

10, 11 e 12 de novembro de 2025

POLITÉCNICO DO PORTO / ISCAP
PORTO - PORTUGAL

APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL EM PROCESSOS DE PRESERVAÇÃO DE ARQUIVOS DIGITAIS

Gigliolla de Lourdes Batista Moura, Universidade Federal de Pernambuco,
<https://orcid.org/0000-0001-9343-8412>, Brasil, gigliolla.moura@ufpe.br

Sandra de Albuquerque Siebra, Universidade Federal de Pernambuco,
<https://orcid.org/0000-0002-0078-6918>, Brasil, sandra.siebra@ufpe.br

Eixo: Impacto das Tecnologias e Informação e Comunicação

1 Introdução

A pandemia de COVID-19, declarada pela Organização Mundial da Saúde (OMS), em março de 2020, ocasionou medidas como a necessidade de isolamento social, fazendo com que empresas, instituições e organizações, em escala global, precisassem acelerar seu processo de transformação digital, a fim de sobreviver a esse contexto. O que provocou a intensificação tanto a produção de arquivos digitais, quanto a realização de processos de digitalização (Nitahara, 2021).

Neste cenário, o problema da gestão e preservação de arquivos em meio digital se tornou um desafio ainda mais latente em empresas, instituições e organizações, a fim de evitar a perda de dados e garantir o acesso a eles a longo prazo (Lira & Siebra, 2021; Formenton, Castro, Gracioso, Furnival, & Simões, 2017). Especialmente porque, de acordo com Gouveia (2021), os processos relacionados à gestão de arquivos em meio digital recebem um impacto profundo das tecnologias e de sua evolução. Com isso, há sempre o risco da obsolescência tecnológica inerente ao hardware, ao software e ao próprio formato dos arquivos digitais, o que requer a aplicação de estratégias de preservação digital (Formeton & Gracioso, 2020)(Ferreira, 2006). Nesse cenário, a inteligência artificial (IA) emerge como ferramenta que pode contribuir com as iniciativas de preservação digital. A IA é

“um ramo da ciência da computação que se propõe a desenvolver sistemas que simulem a capacidade humana de percepção de um problema, identificando seus componentes para, com isso, resolver problemas e propor/tomar decisões” (Lobo, 2017, p. 188). E ela pode possibilitar “a utilização e desenvolvimento de sistemas informáticos capazes de aprender e adaptar-se sem seguir instruções explícitas, utilizando algoritmos e modelos estatísticos para analisar e extrair inferências a partir de padrões nos dados” (Gouveia, 2021, p. 36).

O uso da IA tem potencial para repercutir na mediação entre homem e máquina. De fato, os modelos de Inteligência Artificial podem desenvolver comportamentos similares à inteligência humana, à medida que são alimentados com dados, que remetam ao desempenho desejado. O que é destacado por Neves (2020, p. 187), quando pontua que “o uso de dispositivos inteligentes, IA e da computação cognitiva nos espaços de construção do conhecimento vem avançando paulatinamente, proporcionando novas formas de interação com os sujeitos”.

Também, o uso de IA tem potencial para otimizar os procedimentos envolvidos nas práticas de preservação digital, como endossado por Teel (2024), quando afirma que há uma tendência emergente de incorporar a tecnologia de IA nas práticas de preservação.

De acordo com Teel (2024), algoritmos de IA podem auxiliar na categorização automática, etiquetagem e criação de metadados para conteúdos digitais, facilitando a gestão e a recuperação de informações. Modelos de aprendizado de máquina podem contribuir para a identificação e mitigação da deterioração ou obsolescência digital e atuar como ferramentas para compreender, localizar, resumir e sintetizar o conhecimento preservado.

Colavizza et al. (2021, p. 14) destaca que “em termos de organização, a IA ajuda os arquivos a superar as limitações das unidades arquivísticas tradicionais, mantendo ao mesmo tempo os seus pontos fortes, por exemplo, em termos de proveniência”. Os modelos e ferramentas de IA são técnicas que auxiliam o arquivo e o arquivista, assim como os demais profissionais da informação que lidam com massas documentais, porquanto, “a integração da IA muda os fundamentos teóricos da prática arquivística” (Colavizza et al. 2021, p. 14).

