

Mediação algorítmica da evidência histórica: o HTR no Arquivo Nacional brasileiro

Algorithmic Mediation and the Production of Historical Evidence: HTR at the Brazilian National Archives

Carolina Moraes Souza

Doutora em História Contemporânea pela Universidade de Coimbra, Portugal
Investigadora do Centro de História da Sociedade e da Cultura da Universidade de Coimbra, Portugal
cmoraesouza@gmail.com
<https://orcid.org/0000-0002-1461-9600>
<http://lattes.cnpq.br/6889838456673995>

João Gabriel Moraes Souza

Doutor em Finanças e Métodos Quantitativos pela Universidade de Brasília, Brasil
Professor Substituto do Departamento de Engenharia de Produção da Universidade de Brasília, Brasil
joaogabrielmoura@yahoo.com.br
<https://orcid.org/0000-0003-0685-3082>
<http://lattes.cnpq.br/6798568258786751>

Resumo: O artigo examina os efeitos da incorporação de sistemas de Reconhecimento de Texto Manuscrito (HTR) sobre a produção da evidência histórica, a partir da noção de mediação algorítmica. Inserido no campo da História Digital, o estudo compreende o HTR como método textual situado e como infraestrutura de aceleração, capaz de reorganizar as condições de legibilidade, exploração e circulação dos documentos históricos. A discussão articula mediação técnica, regimes de temporalidade e crítica ao neopositivismo digital. A experimentação conduzida pelo IBICT em parceria com o Arquivo Nacional brasileiro serve de ancoragem empírica, evidenciando como a automação opera de forma diferencial sobre acervos heterogêneos e como os limites técnicos dos sistemas produzem hierarquias invisíveis de acesso ao passado. O caso demonstra que a reorganização algorítmica da evidência histórica constitui um problema epistemológico central da historiografia contemporânea.

Palavras-chaves: História Digital; Mediação algorítmica; Evidência histórica; Reconhecimento de Texto Manuscrito; Neopositivismo digital.

Abstract: This article examines the effects of incorporating Handwritten Text Recognition (HTR) systems on the production of historical evidence, drawing on the concept of algorithmic mediation. Situated within the field of Digital History, the study approaches HTR as a situated textual method and as an infrastructure of acceleration, capable of reorganizing the conditions of legibility, exploration, and circulation of historical documents. The discussion articulates technical mediation, regimes of temporality, and critiques of digital neo-positivism. The HTR experiment conducted by IBICT in partnership with the Brazilian National Archives serves as an empirical anchor, showing how automation operates differentially over heterogeneous collections and how the technical limits of recognition systems produce invisible hierarchies of access to the past. The case demonstrates that the algorithmic reorganization of historical evidence constitutes a central

Keywords: Digital History; Algorithmic mediation;

Introdução

A incorporação de tecnologias computacionais às práticas de pesquisa histórica constitui um processo de longa duração, cujos efeitos não se deixam reduzir à atual difusão de aplicações baseadas em Inteligência Artificial. Desde as primeiras experiências de informatização de acervos e de uso de métodos quantitativos nas Humanidades, historiadores vêm interrogando os impactos da formalização, da automação e da ampliação de escala sobre a produção do conhecimento histórico. Esses debates inaugurais já apontavam que a introdução de instrumentos técnicos altera condições fundamentais do ofício, incidindo sobre critérios de evidência, formas de interpretação e modalidades de mediação entre o pesquisador e as fontes (BUSA, 1980; COHEN; ROSENZWEIG, 2006).

No campo da História Digital, tais questões ganharam densidade à medida que arquivos e instituições de memória passaram a mobilizar métodos computacionais para lidar com grandes massas documentais manuscritas. Entre essas metodologias, os sistemas de Reconhecimento de Texto Manuscrito (*Handwritten Text Recognition* – HTR), desenvolvidos especificamente para documentos manuscritos, vêm ocupando posição estratégica. Ao permitir a conversão automatizada de manuscritos históricos em texto pesquisável, o HTR incide sobre um problema estrutural da investigação histórica: o déficit de legibilidade de acervos extensos, marcados por diversidade paleográfica, instabilidade gráfica e restrições materiais de acesso. A indexação e a busca textual reconfiguram práticas consolidadas de exploração documental, ampliando a escala e a velocidade da pesquisa.

A adoção dessas metodologias ocorre, contudo, em contextos institucionais concretos, atravessados por limitações materiais e humanas que condicionam seus usos e efeitos. No Arquivo Nacional brasileiro, a experimentação com sistemas de HTR está diretamente associada à necessidade de acelerar o tratamento de documentação manuscrita diante de um quadro estrutural de escassez de especialistas, com um número extremamente reduzido de paleógrafos

responsáveis pela transcrição e validação final dos documentos¹. Nesse cenário, o HTR opera como etapa intermediária de pré-processamento, voltada à produção de transcrições preliminares que antecedem o trabalho paleográfico, sem pretender substituí-lo.

O estudo de caso aqui analisado envolve testes de aplicação do HTR, em ambiente experimental, sobre tipologias documentais distintas, como os diários da Missão Cruls, caracterizados por escrita narrativa, científica e descritiva, e certidões administrativas relativas a imigração, nascimento e propriedade, marcadas por maior regularidade formal. Essa heterogeneidade documental permite observar como metodologias digitais baseadas em aprendizado de máquina respondem de maneira diferenciada a regimes históricos de escrita e a funções institucionais específicas. Ao mesmo tempo, evidencia que a aceleração do processo de transcrição introduz critérios implícitos de priorização documental, influenciando quais conjuntos avançam mais rapidamente no fluxo de tratamento, validação e disponibilização.

Partindo dessas questões, este artigo toma a experimentação com sistemas de Reconhecimento de Texto Manuscrito no Arquivo Nacional brasileiro como estudo de caso para refletir sobre as possibilidades e os limites das metodologias digitais na pesquisa histórica. Ao examinar o HTR enquanto infraestrutura de aceleração e mediação do trabalho historiográfico, sustenta-se que a ampliação da legibilidade e da navegabilidade dos arquivos reconfigura práticas de pesquisa e critérios de constituição da evidência, recolocando no centro do debate a necessidade de uma mediação historiográfica consciente, capaz de articular inovação técnica, crítica de fontes e responsabilidade interpretativa.

