Universidade Federal do Rio Grande do Sul
Faculdade de Biblioteconomia e Comunicação
Mestrado em Ciência da Informação

Leonardo Porto de Bittencourt Pereira

Inteligência Artificial e Patrimônio Documental: potencialidades do Transkribus no reconhecimento de escrita manuscrita.

Porto Alegre
2026

Leonardo Porto de Bittencourt Pereira

Inteligência Artificial e Patrimônio Documental: potencialidades do Transkribus no reconhecimento de escrita manuscrita.

Dissertação apresentada como requisito parcial à obtenção do título de Mestre em Ciência da Informação da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul.

Orientador: Prof. Dr. Moisés Rockembach

Porto Alegre

2026

FICHA CATALOGRÁFICA

CIP - Catalogação na Publicação

Pereira, Leonardo Porto de Bittencourt
Inteligência artificial e patrimônio documental: potencialidades do Transkribus no reconhecimento de escrita manuscrita. / Leonardo Porto de Bittencourt Pereira. -- 2026.
99 f.
Orientador: Moisés Rockembach.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Biblioteconomia e Comunicação, Programa de Pós-Graduação em Ciência da Informação, Porto Alegre, BR-RS, 2026.

1. Inteligência Artificial. 2. Patrimônio Documental. 3. Transkribus. 4. Reconhecimento de Escrita Manuscrita (HTR). 5. Gamificação. I. Rockembach, Moisés, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Leonardo Porto de Bittencourt Pereira

Inteligência Artificial e Patrimônio Documental: potencialidades do Transkribus no reconhecimento de escrita manuscrita.

Dissertação apresentada como requisito parcial à obtenção do título de Mestre em Ciência da Informação da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul.
Orientador: Prof. Dr. Moisés Rockembach

Aprovado em: Porto Alegre, 4 de maio de 2026

BANCA EXAMINADORA:

Prof. Dr. Moisés Rockembach
PPGCIN - UFRGS

Prof. Dr. Rafael Port da Rocha
PPGCIN - UFRGS

Profa. Dra. Luciana Monteiro-Krebs
PPGOCTS - UFRJ

Dr. Jonas Ferrigolo Melo
CITCEM - UP

Profa. Dra. Evelin Mintegui (suplente)
PPGCI- FURG

AGRADECIMENTOS

Começo dedicando este trabalho, assim como tudo em minha vida, a Deus, principalmente por ter colocado em meu caminho pessoas generosas que tornaram esta trajetória possível.

Agradeço à minha família, em especial à minha amada mãe, Neusa Teresinha, por todo o apoio incondicional e pelos valores transmitidos.

Ao meu orientador, Prof. Dr. Moisés Rockembach, agradeço não apenas pela orientação desta dissertação, mas por ter sido uma figura central em todo o meu percurso acadêmico. Desde a graduação, passando pela iniciação científica, pelo Trabalho de Conclusão de Curso e, agora, na dissertação, sua confiança e acompanhamento foram fundamentais para o meu desenvolvimento intelectual e profissional.

Ao Programa de Pós-Graduação em Ciência da Informação (PPGCIN) e à Universidade Federal do Rio Grande do Sul (UFRGS), registro meu agradecimento pelo ambiente acadêmico e pelas condições que possibilitaram a realização desta pesquisa.

Aos membros da banca, agradeço pelo tempo dedicado à leitura deste trabalho e pela generosidade em aceitarem contribuir com suas avaliações e considerações.

Aos amigos que atravessaram comigo diferentes momentos desta trajetória, deixo meu agradecimento pelo apoio.

À Lidiane, agradeço pela compreensão e por ter sido presença firme nos momentos em que o percurso exigiu mais do que apenas esforço acadêmico.

A Alexandre Veiga, registro um agradecimento especial pela contribuição inicial que deu origem à ideia que, posteriormente, se desenvolveu e resultou nesta dissertação.

Por fim, este trabalho marca um processo de transformação pessoal, no qual desafios e aprendizados redefiniram minha forma de pensar, pesquisar e me posicionar diante do conhecimento. Agradeço a todos que, de diferentes formas, contribuíram para sua realização.

PORTO DE BITTENCOURT PEREIRA, Leonardo. **Inteligência Artificial e Patrimônio Documental: potencialidades do Transkribus no reconhecimento de escrita manuscrita.** Dissertação em Ciência da Informação Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2026.

RESUMO

O trabalho tem como objetivo refletir criticamente sobre as potencialidades e desafios da Inteligência Artificial (IA), com o uso de Handwriting Text Recognition (HTR) – reconhecimento de texto manuscrito – tomando a ferramenta Transkribus como estudo de caso no contexto da Arquivologia, com o patrimônio documental, Ciência da Informação (CI) e Humanidades Digitais. Busca-se compreender de que modo a tecnologia de HTR pode contribuir para o tratamento, a preservação e o acesso a acervos documentais históricos, ao mesmo tempo em que suscita questões como a mediação da informação. Para tanto, realiza-se uma análise teórico-conceitual sobre o uso da IA em processos de transcrição de manuscritos, com ênfase nas funcionalidades, limitações e implicações do uso do Transkribus. A partir dessa análise, propõe-se o desenvolvimento de um modelo gamificado para o uso da ferramenta, voltado a contextos de ensino e capacitação profissional, que possa, futuramente, fomentar práticas colaborativas e de valorização do patrimônio documental. Portanto, o trabalho busca, assim, contribuir para o debate sobre o papel da IA na mediação entre tecnologia e o patrimônio documental, reforçando a importância da abordagem crítica e informacional frente às transformações digitais.

Palavras-chave: Inteligência Artificial; Patrimônio Documental; Transkribus; Reconhecimento de Escrita Manuscrita (HTR); Gamificação.

PORTO DE BITTENCOURT PEREIRA, Leonardo. **Artificial Intelligence and Documentary Heritage: the potential of Transkribus in handwritten text recognition.** Master's thesis em Ciência da Informação. Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2026.

ABSTRACT

This study aims to critically reflect on the potential and challenges of Artificial Intelligence (AI), through the use of Handwritten Text Recognition (HTR), taking the Transkribus platform as a case study within the context of Archival Science, documentary heritage, Information Science, and Digital Humanities. It seeks to understand how HTR technology can contribute to the processing, preservation, and access to historical documentary collections, while also raising issues related to information mediation. To this end, a theoretical-conceptual analysis is conducted on the use of AI in manuscript transcription processes, with emphasis on the functionalities, limitations, and implications of using Transkribus. Based on this analysis, the study proposes the development of a gamified model for the use of the tool, aimed at educational and professional training contexts, which may, in the future, foster collaborative practices and the appreciation of documentary heritage. Therefore, this work seeks to contribute to the debate on the role of AI in mediating the relationship between technology and documentary heritage, reinforcing the importance of a critical and informational approach in the face of digital transformations.

Keywords: Artificial Intelligence; Documentary Heritage; Transkribus; Handwritten Text Recognition (HTR); Gamification.

LISTA DE FIGURAS

Figura 1 – Exemplo de transcrição no Transkribus 35
Figura 2 – Modelos de transcrição no Transkribus..... 36
Figura 3 – Exemplo de site dentro do Transkribus 39
Figura 4 – Gráfico de referências da 1ª busca 56
Figura 5 – Gráfico de referências da 2ª busca 58
Figura 6 – Avaliação da facilidade de uso do Transkribus 61

LISTA DE QUADROS

Quadro 1 – Procedimentos metodológicos adotados na revisão sistemática 49
Quadro 2 – Levantamento de referências da 1ª busca 54
Quadro 3 – Levantamento de referências da 2ª busca 57
Quadro 4 – Síntese analítica das dificuldades de uso do Transkribus 62
Quadro 5 – Estrutura do modelo gamificado 64
Quadro 6 – Camadas De Mediação No Modelo Gamificado 67
Quadro 7 – Sistema de pontuação e recompensas 69
Quadro 8 – Plano de análise para futura aplicação no instrumento. 72
Quadro 9 – Desempenho de modelos de IA em HTR segundo diferentes estudos 79

LISTA DE ABREVIATURAS E SIGLAS

ADHO	Alliance of Digital Humanities Organizations
BNF	Biblioteca Nacional da França
CAS	Computational Archival Science
CER	Character Error Rate
CI	Ciência da Informação
ICA	Conselho Internacional de Arquivos
CHAT-GPT	Chat Generative Pre-Trained Transformer
CULTAI	Report of the Independent Expert Group on Artificial Intelligence and Culture
FCT	Fundação para a Ciência e a Tecnologia
FBN	Fundação Biblioteca Nacional
HTR	Handwriting Text Recognition
HITL	Human-in-the-Loop
IA	Inteligência Artificial
InterPARES	International Research on Permanent Authentic Records in Electronic Systems
LGPD	Lei Geral de Proteção de Dados
LLM	Large Language Model
MCTI	Ministério da Ciência, Tecnologia e Inovações
OCR	Optical Character Recognition
ONU	Organização das Nações Unidas
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
RGPD	Regulamento Geral de Proteção de Dados
TSP	Transkribus Scholarship Programme
UFRGS	Universidade Federal do Rio Grande do Sul
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura
WER	Word Error Rate

SUMÁRIO

1 INTRODUÇÃO 11
1.1 PROBLEMA DE PESQUISA 13
1.2 OBJETIVOS 13
1.2.1 Objetivo geral 13
1.2.2 Objetivos específicos 14
2 REFERENCIAL TEÓRICO 15
2.1 MEDIAÇÃO DA INFORMAÇÃO E A ATUAÇÃO ALGORÍTMCA NOS PROCESSOS INFORMACIONAIS 20
2.2 PALEOGRAFIA E PATRIMÔNIO DOCUMENTAL 24
2.3 HANDWRITING TEXT RECOGNITION 28
2.4 TRANSKRIBUS COMO AMBIENTE SOCIOTÉCNICO 34
2.5 GAMIFICAÇÃO E FORMAÇÃO EM AMBIENTES DIGITAIS 42
3 PROCEDIMENTOS METODOLÓGICOS 49
3.1 FASE DIAGNÓSTICA E EXPLORATÓRIA 49
3.2 FASE PROPOSITIVA 52
4 ANÁLISE E DISCUSSÃO DOS RESULTADOS 54
5 CONCLUSÃO 81
REFERÊNCIAS 85
APÊNDICE A – MANUAL TRANSKRIBUS 95
ANEXO A – DOCUMENTO 1 96
ANEXO B – DOCUMENTO 2 97
ANEXO C – DOCUMENTO 3 98
ANEXO D – DOCUMENTO 4 99

1 INTRODUÇÃO

Nos últimos anos o avanço da Inteligência Artificial (IA) produziu transformações significativas em diferentes esferas da sociedade. As suas aplicações vão além das fronteiras anteriormente imaginadas, tendo um impacto nos setores econômicos, científicos e culturais, e modificando práticas do cotidiano através da automação de tarefas e da popularização de dispositivos “inteligentes”.

A aceleração e aprimoramento dos usos de IA tornou-se ainda mais evidente no contexto pós-pandemia de COVID-19. Novas práticas, como o trabalho remoto, intensificaram o uso de tecnologias digitais e ampliaram o acesso a ferramentas baseadas em IA. A popularização de ferramentas como o Chat-GPT (Generative Pre-trained Transformer) representa um marco nesse cenário, ao tornar visível e acessível uma tecnologia que, até então, parecia restrita a ambientes altamente especializados. Ademais, nota-se que há uma disputa entre diferentes Big Techs para a criação de sua própria IA, para venda de novos serviços e produtos, ou seja, a IA vem sendo comercializada como um produto e grande diferencial ao usuário consumidor.

Simultaneamente, emergem debates éticos, sociais e epistemológicos, visto que o uso crescente da IA em atividades sensíveis, como o suporte psicológico, a produção de conteúdo com a geração de imagens e vídeos, etc, levantam questões sobre autenticidade, privacidade e segurança da informação. A dificuldade crescente na distinção entre o real e o sintético potencializa os riscos associados à desinformação e à manipulação. Sendo assim, novas narrativas podem ser criadas a partir de manipulação e notícias falsas, com o potencial para interferir e influenciar questões de Estado e seus cidadãos.

Além disso, os estudos sobre IA e seus impactos no contexto social atravessam diversas áreas do conhecimento, tendo adquirido significativo destaque nos últimos anos. Nesse cenário, o presente trabalho propõe refletir e fomentar o debate acerca das potencialidades de uso da IA no âmbito da Arquivologia e das Humanidades Digitais, com ênfase na CI, buscando contribuir para a ampliação e o aprimoramento de processos e técnicas nessas áreas.

Nesta pesquisa, adota-se o uso do termo Handwritten Text Recognition (HTR) para designar os processos de reconhecimento automático de escrita

manuscrita por meio de técnicas de IA. Embora a literatura apresente variações terminológicas, como “reconhecimento de escrita manuscrita” ou “reconhecimento automático de manuscritos”, opta-se, neste trabalho, pela utilização da sigla HTR por sua ampla consolidação no campo internacional e por sua recorrência em plataformas e ferramentas especializadas, como o Transkribus. Essa escolha busca garantir maior consistência conceitual e alinhamento com a produção científica contemporânea, sendo o termo adotado como padrão ao longo da dissertação.

Nesse sentido, ao observar as pesquisas acerca do uso de HTR, notou-se, a partir da análise de literatura, que elas se destacam em volume no contexto europeu, onde há maior consolidação de grupos de pesquisa, financiamento e implementação de tecnologias voltadas ao uso da IA em acervos históricos e patrimônio documental. Contudo, ainda se identifica uma lacuna de estudos que discutam as especificidades de uso, limitações e problemáticas do HTR no contexto latino-americano, especialmente na América do Sul, região caracterizada por ampla diversidade linguística, documental e cultural.

Diante desse cenário, entende-se que este trabalho pode contribuir para ampliar o alcance das discussões sobre HTR, fomentando reflexões e possíveis inovações adequadas ao contexto regional, além de colaborar para a redução de desigualdades no acesso e uso dessas tecnologias.

Portanto, foi construído um referencial teórico organizado em subcapítulos, contemplando discussões sobre mediação da informação, Paleografia, a tecnologia de HTR e a ferramenta Transkribus, além de abordagens sobre gamificação e metodologias colaborativas. Sendo assim, são apresentados os procedimentos metodológicos, que incluem o levantamento bibliográfico e a análise das produções relacionadas ao tema. Com base nessas etapas, propõe-se o desenvolvimento de um modelo gamificado voltado ao uso do Transkribus em contextos de ensino e capacitação profissional, com potencial para estimular práticas colaborativas e valorizar o patrimônio documental.

A proposta de articular as áreas mencionadas fundamenta-se em seu caráter interdisciplinar, especialmente no âmbito da CI, da Arquivologia e das Humanidades Digitais. No campo arquivístico, particularmente nos estudos de Paleografia, observa-se o avanço de discussões sobre o uso de IA por meio de ferramentas

como o Transkribus, que utiliza HTR para a transcrição de manuscritos, com crescente aplicação em acervos de valor histórico.

Por fim, do ponto de vista prático, esta pesquisa também se justifica pela proposição de um modelo de uso gamificado do Transkribus, com potencial aplicação em estudos futuros, visando à redução do tempo de transcrição, ao aprimoramento do letramento de estudantes e profissionais e, consequentemente, à ampliação do acesso da sociedade a fontes primárias de pesquisa.

1.1 PROBLEMA DE PESQUISA

Tendo em vista os argumentos relatados, o problema de pesquisa se dá no questionamento de como a utilização de tecnologias de HTR, no contexto do patrimônio documental, impacta os processos de transcrição e acesso à informação, e de que forma estratégias de gamificação podem contribuir para a capacitação de usuários no uso da ferramenta Transkribus, considerando as exigências de conhecimento paleográfico envolvidas. Desta forma, questionamos:

De que forma o HTR reconfigura o tratamento e o acesso ao patrimônio documental em contextos de paleografia digital e dados em larga escala, no âmbito de estratégias de gamificação voltadas à mitigação da fadiga cognitiva e à capacitação de usuários na validação de transcrições no Transkribus?

1.2 OBJETIVOS

Portanto, a partir dos argumentos expostos, se construiu os objetivos de pesquisa listados a seguir.

1.2.1 Objetivo geral

Analisar as potencialidades, limitações e implicações da tecnologia de HTR no acesso e tratamento do patrimônio documental histórico e, a partir do diagnóstico das demandas de mediação informacional, desenvolver e propor um modelo

gamificado de capacitação focado na otimização do uso da ferramenta Transkribus por usuários não especialistas.

1.2.2 Objetivos específicos

Os objetivos específicos estão organizados de forma a contemplar, de maneira articulada, os eixos teórico, informacional, sociotécnico e formativo que estruturam a análise proposta nesta pesquisa.

a. Mapear e sistematizar a literatura científica nacional e internacional sobre as intersecções entre HTR, Paleografia Digital e patrimônio documental, por meio de revisão sistemática em bases de dados.

b. Avaliar a usabilidade, a curva de aprendizado e as percepções operacionais de usuários iniciantes durante a interação com o Transkribus, a partir de dados primários coletados em um teste piloto exploratório.

c. Estruturar e delinear conceitualmente um modelo gamificado de capacitação, fundamentado nas lacunas operacionais identificadas empiricamente, visando fomentar o engajamento e a literacia informacional no uso do Transkribus.

2 REFERENCIAL TEÓRICO

O presente capítulo tem como objetivo estabelecer os fundamentos teóricos que sustentam a análise proposta nesta pesquisa. Considerando a problemática relacionada à mediação da informação em contextos de uso de tecnologias de HTR aplicadas ao patrimônio documental, busca-se articular diferentes campos do conhecimento, com ênfase na CI, adotada como eixo interpretativo central.

A discussão desenvolvida parte da compreensão de que a incorporação de tecnologias baseadas em IA não representa apenas um avanço técnico, mas implica transformações nos processos de produção, organização e acesso à informação. Nesse sentido, o capítulo organiza-se a partir da articulação entre a CI, a Arquivologia, a Paleografia e as Humanidades Digitais, considerando suas contribuições para a compreensão das dinâmicas contemporâneas de mediação.

Inicialmente, discute-se o papel da CI como campo responsável pela análise dos processos informacionais, com destaque para o conceito de mediação. Em seguida, aborda-se a Arquivologia como eixo estruturante do patrimônio documental, bem como a Paleografia sendo fonte de conhecimento para o treinamento e validação de modelos de HTR. Por fim, situam-se as Humanidades Digitais como campo de articulação no qual se inserem as práticas analisadas.

Dessa forma, o presente capítulo busca oferecer suporte teórico para a compreensão das transformações introduzidas pelo uso de HTR, orientando a análise desenvolvida nos capítulos subsequentes.

Nesse sentido, a CI e a Arquivologia se fortalecem na interdisciplinaridade (Rodrigues; Gomes, 2021) e, por consequência, se aproximam das Humanidades Digitais, com temas de pesquisa sendo compartilhados, e que, segundo Pimenta (2020, p. 16-17) não deveriam ser pensados separadamente. Ademais, Daniel Alves comenta sobre um desafio que lhe ocorre diante os estudos das Humanidades Digitais em diversas iniciativas dentro de departamentos e laboratórios acadêmicos: a limitação da interdisciplinaridade, que, segundo ele, precisa existir nas Humanidades Digitais (Aquino, 2020, p. 7).

Como diversas áreas do conhecimento, a CI e a Arquivologia vêm sendo campos de discussões acerca dos benefícios e consequências da implicação do

uso de novas tecnologias. Questões como os desafios crescentes diante do volume, da complexidade e da diversidade dos dados digitais, possibilitando com que, nesse contexto, a IA surja não apenas como ferramenta técnica, mas como um agente de mudança estrutural, capaz de modificar a forma como a informação é organizada, recuperada e interpretada. Sendo assim, demandando novas competências profissionais e infraestruturas tecnológicas adequadas (Rockembach, 2021).

Marciano (2022) identifica uma nova perspectiva para a área por meio da Arquivologia Computacional (Computational Archival Science – CAS), compreendida como a aplicação de tecnologias computacionais e o desenvolvimento de estudos da Computação no contexto das instituições arquivísticas e de seus acervos. A partir dessa abordagem, questões como a extração e a análise de grandes volumes de dados passam a ganhar destaque, assim como suas implicações para a gestão documental e a preservação dos acervos (Pajares; Tornero; Martin, 2023, p. 1).

Pajares, Tornero e Martin (2023), a partir de um estudo de caso na Universidade de Lérida, na Espanha, afirmam que a junção entre IA e CAS trás novas potencialidades aos arquivistas e profissionais de arquivos ao passo que alia a gestão de documentos com o armazenamento de dados e uso de novas tecnologias como a IA (Pajares; Tornero; Martin, 2023, p. 5). Sendo assim, se faz fundamental que a Arquivologia e as Humanidades Digitais, se relacionem em projetos em prol dos seus desenvolvimentos, e do constante aprimoramento e formação profissional de seus pesquisadores no atual contexto social.

Se faz relevante destacar que, embora o trabalho se proponha a relacionar as três áreas descritas com a IA, o foco da pesquisa é a área da CI. Nesse sentido, a CI é a disciplina que investiga as propriedades e o comportamento informacional, herdando elementos de áreas como a Documentação e a Biblioteconomia. A CI está relacionada com o corpo de conhecimentos vinculados à origem, coleção, organização, armazenamento, recuperação, interpretação, transmissão, transformação e utilização da informação (Borko, 1968).

Já as Humanidades Digitais constituem um campo interdisciplinar que ultrapassa a mera aplicação de tecnologias às áreas tradicionais das Humanidades.

Mais do que incorporar ferramentas digitais, as Humanidades Digitais propõem novas formas de produção e representação do conhecimento, questionando os próprios fundamentos epistemológicos da pesquisa em humanidades. Conforme argumenta Burdick et al. (2012), o digital não é apenas um meio, mas um ambiente que transforma as práticas de leitura, escrita, autoria e colaboração, exigindo novas metodologias e linguagens críticas.

Nesse sentido, as Humanidades Digitais não se restringem a uma dimensão instrumental, mas configuram um espaço de experimentação teórico-metodológica. Autores como Berry (2012) e Drucker (2011) ressaltam o papel das Humanidades Digitais na problematização das próprias tecnologias que empregam. Ao invés de aceitá-las como instrumentos transparentes, as Humanidades Digitais interrogam seus algoritmos, modelos de dados e processos de mediação, evidenciando os vieses e as relações de poder embutidos nas práticas computacionais. Essa abordagem crítica permite que o campo dialogue com questões éticas, políticas e epistemológicas, ampliando o escopo das humanidades na atualidade.

As Humanidades Digitais constituem um campo interdisciplinar que emerge da convergência entre as ciências humanas e as tecnologias digitais, promovendo novas formas de análise, interpretação e acesso a objetos culturais e documentais (Burdick et al., 2012; Berry, 2012). Nesse contexto, o uso de ferramentas computacionais, incluindo técnicas de IA, tem ampliado as possibilidades de tratamento e exploração do patrimônio documental.

No âmbito desta pesquisa, as Humanidades Digitais não são compreendidas como objeto central de investigação, mas como campo de articulação no qual se inserem as práticas analisadas. É nesse ambiente que se desenvolvem iniciativas voltadas à digitalização, transcrição e disponibilização de documentos manuscritos, frequentemente mediadas por tecnologias como o HTR.

A incorporação dessas tecnologias nos projetos de Humanidades Digitais tem permitido ampliar o acesso a grandes volumes de documentos, anteriormente restritos a especialistas. No entanto, essa ampliação do acesso não ocorre de forma neutra, uma vez que os processos de digitalização, modelagem e transcrição envolvem escolhas técnicas e interpretativas que influenciam a forma como os

documentos são representados e compreendidos (Schreibman; Siemens; Unsworth, 2016).

Sendo assim, as Humanidades Digitais configuram-se como espaço privilegiado para observar as transformações na mediação da informação, uma vez que articulam práticas tradicionais de pesquisa com ferramentas digitais avançadas. A transcrição automatizada de manuscritos, por meio do HTR, exemplifica essa articulação, ao integrar conhecimentos paleográficos, práticas arquivísticas e tecnologias computacionais em um mesmo processo.

Além disso, projetos desenvolvidos no âmbito das Humanidades Digitais frequentemente envolvem colaboração entre diferentes perfis profissionais, incluindo pesquisadores, arquivistas, desenvolvedores e usuários, o que evidencia a natureza sociotécnica desses ambientes. Essa característica reforça a compreensão de que a produção do conhecimento nesses contextos resulta da interação entre sujeitos e tecnologias, e não de processos isolados.

A partir dessa perspectiva, as Humanidades Digitais permitem situar o uso do HTR em um contexto mais amplo de transformação das práticas de pesquisa e de acesso à informação. Ao mesmo tempo, evidenciam que a incorporação de tecnologias digitais não elimina a necessidade de mediação humana, mas reconfigura suas formas de atuação. Assim, ao articular Arquivologia, CI e tecnologias de HTR, as Humanidades Digitais configuram-se como campo no qual se evidenciam as dinâmicas contemporâneas de mediação da informação, permitindo compreender como o uso de HTR se insere em processos mais amplos de produção e circulação do conhecimento.

