتاحة، حيث تشمل الأنشطة الجماعية على الإنترنت، ووسائل التواصل الاجتماعي، والرسائل الفورية، بالإضافة إلى المعاملات الإلكترونية، والسجلات الطبية، والرسائل والكتيبات والمقالات والكتب الرقمية، أدى إلى زيادة الطلب على أدوات معالجة اللغة الطبيعية والتعلم الآلي لفلترة النصوص والبحث فيها وترجمتها إلى بيانات مفيدة علميا[3].

ونتيجةً لزيادة القدرة الحاسوبية، والتقدم في مجال التعلم الآلي، وتوفر مجموعات كبيرة من النصوص على الإنترنت، ازداد استخدام الأساليب الحسابية، بما في ذلك تحليل المحتوى بالاعتماد على البرامج الآلية بشكلٍ سريع. حيث أصبح يُطبَّق تحليل المحتوى آليا ويُطوَّر في مختلف التخصصات، مثل علوم الحاسوب، واللغويات، والعلوم السياسية، والاقتصاد، وبشكلٍ متزايد، علوم الاتصال. وتصف أساليب تحليل المحتوى الرقمي، أو ما يُعرف بـــ "النص كبيانات text as data"، منهجًا يتم فيه تحليل النص، تلقائيًا بواسطة برامج آلية، وعلى عكس الترميز اليدوي، لا يُقرأ النص ويُفهم كوحدة واحدة في التحليل الآلي، بل يُجزأ تلقائيًا[4].

تتأسس منهجية التعامل مع "النص بوصفه بيانات"[5] في العلوم الاجتماعية والسياسية حسب ويلكرسون Wilkerson وكاساس Casas على استثمار القدرات الحوسبية لتوسيع آفاق البحث، سواء من خلال تيسير جمع البيانات الضخمة أو إتاحة خيارات متقدمة لتحليلها كميا، وهي عملية تمر عبر أربع مراحل إجرائية متكاملة؛ تبدأ بمرحلة تحصيل النصوص. حيث تتفاوت سبل الحصول على المحتوى الرقمي بين استخدام واجهات برمجة التطبيقات APIs التي تتيح استرجاع بيانات مهيكلة بكفاءة عالية، وبين تقنيات الاستخراج الآلي Scrapingمن صفحات الويب والوثائق الرقمية التي تعتمد على تتبع علامات التنسيق المخفية، وصولاً إلى حلول التي تستلزم تدخلاً بشرياً لتجميع المحتوى وتصنيفه.

وبمجرد الحصول على المادة النصية، تبدأ المرحلة الثانية المتمثلة في تحويل النص إلى بيانات كمية، حيث يسعى الباحث إلى بناء "مصفوفة المصطلحات والوثائق" بعد تحديد وحدة التحليل المناسبة سواء كانت جملة أو فقرة أو وثيقة كاملة. وتتطلب هذه الخطوة معالجة مسبقة دقيقة تشمل إزالة العناصر غير الجوهرية من النصوص، مثل علامات الترقيم أوالكلمات الشائعة، مع ضرورة توخي الحذر المنهجي لضمان عدم حذف سمات قد تكون حاسمة في فهم سياق النص أو دلالاته التعبيرية، وهو ما يمهد الطريق لتطوير نماذج تتجاوز مجرد حصر الكلمات المفردة لتشمل تركيبات لغوية ومعلومات سياقية أوسع.

أما في المرحلة الثالثة المتعلقة بالتحليل الكمي، فإن البحث يراوح بين المقاييس البسيطة التي تتميز بالشفافية وقابلية التكرار، وبين نماذج التعلم الآلي والأساليب الإحصائية المتقدمة. ويظهر هنا تمايز منهجي جوهري؛ حيث يركز الإحصاء التقليدي على اختبار النظريات وتفسير دور المتغيرات المستقلة في التأثير على النتائج، في حين ينصب اهتمام التعلم الآلي على المخرجات ودقة التنبؤ، مما يدفع الباحثين إلى تجريب خوارزميات متنوعة للوصول إلى أعلى درجات الدقة التنبؤية حتى لو ظل الأداء الداخلي للخوارزمية يتسم بقدر من التعقيد غير المفسر.

