Convergências: estudos em Humanidades Digitais

INTELIGÊNCIA ARTIFICIAL E PESQUISA EM HISTÓRIA MEDIEVAL: APLICAÇÕES, DESAFIOS E POSSIBILIDADES

Artificial Intelligence and Medieval History Research: Applications, Challenges and Possibilities

Inteligencia Artificial e Investigación en Historia Medieval: Aplicaciones, Desafíos y Posibilidades

Saymon da Silva Siqueira [1]
Adriana Vidotte [2]

Resumo: Este artigo analisa a integração da Inteligência Artificial (IA) na pesquisa medieval como um evento epistemológico que reconfigura práticas historiográficas. Argumenta-se que a opacidade algorítmica ("caixa-preta") da IA gera uma "crise de justificação", desafiando o raciocínio tradicional do historiador. A análise de aplicações como o Reconhecimento de Texto Manuscrito (HTR) e a Análise de Redes Sociais (SNA) evidencia a necessidade de uma nova "crítica das fontes 2.0", que avalie tanto o documento quanto o processo computacional. Conclui-se que o ofício do historiador se expande: o pesquisador deve se tornar um "crítico do instrumento", aplicando o rigor hermenêutico à própria tecnologia para garantir uma historiografia digital reflexiva e relevante.
Palavras-chave*:* História Medieval. Inteligência Artificial. Humanidades Digitais. Fontes Históricas. Epistemologia.

Abstract: This article analyzes the integration of Artificial Intelligence (AI) in medieval research as an epistemological event that reconfigures historiographical practices. It argues that the algorithmic opacity ("black box") of AI generates a "crisis of justification," challenging the historian's traditional reasoning. The analysis of applications such as Handwritten Text Recognition (HTR) and Social Network Analysis (SNA) highlights the need for a new "source criticism 2.0," which assesses both the document and the computational process. It concludes that the historian's craft is expanding: the researcher must also become a "critic of the instrument," applying hermeneutic rigor to the technology itself to ensure a reflexive and relevant digital historiography.

[1] Mestre em História Ibérica. Doutorando em História pela Universidade Federal de Goiás, Goiânia, Goiás, Brasil.. E-mail: saymon.siqueira@discente.ufg.br; Lattes: http://lattes.cnpq.br/9488185068212989; ORCID iD: https://orcid.org/0000-0002-3557-0682.
[2] Doutora em História. Professora Associada na Faculdade de História do Universidade Federal de Goiás, Goiânia, Goiás, Brasil. E-mail: adrianavidotte@ufg.br; Lattes: http://lattes.cnpq.br/6066743926743863; ORCID iD: https://orcid.org/0000-0001-7463-9980.

Keywords*:* Medieval History. Artificial Intelligence. Digital Humanities. Historical Sources. Epistemology.

Resumen: Este artículo analiza la integración de la Inteligencia Artificial (IA) en la investigación medieval como un evento epistemológico que reconfigura las prácticas historiográficas. Se argumenta que la opacidad algorítmica ("caja negra") de la IA genera una "crisis de justificación", desafiando el razonamiento tradicional del historiador. El análisis de aplicaciones como el Reconocimiento de Texto Manuscrito (HTR) y el Análisis de Redes Sociales (SNA) evidencia la necesidad de una nueva "crítica de fuentes 2.0", que evalúe tanto el documento como el proceso computacional. Se concluye que el oficio del historiador se expande: el investigador debe convertirse también en un "crítico del instrumento", aplicando el rigor hermenéutico a la tecnología misma para garantizar una historiografía digital reflexiva y relevante.
Palabras clave: Historia Medieval. Inteligencia Artificial. Humanidades Digitales. Fuentes Históricas. Epistemología.

Introdução

A integração da Inteligência Artificial (IA)[3] na pesquisa histórica, e notadamente nos estudos medievais, representa não apenas uma evolução metodológica, mas um evento epistemológico que reconfigura o ofício do historiador de maneira fundamental (Crymble, 2021). Ferramentas capazes de transcrever manuscritos em escala, identificar entidades em milhares de documentos e mapear redes sociais complexas prometem um acesso sem precedentes a um passado de fontes vastas e de difícil decifração. Contudo, essa promessa é acompanhada por um profundo desafio aos princípios fundadores da disciplina, forçando um reexame das formas como o conhecimento histórico é construído, validado e justificado, especialmente ao lidar com a complexidade do registro histórico medieval.

Essa complexidade se manifesta de múltiplas formas. A pesquisa medieval lida com uma diversidade de suportes codicológicos (códices, cartas, fragmentos) que impõem desafios de manuseio e acesso, mediados ainda por barreiras linguísticas e paleográficas significativas. Some-se a isso a lacuna documental, uma condição quase onipresente para o período, que coloca a IA em uma posição ambivalente: por um lado, ela oferece a promessa de conectar

[3] A Inteligência Artificial é um campo da ciência da computação dedicado a criar sistemas que simulam a inteligência humana. A abordagem dominante hoje é o Aprendizado de Máquina (Machine Learning), no qual os algoritmos "aprendem" padrões a partir de dados, em vez de serem explicitamente programados com regras (Sommerschield *et al*., 2023). Os avanços da última década são impulsionados principalmente pelo aprendizado profundo (deep learning), um subtipo que utiliza redes neurais com múltiplas camadas. Essa técnica, aliada a um escopo massivo de dados de treinamento e arquiteturas sofisticadas como os transformers, resultou nos atuais Modelos de Linguagem de Grande Porte (*Large Language Model* - LLM), capazes de gerar resultados com um grau de fluidez e coerência notavelmente próximos aos humanos (Sommerschield *et al*., 2023).

fontes fragmentárias e revelar sentidos ocultos; por outro, corre o risco de criar correlações espúrias que mascarem a própria ausência. É precisamente neste terreno de fontes difíceis que a tensão entre a interpretação humanística e a lógica computacional se torna mais aguda e reveladora.

