

PERSPECTIVAS DE APLICAÇÃO DE INTELIGÊNCIA ARTIFICIAL EM ARQUIVOS: NOVAS ABORDAGENS PARA AS PRÁTICAS ARQUIVÍSTICAS

Jéssyca Janiffer Diniz de Almeida (Museu de Astronomia e Ciências
Afins/Instituto do Patrimônio Histórico e Artístico Nacional)

1 INTRODUÇÃO

A Inteligência Artificial (IA) se configura como um ramo da Ciência da Computação que busca explicar e simular comportamentos inteligentes por meio de métodos, ferramentas e sistemas computacionais. Suas soluções são utilizadas para realizar atividades complexas e que demandem raciocínio de forma mais rápida, analítica e assertiva, potencializando os resultados e auxiliando nas tomadas de decisão.

As aplicações de IA estão em ascensão e os investimentos no campo têm possibilitado avanços em distintas áreas do conhecimento e prometem ajudar a resolver desafios globais, como mudanças climáticas, educação, bem-estar público e acesso a cuidados de saúde mais eficazes, eficientes e equitativos. Em menor magnitude, as técnicas de Inteligência Artificial estão revolucionando e beneficiando quase todos os aspectos de nossa sociedade e economia – desde comércio até transporte e segurança cibernética –, sendo cada vez mais implantadas em nossas atividades cotidianas, influenciando a forma como

trabalhamos e nos divertimos e proporcionando benefícios ao nosso bem-estar social (IBM, 2020; OECD, 2021).

A IA pode, por exemplo, auxiliar na área financeira, em áreas como gestão de ativos, negociação financeira ou subscrição de crédito (OECD, 2021); na tomada de decisões clínicas e saúde pública, pesquisa biomédica, desenvolvimento de medicamentos e administração de sistemas de saúde (HASHIGUCHI; SLAWOMIRSKI; ODERKIRK, 2021); no aumento da produtividade na manufatura, acelerando a pesquisa industrial, o treinamento da força de trabalho e auxiliando no gerenciamento das cadeias de suprimentos (NOLAN, 2021); e em nosso dia a dia, por meio de tecnologias de mapeamento, assistentes inteligentes, *chatbots* para atendimento a clientes, filtragem de spam, tradução de idiomas, mecanismos para fornecimento de recomendações automatizadas para programas de TV e música com base nos hábitos de visualização dos usuários, publicidade personalizada e muito mais (ORACLE, 202[?]).

De acordo com Agência dos Estados Unidos para o Desenvolvimento Internacional – USAID (2022, tradução nossa), “o impacto projetado dos aplicativos de IA na economia global até 2030 é equivalente a um aumento no PIB global de 16%, e prevê-se que desempenhe um papel na abordagem de cada um dos Objetivos de Desenvolvimento Sustentável da ONU”. Diante dessa prevalência cada vez maior das tecnologias de inteligência artificial e do aumento do acesso aos dados e ao poder de computação que amplia sua aplicabilidade e acessibilidade, fica o questionamento de como a IA pode auxiliar nas práticas arquivísticas? Que avanços na tecnologia de IA apresentam oportunidades para os arquivos e quais os desafios e problemas de seu uso na área?

Esse trabalho, portanto, tem como objetivo verificar como a Inteligência Artificial tem sido aplicada nos arquivos e suas implicações nas práticas arquivísticas a partir da revisão da literatura. Para tanto, a metodologia consistiu no levantamento de material bibliográfico e análise dos mesmos, identificando os estudos que tratam da aplicação de tecnologias de inteligência artificial no campo da Arquivologia. Foram utilizadas para levantamento a Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI) e o Google Scholar considerando, principalmente, publicações dos últimos cinco anos. O recorte temporal priorizado levou em consideração que as pesquisas, estudos e discussões sobre o uso da IA no campo da Arquivologia brasileira é relativamente recente, embora como verificado por Duranti et. al (2022) não seja uma ideia recente.