Em decorrência, os profissionais de arquivo precisam estar atentos às mudanças tecnológicas, pois trazem desafios, mas também oportunidades de melhorias para a gestão, organização, preservação e acesso a longo prazo de documentos arquivísticos digitais. Pois a incorporação de novas tecnologias, como a IA¹ no contexto arquivístico requerem atualização de conhecimentos, capacitação periódica e desenvolvimento de novas competências.

Isso é essencial para que, não só o arquivista, mas todos os profissionais da informação conheçam melhor o ferramental que pode ser utilizado nas suas atividades cotidianas. Assim, endossamos o posto por Neves (2020, p. 187) “que toda reflexão é bem-vinda quando se trata de novas tecnologias no contexto social”. Neste cenário, esta pesquisa tem como objetivo apresentar aplicações da inteligência artificial (IA) em processos relacionados à preservação de arquivos digitais, com base na produção científica encontrada na base de dados SCOPUS (Elsevier).

Este artigo está dividido em sessões, as quais são: “sessão 2, “A inteligência artificial no contexto informacional e na preservação de

arquivos digitais”, a qual compõe boa parte da literatura desta pesquisa, apresentando os conceitos de IA e suas aplicações como ferramentas tecnológicas e explana a temática deste estudo; sessão 3, “Procedimentos Metodológicos”, descreve a metodologia utilizada para elaboração desta pesquisa; sessão 4, intitulada “Panorama Bibliométrico da Produção Analisada” apresenta os resultados obtidos a partir da seleção de artigos identificados na coleta de dados, pontuando as publicações por ano, os tipos de autoria e co-autoria e as publicações por países, a partir de uma elucidação dos autores com mais artigos publicados e dos periódicos que tiveram mais destaque, neste estudo. A sessão “Considerações Finais” destaca a importância da utilização das ferramentas de Inteligência Artificial, apontando a relevância dessas tecnologias aplicadas em arquivos, para gerir e aperfeiçoar processos de gestão documental, com o objetivo de proporcionar preservação a longo e acesso aos documentos digitais.

Espera-se que este artigo colabore nas reflexões sobre as possíveis aplicações da IA no contexto das práticas da Ciência da Informação.

2 A Inteligência Artificial no Contexto Informacional e na Preservação de Arquivos Digitais

A Inteligência Artificial pode trazer contribuições para o aprimoramento das funções arquivísticas, em especial, este estudo foca em contribuições que possam agilizar ou aprimorar processos relacionados à preservação de arquivos digitais ou que apresentem contribuições no contexto de unidades de informação. Jaillant e Rees (2022, p. 573, tradução nossa) coadunam com esta perspectiva, afirmando que consideram “a IA como um conjunto abrangente de abordagens computacionais utilizadas para tomar decisões e executar tarefas”. Os autores também alegam que, o desempenho das tecnologias de IA em tornar arquivos nato-digitais acessíveis e úteis é apontado como uma importante estratégia para gestores e pesquisadores

interessados nos usos de materiais arquivísticos (Jaillant & Rees, 2022, tradução nossa).

A IA possibilita a reprogramação do modo como a ciência e os pesquisadores operam os **processos de digitalização de acervos e de conjuntos documentais de arquivos**. Pode possibilitar o **reconhecimento de texto, após a digitalização**, tal como é feito na plataforma Transkribus ou permitir a criação de registros auto-descritivos.

Neste sentido, a plataforma *Transkribus*, criada pela READ-COOP, uma cooperativa “baseada na diretiva da UE sobre a Sociedade Cooperativa Europeia (SCE)” (READ-COOP, 2024), “foi desenvolvida no âmbito do projeto “READ” da UE Horizonte 2020 por um consórcio de grupos de investigação líderes de toda a Europa, liderado pela Universidade de Innsbruck” (READ-COOP, 2024). Esta plataforma opera com o reconhecimento de texto de documentos pertencentes à era medieval, através da IA. “Os mecanismos HTR processam a imagem inteira de uma palavra ou linha, digitalizando-a em várias direções e, em seguida, colocando esses dados em uma sequência” (Seaward & Kallio, 2017, p. 1).

As autoras complementam que “os motores HTR são baseados em algoritmos de aprendizado de máquina. A tecnologia precisa ser treinada sendo mostrado exemplos de pelo menos 30 páginas de material transcrito. Isso ajuda a entender os padrões que compõem palavras e os caracteres” (Seaward; Kallio, 2017, p. 2).