O cerne desta transformação reside, de fato, na tensão fundamental entre a tradição hermenêutica do historiador (o *Verstehen*, a compreensão empática e interpretativa) e a lógica matemática dos modelos computacionais. Essa tensão, fundadora do campo (Finnemann, 2014, p. 95), opõe a prática humanística, que lida com o ambíguo e o excepcional, à exigência da computação por modelos "totalmente explícitos e consistentes" (Birnbaum, Bonde e Kestemont, 2017, p. S4). Essa necessidade de formalização pode ser "alienante para estudiosos em campos das humanidades onde o excepcional é frequentemente abraçado" (Birnbaum, Bonde e Kestemont, 2017, p. S4), uma vez que fenômenos quantificáveis são mais tratáveis por algoritmos do que "características hermenêuticas, isto é, aquelas relativas à interpretação" (Birnbaum, Bonde e Kestemont, 2017, p. S7).

A prática histórica canônica alicerça-se numa cadeia de raciocínio transparente e auditável. Diante de uma interpretação, um colega pode inquirir, "Como chegou a esta conclusão?", e o historiador responde indicando passagens textuais, o contexto material do documento e fontes corroborativas. A introdução de algoritmos de "caixa-preta" (*black box*), cuja opacidade pode ser tanto técnica quanto intencional (Burrell, 2016), fratura essa cadeia de justificação. A mesma pergunta, quando dirigida a uma análise gerada por um modelo de *topic modeling* ou uma rede neural, encontra uma resposta que, em parte, é epistemologicamente insatisfatória. Gera-se, assim, uma "crise de justificação", na qual a autoridade do pesquisador é desafiada pela sua capacidade de tornar o processo de análise inteiramente transparente, submetendo a academia a uma "sociedade da caixa-preta" onde os algoritmos que controlam a informação permanecem secretos (Pasquale, 2015).

Este desafio manifesta-se de forma concreta nas principais frentes de aplicação da IA à documentação medieval. O Reconhecimento de Texto Manuscrito (HTR, *Handwritten Text Recognition*) questiona como estabelecer um texto fiável a partir de dados ruidosos e probabilisticos, mediando de forma inédita nosso acesso ao "original". O Reconhecimento de Entidades Nomeadas (NER, *Named Entity Recognition*), ao automatizar a prosopografia, levanta o problema de como identificar e desambiguar atores históricos sem impor categorias anacrônicas. Por sua vez, a Análise de Redes Sociais (SNA, Social Network Analysis) nos força

a questionar como laços sociais são inferidos a partir de coocorrências textuais sem simplificar a natureza complexa das relações humanas medievais. Em cada um desses casos, o risco do que Tim Hitchcock denominou "fechamento do problema" torna-se uma ameaça epistemológica real, pois corremos "o perigo de reinventar o problema para colocar questões que as ferramentas e os dados disponíveis nos permitem responder" (Clavert, 2013).

Diante deste cenário, este artigo procede a uma análise das implicações da IA para a pesquisa em História Medieval. Inicia-se com um enquadramento teórico-metodológico que situa o debate atual no legado das tradições quantitativas da história e define os conceitos-chave para uma "crítica das fontes 2.0" (Zbíral, Sikk e Shaw, 2025). Em seguida, analisa-se um conjunto de aplicações em três áreas centrais: o tratamento de manuscritos, o processamento de texto e a modelagem de redes sociais, demonstrando como cada tecnologia reconfigura práticas historiográficas específicas. Por fim, discute-se as implicações mais amplas dessa transformação para o ofício do historiador, focando nos problemas de opacidade, viés e na necessidade de desenvolver uma historiografia digital mais reflexiva. O objetivo não é rejeitar as novas ferramentas, mas forjar o aparato crítico indispensável para manejá-las com rigor, garantindo que o poder da IA sirva à expansão, e não à restrição, da nossa compreensão do passado.

Enquadramento Teórico-Metodológico

A atual inserção da Inteligência Artificial na pesquisa histórica, embora pareça uma ruptura, é herdeira de uma longa trajetória de diálogos, por vezes tensos, entre a história e os métodos quantitativos. A Escola dos Annales, com sua ênfase na *longue durée* e na história serial, já sinalizava o potencial da análise de grandes conjuntos de dados para superar a história puramente factual ou *événementielle* (Nawiyanto, Handayani e Salindri, 2022). Foi nesse contexto que Emmanuel Le Roy Ladurie proferiu sua célebre e posteriormente matizada provocação de que "o historiador de amanhã será programador ou não será nada", encapsulando o otimismo quantitativo de uma era (Fickers, 2021).

Décadas mais tarde, o movimento conhecido como *History and Computing*, ativo nos anos 1980 e 1990, estabeleceu as bases diretas da história digital contemporânea (Hockey, 2004). Este campo já debatia muitas das questões que hoje ressurgem com nova urgência, como a relação entre o quantitativo e o qualitativo, os desafios da criação de bases de dados históricas e as implicações da utilização de computadores para a prática do historiador. Reconhecer este

legado é crucial para compreender que as críticas ao positivismo e ao reducionismo tecnológico não são novas, mas sim a continuação de um diálogo essencial para a vitalidade da disciplina (Coletto, 2021).