وتختتم هذه العملية بمرحلة التحقق وتقييم الأداء لضمان صدقية الاستدلالات المستخلصة، حيث يتم في نماذج التعلم الخاضع للإشراف مقارنة النتائج التنبؤية بـــ "معايير ذهبية" صنفها البشر مسبقاً مع مراعاة اختبار النموذج على بيانات مستقلة لتفادي الانحياز. وفي المقابل، تعتمد طرق التحقق في النماذج غير الخاضعة للإشراف على مقاربات متعددة الأبعاد تشمل فحص المنطق الداخلي للموضوعات المستخرجة، والتأكد من اتساق المجموعات الناتجة عبر خوارزميات مختلفة، وربط التحولات في محتوى النصوص بالوقائع والأحداث المسجلة في العالم الواقعي، مما يضفي صبغة العلمية والنسقية على معالجة المحتوى الرقمي في العلوم الاجتماعية.

ينتقل البحث الاجتماعي في هذا النموذج من "قراءة النصوص" إلى "نمذجتها حسابياً"، حيث تصبح الأولوية هي القدرة على استخلاص أنماط كلية من أحجام ضخمة من البيانات، لا يمكن للعقل البشري الإحاطة بها يدوياً. ولا يعتبر هذا التحول من "قراءة النصوص" إلى "نمذجتها حسابياً" مجرد استبدال لجهد بشري بآلة، بل هو إعادة صياغة إبستمولوجية لمنطق الاستدلال في العلوم الاجتماعية. ففي القراءة التقليدية، ينصب تركيز الباحث على تفكيك المعاني الضمنية، واستجلاء المقاصد، وتأويل السياقات داخل عينة محدودة ومنتقاة بعناية. أما في النموذج الحسابي، فيتم الانتقال إلى معالجة كميات كبيرة من البيانات، بحيث لا تصبح الغاية هي فهم "ماذا يقول نص معين" فقط، بل اكتشاف "كيف يتشكل الخطاب" عبر آلاف أو ملايين النصوص.

كما يمكن للنماذج اللغوية الكبيرة أن تساعد على أتمتة عملية تحليل المحتوى وتحقيق نتائج مهمة على مستوى فهم وتفسير الظواهر المدروسة. فقد وضعت دراسة جويانيس وزملاؤه[6]، حول أتمتة تحليل المحتوى باستخدام "نموذج ChatGPT" بروتوكولاً منهجياً مبتكراً لاستخدام هذا النموذج. حيث ينطلق الباحثون من فرضية أن الذكاء الاصطناعي، بوصفه نموذجاً لغوياً ضخماً، يمتلك القدرة على تبسيط وتسريع ترميز كميات هائلة من النصوص بكلفة اقتصادية وزمنية أقل، مع تحقيق مستويات صدقية تقارب أو تتجاوز الترميز البشري التقليدي الذي يتسم باستهلاكه الكبير للوقت والموارد.