Embora possa ser considerado um campo fértil, a temática ainda não apresenta uma literatura científica expressiva voltada para a relação da IA com

os estudos e práticas arquivísticas. Assim, o desenvolvimento de um quadro conceitual sobre o tema permitirá apresentar e discutir como a tecnologia vêm sendo proposta e explorada em arquivos, além de promover uma reflexão a respeito de como a transformação digital está reconfigurando as práticas arquivísticas.

2 INTELIGÊNCIA ARTIFICIAL

Definir de forma breve o termo “inteligência artificial” é tarefa complexa, visto que possui muitas definições. A discussão sobre o tema remonta os primórdios da computação, mas devido a sua natureza sempre mutável e a abrangência de uma gama de técnicas e tecnologias de automação em si mesma, descrevê-la deixou de ser simples (ROLAN, et al. 2018). Cunhado em 1956 por John McCarthy, o termo foi pensado como a ciência e a engenharia de fazer máquinas com inteligência semelhante à humana e programas de computador que pudessem entender, inferir e aprender. Em outras palavras, pode-se dizer que Inteligência Artificial se refere à capacidade de um sistema baseado em máquina em, para um determinado conjunto de objetivos definidos por humanos, executar tarefas comumente associadas aos processos intelectuais característicos dos humanos, como a capacidade de raciocinar, descobrir significados, generalizar ou aprender com experiências passadas (COPELAND, 2022; OECD, 2021).

Existem ainda muitos outros termos relacionados à IA, como *deep learning*, *machine learning*, reconhecimento de imagem, processamento de linguagem natural, computação cognitiva, amplificação de inteligência, dentre outros (KUSUMAWATI; SALIM, 2022). Os sistemas de IA são projetados para operar com vários níveis de autonomia para que possam fazer previsões, recomendações ou decisões que influenciem ambientes reais ou virtuais. Frequentemente mencionados, os campos secundários de *machine learning* e *deep learning*, são compostos por algoritmos de IA que buscam criar sistemas especializados e capazes de fazer previsões ou classificações com base em dados de entrada (IBM, 2020).

A tecnologia de *Machine Learning*, também conhecida como aprendizado automático ou ainda aprendizado de máquinas, se concentra no padrão de dados e algoritmos para simular a maneira como os humanos aprendem, melhorando gradualmente sua precisão, a partir da análise dos resultados obtidos e associação de diferentes dados. Já o *Deep Learning* se configura como uma subárea de *Machine Learning* e é utilizado para resolver problemas muito complexos que geralmente envolvem grandes quantidades de dados. Engloba, essencialmente, redes neurais com três ou mais camadas organizadas para reconhecer relacionamentos e

padrões complexos nos dados, de forma a possibilitar um aprendizado a partir dessas análises (IBM, 2020; ROUHIAINEN, 2018).

2.1 A Inteligência Artificial e os Arquivos

A produção acelerada e o consequente acúmulo de documentos nos mais diversos suportes e formatos conduziram ao uso intensivo dos recursos tecnológicos nas práticas e procedimentos arquivísticos. No entanto, apesar da agilidade e flexibilidade que o uso das tecnologias da informação e comunicação proporcionam, os arquivistas precisam lidar com volumes cada vez maiores de documentos digitalizados e natos digitais, tornando as atividades de manutenção, classificação, avaliação, destinação e preservação de documentos de arquivo cada vez menos viáveis de serem realizadas de forma manual.

Dessa forma, a IA se apresenta como uma ferramenta promissora para auxiliar na automatização e otimização das práticas arquivísticas. Rolan et al. (2018) são, inclusive, mais categóricos ao afirmar que tais auxílios tecnológicos, bem como as habilidades e conhecimentos necessários para manejá-los, serão necessários para enfrentar os desafios do gerenciamento de documentos na era digital. Sem dúvidas, como afirmado por Colavizza et al. (2021), o papel dos arquivistas é, então, transformado: precisam do suporte das máquinas para auxiliá-los a trabalhar com a explosão de documentos ao mesmo tempo em que precisam aprender a fazer uso dessas tecnologias, até mesmo para avaliar os resultados alcançados pelas máquinas.