1. História Digital, mediação e regimes de temporalidade

A denominada virada digital tem produzido transformações profundas no campo da História, incidindo simultaneamente sobre as tecnologias de produção e acesso à informação e sobre os fundamentos epistemológicos que sustentam a prática historiográfica. A digitalização

¹ O estudo de caso refere-se a um projeto experimental realizado no Arquivo Nacional, em parceria com o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), voltado à avaliação do uso de sistemas de Reconhecimento de Texto Manuscrito (*Handwritten Text Recognition* – HTR) como etapa de pré-processamento da transcrição de documentos históricos manuscritos, diante da escassez estrutural de especialistas em paleografia na instituição.

de acervos, a multiplicação de fontes nascidas digitais e a consolidação de infraestruturas computacionais de armazenamento, busca e circulação de dados alteraram de modo substantivo as condições de investigação do passado. Como observa Serge Noiret (2015), a conversão de arquivos físicos em arquivos digitais e a incorporação de novos suportes documentais, como bases de dados, plataformas colaborativas, redes sociais e ambientes audiovisuais, ampliaram o conceito de fonte histórica e impuseram aos historiadores a necessidade de revisitar suas metodologias e categorias analíticas.

Nesse novo cenário, o trabalho historiográfico passa a operar em meio a uma constelação de ferramentas computacionais, softwares de análise textual, sistemas de metadados, algoritmos de busca e repositórios digitais, que reorganizam a dinâmica da pesquisa. Essas transformações não se limitam a ganhos de escala ou de eficiência, pois afetam diretamente o modo como o passado é acessado, ordenado e interpretado. A virada digital introduz, assim, questões epistemológicas centrais: a preservação e a autenticidade das fontes digitais, a volatilidade da memória em contextos de rápida obsolescência tecnológica, e o papel da mediação algorítmica na construção das narrativas históricas.

A incorporação de métodos computacionais às Humanidades antecede o atual debate em torno da Inteligência Artificial. Desde a segunda metade do século XX, experiências de indexação, quantificação e análise textual já vinham suscitando reflexões sobre os efeitos da formalização técnica na produção do conhecimento histórico. A contribuição pioneira de Roberto Busa (1980) inscreve-se nesse momento inaugural ao demonstrar que a computação não atua apenas como instrumento auxiliar, mas como elemento capaz de reorganizar escalas de análise, critérios de evidência e percursos interpretativos. A História Digital consolida-se, assim, como um campo voltado à problematização das condições técnicas e epistemológicas do ofício historiográfico, mais do que à simples adoção de novas ferramentas.

Esse processo articula-se diretamente à noção de mediação. A informatização dos arquivos e a difusão de bases de dados ampliaram o acesso às fontes e instituíram novos regimes de organização da informação histórica, nos quais a intervenção técnica passa a desempenhar papel estruturante. Cohen e Rosenzweig (2006) chamaram atenção para o fato de que o uso de ferramentas digitais reconfigura as condições de produção do conhecimento ao introduzir camadas técnicas que intervêm na seleção, classificação e circulação dos documentos. Essa reconfiguração incide também sobre os regimes de temporalidade da escrita da história, na

medida em que desloca a relação entre experiência acumulada e expectativa interpretativa, reordenando os modos pelos quais o passado é mobilizado no presente da pesquisa (KOSELLECK, 2006).

A disponibilidade de mecanismos de busca, indexação e filtragem textual antecipa formas de inteligibilidade antes mesmo do contato aprofundado com o arquivo, orientando a formulação das perguntas, a delimitação dos recortes e a percepção do que aparece como relevante. A expectativa do pesquisador passa a ser moldada por regimes de recuperabilidade técnica, nos quais a possibilidade de localizar rapidamente termos, nomes ou padrões discursivos tende a estruturar o percurso investigativo desde seu início. Nesse contexto, o horizonte de expectativa da pesquisa deixa de ser construído apenas pela experiência acumulada do historiador e passa a ser co-produzido por dispositivos técnicos que sugerem recorrências, lacunas e conexões possíveis.

Esse deslocamento tem implicações diretas para a relação entre tempo da pesquisa e tempo histórico. A antecipação técnica do acesso ao *corpus* tende a comprimir etapas tradicionais de exploração lenta e progressiva da documentação, reconfigurando o ritmo do trabalho historiográfico. A investigação passa a operar em um regime no qual a descoberta do material e a formulação das hipóteses ocorrem de modo simultâneo, mediadas por operações automatizadas que organizam o arquivo segundo critérios de visibilidade e eficiência. Tal dinâmica não elimina o gesto interpretativo, mas o reinscreve em uma temporalidade marcada pela aceleração e pela expectativa de resposta imediata. O desafio historiográfico, nesse cenário, consiste em reconhecer que o horizonte de expectativa da pesquisa já não se constitui apenas no interior da disciplina, mas emerge da interação contínua entre práticas interpretativas e infraestruturas digitais que organizam o acesso ao passado.

Nesse sentido, a História Digital não pode ser compreendida como simples atualização técnica da historiografia, mas como um espaço de rearticulação das temporalidades do ofício. A mediação computacional introduz formas específicas de antecipação, seleção e hierarquização que deslocam o ponto de partida da interpretação. No contexto brasileiro, essas questões assumem densidade particular. A adoção de metodologias digitais ocorre em ambientes marcados por desigualdades institucionais, limitações infraestruturais e assimetrias no acesso aos acervos. Como argumenta Bonaldo (2023), a incorporação de métodos digitais à pesquisa histórica precisa ser pensada a partir das condições materiais de produção do conhecimento,

sob pena de reforçar hierarquias já existentes entre instituições, centros de pesquisa e tipos documentais. A mediação técnica, longe de ser homogênea, reflete disputas geopolíticas e epistemológicas que atravessam o campo das Humanidades Digitais.

A crítica cultural proposta por Fiormonte (2012) contribui decisivamente para esse debate ao evidenciar que protocolos, padrões linguísticos e infraestruturas digitais são historicamente situados. Ao problematizar a pretensão de universalidade dos instrumentos técnicos, o autor demonstra que a metodologia digital carrega marcas culturais, políticas e linguísticas que moldam agendas de pesquisa e regimes de visibilidade. Essa perspectiva dialoga com a proposta de desobediência epistêmica formulada por Mignolo (2008), ao reivindicar a historicidade das categorias analíticas e a necessidade de deslocar o olhar para além dos centros hegemônicos de produção do conhecimento.