Portanto, as Humanidades Digitais devem ser compreendidas não apenas como um conjunto de técnicas aplicadas à cultura, mas como uma área de reflexão sobre como o digital transforma o modo de pensar, pesquisar e narrar o humano. Como pode ser observado por suas características, ambas as áreas podem trazer elementos da disciplina arquivística.

Nesse sentido, a leitura e a compreensão de um manuscrito, sob o ponto de vista arquivístico e informacional, não se esgotam no reconhecimento dos elementos gráficos que o documento carrega. Elas envolvem processos de mediação cognitiva e contextual, nos quais o pesquisador reconstrói significados a

partir da historicidade do suporte, da intencionalidade do autor e das condições de produção e transmissão do documento (Cook, 1997).

Portanto, a CI possibilita que as áreas da Arquivologia e Humanidades Digitais possam convergir, de forma coerente, para auxílio dos profissionais que lidam com patrimônio documental. A introdução de ferramentas de IA nestes campos não representa uma ruptura com a tradição estabelecida, mas, sim, uma mudança de paradigma, o que lhe atribui um caráter tecnológico mais acentuado.

Nesse contexto, faz-se relevante abordar o projeto International Research on Permanent Authentic Records in Electronic Systems (InterPARES), que se constitui de uma iniciativa internacional voltada à pesquisa e desenvolvimento de metodologias que garantam a preservação e autenticidade de documentos arquivísticos digitais. O projeto, que teve seu início em 1999, busca assegurar a integridade de registros eletrônicos e promover a gestão adequada de documentos digitais produzidos em diferentes contextos institucionais e sociais (Interpares Trust AI, 2023).

Reunindo pesquisadores e alunos, o projeto, elaborado na Universidade de British Columbia, no Canadá, desenvolveu-se em distintas fases, cada uma com um foco específico. Na primeira fase, denominada de InterPARES 1 (1999–2001), os pesquisadores concentraram-se na identificação de requisitos conceituais para avaliação e manutenção da autenticidade de documentos digitais produzidos em atividades administrativas e legais (Duranti, 2007).

A segunda fase, InterPARES 2 (2002–2006), abordou documentos digitais gerados em contextos artísticos, científicos e governamentais, incluindo sistemas experimentais e interativos. Já na terceira fase, InterPARES 3 (2007–2012), buscou-se capacitar as organizações responsáveis pela produção e manutenção de documentos digitais, desenvolvendo estratégias para preservação e acesso de longo prazo (Duranti, 2009). Após isso, a fase InterPARES Trust (2013–2017) investigou documentos produzidos e mantidos em ambientes da internet, com ênfase em nuvem e mídias sociais e, atualmente, a fase InterPARES Trust AI (2021–2026) concentra-se na aplicação da IA na avaliação de documentos digitais, visando desenvolver modelos capazes de mensurar a adequação da IA para essa

finalidade (Interpares Trust AI, 2023).

Portanto, diante do exposto, torna-se necessário aprofundar a discussão sobre a mediação da informação no contexto da CI e IA, tomando-a como eixo central para a compreensão dos processos informacionais contemporâneos.

2.1 MEDIAÇÃO DA INFORMAÇÃO E A ATUAÇÃO ALGORÍTMCA NOS PROCESSOS INFORMACIONAIS

A presente pesquisa insere-se no campo da CI, considerando as transformações contemporâneas nos processos de produção, organização e acesso à informação, especialmente em contextos que envolvem o uso de tecnologias digitais. Nesse sentido, o avanço da IA tem impactado significativamente as formas de tratamento e disponibilização da informação, exigindo uma análise que vá além de seus aspectos técnicos (Capurro, 2003; Hjørland, 2002).

No âmbito do patrimônio documental, essas transformações tornam-se particularmente evidentes, uma vez que a incorporação de tecnologias como o HTR modifica práticas tradicionalmente associadas à leitura e à transcrição de documentos manuscritos. Tais práticas, historicamente mediadas por especialistas, passam a incorporar sistemas automatizados que influenciam a forma como a informação é processada e acessada.

A partir da perspectiva da CI, essas mudanças podem ser compreendidas por meio da mediação da informação, entendida como um processo que articula sujeitos, documentos e tecnologias na construção do conhecimento (Gomes, 2014; Araújo, 2018). A introdução de sistemas baseados em IA, contudo, não se limita à automatização de tarefas, mas implica a reconfiguração dessas dinâmicas ao inserir novos agentes no processo informacional.

Dessa forma, o acesso à informação deixa de depender exclusivamente da atuação humana, passando a envolver processos algorítmicos que operam sobre os documentos, influenciando sua leitura, interpretação e representação. Tais sistemas são treinados a partir de dados previamente estruturados, o que evidencia

que sua atuação está condicionada a escolhas técnicas que impactam o resultado informacional (Kitchin, 2017).

Nesse contexto, torna-se necessário compreender que a tecnologia não atua de forma neutra, mas participa da construção da informação, interferindo nos modos de organização e circulação do conhecimento. Assim, esta pesquisa adota a mediação da informação como eixo central de análise, buscando compreender como o uso de tecnologias de HTR no patrimônio documental reconfigura práticas informacionais, a partir de uma perspectiva informacional situada na CI.

A introdução de agentes técnicos amplia o campo da mediação informacional, incorporando mediadores não humanos que operam segundo lógicas de cálculo e previsão estatística. Nesse cenário, a suposta neutralidade dos sistemas de IA pode mascarar valores, escolhas e exclusões codificadas nos algoritmos, exigindo uma abordagem crítica acerca de seus efeitos nos processos informacionais.

Assim, a ação dos algoritmos ultrapassa o campo da automação e da eficiência computacional, passando a influenciar a forma como o conhecimento é produzido, organizado e legitimado. A mediação algorítmica, nesse sentido, emerge como uma lente analítica para compreender as interações entre humanos, dados e sistemas automatizados (Rangel, 2022, 2024), sendo caracterizada pela capacidade dos algoritmos de estruturar relevâncias, hierarquias e significados.

No âmbito da CI, a mediação algorítmica tem sido compreendida como uma reconfiguração das formas tradicionais de mediação, e não sua substituição, uma vez que os processos informacionais passam a ser atravessados por lógicas computacionais como ranqueamento, otimização e predição (Sousa Júnior, 2024). Tais dinâmicas afetam não apenas a recuperação da informação, mas também sua circulação e consumo, introduzindo fenômenos como personalização, filtragem e opacidade (Rangel, 2022, 2024).

Entretanto, grande parte dessas análises permanece centrada nos processos de organização e acesso, dedicando menor atenção às dimensões interpretativas envolvidas na mediação. Essa lacuna torna-se especialmente relevante no campo do patrimônio documental, no qual a mediação não se restringe à recuperação de dados, mas envolve a reconstrução de sentidos a partir de documentos historicamente contextualizados.

Nesse contexto, o uso de ferramentas de HTR, como o Transkribus, evidencia uma forma específica e mais complexa de mediação algorítmica. Diferentemente de sistemas voltados à organização e recuperação, o HTR interfere diretamente na constituição do objeto informacional ao transformar imagens manuscritas em texto legível por máquina.

Esse processo implica que o algoritmo não apenas medeia o acesso, mas participa da reconfiguração da informação, uma vez que erros de reconhecimento, frequentemente mensurados por métricas como o Character Error Rate (CER), podem alterar significativamente o conteúdo disponibilizado. Nessa perspectiva, o algoritmo deixa de ser apenas um mediador e passa a atuar como um agente que intervém na própria materialidade informacional do documento.

Diante disso, abordagens como o Human-in-the-Loop (HITL) tornam-se fundamentais, ao integrar sujeitos humanos nos processos de validação e correção dos sistemas de IA, contribuindo para a qualidade e confiabilidade dos resultados (Amershi et al., 2014).

Essa perspectiva híbrida é corroborada pela literatura recente, que alerta que os sistemas de inteligência artificial aplicados a arquivos não devem ser vistos como substitutos dos profissionais da informação, mas sim como assistentes capazes de otimizar tarefas repetitivas, garantindo que o processo seja sempre validado e supervisionado pela inteligência humana (Sousa; Rockembach, 2025)

No contexto do HTR aplicado ao patrimônio documental, essa abordagem evidencia que a mediação assume um caráter híbrido, no qual algoritmos e profissionais da informação atuam de forma articulada.

Autores como Capurro (2003) e Floridi (2019) reforçam que a informação e suas mediações estão inseridas em contextos éticos e culturais, sendo influenciadas por valores e decisões incorporados às tecnologias. No caso do HTR, essa condição é particularmente evidente, uma vez que os sistemas, ao converterem manuscritos em dados, não apenas ampliam o acesso, mas também introduzem novas formas de opacidade e reconfiguram as condições de leitura e interpretação dos documentos históricos (Nockels; Gooding; Terras, 2024).

Nesse sentido, a literacia crítica em IA, conforme proposta por Arias Hernández e Rockembach (2025), torna-se essencial para a atuação dos profissionais da informação, permitindo compreender as limitações, vieses e

implicações desses sistemas. O arquivista, nesse contexto, passa a desempenhar um papel ampliado, não apenas como mediador entre documento e usuário, mas como intérprete das mediações técnicas que condicionam a produção e circulação da informação.

Diante da complexidade dessas tecnologias, a literatura evidencia a necessidade de que os profissionais da informação agreguem funções inerentes a um eticista digital, responsabilizando-se por orientar a aplicação ética da inteligência artificial, mitigar vieses algorítmicos e assegurar a transparência organizacional nesses novos processos de mediação (Rockembach; Geerts, 2024).

Assim, o conhecimento informacional passa a ser construído de forma híbrida, envolvendo humanos e algoritmos, e o documento assume uma dupla dimensão: material e digital. No campo do patrimônio documental, isso implica reconhecer que a IA não apenas amplia o acesso, mas transforma as formas de leitura e rememoração do passado.

A partir dessa perspectiva, os sistemas de IA exigem o reposicionamento das práticas arquivísticas à luz de princípios como transparência, auditabilidade e responsabilidade (Rockembach, 2020). Enfrentar esses desafios demanda uma abordagem interdisciplinar que articule a CI, os Estudos Críticos de Algoritmos e a Diplomática contemporânea, possibilitando uma análise ética e contextualizada das mediações digitais no patrimônio documental.

As reconfigurações da mediação da informação no contexto da IA tornam-se especialmente relevantes quando situadas no campo do patrimônio documental, onde os processos informacionais não se limitam ao acesso, mas envolvem a interpretação de documentos historicamente complexos. Nesse cenário, a Paleografia emerge como prática fundamental de mediação, evidenciando que a leitura de manuscritos não é um processo direto, mas uma construção interpretativa que se torna ainda mais tensionada com a incorporação de tecnologias como o HTR. Diante disso, no próximo subcapítulo serão abordados os fundamentos da Paleografia e suas relações com o patrimônio documental.

2.2 PALEOGRAFIA E PATRIMÔNIO DOCUMENTAL

O patrimônio documental é compreendido como um conjunto de documentos que possuem valor histórico, cultural e informacional, sendo objeto de preservação, organização e acesso no âmbito das instituições arquivísticas (Bellotto, 2002; Rondinelli, 2005; Jardim, 1995). Nesse sentido, a Arquivologia desempenha papel fundamental, ao fornecer os princípios e as práticas que orientam o tratamento desses documentos ao longo do tempo.

Sendo assim, os documentos manuscritos, em especial, apresentam desafios específicos relacionados à sua leitura, interpretação e transcrição, demandando conhecimentos especializados para sua adequada compreensão. Tradicionalmente, essas atividades são realizadas por meio da Paleografia, área na qual o paleografo se dedica ao estudo das formas de escrita e à interpretação de documentos produzidos em diferentes contextos históricos.

Nesse cenário, a transcrição de documentos manuscritos constitui uma prática central para o acesso à informação, uma vez que possibilita a conversão de conteúdos de difícil leitura em formatos mais acessíveis. Trata-se, portanto, de uma atividade que se insere diretamente nos processos de mediação da informação, ao viabilizar a circulação e o uso do conhecimento contido nos documentos.

Com o avanço das tecnologias digitais, especialmente da IA, essas práticas passam a incorporar novos recursos, como o HTR, que automatiza parte do processo de transcrição. No entanto, a introdução de novas tecnologias não elimina a necessidade dos fundamentos arquivísticos, como a garantia da autenticidade, da integridade e da contextualização dos documentos (Duranti, 1998).

Dessa forma, a utilização de HTR em contextos arquivísticos deve ser compreendida como elemento que se integra às práticas existentes, e não como substituição da atuação profissional. A Arquivologia permanece como eixo estruturante do tratamento do patrimônio documental, enquanto a tecnologia atua como suporte e mediação nos processos de acesso à informação (Bellotto, 2002; Duranti, 1998).

Além disso, a incorporação de sistemas automatizados nos fluxos arquivísticos introduz novas questões relacionadas à interpretação e à

representação dos documentos, uma vez que os resultados gerados pelos modelos algorítmicos dependem de dados de treinamento, parâmetros técnicos e escolhas realizadas durante o processo de uso da ferramenta.

Nesse sentido, a análise do uso de HTR no patrimônio documental demanda uma abordagem que considere não apenas os aspectos técnicos da tecnologia, mas também sua inserção nas práticas arquivísticas e suas implicações para a mediação da informação, com o paleografo.

Considerando que a transcrição de documentos manuscritos constitui prática central no acesso ao patrimônio documental, torna-se necessário aprofundar o papel do conhecimento especializado que historicamente sustenta essa atividade. Nesse contexto, a Paleografia emerge como elemento fundamental, especialmente quando se considera sua relação com o desenvolvimento e a aplicação de tecnologias de HTR.

Retomando às questões sobre patrimônio documental e a tecnologia HTR na aplicação da Paleografia Digital, faz-se necessário rememorar fundamentos básicos e um breve contexto histórico acerca da área e da técnica paleográfica. O estudo da Paleografia, enquanto ciência e técnica voltada à leitura, transcrição e análise de textos manuscritos, requer a compreensão de seus fundamentos históricos e conceituais para se entender plenamente sua adaptação ao contexto digital contemporâneo. Seguindo a definição de Berwanger e Leal (2008), a Paleografia é:

> “[...] o estudo técnico de textos antigos, na sua forma exterior, que compreende o conhecimento dos materiais e instrumentos para escrever, a história da escrita e a evolução das letras, objetivando sua leitura e transcrição.” (Berwanger; Leal, 2008. p. 16)

Nesse sentido, historicamente a Paleografia é dividida em dois grandes períodos: o de formação, entendido como do fim do século XVII à primeira parte do século XVIII, e o período de afirmação, do século XVIII até a atualidade. No século XX, em especial após sua primeira metade, a Paleografia consolidou-se como técnica de aplicação pragmática, incorporada aos currículos universitários e deixando de ser domínio exclusivo de poucos especialistas.

Se faz relevante destacar que, embora suas técnicas sejam usadas em grande parte em documentos datados de períodos antigos, a aplicação de técnicas

paleográficas não se restringe a esses manuscritos, podendo seu uso ser estendido a documentos contemporâneos, desde que estes apresentem elementos de dificuldades de leitura. Além disso, a Paleografia atua como ciência auxiliar da História, do Direito e da Arquivologia.

Além disso, fundamental à Paleografia, a Diplomática é uma disciplina que tem por objetivo o estudo crítico dos documentos, analisando sua forma, estrutura e autenticidade. Surgida no século XVII, com Jean Mabillon, essa área consolidou-se como fundamental para a verificação da veracidade de documentos antigos e para a compreensão de seu contexto de produção. Como afirma Bellotto (2002), a Diplomática "busca compreender o documento em sua totalidade, examinando seus elementos internos e externos para determinar sua gênese, função e valor". Assim, ela ultrapassa a simples análise material e assume um papel interpretativo na investigação das relações entre o conteúdo documental e o ambiente institucional que o originou.

No caso dos documentos manuscritos, a Diplomática é essencial para o reconhecimento das particularidades gráficas, formais e funcionais desses registros. Muitos manuscritos antigos apresentam caligrafias, fórmulas e expressões próprias de um período ou administração específica, o que torna indispensável uma análise contextualizada. Segundo Duranti (1998, p. 17), a Diplomática dedica-se ao estudo da gênese, da forma e da transmissão dos documentos arquivísticos, analisando seus elementos formais e sua relação com os fatos que registram e com seus produtores. Desse modo, a leitura diplomática de um manuscrito não se restringe à paleográfica, mas envolve a reconstrução do contexto de criação, circulação e uso do documento.

O conhecimento sobre a estrutura e a função dos documentos é fundamental para a validação e interpretação por sistemas de que usam HTR. Assim, ao dialogar com a tecnologia, a Diplomática garante que o sentido histórico e a autenticidade dos documentos não se percam no processo de transcrição e análise automatizada. Na atualidade, o avanço das tecnologias e a consolidação da Paleografia Digital ilustram os "novos usos de velhas ciências" mencionados por Duranti (1998, p. 2–5).

O entendimento dessa transição entre tradição e inovação, mencionado por Duranti, pode ser enriquecido pela analogia entre as fases da web e os estágios de

aplicação tecnológica na Paleografia, propostas por Magalhães (2021). Na Paleografia 1.0, o trabalho era predominantemente manual, com apoio limitado da tecnologia, restrito à produção de fac-símiles digitalizados para preservação e acesso remoto. Já a Paleografia 2.0 marca o início da interação mais intensa entre pesquisador e documento, com recursos como bases de dados, hiperlinks e edições digitais que integram o texto e paratexto, como os sinais especiais, os selos e as anotações aproximando-se dos métodos da Crítica Textual e da Diplomática.

Por fim, a Paleografia 3.0 representa a integração plena da tecnologia, podendo destacar-se o uso de ferramentas como o Transkribus, com o uso de HTR (Magalhães, 2021). Essa evolução demonstra que, assim como a web avançou de um ambiente estático, para espaços colaborativos e dinâmicos, a Paleografia também incorporou inovações tecnológicas sem abandonar seus fundamentos tradicionais.

Sendo assim, cabe ao paleógrafo apropriar-se criticamente dessas ferramentas, garantindo a autenticidade, a qualidade interpretativa e a relevância científica de suas análises, preservando o equilíbrio entre tradição e inovação no tratamento do patrimônio documental no contexto tecnológico. Contudo, se faz relevante citar que com a popularização de ferramentas digitais, novos usuários, sem formação técnica, passam a integrar ambientes sociotécnicos como a ferramenta Transkribus. Portanto, é fundamental manter as seguintes recomendações no que tange a transcrição documental:

> -Realizar a transcrição com o maior cuidado, e sem pressa;
> -Reproduzir o documento tal qual feito pelo autor;
> -Respeitar a ortografia da época do documento;
> -Não passar para a forma gráfica atual;
> -Respeitar as peculiaridades do manuscrito;
> -Respeitar o texto: o que está escrito, pois: “Nunca podemos olhar um documento do passado com os olhos do presente" (João E. Franklin Leal);
> -Conhecer a língua em que o documento foi escrito (época);
> -Utilizar o método comparativo;
>
> (Berwanger; Leal, 2008)

Sendo assim, podemos observar elementos de interdisciplinaridade na Paleografia, abrangendo conteúdos e técnicas de áreas como Visão Computacional, com o uso de ferramentas e tecnologias como Aprendizado de

Máquina e uso de Large Language Model (LLM). Essa confluência revela a importância do HTR não apenas como ferramenta técnica, mas como fenômeno sociotecnológico capaz de ampliar o acesso à memória documental e fortalecer os vínculos entre o passado e o presente (Silva et al., 2023).

Para Silva (2025), esse campo fértil amplia as práticas arquivísticas ao permitir que os profissionais dediquem-se menos a tarefas repetitivas e mais à interpretação crítica dos documentos. O uso de transcritores automáticos, como aponta Lose (2025), confere velocidade ao processo de transcrição, mas não elimina a necessidade do olhar humano, que atua como restaurador capaz de decifrar nuances gráficas da mesma forma que um conservador de obras de arte recupera os contornos de uma fotografia antiga desbotada.

No entanto, esse panorama tecnológico ainda apresenta lacunas significativas quando se trata da língua portuguesa. Enquanto o inglês dispõe de vastos conjuntos de dados e modelos consolidados, o português (pt-pt e pt-br) avança em ritmo mais lento devido à escassez de ferramentas adaptadas às suas especificidades gráficas (Silva et al., 2023). Essa desigualdade evidencia não apenas a relevância, mas também a originalidade de pesquisas que busquem suprir tais carências, visto que a inclusão do português nesse campo amplia a diversidade linguística contemplada pela IA e assegura maior autonomia e soberania científica à comunidade de língua portuguesa.

2.3 HANDWRITING TEXT RECOGNITION

O avanço das tecnologias baseadas em IA tem possibilitado o desenvolvimento de sistemas capazes de realizar o reconhecimento automático de escrita manuscrita, conhecidos como HTR. Esses sistemas têm sido amplamente utilizados em projetos de digitalização e acesso ao patrimônio documental, especialmente no contexto das Humanidades Digitais, onde contribuem para ampliar a disponibilidade de fontes históricas (Ciula; Eide, 2017; Burdick et al., 2012).

Do ponto de vista técnico, o HTR baseia-se em modelos de aprendizado de

máquina, particularmente redes neurais, que são treinadas a partir de conjuntos de dados previamente transcritos. Esses modelos aprendem a identificar padrões gráficos presentes na escrita manuscrita, convertendo imagens de documentos em texto digital (Goodfellow; Bengio; Courville, 2016).

Na área chamada de Visão Computacional surge o desenvolvimento do HTR, uma tecnologia derivada de Optical Character Recognition (OCR) que busca transcrever, por meio de aprendizado de máquina, textos manuscritos, transpondo-os para o digital e transformando-os em editáveis. Tal tecnologia possibilita que uma massa documental extensa possa ser transcrita com mais agilidade, promovendo a difusão do seu conteúdo informacional e, por consequência, do patrimônio documental.

Diversos fatores podem interferir para o melhor uso do HTR, entre eles estão: a qualidade da imagem que contém o texto original e a variação na forma como o manuscrito foi feito, já que o mesmo trata-se de caligrafia, portanto podendo variar. A tecnologia de HTR vem se popularizando em anos recentes, e já é usada por diversas plataformas ao redor do mundo, como a Google Cloud Vision, a Microsoft Azure Computer Vision, entre outras. Sendo uma tecnologia recente, diversas pesquisas vêm sendo desenvolvidas em torno do HTR nos últimos anos, contudo, no Brasil ainda têm-se poucas pesquisas sobre o tema no contexto de aplicação no patrimônio documental.

Portanto, pode-se observar que a utilização da tecnologia de HTR já vem impactando de forma significativa o trabalho de arquivistas e profissionais da informação de diferentes áreas, que trabalham ou se interessam por acervos documentais. Em uma palestra sobre o uso de IA para transcrição paleográfica de manuscritos, realizada pela Fundação Biblioteca Nacional (FBN) em outubro de 2025, foi relatado o desenvolvimento de um protótipo de ferramenta de transcrição no contexto brasileiro, mais precisamente no Arquivo Nacional (Canal Fundação Biblioteca Nacional, 2025).

O pesquisador Henrique Denes Hilgenberg Fernandes, coordenador de Planejamento, Acompanhamento e Avaliação do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), lidera a formação de uma equipe para coordenar

o projeto, cujo objetivo é aplicar técnicas modernas de IA, combinando OCR e HTR, para automatizar a transcrição de manuscritos históricos. O foco do projeto está nos documentos da Missão Cruls[1] e nas certidões do Arquivo Nacional.

Sendo assim, a equipe iniciou o trabalho com uma revisão de literatura para identificar as ferramentas disponíveis para o projeto, percebendo que a maioria dos projetos semelhantes estava inserida no contexto europeu. Entre as plataformas analisadas estavam: Transkribus, eScriptorium (Scripta PSL), Arkubdex (Teklia), ABBYY FineReader Server.

Ademais, diversos fatores foram determinantes para a decisão de construir uma ferramenta própria baseada em LLM para o Arquivo Nacional, incluindo: controle e soberania sobre os dados; especialização no domínio paleográfico brasileiro; transparência e auditabilidade; privacidade institucional; autonomia e evolução contínua; confiança e legitimidade (Canal Fundação Biblioteca Nacional, 2025).

Um fato curioso destacado foi que, durante a etapa inicial de aprimoramento dos modelos de transcrição, foi considerada a utilização de IA generativa para melhorar as transcrições. No entanto, a ideia foi descartada devido a preocupações com possíveis vieses introduzidos pelo modelo pré-treinado, como correções ortográficas automáticas que poderiam alterar palavras conforme a ortografia histórica, comprometendo a fidelidade aos documentos originais.

Sendo assim, indo ao encontro de Lose (2025), os avanços tecnológicos não substituem o trabalho do especialista em Paleografia e Diplomática, mas, ao contrário, reafirmam sua posição central como revisor e garantidor da autenticidade do patrimônio documental. Isso significa que o uso HTR não deve eliminar a mediação humana, mas a reposiciona: se antes o trabalho do paleógrafo era essencialmente manual, hoje sua análise é aplicada na validação crítica das transcrições automáticas.

