

PATRIMÔNIO E PRESERVAÇÃO DOCUMENTAL: USOS DA INTELIGÊNCIA ARTIFICIAL NA DIGITALIZAÇÃO E CATALOGAÇÃO DE ACERVOS

Orlando da Silva Martins¹

Resumo: O presente trabalho objetiva analisar as potencialidades e limitações do emprego de inteligência artificial, no que diz respeito a digitalização e catalogação de acervos documentais. Em evidência na contemporaneidade em virtude da sua importância para o entendimento acerca dos processos envolvidos na construção identitária e da mentalidade histórica de determinados segmentos humanos, a pesquisa permite-nos discutir a relevância e os desafios do emprego de inteligências artificiais na preservação patrimonial. Partindo de uma metodologia qualitativa, realizamos o levantamento da documentação técnica vinculada ao projeto “FamilySearch” e publicada entre 2015 e 2020. A partir da análise desta, fomos capazes de identificar os potenciais e problemas enfrentados pelo projeto, no que tange a implementação de inteligências artificiais no processo de preservação do documento e, por conseguinte, do patrimônio histórico-cultural dos segmentos que o produziram.

Palavras-chave: Preservação documental, Inteligência artificial, Patrimônio histórico-cultural.

Introdução

A digitalização e a catalogação de documentos são algumas das formas mais eficazes de assegurar a preservação e o acesso a registros culturais e históricos que, de outra forma, estariam em risco de deterioração irreversível. Distanciando-nos da concepção metódica que coloca a fonte como um fim em si mesma, o processo de preservação implica em uma abordagem que naturalize a leitura crítica do acervo dentro de um fluxo de transformação do documento em patrimônio. Advinda do progressivo alargamento das fontes empregadas na pesquisa histórica e do entendimento do documento como promotor de sentimentos e reflexões, essa percepção elucida um

¹ Graduando em História pela Universidade Estadual do Paraná (UNESPAR) – *Campus* de Campo Mourão.

constante diálogo entre fonte e indivíduo e, portanto, uma convergência entre os registros contidos no documento e o acesso ao patrimônio histórico que ele representa (Le Goff, 1984; Pereira Filho, 2022).

Nesse contexto, o emprego da inteligência artificial (IA) no tratamento arquivístico de acervos históricos e culturais é um dos tópicos mais discutidos nas últimas décadas, dada a sua capacidade de transformar processos tradicionalmente analógicos e demorados em operações mais eficientes e acessíveis (Russel; Norvig, 2016). Contudo, apesar da agilidade e acessibilidade que proporciona, a aplicação da IA na preservação patrimonial encontra desafios consideráveis relacionados ao modo como é encarada. Apesar de realizar a leitura literal, digitalização e indexação dos textos com rapidez, a IA tem dificuldades em reproduzir alguns procedimentos inerentes a ciência histórica, como a re inserção da fonte em seu contexto sociocultural de origem e a interpretação nuançada dos registros que permeiam o documento.

Assim, consoante a importância de discutirmos o impacto dessas transformações nas práticas de preservação documental e o modo como entendemos o acervo, o presente trabalho visa analisar as potencialidades e limitações do emprego de inteligências artificiais, no que tange à digitalização e catalogação de acervos, bem como à conservação do patrimônio histórico-cultural.

Materiais e procedimentos

A metodologia utilizada foi a análise qualitativa da documentação técnica do projeto FamilySearch à luz da literatura sobre o uso de IA, com foco em Deep Learning – ou aprendizagem profunda, em português – e no Reconhecimento Óptico de Caracteres (OCR), aplicados ao contexto da preservação de acervos documentais. O *corpus* empírico foi composto por 89 documentos disponíveis na seção “Comunidade” do site FamilySearch.org, sendo 46 deles provenientes da categoria “Questões Gerais de Pesquisa”, 28 da categoria “Indexação” e 15 das “Questões Gerais”.

De início, foi realizada uma revisão bibliográfica de fontes científicas, manuais técnicos e relatórios institucionais, com vistas a estabelecer as principais características das tecnologias de IA, suas aplicações no patrimônio histórico-cultural e os desafios específicos enfrentados na área.

A partir desse estudo teórico-empírico, a documentação do projeto e a literatura pertinente foram organizadas em duas categorias: (1) as potencialidades da IA, especialmente na digitalização, catalogação e acesso a acervos e (2) as limitações e desafios que comprometem a eficiência e a precisão desses sistemas. Utilizando dados empíricos do projeto *Family Search*, foram observadas as vantagens e as limitações específicas da aplicação dessas tecnologias em um contexto real.