É nesse quadro que a História do Tempo Presente oferece um enquadramento particularmente fecundo. Ao afirmar o presente como espaço legítimo de investigação historiográfica, essa abordagem enfatiza as disputas narrativas, as experiências vividas e as condições contemporâneas de produção da memória. Arendt e Lohn (2022) destacam que uma historiografia do tempo presente exige atenção às experiências e discursos em curso, bem como uma abertura epistemológica capaz de incorporar o Sul Global como perspectiva analítica. A centralidade do presente, nesse contexto, não elimina a historicidade, mas torna visíveis as condições temporais e técnicas que estruturam o acesso ao passado.

A articulação entre História Digital, regimes de temporalidade e mediação historiográfica permite compreender que as metodologias computacionais operam no interior de transformações mais amplas do campo histórico. A ampliação de escala, a reorganização dos tempos da pesquisa e a redefinição dos critérios de evidência configuram um novo horizonte de práticas e problemas. É nesse cenário que se inscrevem as metodologias de reconhecimento automático de textos manuscritos, cuja análise demanda considerar seus resultados empíricos, as formas de mediação e temporalidade que introduzem no trabalho historiográfico.

2. HTR como método textual e infraestrutura de aceleração

A noção de ferramenta como operação historiográfica, proposta por Silveira (2022), permite compreender esse deslocamento de forma mais precisa. Ao processar textos históricos, a ferramenta não se limita à execução de comandos técnicos, mas materializa pressupostos sobre regularidade, relevância e recorrência. O tratamento computacional do documento implica decisões prévias acerca do que deve ser segmentado, estabilizado ou descartado, incorporando ao método escolhas que incidem diretamente sobre a produção do conhecimento histórico. Nesse sentido, a ferramenta atua como uma teoria em funcionamento, ainda que suas premissas permaneçam frequentemente implícitas ao pesquisador.

A incorporação do HTR ao fluxo da pesquisa histórica aprofunda essa formulação ao evidenciar que determinadas perguntas se tornam formuláveis porque o texto manuscrito foi previamente convertido em dado processável. O reconhecimento automático amplia o corpus disponível e reorganiza a economia cognitiva da investigação, favorecendo questões orientadas por recorrência lexical, identificação de entidades e padrões discursivos, reconfigurando a escala possível da pesquisa histórica. O HTR atua, assim, como condição indireta de possibilidade da pergunta histórica, influenciando o tipo de problema que emerge como investigável e deslocando o eixo da pesquisa para aquilo que pode ser tecnicamente reconhecido, comparado e re combinado.

Esse deslocamento incide sobre a distinção entre descoberta exploratória e descoberta orientada. Em regimes tradicionais de leitura paleográfica, o arquivo se apresenta como campo opaco, no qual a formulação das hipóteses decorre de um contato frequentemente lento e contingente com a documentação. A pesquisa mediada por HTR favorece percursos orientados por sugestões técnicas de relevância, nos quais a exploração do corpus ocorre em diálogo contínuo com resultados provisórios de busca e reconhecimento. A ferramenta não substitui a interpretação, mas participa ativamente da configuração das heurísticas que orientam o processo investigativo desde seus momentos iniciais.

Essa dinâmica pode ser interpretada à luz das reflexões de Michel de Certeau (1982) sobre a operação historiográfica. Ao analisar os efeitos da formalização técnica no trabalho histórico, Certeau observa que determinados dispositivos produzem coerências antes mesmo da interpretação propriamente dita, ao organizar o campo do possível por meio de modelos e

procedimentos antecipados. No caso das metodologias digitais, esse efeito torna-se particularmente visível pois o reconhecimento automático estrutura o acesso ao documento, delimitando aquilo que poderá ser comparado, indexado e articulado, e reorganizando o gesto interpretativo a partir dessas condições iniciais.

Os sistemas de HTR operam por meio de modelos de aprendizado de máquina (*machine learning*) treinados para identificar padrões visuais recorrentes e convertê-los em sequências textuais processáveis. O HTR não realiza uma leitura do texto em sentido hermenêutico, ele executa uma tradução técnica, na qual traços gráficos são transformados em unidades reconhecíveis por sistemas computacionais. Essa tradução implica uma transformação substantiva da escrita histórica. O manuscrito, marcado por instabilidades, variações caligráficas e dependência do contexto material de produção, é convertido em texto operável, isto é, em dado passível de indexação, busca e recombinação. Nesse processo, ambiguidades são parcialmente estabilizadas, variações são normalizadas e irregularidades são absorvidas por modelos probabilísticos. A legibilidade resultante não corresponde à restituição da singularidade material do documento, mas à produção de uma forma específica de inteligibilidade, ajustada às exigências das infraestruturas digitais.

A literatura recente sobre processamento automático de documentos históricos tem destacado o caráter situado dessas operações. Estudos dedicados ao reconhecimento de entidades nomeadas e à classificação textual em corpora históricos indicam que o desempenho dos sistemas depende fortemente das características materiais, gráficas e linguísticas dos documentos, bem como das decisões envolvidas no treinamento dos modelos (EHRMANN *et al.*, 2021).² O HTR, portanto, opera de maneira diferencial conforme os regimes históricos de escrita e as tipologias documentais às quais é aplicado.

No campo arquivístico, essa constatação tem sido formulada a partir de outra chave analítica. Ao discutir a aplicação do HTR em arquivos portugueses, Ana Margarida Dias da Silva (2023) observa que a digitalização extensiva, quando dissociada de mecanismos de leitura e indexação textual, produz um acesso limitado à informação. A simples disponibilização de

² Nos arquivos em língua portuguesa, o caráter situado do processamento automático de documentos históricos também tem sido discutido no âmbito da paleografia digital e das Humanidades Digitais. Estudos sobre o uso de sistemas de HTR, em especial da plataforma Transkribus, destacam a dependência dos modelos em relação à materialidade documental, às especificidades gráficas da escrita manuscrita e às decisões metodológicas envolvidas no treinamento, reafirmando o papel central do saber paleográfico e arquivístico na mediação entre automação e crítica das fontes. Ver mais em Lose *et al.* (2024). Transkribus: uma ferramenta de paleografia digital mediando pesquisas em fontes inquisitoriais. *Revista LaborHistórico*.

imagens digitais mantém os documentos legíveis apenas no plano visual, preservando sua opacidade no plano da pesquisa. O HTR emerge, nesse contexto, como tentativa de reorganizar o acesso aos acervos, deslocando a leitura para uma etapa posterior e reconfigurando a temporalidade da investigação.