يعتمد البروتوكول المقترح على تحويل "دليل الترميز Codebook" التقليدي الذي يتضمن المتغيرات، الأوصاف، وقواعد الترميز إلى أوامر مهيكلة Structured Prompts توجه تحليل النموذج اللغوي المعتمد. وتتضمن هذه العملية تحديد دور النموذج كباحث في تحليل المحتوى، وتزويده بتعريفات دقيقة لكل متغير والقيم الممكنة له، مع إلزام النموذج بتقديم الإجابات في قالب محدد قصد تسهيل المعالجة اللاحقة، وتوخي الدقة وتجنب الاختلاق غير المطلوب في المهام التصنيفية. وللتحقق من كفاءة هذا البروتوكول، طُبقت الدراسة على عينة مؤلفة من 980 ملخصاً لبحوث علمية في مجال الاتصال، بهدف تصنيف المقاربات البحثية (كمية، نوعية، أو مختلطة) وتحديد أدوات جمع البيانات. وأظهرت النتائج قدرة عالية لـــ ChatGPT في تمييز الدراسات الكمية) بمعامل F1-score بلغ 0.96 (والدراسات النوعية (بمعامل 0.82)، كما حقق النموذج مستويات اتفاق جوهرية مع المحللين البشر وفقاً لمقاييس "كابا كوهين" و"ألفا كريبندورف". غير أن النموذج واجه تحديات واضحة في تصنيف البحوث ذات المناهج المختلطة (بمعامل 0.60)، وفي التعامل مع الفئات غير المحددة بدقة منهجية واضحة. حيث سجلت أدنى مستويات الأداء (0.26).

وتشير مراجعة الأخطاء التي وقع فيها النموذج إلى أن الصعوبة تكمن غالباً في التداخل المفاهيمي بين المنهجيات أو العلاقات التراتبية بينها؛ حيث خلط النموذج أحياناً بين "تحليل الخطاب" و"تحليل المحتوى"، أو بين "الإثنوغرافيا" و"المقابلة" نظراً لاشتراكها في سمات إجرائية. وخلصت الدراسة إلى أن الذكاء الاصطناعي يمثل أداة دعم قوية وقابلة للتوسع في مشاريع تحليل المحتوى، بشرط الالتزام بتعريفات فئوية واضحة وتصميم أوامر برمجية دقيقة، مما يتيح للباحثين التركيز على المهام التفسيرية والتحليلية المعقدة بدلاً من الاقتصار على عمليات الترميز اليدوي المحدودة.

علاوة على ذلك يتيح الذكاء الاصطناعي للباحثين تحليل الأرشيفات الرقمية الضخمة لرصد تحول القيم الاجتماعية عبر عقود، وهو أمر يستحيل القيام به يدوياً. ففي دراسة لكوزلوفسكي، تادي، وإيفانز، بعنوان "هندسة الثقافة: تحليل المعنى من خلال تضمين الكلمات"[7]، يكشف كوزلوفسكي وزملاؤه عن جدوى ونجاعة الاعتماد على القدرات الحسابية لأدوات الذكاء الاصطناعي في إنتاج فهم أعمق للارتباطات والفئات الثقافية مقارنة بالأساليب التقليدية. وذلك من خلال تحليل التطور المشترك للروابط بين النوع الاجتماعي والطبقة في الولايات المتحدة خلال القرن العشرين، وتحليل مقارن للفروقات التاريخية بين مؤشرات النوع الاجتماعي والطبقة في الولايات المتحدة وبريطانيا.

تنطلق الدراسة من المبادئ اللسانية البنيوية لفرديناند دي سوسير. وتوظف نماذج "تضمين الكلمات Word Embeddings" القائمة على الشبكات العصبية لتمثيل النصوص كفضاءات هندسية عالية الأبعاد، حيث تتحول الكلمات إلى "متجهات" رياضية يعكس تقاربها أو تباعدها المكاني حجم التشابه الدلالي والثقافي بين المفاهيم في الوعي الاجتماعي العام، مما يسمح بتحويل المحتوى الكيفي إلى إحداثيات كمية قابلة للقياس والمقارنة. واستطاع الباحثون استخلاص أبعاد ثقافية مهيكلة من هذه الفضاءات المتجهية، وذلك من خلال تحديد محاور رياضية تمثل ثنائيات اجتماعية كبرى مثل الرجل مقابل المرأة أو الغني مقابل الفقير أو الليبرالي مقابل المحافظ. فمن خلال تقنية الإسقاط الهندسي للكلمات على هذه المحاور، نجح الباحثون في قياس الارتباطات الثقافية الكامنة والتحيزات النوعية والطبقية بدقة حسابية عالية، وهو ما يحول اللغة من مجرد وسيلة للتواصل إلى مختبر سوسيولوجي يمكن من خلاله رسم خرائط هندسية للقيم والتمثلات الاجتماعية السائدة، ورصد كيفية ارتباط مفاهيم معينة بمهن أو صفات أو طبقات اجتماعية محددة دون الحاجة لطرح أسئلة مباشرة على المبحوثين. ولضمان الصرامة العلمية لهذه المقاربة الحسابية، قام لكوزلوفسكي وزملاؤه بمقارنة النتائج المستخلصة من نماذج الذكاء الاصطناعي مع بيانات المسوحات الاجتماعية التقليدية الموثوقة، ليجدوا توافقاً إحصائياً كبيرا يؤكد أن الخلاصات الرياضية في عالم الخوارزميات تعكس بصدق التعبيرات الثقافية في الواقع الاجتماعي.