Outro exemplo de transcrição de documentos medievais, através de reconhecimento de texto manuscrito, é o projeto *In Codice Ratio*, que visa desenvolver novos métodos e ferramentas para conduzir estudos baseados em dados sobre grandes fontes históricas. O projeto é direcionado para as coleções do Arquivo Secreto do Vaticano (*In Codice Ratio*, 2025). Ammirati et al. (2019), afirmam que “a maioria desses documentos é digitalizada como imagens, a transcrição automática é um facilitador do aplicativo de técnicas de descoberta de conhecimento, bem como uma ferramenta útil para o paleógrafo por acelerar

o processo de transcrição” (2019, p. 185, tradução nossa).

Nesse sentido, Jaillant et al (2022, p. 288, tradução nossa), enfatiza que “uma vez que o problema do acesso aos arquivos foi superado, novas ferramentas computacionais podem ser desenvolvidas para ajudar a explorar esses arquivos ou completar a digitalização por conta própria. Jaillant e Rees (2022), reiteram que, a classificação automática de documentos governamentais antes da disponibilização de acesso à informação ou transferência para arquivos nacionais, assim como, experimentos para aperfeiçoamento de descobertas de coleções arquivísticas visuais e ferramentas de busca algorítmica, são alguns tipos de ferramentas de Inteligência Artificial para ampliação de acesso e uso a arquivos digitais. Os autores também ressaltam que, “além disso, redes acadêmicas e de pesquisa, como AURA, AEOLIAN e AI4LAM, reúnem diversos agentes interessados na aplicação da IA a acervos arquivísticos” (Jaillant & Rees, 2022, p. 573).

Desta forma, a disponibilidade das ferramentas de IA para acesso e usabilidade tornam possíveis as aplicações desses métodos em documentos arquivísticos e unidades de informação, como veremos a seguir nesta pesquisa, exemplos de uso de IA, distribuídos nos Quadros. Gouveia (2021) pontua as possibilidades de uso da IA no **contexto da organização e acesso a arquivos**, conforme resumido no Quadro 1.

Quadro 1 – Uso da IA para Organização e Acesso à Arquivos

Organização e acesso a arquivos
A quantidade crescente de materiais de arquivo digitalizados e de base digital exige a automatização e mesmo a reinvenção da organização e do acesso aos registros. A abordagem tradicional pode agora ser complementada por outros meios, centrados principalmente no conteúdo dos registros. Esta mudança não só permite aos utilizadores dar sentido a materiais de arquivo em escalas e complexidade maiores, como o fazer de formas inovadoras.
• Extração e indexação automática de conteúdos
• Registros de arquivo para leitura/aquisição de modo remoto
• Pesquisa e recuperação

Fonte: Adaptado de Gouveia (2021).

Martins (2010) destaca que a IA pode ser empregada para a **classificação automática de conteúdos**, de forma lógica e semanticamente coerente, o que pode facilitar a recuperação da informação. Ou seja, pode auxiliar com a organização dos arquivos, contribuindo para automação de sistemas de informação arquivística, preservação digital dos documentos e acesso. Com relação às atividades do campo da organização da informação, Rockembach (2021) explana que,

as questões da organização do conhecimento, das classificações, taxonomias e ontologias, encontram-se diretamente relacionadas as aplicações de aprendizagem de máquina, ou *Machine Learning*, com o uso de modelos que podem aprender com os dados, sendo que esta aprendizagem pode acontecer de forma supervisionada, semi-supervisionada e não supervisionada (Rockembach, 2021, p. 239).

As ferramentas de IA podem **simplificar e otimizar os processos de classificação, organização e análise arquivística**, atividades estas mapeadas no Quadro 2, a partir de práticas que possibilitem a utilização de “novas ferramentas de mediação homem máquina, como também novos métodos na utilização e análise de arquivos como o uso e exploração de (ciência dos dados) (Gouveia, 2021, p. 36).

Quadro 2 – Uso de IA na Automatização de Processos

Automatizar processos e decisões de manutenção de registros
O potencial de tecnologias de IA promete automatizar partes do fluxo de trabalho dos arquivos. As expectativas relacionadas com este potencial estão predominantemente situadas na Captura e Diversidade dos conteúdos associados. A curadoria manual, de alta qualidade e colaborativa de dados continua a ser uma abordagem bem-sucedida em muitas situações.
Avaliação: os profissionais de manutenção de registros são confrontados com volumes cada vez maiores de registros não estruturados e não categorizados. A IA tem potencial para ajudar a gerir volumes tão grandes e a literatura reporta inúmeras experiências e casos de sucesso.
Tratamento de informação sensível: as instituições devem evitar que a incapacidade de identificar informação sensível torne virtualmente impossível

conceder liberdade de acesso e pedidos de informação ou de permitir a investigação arquivística.