A conversão do manuscrito em texto processável reorganiza, assim, os modos de exploração do arquivo. A possibilidade de busca textual desloca a investigação para práticas orientadas por recorrência, palavras-chave e identificação automática de entidades. O *corpus* deixa de ser delimitado exclusivamente por critérios prévios de seleção e passa a emergir de interações sucessivas entre pergunta, reconhecimento técnico e descoberta documental.

A centralidade da busca textual introduz uma diferença substantiva em relação à leitura sequencial tradicional. Enquanto a leitura paleográfica pressupõe uma imersão contínua no documento, orientada por sua materialidade, linearidade e contexto de produção, a busca opera por descontinuidade, selecionando fragmentos a partir de termos, entidades ou padrões previamente reconhecidos. O arquivo deixa de ser percorrido como narrativa potencial e passa a ser acessado como campo de ocorrências, no qual o sentido emerge da recorrência e da comparação entre fragmentos dispersos. A narrativa histórica construída a posteriori carrega os efeitos desse regime, tendendo a privilegiar regularidades e padrões reconhecíveis em detrimento de zonas menos legíveis ou menos compatíveis com os dispositivos de busca.

O desafio metodológico reside justamente na articulação entre esses dois regimes de leitura: a imersão lenta e contextual do trabalho paleográfico e a exploração orientada pela busca, reconhecendo que ambos produzem formas distintas de acesso ao documento e que seus efeitos se prolongam na escrita e na argumentação historiográfica. Esse modo de exploração altera o ritmo e a sequência do trabalho historiográfico. A leitura extensiva e sequencial cede espaço a procedimentos descontínuos, baseados em aproximações sucessivas, verificações rápidas e retornos constantes ao arquivo. O tempo da investigação passa a ser atravessado por operações de reconhecimento e filtragem que reorganizam a relação entre exploração, seleção e interpretação. A aceleração que daí decorre não se apresenta como valor normativo, mas como efeito estrutural de um método orientado pela recuperabilidade textual. Essa dinâmica sugere que a legibilidade algorítmica não incide de modo uniforme sobre os acervos. Na medida em que o desempenho dos sistemas de HTR depende das características materiais, gráficas e tipológicas dos documentos, é de se esperar que conjuntos produzidos segundo regimes históricos de

escrita mais padronizados respondam de maneira diferente à automação do que manuscritos marcados por maior variação caligráfica e complexidade narrativa. O Arquivo Nacional brasileiro oferece, nesse sentido, um recorte analítico preciso, um acervo marcado por heterogeneidade real, sobre o qual a automação opera de forma desigual e verificável.

Compreender o HTR como método textual e infraestrutura de aceleração requer situá-lo no interior das transformações contemporâneas do ofício historiográfico. O deslocamento do início da investigação para operações técnicas de reconhecimento e busca reorganiza a leitura, a constituição do corpus e a temporalidade da pesquisa. A explicitação dessas mediações permite analisar os filtros e escolhas que sustentam a legibilidade algorítmica e delineia o terreno conceitual a partir do qual se desenvolve, na seção seguinte, a reflexão sobre os regimes temporais e epistemológicos implicados na reorganização algorítmica da evidência histórica.

3. Atualismo, neopositivismo digital e evidência histórica

A historiografia recente³ tem recorrido à noção de atualismo para caracterizar o regime de aceleração que atravessa os processos de produção, circulação e recuperação da informação no mundo contemporâneo. Trata-se de uma forma histórica de estruturação do tempo que incide diretamente sobre a prática historiográfica, ao reconfigurar as articulações entre passado, presente e futuro e ao redefinir os critérios de relevância e inteligibilidade do conhecimento histórico.

Na formulação de Pereira, Nicodemo e Araújo (2022), o atualismo descreve um modo de funcionamento no qual o presente assume centralidade como instância de validação historiográfica. O passado passa a ser mobilizado segundo parâmetros definidos pelas exigências cognitivas, políticas e técnicas do tempo presente, orientadas pela atualização constante, pela rapidez da resposta e pela circulação intensiva de sentidos. Nesse regime, o presente organiza retrospectivamente a leitura do passado, convertendo-se em horizonte

³ A noção de atualismo foi formulada e desenvolvida sobretudo por Valdei Lopes de Araújo e Mateus Henrique de Faria Pereira, no âmbito da teoria da história no Brasil, como categoria destinada a compreender um regime temporal marcado pela centralidade da atualização e pela reconfiguração das articulações entre passado, presente e futuro. As formulações centrais do conceito encontram-se entre outros em: (ARAÚJO e PEREIRA, 2019); (PEREIRA e ARAÚJO, 2016); (PEREIRA e ARAÚJO, 2017)

normativo que condiciona a seleção, a interpretação e a visibilidade das evidências históricas no interior da pesquisa.

A noção de presentismo, formulada por François Hartog (2013), oferece um instrumental complementar para a compreensão dessa dinâmica. Ao descrever um regime de historicidade no qual o presente se torna dominante, Hartog evidencia a absorção progressiva do passado e do futuro em uma temporalidade marcada pela urgência, pela simultaneidade e pela redução do distanciamento histórico. Esse enquadramento permite compreender como o presente deixa de operar como ponto de passagem entre experiências pretéritas e expectativas futuras e passa a funcionar como tempo central de referência e validação, reconfigurando as condições de inteligibilidade do passado.

O atualismo pode ser compreendido como uma expressão específica desse regime de historicidade, intensificada pelas infraestruturas digitais e pela lógica da atualização contínua. Sistemas de busca, indexação e recuperação automatizada reorganizam o acesso ao passado segundo critérios de pertinência imediata, eficiência técnica e visibilidade operacional. A historicidade não se dissolve, mas passa a ser mediada por dispositivos que privilegiam aquilo que pode ser rapidamente mobilizado, reprocessado e reinscrito nos fluxos do presente, produzindo uma relação com o passado marcada pela recorrência funcional e pela reatualização permanente dos sentidos.

