

La programación de la historia: ciencia de datos para el análisis de fuentes históricas

Programming History: Data Science for the Analysis of Historical Sources

Albert Horacio Mestre Díaz¹
Universidad de la República, Uruguay
<https://orcid.org/0009-0009-6181-6824>

DOI: <https://doi.org/10.25032/crh.v12i22.2793>

Enviado: 31/1/2026

Aceptado: 22/6/2026

Resumen: Este artículo analiza la incorporación de técnicas de ciencia de datos al trabajo con fuentes históricas y sus efectos epistemológicos, metodológicos y materiales en la historiografía contemporánea. Argumenta que estas técnicas — que constituyen la base operativa de la inteligencia artificial— no son meros instrumentos neutrales, sino arbitrajes técnicos que reconfiguran el acceso, procesamiento e interpretación del pasado. Muestra que la digitalización amplía la escala del análisis, pero introduce límites vinculados a la incompletitud del registro histórico, la opacidad algorítmica y las infraestructuras desiguales. Sostiene que la abundancia de datos no garantiza mayor verdad histórica y que los modelos computacionales pueden reproducir sesgos técnicos y políticos. Propone una colaboración crítica e interdisciplinaria entre historiadores y

¹ **Albert Mestre Díaz.** Licenciado en Historia por la Universidad de la República (Uruguay) y estudiante del Profesorado de Historia en el Instituto de Profesores Artigas (IPA). Su trabajo se sitúa en la intersección entre la historia de la educación, la filosofía de la educación, la alfabetización digital y las humanidades digitales. Complementa su formación humanística con estudios en Ciencia de Datos y Machine Learning. Sus líneas de investigación exploran las posibilidades de la inteligencia artificial, el análisis computacional y las tecnologías digitales aplicadas a la investigación histórica, la educación y el patrimonio documental. Actualmente proyecta la realización de estudios de posgrado en Ciencia de Datos, con el objetivo de profundizar en el desarrollo de metodologías computacionales aplicadas a la investigación histórica y educativa

científicos de datos y concluye que el giro computacional exige una historiografía reflexiva, situada y consciente de sus mediaciones técnicas.

Palabras clave: ciencia de datos, historia, epistemología, minería de datos.

Abstract: This article examines the incorporation of Data Science techniques into historical source analysis and their epistemological, methodological, and material implications for contemporary historiography. It argues that these techniques –fundamental to the operation of Artificial Intelligence– are not merely neutral instruments but technical mediations that reshape access to, processing of, and interpretation of the past. It shows that digitalization expands analytical scale while introducing limits related to the incompleteness of the historical record, algorithmic opacity, and unequal infrastructures. The article contends that data abundance does not ensure greater historical truth and that computational models reproduce technical and political biases. It advocates for critical interdisciplinary collaboration between historians and data scientists and concludes that the computational turn requires a reflexive, situated historiography aware of its technical mediations.

Keywords: data science, historiography, epistemology, data mining.

1. Introducción

En el último quinquenio, el desarrollo de las novedosas técnicas de abordaje de fuentes se ha vuelto perceptible en múltiples ámbitos de la vida académica, especialmente en las universidades y, en particular, a aquellas vinculadas a las ciencias sociales. Este proceso ha introducido tensiones significativas en la academia, ya que estas tecnologías no solo amplían las capacidades de análisis, sino que cuestionan métodos, formas de producción de conocimiento y prácticas consolidadas.

En el campo de la historia, la incorporación de herramientas computacionales aplicadas a fuentes documentales marca un punto de inflexión respecto de cómo se investiga, se interpreta y se construye el conocimiento histórico. Más que evaluar la inteligencia artificial (IA), este artículo analiza las

condiciones materiales y técnicas que permiten su funcionamiento en el trabajo con fuentes históricas.

En este contexto, este trabajo se propone responder la siguiente pregunta: ¿cómo reconfiguran las herramientas digitales aplicadas a fuentes documentales las prácticas historiográficas contemporáneas, y cuáles son los límites epistemológicos, metodológicos y materiales de esta transformación?

El problema que plantean estas tecnologías no se reduce únicamente a la desigualdad en el acceso según los contextos socioeconómicos de los investigadores, sino que remite también a cuestiones técnicas y epistémicas que exigen niveles crecientes de especialización. Instrumentos como el Reconocimiento Óptico de Caracteres (OCR), el Reconocimiento de Texto Manuscrito (HTR), el Procesamiento del Lenguaje Natural (PLN), el *Machine Learning* (ML) y la Minería de Datos (MD) con frecuencia operan como sistemas opacos: los criterios mediante los cuales procesan y reorganizan la información no siempre son transparentes para el historiador, configurándose como verdaderas cajas negras algorítmicas. Este artículo aborda críticamente dichos insumos técnicos, que se constituyen como bases fundamentales de la IA y vitales para el análisis de fuentes.

En este escenario, la disciplina histórica y la academia en general se ven interpeladas por la necesidad de reflexionar de manera crítica sobre los caminos que abren las novedosas herramientas. La innovación tecnológica ha dado lugar a nuevas perspectivas que desafían los cánones historiográficos, cuestión que ha sido abordada por investigadores como Juan Andrés Bresciano (2000, 2014), Matilde Eiroa (2018) e Ian Milligan (2019, 2020). Sin embargo, este artículo no se limita a revisar dichos aportes: propone analizar específicamente los mecanismos operativos utilizados en la ciencia de datos y su impacto sobre la práctica historiográfica atendiendo tanto a sus potencialidades como a sus límites.

El advenimiento de una época de abundancia en información ha provocado una mutación en la disciplina histórica. Dentro de este contexto, el historiador no solo se ve impulsado a colaborar con especialistas de otras

disciplinas, sino también a familiarizarse con recursos metodológicos de otras áreas. Con la llegada de una era en la que se crea información masivamente, se plantea la necesidad de incluir en el currículo educativo el pensamiento computacional necesario para analizar realidad en constante cambio. Esta es la magnitud con la que los investigadores tendrán que lidiar, por lo que es necesario preparar los fundamentos teóricos y éticos por aquellos que vendrán después.

Todo ello configura un escenario profundamente ambiguo que este trabajo plantea problematizar: ¿cómo investigar y escribir historia en un contexto en el que los datos, las infraestructuras y los algoritmos median el acceso a las fuentes? y, ¿qué transformaciones resultan productivas desde el punto de vista metodológico?, ¿cuáles introducen nuevas controversias y cuáles generan riesgos epistemológicos al incorporar herramientas digitales poco transparentes y difíciles de verificar al trabajo historiográfico?

Por lo tanto, resulta relevante distinguir la creciente articulación entre la informática y la disciplina histórica, así como las consecuencias de no asumir una posición explícita frente a este proceso. La indefinición de las universidades dedicadas a las ciencias sociales y humanas podría derivar, a futuro, en una situación desfavorable si no se adoptan posturas claras en el presente.

En un contexto caracterizado por el desarrollo de hardware cada vez más potente y por la cooperación entre centros de investigación a escala global, la tecnología se ha convertido en un componente central para la producción, gestión y validación del conocimiento. Dentro de este marco, las instituciones formadoras se enfrentan al desafío de posicionarse activamente como agentes del cambio, y no solo como usuarios de dispositivos cuyo funcionamiento, alcances y límites no controlan plenamente.

2. La computación y la investigación histórica

2.1. Mediación técnica y reorganización del conocimiento

El progreso de la telemática modificó ampliamente la investigación científica. También introdujo nuevas tensiones epistemológicas y metodológicas al cambiar cómo se producía, almacenaba y problematizaba la creciente cantidad

de información. A mediados del siglo XX aparecen algunos hitos importantes que permiten pensar la mediación técnica —mediante la implementación de tecnologías— en la reorganización del conocimiento y la externalización de la memoria.