A formalização das Humanidades Digitais como campo distinto consolidou estas práticas. Volumes fundacionais como *A Companion to Digital Humanities* (Schreibman, Siemens e Unsworth, 2004) mapearam o terreno técnico inicial, focado na codificação de textos e na criação de bases de dados. Contudo, foi com a chamada "virada crítica", exemplificada por coletâneas como *Debates in the Digital Humanities* (Gold, 2012), que o campo passou a questionar não apenas o "como fazer", mas o "por que fazer". Essa inflexão trouxe para o centro do debate questões sobre poder, viés e a necessidade de uma prática digital mais reflexiva.

É neste cenário que se inserem obras recentes fundamentais para este artigo, como as de Adam Crymble (2021) sobre as transformações do ofício, de Ian Milligan (2019) sobre a pesquisa na "era da abundância" de fontes, e de Max Kemman (2021) sobre as "zonas de transação" interdisciplinares onde o conhecimento histórico digital é negociado. O objetivo, conforme articulado por projetos que buscam integrar formalmente a crítica das fontes a modelos computacionais, é demonstrar que "os métodos computacionais devem ter seu lugar firme no centro, e não nas margens, da história como disciplina" (Zbíral, Sikk e Shaw, 2025, p. 70, tradução nossa).

No contexto brasileiro, a reflexão sobre estas transformações também tem ganhado corpo, com pesquisadores investigando os impactos da era digital no ofício do historiador (Santos, 2021), os desafios epistemológicos de grandes *corpora* como hemerotecas digitais (Brasil, 2020) e o problema hermenêutico colocado pelas novas tecnologias (Gonçalves, 2022). Alinhado a essa perspectiva crítica, o presente artigo adota como metodologia uma revisão qualitativa e analítica de um conjunto selecionado de projetos e publicações que aplicam IA à História Medieval.

Alinhado a essa perspectiva crítica, o presente artigo adota como metodologia uma revisão qualitativa e analítica de um conjunto selecionado de projetos e publicações. A seleção dos casos de estudo seguiu uma abordagem de amostragem intencional (*purposive sampling*), realizada a partir de um levantamento nas principais bases de dados acadêmicas (*Scopus*, *Web of Science*, *Google Scholar*) e nos anais de conferências da área (ADHO), focando em trabalhos publicados nos últimos cinco anos. Os critérios para a inclusão de um trabalho como exemplo central na análise foram:

1. **Relevância Acadêmica Objetiva:** Verificada pela publicação em periódicos de alto impacto para o campo (como *Speculum*, *Digital Medievalist*, *DHQ*), pela associação com projetos de pesquisa financiados por agências de fomento de prestígio, como a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP/Brasil), Fundação para a Ciência e a Tecnologia (FCT/Portugal)[4], Alliance of Digital Humanities Organizations (AHRC/Alemanha), ou pela apresentação em conferências internacionais de referência.
2. **Representatividade Metodológica:** Escolha de casos que exemplificam de forma clara as principais frentes de aplicação da IA na área (HTR, NER, SNA), permitindo uma análise focada nas reconfigurações de práticas historiográficas específicas.
3. **Transparência Metodológica:** Prioridade para projetos e artigos que detalham abertamente seus pressupostos, modelos de dados, desafios e limitações, tornando possível a análise crítica de suas implicações epistemológicas, que é o foco deste estudo.

O objetivo, portanto, não é conduzir um inventário exaustivo, mas utilizar estes exemplos, selecionados por seu impacto e clareza metodológica, como lentes para examinar as reconfigurações em curso no coração da prática historiográfica.

A Reconfiguração das Práticas Historiográficas: Análise de Aplicações

A transição de uma prática historiográfica solitária, baseada na leitura próxima, para uma abordagem colaborativa e computacionalmente assistida, redefine a essência do trabalho com as fontes medievais. Longe de serem meros auxiliares, os métodos de Inteligência Artificial atuam como parceiros ativos na construção do conhecimento, cujas operações, vieses e lógicas internas precisam ser compreendidas e criticadas. Esta seção analisa três domínios centrais desta transformação: a análise de manuscritos, o processamento de textos e a

[4] O projeto *iForal* — Forais medievais portugueses: uma perspetiva histórica e linguística na era digital (2021–2024) exemplifica a escala de empreendimentos em Humanidades Digitais. Financiado pela Fundação para a Ciência e a Tecnologia (FCT) de Portugal com o valor de 190.459,23€, sua estrutura evidencia a complexidade da colaboração: foi constituído por uma instituição proponente (Universidade de Lisboa), cinco universidades portuguesas parceiras, uma parceria internacional (Université de Lausanne), uma equipe de 23 investigadores de diversas áreas e 6 consultores internacionais, além de prever 6 bolsas de investigação. O objetivo do projeto foi a criação da plataforma EDICOLAB para a edição digital de um vasto corpus de forais régios medievais, demonstrando como a análise de fontes complexas na era digital depende de investimentos substanciais e de uma ampla rede de cooperação.

modelagem de redes sociais. O objetivo é demonstrar, por meio de exemplos concretos, como cada ferramenta não apenas acelera a pesquisa, mas principalmente reconfigura as questões epistemológicas que o historiador deve enfrentar.

Antes de examinar cada domínio em detalhe, o quadro a seguir (Quadro 1) sintetiza o argumento central desta seção. Ele organiza a análise mapeando os principais métodos de IA às práticas historiográficas que eles reconfiguram, contrastando a promessa heurística de cada ferramenta com o desafio epistemológico correspondente. A matriz serve, portanto, como um roteiro para a discussão desenvolvida nas subseções subsequentes.