Jaillant (2022) destaca esta questão como um dos principais desafios a serem enfrentados no caminho desde a produção documental até a sua avaliação, análise e preservação. Segundo a autora, apesar da Ciência de Dados e da IA estarem se tornando ferramentas essenciais, são poucos os acadêmicos treinados para dominar esses métodos de pesquisa, especialmente na área das humanidades, o que tem um grande impacto na formação de futuros profissionais.

A transformação digital está modificando os arquivos e redefinindo a forma como os profissionais da área atuam, requerendo novos saberes e desenvolvimento de novas práticas. Como consequência, a automação na forma de técnicas de inteligência artificial é cada vez mais estudada e elegida como uma possibilidade tanto para dimensionar as atividades tradicionais de manutenção de documentos quanto para experimentar novas maneiras de capturá-los, organizá-los e acessá-los (COLAVIZZA, et al. 2021). Da mesma forma, a própria sustentabilidade da IA também depende de uma gestão de documentos eficiente no ambiente digital. Conforme afirma Kusumawati e Salim (2022), conhecimentos arquivísticos também são necessários para lidar com a enorme

quantidade de dados da qual depende a implementação de aplicativos de IA eficientes.

Segundo Coelho (2022), em palestra proferida pela 6ª Semana Nacional de Arquivos, sobre aplicações de Inteligência Artificial em Arquivos, para implementação em arquivos, os modelos de IA podem servir a diferentes propósitos, tais como: classificação de tipologias documentais, a serem determinadas automaticamente, sem qualquer interferência humana; anonimização de documentos, para que determinados sujeitos não tenham acesso a certas informações sensíveis, mesmo acessando o documento durante o processo de tramitação; extratores de texto que, a partir do processamento de linguagem natural (PLN)¹⁷⁴, permitem extrair automaticamente informações estruturadas de fontes textuais e reconhecimento das entidades nomeadas; e resumos documentais, que permitem a elaboração de resumos do documento automaticamente.

Destaco ainda a possibilidade de aplicação de outras tarefas subjacentes utilizadas em níveis mais complexos de PLN em arquivos, tais como: categorização de conteúdo, para possibilitar a pesquisa e indexação a partir de termos chave; conversão fala-texto e texto-fala, que transforma em texto escrito em comandos de voz e vice-versa, para auxiliar na acessibilidade dos arquivos; e tradução de máquina, que traduz texto ou fala de um idioma para outro automaticamente, para análise de documentos em diferentes idiomas presentes no acervo, especialmente em arquivos pessoais.

Para a aplicação da Inteligência Artificial no tratamento dos documentos arquivísticos digitais, conforme ressaltado por Coelho e Siebra (2022), é preciso levar em consideração algumas necessidades específicas:

- Acesso a dados e documentos reais – a IA precisa ser treinada com documentos verdadeiros para diminuir BIAS;
- Alto nível de segurança – considerações às questões da LGPD e de como se portar com dados sensíveis;
- Hardware especializado – computadores específicos para executar processamentos de redes neurais;
- Equipe multidisciplinar – além de arquivistas e documentaristas é necessário matemáticos para traduzir o conteúdo dos documentos, estatísticos para tratar o volume de dados e especialistas em PLN;

¹⁷⁴ O processamento de linguagem natural (PLN) consiste no desenvolvimento de modelos computacionais para a realização de tarefas que dependem de informações expressas em alguma língua natural, incorporando, para isso, técnicas diversas para interpretar a linguagem humana, desde métodos estatísticos e de *machine learning* a abordagens algorítmicas e baseadas em regras (PEREIRA, [s.d.]; SAS, 2022).