A esse quadro soma-se a contribuição de Hans Ulrich Gumbrecht (2015), para quem a experiência contemporânea do tempo é marcada pela expansão de um presente que já não se orienta prioritariamente pela narrativa histórica, mas pela simultaneidade, pela presença e pela saturação de estímulos. No que Gumbrecht (2015) denomina “presente amplo”, o passado não é propriamente superado nem o futuro claramente projetado, mas ambos permanecem disponíveis como repertório mobilizável no agora. Em contextos mediados por dispositivos digitais, essa experiência tende a se intensificar, produzindo formas de relação com o passado menos orientadas pela distância histórica e mais pela imediaticidade da presença técnica.

Essa transformação incide diretamente sobre o estatuto da evidência histórica. A evidência deixa de emergir prioritariamente da leitura prolongada, da resistência material do documento e da reconstrução contextual de seus sentidos, passando a ser mobilizada a partir de sua capacidade de circular nos regimes técnicos de busca e recuperação. A aceleração não se limita a encurtar o tempo da investigação, pois ela também reorganiza os critérios de relevância,

favorecendo formas documentais mais compatíveis com as infraestruturas em operação e relegando outras a zonas de menor visibilidade.

É nesse ponto que a crítica ao neopositivismo digital, formulada por Gil (2024), adquire densidade analítica. Diferentemente do positivismo clássico, que se sustenta na crença explícita na neutralidade dos fatos, o neopositivismo digital está centrado na naturalização dos procedimentos técnicos que produzem os dados. Algoritmos, modelos estatísticos e sistemas automatizados passam a operar como instâncias implícitas de autoridade, cujos resultados são frequentemente aceitos como evidência suficiente, sem que os pressupostos epistemológicos que os estruturam sejam tornados visíveis. A mediação técnica tende a desaparecer do horizonte analítico, enquanto os produtos da automação se apresentam como dados imediatamente disponíveis ao conhecimento histórico.

A articulação entre atualismo e neopositivismo digital produz um efeito específico sobre a hierarquização da evidência. Aquilo que é reconhecido, indexado e recuperável tende a adquirir estatuto privilegiado no processo investigativo, enquanto zonas de opacidade algorítmica permanecem à margem. Não se trata de uma exclusão deliberada, mas de um efeito estrutural das infraestruturas que organizam o acesso ao arquivo. A evidência histórica passa a ser valorizada por sua compatibilidade com regimes de aceleração e busca, ajustando-se às temporalidades do presente da pesquisa.

Esse deslocamento torna-se particularmente visível quando considerado a partir da experiência concreta do ofício historiográfico no tempo presente. É nesse ponto que as reflexões de Dilton Maynard (2022) oferecem uma contribuição decisiva. Ao analisar a relação entre História, internet e tempo presente, o autor chama atenção para a vivência do historiador em um ambiente marcado simultaneamente pelo excesso de registros e pela instabilidade da preservação. O problema já não é a escassez de fontes, mas a dificuldade de sustentar práticas de crítica, seleção e contextualização em um cenário caracterizado pela abundância informacional e pela obsolescência acelerada dos suportes.

Para Maynard (2022), a aceleração se constitui como uma experiência cognitiva e epistemológica que atravessa o trabalho histórico. O historiador do presente opera sob a pressão constante da atualização, da circulação rápida e da expectativa de resposta imediata, o que tende a fragilizar os tempos longos da análise e da reflexão. A impossibilidade de abarcar a totalidade dos registros digitais não é um problema inédito, mas a naturalização dessa

impossibilidade como condição normal da pesquisa altera profundamente o modo como a evidência é reconhecida e mobilizada.

Nesse contexto, a autoridade historiográfica passa a ser tensionada por novas formas de validação. A crítica documental clássica, fundada na análise da proveniência, da materialidade e das condições de produção dos registros, passa a coexistir com critérios apoiados na eficácia técnica dos sistemas de acesso. A evidência prontamente recuperável adquire centralidade, enquanto documentos menos compatíveis com os dispositivos técnicos tendem a permanecer fora dos percursos analíticos mais frequentes. O risco não reside na ausência de acesso, mas na conversão da eficiência técnica em critério tácito de validação do conhecimento histórico.

A reorganização algorítmica da evidência histórica interpela, assim, o ofício do historiador em sua dimensão epistemológica. Tornar visíveis as mediações técnicas que estruturam o acesso ao passado não implica rejeitar a automação, mas reconhecer que a evidência não emerge espontaneamente dos dados. Ela é produzida por meio de escolhas técnicas, institucionais e conceituais que precisam ser explicitadas e discutidas. É nesse movimento que se preserva a espessura histórica da evidência, mesmo em ambientes marcados pela aceleração, pelo excesso informacional e pela instabilidade temporal.

Ao deslocar o debate das promessas de eficiência para as condições de produção do conhecimento histórico, esta seção estabelece o quadro conceitual necessário para compreender o Arquivo Nacional brasileiro como estudo de caso.

4. O Arquivo Nacional brasileiro como estudo de caso da mediação algorítmica da evidência

A experimentação com sistemas de Reconhecimento de Texto Manuscrito no Arquivo Nacional brasileiro permite observar, em escala institucional concreta, como a mediação algorítmica incide sobre a produção da evidência histórica em contextos marcados por forte heterogeneidade documental. O caso não funciona apenas como ilustração da incorporação de metodologias digitais, mas como recorte analítico capaz de tornar visíveis os efeitos diferenciais da automação em acervos historicamente desiguais e tipologicamente instáveis.

O projeto desenvolvido em parceria entre o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e o Arquivo Nacional, executado entre janeiro de 2023 e outubro de 2025, insere-se em um cenário institucional caracterizado por constrangimentos estruturais bem definidos: a presença de extensos conjuntos manuscritos, a diversidade de tipologias documentais e a limitação do número de profissionais especializados em paleografia responsáveis pela transcrição, revisão e validação final dos documentos. O HTR foi integrado ao fluxo institucional como etapa de pré-processamento, voltada à produção de transcrições preliminares e à ampliação da navegabilidade dos fundos, sem pretensão de substituir o trabalho especializado (IBICT, 2025). O resultado operacional do projeto foi a entrega do protótipo IBYCT (*Intelligent Byte and Character Transcription*) ao Arquivo Nacional, desenvolvido e treinado em infraestrutura dedicada da própria instituição, e acompanhado de sessão de capacitação técnica junto à equipe do AN em novembro de 2024. A automação aparece, assim, como resposta institucional à escala e à complexidade do acervo, ancorada em condições materiais e humanas concretas. Essa integração reorganiza as condições de acesso ao arquivo.