سمح هذا النموذج بالكشف عن كيفية تحول الصور النمطية وتغير مكانة المجموعات الاجتماعية عبر الزمن، مما يبرهن على أن النمذجة الحسابية لا تكتفي بتسريع التحليل، بل تقدم رؤى سوسيولوجية عميقة حول ديناميكيات التغير الثقافي الكبرى التي تعجز القراءات اليدوية عن الإحاطة بها. وبأن الأدوات المنهجية التي يتم تطويرها بالاعتماد على أدوات الذكاء الاصطناعي، تفتح آفاقاً غير مسبوقة للبحث الاجتماعي، عبر الجمع بين القدرة على معالجة البيانات الضخمة وبين العمق النظري في فهم الهوية والطبقة والثقافة. وبذلك، توفر هذه المقاربة أداة موضوعية لاستنطاق الأرشيفات الرقمية والتاريخية وتجاوز محدودية العينات في الدراسات التقليدية.

ثانيا: التحديات المنهجية والأخلاقية لتوظيف الذكاء الاصطناعي في تحليل المحتوى الرقمي

على الرغم من الآفاق الواسعة التي تفتحها أدوات الذكاء الاصطناعي في تطوير تقنيات تحليل المحتوى الرقمي وتوسيع نطاق العينات المدروسة، إلاّ أن الممارسة التطبيقية تكشف عن مجموعة من التحديات المنهجية والتقنية والأخلاقية التي ينبغي العمل على تجاوزها. فبقدر ما توفره هذه الأدوات من سرعة في الترميز، بقدر ما تصطدم بإكراهات ترتبط بطبيعة "المنطق الخوارزمي" الذي قد يعجز عن استيعاب التداخلات المفاهيمية المعقدة أو الفئات المنهجية المتداخلة، كما رأينا في المحور السابق. فعلى سبيل المثال، تطرح الممارسة المنهجية لتحليل "النص بوصفه بيانات" مجموعة من التحديات والإكراهات التي تستوجب حذراً ابستمولوجياً، وذلك بدءا بمرحلة تجميع النصوص أو المادة التي ستخضع لتحليل المحتوى. حيث تبرز صعوبات تقنية تتعلق بمدى توفر واجهات برمجة التطبيقات (APIs) أو تعقيد عمليات الاستخراج الآلي Scraping من مصادر غير متجانسة، مما قد يستهلك وقتاً طويلاً ويؤثر على شمولية العينة. وعند الانتقال من النص إلى البيانات، فإن أحد أكبر التحديات المنهجية يكمن في عملية "المعالجة المسبقة"، حيث إن القرارات المتعلقة بحذف الكلمات الشائعة أو علامات الترقيم قد تؤدي إلى فقدان سمات جوهرية لتحليل السياق، مثل أدوات النفي التي قد تغير المعنى الكلي للنص في دراسات تحليل المواقف السياسية على سبيل المثال.