[1]A Missão Cruls foi uma expedição científica realizada entre 1892 e 1893, sob a liderança do astrônomo belga naturalizado brasileiro Luís Cruls. Seu objetivo principal era delimitar a área onde seria instalada a futura capital federal, conforme previa a Constituição Republicana de 1891. MONTE-MÓR, Roberto Luís de Melo. Brasília: da utopia à cidade. Belo Horizonte: Editora UFMG, 2003.

Ainda, como aponta Silva (2025), a tecnologia amplia as possibilidades de acesso, mas é o conhecimento humano que assegura a fidedignidade do processo. A integração entre máquinas e especialistas, portanto, deve ser compreendida como complementar, em que a IA potencializa a velocidade e o alcance, enquanto o humano preserva a precisão e a legitimidade das fontes.

Com o objetivo de situar o debate no contexto contemporâneo, apresentamos a seguir um conjunto de trabalhos que aplicaram a tecnologia de HTR em suas investigações. Alguns desses estudos a exploram de forma empírica, demonstrando suas potencialidades e desafios na prática. Essa breve revisão busca delinear o estado da arte da tecnologia de HTR, destacando como a mesma vem sendo incorporada às práticas de pesquisa, tratamento e mediação do patrimônio documental.

Já citado anteriormente nesta pesquisa, Nockels et al., (2024) analisaram a precisão e a eficiência do HTR na digitalização de manuscritos, destacando o potencial desta tecnologia para transformar a forma como interagimos com os acervos de bibliotecas, arquivos e museus. Através de uma revisão sistemática da literatura, os autores argumentam que a simples introdução do HTR não garantirá uma revolução na investigação acadêmica, mas que as suas aplicações e efeitos sociais são complexos e multifacetados.

Sendo assim, o estudo propõe um conjunto de recomendações para investigadores e instituições, visando orientar futuras investigações e práticas que considerem tanto os benefícios como os desafios associados à utilização do HTR. Portanto, em suas conclusões, destacam a necessidade de uma abordagem crítica em relação ao HTR, enfatizando que a sua implementação deve ser fundamentada numa compreensão aprofundada do seu impacto sobre a produção de conhecimento.

Ademais, os autores alertam para o risco de simplificar a narrativa que associa a eficiência do HTR com transformações profundas na pesquisa histórica. Em vez disso, dizem que a tecnologia deve ser compreendida como um instrumento que, quando utilizado de forma apropriada, tem o potencial de enriquecer as interações com documentos históricos e fomentar o surgimento de novas formas de

investigação ao patrimônio documental.

O estudo analisa a diversidade no acesso às plataformas de Humanidades Digitais, com foco no Programa de Bolsa do Transkribus (Transkribus Scholarship Programme/TSP). Utilizando uma abordagem mista, eles investigam a demografia dos usuários do TSP, o trabalho possibilitado pelo programa e como isso pode informar um acesso mais equitativo às plataformas de Humanidades Digitais (Nockels et al., 2024).

As principais descobertas incluem que a maior parte dos candidatos ao TSP é composta por estudantes, principalmente de universidades europeias. Embora os beneficiários venham de instituições que não têm participação acionária na plataforma Transkribus, ainda existem desafios em termos de acessibilidade (Nockels et al., 2024).

O estudo propõe uma série de recomendações para provedores de plataformas e agências de financiamento, destacando a necessidade de suporte financeiro e treinamento adequado para promover uma maior diversidade na pesquisa. As recomendações incluem o desenvolvimento de políticas transparentes para acesso gratuito à plataformas e a comunicação contínua com os usuários para entender melhor suas necessidades, tal questão pode ajudar a fomentar o desenvolvimento de novas pesquisas em contextos de países economicamente deficitários.

Ao encontro de tais questões, os autores Colavizza et al. (2021) buscaram, em seu trabalho, compreender e explorar o impacto da IA e a sua crescente integração nos processos arquivísticos, impulsionada pela digitalização massiva e pela necessidade de automação. Segundo os autores, a adoção da IA tem vindo a transformar o papel dos arquivistas, que agora precisam interpretar e validar os resultados gerados por algoritmos, além de enfrentar desafios éticos, como o viés algorítmico e a transparência. O artigo destaca o modelo Records Continuum para compreender o impacto da IA em todas as etapas do ciclo de vida documental.

Além da automação da catalogação e descrição de documentos, a IA é também utilizada para otimizar a acessibilidade e a pesquisa em arquivos, com ferramentas como o OCR e a análise semântica. A utilização da IA para lidar com o

grande volume de documentos digitais gerados na era da internet, incluindo redes sociais e registros automatizados, vem sendo exponencialmente aumentado.

Contudo, a pesquisa salienta a necessidade de uma análise crítica sobre o impacto da IA, destacando a importância de uma participação mais ativa dos arquivistas na concepção dessas tecnologias, de modo a garantir o respeito aos princípios arquivísticos, tais como a autenticidade e a proveniência. Os autores também ressaltam a interdisciplinaridade como elemento fundamental para o sucesso da aplicação da IA na gestão arquivística (Colavizza et al., 2021).

Ainda pensando nas possíveis aplicações da IA no contexto do patrimônio documental, Maness e Pham (2022) abordam a utilização de HTR em documentos históricos na coleção de nome Jewish Consumptives Relief Society, um sanatório para tuberculosos que esteve em funcionamento entre 1904 e 1954 em Denver, no Colorado (EUA). O projeto recorreu à plataforma Transkribus para a transcrição de registos manuscritos e para a sua subsequente análise computacional.

Além dos benefícios desta tecnologia para a investigação acadêmica e o ensino, os autores salientam desafios éticos na disponibilização desses dados, que contêm informações médicas e pessoais de indivíduos vulneráveis, defendendo a intervenção humana para garantir que a utilização destas tecnologias respeite as comunidades envolvidas (Maness; Pham, 2022).

Ademais, na Biblioteca Nacional da França (BNF), o uso de HTR aplicados a diferentes acervos vem permitindo a implementação de programas experimentais, principalmente no portal Gallica (Bibliothèque Nationale De France, [s.d.]), que compreende e disponibiliza centenas de coleções manuscritas de documentos datados do século XVII e além. Consequentemente, considerando o passado histórico da França enquanto nação e a sua reconhecida valorização das artes, da literatura e do património documental, observa-se que o portal Gallica, aliado à aplicação de tecnologias de HTR nos acervos de documentos históricos, evidencia a relevância do tema na atualidade.

Portanto, estas iniciativas não apenas permitem uma reavaliação e reinterpretação do passado da história francesa e da Europa, como também contribuem para a preservação, difusão e democratização do acesso às fontes

primárias. Desta forma, simultaneamente à recuperação da memória escrita da França, é gerado um ambiente de inovação que consolida a posição do país como referência internacional na integração entre a tradição cultural e as tecnologias digitais emergentes.

No contexto brasileiro, o dataset LHTR.br constitui um marco ao oferecer dados anotados em português brasileiro, incluindo letras acentuadas e símbolos especiais que representam desafios adicionais para os algoritmos (Silva et al., 2023). Esses esforços ainda são pontuais, mas reforçam a importância da expansão de recursos para a língua portuguesa como elemento diferenciador e estratégico no avanço do HTR.

Encerrando este conjunto de relatos, observa-se que as experiências apresentadas oferecem um panorama de diferentes contextos e abordagens sobre o uso da tecnologia HTR em diferentes contextos e abordagens, permitindo assim compreender tanto as potencialidades quanto as limitações envolvidas no uso de HTR.

Portanto, a partir do que foi apresentado sobre a tecnologia de HTR, o próximo subcapítulo tem como foco a ferramenta Transkribus. Enquanto neste capítulo foram examinados casos que aplicam a HTR em diversas perspectivas, o seguinte dedicará atenção à estrutura, funcionamento e relevância do Transkribus como ferramenta de IA e ambiente de aplicação prática e de mediação informacional baseada em transcrição de documentos.

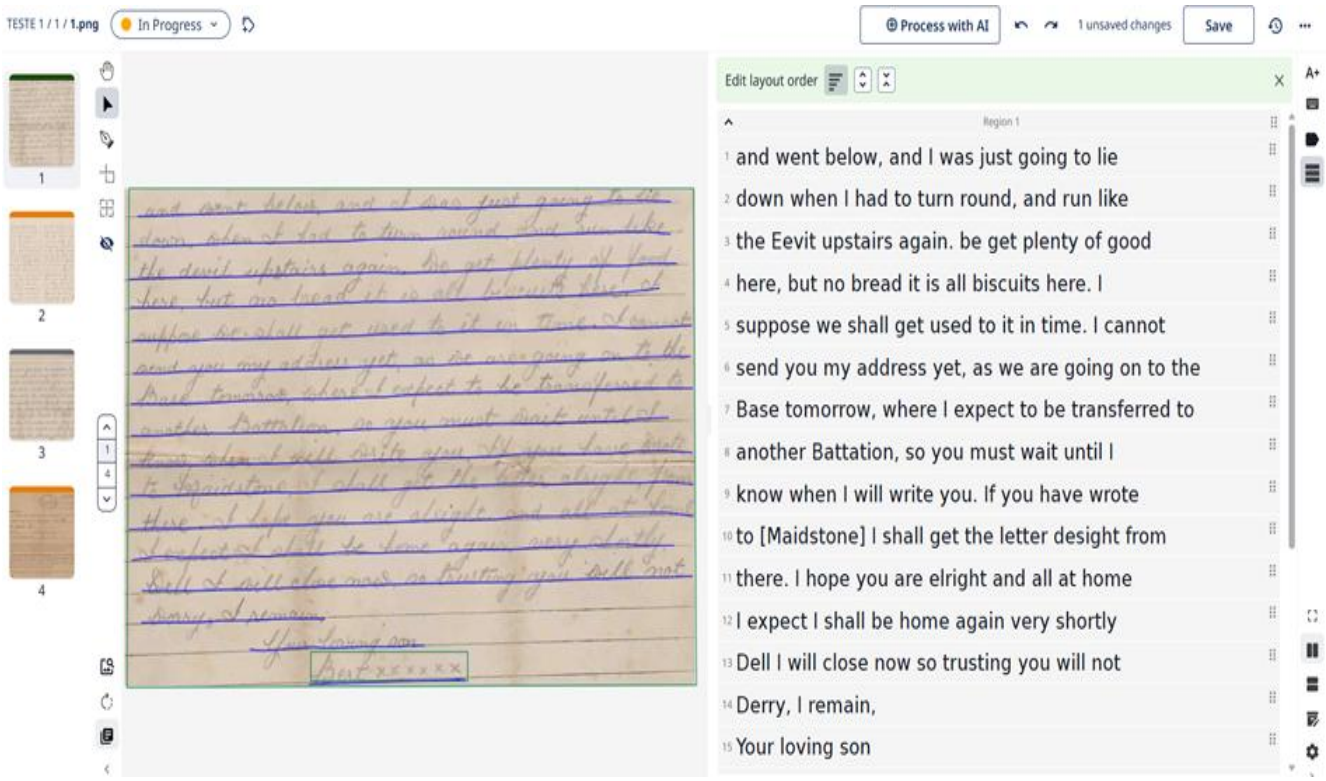
2.4 TRANSKRIBUS COMO AMBIENTE SOCIOTÉCNICO

A ferramenta Transkribus busca, por meio de HTR, realizar transcrições paleográficas digitais, assim processando documentos digitalizados e aplicando modelos de transcrição disponibilizados na própria plataforma. Para tal, se faz necessário o upload de pelo menos quatro documentos, assim como a consequente escolha de modelo treinado para aplicação e escaneamento. Estes modelos, disponibilizados dentro da plataforma, são baseados em aprendizado de máquina,

e permitem a separação em linhas de texto, transcrevendo e, quando necessário, possibilitando que o usuário corrija as interpretações erradas ou falhas do sistema.

As falhas mais comuns se dão em razão da variabilidade de diferenças em uma própria caligrafia, acentuação e pontuação e, em alguns casos, como consequência, letras podem ser repetidas ou mal interpretadas pelo sistema.

Figura 1 – Exemplo de transcrição no Transkribus



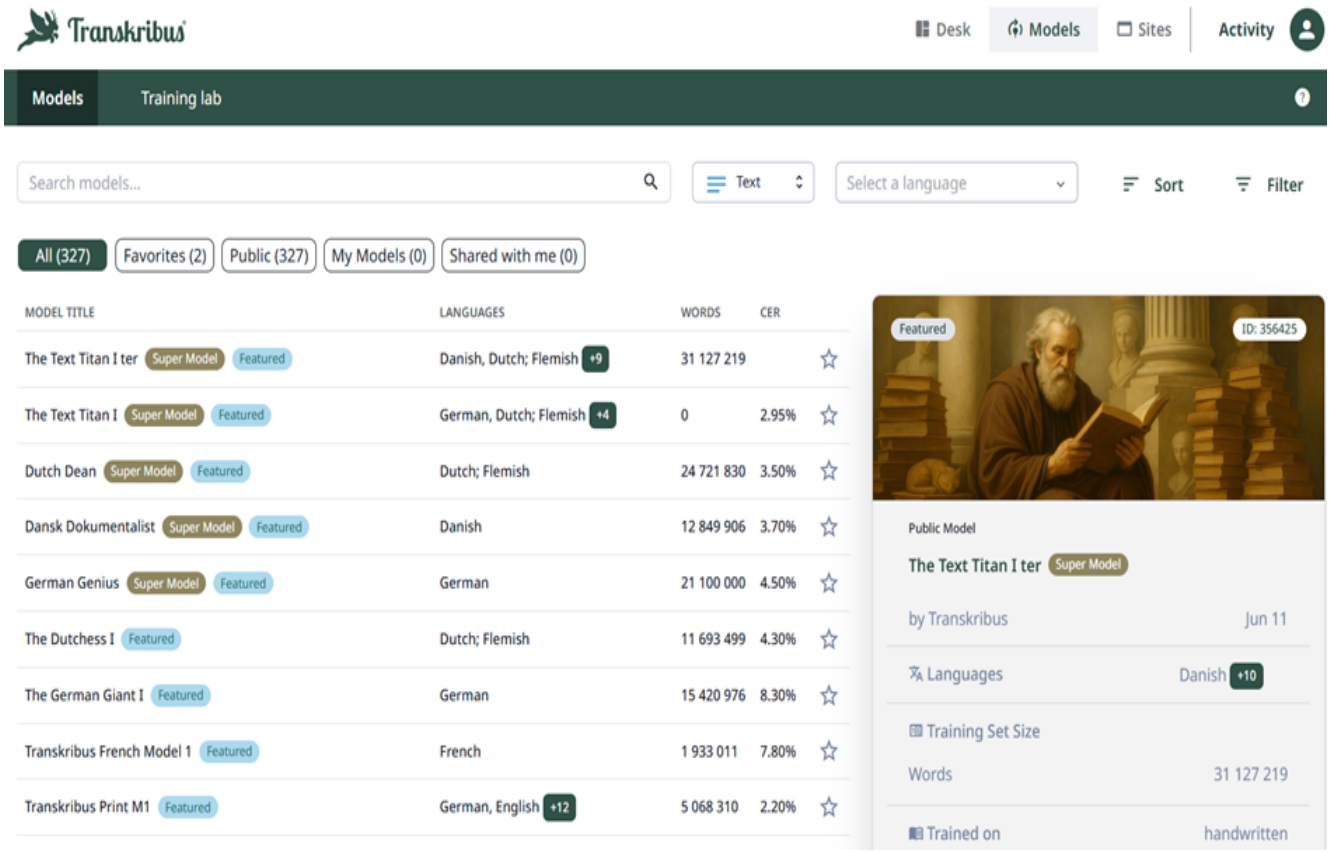
Fonte: Transkribus (2025).

Portanto, cada modelo deve apresentar informações detalhadas, como sua precisão em CER, data de criação, contexto de criação, número de palavras usadas para treinamento, idiomas e tipo de documento a que se aplica (manuscritos, datilografados ou impressos) entre outras (READ-COOP, 2024).

A precisão varia de 0% a 100%, com valores mais baixos indicando maior precisão. Por exemplo, o modelo "The English Eagle" apresenta um CER de 8,30%, significando que, em média, oito caracteres por cada cem são transcritos

incorretamente. Nesse sentido, a criação de modelos eficazes irá depender de grandes volumes de documentos já transcritos de forma correta, e de uma diversidade de caligrafias, assim otimizando os resultados da transcrição pelo sistema.

Figura 2 – Modelos de transcrição no Transkribus



Fonte: Transkribus (2025).

Se faz relevante destacar que a plataforma Transkribus não se enquadra na categoria de software livre, contudo, a troca de experiências entre os usuários é fundamental. Os modelos públicos, criados por diversos usuários, podem ser editados para receberem contribuições de outros usuários da plataforma. Nesse sentido, é essencial seguir os protocolos estabelecidos pelo criador do modelo, garantindo a consistência dos dados.

Sendo assim, o desenvolvimento de modelos com alto grau de precisão requer a atuação de um grupo de pesquisadores qualificados, preferencialmente inseridos em núcleos ou redes de pesquisa consolidadas, condição que favorece maior rigor metodológico e aprimora a qualidade das informações utilizadas no treinamento dos modelos. Portanto, a falta de incentivo financeiro pode limitar a criação de novos modelos aplicados ao contexto de países com menor condição financeira.

Ademais, o Transkribus já disponibiliza modelos específicos para o português (pt.pt), como o “General Portuguese M1”, que obteve um CER de 3,80%. Modelos de transcrição vêm sendo usados pela Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa, Portugal (Baudry, 2024). Um caso exemplar é o modelo "TraPriInq", desenvolvido pela Fundação para a Ciência e a Tecnologia (FCT) em Portugal. Este modelo foi treinado com mais de 6 mil páginas de documentos históricos da Inquisição Portuguesa, produzidos entre 1536 e 1821, alcançando um CER de 5,20%, sendo assim, altamente eficaz para este tipo de acervo.

A avaliação da precisão de modelos de transcrição é realizada por meio do CER e do Word Error Rate (WER), ambas derivadas da distância de Levenshtein. Essa métrica é usada em sistemas de correção ortográfica, reconhecimento de texto e buscas aproximadas, permitindo que a máquina identifique palavras parecidas, mesmo quando há erros de digitação ou variações na escrita (Alkendi et al., 2024).

Contudo, mais do que a arquitetura dos modelos, a qualidade do material de origem se mostra determinante para o melhor desempenho do HTR, ou seja, documentos degradados, com rasuras, manchas e sobreposição de caracteres, desafiam os algoritmos em sua capacidade de reconhecimento. A corrosão da tinta, por exemplo, impacta mais fortemente do que o simples desvanecimento do papel, pois quando o contraste desaparece, o sistema perde completamente sua referência de leitura (Roling, 2020).

Nesse sentido, técnicas de pré-processamento, como a compensação de iluminação ou a binarização (separação entre a tinta e o fundo da imagem), funcionam como uma restauração digital preliminar, permitindo que os algoritmos mantenham a legibilidade mesmo em condições adversas (Alkendi et al., 2024).

Ainda no contexto educacional, mais precisamente na Universidade de Leiden, na Holanda, estudantes aprenderam a transcrever manuscritos e treinar modelos de HTR. Alcançando taxas de erro inferiores a 4% para documentos manuscritos e 2% para documentos impressos (Caers, 2024). Ademais, no 33º simpósio BOBCATSSS, realizado em Istambul (2025), foram apresentados resultados que indicam a eficiência do uso do Transkribus em acervos de documentos históricos.

O trabalho comparou os resultados de um modelo de transcrição de nome “German Giant I”, e uma transcrição feita manualmente. Os resultados de aferição de precisão foram de 4,48%, indicando que cerca de quatro caracteres por cada cem são transcritos incorretamente, ou seja, o modelo mostrou alta taxa de precisão (Freiling; Rether, 2025).

No contexto de aplicação, o Transkribus tem sido utilizado em projetos de pesquisa voltados ao patrimônio documental, evidenciando seu potencial não apenas técnico, mas também científico. O estudo de Lose et al. (2024), ao desenvolver modelos de HTR para documentos inquisitoriais em língua portuguesa, demonstra que a plataforma permite não apenas a automatização da transcrição, mas também a ampliação do acesso às fontes e o desenvolvimento de novas investigações em diferentes áreas do conhecimento.

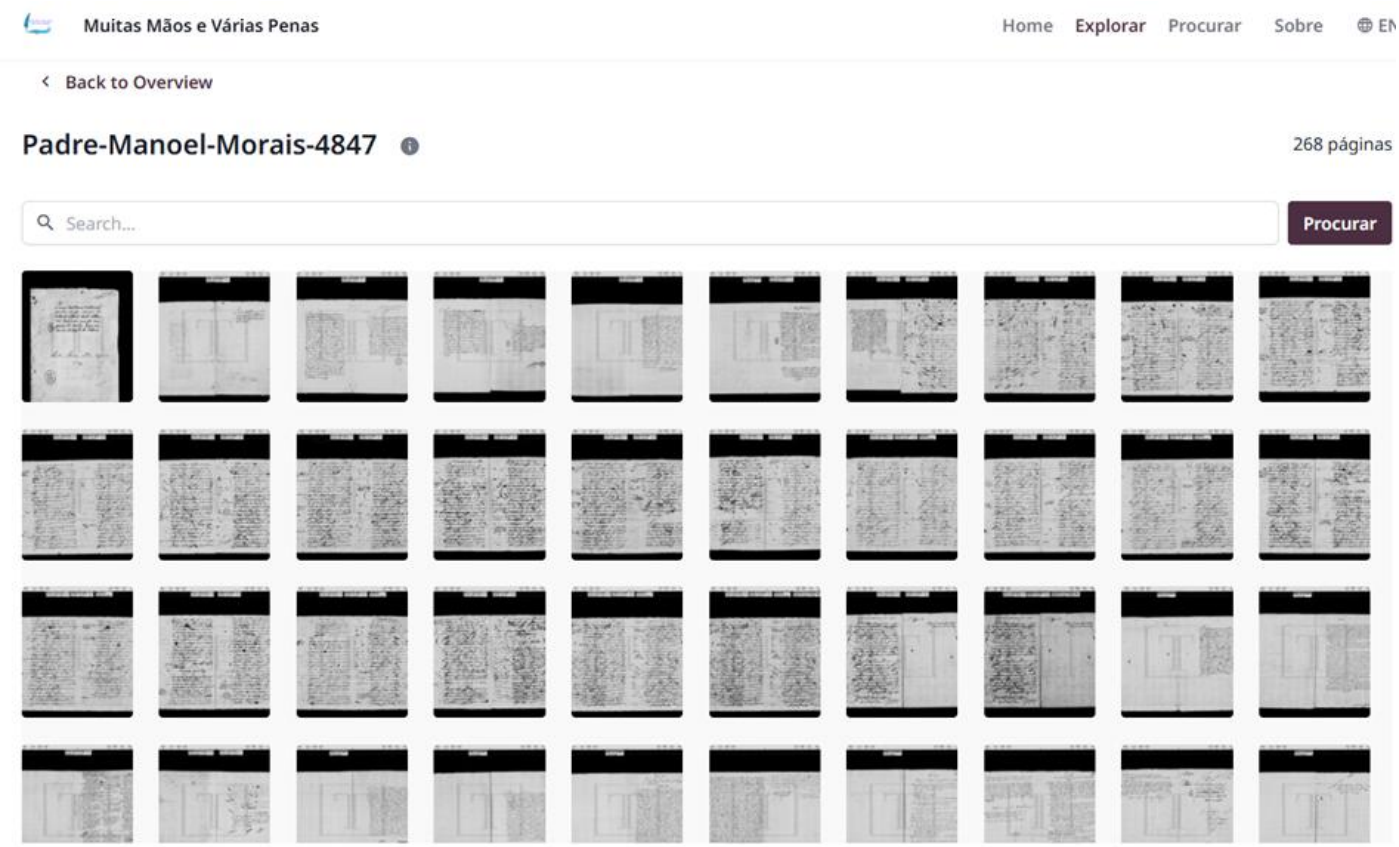
Sendo assim, tais iniciativas demonstram o evidente potencial da ferramenta Transkribus e do HTR no fomento à preservação e à análise de documentos históricos ao redor do mundo, integrando diferentes contextos linguísticos e culturais. Assim, o Transkribus emerge como uma plataforma que não só possibilita, mas também fomenta a captura e troca de modelos de transcrição, de modo a otimizar as transcrições paleográficas em todo o mundo.

A aplicação desta técnica de Paleografia Digital e a sua otimização em diferentes conjuntos de documentos de diversas origens (década, país, idioma) e culturas pode auxiliar até mesmo em causas humanitárias e reivindicação de direitos, já que busca difundir manuscritos antes limitados ao acervo físico.

Além do HTR, para transcrição de documentos, cabe destacar que o Transkribus oferece também a possibilidade de criação de sites, espaços virtuais nos quais instituições ou pesquisadores podem disponibilizar seus acervos, ou parte deles, para a comunidade acadêmica e para o público em geral. Esses sites

funcionam como uma vitrine dos acervos, permitindo não apenas o acesso remoto aos documentos, mas também a exploração dos resultados das transcrições e da aplicação de HTR, ampliando as possibilidades de difusão e uso do patrimônio documental.