Mantido pela Igreja de Jesus Cristo dos Santos dos Últimos Dias, o FamilySearch é um dos maiores e mais reconhecidos sites de genealogia e história familiar do mundo. Fundado em 1894 sob o nome *Genealogical Society of Utah*, o projeto inicialmente visava preservar registros genealógicos e promover o interesse pela história familiar entre os membros da igreja. Com o passar dos anos, o projeto evoluiu para alcançar um público global, resultando em uma plataforma digital que permite a qualquer pessoa, independentemente de religião, pesquisar e documentar suas linhagens familiares.

A trajetória da FamilySearch está profundamente ligada à inovação na preservação documental. Nos anos 1930, a instituição começou a usar microfilmagem para arquivar documentos históricos e genealógicos. Essa iniciativa resultou em uma coleção massiva de microfimes de registros civis, eclesiásticos e censitários de diversas partes do mundo. Em 1999, com o avanço da tecnologia, o site FamilySearch.org foi lançado, permitindo acesso digital a milhões de registros, democratizando e ampliando significativamente o acesso à informação genealógica.

O impacto de FamilySearch na preservação documental e patrimonial é muito grande e abrangente. Ao digitalizar e disponibilizar registros históricos, como certidões de nascimento, casamento, óbito, registros de imigração e censos, o site não só preserva esses documentos, como também permite a construção e o acesso à uma memória coletiva muitas vezes silenciada (Pollak, 1989). Além disso, o projeto também colabora com arquivos e bibliotecas ao redor do mundo, contribuindo para a preservação de documentos que muitas vezes são únicos e vulneráveis. Esse trabalho de cooperação envolve desde a digitalização até a capacitação técnica dos gestores de acervos, o que torna o site um agente ativo e estratégico na preservação do patrimônio documental.

Resultados e discussões

A IA tem contribuído significativamente para a transformação do modo como acervos documentais são digitalizados, catalogados e acessados pelo público. No contexto do FamilySearch, as tecnologias de aprendizado profundo e OCR se mostraram fundamentais para a transcrição de documentos históricos e a criação de ferramentas de pesquisa acessíveis aos pesquisadores e familiares interessados em genealogia.

O OCR é utilizado para converter documentos impressos e manuscritos em texto digital, tornando acessíveis informações que antes estavam limitadas a suportes físicos e tornando-as “pesquisáveis” a partir de filtros como indexadores e palavras-chave. Já o aprendizado profundo constitui um conceito mais amplo relacionado ao funcionamento de algoritmos em relação ao seu grau de autonomia. A partir de redes neurais, esses são capazes de identificar padrões e tomar ações baseadas na experiência incorporada pela repetição e *feedback* (Bezerra, 2016).

A aplicação de algoritmos de aprendizado profundo (Deep Learning) representa um avanço significativo no reconhecimento de caracteres complexos. Sendo os bases do aprendizado profundo, os sistemas de redes neurais convolucionais (CNNs), têm sido especialmente eficazes na identificação de padrões em manuscritos e documentos antigos, superando algumas limitações encontradas nos métodos tradicionais (Lecun et al., 2015). Além disso, a capacidade de aprendizado dos algoritmos de Reconhecimento Óptico de Caracteres (OCR) permite que eles se adaptem ao longo do tempo, melhorando a precisão e facilitando o reconhecimento de características mais contextuais, como a caligrafia, que é comum em documentos de diferentes períodos e culturas (Smith, 2007).

Durante a análise da documentação, foram identificados 17 casos em que a capacidade de indexação automática da IA foi elogiada pela comunidade de usuários. Em uma das publicações, referente a um livro paroquial português de 1817, o usuário comenta que realizou buscas por um familiar em junho de 2017, mas não obteve nenhum retorno. Porém, ao empreender nova tentativa em maio de 2020, o livro que continha o assento de batismo de seu ascendente havia sido digitalizado e as informações do seu interior já se encontravam indexadas pelo FamilySearch.