A legibilidade dos documentos passa a ser produzida por uma cadeia de operações técnicas, reconhecimento automático, segmentação textual e indexação, que antecede a leitura historiográfica. A evidência histórica emerge, portanto, já mediada por critérios de recuperabilidade definidos no interior da infraestrutura técnica, deslocando o ponto inicial da pesquisa para um estágio anterior à interpretação propriamente dita.

Os testes foram realizados sobre dois conjuntos documentais com características históricas e tipológicas distintas. O primeiro compreende os diários de campo do engenheiro Hastimphilo de Moura, produzidos durante a Comissão Exploradora do Planalto Central entre 1892 e 1894, expedição que resultaria na demarcação do quadrilátero destinado à futura capital federal. Trata-se de uma escrita narrativa, científica e descritiva, produzida por um único autor ao longo do tempo, o que introduz variação caligráfica interna associada ao ritmo e às condições de cada expedição. O Diário número 1 foi utilizado para treinamento dos modelos, enquanto o Diário número 2, contextualmente similar, mas distinto em conteúdo, serviu como conjunto de teste, permitindo avaliar a capacidade de generalização dos algoritmos (IBICT, 2025). O segundo conjunto é formado por certidões do Arquivo Nacional, como registros de propriedade, nascimento e outros documentos de natureza administrativa e jurídica, produzidos por

múltiplos redatores em contextos institucionais distintos e ao longo de diferentes períodos históricos. Essa heterogeneidade de mãos, funções e épocas de produção caracteriza um regime de escrita radicalmente diferente do primeiro conjunto, ainda que ambos sejam manuscritos históricos em língua portuguesa. A seleção deliberada dessas duas tipologias permitiu observar como os sistemas de HTR respondem a regimes históricos de escrita com lógicas de produção documental distintas, tornando o experimento um caso analiticamente produtivo para além de sua função preservacionista.

Os resultados obtidos tornam visível a incidência diferencial da automação sobre os dois conjuntos. O modelo baseado em arquitetura CRNN, treinado com os diários da Missão Cruls, alcançou uma Taxa de Erro de Caracteres (CER) de 0,100, indicando que aproximadamente 90% dos caracteres foram reconhecidos corretamente. O mesmo modelo, treinado com as certidões do Arquivo Nacional, apresentou CER de 0,205, o dobro do erro verificado no conjunto anterior (IBICT, 2025). Esse contraste é historiograficamente significativo porque inverte uma expectativa intuitiva: os documentos administrativos, produzidos segundo fórmulas relativamente estáveis e reiterativas, demonstraram maior dificuldade de reconhecimento do que os diários narrativos. A explicação reside não na função dos documentos, mas em seu regime de produção gráfica: enquanto os diários apresentam a escrita de um único autor ao longo do tempo, as certidões reúnem múltiplos redatores com caligrafias distintas, produzidas em períodos históricos diferentes, o que amplifica a variabilidade visual com a qual o modelo precisa lidar. A regularidade funcional do documento não garante, portanto, regularidade gráfica e é essa distinção que os algoritmos de reconhecimento precisam enfrentar. O dado do CER inicial de 0,62⁴ nas certidões, registrado antes do retreinamento que envolveu ajustes no pipeline de pré-processamento e na curadoria dos dados de treinamento, reforça que o desempenho dos sistemas de HTR depende tanto das decisões técnicas sobre os dados quanto da arquitetura dos modelos (IBICT, 2025). A aceleração promovida pelo HTR não se distribui, portanto, de forma homogênea no interior dos fundos. Assim, certos conjuntos tornam-se imediatamente recuperáveis, enquanto outros permanecem dependentes de mediações

⁴ O projeto testou dois modelos de arquiteturas distintas: CRNN (*Convolutional Recurrent Neural Network*, rede neural que combina processamento visual e sequencial) e TrOCR (modelo baseado em arquitetura *Transformer*, originalmente desenvolvido para reconhecimento óptico de caracteres). Os resultados apresentados referem-se ao modelo CRNN, cujo desempenho foi o objeto de análise do relatório técnico disponível (IBICT, 2025). A comparação sistemática entre arquiteturas, embora relevante do ponto de vista técnico, ultrapassa o escopo historiográfico deste artigo.

posteriores, produzindo ritmos diferenciados de acesso e explorabilidade no interior do mesmo acervo.

Um dos limites identificados durante a validação operacional do protótipo IBYCT oferece um exemplo preciso do que significa opacidade algorítmica em contexto arquivístico concreto. Imagens contendo selos, elemento comum em documentos administrativos e jurídicos, provocavam falha de segmentação no sistema, que retornava o erro "*No Lines were predicted by the model*", sem produzir transcrição nem sinalizar erro mensurável ao usuário (IBICT, 2025). O documento existia, estava digitalizado e era visualmente legível, mas permanecia fora do alcance da busca textual. Esse tipo de limite não aparece nas métricas de CER ou WER (*Word Error Rate*), que pressupõem que o modelo produziu alguma saída avaliável. Trata-se de um silêncio técnico: o sistema simplesmente não reconhece que há texto a ser lido. Para o historiador, esse fenômeno tem implicações diretas sobre a constituição do corpus pesquisável, pois os documentos afetados por essas limitações tendem a permanecer invisíveis nos circuitos de busca, não por decisão interpretativa, mas como efeito estrutural da infraestrutura técnica. A crítica ao neopositivismo digital, aplicada a esse caso concreto, revela uma dimensão que as métricas de desempenho tendem a obscurecer: o problema da evidência ausente. Mais do que a confiança excessiva nos resultados produzidos, o que está em jogo é a invisibilidade dos documentos que o sistema simplesmente não processa.

Essa dinâmica dialoga diretamente com as discussões sobre mediação historiográfica desenvolvidas nas seções anteriores. A reorganização algorítmica do acesso não redefine o valor histórico dos documentos, mas condiciona sua circulação analítica no presente da pesquisa. A evidência que se torna prontamente recuperável tende a ser mobilizada nos estágios iniciais da investigação, enquanto outros conjuntos permanecem menos visíveis em função de sua menor compatibilidade com regimes de aceleração técnica, o que incide sobre os percursos efetivos da pesquisa histórica.