El historiador e ingeniero norteamericano Vannevar Bush, con su ensayo *As we may think* (1945), así como el sacerdote italiano Roberto Busa con su proyecto *Index Thomisticus* (1946), abrieron la posibilidad del uso de sistemas electromecánicos en las ciencias humanas. Bush propuso la utilización de un escritorio mecatrónico, destinado a guardar información y acceder eficazmente a ella en una instancia posterior. Busa, en cambio, ideó un proyecto cuyos resultados no se darían a conocer hasta —casi— treinta años más tarde: se indexaron la totalidad de las palabras contenidas en las obras de Tomás de Aquino, que totalizaron más de once millones de registros.

2.2. Informática, cuantificación y reacción

Durante las décadas de 1960 y 1970 se comenzó a asociar la informática con la historia y se manifestaron disyuntivas entre el optimismo tecnológico y las prácticas. Entre algunos de los casos más representativos, los historiadores Emmanuel Le Roy Ladurie, Robert Fogel y Stanley Engerman llevaron a cabo intervenciones teóricas y prácticas en esta área del conocimiento. Le Roy Ladurie fue conocido por su marcada perspectiva tecno-optimista donde «el historiador del mañana será programador o no será nada», lo que manifiesta el advenimiento de una reconfiguración de la generación del conocimiento causada por la tecnología (Pons 50).

Por otro lado, Fogel y Engerman aplicaron métodos cuantitativos y estadísticos de manera intensiva en su proyecto *Time on the Cross* (1974), con lo que promovieron una ambiciosa —y polémica— renovación cliométrica basada en el procesamiento de grandes volúmenes de datos. Sin embargo, el impacto generado por este proyecto produjo un efecto de rebote que contribuyó a frenar el avance de la cuantificación histórica y a reforzar la desconfianza hacia los ordenadores y su papel en las humanidades (Pons 56).

El historiador español Anacleto Pons (99-102), quien hace un repaso sobre la historia de la computación y su influencia en la historia, plantea que a finales del siglo XX el exceso del determinismo cuantitativo provocó una fuerte reacción neoludita, en el caso de historiadores como David Noble o Gertrude Himmelfarb, y escéptica como con Lawrence Stone. Himmelfarb, por ejemplo, comentaba que la computación traía consigo preocupaciones: se corre el peligro, dice la autora estadounidense, de olvidar la prioridad de entender la herramienta primero antes de usarla.

En este contexto de transición, donde las herramientas del investigador histórico han mutado y ocasionado tensiones en el mundo académico, Bresciano (2015) comenta que el desarrollo de una economía en red conlleva una profunda transformación en los métodos de producción del conocimiento (18). Es en este nuevo umbral donde se sitúa el problema que abordará las siguientes secciones, dedicadas a analizar el alcance, los límites, las implicaciones epistemológicas y metodológicas de estos insumos.

3. Herramientas para el análisis documental

En las secciones posteriores se abordarán diversos recursos computacionales para el análisis de fuentes históricas que suelen presentarse bajo el paraguas de la denominada IA. Sin embargo, este trabajo no se centrará en ella como objeto de examen, sino en las partes técnicas que la hacen posible y que intervienen en el trabajo con fuentes documentales. Esta elección responde a dos razones principales.

En primer lugar, problematizar la IA como tal en relación con la historia exigiría necesariamente explicar en detalle su arquitectura técnica, sus métodos de entrenamiento y sus límites epistemológicos y éticos. Esto implicaría analizar plataformas comerciales cuyos modelos son opacos, privativos y basados en la apropiación de datos de terceros, además de discutir las implicancias políticas de entregar información histórica a corporaciones privadas.

Abordar este problema obligaría a recorrer todas las herramientas que se desarrollarán en las secciones, pero sin la profundidad analítica necesaria para evaluar sus riesgos heurísticos y metodológicos específicos. Por ello, se optó por

invertir el enfoque: en lugar de partir de la IA como abstracción general, se analizan primero sus componentes técnicos fundamentales, con el objetivo de iluminar tanto su funcionamiento como sus efectos sobre la práctica historiográfica.

En segundo lugar, se sostiene que lo que realmente transforma el trabajo con fuentes históricas no es solo la IA, sino las técnicas concretas de mediación digital que emplea y que reconfiguran la heurística, la escala y los criterios de interpretación del historiador. Esto no implica negar la utilidad como asistente de escritura, búsqueda o síntesis documental; sin embargo, dichas aplicaciones exceden los objetivos de este artículo, cuyo foco se sitúa en las herramientas que modifican materialmente la relación entre el especialista histórico y la fuente.

3.1. Digitalización y mediación de la fuente

La incorporación de recursos informáticos al trabajo historiográfico dio lugar a la consolidación de un conjunto de enfoques metodológicos y técnicos que la transformaron. Más allá de los debates generales en torno a sus implicancias, la continua diversificación de los archivos con la popularización del World Wide Web (WWW) complejizó las disciplinas existentes y creó ciencias auxiliares nuevas. El enfoque, por tanto, ya no solamente era aprender sobre el objeto de estudio, sino aprender a cuestionar las metodologías que transformaron su oficio.

Dos de las herramientas más conocidas por la simplicidad en su labor es el OCR y el HTR: consisten en dos técnicas con fundamentos similares, pero con alcances y limitaciones distintas que reconfiguran la heurística y la naturaleza misma de la fuente documental tradicional (Ballesteros Estrada *et al.* 27-28). Ambas traducen la realidad material del documento a bits interpretables por una computadora, luego identifica dónde se encuentra el texto y el orden de las oraciones, donde posteriormente las homogeneiza y reconoce mediante una comparación de los caracteres con un servidor de datos.

3.2. Machine Learning y dependencia técnica

El OCR resulta deficiente de no darse las condiciones apropiadas: texto legible, oraciones alineadas horizontalmente, tipografía regular y soporte de

papel en buenas condiciones (Graham *et al.* 54). Es por ello que se ha implementado ML, que consiste en aplicar un conjunto de métodos que permiten a las computadoras aprender patrones a partir de grandes volúmenes de datos.² Desde una perspectiva historiográfica, esta transformación modifica la relación del investigador con las fuentes. Documentos que antes requerían una lectura individual pueden incorporarse a un análisis sistemático a gran escala, lo que permite ampliar las preguntas que es posible formularles.

La necesidad de reconocer textos manuscritos llevó a la creación del HTR: una imbricación entre ML, interpretación humana y decisiones heurísticas. Este método se diferencia del OCR al introducir un cambio cualitativo en su funcionamiento: se apoya en modelos de aprendizaje automático entrenados sobre conjuntos específicos de datos. En otras palabras, se utilizan insumos analíticos sobre caracteres escritos en distintos soportes y calidades con la finalidad de disminuir el coeficiente de error del resultado. Iniciativas como el proyecto *tranScriptorium*, llevada a cabo por la Universitat Politècnica de Valencia, revelaron los avances y limitaciones de esta tecnología: por un lado, la necesidad de un personal especializado para la creación de los algoritmos y, por otro, el HTR es dependiente del lenguaje por lo que los resultados son variables.

El desarrollo de estas estructuras escapa a la formación tradicional de investigadores sociales: hay una necesaria transdisciplinariedad que refuerza las lógicas de cooperación con especialistas en matemáticas e informática. Estos cambios de paradigma no se limitan a los estudios históricos de la contemporaneidad, sino que inciden también en la manera de leer y problematizar procesos históricos de distintas temporalidades. Desde esta perspectiva, la generación de conocimiento histórico se transforma en una doble dimensión: no solo en cómo se produce el conocimiento, sino también en cómo se valida y qué tan trazable es.