Figura 3 – Exemplo de site dentro do Transkribus



Fonte: Transkribus (2025).

Já é possível observar, dentro do Transkribus, a existência de diversos sites públicos mantidos por instituições nacionais e internacionais, nos quais se encontram acervos de documentos históricos, muitos deles já transcritos e acessíveis a qualquer interessado. Esses exemplos demonstram o potencial do recurso, tanto para a preservação digital e difusão dos acervos, quanto para o fortalecimento da pesquisa em diferentes áreas do conhecimento, além de possibilitar ao público em geral uma aproximação com fontes documentais antes restritas a espaços físicos específicos.

Essa funcionalidade se apresenta, portanto, como um recurso importante para a democratização do conhecimento, já que promove maior visibilidade para acervos que, muitas vezes, permanecem restritos em arquivos ou bibliotecas. A disponibilização em ambiente digital não só garante maior alcance geográfico e social, como também fomenta pesquisas interdisciplinares, estimulando colaborações entre áreas como a CI, a Arquivologia e as Humanidades Digitais.

No entanto, cabe destacar que a criação e manutenção de sites no Transkribus estão condicionadas a modelos de acesso pagos, o que pode representar uma limitação significativa, especialmente para instituições e pesquisadores com recursos financeiros restritos. Arquivos públicos de pequeno porte, iniciativas independentes e projetos comunitários frequentemente não dispõem de orçamento para custear esse tipo de serviço, o que tende a restringir o potencial de democratização do acesso ao patrimônio documental. Assim, um recurso com potencial impacto social amplo pode acabar concentrado em grupos que possuem financiamento.

Ainda que o Transkribus ofereça um plano gratuito, com a concessão de 50 créditos mensais, essa modalidade atende sobretudo a usos iniciais, como testes ou projetos de pequena escala (READ-COOP, 2024). Embora esses créditos possibilitem avanços pontuais, eles se mostram insuficientes para demandas, que envolvem grandes volumes documentais ou fluxos contínuos de trabalho.

Para atender a essas demandas, a plataforma disponibiliza planos pagos, como Scholar, Team e Organization, que ampliam as capacidades de armazenamento, processamento, treinamento de modelos e colaboração entre usuários. Além disso, há a possibilidade de aquisição de créditos adicionais sob demanda e a oferta de créditos gratuitos por meio de programas específicos, como o Transkribus Scholarship Programme (TSP), voltados a estudantes e docentes.

Nesse cenário, observa-se que as condições de acesso à tecnologia não são distribuídas de forma equitativa, o que pode produzir assimetrias no uso de ferramentas de HTR no campo do patrimônio documental. Instituições com maior disponibilidade de recursos tendem a ampliar sua capacidade de processamento, descrição e disponibilização de acervos, enquanto aquelas com restrições orçamentárias enfrentam limitações técnicas e operacionais. Como consequência, há impactos não apenas no acesso aos documentos, mas também na própria produção de conhecimento, uma vez que diferentes atores passam a dispor de condições desiguais para interpretar, transcrever e difundir informações provenientes desses acervos.

Dessa forma, a incorporação de tecnologias como o Transkribus, embora amplie significativamente as possibilidades de tratamento de documentos manuscritos, também evidencia a necessidade de reflexão crítica sobre os condicionantes econômicos que permeiam seu uso, especialmente no que diz respeito à democratização do acesso e às desigualdades no acesso e uso da informação no campo do patrimônio documental.

Portanto, ao relacionar essas condições de acesso ao conceito de difusão da informação, percebe-se que a gratuidade limitada dos créditos contribui para a circulação inicial da informação em pequenos acervos, mas as barreiras financeiras impostas nos planos mais amplos reduzem o alcance social da difusão. Ainda sobre o conceito de difusão da informação, González de Gómez (2002, p. 63) observa que a difusão da informação deve ser entendida como um processo de mediação que “reorienta os fluxos informacionais para públicos diversificados”.

Essa ideia aproxima-se do conceito de difusão arquivística que, conforme Fonseca (2005), constitui-se como uma das funções sociais do arquivo, justamente por permitir que a informação documental cumpra seu papel de bem cultural e social. Nesse contexto, a existência de limitações financeiras no Transkribus pode ser compreendida como um entrave, pois restringe a difusão àqueles que dispõem de recursos. Portanto, embora a plataforma contribua para a preservação e disponibilização de documentos históricos, o acesso remunerado pode distanciar a perspectiva de difusão inclusiva da informação e da função social dos arquivos e de seus acervos, especialmente quando pensamos em contextos socioeconômicos restritos.

Concluída a análise do Transkribus, o capítulo seguinte abordará os princípios da gamificação no âmbito do patrimônio documental. Essa discussão servirá de base para a metodologia apresentada posteriormente, na qual será proposto um modelo gamificado que incorpora o uso do Transkribus como elemento central de experimentação e aprendizagem.

2.5 GAMIFICAÇÃO E FORMAÇÃO EM AMBIENTES DIGITAIS

A gamificação pode ser compreendida como o uso de elementos de jogos em contextos não lúdicos, com o objetivo de promover engajamento, motivação e aprendizagem (Deterding et al., 2011). No contexto educacional, essa abordagem tem sido utilizada como estratégia para estimular a participação ativa dos sujeitos e favorecer processos de construção do conhecimento, especialmente em contraposição a modelos tradicionais centrados na passividade discente.

No âmbito da formação em CI, a gamificação apresenta potencial para contribuir com o desenvolvimento de competências relacionadas ao uso crítico de tecnologias, na medida em que permite a simulação de situações reais de atuação profissional. Por meio de desafios, feedbacks e atividades estruturadas, é possível criar ambientes de aprendizagem que aproximam teoria e prática, favorecendo a participação ativa e a aprendizagem baseada em interação contínua (Lee; Hammer, 2011; Alves; Minho; Diniz, 2014; Silva; Sales; Castro, 2019).

Entretanto, é importante destacar que a gamificação não deve ser compreendida apenas como mecanismo de motivação, mas como estratégia pedagógica que precisa estar alinhada a objetivos de aprendizagem claramente definidos. A simples inserção de elementos de jogo não garante a efetividade do processo formativo, sendo necessário considerar aspectos como a estrutura das atividades, a progressão das tarefas e os resultados esperados (Kapp, 2012). Além disso, sua efetividade depende de planejamento e acompanhamento adequados, podendo inclusive gerar efeitos limitados ou superficiais quando aplicada de forma descontextualizada (Alves; Minho; Diniz, 2014).

No campo das evidências empíricas, há indícios de que a gamificação pode impactar positivamente a aprendizagem. Em investigação quase-experimental no ensino de Física, estudantes submetidos a uma abordagem gamificada apresentaram ganho de aprendizagem superior em comparação ao modelo tradicional, indicando efeitos relevantes no desempenho acadêmico (Silva; Sales; Castro, 2019). No entanto, tais resultados devem ser interpretados com cautela, considerando sua limitação contextual e a necessidade de ampliação de estudos empíricos na área.

No contexto desta pesquisa, a gamificação é compreendida como estratégia para o desenvolvimento de competências relacionadas ao uso de HTR, especialmente no que se refere à validação de transcrições, à escolha de modelos e à interpretação dos resultados. Ao simular situações de uso dessas tecnologias, a proposta formativa busca promover uma compreensão crítica dos processos de mediação algorítmica.

Além disso, a utilização de abordagens gamificadas permite explorar a dimensão colaborativa do trabalho com documentos, incentivando a interação entre participantes e a construção coletiva do conhecimento. Essa característica aproxima a proposta das práticas observadas em ambientes digitais colaborativos, nos quais a produção da informação ocorre de forma distribuída e participativa, envolvendo os usuários como agentes ativos no processo de construção e validação dos dados (Benkler, 2006; Prahalad; Ramaswamy, 2004; Busciantella-Ricci; Scataglini, 2024).

No âmbito das unidades de informação, como arquivos, bibliotecas e museus, a gamificação surge como uma estratégia frente ao desafio contemporâneo de engajamento, especialmente diante da necessidade de garantir a preservação e a difusão de acervos culturais. A transcrição de documentos históricos exemplifica esse cenário, uma vez que se trata de uma atividade que exige atenção prolongada, persistência e competências específicas, sendo frequentemente marcada por dificuldades de engajamento (Silva; Sales; Castro, 2019).

Ademais, López-Nores et al (2019), por exemplo, apresentam o sistema CuneiForce como uma proposta que combina gamificação e crowdsourcing para envolver usuários não especialistas na anotação de tábuas cuneiformes ainda não analisadas, demonstrando o potencial da participação coletiva para apoiar a pesquisa em patrimônio cultural.

Nesse sentido, a gamificação pode contribuir para reposicionar o usuário no centro do processo de construção do conhecimento, em consonância com os pressupostos da Competência em Informação, que envolvem identificar necessidades informacionais, buscar, avaliar, interpretar e aplicar criticamente a informação (Nascimento; Santos; Mata, 2024). As dinâmicas presentes nos jogos, ao integrarem objetivos, desafios e feedback contínuo, favorecem o desenvolvimento dessas habilidades em contextos de ensino e pesquisa (Alves; Minho; Diniz, 2014).

Além disso, a gamificação pode ser compreendida também como uma estratégia de mediação da informação, na medida em que atua não apenas na disponibilização de conteúdo, mas na forma como os sujeitos interagem, interpretam e atribuem sentido à informação. Ao estruturar experiências baseadas em participação ativa, feedback contínuo e progressão de tarefas, cria-se um ambiente no qual o usuário deixa de ser receptor e passa a atuar como agente no processo informacional, aproximando-se de abordagens que compreendem a mediação como elemento central na construção do conhecimento (Lloyd, 2010; Bawden, 2001).

Essa perspectiva torna-se particularmente relevante no contexto do uso de IA aplicada com o HTR, uma vez que a ampliação do acesso por meio da transcrição automatizada não garante, por si só, a compreensão dos documentos. Assim, a gamificação, enquanto estratégia de mediação, pode contribuir para transformar o acesso técnico em experiência significativa, promovendo maior envolvimento dos sujeitos com o patrimônio documental.

Elementos como pontos, insígnias, rankings e feedback imediato estruturam ambientes que incentivam não apenas a participação, mas também a autonomia e a aprendizagem autodirigida (Brazil; Albagli, 2020; Cativelli; Monsani; Juliani, 2016). No entanto, tais mecanismos devem ser utilizados com cautela, uma vez que podem tanto estimular quanto desmotivar participantes, especialmente quando associados a dinâmicas excessivamente competitivas (Alves; Minho; Diniz, 2014).

Nos arquivos e museus, a gamificação tem se destacado em projetos de crowdsourcing e ciência cidadã, nos quais voluntários colaboram com a transcrição e organização de acervos. Contudo, observa-se que o engajamento tende a se concentrar em um grupo reduzido de participantes altamente ativos, enquanto a maioria contribui de forma esporádica, evidenciando limites dessas estratégias (Eveleigh et al., 2014 apud Brazil; Albagli, 2020).

Diante disso, a articulação entre gamificação e IA abre novas possibilidades, como a personalização das experiências de aprendizagem e o ajuste do nível de dificuldade das tarefas de acordo com as competências dos participantes, favorecendo a manutenção do engajamento (Koivisto; Hamari, 2019). Além disso, a IA pode atuar como suporte à mediação, oferecendo feedback contínuo e contribuindo para processos formativos mais dinâmicos (Chen; Lin, 2020 apud Soares Netto; Duque, 2025).

Assim, o uso articulado de gamificação e HTR pode contribuir não apenas para a organização e o acesso a acervos documentais, mas também para a formação de usuários capazes de interagir criticamente com essas tecnologias. Em contextos formativos, como workshops, essa integração possibilita a simulação de desafios paleográficos complexos, estimulando a reflexão e a aprendizagem coletiva.

Contudo, tais práticas devem ser conduzidas sob orientação crítica e ética, garantindo que a participação dos usuários seja reconhecida e revertida em acesso ao conhecimento produzido. Nesse sentido, destaca-se a importância do letramento informacional e tecnológico dos profissionais que atuam com IA no patrimônio documental (Bawden, 2001; Lloyd, 2010; Rockembach, 2020).

Além disso, questões relacionadas à ética e à privacidade de dados tornam-se centrais nesse contexto, especialmente diante do uso de sistemas baseados em IA. Como destaca Rockembach (2020), é fundamental considerar legislações como a LGPD e o RGPD, a fim de mitigar riscos relacionados à privacidade, à discriminação e à segurança da informação.

Ainda que a gamificação seja amplamente utilizada como estratégia para promover engajamento em ambientes digitais, é importante reconhecer que o envolvimento dos usuários não se restringe à incorporação de elementos típicos de jogos. Abordagens baseadas na participação ativa, como aquelas fundamentadas em práticas de co-design, também têm demonstrado potencial para mobilizar os sujeitos em tarefas complexas, por meio da colaboração, da reflexão e da construção coletiva de soluções. Nesse contexto, experiências baseadas em oficinas participativas evidenciam que fatores como interação social, senso de contribuição e possibilidade de influência sobre o sistema podem atuar como mecanismos de engajamento tão relevantes quanto sistemas de pontuação ou recompensas.

Nesse sentido, estudos que exploram metodologias participativas demonstram que a interação direta dos usuários com sistemas digitais, por meio de atividades colaborativas de análise e proposição, permite identificar limitações, aprimorar interfaces e promover maior compreensão dos processos informacionais envolvidos (Rockembach et al., 2025; Monteiro-Krebs, 2022; Melo, 2025). Tais experiências evidenciam que elementos como clareza, transparência e controle do usuário são centrais para a construção de confiança e usabilidade, reforçando a ideia de que o engajamento pode emergir não apenas de estímulos lúdicos, mas também da participação ativa no processo de mediação.

Retomando as metodologias colaborativas e os processos de gamificação, em outubro de 2024 ocorreu o AI4Culture Workshop, evento voltado à colaboração comunitária por meio da exploração de plataformas colaborativas que incentivam a participação pública na validação e no enriquecimento de transcrições de documentos históricos.

Nesse evento foi apresentado o Transkribathon (Europeana Foundation, [s.d.]), uma plataforma que visava à maratona de transcrição colaborativa e competitiva promovida pela plataforma, com o objetivo de incentivar a transcrição, anotação e enriquecimento de documentos históricos manuscritos e impressos. A plataforma combina elementos de gamificação, ciência cidadã e educação para promover o engajamento de diversos públicos, desde pesquisadores a estudantes no trabalho com acervos históricos disponíveis em seu site.

Sendo assim, iniciativas como o Transkribathon evidenciam o potencial da articulação entre gamificação, ciência cidadã e plataformas digitais colaborativas, promovendo o engajamento de diferentes públicos na transcrição e enriquecimento de documentos históricos. Tais experiências reforçam a relevância de estratégias que integrem tecnologia, participação e mediação, ampliando as possibilidades de acesso e apropriação do patrimônio documental.

Neste capítulo, buscou-se apresentar uma contextualização teórica a partir de autores e estudos da área de gamificação, evidenciando seus principais conceitos e potencialidades no contexto educacional e informacional. Esses conteúdos constituem o fundamento para a proposição de um modelo gamificado de uso do Transkribus, que será desenvolvido nos capítulos seguintes. Na sequência, serão apresentados os procedimentos metodológicos que orientam a construção e análise dessa proposta.

3 PROCEDIMENTOS METODOLÓGICOS

A presente pesquisa foi estruturada em duas fases metodológicas consecutivas e interdependentes, com o objetivo de articular diagnóstico empírico e proposição teórica no contexto do uso de IA aplicada ao patrimônio documental.

A primeira fase, de caráter diagnóstico e exploratório, compreendeu uma análise crítica fundamentada em revisão sistemática da literatura, articulada a uma experiência piloto com 48 estudantes inseridos em uma dinâmica de uso da ferramenta Transkribus. Essa etapa refere-se a resposta ao objetivo especifico b, que teve como finalidade avaliar a usabilidade, a curva de aprendizado e as percepções operacionais de usuários iniciantes durante a interação com o Transkribus, a partir de dados primários coletados em um teste piloto exploratório.

Com base nesses elementos, desenvolveu-se a segunda fase metodológica da pesquisa, de caráter propositivo. Essa etapa foi construída a partir dos achados da fase anterior, tomando como referência o levantamento bibliográfico e as dificuldades e limitações identificadas na interação dos usuários com o Transkribus, de modo a subsidiar a elaboração de uma proposta teórica orientada à mediação de tais questões.

3.1 FASE DIAGNÓSTICA E EXPLORATÓRIA

A seguir, a revisão sistemática de literatura é descrita no Quadro 1:

Quadro 1 – Procedimentos metodológicos adotados na revisão sistemática

Objetivo	Mapear trabalhos nacionais e internacionais que abordem a temática de HTR, Transkribus, gamificação relacionados ao contexto de Patrimônio Documental, buscando assim refletir sobre as potencialidades da ferramenta, assim como propor um uso gamificado da mesma.

Equações da pesquisa	Análise da literatura; refletir sobre a potencialidade da ferramenta; Proposição de uso gamificado do Transkribus.
Âmbito da pesquisa	Dimension.ai; BRAPCI (Base de dados em Ciência da Informação)
Critérios de inclusão	Artigos de revistas científicas lançados até outubro de 2025.
Critérios de exclusão	Artigos em idiomas que não sejam o espanhol, inglês e português; Artigos sem acesso aberto; Artigos com menos de 1 citação.
Critérios de validade metodológica	Verificação dos critérios de inclusão e exclusão.
Resultados	Descrição da pesquisa e registro dos procedimentos.
Tratamento de dados	Análise com o apoio de fichamentos para organização dos dados.

Fonte: Elaborado pelo autor (2025).

Para a compilação bibliográfica, foram usados os repositórios Dimensions.ai[2] e a BRAPCI[3]. Foram realizadas duas buscas no Dimensions.ai. Na primeira foram utilizados os respectivos termos: "Handwriting recognition" OR "Handwritten text recognition" AND "Archival". Com a utilização do filtro “Title and abstract” aplicado aos termos, a busca resultou em 32 trabalhos. Contudo, 7 foram excluídos da pesquisa, pois mantinham acesso restrito (pago) ou foram escritos em alfabeto cirílico. Sendo assim, 26 artigos foram exportados através do Dimension.ai para o Quadro 2 e, após isso, foram mantidos os artigos que

[2] DIMENSIONS. *Dimensions*. Disponível em: https://www.dimensions.ai/. Acesso em: 27 jan. 2026.

[3] BRAPCI. *Base de Dados em CI*. Disponível em: https://brapci.inf.br/home. Acesso em: 27 jan. 2026.

tinham pelo menos 1 citação, com as respectivas informações: Título; Ano de publicação; Autores e Número de citações.

Já a segunda busca no Dimensions.ai utilizou os termos: "Transkribus" AND "Archival", com a utilização do filtro “Title and abstract” aplicado. A busca gerou 26 trabalhos. Porém, 6 desses eram repetidos da primeira consulta e outros 8 trabalhos estavam em alfabeto cirílico, ou em idiomas como italiano, alemão e eslovaco. Portanto, a partir disso, foram mantidos 12 trabalhos, dos quais representamos no Quadro 3 os que tinham pelo menos 1 citação.

Na BRAPCI, os termos de pesquisa foram: Handwriting Text Recognition; Transkribus; Paleografia Digital. Obteve-se somente 2 textos, ambos do periódico Cadernos BAD (Portugal), datados de 2025.

Além disso, ainda na fase diagnóstica e exploratória da pesquisa, após o levantamento bibliográfico, foi realizado um exercício piloto com 48 estudantes, com o objetivo de testar a aplicação da ferramenta Transkribus. Ao final da atividade, foi aplicado uma pesquisa de opinião de forma voluntária e anônima, obtendo-se retorno de 10 participantes. A pesquisa teve como objetivo captar percepções iniciais acerca da experiência, do uso da tecnologia e das estratégias adotadas.

A atividade consistiu na utilização do Transkribus para a transcrição de documentos relacionados à Segunda Guerra Mundial, envolvendo etapas estruturadas de registro na plataforma (a partir da leitura de um manual, ver Apêndice A), organização de coleções documentais, aplicação de modelos automáticos de transcrição e posterior correção manual dos resultados.

A metodologia adotada se construiu a partir da seleção de quatro documentos (Anexo A, Anexo B, Anexo C, e Anexo D), sendo três manuscritos e um híbrido, a aplicação de modelos de HTR e a avaliação dos resultados, possibilitando não apenas a familiarização dos estudantes com a tecnologia, mas também a análise de suas potencialidades e limitações em contexto educacional.

A coleta de dados ocorreu de forma totalmente anônima, sem possibilidade de identificação direta ou indireta dos respondentes, caracterizando-se como pesquisa de opinião com participantes não identificados. Nesse sentido, o estudo enquadra-se no disposto na Resolução do Conselho Nacional de Saúde nº

510/2016, estando dispensado de submissão ao sistema CEP/CONEP (BRASIL, 2016).

O instrumento de coleta consistiu em um questionário estruturado, composto por questões fechadas e abertas, abordando os seguintes aspectos: avaliação geral da atividade; percepção sobre o uso do Transkribus; nível de dificuldade na realização das tarefas; interesse em utilizar a tecnologia em outros contextos; e espaço aberto para comentários, sugestões e críticas.

Entre as questões fechadas, foram utilizadas escalas de avaliação (níveis de satisfação e facilidade de uso) e perguntas objetivas sobre a experiência dos participantes. Já as questões abertas possibilitaram a expressão livre das percepções dos estudantes acerca da atividade.

3.2 FASE PROPOSITIVA

O modelo gamificado foi projetado para apresentar aos participantes conceitos introdutórios de Paleografia, bem como oferecer treinamento inicial no uso da ferramenta Transkribus e discutir seus possíveis impactos no contexto de documentos históricos. Ressalta-se, contudo, que os participantes do estudo piloto não possuem formação prévia em Paleografia, sendo, em sua maioria, estudantes sem experiência anterior na leitura e interpretação de manuscritos históricos. Essa característica não é tratada como limitação, mas como elemento constitutivo de sua proposta analítica.

Nesse sentido, a ausência de conhecimento prévio permite observar de forma mais evidente o papel da mediação tecnológica proporcionada pelo HTR, especialmente no que se refere à sua capacidade de apoiar processos iniciais de aproximação com documentos manuscritos. Ao mesmo tempo, essa condição também evidencia os limites dessas ferramentas, uma vez que a interpretação documental não se restringe ao reconhecimento de caracteres, mas envolve dimensões contextuais, históricas e cognitivas que extrapolam o processamento automatizado.

Se faz relevante destacar ainda, que o objetivo do instrumento gamificado não consiste em avaliar o desempenho paleográfico absoluto dos participantes, tampouco comparar níveis de proficiência técnica na leitura de manuscritos. O foco recai sobre

a análise das condições de interação entre sujeitos iniciantes e documentos históricos, em dois cenários distintos: a transcrição manual e a transcrição mediada por HTR.

Dessa forma, busca-se compreender aspectos como percepção de dificuldade, tempo despendido, taxa de erro e experiência de uso, considerando a mediação da informação como eixo central de análise. Tal delimitação permite reduzir a interferência de variáveis relacionadas à formação prévia dos participantes, direcionando a investigação para o papel da tecnologia enquanto agente mediador no acesso ao patrimônio documental.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

A análise dos resultados desta pesquisa evidenciou, por um lado, a existência de um movimento internacional recente voltado à formulação de diretrizes para a governança da IA em contextos educacionais, culturais, patrimoniais e arquivísticos e, por outro, a materialização concreta dessas discussões a partir da experiência prática com o uso da tecnologia de HTR no ambiente do Transkribus. A seguir, como resultado da primeira busca de referências temos o Quadro 2.