Metadados: questões de descrição e extração de metadados para auxiliar os criadores de registos do fornecimento de metadados(classificação), aumentando a qualidade dos metadados existentes (catálogos).

Fonte: Adaptado de Gouveia (2021).

A IA também pode ser **utilizada para a recolha de dados em fontes diversas**, dando origem a novas formas de arquivo (Gouveia, 2021), conforme descrito no Quadro 3.

Quadro 3 – Arquivos digitais em novos formatos

Novas formas de arquivos digitais
Facilitar novos tipos de recolha de dados, como nos sistemas integrados, sistemas inteligentes, com recurso à Internet das Coisas (IoT). A informação resultante é capturada e organizada, de acordo com a conformidade a leis, regulamentos e expectativas relativamente à privacidade e outros valores públicos. Os princípios de manutenção de registos, geralmente aceites para arquivo (responsabilidade, transparência, integridade, proteção, conformidade, disponibilidade, retenção e disposição), fornecem um referencial para a organização destas novas formas de arquivo digital.

Fonte: Adaptado de Gouveia (2021).

Neves (2020), com base na análise da literatura científica internacional, pontua algumas das perspectivas de uso da IA nas unidades de informação que vem transformando a forma de trabalho dos profissionais da informação. A autora destaca as possibilidades de “melhoria do processamento de informações, o apoio à pesquisa, técnicas do aprendizado de máquina incorporados aos sistemas de automação” (Neves, 2020, p. 200). Ela também menciona o uso que pode ser feito da IA na **criação de assistentes inteligentes, que podem ser utilizados na interação dos usuários** com as unidades de informação. O que já era pontuado por Bohle (2018) e Williams (2019) quando destacaram que as técnicas e recursos de IA, como a computação cognitiva² podem servir de apoio à pesquisa dos usuários e como formas alternativas de interação automatizada. *O National Archives and Records Administration (NARA) dos Estados Unidos* vem se empenhando em produzir uma solução fundamentada em Inteligência Artificial e aprendizagem de máquina, para **disponibilização de documentos digitalizados e nato-digitais** em seu Catálogo, a partir do

registro de metadados, fazendo uso de **pesquisa personalizada** (vide Quadro 4) e **registros auto descritivos** (vide Quadro 5) (NARA, 2021).

Quadro 4 – Uso de IA/Aprendizagem de Máquina para Pesquisa Personalizada

Fonte: Adaptado de NARA, Arquivos Nacionais Catálogo (2021)

É importante mencionar que os documentos precisam de sistemas que organizem os metadados dos registros pós digitalização. Jaillant e Caputo (2022), declaram que, “esses arquivos de metadados apresentam informações contextuais, incluindo a data em que o item foi digitalizado, os dados técnicos, material usado para capturar as imagens e o método usado para transcrições (Jaillant & Caputo, 2022, p. 828, tradução nossa)”.

Quadro 5 – Uso de IA/Aprendizagem de Máquina para criação de Registros/Metadados Auto descritivos

Registros Autodescritivos
<ul style="list-style-type: none"> As agências transferem registros digitalizados e eletrônicos, para os quais os arquivistas atualmente precisam descrever manualmente para o Catálogo
<ul style="list-style-type: none"> O NARA busca explorar o processamento de objetos de arquivo brutos, como PDFs, imagens e e-mails (também conhecidos como “objetos digitais”) para automaticamente produzir metadados descritivos para acesso público com o mínimo de intervenção arquivista

Fonte: Adaptado de NARA, Arquivos Nacionais Catálogo (2021)

O NARA apresenta alguns casos de uso de IA e aprendizagem de máquina, conforme especificado no Quadro 6.

Quadro 6 – Abordagem utilizada para os usos de caso para IA/Aprendizagem de Máquina

<ul style="list-style-type: none"> Identificar e resolver problemas de qualidade de dados, como preconceitos
<ul style="list-style-type: none"> Anonimização versus personalização da pesquisa fácil de usar
<ul style="list-style-type: none"> Algoritmos, estruturas e ferramentas para criar soluções de IA para os casos de uso acima, indicando o resultado, abordando os pontos fortes e as limitações.
<ul style="list-style-type: none"> Comparação de recursos fornecidos por diferentes provedores de nuvem, como AWS, Azure, IBM, Google etc.
<ul style="list-style-type: none"> COTS ou SaaS/PaaS/IaaS baseado em nuvem versus ferramentas não baseadas em nuvem ou estruturas.