Quadro 1 - Matriz de Reconfiguração Epistemológica na Pesquisa Medieval com IA

Método de IA	Prática Historiográfica Reconfigurada	Promessa Heurística	Desafio Epistemológico Central
HTR (Reconhecimento de Texto Manuscrito)	Paleografia, Diplomática, Crítica Textual	Transcrição em massa e acesso a vastos corpora documentais.	A natureza probabilística do texto; a gestão da incerteza e a "crise de justificação" na validação da transcrição.
NER (Reconhecimento de Entidades Nomeadas)	Prosopografia, Toponímia, Crítica Interna	Extração e indexação automática de atores e lugares em larga escala.	O risco do anacronismo na definição de categorias; o viés do processo de anotação e o "salto hermenêutico" para validar entidades.
SNA (Análise de Redes Sociais)	História Social, Prosopografia, Análise de Estruturas	Visualização e quantificação de sistemas sociais complexos.	A inferência do "laço social" a partir da coocorrência textual; a rede como modelo heurístico e não como espelho da realidade.

Fonte: os autores, 2025.

Da Paleografia à Análise Visual: A Transcrição e a Materialidade em Debate

O trabalho com manuscritos medievais, pilar da erudição clássica, é talvez a área mais visivelmente impactada pela IA. O Reconhecimento de Texto Manuscrito (HTR), em particular, emergiu como uma tecnologia transformadora, com plataformas como a *Transkribus*[5] permitindo a transcrição de vastos *corpora* documentais em uma fração do tempo exigido pela paleografia manual (Mühlberger *et al.*, 2019). Contudo, essa aceleração introduz uma complexidade fundamental. O resultado de um modelo de HTR não é um texto definitivo, mas

[5] Transkribus é uma plataforma e cooperativa europeia, desenvolvida originalmente pela Universidade de Innsbruck (Áustria), que oferece um conjunto integrado de ferramentas para a digitalização, reconhecimento de texto manuscrito (HTR), transcrição e pesquisa em documentos históricos. É amplamente utilizada em projetos de humanidades digitais por permitir que pesquisadores e instituições treinem seus próprios modelos de IA, adaptados às particularidades das escritas com as quais trabalham.

uma "testemunha" textual probabilística, repleta de incertezas e erros que precisam ser metodologicamente tratados (Stutzmann *et al*., 2019). A questão epistemológica, portanto, desloca-se da habilidade caligráfica do paleógrafo para a capacidade do historiador de auditar a opacidade do modelo algorítmico, sua reprodutibilidade e suas métricas de erro (Beatrice, 2023).

A gestão dessa incerteza tornou-se um campo de inovação metodológica. Em vez de uma correção manual exaustiva, projetos como o HORAE[6], que estuda os Livros de Horas medievais, propõem uma solução sofisticada. Eles tratam a transcrição ruidosa do HTR como uma nova testemunha manuscrita dentro de uma tradição textual, utilizando algoritmos de detecção de reutilização de texto para alinhar e corrigir os dados gerados automaticamente contra edições críticas e outros manuscritos conhecidos. A ambição é, assim, "criar dados corretos a partir de dados difusos, incorretos e gerados automaticamente" (Stutzmann *et al*., 2019, p. 5, tradução nossa). Essa abordagem integra a nova tecnologia à tradição filológica da colação e da estematologia, em vez de simplesmente substituir uma pela outra, demonstrando uma prática digital criticamente informada.

A reconfiguração da análise de manuscritos, contudo, transcende a palavra escrita. Redes neurais são agora empregadas para detectar e classificar automaticamente elementos visuais, como iluminuras, agilizando o trabalho de historiadores da arte (Aouinti *et al*., 2022). Outros estudos utilizam a análise computacional do estilo da escrita para estimar a datação de manuscritos hebraicos, complementando a perícia paleográfica com modelos de regressão estatística (Madi *et al*., 2024). Embora promissoras, essas abordagens correm o risco do que se pode chamar de "essencialismo do ecrã", onde a análise se foca apenas na imagem digital, abstraindo-se da materialidade do códice, como o toque do pergaminho ou as marcas de uso. A crítica das fontes na era digital, portanto, expande-se para uma "Crítica Fonte-Dado", um processo que exige a análise de toda a cadeia de produção, desde o artefato físico, passando por sua digitalização e transcrição, até o modelo de dados que o estrutura (Zbíral, Sikk e Shaw, 2025).

[6] O projeto HORAE (Hours Recognition & Analysis Engine) é uma iniciativa de pesquisa interdisciplinar, liderada por especialistas como o paleógrafo Dominique Stutzmann, focada no estudo dos Livros de Horas medievais. Sua principal contribuição metodológica é o tratamento da incerteza do HTR: em vez de corrigir os erros manualmente, o projeto utiliza algoritmos de detecção de reutilização de texto para alinhar as transcrições ruidosas com edições críticas e outras testemunhas textuais conhecidas, tratando a saída da IA como uma nova variante a ser analisada.

Da Crítica Interna à Estilometria: A Hermenêutica e a Autoria em Larga Escala

Uma vez que o texto é transcrito, um novo conjunto de ferramentas de Processamento de Linguagem Natural (PLN) permite análises que reconfiguram a crítica interna das fontes. O Reconhecimento de Entidades Nomeadas (NER), por exemplo, automatiza a tarefa classicamente laboriosa da prosopografia e da toponímia. Estudos como o de Chastang, Aguilar e Tannier (2021) demonstram a capacidade de modelos de IA para extrair nomes de pessoas e lugares de milhares de cartas diplomáticas medievais, viabilizando análises em uma escala antes impraticável. Contudo, essa automação embute decisões epistemológicas cruciais. A definição das próprias categorias de "entidades" a serem reconhecidas exige uma sensibilidade histórica para evitar o anacronismo, e o processo de anotação manual, necessário para treinar os modelos, pode introduzir vieses que se propagam e se solidificam como evidência nos resultados finais (Ehrmann *et al*., 2023).