أما فيما يخص مرحلة التحليل الكمي، فيشير غريمر وستيوارت[8] Grimmer and Stewart إلى وجود فجوة بين الأساليب الإحصائية التقليدية ونماذج التعلم الآلي؛ فبينما يركز الإحصاء على التفسير وفهم العلاقات بين المتغيرات، يميل التعلم الآلي نحو إعطاء الأولوية للتنبؤ ودقة المخرجات، وهو ما قد يؤدي إلى نتائج دقيقة إحصائياً، وفقيرة من حيث التفسير النظري العميق لأسباب ظهور هذه الأنماط. ويظهر هذا التحدي بوضوح في النماذج التي تعتمد على ميزات لغوية قد لا ترتبط مباشرة بالظاهرة المدروسة، مما يجعل الباحث أمام معضلة "الصندوق الأسود" حيث يصعب شرح الآلية التي اتخذت بها الآلة قرارها التصنيفي.

وتعد مرحلة التحقق وتقييم الأداء التحدي الأكثر حرجاً في هذه المنهجية، إذ يحذر الباحثان من الاعتقاد بأن الأتمتة تعني الاستغناء عن الجهد البشري؛ ففي التعلم الخاضع للإشراف، يظل الباحث مقيداً بجودة "المعيار الذهبي" الذي وضعه المحللون البشر، مع وجود خطر دائم يتمثل في "الإفرط في التخصيص"، حيث ينجح النموذج في تحليل البيانات التدريبية ويفشل في التعميم على نصوص جديدة. أما في التعلم غير الخاضع للإشراف، فإن الخطر يكمن في ذاتية التأويل، حيث يقع على عاتق الباحث إثبات أن المجموعات التي استخرجها البرنامج بشكل آلي، تمثل مواضيع حقيقية ومنطقية في الواقع السوسيولوجي، وليست مجرد تكتلات إحصائية عشوائية ناتجة عن تكرار الكلمات دون معنى سياقي.

كما تتمثل إحدى الإكراهات المرتبطة بتحليل المحتوى الرقمي في أن العديد من الشركات تعمل باستمرار على تقليص البيانات التي يمكن استخلاصها من منصاتها، وذلك بسبب الالتزامات التنظيمية (مثل اللائحة العامة لحماية البيانات في الاتحاد الأوروبي)، أو بسبب فضائح بعض الشركات في التعامل مع البيانات الخاصة بمستخدميها (مثل فضيحة كامبريدج أناليتيكا وفيسبوك)، وهو ما يترتب عنه إغلاق مسارات بحثية مهمة. وبالتالي فإن قابلية وفرت البيانات للتغييرات التعسفية وغير المتوقعة من قِبل جهات فاعلة خاصة، تجعل هذا النظام غير موثوق به بطبيعته، وقد يكون متحيزاً في العلوم التي ينتجها.

علاوة على أن البيانات الناتجة عن العديد من المنصات الاستهلاكية ليست دائما مناسبة للأغراض البحثية. إذ قد لا يمثل مستخدمو خدمات المنصات عبر الإنترنت عامة السكان، كما قد يكون سلوكهم متحيزاً بطرق غير معروفة. نظراً لكون المنصات لم تصمم للإجابة على الأسئلة البحثية، فقد لا يتم جمع البيانات الأكثر صلة بموضوع البحث (على سبيل المثال، الباحثون المهتمون بنشر المعلومات يحسبون التغريدات المعادة لأن هذا هو ما يتم تسجيله)، أو يمكن جمعها بطريقة تربكها عناصر أخرى من النظام (على سبيل المثال، يتم الخلط بين الاستنتاجات حول تفضيلات المستخدم من خلال تأثير خوارزميات التصنيف والتوصيات الخاصة بالشركة). وقد يتغير تصميم المنصات وميزاتها وتسجيل البيانات وإستراتيجية الوصول إلى البيانات في أي وقت، نظرا لعدم التزام هذه المنصات باتساق والانسجام. وبالإضافة إلى أن البيانات المحيّنة غالبًا ما تكون غير متاحة أمام الباحثين، بسبب المخاوف المتعلقة بالخصوصية والملكية الفكرية، فقد تصبح غير متاحة في المستقبل، مما يعيق إمكانية تكرار النتائج وتكرارها[9].