Quadro 2 – Levantamento de referências da 1ª busca

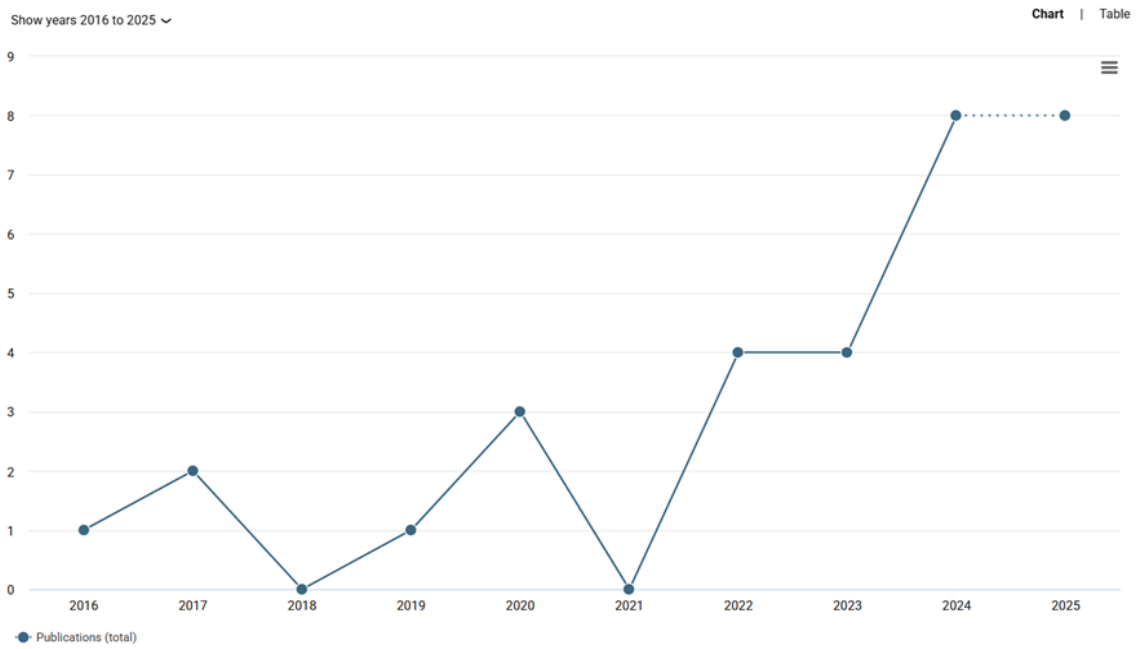
Título	Ano	Autor(es)	Citações
Transforming scholarship in the archives through handwritten text recognition	2019	Muehlberger, Guenter; Seaward, Louise; Terras, Melissa; et al.	99
ICDAR2017 Competition on Handwritten Text Recognition on the READ Dataset	2017	Sánchez, Joan Andreu; Romero, Verónica; Toselli, Alejandro H.; Villegas, Mauricio; Vidal, Enrique	53
Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of Transkribus in published research	2022	Nockels, Joe; Gooding, Paul; Ames, Sarah; Terras, Melissa	46
Automatic CNN-Based Arabic Numeral Spotting and Handwritten Digit Recognition by Using Deep Transfer Learning in Ottoman Population Registers	2020	Can, Yekta Said; Kabadayı, M. Erdem	21
CNN-Based Page Segmentation and Object Classification for Counting Population in Ottoman Archival Documentation	2020	Can, Yekta Said; Kabadayı, M Erdem	13
Experimenting with Training a Neural Network in Transkribus to Recognise Text in a Multilingual and Multi-Authored Manuscript Collection	2023	Capurro, Carlotta; Provatorova, Vera; Kanoulas, Evangelos	9

The implications of handwritten text recognition for accessing the past at scale	2024	Nockels, Joseph; Gooding, Paul; Terras, Melissa	8
Handwritten text recognition using deep learning techniques: A survey	2024	Rakesh, S.; Reddy, P. Kushal; Prashanth, V.; Reddy, K. Srinath	7
Exploring Data Provenance in Handwritten Text Recognition Infrastructure: Sharing and Reusing Ground Truth Data, Referencing Models, and Acknowledging Contributions. Starting the Conversation on How We Could Get It Done	2024	Romein, C. Annemieke; Hodel, Tobias; et al.	5
Recognizing text lines in handwritten archival document images using octave convolutional and attention recurrent neural networks	2024	Mechi, Olfa; Mehri, Maroua; Ingold, Rolf; Essoukri Ben Amara, Najoua	2
TexT - Text Extractor Tool for Handwritten Document Transcription and Annotation	2017	Hast, Anders; Cullhed, Per; Vats, Ekta	2
Digitization Of Tulu Handwritten Scripts- A Literature Survey	2024	Sandhya Bangera; Subrahmanya Bhat	1
AI and archive. Handwritten Text Recognition Applied to Patrimonial Holdings: An Example of 10 diaries written by Spanish Republican Teachers in 1932	2022	Pajares, Pepita Raventós; Tornero, Celio Hernández; Martin, Meritxell Simon	1
Handwritten Text Recognition of Historical Manuscripts Using Transformer-Based Models	2025	Meoded, Erez	1
IA y archivo	2023	Raventós Pajares, Pepita; Hernández Tornero, Celio; Simon Martin, Meritxell	1
Introduction: Content, Design and Structure of Major Databases with Historical Longitudinal Population Data	2023	Alter, George; Mandemakers, Kees; Vézina, Hélène	1

Fonte: Dimensions.ai (2025)

Nota-se, a partir da Figura 4, que há um crescimento acentuado das pesquisas sobre a temática de HTR na última década, especialmente nos últimos 5 anos.

Figura 4 – Gráfico de referências da 1ª busca



Fonte: Dimensions.ai (2025)

Portanto, indicando não apenas um crescimento quantitativo, mas também uma consolidação recente do campo. Esse comportamento sugere que o avanço do HTR está diretamente vinculado ao amadurecimento das tecnologias de IA, especialmente no que se refere a modelos de aprendizado profundo, que passaram a oferecer melhores resultados em documentos históricos. A partir da segunda busca, foi construído o Quadro 3:

Quadro 3 – Levantamento de referências da 2ª busca

Título	Ano	Autor(es)	Citações
Artificial Intelligence in archival and historical scholarship workflow: HTS and ChatGPT	2023	Spina, Salvatore	8
Transcribing Foucault's handwriting with Transkribus	2019	Massot, Marie-Laure; Sforzini, Arianna; Ventresque, Vincent	6
Are Digital Humanities platforms facilitating sufficient diversity in research? A study of the Transkribus Scholarship Programme	2024	Nockels, Joseph; Gooding, Paul; Terras, Melissa	3
Sobre Big Data E Neopositivismo Digital Na Pesquisa Em História	2024	Gil, Tiago Luís	3
Transkribus im Archiv – Ein polnisch-deutsches Projekt zur Handschriftentexterkennung an historischen Dokumenten	2021	Alvermann, Dirk; Gut, Paweł	2
From Digitization and Images to Text and Content: Transkribus as a Case Study	2024	Prebor, Gila	1

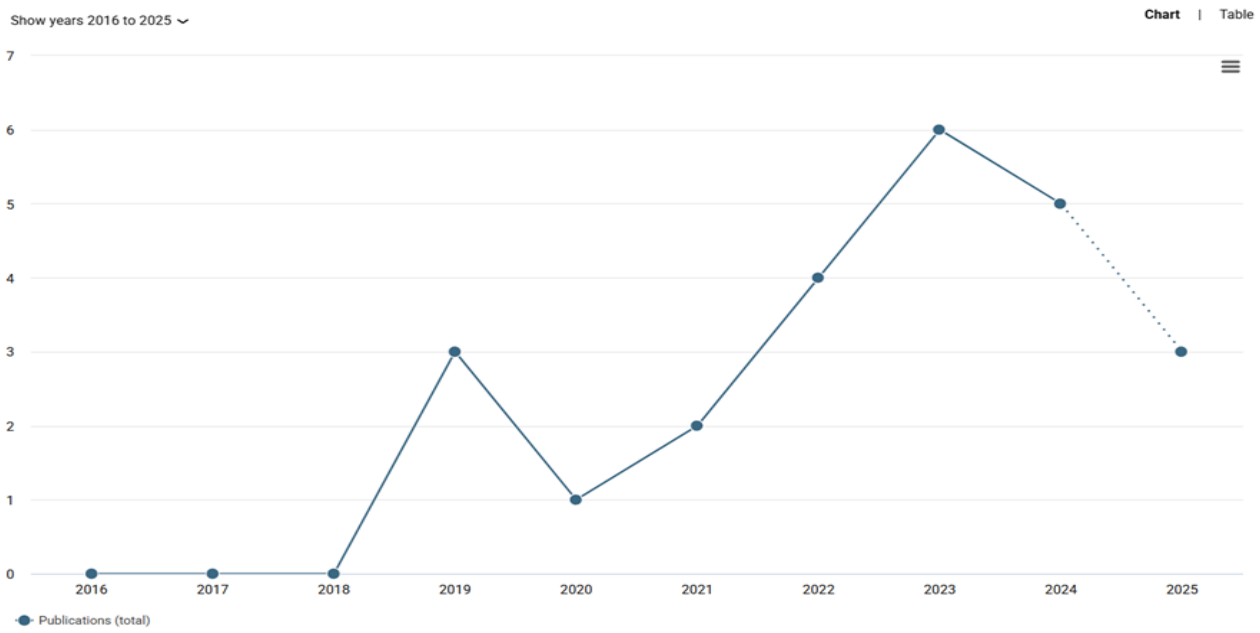
Fonte: Dimensions.ai (2025)

Mais do que um simples aumento no volume de estudos, esse recorte temporal revela uma inflexão no interesse científico, em que o HTR deixa de ser uma tecnologia experimental e passa a ocupar um espaço mais estruturado nas agendas de pesquisa, particularmente em iniciativas voltadas ao patrimônio documental e às Humanidades Digitais. Ao mesmo tempo, esse crescimento recente também reforça o caráter ainda em desenvolvimento da área, o que implica

reconhecer tanto suas potencialidades quanto suas limitações, especialmente quando analisadas sob a perspectiva da mediação da informação no campo da CI.

Assim como na primeira busca, nota-se na Figura 5 que também houve um crescimento acentuado de trabalhos sendo desenvolvidos em anos recentes.

Figura 5 – Gráfico de referências da 2ª busca



Fonte: Dimensions.ai (2025)

Entre 2023 e 2026, organismos como UNESCO, ONU e OCDE, bem como instâncias técnico-profissionais como o Conselho Internacional de Arquivos (ICA) e o InterPARES, passaram a produzir recomendações que convergem em torno de princípios como centralidade humana, supervisão institucional, transparência, mitigação de vieses e proteção da diversidade cultural, ainda que não constituam um regime unificado de governança (UNESCO, 2023; OCDE, 2023).

Nesse cenário, a experiência realizada a partir do AI4Culture Workshop e da plataforma Transkribathon (Europeana Foundation, [s.d.]), articulada ao desenvolvimento teórico desta pesquisa, possibilitou a realização de um estudo com 48 estudantes da disciplina de “Introdução às Humanidades Digitais”, em uma universidade portuguesa. Esse tipo de abordagem prática dialoga diretamente com

as diretrizes da UNESCO, que apontam a necessidade de desenvolver competências críticas e operacionais no uso da IA, integrando dimensões técnicas, éticas e pedagógicas (UNESCO, 2024a; UNESCO, 2024b).

Com base no referencial teórico apresentado, especialmente nas discussões acerca da mediação da informação, do uso de IA em acervos documentais e das estratégias de gamificação em contextos educacionais, foi estruturado o modelo gamificado proposto. Tal modelo busca articular dimensões técnicas, perceptivas e analíticas do uso do HTR, permitindo uma abordagem integrada do fenômeno investigado.

Sendo assim, o levantamento de dados se deu a partir da análise de trabalhos e experiências descritos na literatura no âmbito do patrimônio documental, com foco em iniciativas que articulam tecnologias digitais, ensino e acesso a acervos históricos. Esses referenciais subsidiaram a construção do modelo gamificado e orientaram a definição de variáveis analíticas, especialmente no que diz respeito à interação entre usuários iniciantes e sistemas de HTR.

O modelo gamificado utiliza documentos da Segunda Guerra Mundial (Anexo A, Anexo B, Anexo C e Anexo D), selecionados a partir de acervos de livre acesso ao público. A escolha desse recorte temporal justifica-se por se tratar de um evento amplamente conhecido e consolidado no imaginário coletivo, frequentemente representado em diversas mídias, como filmes, literatura e jogos. Essa presença cultural favorece um reconhecimento imediato do contexto histórico, reduzindo barreiras cognitivas e interpretativas no processo de aprendizagem.

Dessa forma, ao utilizar documentos desse período, o modelo gamificado se beneficia de um contexto narrativo familiar, permitindo que o foco dos participantes se concentre não na compreensão de um contexto histórico desconhecido, mas na experiência de análise, interpretação e interação com as fontes documentais em si. Isso contribui para isolar, ainda que parcialmente, as variáveis relacionadas ao uso da tecnologia e à mediação da informação, reduzindo interferências decorrentes de dificuldades contextuais.

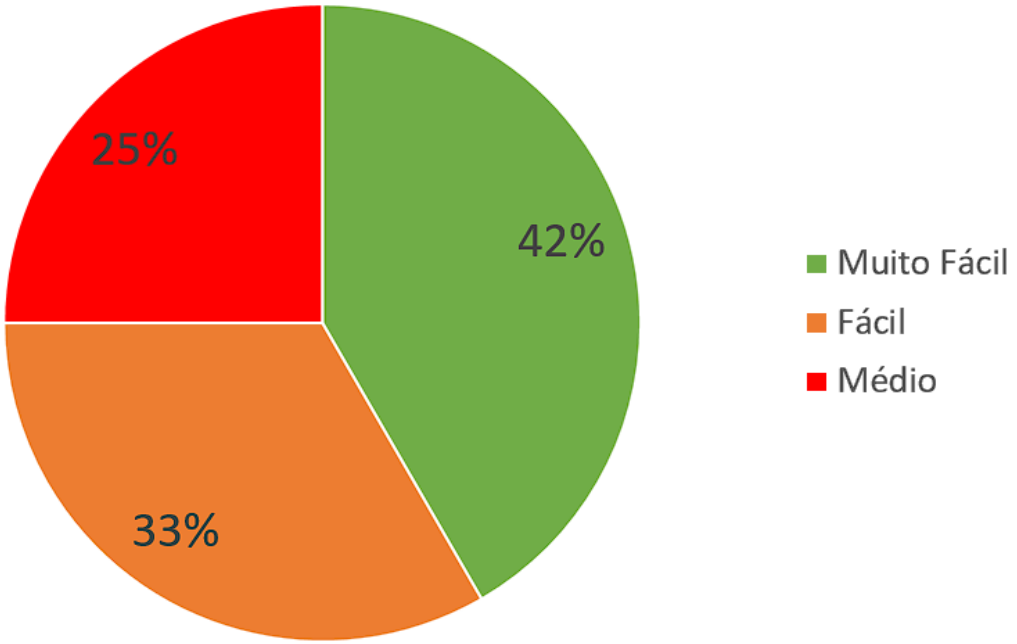
Por fim, a caracterização do nível de experiência dos participantes em Paleografia torna-se um elemento central para a análise dos resultados, permitindo compreender como sujeitos sem formação específica interagem com documentos manuscritos em diferentes condições, manualmente e com apoio de ferramentas de

HTR e quais são as implicações dessa interação para o acesso, uso e interpretação do patrimônio documental.

Nesse sentido, a opção por participantes sem formação prévia em Paleografia não apenas se justifica metodologicamente, mas também reflete transformações contemporâneas no acesso ao patrimônio documental, marcadas pela crescente presença de tecnologias baseadas em IA. Assim, o estudo aproxima-se de cenários reais de uso, nos quais sujeitos não especialistas passam a interagir com documentos históricos mediadas por sistemas automatizados, tensionando os limites entre acesso, interpretação e mediação da informação no contexto digital.

Os dados coletados a partir dos questionários são apresentados por meio da Figura 6, que sintetiza a percepção geral dos participantes, e do Quadro 4, que organiza as principais dificuldades relatadas. Esses elementos têm como finalidade oferecer uma visão inicial das experiências dos usuários no uso do Transkribus, sem, contudo, esgotar sua interpretação.

Figura 6 – Avaliação da facilidade de uso do Transkribus



Fonte: Elaborado pelo autor (2026).

Nesse sentido, a avaliação da facilidade de uso do Transkribus revela um resultado predominantemente positivo por parte dos participantes, uma vez que 42% consideraram a ferramenta de muito fácil usabilidade, 33% a classificaram como fácil e 25% indicaram um nível de dificuldade mediano, com nenhum dos alunos respondendo as categorias “difícil” ou “muito difícil”, por este motivo estas categorias foram excluídas da Figura 6.

Sendo assim, esses dados sugerem que, embora a ferramenta seja amplamente percebida como acessível, ainda existem elementos de complexidade que impactam parte dos usuários, especialmente em fases iniciais de uso. Ademais, no Quadro 4 os resultados indicam que o principal ponto de tensão no uso do HTR não está na automação da transcrição, mas na etapa de validação, onde se reintroduz a necessidade de mediação humana qualificada.

Quadro 4 – Síntese analítica das dificuldades de uso do Transkribus

Categoria	Frequência	Interpretação
Edição de erros de transcrição	70%	Principal gargalo operacional; exige grade nível de mediação humana
Navegação na interface	20%	Dificuldade inicial de usabilidade
Escolhas de modelos de transcrição	10%	Complexidade técnica
Carregamento de documentos	10%	Problemas pontuais

Fonte: Elaborado pelo autor (2026).

Para a construção do Quadro 4, foi possibilitado que os estudantes pudessem escolher mais de uma categoria, caso fosse necessário. Embora a ferramenta seja percebida como acessível e eficiente, seu uso evidencia limitações que exigem intervenção crítica por parte do usuário. Além disso, as dificuldades associadas à revisão das transcrições, à navegação inicial e à compreensão dos modelos indicam que o desafio não reside apenas no acesso à tecnologia, mas na mediação do processo de uso e interpretação.

Sendo assim, os dados obtidos por meio dos questionários evidenciaram que as dificuldades de engajamento, a fadiga inerente à leitura paleográfica e as limitações relacionadas à usabilidade não se configuram como obstáculos pontuais, mas como elementos estruturais da interação entre usuários iniciantes e ferramentas de HTR.

Nesse sentido, tais aspectos foram assumidos como insumos primários para a pesquisa, não com o intuito de validar soluções, mas de fundamentar a identificação das principais barreiras de usabilidade presentes no uso não mediado da ferramenta. A partir desse diagnóstico, tornou-se possível orientar, de forma mais precisa, a construção de uma proposta teórica que respondesse diretamente

às dificuldades observadas, estabelecendo uma relação clara entre problema identificado e solução projetada.

Nesse sentido, a proposta de gamificação foi concebida como uma resposta às limitações identificadas na etapa diagnóstica, tendo suas mecânicas, como missões, pontuações e recompensas colaborativas, estruturadas para mitigar os obstáculos evidenciados. Sendo assim, o modelo não se configura como uma proposição arbitrária, mas como uma construção metodologicamente orientada a partir dos problemas observados no uso do Transkribus em contextos de aprendizagem.

A estrutura do modelo gamificado organiza-se em cinco etapas: introdução, treinamento, execução, validação e síntese, conforme apresentado no Quadro 6. Essa organização visa simular um ambiente de aprendizagem no qual os participantes assumem o papel de pesquisadores, articulando prática, reflexão e análise crítica. Tal estrutura dialoga com princípios de progressão, feedback e construção ativa do conhecimento, amplamente discutidos na literatura de gamificação.

Quadro 5 – Estrutura do modelo gamificado

Etapa	Descrição	Elementos de Gamificação	Finalidade	Fundamentação Teórica
Introdução	Contextualização dos documentos e definição dos objetivos. (transcrição e análise)	Narrativa inicial, definição de missão.	Engajamento inicial e orientação da atividade.	Deterding et al. (2011); Werbach e Hunter (2012).
Treinamento	Apresentação dos conceitos básicos de Paleografia e do funcionamento do Transkribus.	Preparação dos participantes. (onboarding)	Nívelamento.	Kapp (2012).
Execução - Transcrição manual	Realização da transcrição sem auxílio automatizado.	Missões por níveis de dificuldade, sistema de pontos.	Engajamento e estímulo à persistência em tarefa complexa.	Werbach e Hunter (2012); Deterding et al. (2011); Hamari e Koivisto (2019).
Execução - Transcrição com HTR	Uso do Transkribus para transcrição dos mesmos documentos.	Missões por níveis, comparação de desempenho.	Compreensão das potencialidades e limitações da HTR.	Werbach e Hunter (2012); Deterding et al. (2011); Muehlberger et al. (2019); Terras et al. (2024).
Validação	Revisão crítica das transcrições.	Pontuação baseada em métricas (tempo, precisão, CER), progressão.	Avaliação de desempenho e desenvolvimento crítico.	Werbach e Hunter (2012); Hamari e Koivisto (2019).

Síntese	Discussão final e comparação entre os métodos.	Feedback contínuo, reforço de aprendizagem.	Consolidação do conhecimento e reflexão crítica.	Kapp (2012); Deterding et al. (2011).

Fonte: elaborado pelo autor (2026).

Do ponto de vista teórico, o modelo fundamenta-se na compreensão da gamificação como o uso de elementos de jogos em contextos não lúdicos (Deterding et al., 2011), sendo operacionalizada por meio de estruturas como narrativa, missões, pontuação e progressão. Tais elementos são organizados conforme abordagens de design de sistemas gamificados (Werbach; Hunter, 2012), que enfatizam o alinhamento entre mecânicas, objetivos e experiência do usuário. No contexto educacional, essa abordagem é complementada pelas contribuições de Kapp (2012), que destaca o potencial da gamificação para promover engajamento, reduzir a carga cognitiva inicial e favorecer a aprendizagem ativa.

Além disso, estudos empíricos, como os de Hamari e Koivisto (2019), indicam que elementos como pontos, níveis e desafios contribuem para a manutenção do engajamento em atividades complexas e repetitivas, como a transcrição paleográfica. Nesse contexto, a gamificação atua como estratégia para sustentar a participação dos sujeitos diante da fadiga cognitiva inerente à tarefa. A organização em etapas permite, ainda, compreender o instrumento como um processo contínuo de mediação da informação, no qual diferentes momentos, desde a introdução até a síntese, contribuem para a construção de conhecimento.

Nesse processo, a inserção da etapa de transcrição com HTR articula o modelo gamificado à dimensão tecnológica da pesquisa, evidenciando que o uso de sistemas como o Transkribus não elimina a mediação humana, mas a reconfigura, deslocando o foco da leitura direta para a validação e interpretação das saídas algorítmicas. Assim, a proposta não se limita à aplicação de uma ferramenta, mas configura-se como um dispositivo de mediação sociotécnica, no qual diferentes agentes, humanos e algorítmicos, participam da produção, transformação e validação da informação.

Faz-se relevante destacar que o modelo gamificado proposto não se limita à organização de etapas operacionais, podendo ser compreendido, também, como

um sistema de mediação da informação no qual diferentes agentes, humanos e algorítmicos, participam da produção, transformação e validação dos registros documentais.

Nesse sentido, o Quadro 6 permite identificar distintas camadas de mediação, que se deslocam ao longo da atividade, passando de uma mediação predominantemente contextual e formativa para uma mediação cognitiva e, posteriormente, algorítmica e crítica. Tal dinâmica evidencia que o uso de tecnologias de HTR não elimina a mediação humana, mas a reconfigura, redistribuindo funções entre sujeitos e sistemas.

Sendo assim, a introdução da IA no processo de transcrição desloca o eixo da mediação da leitura direta para a validação e interpretação de saídas algorítmicas, exigindo do profissional da informação não apenas competências técnicas, mas também capacidade crítica para lidar com limitações, opacidades e vieses dos sistemas. Dessa forma, o modelo proposto explicita que a informação, no contexto da IA, não é simplesmente recuperada, mas continuamente produzida em um circuito de mediações sociotécnicas.

Quadro 6 – Camadas De Mediação No Modelo Gamificado

Etapa	Tipo de Mediação Predominante	Agentes Envolvidos	Operação Informacional	Papel da IA	Papel do Humano	Tensões Críticas
Introdução	Mediação contextual.	Professor, sistema e participante.	Enquadramento do documento e definição de sentido inicial.	Ausente.	Interpretação inicial e construção de sentido.	Dependência do contexto para compreensão.
Treinamento	Mediação formativa.	Professor, interface e participante.	Aquisição de repertório (Paleografia + Transkribus).	Suporte indireto (interface).	Aprendizagem e adaptação cognitiva.	Curva de aprendizagem e letramento digital.
Execução – Manual	Mediação cognitiva.	Participante	Decodificação e interpretação direta do manuscrito.	Ausente.	Central (leitura, inferência, interpretação).	Fadiga, ambiguidade, subjetividade.
Execução – HTR	Mediação algorítmica.	IA (HTR) e participante.	Processamento automático e geração de texto.	Central (reconhecimento e predição).	Validação, interpretação e correção.	Opacidade do modelo, erros sistemáticos.
Validação	Mediação crítica.	Participante e sistema.		Apoio (métricas).	Julgamento crítico e revisão.	Confiança excessiva na IA e rejeição à IA.

			Comparação, correção e avaliação (CER, tempo, precisão).			
Síntese	Mediação reflexiva.	Participante, grupo e professor.	Produção de sentido sobre o processo.	Indireta.	Reflexão e ressignificação da experiência.	Compreensão (ou não) da mediação algorítmica.

Fonte: Elaborado pelo autor (2026)

Sendo assim, como forma de operacionalizar os elementos de gamificação, foi desenvolvido um sistema de pontuação estruturado em níveis de dificuldade progressiva, associado a objetivos cognitivos específicos. Esse sistema tem como finalidade orientar a participação dos usuários, oferecendo feedback contínuo e incentivando a progressão nas tarefas propostas, conforme apresentado no Quadro 7.