O estudo de caso pode ser compreendido, assim, à luz dos regimes de temporalidade que atravessam as práticas arquivísticas e historiográficas contemporâneas. A aceleração promovida pelo HTR responde a demandas institucionais de gestão documental e ampliação do acesso, ao mesmo tempo em que reorganiza as expectativas em relação ao tempo da pesquisa histórica. O arquivo passa a operar segundo uma temporalidade híbrida, na qual a rapidez da indexação

convive com a lentidão da leitura crítica, produzindo um descompasso que precisa ser continuamente administrado pelo historiador.

Esse cenário também ilumina os riscos apontados pela crítica ao atualismo e ao neopositivismo digital. A eficiência técnica dos sistemas de reconhecimento pode favorecer a naturalização dos resultados produzidos, deslocando para segundo plano a reflexão sobre os filtros, modelos e decisões que sustentam a legibilidade algorítmica (GIL, 2024). No caso analisado, o bom desempenho do HTR torna ainda mais necessária a explicitação de seus pressupostos, evitando que a eficácia operacional se converta em critério tácito de autoridade da evidência histórica.

O Arquivo Nacional brasileiro, enquanto instituição histórica concreta, torna visível a articulação entre mediação técnica, heterogeneidade documental e produção da evidência. O estudo de caso demonstra que a incorporação do HTR não se reduz a uma questão de inovação tecnológica, mas envolve escolhas institucionais que afetam o ritmo, a escala e os percursos da pesquisa histórica. Ao evidenciar os efeitos diferenciais da automação sobre acervos heterogêneos, o caso contribui para compreender como a reorganização algorítmica da evidência se torna um problema central da historiografia contemporânea, exigindo uma mediação historiográfica atenta tanto às fontes quanto às infraestruturas que as tornam legíveis.

Considerações finais

A incorporação de sistemas de Reconhecimento de Texto Manuscrito às práticas arquivísticas e historiográficas altera de modo significativo as condições de acesso, legibilidade e mobilização da evidência histórica em ambientes mediados por infraestruturas digitais. Mais do que uma ampliação técnica do alcance dos acervos manuscritos, trata-se de um processo que incide diretamente sobre dimensões centrais do ofício do historiador, ao reconfigurar os pontos de entrada no arquivo e os percursos possíveis da pesquisa.

Ao longo deste artigo, sustentou-se que o HTR deve ser compreendido como um método textual situado e como uma infraestrutura de aceleração. Sua eficácia técnica desloca os problemas clássicos da crítica historiográfica para um estágio anterior à interpretação, no qual a evidência passa a ser produzida e organizada por cadeias técnicas de reconhecimento,

segmentação, indexação e busca. A conversão do manuscrito em texto processável institui uma camada preliminar de mediação que condiciona os ritmos da investigação e influencia a forma como os documentos são explorados no presente da pesquisa.

Essa reorganização da evidência insere-se em regimes contemporâneos de temporalidade marcados pela aceleração e pela centralidade do presente. A articulação entre atualismo e mediação algorítmica permite compreender como o passado é continuamente reprocessado a partir das demandas do agora, orientado por infraestruturas que privilegiam rapidez, recorrência e navegabilidade. As durações históricas não são suprimidas, mas passam a conviver com temporalidades técnicas que comprimem expectativas e reorganizam os tempos do trabalho intelectual.

O estudo de caso do Arquivo Nacional brasileiro torna esses processos particularmente visíveis. A experimentação com HTR, realizada em um contexto caracterizado pela heterogeneidade dos fundos, pela diversidade tipológica e por constrangimentos institucionais, evidencia que a aceleração técnica opera de forma desigual no interior do acervo. Certos conjuntos documentais tornam-se rapidamente exploráveis, enquanto outros permanecem dependentes de mediações mais lentas e intensivas, o que produz assimetrias nos percursos de acesso e na circulação da evidência histórica.

Esse diagnóstico desloca o debate para o exame das condições concretas de produção da evidência em ambientes digitais. A crítica ao neopositivismo digital, desenvolvida na análise do caso, aponta para uma exigência metodológica concreta: quanto mais eficaz o sistema, menor a fricção que alerta o pesquisador para a mediação em curso.

A análise do HTR como infraestrutura de aceleração permite compreender a velocidade como operador epistemológico. A aceleração redefine expectativas institucionais, orienta práticas de pesquisa e interfere nos critérios de relevância histórica. A preservação da espessura da evidência exige estratégias conscientes de desaceleração analítica, capazes de reinscrever o texto processável às suas condições históricas de produção, às marcas materiais que o HTR estabilizou ou descartou, e aos silêncios que nenhum modelo probabilístico torna visíveis.

Nesse cenário, a mediação historiográfica assume papel ampliado. Em ambientes atravessados por algoritmos, a responsabilidade do historiador abrange tanto a crítica das fontes quanto a análise das infraestruturas que as tornam visíveis, pesquisáveis e operáveis. O reconhecimento do caráter situado das ferramentas digitais, assim como das desigualdades

documentais e institucionais que atravessam sua adoção, constitui condição para uma prática historiográfica atenta às hierarquias produzidas no interior do próprio processo de pesquisa.

A experimentação com o Reconhecimento de Texto Manuscrito no Arquivo Nacional brasileiro evidencia, desse modo, que a reorganização algorítmica da evidência histórica se configura como um dos desafios centrais da historiografia contemporânea. Esse processo intensifica questões clássicas sob novas configurações técnicas e institucionais, convocando uma reflexão renovada sobre temporalidade, mediação e autoridade do conhecimento histórico. É nesse terreno, estruturado pela interação entre aceleração técnica e crítica historiográfica, que se delineiam hoje as condições de produção do passado como objeto de conhecimento.