كما تطرح البيانات الرقمية إشكالا منهجيا حسب ماثيو سالغانيك يتمثل في طبيعة "البيانات الضخمة" بوصفها بياناتغير مصممة لأغراض البحث العلمي. فخلافاً للاستبيانات التقليدية التي يصممها الباحث لقياس متغيرات محددة، فإن المحتويات الرقمية تكون أحيانا ناتجة عن أهداف تجارية أو إدارية، مما يجعلها تعاني من مشكلة وهي تفتقر إلى المتغيرات الديموغرافية والاجتماعية الضرورية للفهم والتفسير السوسيولوجي. وهو ما يطرح مشكلة " التمثيلية ". إذ يؤكد سالغانيك أن مستخدمي المنصات الرقمية لا يمثلون بالضرورة مجتمع الدراسة، مما يضعف من "الصدق الخارجي" للنتائج ويجعل تعميمها مجازفة منهجية ما لم يتم تصحيح الانحيازات عبر أدوات إحصائية متقدمة[10].

ويضيف سالغانيك مشكلا منهجياً آخر يتعلق بـــ التشويش الخوارزميAlgorithmic Confounding؛ حيث يوضح أن السلوك البشري على المنصات الرقمية ليس سلوكاً "طبيعياً" صرفاً، بل هو سلوك مشروط بخوارزميات المنصة (مثل أنظمة التوصية أو ترتيب المنشورات). فعندما يحلل الباحث المحتوى الرقمي، فإنه في الواقع يحلل مزيجاً من الإرادة البشرية والتصميم الخوارزمي، مما يصعّب من عملية عزل المتغيرات وفهم الدوافع الاجتماعية الحقيقية خلف إنتاج هذا المحتوى، ويجعل من البيانات الرقمية "أهدافاً متحركة" تتغير بتغير سياسات المنصات وخوارزمياتها.

أما على الصعيد الأخلاقي، فيحدد سالغانيك أربعة مبادئ أساسية يجب أن تحكم تحليل المحتوى الرقمي: احترام الأشخاص (استقلالية المبحوث والموافقة المستنيرة)، والإحسان (تقليل المخاطر وتعظيم الفوائد)، والعدالة (توزيع أعباء وفوائد البحث بشكل عادل)، واحترام القانون والمصلحة العامة. ويرى أن التحدي الأكبر يكمن في "المخاطر المعلوماتية"، حيث إن تقنيات الذكاء الاصطناعي قادرة على إعادة

كشف هوية الأفراد حتى في البيانات التي يُفترض أنها مجهولة، مما يضع خصوصية الأفراد في مهددة ويفرض على الباحثين مسؤولية أخلاقية تتجاوز مجرد الامتثال للقواعد الأخلاقية والقانونية التقليدية.

وفي سياق الربط بين المنهجي والأخلاقي، يناقش سالغانيك معضلة "التعتيم"، حيث تحتكر الشركات الكبرى البيانات الضخمة، مما يخلق فجوة معرفية بين الباحثين الأكاديميين والبيانات الرقمية الضرورية لمعالجة الإشكاليات العلمية التي يطرحونها. هذا الاحتكار لا يهدد فقط قابلية التكرار الدراسات العلمية للتأكد من صدق الخلاصات النظرية المبنية عليها في البحث العلمي، بل يثير تساؤلات حول ديمقراطية المعرفة ومن يملك الحق في تحليل وفهم السلوك الاجتماعي.