² Para ilustrar mejor su funcionamiento conviene hacer una aclaración previa: el ML no aprende en un sentido cognitivo, sino que ajusta modelos a partir de operaciones estadísticas y matemáticas aplicadas sobre grandes volúmenes de información. Análogamente al trabajo del investigador histórico que, tras la lectura de grandes volúmenes documentales, identifica regularidades en las formas de escritura, los temas que se repiten o cambios en el vocabulario.

4. Heterogeneidad documental y límites epistemológicos

4.1. Heterogeneidad y lingüística de las fuentes

La heterogeneidad material, lingüística y de contexto de las fuentes históricas no constituye un obstáculo accidental para la aplicación de herramientas computacionales, sino un rasgo del trabajo historiográfico que obliga a repensar en forma crítica los supuestos de regularidad, estabilidad y neutralidad sobre los que estas tecnologías suelen apoyarse. Lejos de reducirse a un inconveniente rigurosamente técnico, esta heterogeneidad es inseparable de los procesos socioculturales y políticos que estipulan tanto la producción de los registros como su posterior exégesis historiográfica.

Las fuentes históricas pueden entenderse como rastros materiales o simbólicos de la actividad humana que permiten el análisis diacrónico del pasado (Bresciano, *La documentación electrónica* 285). Estos rastros no poseen un significado intrínseco: adquieren su condición de fuente en el marco de una operación intelectual a través de un investigador que los interroga desde problemas, categorías y perspectivas situadas históricamente. En este sentido, la fuente no es un dato neutro, sino una construcción historiográfica que emerge de la relación entre el investigador, el objeto y el contexto de estudio.

Esta condición construida de la fuente se expresa con una marcada diversidad de soportes y modalidades de transmisión. Asimismo, pueden pertenecer a ámbitos públicos o privados, encontrarse digitalizadas o haber sido producidas originalmente en entornos digitales. Esta pluralidad material y lingüística, que constituye una riqueza heurística para la disciplina histórica, introduce al mismo tiempo conflictos significativos cuando se los intenta someter a procesos de análisis masivo.³

³ La normalización es un proceso técnico y metodológico que estandariza, es decir, homogeniza datos heterogéneos para hacerlos computables. En el oficio del historiador, este procedimiento transforma la irregularidad presente en los documentos históricos en una estructura de datos que permita su análisis. Desde el punto de vista epistemológico, Leonardo Barleta (2015) comenta que una normalización puede resultar en una excesiva formalización que obliga a registros históricos y caóticos a encajar en modelos universales.

Digitalizar documentos no constituye un duplicado exacto de la realidad, es más bien una transformación que podría conllevar la pérdida del aspecto material de la fuente. Una desventaja notoria es que, al carecer de contacto directo con ellas, se dificulta la posibilidad de encontrar cuestiones que podrían ser notorias de ser consultadas en el archivo o la biblioteca: anotaciones marginales que fueron excluidas por el algoritmo, desgaste, tipo de tinta —ferrosa, gel, carbón, etcétera—. Este proceso de transformación en bits resuelve cuestiones de conservación y consulta: millones de copias pueden ser resguardadas en un almacenamiento digital y pueden ser consultadas por todo el mundo al mismo tiempo, sin reservar una cita uno a la vez.

La transformación de un documento tridimensional en uno bidimensional conlleva desventajas que a pesar de ser claras no son permanentes y podrían desaparecer de manera parcial o por completo con el perfeccionamiento de la tecnología. Estas desventajas son varias: (a) durante el proceso de digitalización la imagen puede adquirir ruido visual; (b) detalles físicos pueden perderse; (c) la arquitectura de un archivo tradicional es insuficiente para alojar lo necesario para almacenar digitalmente las fuentes, y (d) se pueden perder, borrar o infectar los documentos digitales.

Estos aspectos son ampliados por la disimilitud entre las carreras en ciencias sociales o humanidades, entre aquellas que enseñan sobre innovación tecnológica o las que la presentan de manera tácita. Esto evoca un riesgo cuando se utilicen técnicas propias de la informática, lo que resulta en un peligro metodológico para la investigación. Pensar la estructura, construir y presentar una base de datos es un asunto complejo, más aún si es un repositorio en constante crecimiento: un riesgo podría ser el de un investigador que fuerce la información histórica para encajar en categorías rígidas y produzca, como resultado, una estandarización artificial que puede comprometer la imparcialidad de los resultados (Barleta 126).

Herramientas como el OCR/HTR crean una ilusión de completitud, porque son capaces de leer un libro en segundos. No solo eso, sino que con órdenes más específicas y añadiendo más procesos algorítmicos se puede solicitar esquemas, resúmenes de audio, tablillas de datos, etcétera. Habría que hacer una

aclaración: el OCR y el HTR no interpelan los documentos con los que trabajan, únicamente reconocen sus caracteres y su efectividad es acorde a cómo, para qué y dónde fueron creados. Sería contraproducente emplear un software de reconocimiento de caracteres que está configurado para tipografías cursivas de la Edad Media en Europa y no para fuentes digitales en japonés.

4.2. Procesamiento del Lenguaje Natural, minería de datos y análisis de redes

El siguiente paso en el tratamiento de fuentes digitales es la aplicación de técnicas de procesamiento del lenguaje. Esta etapa resulta fundamental pues las computadoras no comprenden el lenguaje humano de manera inmediata. La dificultad es aún mayor en el caso de fuentes históricas, caracterizadas por ambigüedades, cambios lingüísticos, irregularidades y contextos diversos. En consecuencia, el lenguaje natural —es decir, el idioma tal y como se presenta en los documentos históricos— plantea desafíos que requieren de la conformación de equipos de ingenieros para ocuparse de lo técnico y de investigadores sociales para afinar la exactitud de los resultados.

Para sortear este obstáculo, el PLN identifica regularidades en los textos. Estas regularidades no poseen un carácter universal, sino que se encuentran condicionadas por contextos históricos y lingüísticos concretos. De esta manera, emerge un problema epistemológico importante: los modelos algorítmicos aprenden de datos, pero, ¿de dónde son estos datos? Estos suelen reflejar lenguas normalizadas, períodos o expresiones semánticas recientes y contextos específicos.

En consecuencia, el examen mediante algoritmos puede proyectar sobre las fuentes históricas patrones ajenos a su tiempo y favorecer, así, lecturas del pasado construidas con regularidades del presente. Para crear algoritmos de reconocimiento y comprensión de texto los insumos proporcionados son de contextos específicos o en un solo lenguaje, como el idioma inglés. El anacronismo algorítmico es un desafío sustancial, pero esperado, producto de la adopción de nuevas tecnologías en la historia. Para el académico, esto implica que la interpretación de las fuentes deja de depender exclusivamente de la lectura

directa y comienza a estar mediada por modelos que clasifican y jerarquizan información según criterios informáticos.

La opacidad técnica es un elemento no menor que Anaclet Pons identifica como una caja negra.⁴ Esta falta de inteligibilidad técnica plantea desafíos para una disciplina que tradicionalmente ha valorado la explicitación de las fuentes, los procedimientos de obtención y exégesis, y los criterios de interpretación utilizados. Esta caja negra algorítmica desplaza la autoridad del historiador sobre decisiones de diseño e implementación que a veces no son neutrales y que a menudo responden a lógicas comerciales o de eficiencia computacional.

Al delegar la organización y jerarquización de los datos en estos sistemas, el investigador incurre en un riesgo porque los resultados que devuelven los algoritmos son en realidad regularidades estadísticas. Este fenómeno no es insondable, pero sí muy costoso de solucionar y, hasta el momento de esta publicación, la mejor solución sigue siendo la implementación de equipos locales para afrontar desafíos específicos.