- Curadoria digital – especialistas do tema da matéria em questão (COELHO, SIEBRA, 2022).

Diante das diferentes tecnologias, modelos e ferramentas que compõem o campo da Inteligência Artificial, percebe-se que múltiplas funções podem ser desempenhadas para auxiliar no tratamento e gerenciamento dos documentos arquivísticos. Apesar de que os projetos de IA envolvem ainda um grau de experimentação, diversas iniciativas têm sido desenvolvidas por pesquisadores e instituições internacionais e nacionais, de forma a explorar e testar o potencial da tecnologia. Alguns casos foram identificados e compilados para ilustrar como estas tecnologias podem ser aplicadas nos Arquivos.

2.1.1 Aplicações de IA em Arquivos

Os campos de estudo da IA para aplicação em arquivos envolvem mineração de dados (*data mining*) ou mineração de textos (*text mining*), processamento de linguagem natural, sistemas especialistas; computação cognitiva; e visão computacional centrada em análises e processamento de imagens (COELHO, SIEBRA, 2022).

Um exemplo de aplicação é o projeto do Arquivo Nacional da Austrália, que investiga como criar e emitir autorizações de eliminação e recolhimento de documentos em um formato gerenciável e executável pelos sistemas digitais, com o mínimo envolvimento humano (ROLAN et al, 2019). Há também o experimento de Vellino e Alberts (2016), que utilizou métodos automatizados para auxiliar na avaliação e classificação automática de documentos de e-mail; e a prova de conceito de Büttner (2017), que descreveu como métodos automatizados podem ser aplicados para identificar relacionamentos entre os documentos digitais e como essas tecnologias podem ajudar a fornecer contextos adicionais. O trabalho de Shang et al (2019) propõe um método de classificação baseado em computação paralela distribuída utilizando algoritmos que podem melhorar continuamente a precisão da classificação, alcançando resultados promissores na simulação, realizando com grande precisão a classificação de documentos. Já Kestemont, Christlein e Stutzmann (2017), se concentram na paleografia, numa pesquisa que visou a identificação automatizada, a partir de reproduções fotográficas, de tipos de escrita em manuscritos medievais.

As tecnologias de IA também permitem a captura e organização de documentos de eventos que normalmente escapariam das políticas de aquisição autodefinidas dos arquivos institucionais, apoiando práticas mais democráticas e inclusivas, como por exemplo, ao preservar e disponibilizar documentos de comunidades sub-representadas e grupos minoritários (GUPTA, 2020).

Em outros casos, tecnologias e métodos de IA são propagados para analisar as novas e vastas coleções de documentos natos digitais que surgem dentro e fora das instituições arquivísticas. Connelly [2020] discute o uso de técnicas de PNL para extrair metadados de documentos diplomáticos do banco de dados Freedom of Information Archive. Eles demonstram o potencial dessa abordagem ao criar um índice de “importância do país” no contexto das prioridades da política externa dos EUA.

No panorama nacional, têm-se como exemplo o “MPRJ em Mapas”, apresentado no Arquivo Nacional em 2019 por Henrique de Andrade, líder da Coordenação de Análise, Diagnóstico e Geoprocessamento do Ministério Público do Estado do Rio de Janeiro – MPRJ. A palestra cuja temática foi centrada no uso de inteligência artificial no serviço público para a classificação de documentos, exibiu o projeto que trata de um conjunto de ferramentas e plataformas de análise, diagnóstico e georreferenciamento a partir do qual são feitos cruzamentos das múltiplas bases de dados para compartilhamento de informações, gestão e diagnóstico (MINISTÉRIO PÚBLICO DO ESTADO DO RIO DE JANEIRO, 2020).