Custos de licenciamento relacionados, capacidades operacionais e requisitos apoiar.
<ul style="list-style-type: none"> Capacidades de armazenamento e indexação independentes de formato/tipo de dados.
<ul style="list-style-type: none"> Projeto de pipeline para desenvolvimento e entrega de solução de IA/ML para produção
<ul style="list-style-type: none"> Atividade de pós-produção (atividade operacional) necessária, como infraestrutura monitoramento, depuração, orquestração de trabalhos, etc.
<ul style="list-style-type: none"> Áreas para considerações de custos

Fonte: Adaptado de NARA (2021)

Do Quadro 6 destaca-se que, quanto aos documentos digitalizados, o NARA especifica que “está usando IA para **encontrar e remover automaticamente informações pessoais sensíveis** de seus registros históricos digitalizados, ou seja, realizar a anonimização de dados. Isso tornará o processo mais rápido, mais preciso e protegerá melhor a privacidade, ao mesmo tempo em que permitirá maior acesso público a esses registros” (NARA, 2025).

Pesquisa Personalizada
<ul style="list-style-type: none"> À medida que cresce o volume de conteúdo disponível para pesquisa em nosso Catálogo, torna-se cada vez mais difícil para os usuários encontrarem o que procuram
<ul style="list-style-type: none"> A NARA procura explorar o fornecimento aos usuários dos Arquivos Nacionais Catálogo com uma experiência de pesquisa personalizada para fornecer informações relevantes resultados de pesquisa e faz recomendações para registros relacionados com base sobre o comportamento e interesses do usuário

Jaillant (2022, p. 433, tradução nossa, grifo nosso), aponta que,

Outro uso de IA/ML em coleções de arquivos é prever tipos de arquivos. Santhilata Venkata (Especialista em Preservação Digital, The National Archives UK) observa que sua instituição geralmente recebe registros de dados corrompidos, que são arquivos de texto. A Inteligência artificial pode examinar automaticamente o conteúdo do arquivo e, em seguida, fazer previsões em tipos de arquivos – o que é útil no contexto de preservação e acesso digital (Venkata 2021). Melhorar os metadados e prever os tipos de arquivos são, obviamente,

muito importantes, mas **talvez a aplicação mais importante de IA/ML em arquivos seja sua capacidade de identificar materiais sensíveis.**

Além das práticas e técnicas já citadas que contribuem para a preservação digital dos arquivos digitais, outros métodos e procedimentos através da IA estão sendo desenvolvidos e aplicados em acervos documentais, como no caso do Projeto *Sphaera*, do Instituto Max Plank de História da Ciência, na Alemanha. Este projeto possui um banco de dados de acesso aberto, que permite acesso a mais de 320 livros, dos períodos de 1472 à 1650, a estimativa é que demais obras sejam adicionadas à coleção.

O projeto tem a pretensão de percorrer o sistema científico entre os séculos XIII e XVII, quando as pesquisas eram mais centradas em conhecimentos cosmológicos, no intuito de compartilhar a identidade científica da Europa (*Center for Science and Society, Columbia University, 2024*). Os textos são digitalizados e interpretados através de um sistema computacional que recebe os dados dos conjuntos de documentos, “esses dados foram então colocados em um gráfico, o qual incorporou automaticamente essas conexões individuais em uma rede contendo todos os registros. O intuito era treinar um método de *machine learning* que conseguisse **sugerir conexões entre textos**” (MIT Technology Review, 2023).

No contexto digital, é necessário que sejam aplicados protocolos de segurança, mantendo os documentos em ambientes seguros, para que possam ser acessíveis. Nesse sentido, a Inteligência Artificial pode **contribuir para a segurança e manutenção dos dados, assegurando que a integridade e a veracidade das informações** estejam inalteráveis.