Las limitaciones de herramientas comerciales resultan evidentes en proyectos sensibles que requieren altos estándares de privacidad y gobernanza digital. Un ejemplo de ello es Nuestra Memoria, un proyecto paradigmático desarrollado en Chile para el tratamiento de fuentes históricas vinculada a la última dictadura (1973-1990). Mediante la combinación de técnicas de OCR, HTR y transcripción manual, la iniciativa logró digitalizar y procesar más de cuarenta mil páginas de expedientes judiciales. Sobre esta base, se desarrollaron modelos de lenguaje especializado que permitieron analizar corpus documentales a gran escala y preservaron al mismo tiempo la privacidad del conocimiento.

La magnitud del proyecto se refleja en su capacidad para procesar y articular múltiples tipos de fuentes históricas. Hasta el momento, se analizaron

⁴ La caja negra se define como un sistema o dispositivo cuyas partes internas encargadas de la toma de decisiones son opacas para el usuario: se conoce el insumo que ingresa y el resultado, pero no la lógica interna que vincula a ambos. Anaclet Pons menciona que existe una *algocracia*, un supuesto epistemológico que considera que los algoritmos son inmunes al prejuicio. El sistema de PageRank de Google posee un sesgo porque muestra aquellos actores o páginas web más visibles. En una investigación histórica, esto significaría sobredimensionar la importancia de ciertas fuentes sobre las demás.

más de 1.600.000 fotogramas de películas, colecciones fotográficas, recortes de prensa y miles de expedientes judiciales. A ello se suma otro insumo, y es el desarrollo de bases de datos especializadas para la identificación y seguimiento de víctimas presentes en las fuentes. Recientemente se incorporó al proyecto un corpus de más de 500 horas de registros sonoros históricos con la aplicación de técnicas de reconocimiento de voz e identificación, lo que permite hacer consultas avanzadas sobre personas, situaciones puntuales o relaciones.

Proyectos encomiables como este resultan particularmente problemáticos para la historiografía tradicional, dado que la ingeniería informática no forma parte del currículo formativo del profesional en historia. Es por esto, que la mejor solución encontrada al día de hoy ha sido la conformación de equipos formados en distintas disciplinas. Surge entonces una cuestión clave: ¿cómo validar el entrenamiento de estos modelos y detectar sesgos en herramientas cuyo funcionamiento interno permanece, en gran medida, opaco?⁵ Tal y como señala Bresciano, estos problemas no se limitan a lo técnico, sino que inciden en forma directa en la manera en que se produce y valida el conocimiento histórico (*La historiografía digital* 22-23).

La metodología es acertadamente cuestionada porque en la medida que analizamos procesos históricos que incluyen registros digitales masivos estos mecanismos se vuelven indispensables. La desmaterialización de las fuentes introduce nuevas tensiones: la trazabilidad y verificabilidad de los documentos se vuelve más compleja. La mediación algorítmica entre el investigador y la fuente puede generar una ilusión de transparencia que oculta decisiones técnicas o teóricas previas.

En este sentido, proyectos como Nuestra Memoria implementan en su metodología la sinergia entre humano y artefacto. Para satisfacer sus objetivos,

⁵ Desde una perspectiva historiográfica y epistemológica, el sesgo no debe entenderse únicamente como un error técnico, sino como una mediación que incorpora juicios humanos, valores sociales y relaciones de poder en los procedimientos digitales. Este fenómeno desafía la supuesta neutralidad de la tecnología revelando que tecnologías como la IA conllevan decisiones de diseño que a menudo son opacas y no responden a un método científico. Los sesgos se pueden manifestar, por ejemplo, en la selección de datos de entrenamiento o en los criterios de evaluación de relevancia al analizar información. Evaluar si es un sesgo peligroso para la objetividad de una investigación depende exclusivamente del juicio humano (Abelson *et al.* 145).

mantener la humanización de los datos e imponer un estándar de precisión, se implementan dos conceptos altamente técnicos: *Human in the Loop* (HITL) y el sistema H-LAM/T.⁶ Ambos marcos proponen que la tecnología no debe sustituir la interpretación, sino actuar como una ayuda al investigador manteniendo bajo control la ética y la metodología.

En otras palabras, la IA o las herramientas que la componen son asistentes y no autores de los resultados: hay una fuerte mitigación de la opacidad algorítmica ya que cada respuesta es trazable y permite que el investigador verifique la respuesta. Estos sistemas son aún más potenciados en contextos de cooperación entre distintas disciplinas, donde el HITL no solamente incluye al programador, sino a equipos que aportan a la hermenéutica.

Las metodologías H-LAM/T y HITL transforman la relación entre el cientista social y su herramienta desplazando al ordenador de su papel tradicional como herramienta de cálculo hacia un asistente para el investigador. Lo que, es más, requiere una supervisión hermenéutica constante que permita contextualizar y corregir las inferencias que pueden producir los modelos: la lectura crítica de las fuentes permanece como una tarea irreductiblemente humana.

Volviendo al procesamiento computacional, esta tiene una secuencia de operaciones que suceden en orden: digitalización → OCR/HTR/Audio a texto → Minería de Datos/PLN → Grafo de conocimiento y análisis de redes → Concreción y validación (HITL) → Preguntas y respuestas. En este recorrido, la MD constituye un punto importante ya que no se emplea en solo una etapa, sino que puede estar presente en varias.

⁶ Desde un punto de vista de la investigación histórica, los sistemas HITL permiten mantener la intervención humana en etapas críticas del análisis. En lugar de delegar por completo la interpretación a modelos computacionales, el historiador participa activamente en la validación y contextualización de los resultados. Esta característica es indispensable para la disciplina histórica donde la ambigüedad de las fuentes dificulta la aplicación de procedimientos automatizados. En cambio, desde una perspectiva metodológica, los sistemas H-LAM/T transforman la interacción entre el historiador y el archivo digital. Mientras que las consultas tradicionales requieren búsquedas predeterminadas y categorías definidas manualmente, estos modelos permiten formular preguntas y explorar conexiones documentales de manera fluida. Esto aproxima al investigador a ser más creativo con la interrogación documental.

Desde una perspectiva historiográfica, la minería de datos amplía la escala en la que pueden formularse hipótesis. Mientras que la lectura tradicional puede centrarse en documentos específicos o relativamente acotados, estas técnicas permiten examinar miles de registros en simultáneo y detectar tendencias difíciles de advertir mediante métodos convencionales. Esta ampliación, sin embargo, reabre la tensión entre la identificación de patrones globales y la comprensión de experiencias históricas particulares.

El uso de MD no es por sí mismo una automatización, sino un mapeo de los medios analíticos tradicionales resultando en la creación de *meta-medios* (Rota 4). Este giro computacional obliga al historiador a transitar de una cultura de la escasez, a una de la abundancia informativa donde el desafío más grande no es encontrar información, sino procesar volúmenes masivos de conocimiento.

Los beneficios que introduce la MD como herramienta en el análisis histórico son múltiples y operan sobre distintos tipos de información:

- I. Permite el análisis sistemático de metadatos —fechas, enlaces, procedencia— que facilitan la reconstrucción de trayectorias documentales y cronologías dispersas.
- II. El procesamiento automatizado de imágenes posibilita la extracción de características visuales a partir de grandes conjuntos de archivos, que habilita la detección de patrones en culturas iconográficas específicas. En el caso de Nuestra Memoria los algoritmos pueden identificar la situación y los ítems en una imagen con gran precisión: identificar a militares, autos o animales e inclusive qué sucede en la imagen.
- III. El procesamiento masivo de texto permite analizar en lapsos reducidos millones de palabras provenientes de corpus documentales heterogéneos, mientras que la integración el PLN junto con la MD posibilitan la identificación de temáticas latentes o recurrencias léxicas.
- IV. Prácticas como la *lectura distante* permiten abordar grandes volúmenes de información sin recurrir a la lectura humana individual de cada documento; así, se desplaza el foco desde el

contenido singular hacia las regularidades y relaciones que emergen del conjunto colectivo de documentos.