إن هذه التحديات التي تواجه توظيف تقنيات وأدوات الذكاء الاصطناعي في تحليل المحتوى الرقمي تفرض على الباحث في العلوم الاجتماعية الحذر من التحول إلى مجرد "خادم للخوارزميات"، والعمل على تبني مقاربة نقدية تتجاوز الاستخدام التقني والسطحي للذكاء الاصطناعي، نحو فهم "المنطق البرمجي" الذي يحكم هذه الأدوات؛ وهذا لا يعني التحول خبير في البرمجة، بل فقط امتلاك القدرة على فهم كيفية عمل البرامج والأدوات التقنية المعتمدة على المنطق الخوارزمي، وإدراك كيف تعيد تشكيل البيانات والواقع الاجتماعي المدروس، قصد التعامل معها بأفضل طريقة ممكنة. وذلك لأن هذه المعرفة تعتبر ضرورية لدى الباحث لإخضاع المخرجات التي تقدمها هذه الأدوات لمعايير "الصدق العلمي"، بحيث يظل التأويل النظري هو المرجعية النهائية، لا الأرقام المجردة التي تفرزها الآلة. فالبيانات الرقمية ليست مجرد أرقام، بل امتداد لهويات المبحوثين وخصوصياتهم. وبذلك، تصبح الأداة الخوارزمية تحت سلطة الباحث في العلوم الاجتماعية، تعمل كعدسة مكبرة تخدم الفهم العميق للظواهر، مع الالتزام الدائم بالشفافية حول حدود هذه الأداة وتحيزاتها، مما يضمن أن يظل البحث فعلاً معرفياً إنسانياً في مقصده، وحسابياً في وسيلته فقط.

خاتمة:

إن التحولات الناتجة عن الثورة الرقمية لا تعكس مجرد تحول في حجم البيانات فحسب، بل هي دعوة لإعادة صياغة العدة المنهجية التي يتسلح بها الباحثون في العلوم الاجتماعية، حينما يقدمون على دراسة الظواهر الاجتماعية المعاصرة، خاصة تلك المرتبطة بالرقمي. فقد كشفت هذه الورقة أن انتقال المحتوى من صيغته المادية الساكنة إلى "أثر رقمي" يتطلب تجاوز أدوات التحليل الكلاسيكية نحو النمذجة الحسابية المدعومة بالذكاء الاصطناعي. وإن هذا التحول لم يمنح الباحثين قدرة فائقة على معالجة التدفقات المعلوماتية الضخمة فحسب، بل أتاح لهم استكشاف أنماط جديدة من العلاقات بين المتغيرات كلية كانت تتواري خلف كثافة النصوص وتستعصي على الإدراك البشري المعتمد على قدراته العقلية واليدوية البسيطة. إذ من شأن الذكاء الاصطناعي أن يمكّن الباحثين من معالجة كميات هائلة من البيانات الرقمية (سجلات منصات التواصل، سجلات الهواتف، بيانات المعاملات، نصوص، صور...إلخ). وهو ما يفتح إمكانيات لدراسة التفاعلات الاجتماعية على نطاق واسع لم يكن ممكناً في عصر ما قبل ثورة الذكاء الاصطناعي.

كما أن تقنيات التعلم الآلي ونمذجة الشبكات والمحاكاة واسعة النطاق، تسمح باختبار فرضيات معقدة، وباكتشاف أنماط غير مرئية، وإجراء تجارب ميدانية رقمية على مجموعات كبيرة. وهو ما يحوّل العلوم الاجتماعية من اعتماد تقليدي على استبيانات ومقابلات إلى مزيج من الملاحظة التجريبية والتحليل الحسابي. علاوة على أن نماذج الذكاء الاصطناعي قادرة على التنبؤ بانتشار الأفكار أو الأمراض أو اتجاهات ومواقف الأفراد بشكل أسرع وأكثر دقة عند توفر بيانات كافية، ما يجعل النتائج قابلة للتطبيق في السياسات العامة .