Quadro 7 – Sistema de pontuação e recompensas

Missão/Nível de Dificuldade	**Descrição de Tarefa**	**Recompensas e Pontuação**	**Objetivo Cognitivo e Lúcido**
Missão 1 - Fácil	Transcrição de fragmentos curtos e legíveis.	+10 pontos por linha correta.	Desenvolver familiaridade com o processo de transcrição manual e com o HTR.
Missão 2 - Intermediário	Transcrição de textos parcialmente degradados.	+5 pontos por correção de erro de HTR.	Estimular atenção aos detalhes e habilidade de revisão colaborativa.
Missão 3 - Difícil	Revisão e validação de transcrições complexas, com alto grau de dificuldade, incluindo excesso de abreviaturas, rasuras e interferências visuais.	+50 pontos bônus por documento completo. -5 pontos por linhas não revisadas.	Promover persistência, precisão e consolidação das competências adquiridas.

Fonte: Elaborado pelo autor (2026).

O sistema de pontuação organiza a experiência em três níveis progressivos, incentivando persistência e feedback contínuo (Deterding et al., 2011). Contudo,

ressalta-se que, embora o sistema de pontuação contribua para o engajamento e a organização da atividade, ele não deve ser compreendido como um mecanismo neutro. A atribuição de recompensas pode influenciar o comportamento dos participantes, orientando suas ações para a obtenção de pontos, o que deve ser considerado na interpretação dos resultados em uma eventual aplicação do modelo.

A mecânica de perda de pontos apresentada no Quadro 7 possui caráter conceitual e deverá ser validada empiricamente em estudos futuros. Embora sua inclusão tenha sido concebida como um mecanismo para desencorajar validações descuidadas, pesquisas sobre iniciativas de crowdsourcing em instituições de memória indicam que estratégias baseadas predominantemente em reforço positivo, como reconhecimento, recompensas e feedbacks construtivos, tendem a ser mais eficazes para manter o engajamento e a participação contínua de voluntários.

Portanto, a adoção de mecanismos de penalização deverá ser avaliada quanto aos seus efeitos sobre a motivação, a permanência dos participantes e a qualidade das contribuições em aplicações futuras do modelo proposto (Ridge, 2014).

Os sistemas de recompensas, as missões e a pontuação descritas na gamificação (Quadro 7) não devem ser compreendidos como meros artifícios lúdicos ou elementos acessórios ao processo, mas como dispositivos estruturantes de uma estratégia de mediação sociotécnica. Em um contexto no qual a transcrição de manuscritos históricos se configura como uma atividade repetitiva, cognitivamente exigente e, por vezes, extenuante, tais mecanismos operam no sentido de sustentar o engajamento do sujeito ao longo da tarefa.

Dessa forma, a gamificação atua não apenas sobre a dimensão motivacional, mas também sobre a organização da experiência informacional, orientando percursos, oferecendo feedback contínuo e reconfigurando a relação entre o usuário e o sistema de HTR. Nesse sentido, trata-se de uma mediação intencionalmente projetada, que busca equilibrar as limitações da automação com a permanência indispensável da intervenção humana no processo de interpretação e validação documental.

Nesse sentido, considerando a natureza propositiva do modelo gamificado, foi elaborado um plano de análise que explicita as possibilidades de tratamento dos

dados em uma futura aplicação. Tal plano estrutura-se a partir da articulação entre dados quantitativos e qualitativos, permitindo uma abordagem integrada do fenômeno investigado, contemplando dimensões como desempenho técnico, percepções dos participantes e impacto dos elementos de gamificação.

Para tanto, foram definidos eixos analíticos que possibilitam examinar comparativamente as modalidades de transcrição manual e assistida por HTR, bem como compreender as implicações do uso dessas tecnologias no contexto da mediação da informação. A organização desses eixos busca integrar diferentes tipos de evidência, favorecendo uma análise mais abrangente e consistente do fenômeno.

Ressalta-se que o plano de análise apresentado refere-se a um cenário de aplicação futura do modelo, não correspondendo a dados empíricos coletados no âmbito desta pesquisa. A seguir, no Quadro 8, apresenta-se a estrutura do plano de análise proposto.

Quadro 8 – Plano de análise para futura aplicação no instrumento.

Eixo de Análise	Tipo de Dado Previsto	Produto Esperado (Tabela/Gráfico/Quadro)	Técnica de Análise	Finalidade Analítica
1. Análise Quantitativa.	Métricas de tempo de execução e taxa de erro (CER) nas tarefas de transcrição manual e assistida por HTR.	Gráfico de barras comparando o tempo médio por Nível de Dificuldade. (Manual vs. HTR)	Estatística descritiva e comparação de médias.	Avaliar o desempenho técnico projetado entre modalidades e identificar ganhos potenciais de tempo e precisão.
		Tabela comparativa do CER médio por modalidade.		
2. Análise Qualitativa. **(Percepções de Uso)**	Respostas a questionários. (Escala Likert e questões abertas)	Gráficos percentuais (pizza ou barras) para itens de escala Likert.	Análise de conteúdo e estatística descritiva simples.	Compreender percepções de usabilidade, clareza e confiança no instrumento.
		Quadro-síntese com categorias temáticas das respostas abertas.		

3. Análise de Engajamento.	Pontuação obtida nas missões, níveis de dificuldade e percepção de engajamento.	Tabela de correlação entre pontuação, CER e percepções qualitativas. Gráfico de dispersão mostrando relação entre desempenho e engajamento.	Correlação exploratória e análise interpretativa.	Investigar o impacto dos elementos gamificados (pontos, missões, desafios) sobre o engajamento e a qualidade da tarefa.
4. Estrutura de Níveis e Sistema de Pontos.	Parâmetros de design do jogo. (Missões e recompensas)	Tabela conceitual com regras de pontuação e progressão por missão.	Análise comparativa de design e modelagem conceitual.	Demonstrar coerência interna do sistema de gamificação e sua integração com os objetivos do instrumento.
5. Síntese Integrativa	Resultados agregados das análises anteriores.	Quadro integrador com achados esperados e implicações para o design do instrumento.	Análise interpretativa e triangulação de dados secundários.	Consolidar a coerência entre desempenho projetado, percepções de usabilidade e engajamento lúcido.

Fonte: elaborado pelo autor (2026).

O modelo gamificado foi projetado em cinco etapas sequenciais (Quadro 5 e Quadro 6), contemplando momentos de introdução, treinamento, execução de tarefas, validação e reflexão. Essa organização busca simular um ambiente de aprendizagem estruturado, no qual os participantes são inseridos em uma narrativa de missão, assumindo o papel de pesquisadores responsáveis pela recuperação de informações em documentos históricos.

A utilização de narrativas e objetivos claros está alinhada à literatura de gamificação, que aponta a importância de contextos significativos para a construção do engajamento e da motivação dos participantes (Kapp, 2012). Nesse sentido, a escolha de documentos da Segunda Guerra Mundial contribui para a criação de um cenário reconhecível, reduzindo barreiras cognitivas e favorecendo a imersão na atividade.

Na etapa de treinamento técnico, propõe-se a introdução ao uso da plataforma Transkribus, acompanhada de exercícios guiados de transcrição manual e assistida por HTR. Essa etapa dialoga com princípios de progressão gradual e aprendizagem orientada, considerados fundamentais para a efetividade de experiências gamificadas (Werbach; Hunter, 2012).

O núcleo da experiência consiste na execução de tarefas organizadas em níveis de dificuldade progressiva, incorporando elementos como sistema de pontos, missões e desafios. A estruturação em níveis e recompensas segue modelos amplamente discutidos na literatura, nos quais a progressão é utilizada como mecanismo de feedback contínuo e incentivo à persistência (Deterding et al., 2011; Kapp, 2012).

Na etapa de validação, propõe-se a comparação entre os resultados obtidos, bem como a identificação de limitações do sistema de HTR, acompanhada da aplicação de instrumentos de coleta de percepção. O uso de feedback e revisão é apontado como componente essencial em sistemas gamificados, na medida em que permite aos participantes refletirem sobre seu desempenho e ajustarem suas estratégias (Werbach; Hunter, 2012).

Por fim, a etapa de síntese prevê a integração dos resultados e a realização de discussões orientadas, permitindo que os participantes reflitam sobre as implicações do uso do HTR. Nesse contexto, a gamificação atua não

apenas como elemento motivacional, mas como estrutura organizadora da experiência de mediação da informação.

A literatura de gamificação indica que a utilização de elementos de jogos pode favorecer o engajamento e a participação ativa dos usuários, especialmente quando associada a objetivos claros e sistemas de feedback (Deterding et al., 2011; Kapp, 2012). No entanto, esses mesmos elementos podem influenciar o comportamento dos participantes, orientando suas ações para a obtenção de recompensas, o que pode impactar a forma como interagem com o conteúdo (Werbach; Hunter, 2012).

Nesse sentido, a gamificação não deve ser compreendida como mecanismo neutro, mas como um dispositivo que estrutura a experiência e influencia os modos de interação com a informação. Essa compreensão é particularmente relevante no contexto do HTR, no qual a mediação da informação envolve não apenas a leitura técnica, mas também processos interpretativos e cognitivos.

De um total de 48 estudantes, 10 responderam à pesquisa, e os resultados indicaram que 8 deles consideraram a utilização do Transkribus fácil ou muito fácil, enquanto uma parcela menor relatou dificuldades relacionadas principalmente à escolha dos modelos de transcrição e à etapa de correção manual. A facilidade percebida pode ser interpretada como indicativo da acessibilidade da interface e da maturidade da ferramenta em termos de usabilidade, reforçando seu potencial como recurso pedagógico.

Por outro lado, as dificuldades apontadas revelam um aspecto crítico: o uso efetivo da tecnologia não é completamente automatizado e exige competências específicas de interpretação, validação e tomada de decisão, especialmente no que se refere à seleção de modelos e à revisão dos resultados.

Esse dado é particularmente relevante quando articulado às diretrizes internacionais, que enfatizam que a IA não deve ser compreendida como substituta da ação humana, mas como tecnologia que requer supervisão e mediação qualificada (UNESCO, 2023; OCDE, 2024). A necessidade de correção manual observada no experimento reforça a perspectiva de que o HTR opera dentro de uma lógica de colaboração humano-máquina, aproximando-se de

abordagens como o Human-in-the-Loop, nas quais o julgamento humano permanece central no processo.

Além disso, a percepção de agilidade no processamento e a navegabilidade da plataforma evidenciam o potencial do Transkribus para o ensino de técnicas paleográficas e para a inserção de tecnologias digitais no campo das Humanidades Digitais e da CI. No entanto, à luz das discussões propostas pela OCDE, tais benefícios devem ser analisados em conjunto com possíveis riscos, como dependência tecnológica, superficialização de habilidades e desigualdade de acesso (OCDE, 2023; OCDE, 2024).

A experiência também confirma o potencial do Transkribus como ferramenta eficaz para o enfrentamento dos desafios da transcrição paleográfica, contribuindo para a ampliação do acesso a documentos históricos e para o desenvolvimento de competências técnicas e analíticas. Esse resultado foi corroborado pela apresentação do estudo no evento Digital Humanities Conference 2025 (ADHO), indicando sua relevância no campo das Humanidades Digitais, bem como na Arquivologia e na CI.

A partir dessa experiência, emergiu a proposta de desenvolvimento de um modelo teórico de gamificação aplicado ao uso do Transkribus, com o objetivo de promover o engajamento de estudantes e profissionais no processo de transcrição documental. Tal proposta dialoga com as diretrizes internacionais ao buscar integrar tecnologia, aprendizagem ativa e desenvolvimento de competências, evitando uma abordagem meramente instrumental da IA (UNESCO, 2024c).

Entretanto, ao articular esses resultados com o relatório Impact of Artificial Intelligence on Culture, publicado pelo grupo CULTAI, observa-se a necessidade de uma leitura mais crítica. O documento aponta que a IA tem avançado em ritmo superior à capacidade de regulação e governança cultural, gerando um cenário de descompasso estrutural (UNESCO, 2025).

Nesse sentido, embora ferramentas como o Transkribus apresentem benefícios concretos em termos de acesso e eficiência, sua adoção também deve ser analisada com foco em questões como concentração tecnológica, dependência de plataformas e possíveis impactos na diversidade cultural.

A concessão de créditos gratuitos para o desenvolvimento desta pesquisa, por exemplo, evidencia tanto a viabilidade de acesso inicial à ferramenta quanto a existência de modelos de negócio que podem limitar seu uso em larga escala, especialmente em contextos com menor disponibilidade de recursos. Esse aspecto dialoga com as preocupações apontadas pelo relatório CULTAI acerca da redistribuição desigual de valor na economia digital e da necessidade de garantir acesso equitativo às tecnologias culturais (UNESCO, 2025).

No campo arquivístico, essa discussão se aprofunda ao considerar princípios como autenticidade, contexto e confiabilidade, conforme destacado por iniciativas do ICA e do InterPARES (ICA, 2024; Interpares, 2023). A utilização de HTR não implica apenas ganhos operacionais, mas também desafios relacionados à preservação do contexto documental e à validação dos processos automatizados, reforçando a necessidade de avaliação crítica e institucional da tecnologia.

Dessa forma, a proposta de desenvolvimento de um modelo gamificado, estruturado em cinco etapas (Introdução, Treinamento, Execução, Validação e Síntese), incorpora, em seu núcleo operacional, dois momentos complementares de transcrição: a paleográfica manual e a digital com HTR. Essa articulação apresenta-se como uma tentativa de equilibrar as dimensões técnica e interpretativa do processo, de modo que, ao combinar prática tradicional e uso de tecnologia, o modelo busca não apenas promover o uso de ferramentas baseadas em IA, mas também desenvolver competências críticas, alinhando-se às recomendações internacionais de formação e governança.

Por fim, os resultados evidenciam uma tensão central: enquanto as diretrizes globais enfatizam a necessidade de governança crítica e uso responsável da IA, sua incorporação em contextos culturais e educacionais já ocorre de forma acelerada e, muitas vezes, anterior à consolidação dessas estruturas (ONU, 2024; UNESCO, 2025). Nesse sentido, a experiência de uso do Transkribus não deve ser compreendida apenas como aplicação tecnológica bem-sucedida, mas como parte de um processo mais amplo de transformação sociotécnica, no qual estão em disputa questões de acesso, poder, representação e produção de conhecimento.

Assim, conclui-se que o uso do HTR no patrimônio documental deve ser orientado não apenas por critérios de eficiência, mas por uma abordagem crítica e contextualizada, capaz de integrar tecnologia, mediação e governança, assegurando que a IA contribua efetivamente para a ampliação do acesso e da diversidade no campo do patrimônio. Ademais, a literatura recente evidencia que os LLMs têm apresentado avanços relevantes nos processos de uso de HTR, especialmente pela capacidade de operar em cenários zero-shot, isto é, sem a necessidade de treinamento específico para um determinado tipo de manuscrito.

Nesses casos, os modelos utilizam conhecimentos previamente adquiridos durante seu treinamento em larga escala para inferir padrões linguísticos e visuais, o que permite a realização de transcrições mesmo em contextos não previamente especializados. No entanto, essa mesma característica também impõe limitações, uma vez que o desempenho tende a variar conforme o idioma, o tipo de documento e a proximidade com os dados previamente internalizados pelo modelo.

Nesse contexto, pesquisas têm recorrido a estratégias de benchmarking, isto é, à avaliação comparativa de modelos a partir de conjuntos de dados e métricas padronizadas, como o CER, com o objetivo de medir a precisão das transcrições em diferentes cenários. LLMs open-source, como LLaMA, Mistral e Falcon, apresentam desempenho variável, dependendo da arquitetura e da capacidade multimodal, não sendo originalmente projetados para tarefas de HTR. Conforme sintetizado no Quadro 9, os resultados indicam que LLMs multimodais alcançam desempenhos variáveis em benchmarks multilíngues, enquanto modelos especializados em HTR, como os utilizados no Transkribus, tendem a atingir valores mais baixos e estáveis de CER, ou seja, são mais precisos nas transcrições.

Quadro 9 – Desempenho de modelos de IA em HTR segundo diferentes estudos

Modelo	Categoria	Dataset / Contexto	CER (%)
GPT-4º.	LLM multimodal.	Benchmarks diversos (sem treinamento prévio, multilíngue).	9% – 24%
Claude 3.5.	LLM multimodal.	Benchmarks diversos (sem treinamento prévio, multilíngue).	7% – 21%
LLMs open-source.	LLM multimodal.	Benchmarks diversos.	20% (variável)
Transkribus (HTR+).	Modelo especializado.	Modelos treinados em datasets específicos.	5%
Transkribus (modelos gerais).	Modelo especializado.	Modelos pré-treinados em corpora históricos multilíngues (ex.: alemão, holandês).	7% – 11%

Fonte: Levchenko (2025); Crosilla, Klic e Colavizza (2025).

A própria documentação do Transkribus destaca que o CER corresponde à porcentagem de caracteres incorretamente reconhecidos em relação a uma transcrição de referência, sendo um indicador direto da qualidade do modelo. Nesse sentido, resultados abaixo de 10% podem ser considerados altamente eficientes para transcrição automática, enquanto valores entre 20% e 30% ainda permitem usos como busca textual, embora com limitações.

Além disso, o processo de treinamento desses modelos pode ser acompanhado por meio de curvas de aprendizagem, nas quais o CER tende a diminuir progressivamente à medida que o modelo é exposto a mais dados e

ciclos de treinamento, evidenciando a relação direta entre especialização e desempenho.

Sob a perspectiva da mediação da informação, esses aspectos evidenciam que a qualidade da transcrição não depende apenas da capacidade técnica do modelo, mas também do processo de construção e refinamento que envolve dados, parâmetros e intervenção humana. Diferentemente dos LLMs em regime zero-shot, cuja atuação se baseia em padrões generalistas, os modelos treinados em ferramentas como o Transkribus incorporam progressivamente especificidades paleográficas e contextuais, aproximando-se mais das práticas interpretativas envolvidas na leitura de documentos históricos.

Ademais, a literatura aponta que há implicações diretas entre o desempenho medido pelo CER e a experiência do usuário. Em cenários nos quais a taxa de erro ultrapassa determinados limiares, o esforço necessário para correção manual pode comprometer a eficiência do processo. Nesse sentido, transcrições automáticas com CER superior a 20% tendem a exigir um tempo de revisão que pode tornar a atividade mais exaustiva do que a própria transcrição paleográfica realizada manualmente, especialmente em contextos de maior complexidade documental (Serratore et al., 2025).

Sendo assim, essa relação evidencia que a adoção de IA, mais especificamente do HTR, não deve ser avaliada apenas em termos de automação, mas também de usabilidade e impacto no trabalho intelectual do usuário. Portanto, a utilização de IA generativa no HTR pode ser compreendida como parte de um processo ampliado de mediação, no qual a interpretação do documento não é realizada exclusivamente pela máquina, mas emerge da interação entre sistemas automatizados e sujeitos humanos. Dessa forma, o papel do pesquisador, arquivista ou usuário não é eliminado, mas ressignificado, assumindo uma função crítica na validação, contextualização e atribuição de sentido às transcrições produzidas.

Nesse cenário, reforça-se a compreensão de que os LLMs não substituem os sistemas especializados de HTR, mas se inserem como tecnologias complementares em um ecossistema sociotécnico mais amplo (Levchenko, 2025; Crosilla; Klic; Colavizza, 2025).

5 CONCLUSÃO

Diante do desenvolvimento deste trabalho, tornou-se evidente que a incorporação de tecnologias de HTR no contexto do patrimônio documental não pode ser compreendida apenas como um avanço técnico, mas como uma transformação nos modos de produção, mediação e acesso à informação. Ainda que esses sistemas alcancem elevados índices de eficiência, sua operação permanece intrinsecamente vinculada ao conhecimento humano que os sustenta, especialmente no que se refere à Paleografia. O treinamento dos modelos, baseado em dados previamente estruturados por especialistas, evidencia que a automação não substitui o saber paleográfico, mas o incorpora e o ressignifica em novas dinâmicas informacionais.

Nesse sentido, o uso de tecnologias baseadas em IA revela simultaneamente potencialidades e limitações. Se, por um lado, amplia o acesso e otimiza processos, por outro, ainda enfrenta dificuldades na interpretação de manuscritos complexos. Isso exige do usuário não apenas domínio técnico, mas também letramento histórico e arquivístico. A leitura documental, sob a perspectiva da CI, não se reduz ao reconhecimento de padrões gráficos; ela envolve processos de mediação cognitiva e contextual que permanecem fundamentais. Assim, a mediação da informação não é eliminada, mas reconfigurada na interação entre sistemas automatizados e sujeitos interpretativos.

A presente pesquisa partiu do questionamento acerca de que forma o HTR reconfigura o tratamento e o acesso ao patrimônio documental em contextos de Paleografia Digital e dados em larga escala, no âmbito de estratégias de gamificação voltadas à mitigação da fadiga cognitiva e à capacitação de usuários na validação de transcrições no Transkribus. A partir do percurso desenvolvido, compreende-se que o HTR amplia significativamente as possibilidades de acesso e processamento de documentos manuscritos, ao automatizar etapas tradicionalmente dependentes de leitura especializada.

Entretanto, essa contribuição não ocorre de forma neutra. O uso de HTR reconfigura os processos de transcrição e acesso à informação ao introduzir uma camada algorítmica de mediação, na qual a informação passa a ser parcialmente produzida e filtrada por modelos treinados sobre dados específicos. Embora

ferramentas como o Transkribus possibilitem ganhos em escala e velocidade, também introduzem limitações relacionadas à acurácia, à opacidade dos modelos e à dependência de conhecimento especializado.

Dessa forma, conclui-se que o HTR não elimina a necessidade de mediação humana, mas a reconfigura e a complexifica. O profissional da informação permanece como agente fundamental na interpretação, validação e contextualização dos dados produzidos pelos sistemas, atuando em uma lógica de Human-in-the-Loop.

O objetivo geral desta pesquisa consistiu em analisar as potencialidades, limitações e implicações da tecnologia de HTR no acesso e no tratamento do patrimônio documental histórico e, a partir do diagnóstico das demandas de mediação informacional, desenvolver e propor um modelo gamificado de capacitação focado na otimização do uso da ferramenta Transkribus por usuários não especialistas.

Nesse contexto, a gamificação apresenta-se como uma estratégia promissora de capacitação, ao estruturar o processo de aprendizagem por meio de desafios progressivos, sistemas de feedback e mecanismos de engajamento. Tal abordagem favorece não apenas a familiarização operacional com a ferramenta, mas também a construção de uma compreensão mais crítica acerca das potencialidades e limitações do HTR. Ainda assim, a gamificação não deve ser compreendida como solução plena, mas como uma resposta situada às dificuldades identificadas, atuando como uma camada de mediação que busca mitigar barreiras de uso sem eliminar as limitações estruturais da própria tecnologia.

No que se refere aos resultados alcançados, a pesquisa consolidou dois eixos complementares que dialogam diretamente com os objetivos propostos. Em um primeiro momento, na fase diagnóstica e exploratória, foi possível constatar de forma sistemática as barreiras de usabilidade tecnológica enfrentadas por usuários não especialistas no uso de ferramentas baseadas em IA, como o Transkribus. Esses achados evidenciam que dificuldades operacionais, limitações na compreensão das funcionalidades da plataforma e a necessidade de conhecimentos prévios, ainda que básicos, constituem entraves concretos à apropriação dessas tecnologias.

Ademais, aspectos como a fadiga cognitiva, a repetitividade das tarefas e a insegurança diante dos resultados gerados pelo HTR reforçam que a incorporação da IA em contextos de patrimônio documental não pode ser compreendida apenas

sob a lógica da eficiência técnica, exigindo atenção às dimensões cognitivas, formativas e de mediação da informação.

Em um segundo momento, na fase propositiva, a pesquisa avançou na síntese criativa de um framework de gamificação, concebido como resposta direta às lacunas identificadas empiricamente. Esse modelo estrutura-se como uma proposta de mediação sociotécnica orientada à capacitação, ao engajamento e ao desenvolvimento de competências no uso do Transkribus, articulando elementos como progressão por níveis, missões, feedback contínuo e sistemas de recompensa. Mais do que um recurso instrumental, o framework configura-se como um legado aplicável, com potencial de adaptação em diferentes contextos institucionais, especialmente em políticas de letramento informacional voltadas a instituições de memória.

Nesse sentido, a proposta contribui não apenas para a melhoria da experiência do usuário, mas também para a consolidação de estratégias formativas que reconheçam a centralidade da mediação humana no uso de tecnologias baseadas em IA. Ademais, se faz relevante ressaltar algumas limitações da pesquisa, especialmente diante do caráter dinâmico do objeto investigado, ao situar esta investigação em um campo marcado pela rápida evolução tecnológica, reconhece-se que seus resultados se inserem em um contexto de constante atualização. A dinâmica de desenvolvimento de ferramentas como o Transkribus impõe limites à consolidação de análises de longo prazo. Ao mesmo tempo, evidencia a necessidade de abordagens críticas no uso dessas tecnologias.

Além disso, outro limite refere-se às desigualdades linguísticas, uma vez que o português ainda apresenta menor disponibilidade de modelos e recursos quando comparado ao inglês e outros idiomas. Soma-se a isso o caráter provisório das análises, considerando a constante atualização das ferramentas, o que limita avaliações de longo prazo. Assim como a baixa adesão de respostas aos questionários. Do ponto de vista prático, a não aplicação do modelo gamificado configura uma limitação relevante, mantendo a proposta em nível conceitual.