Constantemente integrando novos documentos e acervos ao seu repositório virtual, o projeto vale-se da IA a fim de tornar o tratamento arquivístico um processo mais dinâmico e veloz. Sem a automação trazida pela inteligência artificial, o processo de digitalização e indexação continuaria possível, contudo, mais compassado e dependente

Patrocínio Master

Gestão

Realização

de mão-de-obra humana. O impasse é que justamente nessa dinamização do acesso e disseminação da informação reside a sociedade em rede, momento da história em que, segundo Castells (1999), nos organizamos em torno de espaços de fluxo comunicacional constantes que influenciam a forma como interagimos e nos conectamos com o meio social.

Logo, outro ponto relevante é a sua capacidade de categorizar e filtrar automaticamente vastas quantidades de dados. Em acervos extensos, como o FamilySearch, a IA organiza automaticamente documentos em categorias predefinidas, o que agiliza a busca de informações específicas, como certidões de nascimento, casamento e óbito, poupando tempo e esforço em processos que, anteriormente, demandariam horas de trabalho manual. Esse potencial de automação e avanço rumo a uma autogestão é essencial para democratizar o acesso a acervos históricos, vez que atende a uma crescente demanda por dados e informações, contribuindo para a preservação da memória coletiva (Unesco, 2003).

Contudo, apesar dos avanços, a IA enfrenta desafios persistentes quando aplicada ao campo da preservação do patrimônio histórico-cultural. Os algoritmos de IA têm dificuldades em interpretar contextos históricos complexos e em entender nuances culturais que podem impactar a leitura dos dados. Termos que mudaram de significado ao longo do tempo, por exemplo, frequentemente causam problemas para a IA, que pode interpretar erroneamente palavras, expressões ou mesmo significados históricos de forma literal, sem a sensibilidade necessária para compreender o contexto original (Mullainathan; Spiess, 2017).

Consoante a isso, Cruz e Peixoto (2007) chamam atenção para a importância de lançarmos um olhar sensível sobre o documento. Apesar de trabalharem especificamente com periódicos, sua análise permite-nos traçar um paralelo com as IAs e evidencia a importância da adoção de procedimentos teórico-metodológicos a fim de compensar a ausência de sensibilidade da máquina. Sem esse suporte, a pesquisa constatou que o algoritmo pode ser falho em capturar os matizes que permeiam a conjuntura sociocultural de uma população em certo período histórico e, portanto, pode resultar em uma interpretação imprecisa e fragilizada dos fatos.

Nesse sentido, a necessidade de supervisão humana é um aspecto essencial na aplicação da IA no processo de preservação do patrimônio histórico-cultural. Ao encontro

disso, Gomez (2019) destaca a importância da colaboração entre IA e especialistas humanos para evitar distorções e falhas que possam comprometer a integridade da análise historiográfica. Embora a IA seja eficaz em processos automatizados, a interpretação de documentos históricos ainda exige a presença de humanos que possam validar os dados e corrigir possíveis erros. Em um campo onde a precisão é essencial, a supervisão humana atua como um filtro crítico para garantir a qualidade e a autenticidade dos dados preservados.

Considerações finais

Ao longo deste estudo, a aplicação de Inteligência Artificial (IA) na digitalização, catalogação e conservação do patrimônio histórico-cultural revelou-se um campo em franca expansão, com potencial significativo para transformar práticas tradicionais e ampliar o acesso aos acervos documentais. O uso de tecnologias como o aprendizado profundo (Deep Learning) e o Reconhecimento Óptico de Caracteres (OCR) oferece meios eficazes para converter e indexar grandes volumes de documentos, possibilitando um novo nível de acessibilidade e rapidez no processamento de dados. Nesse contexto, a análise do projeto FamilySearch ilustra os avanços e os desafios inerentes à implementação dessas tecnologias na preservação patrimonial, destacando a necessidade de um equilíbrio entre automação e intervenção humana.

A pesquisa evidenciou como a IA, por meio do Deep Learning e do OCR, consegue identificar padrões de escrita e caracteres em documentos históricos, mesmo naqueles cuja preservação física já se encontra ameaçada. Esse avanço é particularmente útil no reconhecimento de manuscritos e fontes antigas que possuem alta variabilidade em aspectos como caligrafia, estrutura e formato. A capacidade de aprender e adaptar-se a diferentes estilos de escrita permite ao OCR evoluir conforme novos dados são incorporados ao sistema, tornando-o uma ferramenta cada vez mais precisa e adaptável às particularidades de cada acervo.