Referências bibliográficas

- ARAUJO, Valdei Lopes de; PEREIRA, Mateus Henrique de Faria (2019). *Atualismo 1.0: como a ideia de atualização mudou o século XXI*. 2 ed. Vitória: Milfontes; Mariana: Editora da SBTHH.
- ARENDT, Silvia; LOHN, Reinaldo (2022). Nas margens do tempo: a contribuição da UDESC para a história do tempo presente no Brasil. In: MÜLLER, Angélica; IEGELSKI, Francine (orgs.). *História do tempo presente: mutações e reflexões*. Rio de Janeiro: Editora FGV.
- BONALDO, Rodrigo Bragio (2023). História mais do que humana: descrevendo o futuro como atualização repetidora da inteligência artificial. *História (São Paulo)*, São Paulo, v. 42, e2023037. Disponível em: <<https://www.scielo.br/j/his/a/WbLs9ZhxbKKvD69smKVQ3Vh>>. Acesso em: 15 jan. 2026.
- BUSA, Roberto (1980). The annals of humanities computing: the *Index Thomisticus*. *Computers and the Humanities*, Dordrecht, v. 14, n. 2, pp. 83–90.
- CERTEAU, Michel de (1982). A operação historiográfica. In: CERTEAU, Michel de. *A escrita da história*. Rio de Janeiro: Forense Universitária.
- COHEN, Daniel J.; ROSENZWEIG, Roy (2006). *Digital history: a guide to gathering, preserving, and presenting the past on the web*. Philadelphia: University of Pennsylvania Press.
- EHRMANN, Maud; HAMDI, Ahmed; PONTES, Elvys Linhares; ROMANELLO, Matteo; DOUCET, Antoine (2021). Named entity recognition and classification on historical documents: a survey. *arXiv preprint*. Disponível em: <<https://arxiv.org/abs/2109.11406>>. Acesso em: 10 jan. 2026.
- FIORMONTE, Domenico (2012). Towards a cultural critique of digital humanities. *Historical Social Research*, Köln, v. 37, n. 3, pp. 59–76. Disponível em: <<https://www.jstor.org/stable/41636597>>. Acesso em: 25 jan. 2026.
- GIL, Tiago (2024). Sobre big data e neopositivismo digital na pesquisa em História. *Almanack*, Guarulhos, n. 36. Disponível em:

- <<https://www.scielo.br/j/alm/a/vbZwHsRnVdLBD7PWW9Qcjts>>. Acesso em: 12 jan. 2026.
- GUMBRECHT, Hans Ulrich (2015). *Nosso amplo presente: o tempo e a cultura contemporânea*. Tradução de Ana Isabel Soares. São Paulo: Editora da UNESP.
- HARTOG, François (2013). *Regimes de historicidade: presentismo e experiências do tempo*. Tradução de Andréa Stahel M. da Silva e Ivone C. Benedetti. Belo Horizonte: Autêntica.
- INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT) (2025). *Projeto de inteligência artificial para transcrição paleográfica de manuscritos*: relatório final. Brasília, DF. Relatório técnico institucional não publicado.
- KOSELLECK, Reinhart (2006). *Futuro passado: contribuição à semântica dos tempos históricos*. Tradução de Wilma Patrícia Maas e Carlos Almeida Pereira. Rio de Janeiro: Contraponto; Ed. PUC-Rio.
- LOSE, Alícia Duhá; SANTOS, João Guilherme Veloso Andrade dos; JESUS, Leonardo Coelho Marques de; MAGALHÃES, Livia Borges Souza; XAVIER, Lucia Furquim Werneck (2024). Transkribus: uma ferramenta de paleografia digital mediando pesquisas em fontes inquisitoriais. *Revista LaborHistórico*, Rio de Janeiro, v. 10, n. 1, pp. 1–22. Disponível em: <<https://revistas.ufrj.br/index.php/lh/article/view/63285>>. Acesso em: 15 jun. 2026.
- MAYNARD, Dilton (2022). Farejar carne humana em meio a bytes: a internet, a história e o tempo presente. In: MÜLLER, Angélica; IEGELSKI, Francine (orgs.). *História do tempo presente: mutações e reflexões*. Rio de Janeiro: Editora FGV.
- MIGNOLO, Walter D. (2008). Desobediência epistêmica: a opção descolonial e o significado de identidade em política. *Cadernos de Letras da UFF*, Niterói, n. 34, pp. 287–324. Disponível em: <<https://periodicos.uff.br/gragoata/article/view/33191>>. Acesso em: 12 jan. 2026.
- NOIRET, Serge (2015). História pública digital. In: ALMEIDA, Juniele Rabêlo de; ROVAL, Marta G. (orgs.). *Introdução à história pública*. São Paulo: Letra e Voz.
- PEREIRA, Mateus Henrique de Faria, ARAUJO, Valdei Lopes de (2016). Reconfigurações do tempo histórico: presentismo, atualismo e solidão na modernidade digital. *Revista da Universidade Federal de Minas Gerais*, Belo Horizonte, v. 23, n. 1–2, pp. 270–297.
- PEREIRA, Mateus Henrique de Faria; ARAUJO, Valdei Lopes de (2017). Actualismo y presente amplio: breve análisis de las temporalidades contemporáneas. *Desacatos*, Ciudad de México, n. 55, pp. 12–27.
- PEREIRA, Mateus Henrique de Faria; NICODEMO, Thiago Lima; ARAUJO, Valdei Lopes de (2022). A indústria das *fake news* como um problema historiográfico: atualismo e política em um presente agitado. In: MÜLLER, Angélica; IEGELSKI, Francine (orgs.). *História do tempo presente: mutações e reflexões*. Rio de Janeiro: Editora FGV, pp. 115–140.
- SILVA, Ana Margarida Dias da (2023). Transcrições automáticas nos arquivos distritais portugueses: acelerar o acesso à informação. *Cultura: Revista de História e Teoria das Ideias*, n. 41–42, pp. 269–286. Disponível em: <<https://revistas.fcsh.unl.pt/cultura/article/view/1226>>. Acesso em: 5 jan. 2026.
- SILVEIRA, Pedro Telles da (2022). O que é uma ferramenta historiográfica? *História da Historiografia*, v. 15, n. 40, pp. 219–231. Disponível em: <<https://www.scielo.br/j/hh/a/7cjzV4WPGsd3v7gWLB3jwTg>>. Acesso em: 5 jan. 2026.

Declaração de Uso de Inteligência Artificial Generativa

Os autores do artigo acima identificado declaram que ao longo da elaboração deste manuscrito foram utilizadas as seguintes ferramentas de inteligência artificial: Claude (Anthropic), ChatGPT (OpenAI) e Elicit, com a finalidade de: revisão de linguagem, aprimoramento do estilo textual e da clareza; apoio à revisão editorial e verificação de conformidade com normas de submissão, pesquisa bibliográfica e identificação de referências sobre o tema. Após o uso das ferramentas, os autores revisaram e editaram criticamente todo o conteúdo, assumindo total e integral responsabilidade pelo manuscrito publicado, inclusive por eventuais imprecisões ou incorreções.