As práticas de IA estão em desenvolvimento em diversos seguimentos e áreas do conhecimento, incluindo a Ciência da Informação. A utilização dos métodos, técnicas e ferramentas de IA tem potencial para aprimorar a maneira como os arquivos digitais são gerenciados e preservados, além de possibilitar maior agilidade na recuperação da informação e na execução de processos (como

os de anonimização), garantindo qualidade arquivística aos documentos, mantendo sua integridade e autenticidade, contribuindo com a preservação digital e o acesso a longo prazo. Porém, Hernandez, Fewster e Penniman (2024), ressaltam que,

embora a IA e tecnologias semelhantes tenham amplas aplicações no domínio arquivístico, existem lacunas na transformação de experiências com IA em práticas e infraestruturas de arquivo duradouras. Nós discutimos que a colmatação destas lacunas virá de pelo menos três fontes: (1) profissionais de arquivo competentes o suficiente para integrar IA em seu fluxo de trabalho; (2) tecnologias comerciais de arquivamento e manutenção de registros que efetivamente incorporem IA em seus produtos; e (3) pesquisadores que contribuem com estudos relevantes e críticos para a ética e co-modelação eficaz da IA e da ciência e prática arquivística (Hernandez, Fewster, & Penniman 2024, p. 36, tradução nossa).

Assim, endossamos que, nesse cenário de uso de tecnologias avançadas como a IA, é preciso ter profissionais qualificados e atualizados, o que exige atualização constante de conhecimentos; é preciso ter a infraestrutura necessária; e que se intensifiquem as pesquisas sobre o uso da IA, assim como sobre as questões morais e éticas de sua utilização, em âmbito acadêmico e profissional.

Lorenz e Konečný (2023, p. 338, tradução nossa), relatam que, “com a inserção da inteligência artificial nos arquivos e nas práticas curatoriais, a importância dos aspectos éticos da curadoria de dados — especialmente ao lidar com materiais sensíveis ou controversos — vem sendo cada vez mais reconhecida”. Nesses aspectos, estão incluídos a privacidade, consentimento e vieses, elementos os quais necessitam ser considerados, para tornar um arquivo confiável, respeitando assim os direitos da sociedade, a qual possui seus dados representados em arquivos (Lorenz & Konečný, 2023, tradução nossa).

Por fim, vale mencionar nas indicações de uso e práticas de IA o trabalho realizado pela Federação Internacional de Associações e Instituições de Bibliotecas (IFLA)³. Ela produz estudos e aplicações com foco em bibliotecas, mas também aborda usos e práticas para arquivos digitais. A IFLA também possui a Seção de Preservação e Conservação (PRESCONS), que trata sobre a disponibilidade à longo prazo das coleções de patrimônios documentais em formatos físico e digitais. A PRESCONS juntamente com outras organizações como a International Council on Archives (ICA), trabalham para conservar e preservar coleções patrimoniais.

Neste seguimento, o projeto InterPARES⁴ mantém em desenvolvimento pesquisas voltadas às tecnologias de IA e suas interpretações e aplicações em instituições de arquivo, discutindo desafios, riscos e benefícios da utilização responsável de ferramentas de IA. Pesquisas recentes realizadas nesse contexto apontam para possibilidades de **visualização de imagens de arquivo com Inteligência Artificial**, construção de **IA confiáveis em sistemas de gerenciamento de registros e arquivamento**, digitalização e IA para arquivos de documentos patrimoniais, dentre outras vertentes de aplicação. Consequente, os principais usos da IA encontrados na literatura foram, a visualização de imagens de arquivo com Inteligência Artificial e ferramentas de IA confiáveis para sistemas de gerenciamento de registros e arquivamento. Estes processos de Inteligência Artificial em arquivos viabiliza a usabilidade da informação e da organização do acervo em instituições de arquivo, servindo de auxílio para ajudar os profissionais da informação, no entanto, os mesmos precisam estar capacitados e preparados para trabalhar com as tecnologias que essas ferramentas exigem.

3 Procedimentos Metodológicos

Este estudo teve como base a abordagem bibliográfica, que segundo Prodanov e Freitas (2013, p. 54), é aquela “elaborada a partir de material já publicado, com o objetivo de colocar o pesquisador em contato direto com

todo material já escrito sobre o assunto da pesquisa”.

A pesquisa é exploratória, porque “tem como finalidade proporcionar mais informações sobre o assunto que vamos investigar, possibilitando sua definição e seu delineamento, isto é, facilitar a delimitação do tema da pesquisa” (Prodanov & Freitas, 2013, p. 51-52).