O desafio se aprofunda ao considerar a natureza dos documentos medievais. A vasta variação paleográfica e linguística, com múltiplas grafias para um mesmo nome ou topônimo, representa um obstáculo significativo para os modelos estatísticos. Como demonstram estudos sobre NER multilingue, a tendência desses modelos é uniformizar as variantes para classificá-las sob um único rótulo, um procedimento que, embora tecnicamente necessário, pode apagar nuances importantes sobre práticas escribais, dialetos regionais ou a evolução de uma identidade ao longo do tempo (Aguilar *et al*., 2022). O ganho em escala, portanto, pode vir ao custo de uma perda de especificidade, exigindo do historiador uma análise crítica sobre que informação é preservada e que informação é abstraída pelo modelo.

De forma semelhante, a estilometria computacional moderniza a tradicional questão da atribuição de autoria. Se no século XV Lorenzo Valla empregou uma análise filológica primorosa para desmascarar a Falsidade da Doação de Constantino, hoje algoritmos podem comparar traços estilísticos em *corpora* massivos para identificar semelhanças autorais. Trabalhos recentes aplicam modelos de *transformers*[7] para analisar textos em latim medieval, conseguindo agrupar obras de um mesmo autor ou círculo sem supervisão humana prévia e produzindo "agrupamentos interpretáveis e de alta qualidade" (De Langhe, De Clercq e Hoste,

[7] Na área de Inteligência Artificial, *transformers* referem-se a uma arquitetura de rede neural introduzida em 2017. Sua principal inovação é o "mecanismo de atenção", que permite ao modelo ponderar a importância de diferentes palavras em uma frase para compreender o contexto de forma mais eficaz. Essa arquitetura é a base da maioria dos Modelos de Linguagem de Grande Porte (LLMs) contemporâneos, como o BERT (*Bidirectional Encoder Representations from Transformers*) e a série GPT (*Generative Pre-trained Transformer*).

2024, p. 63). O surgimento de um agrupamento estatisticamente robusto, contudo, não constitui uma prova de autoria. Ele inaugura um diálogo necessário entre a aparente objetividade da métrica computacional e a subjetividade informada do juízo filológico, que deve validar se a similaridade estilística detectada corresponde a uma conexão histórica genuína ou se é um artefato do modelo.

Tanto no NER quanto na estilometria, o historiador se depara com o que pode ser descrito como um "salto hermenêutico". Os resultados gerados, sejam eles entidades classificadas ou agrupamentos de textos, são construções estatísticas, não fatos históricos. Cabe ao pesquisador a tarefa interpretativa de conferir significado a esses padrões, conectando-os ao contexto e validando-os com o retorno à leitura próxima dos documentos. A IA, nesse sentido, não oferece respostas, mas sim novas proposições em larga escala que exigem um novo nível de escrutínio crítico.

Da Prosopografia à Análise de Redes: A Abstração das Relações Sociais

Finalmente, a Análise de Redes Sociais (SNA) eleva a prospecção de relações humanas a um novo patamar de abstração e escala, superando os limites da prosopografia tradicional. Se antes o historiador mapeava genealogias e relações patrono-cliente de forma manual, a SNA permite visualizar e quantificar a estrutura de vastos sistemas sociais a partir de dados extraídos das fontes. Projetos como o brasileiro *HisCoMM*[8], que mapeia as conexões no Mediterrâneo medieval, incluem em seu escopo metodológico o uso de ferramentas de SNA para recompor interações históricas em um nível transregional (Universidade de São Paulo, [2025]). A força desta abordagem reside em sua capacidade de revelar padrões invisíveis a uma leitura linear. Análises de redes em narrativas como a *História Eclesiástica* de Beda, por exemplo, demonstraram como personagens femininas, embora aparentemente periféricas, ocupavam posições estruturais cruciais como intermediárias de informação (Buchnea e Elsanah, 2022).

O desafio epistemológico central da SNA, contudo, reside na natureza da inferência. Uma coocorrência textual, como a menção de dois indivíduos em uma mesma carta, não constitui, em si, um laço social. É o historiador quem realiza o ato interpretativo de transformar

[8] HisCoMM é a sigla para o projeto temático "Uma História Conectada da Idade Média. Comunicação e Circulação a partir do Mediterrâneo", financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e coordenado pela Universidade de São Paulo (USP). O projeto utiliza a metodologia da História Conectada e um núcleo dedicado a métodos digitais, como georreferenciamento e análise de redes, para reinterpretar as dinâmicas de circulação no Mediterrâneo medieval, afirmando-se como uma referência na aplicação das Humanidades Digitais à História Medieval no Brasil.

essa proximidade textual em uma aresta no grafo[9], definindo sua natureza: é um vínculo de parentesco, vassalagem, comércio ou mera casualidade? Como argumenta De Valeriola (2022), o próprio desenho da rede, a decisão sobre o que constitui um "nó" e uma "aresta", já embute um argumento histórico e um conjunto de pressupostos teóricos. A visualização final, portanto, não é um espelho da realidade social, mas um modelo construído sobre uma série de decisões interpretativas que precisam ser tornadas explícitas.

Ciente deste desafio, a pesquisa mais rigorosa na área tem desenvolvido modelos de dados mais complexos para capturar a heterogeneidade das relações. O projeto DISSINET[10], por exemplo, ao modelar dados dos registros da Inquisição, vai além de simples conexões binárias. Seu modelo permite registrar a perspectiva de quem relata a informação, o nível de certeza da afirmação e a modalidade da relação, integrando a crítica das fontes diretamente na estrutura quantitativa dos dados (Zbíral, Sikk e Shaw, 2025). Essa abordagem metodológica representa um avanço significativo, pois permite que as incertezas e os vieses das fontes sejam formalmente analisados, em vez de serem relegados a notas de rodapé qualitativas.