بيد أن الوعود التي يقدمها الذكاء الاصطناعي، وخاصة نماذج اللغات الكبيرة (LLMs) وتقنيات معالجة اللغات الطبيعية (NLP)، تظل مشروطة بحذر إبستمولوجي صارم. فقد أظهرت الدراسة أن الأتمتة لا تعني "حياد الآلة"، بل هي عملية محفوفة بالمخاطر التقنية والمنهجية؛ بدءاً من مترلقات المعالجة المسبقة التي قد تجتث السياق من جذوره، وصولاً إلى معضلة "الصندوق الأسود" وغياب التفسير النظري للمخرجات التنبؤية. فالأخطاء المنهجية التي يؤدي إليها استعمال النماذج اللغوية الكبرى في تحليل المحتوى الرقمي التحديات، كما هو الحال بالنسبة للتمييز بين المناهج المعتمدة في الدراسات والأبحاث كما أشارت الدراسة المحال عليها في المحور الأول من هذه الورقة على سبيل المثال، تؤكد أن الذكاء الاصطناعي يظل "شريكاً تقنياً" يحتاج دوماً إلى توجيه صارم من قبل للباحث.

فضلا عن أن البيانات الرقمية لا تمثل مجتمع الدراسة بالضرورة؛ فخوارزميات المنصات قد تُضخّم تحيّزات أو تُغيّب فئات معينة، ما يهدد صلاحية الاستنتاجات إذا لم تُصحّح منهجيا. كما أن استخدام بيانات حساسة يثير مخاطر انتهاك الخصوصية، ويستدعي أطر أخلاقية وقانونية جديدة لحماية الأفراد وضمان ثقة الجمهور. المقال يؤكد ضرورة تحسين مشاركة البيانات وأخلاقيات البحث. فضلا عن أن الجامعات والمؤسسات البحثية غالباً غير مجهزة تقنياً أو تنظيمياً للتعامل مع البيانات الضخمة أو التعاون مع القطاع الخاص؛ هناك حاجة لإعادة تنظيم المناهج وبناء بنى تحتية مشتركة وآليات تعاون واضحة.

وبالتالي، فأن مستقبل تحليل المحتوى الرقمي في العلوم الاجتماعية لا يكمن في الاختيار بين اليدوي والحسابي، بل في تبني مقاربة هجينة تزاوج بين دقة الآلة في النمذجة وعمق الإنسان في التأويل. إن الحفاظ على العلمية في عصر الخوارزميات يتطلب إرساء بروتوكولات صارمة للتحقق من أداء البرامج والنماذج اللغوية في تحليل المحتوى، والالتزام بشفافية منهجية تجعل من الذكاء الاصطناعي أداة لتعزيز "الخيال السوسيولوجي" لا لاستبداله. فالهدف النهائي ليس اختصار الوقت والتكلفة، بل الوصول إلى فهم أعمق وأكثر شمولاً لتعقيدات المجتمع الرقمي المعاصر.

الهوامش:

[1] Cioffi-Revilla, C. (2010). Computational social science. Wiley Interdisciplinary Reviews: Computational Statistics, 2(3), p 260.

[2] Krippendorff, K. (2019). Content Analysis: An Introduction to Its Methodology. Content Analysis: An Introduction to Its Methodology, p 34.

[3] Kozlowski, A. C., Taddy, M., & Evans, J. A. (2018). The Geometry of Culture: Analyzing Meaning through Word Embeddings. American Sociological Review, 84(5), p 905.

[4] Hase, V. (2023). Automated Content Analysis. Standardized Content Analysis in Communication Research, p 24.

[5] Wilkerson, J., & Casas, A. (2017). Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges. Annual Review of Political Science, V. 20, p 530.

[6] Domínguez-Diaz, A., Goyanes, M., & de-Marcos, L. (2025). Automating content analysis of scientific abstracts using ChatGPT: A methodological protocol and use case. MethodsX, 15.

[7] Kozlowski, A. C., Taddy, M., & Evans, J. A. (2018). The Geometry of Culture: Analyzing Meaning through Word Embeddings. American Sociological Review, 84(5), 905–949.

[8] Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis, 21(3).

[9] Lazer, D. M. J., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational social science: Obstacles and opportunities. Science (New York, N.Y.), 369(6507), p 1060.

[10] Salganik, M. J. . (2017). Bit by bit: Social research in the digital age. Princeton University Press.