Possui natureza quantitativa e qualitativa. A pesquisa quantitativa “considera que tudo pode ser quantificável, o que significa traduzir em números opiniões e informações para classificá-las e analisá-las” (Prodanov & Freitas, 2013, p. 69). E a pesquisa qualitativa, “considera que há uma relação dinâmica entre o mundo real e o sujeito, isto é, um vínculo indissociável entre o mundo objetivo e a subjetividade do sujeito que não pode ser traduzido em números” (Prodanov & Freitas, 2013, p. 70).

A coleta de dados foi realizada na base de dados SCOPUS (Elsevier), a partir de um recorte usando os termos de busca “Inteligência Artificial”, “Arquivos Digitais”, “Arquivos em Meio Digital”, em português e inglês, no período de 2020 à 2024, a fim de considerar os últimos cinco anos de produções. Uma vez que, desde a pandemia de Covid-19, iniciada em 2020, tanto houve uma maior produção de arquivos em meio digital, quanto como se intensificaram as pesquisas relacionadas à IA e suas ferramentas (Chenrui et al., 2024). Ressalta-se que primeiramente se combinou o termo “Inteligência Artificial” com “documentos digitais”, porém, curiosamente, os resultados recuperados tinham mais relação com ações dentro de bibliotecas. Ao substituir por “arquivos digitais”, foram recuperadas mais ações no contexto de Arquivos.

Foram recuperados 48 (quarenta e oito) trabalhos sobre os termos utilizados. Destes, 18 (dezoito) focavam em outras temáticas relacionadas a Biblioteconomia, Museologia, Internet das Coisas (IoT), Código Aberto, Infraestruturas de Dados, Gestão de Qualidade, entre outras, sem descrever necessariamente usos e técnicas de IA aplicadas a arquivos digitais para contribuir nos processos de preservação digital.

Adicionalmente, havia uma publicação do ano de 2021 que possuía falhas em sua legitimidade e precisou ser retratado pela instituição que o publicou, desta forma e, por este motivo, também foi retirado desta pesquisa, totalizando 29 (vinte e nove) artigos completos em periódicos ou eventos, a serem analisados.

Os artigos foram lidos na íntegra para proporcionar o alcance do objetivo, como também foram analisados utilizando o método bibliométrico, para que fosse possível fazer a mensuração e dar um panorama da produção científica sobre Inteligência Artificial e arquivos digitais analisada. Estes resultados deram origem tanto ao que foi apresentado na seção 2, como ao que será apresentado na seção 4.

Ressalta-se que, para complementação dos usos e técnicas de IA identificados no recorte de produção analisado, foram também lidos e utilizados, livros, artigos que faziam parte das referências dos artigos analisados e sites específicos dos projetos citados nos artigos.

4 Panorama Bibliométrico da Produção Analisada

Para dar um panorama das 29 (vinte e nove) publicações analisadas, no gráfico 1 é possível ver como elas se distribuíram por ano.

É possível observar que o gráfico 1 apresenta um crescimento das publicações, a partir do ano de 2022, pós-pandemia, constando de 11 (onze) publicações, sendo Jaillant, L. a autora que mais publicou nesse referido ano. Houve um declínio na produção no ano seguinte, 2023, constando de 7 (sete) artigos, respectivamente, para depois ocorrer um novo aumento das publicações no ano de 2024, com um total de 10 publicações. Percebe-se que, nesta base de dados, a temática praticamente não teve produções no período pandêmico (2020-2021). Sendo a única produção do período a intitulada *Eurasian Latin Archive*, do autor Carbé, E., do *Dipartimento di Filologia e Critica delle Letterature Antiche e Moderne*, da Universidade de Siena.

Colavizza et al. (2021) pontuam que, ao longo dos anos, as pesquisas relacionadas à IA aplicada aos processos de manutenção de

registros de arquivos tendiam a aumentar, o que foi observado nos resultados obtidos. Também, o quantitativo de produções, denotam os avanços nas pesquisas relacionando IA e arquivos digitais destes países, podendo ser citados como exemplo os estudos sobre a IA aplicada ao Arquivo Nacional dos Estados Unidos; os estudos na Europa sobre os avanços da plataforma *Transkribus*, relacionados à digitalização de acervos e reconhecimento de textos manuscritos; o trabalho do projeto *In Codice Ratio*, direcionado ao Arquivo Secreto do Vaticano; além das pesquisas realizadas no contexto dos projetos *Iron Mountain*, do Reino Unido e da rede AURA, do Reino Unido e Irlanda.