Em última análise, uma rede social gerada por computador deve ser compreendida como um poderoso modelo heurístico, e não como uma representação fotográfica da sociedade medieval. Sua principal função não é fornecer respostas definitivas, mas gerar novas questões e hipóteses que, por sua vez, devem ser rigorosamente investigadas através de métodos qualitativos e da leitura próxima das fontes. A rede social computacional é um mapa que revela padrões potenciais, mas é o trabalho hermenêutico do historiador que confere a esse mapa profundidade, contexto e significado histórico.

Implicações, Limites e o Ofício do Historiador

A análise das aplicações da IA na pesquisa medieval revela um padrão consistente: cada avanço tecnológico introduz uma nova camada de mediação entre o historiador e a fonte,

[9] No jargão da teoria dos grafos, que fundamenta a Análise de Redes Sociais, um *nó* (ou vértice) representa uma entidade individual no interior da rede (ex: uma pessoa, uma instituição, um lugar). Uma *aresta* (ou laço) representa a relação ou conexão que une dois nós (ex: um casamento, uma correspondência, uma transação comercial). O *grafo* é a estrutura abstrata completa, composta pelo conjunto de nós e arestas, que modela e visualiza a rede como um todo.

[10] DISSINET (*The Dissident Networks Project*) é um projeto de pesquisa focado nos registros da Inquisição medieval, notável por sua abordagem metodológica à crítica das fontes. Seu modelo de dados foi desenhado para registrar formalmente não apenas os dados factuais dos documentos (pessoas, lugares, eventos), mas também as incertezas e os vieses inerentes a eles, capturando variáveis como a perspectiva do depoente, o nível de certeza de uma afirmação e sua modalidade. Essa abordagem, denominada pelos próprios autores como "crítica de fontes 2.0", permite que a análise historiográfica seja integrada diretamente à estrutura quantitativa dos dados.

gerando implicações que transcendem a mera otimização de tarefas. A questão fundamental que emerge é como o historiador pode sustentar sua autoridade interpretativa quando os processos que geram suas evidências são, em parte, opacos. Esta "crise de justificação", resultante da natureza de "caixa-preta" de muitos algoritmos (Burrell, 2016), exige uma redefinição do que constitui uma prática metodológica transparente. A opacidade não é apenas uma limitação técnica, mas um fenômeno com dimensões institucionais e sociais que desafia a capacidade da comunidade acadêmica de auditar e validar a pesquisa de forma plena.

Talvez o risco mais profundo não seja a dificuldade de explicar os resultados, mas sim a sutil reorientação da própria agenda de investigação. O conceito de "fechamento do problema" (*problem closure*), articulado por Tim Hitchcock, alerta para o perigo de os pesquisadores passarem a privilegiar questões que as ferramentas computacionais são capazes de responder, em detrimento daquelas que a disciplina historicamente considera relevantes (Clavert, 2013). Se a tecnologia é excelente para identificar padrões de coocorrência, mas inepta para analisar a ironia, o subtexto ou a intenção não declarada, a curiosidade intelectual do historiador pode ser sutilmente canalizada para o quantificável, deixando de lado a complexidade da experiência humana que não se traduz em dados estruturados. A ferramenta, assim, deixa de ser um meio para se tornar um fator que molda o próprio horizonte do que é considerado pesquisável.

Adicionalmente, a discussão sobre o viés algorítmico deve ser conduzida com maior sofisticação. É preciso diferenciar, no mínimo, três camadas de viés. Primeiro, o viés inerente às fontes medievais, que refletem as perspectivas de uma elite letrada. Segundo, o viés do arquivo e da digitalização, que resulta das decisões sobre o que preservar e como representar digitalmente. Terceiro, o viés do próprio modelo computacional, presente nos dados de treinamento, na arquitetura do algoritmo e nas escolhas de seus parâmetros. Abordagens críticas, como as propostas pelo feminismo de dados, argumentam que esses vieses não são meros erros técnicos a serem corrigidos, mas manifestações de estruturas de poder. Afinal, "a ciência de dados é uma forma de poder" (D'Ignazio e Klein, 2020, p. 6, tradução nossa), e sua aplicação acrítica corre o risco de reforçar narrativas históricas hegemônicas e marginalizar ainda mais as vozes já sub-representadas nas fontes.

Por fim, a adoção destas tecnologias não ocorre em um vácuo material. A infraestrutura da IA possui custos e limites concretos que precisam ser considerados. A predominância da língua inglesa nos grandes modelos de linguagem e nos dados de treinamento cria uma desigualdade epistêmica, prejudicando o desempenho de ferramentas em fontes escritas em

português medieval, galego ou outras línguas historicamente marginalizadas (Nyhan e Ortolja-Baird, 2023). Soma-se a isso a disparidade de acesso a recursos computacionais de alto desempenho entre instituições do Norte e do Sul Global, e o crescente custo ambiental associado ao consumo de energia e água das infraestruturas de IA (Dhar, 2020). O medievalista digital precisa, portanto, ponderar o custo-benefício epistemológico de suas escolhas, questionando se o ganho analítico justifica seus custos materiais e sociais.