Além disso, a IA possibilita a categorização automática de documentos, o que representa um ganho de produtividade ao permitir a indexação de informações de forma organizada e acessível. No caso do FamilySearch, a automação de categorias, como certidões de nascimento, casamento e óbito, facilita a busca e o acesso a registros específicos, democratizando o acesso ao conhecimento histórico e genealógico para um

público global. Ao transformar processos analógicos em digitais, a IA atende diretamente a essa demanda por informação, promovendo o compartilhamento do patrimônio de maneira mais ágil e eficaz.

No entanto, o estudo também apontou limitações importantes, como a dificuldade da IA em interpretar nuances contextuais e históricas presentes nos documentos. Em muitos casos, termos e expressões que mudaram de significado ao longo do tempo são interpretados de maneira literal, comprometendo possíveis compreensões acerca do conteúdo dessas fontes. Deste modo, esta falta de sensibilidade para captar o contexto original exige uma abordagem metodológica mais abrangente, que compense as falhas interpretativas da IA.

A análise do FamilySearch exemplifica bem essa limitação, uma vez que os sistemas de IA aplicados na plataforma, apesar de eficazes, não substituem a necessidade de supervisão humana. Os especialistas são essenciais para a validação dos dados e para corrigir eventuais imprecisões geradas por interpretações automáticas. A IA, por mais desenvolvida que seja, ainda não é capaz de substituir a análise contextual e o julgamento crítico que são característicos da historiografia. Essa supervisão humana assegura que os documentos sejam preservados e interpretados à luz do seu contexto original, garantindo a integridade e o acesso ao patrimônio cultural salvaguardado.

Assim, a utilização de IA na digitalização e catalogação de acervos representa tanto uma revolução quanto um desafio para a preservação do patrimônio histórico-cultural. Enquanto a IA expande a acessibilidade e dinamiza o processo de preservação, seu uso deve ser cuidadosamente balanceado com a supervisão de um pesquisador, garantindo que as interpretações sejam mais precisas e o contexto histórico da produção documental preservado. A IA, nesse sentido, não substitui a intervenção humana, mas complementa as práticas tradicionais, ampliando a capacidade de preservação de documentos históricos e, por conseguinte, do patrimônio cultural que constituem.

Referências

BEZERRA, Eduardo. Introdução à aprendizagem profunda. **Artigo–31º Simpósio Brasileiro de Banco de Dados–SBBD2016–Salvador**, 2016.

CASTELLS, Manuel. A cultura da virtualidade real: a integração da comunicação eletrônica, o fim da audiência de massa e o surgimento de redes interativas. In: CASTELLS, Manuel. **A sociedade em rede**. 6. ed. São Paulo: Paz e Terra, 1999.

CRUZ, Heloisa de Faria; PEIXOTO, Maria do Rosário da Cunha. Na oficina do historiador: conversas sobre história e imprensa. **Revista Projeto História**, n. 35, p. 253-270, 2007.

GOMEZ, Pedro. Digital Preservation and Historical Context: The Human Factor in AI. **Journal of Historical Archives**, v. 25, n. 3, p. 145-162, 2019.

LECUN, Yann et al. Deep Learning. **Nature**, v. 521, n. 7553, p. 436-444, 2015.

LE GOFF, Jacques. Documento/ Monumento. **Enciclopédia Einaudi**: memória-história. Lisboa: Imprensa Nacional/ Casa da Moeda, 1984. v. 1. p. 95-106.

MULLAINATHAN, Sendhil; SPIESS, Jonas. Machine Learning: An Applied Econometric Approach. **Journal of Economic Perspectives**, v. 31, n. 2, p. 87-106, 2017.

PEREIRA FILHO, Hilario Figueiredo. Documento e patrimônio entre usos e reflexões. **Tempo Social**, São Paulo, Brasil, v. 34, n. 3, p. 295–313, 2022.

POLLAK, Michael. Memória, Esquecimento, Silêncio. **Revista Estudos Históricos**, Rio de Janeiro, v. 2, n. 3, 1989.

RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence: A Modern Approach**. 3. ed. New Jersey: Prentice Hall, 2016.

SILVA, Mariana; SANTOS, Ricardo. Desafios na Digitalização de Acervos Culturais em Países em Desenvolvimento. **Revista Brasileira de Preservação Cultural**, v. 7, n. 1, p. 30-50, 2020.

SMITH, Ray. An Overview of the Tesseract OCR Engine. In: **Proceedings of the Ninth International Conference on Document Analysis and Recognition**. Washington: IEEE, 2007.

UNESCO. **Charter on the Preservation of the Digital Heritage**. Paris: UNESCO, 2003.