Cabe destacar ainda a instituição recente do novo Projeto InterPARES: o I Trust AI, que teve início em 2021 e se destina “a projetar, desenvolver e alavancar a IA para apoiar a disponibilidade e acessibilidade contínuas de documentos arquivísticos públicos confiáveis” (DURANTI, et al. 2022, p. 02), por meio de uma parceria sustentável e contínua entre a academia, as instituições arquivísticas, os profissionais de arquivo e a indústria, para produção de pesquisas originais e treinamento de estudantes e outros profissionais altamente qualificados. Dividido em cinco etapas a serem cumpridas anualmente, o projeto visa identificar tecnologias específicas de IA que possam lidar com os desafios dos arquivos; determinar os benefícios e riscos do uso dessas tecnologias em arquivos; assegurar que os conceitos e princípios arquivísticos sejam a base do desenvolvimento responsável de uma IA; e validar os resultados do por meio de estudos de caso e demonstrações.

3 CONSIDERAÇÕES FINAIS

O campo da Inteligência Artificial (IA) teve um avanço considerável nos últimos anos, gerando resultados e criando possibilidades em um número cada vez maior de domínios. A ideia de propor o uso de ferramentas e modelos que utilizem tecnologias de IA para automatizar e facilitar as práticas arquivísticas é convidativa e desafiadora. No entanto, o uso de novas tecnologias também representa desafios e dúvidas em relação ao seu impacto nas práticas profissionais, bem como em sua escalabilidade ao longo do tempo. Embora seu

potencial seja notável e já sejam identificáveis casos interessantes de sua aplicabilidade com diferentes propósitos, percebe-se que o conceito tem um longo caminho a percorrer para amadurecer e se adaptar a fim de otimizar as práticas arquivísticas.

O uso de aplicativos de inteligência artificial não pretende substituir o papel dos arquivistas como um todo. Embora na prática as tecnologias da informação e comunicação dominem as práticas arquivísticas, a aplicação da IA visa auxiliar os arquivistas no cumprimento de suas funções, devido, especialmente, ao grande volume de documentos. O papel dos arquivistas é, portanto, transformado, à medida em que complexos desafios vêm se impondo aos arquivos, especialmente decorrentes do crescimento do volume de conteúdo digital e da rápida obsolescência de *hardwares* e *softwares*, do uso de sistemas comerciais e proprietários, e dos múltiplos formatos digitais, o que resulta em outros tipos de problemas associados ao ciclo de vida digital, à garantia de autenticidade dos documentos digitais e à sua permanência e acesso a longo prazo.

Embora muitos dos riscos potenciais associados à IA em arquivos não sejam exclusivos dessa inovação, o uso de tais técnicas pode ampliar essas vulnerabilidades, dada a extensão da complexidade das técnicas empregadas, sua adaptabilidade dinâmica e seu nível de autonomia.

Apesar de vários desafios dificultarem sua implantação, em especial devido a falta de competências e conhecimento destas tecnologias pelos profissionais de arquivo, os riscos emergentes da implantação de técnicas de IA precisam ser identificados e mitigados para apoiar e promover o uso de IA responsável e a formulação de políticas que avaliem as implicações dessas novas tecnologias, identifiquem os benefícios e riscos relacionados ao seu uso e orientem o design e as boas práticas para seu uso.

REFERÊNCIAS

BÜTTNER, Gesa. Auto-classification in an international organization: report from a feasibility study. **Comma**. (2): 15–26. 2017.

COELHO, Barbara; SIEBRA, Sandra. **6ª SNA - Aplicações da inteligência artificial em arquivos**. Youtube, 09 Jun. 2022. Disponível em <https://www.youtube.com/watch?v=9u9iQ17Ys48>.

COLAVIZZA, Giovanni; BLANKE, Tobias; JEURGENS, Charles; NOORDEGRAAF, Julia. Archives and AI: an overview of current debates and future perspectives. **J. Comput. Cult. Herit**. 15, 1, Article 4, 2021. 15f.