Diante desses desafios, o ofício do historiador não se vê diminuído, mas profundamente expandido. O pesquisador da era digital deve ser, para além de um intérprete de fontes, um "crítico do instrumento". Essa nova responsabilidade exige uma literacia que permita investigar e questionar toda a cadeia tecnológica, desde a materialidade da fonte até a lógica do algoritmo. A superação da "crise de justificação" passa por uma transparência metodológica radical, onde as decisões de modelagem, os vieses dos dados de treinamento e os limites das ferramentas são explicitados como parte integrante do argumento histórico. É essa prática reflexiva que garante que a tecnologia permaneça uma serva da investigação histórica, e não sua mestra.

Considerações Finais

Este artigo buscou demonstrar que a crescente adoção da Inteligência Artificial na pesquisa em História Medieval transcende a simples incorporação de novas ferramentas. Trata-se de um fenômeno que impõe uma profunda reflexão sobre a epistemologia do conhecimento histórico. Ao analisarmos as aplicações em manuscritos, textos e redes sociais, constatamos que cada avanço computacional que promete eficiência e escala também introduz novas camadas de abstração, opacidade e vieses potenciais. A prática historiográfica, consequentemente, é desafiada em seu núcleo: a capacidade do historiador de manter uma cadeia de justificação transparente desde a fonte até o argumento final.

A conclusão central que emerge desta análise é que a resposta a este desafio não reside em uma aceitação ingênua da tecnologia nem em uma recusa reativa a ela. O caminho para uma historiografia digital robusta e relevante passa pelo desenvolvimento de uma prática crítica e reflexiva. O papel do medievalista se expande. Ele não é mais apenas um intérprete de fontes, mas deve se tornar também um auditor de algoritmos, um crítico de modelos de dados e um curador consciente dos vieses embutidos em todo o processo tecnológico. A transparência metodológica, a documentação rigorosa das escolhas de modelagem e um diálogo

interdisciplinar constante deixam de ser opcionais para se tornarem componentes essenciais do próprio rigor histórico.

Em última análise, a IA força a disciplina a um saudável exercício de autoconsciência, obrigando-nos a revisitar e a explicitar os pressupostos que sempre guiaram nosso ofício. O futuro de uma pesquisa medieval inovadora e intelectualmente honesta dependerá de nossa habilidade de manejar estas poderosas ferramentas sem sucumbir às suas armadilhas epistemológicas. A tarefa é forjar uma abordagem que alie o poder computacional à profundidade hermenêutica, garantindo que a tecnologia sirva para ampliar nossa interrogação sobre o passado, e não para ditar as respostas.

Referências

AGUILAR, J.; et al. Multilingual Named Entity Recognition for Medieval Manuscripts. *In*: **Proceedings of the 3rd workshop on language technologies for historical and ancient languages** *(LT4HALA 2022)*, Marseille, 2022. p. 261-274.

AOUINTI, F.; EYHARABIDE, V.; FRESQUET, X.; BILLIET, F. Illumination detection in IIIF medieval manuscripts using deep learning. **Digital Medievalist,** v. 15, n. 1, p. 1–18, 2022. Disponível em: https://journal.digitalmedievalist.org/article/id/8073/. Acesso em: 18 ago. 2025.

BEATRICE, C. The challenges of HTR model training. **Journal of Data Mining and Digital Humanities**, 2023.

BIRNBAUM, D. J.; BONDE, S.; KESTEMONT, M. The Digital Middle Ages: An Introduction. **Speculum**, v. 92, n. S1, p. S1-S38, out. 2017. Disponível em: https://www.journals.uchicago.edu/doi/full/10.1086/694236. Acesso em 18 ago. 2025.

BRASIL, E. História digital: reflexões a partir da hemeroteca digital. **Estudos Históricos**, Rio de Janeiro, v. 33, n. 69, p. 531-551, 2020. Disponível em: https://periodicos.fgv.br/reh/article/view/79933. Acesso em: 18 ago. 2025.

BUCHNEA, E.; ELSANH, Z. Historical social network analysis: Advancing new directions for international business research. **Journal of World Business**, v. 57, n. 5, 2022.Acesso em: https://www.sciencedirect.com/science/article/pii/S096959312200018X. Acessoe m: 18 ago. 2025.

BURRELL, J. How the machine “thinks”: Understanding opacity in machine learning algorithms. **Big Data & Society**, v. 3, n. 1, 2016. Disponível em: https://journals.sagepub.com/doi/full/10.1177/2053951715622512. Acesso em: 18 ago. 2025.

CHASTANG, P.; AGUILAR, S. T.; TANNIER, X. A named entity recognition model for medieval Latin charters. **DHQ: Digital Humanities Quarterly**, v. 15, n. 4, 2021.Disponível em: https://orbilu.uni.lu/bitstream/10993/59645/1/000574.pdf. Acesso em: 18 ago. 2025.

CLAVERT, F. Sur trois textes: épistémologie de l'histoire et Humanités numériques. **L'histoire contemporaine à l'ère numérique**, 20 dez. 2013. Disponível em: https://histnum.hypotheses.org/1968. Acesso em: 04 ago. 2025.

COLETTO, G. M. **On Making in the Digital Humanities**: A study of digital-born scholarly outputs. Tese (Doutorado em Digital Humanities) – University College London, Londres, 2021. Disponível em: https://discovery.ucl.ac.uk/id/eprint/10166341/1/On-making-in-the-digital-humanities.pdf. Acesso em: 18 ago. 2025.

CRYMBLE, A. **Technology and the Historian**: Transformations in the Digital Age. Urbana: University of Illinois Press, 2021.

DE LANGHE, L.; DE CLERCQ, O.; HOSTE, V. Unsupervised Authorship Attribution for Medieval Latin using Transformer-Based Embeddings. *In*: **Proceedings of the third workshop on language technologies for historical and ancient languages (lt4hala) @ lrec-coling-2024**, Turin, 2024. p. 57-64.