Este desplazamiento desde el documento individual hacia el conjunto de datos abre el camino a enfoques que no solo buscan identificar patrones, sino que encuentran vínculos estructurales entre actores, conceptos y eventos. Los insumos algorítmicos aquí descritos han sido el resultado de décadas de desarrollo e implementación y también es necesario entenderlas para plantear el abordaje del análisis de redes.

Este recurso digital constituye una metodología orientada a representar y examinar relaciones entre actores, instituciones, eventos y fenómenos históricos. Este enfoque parte del supuesto de que estas relaciones pueden modelarse como estructuras compuestas por nodos y enlaces (Graham *et al.* 194). Los nodos representan entidades definidas por el investigador, mientras los enlaces son los vínculos que las conectan permitiendo comprender procesos históricos como configuraciones en un campo más grande antes que como hechos aislados.

En cuanto al historiador, su rol puede desplazarse hacia un nivel elevado de abstracción debido a la magnitud de los registros analizados. Shawn Graham *et al.* denominan a este riesgo la falacia del «bosque y los árboles», es decir, a la posibilidad de perder de vista la singularidad de los individuos y de los acontecimientos al privilegiar patrones globales. Esta forma de representación desplaza parcialmente la atención desde los individuos hacia estructuras relacionadas y reabre debates sobre la tensión entre los enfoques centrados en actores y aquellos orientados a procesos más amplios.

En este sentido, la complejidad técnica no sustituye el análisis crítico de las fuentes: una red no puede condensar las ambigüedades, los conflictos interpretativos, las experiencias vividas por los actores ni el conocimiento historiográfico acumulado sobre un campo de estudio. Del mismo modo, estas herramientas de indagación deben someterse a las mismas preguntas críticas que cualquier otro método historiográfico: ¿qué entendemos como nodo o como enlace?, ¿por qué se seleccionan esas relaciones y no otras?, ¿qué supuestos teóricos y contextuales orientan dichas decisiones?, ¿qué sacrificamos o ignoramos con los nodos y enlaces seleccionados?

Especial atención debería ser dedicada hoy en día a no caer en supuestos inamovibles. Es necesaria la continua actitud crítica a las herramientas, y a nuestras propias investigaciones, para no replicar de manera inmoral resultados propiciados por un código. Desde una perspectiva técnica y ética, estas limitaciones sugieren que el análisis de redes debería utilizarse como un elemento que ayude al investigador.

Su potencial historiográfico depende de su integración crítica con otros métodos, de la transparencia de sus supuestos y del investigador para contextualizar e interpretar los resultados producidos algorítmicamente. Como resultado, la exploración masiva no sustituye la comprensión historiográfica, sino que funciona como una herramienta exploratoria preliminar o como ayudante, cuyo valor dependerá de la verificabilidad de los resultados.

5. ¿Pueden pensar los datos? Ciencia de datos y la historia

5.1. Ciencia de datos y la disciplina histórica: ensamblaje técnico y epistemológico

¿Qué es la ciencia de datos (CD)? La CD no produce conocimiento por sí misma, sino que reconfigura las condiciones técnicas bajo las cuales este puede ser generado, organizado y analizado.⁷ Esta área utiliza todas las tecnologías digitales antes mencionadas, e inclusive más. Se consolida como una disciplina académica en la última década y su emergencia está estrechamente vinculada al crecimiento exponencial de la capacidad de almacenamiento, procesamiento y circulación de información digital. Es un área heterogénea e interdisciplinaria que reordena jerarquías de saber y se aplica a ámbitos tan diversos como la medicina o la administración. Desde la perspectiva historiográfica, esta transformación tiene efectos directos sobre la práctica investigativa.

Más que una especialización, puede entenderse como un modo de producción de conocimiento. No constituye un paradigma cerrado, sino un

⁷ Hoy en día, los datos se generan de manera constante y ubicua: sistemas de tráfico, transacciones económicas, rutas aeronáuticas, plataformas digitales y dispositivos personales producen volúmenes de información sin precedentes. Este fenómeno ha transformado la información en un recurso estratégico, no solo en términos económicos, sino también epistemológicos, al redefinir las formas en que ciertas instituciones observan, registran y analizan la realidad. En este escenario, la CD no se presenta únicamente como un conjunto de herramientas técnicas, sino como un enfoque que transforma los modos contemporáneos de producción de conocimiento.

conjunto de instrumentos polivalentes que pueden incorporarse a distintos campos, incluida la historia. Su aplicación en investigaciones históricas puede potenciar la heurística mediante la MD en al menos tres niveles: (a) fuentes nativas digitales, (b) fuentes digitalizadas y (c) fuentes audiovisuales. En otros términos, amplía la escala, pero exige mayor atención a la mediación técnica, la formalización y la cuantificación, así como una dependencia de infraestructuras.

Sin embargo, estas tecnologías no son neutrales ni universales. Su funcionamiento depende de infraestructuras específicas, de los idiomas en los que fueron desarrolladas y de las capacidades económicas necesarias para entrenarlas y mantenerlas; factores que pueden introducir sesgos algorítmicos. Frente a esta situación, la participación de equipos interdisciplinarios con conocimientos en los contextos que se pretende trabajar es indispensable.

Más allá del documento físico, la introducción de la CD exige una sofisticación de la mirada crítica del investigador con tal de navegar en esta abundancia informativa con integridad, y evitar naufragar en tesis deterministas o tecno-solucionistas. Para esto, la historia se enfrenta a la apertura de una caja negra donde los resultados de un buscador no responden necesariamente a las intenciones del historiador, sino a los criterios de la herramienta (Pons 100). Formar una técnica crítica permite detectar esa *algocracia*, que a menudo invisibiliza voces subyacentes enviándolas a un lugar recóndito de la página de resultados (Sayers 32).

Esta ilusión aritmética es uno de los problemas centrales asociados al surgimiento del *Big Data*. En este contexto, se consolida una retórica de objetividad basada en la magnitud de evidencia disponible y documentación analizada, aunque ello no elimina la mediación interpretativa inherente al estudio histórico. Autores como Ian Milligan y Matthew Jockers han cuestionado esta premisa al señalar que la acumulación de información no produce verdad por sí misma, sino conjuntos de rastros procesados que continúan siendo vestigios del pasado.

La falacia de la objetividad se intensifica a medida que aumentan tanto la complejidad como el volumen de información analizada, un caso de ello es *Time on the Cross* (1974). Pero los algoritmos, aun manejando cantidades ingentes de información, reflejan los sesgos de sus arquitectos. Por esta razón, la exégesis

histórica requiere incorporar una instancia adicional de reflexión crítica, poniendo especial énfasis en una transparencia metodológica.

Por otra parte, un problema derivado del uso de estas herramientas, es la soberanía de los datos. La mayor parte de las técnicas o del software empleados para el análisis de información de manera masiva han sido desarrollados en Europa y América del Norte. Este es un problema que desafía la soberanía de datos y la exactitud histórica porque usamos una estructura del Norte Global para analizar culturas del Sur. Como resultado, ciertos datos o huellas digitales pueden quedar invisibilizadas limitando la comprensión de realidades puntuales (Milligan, *History in the Age* 12). En América Latina, iniciativas como Nuestra MemorIA en Chile y el proyecto Cruzar en Uruguay representan esfuerzos orientados a responder ante necesidades éticas, políticas y metodológicas vinculadas al tratamiento de archivos producidos por los regímenes dictatoriales.