COPELAND, B.J. **Artificial intelligence**. Encyclopedia Britannica, 18 Mar. 2022. <https://www.britannica.com/technology/artificial-intelligence>.

DURANTI, Luciana; ABDUL-MAGEED, Muhammad; HOFMAN, Darra Hofman; SULLIVAN, Peter. I Trust AI, el nuevo proyecto de investigación de InterPARES. *School of Archival Yearbook*. n. 13. 2021. p. 36-55.

GUPTA, Abhishek; KAPOOR, Nikitasha. **Comprehensiveness of archives: a modern AI-enabled approach to build comprehensive shared cultural heritage**. In: Workshop Datafication and Cultural Heritage – Provocations, Threats, and Design Opportunities. 2020. Siegen, Germany.

HASHIGUCHI, Oliveira, SLAWOMIRSKI, T., L.; ODERKIRK J. **Laying the foundations for artificial intelligence in health**. OECD Health Working Papers, n.128, OECD, 2021.

IBM CLOUD EDUCATION. **O que é inteligência artificial?** [online]. 2020. <https://www.ibm.com/br-pt/cloud/learn/what-is-artificial-intelligence#toc-aplicativo-LCOINXXr>.

JAILLANT, Lise. **Introduction**. In: JAILLANT, Lise (Org). Archives, access and artificial intelligence: working with born-digital and digitized archival collections. Bielefeld University Press. 2022. 225p.

KESTEMONT, Mike; CHRISTLEIN, Vincent; STUTZMANN, Dominique. Artificial paleography: computational approaches to identifying script types in medieval manuscripts. *Speculum*. v. 92, n. S1, 2017.

KUSUMAWATI, Lisnanda; SALIM, Tamara Adriani. Artificial intelligence and knowledge management implementation in electronic archives preservation: literature review. *International Journal of Innovative Science and Research Technology*. v. 7, n. 5. 2022.

MINISTÉRIO PÚBLICO DO ESTADO DO RIO DE JANEIRO. **PGJ apresenta o "MPRJ em Mapas" e o projeto "Farol" ao presidente das Comissões de Controle Administrativo e Financeiro e de Enfrentamento da Corrupção do CNMP**. 2020.

NOLAN, A. **Artificial intelligence, its diffusion and uses in manufacturing**. Going Digital Toolkit Note. n. 12, 2021. https://goingdigital.oecd.org/data/notes/No12_ToolkitNote_AI&Manufacturing.pdf.

OECD. **OECD Business and finance outlook 2021: AI in business and finance**. OECD Publishing, Paris. 2021.

PEREIRA, Silvio do Lado. **Processamento de Linguagem Natural**. [s.d.] <https://www.ime.usp.br/~slago/IA-pln.pdf>.

ROLAN, G.; HUMPHRIES, G.; JEFFREY, L.; SAMARAS, E.; ANTISOPOVA, T.; STUART, K. More human than human? Artificial intelligence in the archive. *Archives and Manuscripts*, 1–25. 2018.

ROUHIINEN, Lasse. **Inteligencia artificial: 101 cosas que debes saber hoy sobre nuestro futuro**. Alienta Editorial. 2018.

SAS. **Processamento de linguagem natural: o que é e qual sua importância?** [online]. https://www.sas.com/pt_br/insights/analytics/processamento-de-linguagem-natural.html. html.

Shang, Erbo; LIU, Xiaohua; WANG, Hailong; RONG, Yangfeng; LIU, Yuerong. **Research on the application of artificial intelligence and distributed parallel computing in archives classification**. IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 4th. 2019, pp. 1267-1271.

USAID. **Artificial intelligence action plan: charting the course for responsible AI in USAID programming**. 2022. https://www.usaid.gov/sites/default/files/documents/USAID_Artificial_Intelligence_Action_Plan.pdf.

VELLINO, André; ALBERTS, Inge. Assisting the appraisal of e-mail records with automatic classification. **Records Management Journal**. 26 (3): 293–313. 2016.