DE VALERIOLA, S. A network approach to medieval documents: the heterogeneity of relationships. **Digital Medievalist,** v. 15, 2022. Disponível em: https://journal.digitalmedievalist.org/article/id/8070/. Acesso em: 18 ago. 2025.

DHAR, P. The carbon impact of artificial intelligence. **Nature Machine Intelligence**, v. 2, n. 8, p. 423–425, 2020. Disponível em: https://www.nature.com/articles/s42256-020-0219-9. Acesso em: 18 ago. 2025.

D'IGNAZIO, C.; KLEIN, L. F. **Data Feminism**. Cambridge, MA: MIT Press, 2020.

EHRMANN, M.; et al. Named Entity Recognition, Linking, and Discovery in Historical Documents and Settings: A Survey. **ACM Computing Surveys**, 2023.

FICKERS, A. The Annales School and the Digital Turn. *In*: ZAAGSMA, G. (ed.). **History in the Digital Age**. Berlin: De Gruyter Oldenbourg, 2021. p. 15-32.

FINNEMANN, N. O. Digital humanities and networked digital media. **MedieKultur: Journal of media and communication research**, v. 30, n. 57, p. 94-114, 2014. Disponível em: https://www.mediekultur.dk/article/view/15592. Acesso em: 18 ago. 2025.

GOLD, M. K. (ed.). **Debates in the Digital Humanities**. Minneapolis: University of Minnesota Press, 2012.

GONÇALVES, M. The hermeneutic problem posed by digital humanities. **Revista de Teoria da História**, v. 10, n. 1, 2022. Disponível em: https://revistas.ufg.br/teoria/article/view/70880. Acesso em: 18 ago. 2025.

GUPTA, A.; HU, Y.; MA, J.; LI, M. **AI's excessive water consumption threatens to drown out its benefits**. United Nations Sustainable Development Group, 2024.

HOCKEY, S. The History of Humanities Computing. *In*: SCHREIBMAN, S.; SIEMENS, R.; UNSWORTH, J. (ed.). **A Companion to Digital Humanities**. Oxford: Blackwell, 2004. p. 3-19.

KEMMAN, M. **Trading Zones of Digital History**. Berlin; Boston: De Gruyter, 2021.

MADI, B. et al. Automated dating of medieval manuscripts with a new dataset. *In*: MOUCHÈRE, H.; ZHU, A. (ed.). **Document Analysis and Recognition – ICDAR 2024 Workshops**. Cham: Springer, 2024.

MILLIGAN, I. **History in the Age of Abundance? How the Web Is Transforming Historical Research**. Montreal; Kingston: McGill-Queen's University Press, 2019.

NAWIYANTO; HANDAYANI, S. A.; Salindri, D. Annales School of History: Its Origins, Development and contributions. **IOSR Journal of Humanities and Social Science**, v. 27, n. 5, p. 39-45, 2022. Disponível em: https://www.iosrjournals.org/iosr-jhss/papers/Vol.27-Issue5/Ser-8/G2705083945.pdf. Acesso em: 18 ago. 2025.

MÜHLBERGER, G.; et al. Transforming scholarship in the archives through handwriting text recognition: Transkribus as a case study. **Journal of Documentation**, v. 75, n. 5, p. 954-976, 2019. Disponível em: https://www.emerald.com/jd/article/75/5/954/206914/Transforming-scholarship-in-the-archives-through. Acesso em: 18 ago. 2025.

NYHAN, J.; ORTOLJA-BAIRD, A. Data is never raw: ethics and biases in digital cultural heritage collections. **Science Museum Group Blog**, 20 jun. 2023.

PASQUALE, F. **The Black Box Society**: The Secret Algorithms that Control Money and Information. Cambridge, MA: Harvard University Press, 2015.

SANTOS, G. N. Por uma história digital: o ofício do historiador na era da internet. **Tempo e Argumento**, v. 13, n. 31, 2021.

SCHREIBMAN, S.; SIEMENS, R.; UNSWORTH, J. (ed.). **A Companion to Digital Humanities**. Oxford: Blackwell, 2004.

SOMMERSCHIELD, Thea; ASSAEL, Yannis; PAVLOPOULOS, John; STEFANAK, Vanessa; SENIOR, Andrew; DYER, Chris; BODEL, John; PRAG, Jonathan; ANDROUTSOPOULOS, Ion; DE FREITAS, Nando. Machine learning for ancient languages: a survey. **Computational Linguistics**, Cambridge, MA, v. 49, n. 3, p. 703-747, set. 2023. Disponível em: https://doi.org/10.1162/coli_a_00481. Acesso em: 18 ago. 2025.

STUTZMANN, D. et al. Integrated DH. Rationale of the HORAE Research Project. *In*: **ADHO - 2019 "COMPLEXITIES",** Utrecht, 2019.

UNIVERSIDADE DE SÃO PAULO. **Uma História Conectada do Mediterrâneo Medieval – Objetivos**. [2025]. Disponível em: https://sites.usp.br/historiaconectada/objetivos/. Acesso em: 04 ago. 2025.

ZBÍRAL, D.; SIKK, K.; SHAW, R. L. J. Bridging Digital History Methods and Source Criticism: A Research Agenda for the Study of Inquisition Records. *Acta Historica* **Tallinnensia,** v. 31, n. 1, p. 46-72, 2025. Disponível em: https://doi.org/10.3176/hist.2025.1.02. Acesso em: 18 ago. 2025.

Recebido em: 9 de abril de 2025
Aceito em: 23 de agosto de 2025