En el caso uruguayo, el proyecto Cruzar impulsó el desarrollo de LUISA (Leyendo Unidos para Interpretar loS Archivos), una iniciativa destinada al procesamiento de archivos militares de la última dictadura en el país (1973-1985). De manera similar a Nuestra MemorIA, este proyecto busca ampliar y democratizar el acceso a la información contenida en los fondos documentales y facilita su exploración, análisis e interpretación mediante técnicas digitales como las ya descritas.

Es precisamente en este punto donde enfoques como HITL y H-LAM/T son imprescindibles, pues logran justificar el uso de investigadores, y no solo programadores: el científico social es especialmente importante porque logra producir y verificar datos fehacientes trabajando a la par con colegas de otras disciplinas. La relevancia de estas perspectivas metodológicas no se limita a la protección y la soberanía, sino que responden a exigencias y técnicas propias del trabajo con fuentes históricas. La heterogeneidad, la irregularidad y la complejidad de estos materiales suelen requerir herramientas desarrolladas para contextos específicos. Proyectos emblemáticos como los creados en Chile y Uruguay demuestran el potencial del trabajo entre disciplinas y del análisis de fuentes.

5.2. Ventajas heurísticas y cambios de escala en la investigación histórica

La actual abundancia de fuentes —tanto digitales nativas como digitalizadas— redefine las condiciones de acceso al pasado. Archivos antes dispersos o de difícil consulta se vuelven potencialmente accesibles a escala global, pero esta disponibilidad no es homogénea. Coexisten repositorios masivamente digitalizados con colecciones fragmentarias y con vastos volúmenes de documentación nacida de manera digital cuya preservación a largo plazo sigue siendo incierta. La expansión del archivo, por tanto, no reproduce las asimetrías, sino que las reordena.

En este contexto emerge lo que Serge Noiret (2015) denomina una historia digital «glocal», donde lo local y lo global se encuentran entrelazados por infraestructuras digitales y circuitos transnacionales de información (70). Esta perspectiva tensiona la noción clásica de historia local: los procesos microhistóricos no pueden comprenderse sin atender a redes más amplias de circulación de registros históricos, personas, significados y contextos. Al mismo tiempo, la huella digital de individuos, instituciones y comunidades amplía el campo de observación historiográfica más allá de los archivos tradicionales.

Para la historia, esta transformación supone un aumento de la complejidad en el estudio de las redes sociales asociadas a sujetos y eventos históricos. Las relaciones ya no se observan solamente a través de su dimensión física o documental, sino también mediante los rastros digitales que producen. Este tránsito de lo analógico a lo digital ha sido incorporado con cautela: la Universidad de Leiden constituye un caso relevante, así como un antecedente significativo en la institucionalización de prácticas de gestión de datos históricos.

En el 2019, el Instituto de Historia de la Universidad de Leiden (2016) adoptó lineamientos formales sobre la gestión de datos que pueden resumirse en tres principios (1):

- I. *Todos los datos durante la investigación deben ser almacenados de manera segura.* Esto debe permitir la integridad, la disponibilidad y la confidencialidad de los datos.
- II. *La documentación de la investigación puede ser reutilizable.* Los datos pueden almacenarse siempre y cuando sea de manera accesible, reutilizable y sustentable. Luego de completar la pesquisa,

se pueden agregar los metadatos, documentación y software utilizados durante el proceso.

- III. *Almacenar los datos éticamente.* El acopio de información debe seguir una normativa internacional, y el almacenaje debe ser de al menos diez años.

Estas políticas no son meramente administrativas, sino que redefinen lo que cuenta como evidencia histórica al transformar el proceso de pesquisa en un flujo documentado, auditable y reproducible. Aun así, el desarrollo y continuidad de las normativas dependen en gran medida de los recursos disponibles. En este sentido, el financiamiento se convierte en una variable estratégica que condiciona el acceso a infraestructura y equipos interdisciplinarios necesarios.

En la medida en que los métodos sean documentados y transparentes, ciertos procedimientos pueden volverse auditables y reproducibles. Idealmente, la colaboración entre historia y CD podría ampliar la autonomía metodológica del primero, aunque en la práctica muchas instituciones seguirán dependiendo de infraestructuras externas. No obstante, esta autonomía es desigual y depende de recursos técnicos, financiamiento e infraestructura que no todas las instituciones poseen. Las fortalezas de un grupo de profesionales en ambas carreras pueden ser varias:

- I. *Heurística.* En términos heurísticos, la CD puede ampliar el horizonte documental del historiador al permitir la exploración sistemática de grandes corpus textuales y visuales. Con la mediación constante del investigador histórico, los modelos computacionales pueden revelar recurrencias, anomalías y patrones que serían invisibles con lecturas individuales. En otras palabras, es a través de fuentes que se dota de sentido los vestigios de la actividad humana (Bresciano, *La investigación histórica...* 48). El aprovechamiento de estas herramientas depende de la colaboración interdisciplinaria, permitiendo sobrepasar los límites del examen histórico tradicional. Por ejemplo, el *Old Bailey Project*, una iniciativa inglesa, donde se indexaron 197.752 litigios, que datan entre 1674 y 1913 y que totalizan más de 127 millones de palabras, permitió conocer la vida de aquellos que no eran pertenecientes a la elite. Este caso muestra que la CD no

sustituye la comprensión histórica, sino que amplía el archivo social visible y desplaza el foco desde las elites hacia sectores históricamente subrepresentados.

- II. *Hermenéutica*. La CD permite sistematizar y analizar patrones previamente identificados de manera cualitativa, pero no puede sustituir la interpretación hermenéutica propia del trabajo historiográfico. En toda investigación orientada a estudiar el pasado, la producción y el tratamiento de los datos plantean interrogantes éticas tanto sobre los procedimientos y los resultados. Asimismo, la CD debe enfrentarse a problemas frecuentes de las fuentes históricas, como la presencia de información incompleta o faltante, así como la necesidad de construir bases de conocimiento y practicar una gobernanza digital sobre la información. Dentro de este contexto, el investigador social procura desarrollar una metodología transparente y justificable mientras el científico de datos debe garantizar procedimientos auditables.
- III. *Epistemología*. La historia como la CD son compatibles en la medida de que se respeten las técnicas del procesado de registros mediante una metodología transparente. La imbricación entre ambas puede generar nuevas líneas de exploración de archivos, metodologías y técnicas más transparentes. La hermenéutica histórica se apoya en una lectura ética e intensiva, una selección clara de fuentes y un trabajo contextual previo profundo. La CD inspira al científico social a explicitar sus supuestos teóricos y metodológicos; así, fortalece la reflexividad del proceso de investigación. Adicionalmente, vuelve visible la mediación técnica del algoritmo y subraya el problema de la trazabilidad de los datos.

Esta relación previene el tecnosolucionismo al subrayar que ningún modelo es neutral ni universal. La base de datos no es un espejo del pasado, sino una hipótesis que requiere interpretación histórica constante. En este sentido, el rol del historiador no se diluye, sino que se vuelve más central: su tarea no es leer menos, sino interpretar con los nuevos instrumentos en condiciones de una abundancia informativa.