Diante dessas limitações, destacam-se possibilidades de pesquisas futuras, entre elas, a aplicação empírica do modelo proposto, permitindo avaliar engajamento, desempenho e compreensão crítica por meio de métricas quantitativas e qualitativas. Estudos comparativos entre diferentes perfis de usuários também podem aprofundar

a compreensão sobre o papel do conhecimento prévio na interação com sistemas de HTR.

Além disso, investigações voltadas ao desenvolvimento e avaliação de modelos de HTR para documentos em língua portuguesa podem contribuir para ampliar o acesso ao patrimônio documental. Por fim, sugere-se o aprofundamento das discussões sobre mediação algorítmica, especialmente no que se refere à constituição da informação em ambientes mediados por IA.

O trabalho conclui que a adoção de ferramentas como o Transkribus se justifica não apenas pela sua eficácia de CER, mas por sua integridade diplomática: são arquiteturas algorítmicas projetadas para o reconhecimento literal e não para a geração sintética, respeitando a materialidade informacional do registro de arquivo. A ampliação desta discussão reforçará o posicionamento crítico dos profissionais da informação diante da voracidade atual do mercado pelas Big Techs e pelas IAs de linguagem natural.

Portanto, esta dissertação, busca cumprir a missão de repensar a Arquivologia e a CI no século XXI, posicionando o arquivo não como um depósito inerte, mas como um ambiente sociotécnico vivo e computacionalmente ativado. Nesse sentido, o modelo gamificado proposto não se encerra como resultado desta pesquisa, mas se projeta como instrumento de continuidade investigativa. Ao articular tecnologia, mediação da informação e formação de usuários, este trabalho contribui para o fortalecimento de uma abordagem crítica no uso de tecnologias no contexto documental, reafirmando o papel da CI na mediação desses processos em constante transformação.

REFERÊNCIAS

ALKENDI, W. et al. Advancements and Challenges in Handwritten Text Recognition: A Comprehensive Survey. J. Imaging, Basel, v. 10, n. 1, p. 18, 2024. DOI: 10.3390/jimaging10010018. Disponível em:https://pubmed.ncbi.nlm.nih.gov/38249003/. Acesso em: 24 de jul. 2024.

ALVES, Lynn Rosalina Gama; MINHO, Maria Rita; DINIZ, Marcelo Viana. Gamificação: diálogos com a educação. In: FADEL, Luciane Maria et al. (org.). *Gamificação na educação*. São Paulo: Pimenta Cultural, 2014. p. 73–95. Disponível em: https://www.researchgate.net/publication/304011416. Acesso em: 27 mar. 2026.

AMERSHI, Saleema; CAKMAK, Maya; KNOX, W. Bradley; KULESZA, Todd. Power to the people: the role of humans in interactive machine learning. *AI Magazine*, [S.l.], v. 35, n. 4, p. 105–120, 2014. Disponível em: https://ojs.aaai.org/index.php/aimagazine/article/view/2513. Acesso em: 04 mar. 2026.

AQUINO, I. Digital Humanities como uma comunidade de práticas: entrevista com o professor Daniel Alves (IHC / NOVA FCSH). Revista Aedos, [S. l.], v. 12, n. 26, p. 740–761, 2020. Disponível em: https://seer.ufrgs.br/index.php/aedos/article/view/105132. Acesso em: 2 mar. 2023.

ARAÚJO, Carlos Alberto Ávila. Mediação da informação: aportes teóricos e conceituais. Informação & Informação, Londrina, v. 23, n. 2, p. 17–40, maio/ago. 2018. Disponível em: https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/29621. Acesso em: 20 set. 2025.

ARIAS HERNÁNDEZ, R., ROCKEMBACH, M. Building trustworthy AI solutions: integrating artificial intelligence literacy into records management and archival systems. AI & Soc (2025). Disponível em: https://www.researchgate.net/publication/388723352_Building_trustworthy_AI_solutions_integrating_artificial_intelligence_literacy_into_records_management_and_archival_systems. Acesso em: 24 de jul. 2024.

BAUDRY, H. Transcrever os processos da Inquisição Portuguesa (1536–1821). In: FERREIRA, A. P.; GARCIA, L. D.; DORES, M.; SEQUEIRA, O. (org.). *Between lines and notarial marks*. Évora: Publicações do CIDEHUS, 2024. Disponível em: https://doi.org/10.4000/11qiq. Acesso em: 27 ago. 2024.

BAWDEN, David. Information and digital literacies: a review of concepts. Journal of Documentation, Bingley, v. 57, n. 2, p. 218–259, 2001. Disponível em: https://www.emerald.com/insight/content/doi/10.1108/EUM0000000007083/full/html. Acesso em: 27 mar. 2026.

BELLOTTO, Heloísa Liberalli. Arquivos permanentes: tratamento documental. 3. ed. Rio de Janeiro: FGV, 2002. Acesso em: 29 set. 2025.

BENKLER, Yochai. *The wealth of networks: how social production transforms markets and freedom*. New Haven: Yale University Press, 2006. Disponível em: https://www.benkler.org/Benkler_Wealth_Of_Networks.pdf. Acesso em: 27 mar. 2026.

BERRY, David M. *Understanding digital humanities*. London: Palgrave Macmillan, 2012. Disponível em: https://link.springer.com/book/10.1057/9780230371934. Acesso em: 07 nov. 2025.

BERWANGER, Ana Regina; LEAL, João Eurípedes Franklin. *Noções de paleografia e diplomática*. 3. ed. Santa Maria: UFSM, 2008. Acesso em: 20 mar. 2025.

BIBLIOTHÈQUE NATIONALE DE FRANCE. Intelligence artificielle et collections de la BnF : l'exemple de la HTR (Handwritten Text Recognition). Paris: BnF, [s.d.]. Disponível em: https://www.bnf.fr/fr/intelligence-artificielle-et-collections-de-la-bnf-lexemple-de-lhtr-handwritten-text-recognition. Acesso em: 24 ago. 2025.

BORKO, H. Ciência da Informação: O que é isto? American Documentation Vol.19, N.1, P.3-5, Jan. 1968. Acesso em: 01 de mar. 2023.

BRAPCI. Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação. [S.l.: s.n.], [s.d.]. Disponível em: https://brapci.inf.br/home. Acesso em: 27 jan. 2026.

BRASIL. Conselho Nacional de Saúde. Resolução nº 510, de 7 de abril de 2016. Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais. Diário Oficial da União: seção 1, Brasília, DF, 24 maio 2016. Disponível em: https://conselho.saude.gov.br/resolucoes/2016/Reso510.pdf. Acesso em: 27 mar. 2026.

BRAZIL, A. L.; ALBAGLI, S. Os usos da gamificação na mobilização cognitiva da ciência cidadã online. Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 25, p. 1-21, 2020. Disponível em:https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2020.e66373. Acesso em: 20 set. 2025.

BURDICK, ANNE et al. Digital humanities. Cambridge: MIT Press, 2012. Disponível em: https://mitpress.mit.edu/9780262018470/digital-humanities/. Acesso em: 17 mar. 2026.

BUSCIANTELLA-RICCI, Daniele; SCATAGLINI, Sofia. Research through co-design. *Design Science*, Cambridge, v. 10, e3, 2024. Disponível em: https://doi.org/10.1017/dsj.2023.35. Acesso em: 27 mar. 2026.

CAERS, Bram. Teaching handwritten text recognition: can new technologies save old skills? *Quaerendo*, Leiden, v. 54, n. 2-3, p. 198–209, 2024. Disponível em: https://brill.com/view/journals/qua/54/2-3/article-p198_6.xml. Acesso em: 02 dez. 2024.

CANAL FUNDAÇÃO BIBLIOTECA NACIONAL. Ciclo de Palestras sobre IA e Bibliotecas [vídeo]. YouTube, 13 out. 2025. Disponível em: https://www.youtube.com/watch?v=VxivWqMa6ys&t=2041s. Acesso em: 05 nov. 2025.

CAPURRO, R. Epistemologia e Ciência da Informação. Perspectivas em Ciência da Informação, v. 8, n. 1, 2003. Acesso em: 04 jun. 2025.

CATIVELLI, A. S.; MONSANI, D.; JULIANI, J. P. Gamificação em bibliotecas: despertando a motivação nos usuários. Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, v. 21, n. 45, p. 70-81, jan./abr., 2016. Disponível em:https://www.redalyc.org/journal/147/14743480006/. Acesso em: 04 set. 2025.

CHEN, L., CHEN, P., LIN, Z. (2020) Artificial Intelligence in Education: A Review. IEEE Access, 8, 75264-75278. Disponível em:https://ieeexplore.ieee.org/document/9069875. Acesso em: 04 set. 2025.

CIULA, Arianna; EIDE, Øyvind. Modelling in digital humanities: signs in context. Digital Scholarship in the Humanities, Oxford, v. 32, n. suppl_1, p. i33–i46, 2017. Disponível em: https://academic.oup.com/dsh/article/32/suppl_1/i33/3864960. Acesso em: 15 mar. 2026.

COLAVIZZA, Giovanni et al. Archives and AI: An Overview of Current Debates and Future Perspectives. Journal on Computing and Cultural Heritage, v. 15, n. 1, 2021. Disponível em:https://dl.acm.org/doi/full/10.1145/3479010. Acesso em: 04 dez. 2024.

COOK, Terry. What is past is prologue: a history of archival ideas since 1898, and the future paradigm shift. *Archivaria*, Ottawa, n. 43, p. 17–63, Spring 1997.

CROSILLA, Giorgia; KLIC, Lukas; COLAVIZZA, Giovanni. Benchmarking large language models for handwritten text recognition. Journal of Documentation, v. 81, n. 7, p. 334-354, 2025. Disponível em: https://www.emerald.com/jd/article/81/7/334/1275080/Benchmarking-large-language-models-for-handwritten. Acesso em: 16 mar. 2026.

DETERDING, Sebastian; DIXON, Dan; KHALED, Rilla; NACKE, Lennart. From game design elements to gamefulness: defining “gamification”. In: INTERNATIONAL ACADEMIC MINDTREK CONFERENCE, 15., 2011, Tampere. Proceedings [...]. Tampere: ACM, 2011. p. 9–15. Disponível em: https://doi.org/10.1145/2181037.2181040. Acesso em: 25 mar. 2026.

DIGITAL SCIENCE. (2018). Dimensions [Software]. Disponível em: https://app. dimensions.ai. Acesso em: 07 nov. 2024.

DRUCKER, Johanna. Humanities approaches to graphical display. *Digital Humanities Quarterly*, [S.l.], v. 5, n. 1, 2011. Disponível em: https://dhq.digitalhumanities.org/vol/5/1/000091/000091.html. Acesso em: 07 nov. 2025.

DURANTI, Luciana. Diplomatics: new uses for an old science. Lanham: Scarecrow Press, 1998. Acesso em: 27 jan. 2026.

DURANTI, Luciana. The InterPARES 2 Project (2002–2007): an overview. *Archivaria*, Ottawa, n. 64, p. 113–121, 2007. Disponível em: https://archivaria.ca/index.php/archivaria/article/view/13155. Acesso em: 03 abr. 2026.

DURANTI, Luciana. From digital diplomatics to digital records forensics. *Archivaria*, Ottawa, n. 68, p. 39–66, 2009. Disponível em: https://archivaria.ca/index.php/archivaria/article/view/13229. Acesso em: 03 abr. 2026.

EUROPEANA FOUNDATION. [s.d.] Europeana Transcribe. Disponível em:https://europeana.transcribathon.eu/. Acesso em: 2 set. 2025.

EUROPEANA FOUNDATION. Transkribathon. [s.d.]. Disponível em: https://transcribathon.eu/. Acesso em: 20 mar. 2026.

EUROPEANA. (2024). Story of Antoni Wajda. Transcribathon. Disponível em: https://europeana.transcribathon.eu/documents/story/?story=149448. Acesso em: 2 dez. 2024.

EVELEIGH, A.; JENNETT, C.; BLANDFORD, A.; BROHAN, P.; COX, A. L. Designing for dabblers and deterring drop-outs in citizen science. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’14), New York, NY, USA, 2014. p. 2985-2994. DOI: 10.1145/2556288.2557262. Disponível em:https://dl.acm.org/doi/10.1145/2556288.2557262. Acesso em: 04 set. 2025.

FLORIDI, Luciano. The Logic of Information: A Theory of Philosophy as Conceptual Design. Oxford: Oxford University Press, 2019. Disponível em: https://academic.oup.com/book/27824. Acesso em: 07 nov. 2025.

FONSECA, Maria Odila Kahl. Arquivologia e Ciência da Informação. Rio de Janeiro: FGV, 2005. Acesso em: 04 jun. 2025.

FREILING, Janna; RETHER, Tanja. Investigating the accuracy of AI-supported handwritten text recognition. In: 33rd BOBCATSSS Symposium, Istanbul, 2025. Disponível em:https://bobcatsss2025.org.tr/accepted-submissions/. Acesso em: 20 fev. 2025.

GOMES, Henriette Ferreira. Mediação da informação e o papel social do bibliotecário. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 19, n. especial, p. 64–74, 2014. Disponível em: https://periodicos.ufmg.br/index.php/pci/article/view/22362. Acesso em: 25 mar. 2026.

GONZÁLEZ DE GÓMEZ, Maria Nélida. A ciência da informação no campo científico. Ciência da Informação, Brasília, v. 31, n. 1, p. 60-70, jan./abr. 2002. Acesso em: 20 jun. 2025.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep learning*. Cambridge: MIT Press, 2016. Disponível em: https://www.deeplearningbook.org. Acesso em: 27 mar. 2026.

HJØRLAND, Birger. Domain analysis in information science: eleven approaches – traditional as well as innovative. *Journal of Documentation*, Bingley, v. 58, n. 4, p. 422–462, 2002. Acesso em: 24 de jul. 2025.

ICA. International Archives Week 2024: CyberArchives. Paris: International Council on Archives, 2024. Disponível em:https://www.ica.org/. Acesso em: 20 mar. 2026.

INTERPARES. InterPARES Trust AI – Research and findings. 2023. Disponível em: https://interparestrustai.org/. Acesso em: 20 mar. 2026.

JARDIM, José Maria. A invenção da memória nos arquivos públicos. Ciência da Informação, Brasília, v. 24, n. 2, p. 1–9, 1995. Acesso em: 04 mar. 2026.

KAPP, Karl M. The gamification of learning and instruction: game-based methods and strategies for training and education. San Francisco: Pfeiffer, 2012. Disponível em: https://www.wiley.com/en-us/The+Gamification+of+Learning+and+Instruction%3A+Game-based+Methods+and+Strategies+for+Training+and+Education-p-9781118096345. Acesso em: 27 mar. 2026.

KITCHIN, Rob. Thinking critically about and researching algorithms. *Information, Communication & Society*, Abingdon, v. 20, n. 1, p. 14–29, 2017. Disponível em: https://doi.org/10.1080/1369118X.2016.1154087. Acesso em: 20 set. 2026

KOIVISTO, J.; HAMARI, J. The rise of motivational information systems: A review of gamification research. International Journal of Information Management, v. 45, p. 191–210, 2019. Disponível em:https://www.sciencedirect.com/science/article/pii/S0268401217305169. Acesso em: 20 set. 2025.

LEE, Joey J.; HAMMER, Jessica. Gamification in education: what, how, why bother? *Academic Exchange Quarterly*, [S.l.], v. 15, n. 2, p. 146–151, 2011. Disponível em: https://www.researchgate.net/publication/258697764_Gamification_in_Education_What_How_Why_Bother. Acesso em: 27 mar. 2026.

LEVCHENKO, Maria A. Evaluating LLMs for historical document OCR: a methodological framework for digital humanities. In: PROCEEDINGS OF THE FIRST WORKSHOP ON NATURAL LANGUAGE PROCESSING AND LANGUAGE MODELS FOR DIGITAL HUMANITIES. Varna, Bulgaria: INCOMA Ltd., 2025. p. 75-85. Disponível em: https://aclanthology.org/2025.lm4dh-1.7.pdf. Acesso em: 16 mar. 2026.

LLOYD, Annemaree. Information literacy landscapes: information literacy in education, workplace and everyday contexts. Oxford: Chandos Publishing, 2010. Disponível em: https://www.sciencedirect.com/book/9781843345169/information-literacy-landscapes. Acesso em: 27 mar. 2026.

LÓPEZ-NORES, Martín et al. Cuneiforce: involving the crowd in the annotation of unread Mesopotamian cuneiform tablets through a gamified design. In: CONFERENCE ON E-BUSINESS, E-SERVICES AND E-SOCIETY, 2019. *Proceedings…* Cham: Springer International Publishing, 2019. p. 385–396. Disponível em: https://doi.org/10.1007/978-3-030-17740-9_30. Acesso em: 29 set. 2025.

LOSE, A. D. Transcritores automáticos de texto e transcrição colaborativa agilizando o trabalho de leitura paleográfica e identificação de informações textuais para acervos históricos: o exemplo do Mooc Paleografia na Rede Wiki. Cadernos BAD, [S. l.], n. 1-2, 2025. DOI: 10.48798/cadernosbad.3090. Disponível em: https://publicacoes.bad.pt/revistas/index.php/cadernos/article/view/3090/2767. Acesso em: 20 set. 2025.

LOSE, ALÍCIA DUHÁ; SANTOS, João Guilherme Veloso Andrade dos; JESUS, Leonardo Coelho Marques de; MAGALHÃES, Lívia Borges Souza; FROUXE, Lucia. Transkribus: uma ferramenta de paleografia digital mediando pesquisas em fontes inquisitoriais. Revista LaborHistórico, v. 10, n. 1, e63285, 2024. DOI: https://doi.org/10.24206/lh.v10i1.63285. Acesso em: 21 out. 2025.

MAGALHÃES, LÍVIA BORGES SOUZA. Da 1.0 até a 3.0: a jornada da Paleografia no mundo digital. LaborHistórico, Rio de Janeiro, v. 7, n. 3, p. 279-295, set./dez. 2021. Disponível em:https://revistas.ufrj.br/index.php/lh/article/view/42368. Acesso em: 03 de jul. 2025.

MANESS, Jack; PHAM, Kim. Just Because We Can Doesn't Mean We Should: On Knowing and Protecting Data Produced by the Jewish Consumptives' Relief Society. Humanist Studies & the Digital Age, v. 7, n. 1, 2022. Disponível em:https://journals.oregondigital.org/hsda/article/view/5761. Acesso em: 05 dez. 2025.

MARCIANO, Richard. Afterword: towards a new discipline of computational archival science (CAS). In: JAILLANT, Lise (org.). *Archives, access and artificial intelligence: working with born-digital and digitized archival collections*. Bielefeld: Bielefeld University Press, 2022. p. 205–218. Disponível em: https://doi.org/10.1515/9783839455845-009. Acesso em: 03 abr. 2026.

MELO, JONAS FERRIGOLO. *Consentimento informado e proteção de dados pessoais: proposta de um modelo de gerenciador de consentimento para plataformas digitais*. Porto: Universidade do Porto, 2025. Tese (Doutorado em Informação e Comunicação em Plataformas Digitais) – Faculdade de Letras da Universidade do Porto, 2025. Disponível em: https://repositorio-aberto.up.pt/handle/10216/171014. Acesso em: 27 mar. 2026.

MONTEIRO-KREBS, Luciana. *Recommender algorithms in academic social networks*. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2022. Tese (Doutorado em Comunicação e Informação) – Universidade Federal do Rio Grande do Sul, 2022. Disponível em: https://lume.ufrgs.br/handle/10183/238299. Acesso em: 27 mar. 2026.

MONTE-MÓR, Roberto Luís de Melo. *Brasília: da utopia à cidade*. Belo Horizonte: Editora UFMG, 2003. Acesso em: 02 nov. 2025.

MUEHLBERGER, Günter et al. Transforming scholarship in the archives through handwritten text recognition: the Transkribus platform. Journal of Documentation, Bingley, v. 75, n. 5, p. 954–976, 2019. Disponível em: https://www.emerald.com/insight/content/doi/10.1108/JD-07-2018-0114/full/html. Acesso em: 27 mar. 2026.

NASCIMENTO, A. C.; SANTOS, J. S. J.; MATA, M. L. da. O uso da gamificação para o desenvolvimento da competência em informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 2024, Aracaju. Anais [...]. Aracaju: 2024. Disponível em: https://brapci.inf.br/v/307247. Acesso em: 02 set. 2025.

NOCKELS, J., GOODING, P., & TERRAS, M. (2024). Are Digital Humanities platforms facilitating sufficient diversity in research? A study of the Transkribus Scholarship Programme. Digital Scholarship in the Humanities, fqae018. Disponível em:https://academic.oup.com/dsh/article/40/Supplement_1/i46/7646909. Acesso em: 02 dez. 2024.

NOCKELS, J., GOODING, P., & TERRAS, M. (2024). The implications of HTR for accessing the past. Journal of Documentation, v. 80, n. 7, p. 148-167, 2024. DOI: 10.1108/JD-09-2023-0183. Disponível em: https://www.emerald.com/jd/article-pdf/80/7/148/9597977/jd-09-2023-0183.pdf. Acesso em: 02 dez. 2024.

OCDE. OECD Digital Education Outlook 2023. Paris: OECD Publishing, 2023. Disponível em: https://www.oecd.org/education/oecd-digital-education-outlook-2023/. Acesso em: 20 mar. 2026.

OCDE. Shaping the future of digital education: AI and education. Paris: OECD Publishing, 2024. Disponível em: https://www.oecd.org/education/. Acesso em: 20 mar. 2026.

ONU. Governing AI for Humanity: Final Report. New York: United Nations, 2024. Disponível em: https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf. Acesso em: 20 mar. 2026.

PIMENTA, R. M. Por que Humanidades Digitais na Ciência da Informação? Perspectivas pregressas e futuras de uma prática transdisciplinar comum. Informação & Sociedade: Estudos, [S. l.], v. 30, n. 2, 2020. DOI: 10.22478/ufpb.1809-4783.2020v30n2.52122. Disponível em: https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/52122. Acesso em: 2 mar. 2023. Acesso em: 02 fev. 2025.

PRAHALAD, C. K.; RAMASWAMY, Venkat. *The future of competition: co-creating unique value with customers*. Boston: Harvard Business School Press, 2004. Disponível em: https://hbr.org/2004/01/co-creating-unique-value-with-customers. Acesso em: 27 mar. 2026..

RANGEL, Taynã Ladislau. *Mediação algorítmica e circulação da informação em ambientes digitais*. João Pessoa: Universidade Federal da Paraíba, 2024. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal da Paraíba, 2024. Disponível em: https://repositorio.ufpb.br/jspui/handle/123456789/32121. Acesso em: 02 fev. 2026.

RANGEL, Taynã Ladislau. *Mediação da informação em ambientes digitais: desafios e perspectivas*. João Pessoa: Universidade Federal da Paraíba, 2022. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) – Universidade Federal da Paraíba, 2022. Disponível em: https://repositorio.ufpb.br/jspui/. Acesso em: 02 fev. 2026.

RAVENTÓS PAJARES, P., HERNÁNDEZ TORNERO, C., & SIMON MARTIN, M. (2023). AI and archive. Handwritten Text Recognition Applied to Patrimonial Holdings: An Example of 10 diaries written by Spanish Republican Teachers in 1932. Archives and Records Management, University of Lleida. Disponível em: https://www.researchgate.net/publication/367456023_AI_and_archive_Handwritten_Text_Recognition_Applied_to_Patrimonial_Holdings_An_Example_of_10_diaries_written_by_Spanish_Republican_Teachers_in_1932. Acesso em: 02 dez. 2024.

READ-COOP (2024) How to improve the CER of your model. Disponível em:https://blog.transkribus.org/en/how-to-improve-the-cer-of-your-model. Acesso em: 28 nov. 2024.

RIDGE, Mia (Org.). CROWDSOURCING OUR CULTURAL HERITAGE. FARNHAM: ASHGATE PUBLISHING, 2014. Acesso em: 20 jun. 2026.

ROCKEMBACH, M; GEERTS, D. Eticista digital: uma função emergente no campo da informação. Boletim do Arquivo da Universidade de Coimbra, v. 37, n. 1, p. 75-93, 2024. https://doi.org/10.14195/2182-7974_37_1_3. Acesso em: 1 jul. 2026.

ROCKEMBACH, M. Ciência da informação e inteligência artificial: um caminho para arquivos e bibliotecas inteligentes. In: Congresso ISKO Espanha-Portugal (5.: 2021: Lisboa): Organização do conhecimento no horizonte 2030: desenvolvimento sustentável e saúde. Atas..[recurso eletrônico]. Lisboa: Universidade de Lisboa, 2021. Disponível em: https://dialnet.unirioja.es/servlet/articulo?codigo=8411211. Acesso em 22 fev 2023.

ROCKEMBACH, M. Estudos de usuários de arquivo e os desafios da Lei Geral de Proteção de Dados. Acervo, Rio de Janeiro, v. 33, n. 3, p. 102-115, 19 ago. 2020. Disponível em: https://revista.an.gov.br/index.php/revistaacervo/article/view/1554/1541. Acesso em: 05 nov. 2025.