5.3. *Los límites del giro computacional en la historiografía*

Si bien la CD ha mostrado un potencial significativo para ampliar la escala, el alcance y la profundidad del análisis histórico, sus posibilidades no son ilimitadas. Empero, su incorporación al trabajo historiográfico introduce una serie de restricciones epistemológicas, técnicas y materiales que deben ser críticamente interpeladas. Esta sección se propone someter a revisión dichos límites para mostrar que no se trata de obstáculos contingentes u operativos, sino de condicionantes estructurales que moldean cuáles consultas pueden formularse; qué evidencias resultan visibles y qué formas de conocimiento se privilegian.

Antes que nada, es necesario plantear los límites epistemológicos de la historia con datos. Es una disciplina sobre huellas y rastros interpretados; y por más información generada que exista, la tarea historiográfica continúa siendo la de interpretar los conocimientos encontrados. El registro histórico es fundamentalmente incompleto por lo que la abundancia digital y de información no garantiza una verdad científica, sino que estimula a procesar los vestigios. En el sentir de Graham *et al.*, la transformación de los medios y de la metodología es inherentemente parte del quehacer histórico (33).

Estamos entonces en frente a una transformación de medios que intenta abarcar a través de herramientas, la vastedad de la transformación del trabajo y vida del ser humano. La caja negra, concepto ya utilizado en este trabajo, es la pregunta que necesita ser resuelta: todos los recursos informáticos descritos son de uso casi cotidiano en la vida de un científico de datos, pero en la historia son metodologías nuevas que expanden bruscamente nuestra formación. Entonces las preguntas son: ¿en qué modifica la digitalización el quehacer histórico?, ¿qué podemos aportar a otras disciplinas?, ¿qué pueden aportarnos otras disciplinas?

No se trata de rechazar la historia, sino de reconocer y aceptar que está experimentando mutaciones derivadas de cambios en su objeto de estudio. Los fenómenos históricos no pueden traducirse a datos estructurados, tienen que ser dinámicos y maleables. A comparación de las renovaciones historiográficas del siglo XIX y XX, ahora es turno de la disciplina histórica que una tecnología

revolucione su metodología y epistemología. En contraste con las ciencias naturales que durante el último siglo encontraron diversas actualizaciones en sus herramientas, el historiador pudo usufructuar mayormente de medios que aceleraban la producción de texto como la mecanografía o que conservaban mejor los archivos.

Los profesionales en historia deben asumir que su criterio interpretativo es hoy más importante que nunca. Frente al surgimiento de medios digitales cada vez más capaces, tienen tanto la oportunidad como la inevitable responsabilidad de participar activamente en su diseño, implementación y evaluación.

La complejidad puede alimentar una ilusión de objetividad. Corrientes como el *cliodinamismo* o el *Big Data* no eliminan la interpretación, aunque tienden a socavar las singularidades de las historias. La cantidad y complejidad de los datos pueden sugerir mayor rigor, pero no equivale a mayor verdad histórica; además, cuestionar los resultados de un algoritmo que analizó millones de palabras exige un esfuerzo crítico considerable. Los modelos computacionales no descubren la historia, sino que reconfiguran la evidencia disponible conforme a supuestos técnicos y decisiones previas.

Por otro lado, la disponibilidad de material amplía el abanico de posibilidades de estudio. Empero, el uso de procedimientos automatizados exige atención a las relaciones de poder que atraviesan tanto la producción de fuentes, como el diseño de las tecnologías empleadas en su análisis. Los supuestos teóricos incorporados en los modelos y algoritmos pueden influir en los resultados obtenidos reproduciendo sesgos preexistentes o generando interpretaciones poco precisas. Como sostiene Pons, el historiador no puede limitarse a ser un consumidor pasivo de la tecnología, sino que debe asumir una posición crítica y reflexiva frente a las herramientas que utiliza (39).

Asimismo, recursos como el PLN y la MD suelen entrenarse con corpus textuales contemporáneos donde la lengua inglesa es predominante. Esto aumenta el riesgo de interpretar fuentes históricas mediante categorías lingüísticas y conceptuales propias del presente, cuestión que favorece la aparición de anacronismos algorítmicos. Obstáculos similares yacen en el análisis

de redes, donde la simplificación de fenómenos complejos o la generación de conexiones basadas exclusivamente en criterios formales pueden producir representaciones alejadas de las dinámicas históricas efectivas.

En conjunto y para finalizar esta sección, estas limitaciones no invalidan el uso de herramientas digitales en la investigación histórica. Por el contrario, ponen de relieve la necesidad de una historiografía consciente de las técnicas que utiliza, de sus metodologías y sus materiales.

5.4. *La ilusión de objetividad: el debate en torno a «Time on the Cross»*

El libro *Time on the Cross: The Economics of American Negro Slavery* (1974) de Robert Fogel y Stanley Engerman fue un escrito dividido en dos volúmenes que buscó reinterpretar un objeto histórico y alterar el estatus epistemológico de la disciplina con una consolidación de la cliometría. Los autores buscaron hacer una relectura de la esclavitud en el Sur de los Estados Unidos con dos intenciones: primero, querían desafiar la visión tradicional «sentimental» o «racista» que retrataba la esclavitud como una institución ineficiente y basada en coacción racional,⁸ además, buscaban demostrar que el sistema era una empresa nacional, eficiente y económicamente dinámica; segundo, querían atraer la atención de eruditos al uso de la cliometría.⁹

Según la crítica formulada por Thomas L. Haskell (1975), esta tarea fue un ejemplo temprano de una fábrica de erudición: Fogel y Engerman emplearon docenas de asistentes, sobre todo formados en economía, para rastrear y codificar miles de registros de censos de población, sucesiones y diarios de plantaciones. Asimismo, el caso evidenció una tensión persistente entre la búsqueda de regularidades estadísticas y la necesidad de preservar la complejidad de las

⁸ « [El libro] challenges virtually every assumption that has been made about the management of slaves, their work habits, their domestic welfare [...] the economy of the antebellum South in general» (Fogel y Engerman, *Time on the Cross: The Economics of...* [en contratapa]).

⁹ «A series of rapid advances in economics, statistics, and applied mathematics, together with the availability of high-speed computers, put information long locked in obscure archives at the disposal of a new generation of scholars. The review based on the new techniques and hitherto neglected sources has contradicted many of the most important propositions in the traditional portrayal of the slave system. As significant as the correction of past errors is the new information brought to light on the conditions of black bondage. Though the investigations are still in progress, enough have been completed so that the main features of the actual operation of the slave economy are now clear» (Fogel y Engerman, *Time on the Cross: The Economics of...* 4).

experiencias históricas individuales. La reducción de fenómenos sociales amplios a variables cuantificables permitió nuevas formas de análisis, pero también generó cuestionamientos sobre aquello que permaneció fuera de la medición.

Varios autores, como Herbert Gutman y Thomas Haskell identificaron fallas profundas en el trabajo investigativo. Uno de los principales deméritos residía en que gran parte de las estimaciones dependían de suposiciones previas. Esto ocasionaba que las cifras o hipótesis bases de la investigación fueran determinantes: equivocarse en la cantidad de esclavos vendidos, por ejemplo, sería devastador para las conclusiones. Además, contiene sesgos de selección de datos que, en conjunto con las conclusiones holísticas, terminó por desacreditar algunos hallazgos: algunas afirmaciones de la jerarquía laboral de los esclavos se sustentaban en muestras limitadas y no representativas del conjunto de plantaciones analizadas.

Otras críticas apuntaron también a los errores de cálculo y a problemas de contextualización de las fuentes primarias, lo que habría favorecido interpretaciones poco honestas ya que se buscaba ajustarlas a los supuestos teóricos del modelo. Como señaló Richard K. Vedder (2024), la aparente confianza en la capacidad explicativa de los indicadores estadísticos generó la percepción de que una experiencia histórico tan compleja y traumática como la esclavitud estaba siendo suavizada mediante ecuaciones y promedios (680).

Más allá de las controversias específicas, el debate en torno a esta investigación resulta especialmente relevante porque anticipa varios problemas que hoy reaparecen en el contexto del giro computacional. Las discusiones sobre la selección de datos, los supuestos incorporados en los modelos, la transparencia metodológica y los límites de la cuantificación muestran que las tensiones entre medición e interpretación no son exclusivas de la IA. La experiencia de la cliometría constituye un antecedente historiográfico que permite comprender de manera crítica los desafíos contemporáneos asociados al uso de herramientas digitales en la tarea histórica.

6. Conclusión

Este artículo abordó un conjunto de instrumentos aplicados al análisis de fuentes históricas que constituyen la base técnica de la denominada IA. En el caso de Nuestra Memoria, el proyecto pone en tela de juicio algunos de los desafíos éticos asociados al uso de herramientas computacionales en historia. Los expedientes judiciales de la dictadura contienen información vinculada a experiencias traumáticas, desapariciones, terrorismo de Estado, procesos de búsqueda de desaparecidos y casos judiciales aún abiertos. Por ello, las decisiones relativas a la digitalización, almacenamiento, procesamiento y difusión de la información poseen una dimensión ética importante.

La posibilidad de analizar grandes volúmenes de datos con ayuda de la IA amplía las capacidades de investigación, pero obliga también a considerar cuestiones como privacidad y soberanía. En este sentido, los problemas de sesgo y normalización conllevan un profundo peso ético.

La experiencia entre patrones generales y experiencias particulares no constituye un problema nuevo para la disciplina histórica, como ya se reiteró diversas veces. Las discusiones en cuanto a la producción histórica de larga duración o basada en cliometría ya habían planteado interrogantes. Del mismo modo, la microhistoria reaccionó frente a estas tendencias reivindicando la capacidad explicativa de casos particulares y contextos específicos. Visto desde esta perspectiva, el giro computacional no se configura como una ruptura absoluta, sino que replantea discusiones preexistentes.

En conclusión, el llamado giro computacional no supone una sustitución de la práctica historiográfica por sistemas automatizados y autónomos, sino una transformación de las condiciones sobre las cuales se produce el conocimiento histórico. El desafío no consiste únicamente en adoptar estas tecnologías, sino en comprender sus alcances y supuestos. El historiador debe reforzar la importancia en la interpretación y la reflexión metodológica para mantener estándares altos de producción académica.

Fuentes

Fogel, Robert William, y Stanley L. Engerman. *Time on the Cross: The Economics of American Negro Slavery*. Little, Brown and Company, 1974.

Referencias

Abelson, Hal, Ken Ledeen, and Harry Lewis. *Blown to Bits: Your Life, Liberty, and Happiness After the Digital Explosion*. Addison-Wesley, 2008.

Ballesteros Estrada, Silvia Socorro, et al. «Los problemas de identificación de caracteres OCR para la recuperación de texto en el libro antiguo: un análisis de caso en el Fondo Antiguo de la Biblioteca Central, UNAM». *Biblioteca Universitaria*, vol. 15, n.º 1, 2012, pp. 25-34.
<https://doi.org/10.22201/dgb.0187750xp.2012.1.39>

Barleta, Leonardo. «Bases de datos y organización de la información histórica: por un modelo (supuestamente) universal», en Juan Andrés Bresciano y Tiago Gil *La historiografía ante el giro digital: Reflexiones teóricas y prácticas metodológicas*, Ediciones Cruz del Sur, 2015, pp. 111–134.

Bocanegra Barbecho, Lidia, y Nuria Rodríguez Ortega. «Nuria Rodríguez Ortega. Un compromiso con las humanidades digitales». *Quiroga*, n.º 14, 2018, pp. 142-155.

Bresciano, Juan Andrés. «Los estudios históricos en la sociedad de la información», en Juan Andrés Bresciano y Tiago Gil, *La historiografía ante el giro digital: Reflexiones teóricas y prácticas metodológicas*, Ediciones Cruz del Sur, 2015, pp. 13–40.

---. «La Historia local en tiempos de Internet. Nuevos cauces para una especialización disciplinaria. » *Tempo e Argumento*, vol. 6, no. 12, mayo-agosto 2014, pp. 5–22

---. «La documentación electrónica y el estudio del presente: posibilidades y desafíos», en *El tiempo presente como campo historiográfico. Ensayos teóricos y estudios de casos*, compilado por Juan Andrés Bresciano, Ediciones Cruz del Sur, 2010, pp. 283–304.

---. *La investigación histórica y las nuevas tecnologías*. Librería de la Facultad de Humanidades y Ciencias de la Educación, 2000.

Bresciano, Juan Andrés, y Tiago Gil (Compiladores). *La historiografía ante el giro digital. Reflexiones teóricas y prácticas metodológicas*, Cruz del Sur, 2015.

Eiroa, Matilde. «El pasado en el presente: el conocimiento historiográfico en las fuentes digitales». *Ayer*, vol. 110, n.º 2, 2018, pp. 83-109. <https://e-archivo.uc3m.es/rest/api/core/bitstreams/4ce4bf3b-2cdo-489c-a54a-0c5caf1d1d13/content>

Graham, Shawn, et al. *Exploring Big Historical Data: The Historian's Macroscope*. Imperial College Press, 2015.

Gutman, Herbert G. *Slavery and the Numbers Game: A Critique of "Time on the Cross"*. University of Illinois Press, 1975.

- Haskell, Thomas L. «The True and Tragical History of “Time on the Cross”». *The New York Review of Books*, vol. 22, n. 15, 1975, pp. 1-12.
https://benschmidt.org/dighist17/Readings/haskell_1975_time.pdf
- Leiden University, Institute of History. *Data Management Guidelines for Historical Research*, 2019.
<https://www.universiteitleiden.nl/binaries/content/assets/geesteswetenschappen/institute-for-history/guidelines-data-management.pdf>
- Milligan, Ian. *History in the Age of Abundance? How the Web Is Transforming Historical Research*. McGill-Queen's UP, 2019.
- . «La historia en la era de la abundancia: archivos web e investigación histórica». *Historia y Memoria*, n.º especial, 2020, pp. 235-269.
- Noiret, Serge. «Historia digital e Historia pública», en *La historiografía ante el giro digital: Reflexiones teóricas y prácticas metodológicas*, editado por Juan Andrés Bresciano y Tiago Gil, Ediciones Cruz del Sur, 2015, pp. 41-76.
- Pons, Anaclet. *El desorden digital: guía para historiadores y humanistas*, Siglo XXI, 2013.
- Proyecto Nuestra Memoria. *Explorando los archivos de la memoria*. Pontificia Universidad Católica de Chile e Instituto Milenio Fundamentos de los Datos, 2024.
- Rota, Alleson Ramón. «El uso de la minería de datos como heurística para la teoría de la historia y la historia de la historiografía». *Amoxthli*, 2021, pp. 1-17. <https://doi.org/10.38123/amox7.205>
- Sayers, Jentery. «Bringing Trouvé to Light: Speculative Computer Vision and Media History». En Kevin Kee y Timothy J. Compeau, editores, *Seeing the Past with Computers: Experiments with Augmented Reality and Computer Vision for History*. University of Michigan Press, 2019, pp. 32-49.
- Universidad de la República. *Proyecto Cruzar: Sistematización de archivos del pasado reciente*. Montevideo: Facultad de Información y Comunicación y Facultad de Ingeniería, 2017.
- Vedder, Richard K. «“Time on the Cross” at Fifty». *The Independent Review*, vol. 28, n.º 4, 2024, pp. 673-681.