ROCKEMBACH, Moisés; EVEN, Priscilla Van; MELO, Jonas Ferrigolo; MONTEIRO-KREBS, Luciana. Metodologias de co-design em Ciência da Informação e Comunicação. In: SILVA, Fabiano Couto Corrêa da; BARROS, Thiago Henrique Bragato (org.). *Caminhos metodológicos em Ciência da Informação: fundamentos e*

aplicações. São Paulo: Pimenta Cultural, 2025. Disponível em: https://www.researchgate.net/publication/396592764. Acesso em: 27 mar. 2026.

RODRIGUES, S; GOMES, P. Arquivologia e educação: múltiplas abordagens. P2P E INOVAÇÃO, [S. l.], v. 7, n. 2, p. 63–87, 2021. DOI: 10.21721/p2p.2021v7n2.p63-87. Disponível em:https://revista.ibict.br/p2p/article/view/5628. Acesso em: 28 fev. 2023.

ROLING, M. Does Handwriting Text Recognition Work for Damaged Archives? Independent Research, [S. l.], p. 1-11, mar. 2020. Disponível em: https://www.researchgate.net/publication/340117708_Does_Handwriting_Text_Recognition_Work_for_Damaged_Archives. Acesso em: 28 mai. 2025.

RONDINELLI, Rosely Curi. Gerenciamento arquivístico de documentos eletrônicos. Rio de Janeiro: FGV, 2005. Acesso em: 04 mar. 2026.

SCHREIBMAN, Susan; SIEMENS, Ray; UNSWORTH, John (org.). *A new companion to digital humanities*. Malden: Wiley-Blackwell, 2016. Disponível em: https://onlinelibrary.wiley.com. Acesso em: 27 mar. 2026.

SERRATORE, G. et al. *An HTR model for automated transcriptions of the Girifalco asylum archive*. In: CONFERENCE ON LANGUAGE, DATA AND KNOWLEDGE (LDK), 5., 2025, Naples. *Proceedings…* Naples: Unior Press, 2025. Disponível em: https://aclanthology.org/2025.ldk-1.30.pdf. Acesso em: 27 mar. 2026.

SILVA, A. M. D. da. A Inteligência Artificial no acesso à informação em documentos manuscritos. Cadernos BAD, [S. l.], n. 1-2, 2025. DOI: 10.48798/cadernosbad.3095. Disponível em: https://publicacoes.bad.pt/revistas/index.php/cadernos/article/view/3095/2766. Acesso em: 28 mai. 2025.

SILVA, G. H. C. da. et al. LHTR.br: em Busca de um Conjunto Anotado de Textos Manuscritos em Português. In: DATASET SHOWCASE WORKSHOP (DSW), 5., 2023. Proceedings… [S. l.]: SBC, 2023. Disponível em: https://sol.sbc.org.br/index.php/dsw/article/view/25501. Acesso em: 28 mai. 2025.

SILVA, Robson Santos da; SALES, José Aires de Castro Filho; CASTRO, Felipe César Viana de. Gamificação como estratégia de aprendizagem ativa no ensino superior. *Revista Tecnologias na Educação*, [S.l.], v. 11, n. 30, p. 1–13, 2019. Disponível em: https://tecedu.pro.br/revistas/index.php/tecedu/article/view/207. Acesso em: 27 mar. 2026.

SOARES NETTO, A. F.; DUQUE, C. G. Inteligência artificial aplicada em aprendizado orientado a problema, para recuperação da informação em ambiente gamificado, na trilha de conhecimento de licitações públicas de tecnologias, sob ótica da gestão de riscos. Informação & Informação, Londrina, v. 30, n. 1, p. 403-430, jan./mar. 2025. Disponível em:https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/51122. Acesso em: 06 set. 2025.

SOUSA JUNIOR, Manoel Bezerra de. *Perspectivas da mediação algorítmica na Ciência da Informação*. Fortaleza: Universidade Federal do Ceará, 2024. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Federal do Ceará, 2024. Disponível em: https://repositorio.ufc.br/handle/riufc/77296. Acesso em: 04 fev. 2026.

SOUSA, N; ROCKEMBACH, M. Chatbots na Ciência da Informação: aplicações em arquivos e bibliotecas. Páginas a&b: arquivos e bibliotecas, Porto, n. 23, p. 314–330, 2025. Disponível em: https://ojs.letras.up.pt/index.php/paginasaeb/article/view/15131. Acesso em: 1 jul. 2026.

UNESCO. AI competency framework for students. Paris: UNESCO, 2024b. Disponível em: https://www.unesco.org/en/articles/ai-competency-framework-students. Acesso em: 20 mar. 2026.

UNESCO. AI competency framework for teachers. Paris: UNESCO, 2024a. Disponível em: https://www.unesco.org/en/articles/ai-competency-framework-teachers. Acesso em: 20 mar. 2026.

UNESCO. Framework for Culture and Arts Education. Paris: UNESCO, 2024c. Disponível em: https://www.unesco.org/sites/default/files/medias/fichiers/2024/02/WCCAE_UNESCO%20Framework_EN_0.pdf. Acesso em: 20 mar. 2026.

UNESCO. Guidance for generative AI in education and research. Paris: UNESCO, 2023. Disponível em: https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research. Acesso em: 20 mar. 2026.

UNESCO. Impact of Artificial Intelligence on Culture. Paris: UNESCO, 2025. Disponível em: https://unesdoc.unesco.org/. Acesso em: 20 mar. 2026.

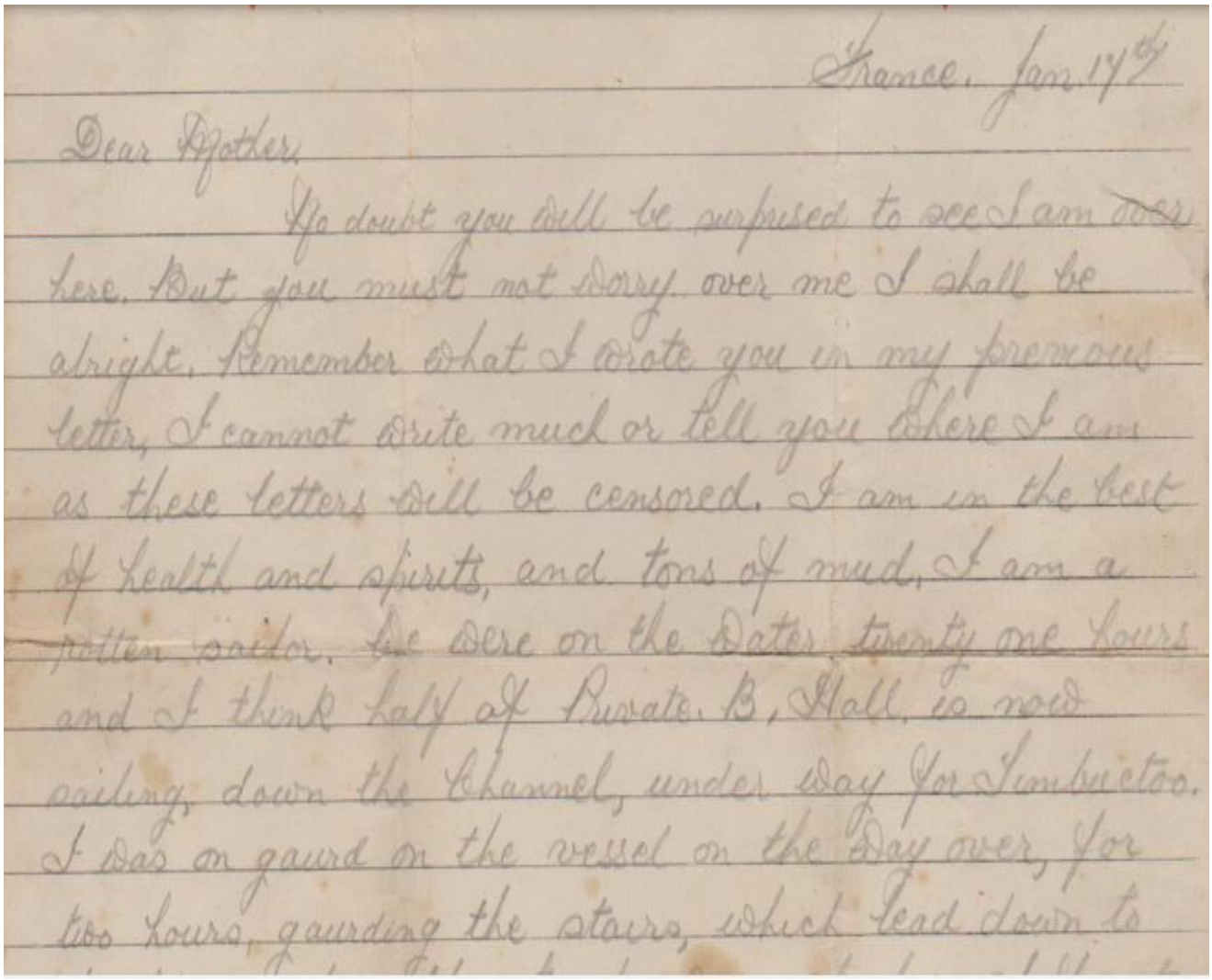
WERBACH, K.; HUNTER, D. For the Win: How Game Thinking can Revolutionize your Business. Wharton Digital Press, 2012. Acesso em: 2 set. 2025. WILSON, T. D. Human Information Behavior. Informing Science, v. 3, n. 2, 2000. Disponível em: https://www.researchgate.net/publication/273946893_For_the_Win_How_Game_Thinking_can_Revolutionize_your_Business. Acesso em: 2 mar. 2025

APÊNDICE A – MANUAL TRANSKRIBUS

Para realizar o exercício se faz necessário o registro na plataforma, com login e senha próprios. Além disso, também é necessário se ter ao menos 4 documentos para se realizar a transcrição.

1. Depois de entrar na plataforma com login e senha, clique em “Criar coleção’’ e dê um nome a coleção. Após, clique em “Upload Document” e faça o upload de pelo menos 4 documentos na plataforma.
2. Após, atualize a página e os documentos selecionados devem aparecer em uma pasta.
3. Abra a pasta e selecione 1 dos documentos. O documento selecionado deve aparecer no lado esquerdo da tela. Ao lado direito clique em “Iniciar transcrição automática”.
4. Abrirá uma página com diversos modelos de transcrição para serem aplicados ao documento. Selecione conforme o idioma original do documento. (Documentos em inglês recomendo o uso do modelo de nome: “The English Eagle”.) Basta usar o filtro no lado direito da página e escrever o nome do modelo.
5. Após a escolha do modelo, clique em “Iniciar o reconhecimento” na parte superior direita. Aguarde alguns segundos enquanto a plataforma irá processar a leitura do documento.
6. Após a leitura, a transcrição deve aparecer ao lado do documento. É possível ler e editar trechos em que a transcrição não teve acerto, palavras, pontuação etc.
7. Após edição da transcrição, na parte superior, clique em “Guardar” para salvar o documento.

ANEXO A – DOCUMENTO 1



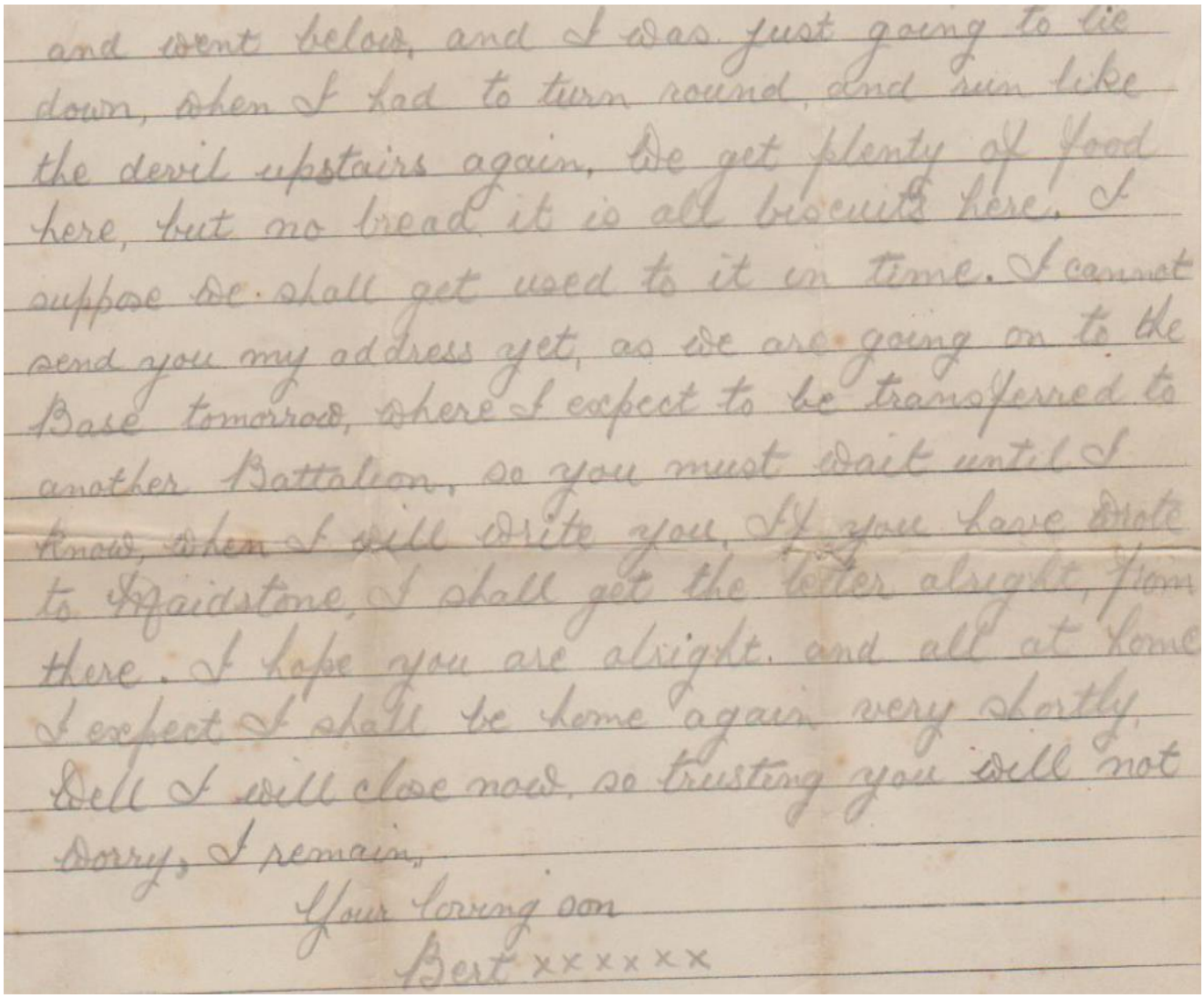
France. Jan. 14th

Dear Mother

No doubt you will be surprised to see I am over here. But you must not worry over me I shall be alright. Remember what I wrote you in my previous letter, I cannot write much or tell you where I am as these letters will be censored. I am in the best of health and spirits, and tons of mud. I am a rotten sailor. We were on the Water twenty one hours and I think half of Private. B. Hall. is now sailing down the Channel, under way for Timbuctoo. I was on gaurd on the vessel on the Day over, for two hours, gaurding the stairs, which lead down to

Fonte: Europeana (2024).

ANEXO B – DOCUMENTO 2



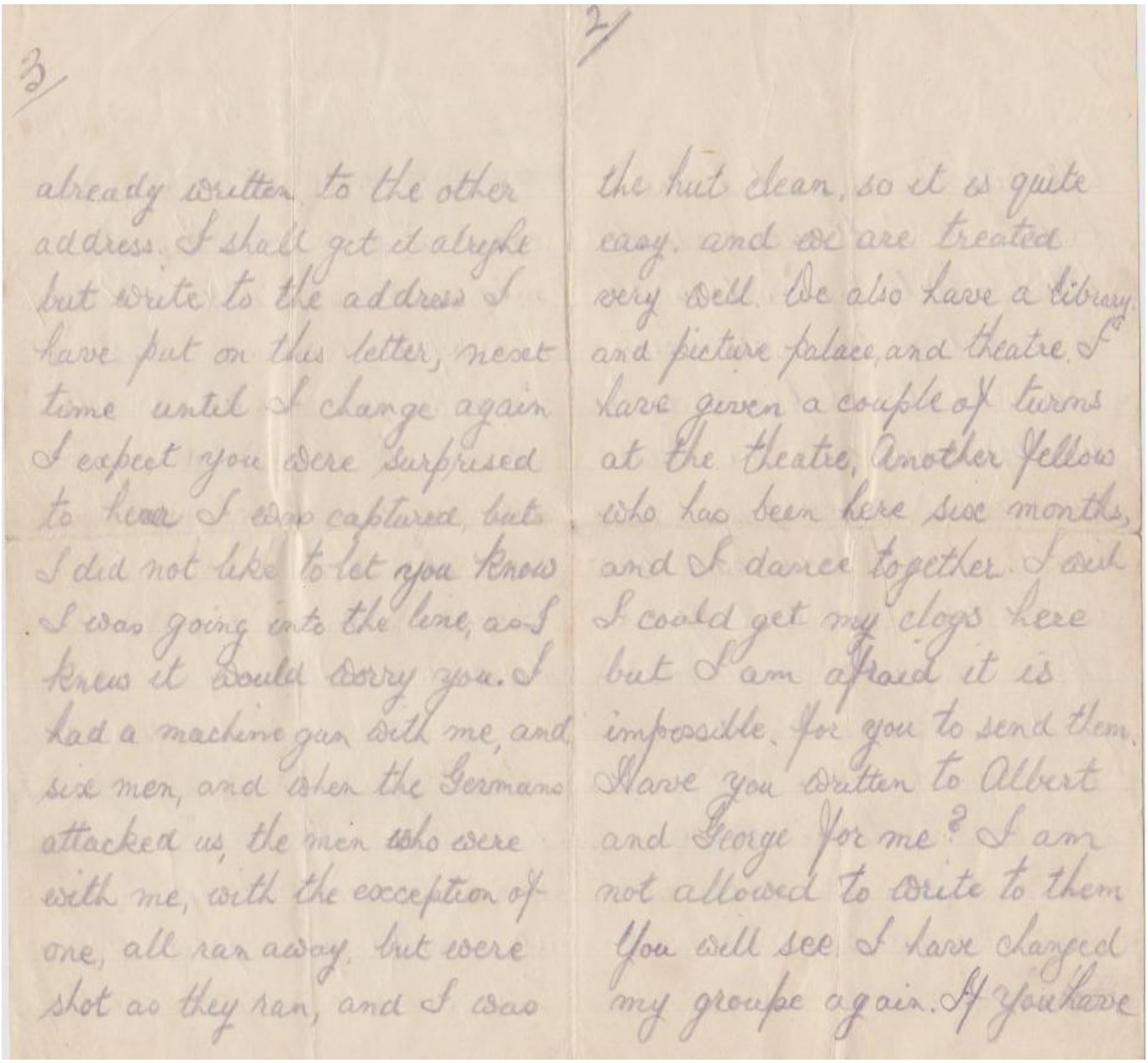
and went below, and I was just going to lie down, when I had to turn round and run like the devil upstairs again. We get plenty of food here, but no bread it is all biscuits here. I suppose we shall get used to it in time. I cannot send you my address yet, as we are going on to the Base tomorrow, where I expect to be transferred to another Battalion, so you must wait until I know, then I will write you. If you have wrote to Maidstone, I shall get the letter alright, from there. I hope you are alright, and all at home I expect I shall be home again very shortly. Well I will close now, so trusting you will not worry, I remain,

Your loving son

Bert xxxxxx

Fonte: Europeana (2024).

ANEXO C – DOCUMENTO 3



3/

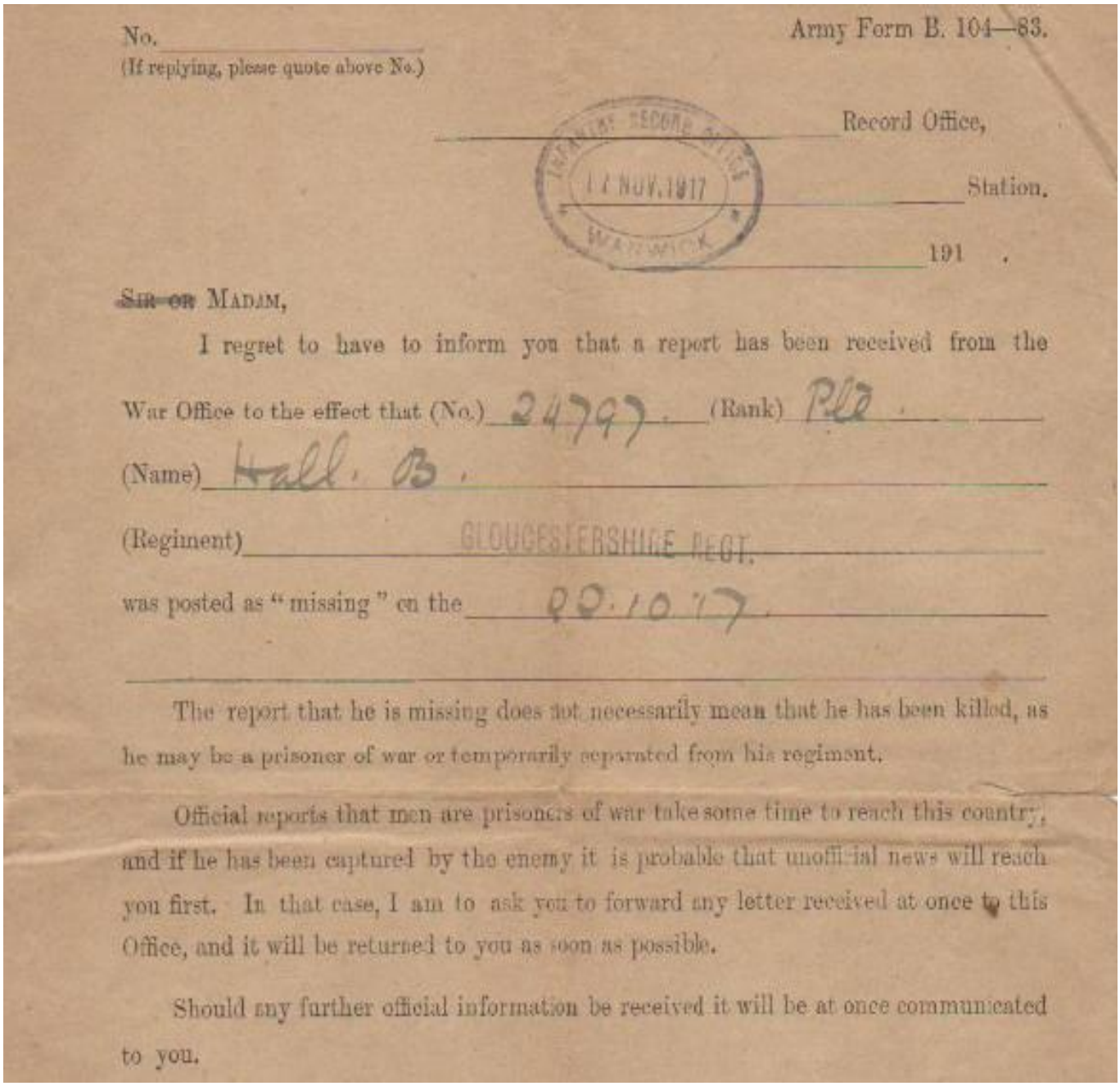
already written to the other address I shall get it alright but write to the address I have put on this letter, next time until I change again I expect you were surprised to hear I was captured, but I did not like to let you know I was going into the line, as I knew it would worry you. I had a machine gun with me, and six men, and when the Germans attacked us, the men who were with me, with the exception of one, all ran away, but were shot as they ran, and I was

2/

the hut clean, so it is quite easy, and we are treated very well. We also have a library and picture palace, and theatre. I have given a couple of turns at the theatre, another fellow who has been here six months, and I dance together. I wish I could get my clogs here but I am afraid it is impossible for you to send them. Have you written to Albert and George for me? I am not allowed to write to them You will see I have changed my groupe again. If you have

Fonte: Europeana (2024).

ANEXO D – DOCUMENTO 4



No. _____ Army Form B. 104—83.
(If replying, please quote above No.)

_____ Record Office,

_____ Station.

_____ 191 .

~~Sir or~~ Madam,

I regret to have to inform you that a report has been received from the War Office to the effect that (No.) 24797. (Rank) Pte.
(Name) Hall. B.
(Regiment) GLOUCESTERSHIRE REGT.
was posted as "missing" on the 22.10.17.

The report that he is missing does not necessarily mean that he has been killed, as he may be a prisoner of war or temporarily separated from his regiment.

Official reports that men are prisoners of war take some time to reach this country, and if he has been captured by the enemy it is probable that unofficial news will reach you first. In that case, I am to ask you to forward any letter received at once to this Office, and it will be returned to you as soon as possible.

Should any further official information be received it will be at once communicated to you.

Fonte: Europeana (2024).