O gráfico 2 apresenta o tipo de co-autoria de publicação das produções analisadas.

É possível observar neste gráfico que a maioria dos trabalhos analisados neste estudo foram publicados em dupla (11 (onze) produções) ou de forma individual (9 (nove) produções). Tanto em dupla (com co-autores variados), quando individualmente, a autora Jaillant, L. se destaca como a mais produtiva, tendo três publicações em dupla e três individuais. Observou-se uma variedade de autores publicando sobre o tema, mas todos como autor individual ou em coautoria em apenas uma única publicação. Sendo as exceções apenas Jaillant, L. (com o total de seis publicações, como já mencionado) e Modiba, M. (com o total de duas publicações em coautoria).

Lise Jaillant é professora associada em Humanidades Digitais, na Universidade de Loughborough. A autora lidera desde 2020 quatro projetos com temáticas associadas à IA, dentre estes, o projeto rede AURA (Arquivos no Reino Unido/República da Irlanda e IA). Suas 3 publicações individuais, são: *Introduction to the Special Issue: Using Visual AI Applied to Digital Archives*; *More Data, Less Process: A User-Centered Approach to Email and Born-Digital Archives*; e *How can we make born-digital and digitised archives more accessible? Identifying obstacles and solutions*.

Suas 3 publicações em dupla, são uma com Aske, K., intitulada: *Are Users of Digital Archives*

Ready for the AI Era? Obstacles to the Application of Computational Research Methods and New Opportunities; uma com o autor Caputo, A, com o título de: *Unlocking digital archives: cross-disciplinary perspectives on AI and born-digital data* e, por fim, uma com Rees, A, intitulada: *Applying AI to digital archives: trust, collaboration and shared professional ethics*.

O autor Mashilo Modiba atua como professor no Departamento de Ciência da Informação da Universidade da África do Sul e aparece em duas publicações em coautoria. Uma com Shekgola, M., denominada: *Artificial Intelligence Embedded Cloud Computing Technology for the Management of Digital Archives in the Fifth Industrial Revolution in South Africa*; e outra com Schellnack-Kelly, I., intitulada: *Developing smart archives in society 5.0: Leveraging artificial intelligence for managing audiovisual archives in Africa*.

A maioria dos periódicos onde as publicações ocorreram são originários dos Estados Unidos da América, como pode ser observado no Gráfico 3. Eles tiveram um quantitativo de 6 produções científicas, distribuídas nos seguintes periódicos americanos: *Journal on Computing and Cultural Heritage*, *Journal of Social Archaeology*, *American Archivist*, *Archival Science*, *Digital Humanities Quarterly*, *Information Development*.

O segundo país com mais produções foi a Inglaterra, totalizando 4 (quatro) publicações, dispostas nos periódicos britânicos: *AI and Society*, *Digital Scholarship in the Humanities*, *Asian Journal of Law and Society* e na *Conferência Internacional ICENIT 2022*. Adicionalmente observa-se que houve publicação em periódicos de quatro continentes: Europeu, Americano, Asiático e Africano.

Observou-se nos achados que os periódicos que mais tiveram publicações sobre a temática pesquisada foram: *AI and Society*, *Journal on Computing and Cultural Heritage* (JOCCH) e *Archival Science*, com duas publicações cada. Diversos outros periódicos foram encontrados com apenas uma publicação cada.

O periódico *AI and Society*, “se concentra em questões sociais, incluindo o design, uso,

gerenciamento e política de informação, comunicações e novas tecnologias de mídia, com ênfase particular em implicações culturais, sociais, cognitivas, econômicas, éticas e filosóficas” (*AI and Society*, 2025). O periódico possui abordagem voltada às tecnologias digitais e suas vertentes, e é uma revista do grupo *Springer Nature*, com alta classificação.

A revista *Journal on Computing and Cultural Heritage* (JOCCH), têm suas publicações voltadas para a área de tecnologias de informação e comunicação (TIC). Abrange subáreas distintas como o patrimônio tangível, como documentação, organização, interpretação e comunicação de artefatos e coleções (incluindo arquivos digitais, catálogos e bibliotecas); e patrimônio intangível, como música, performance, narrativa e mitologia (ACM, 2025). O JOCCH é indexado na base de dados SCOPUS (Elsevier) e na Web of Science, sendo abordado nas principais coleções desta última.

Por fim, a *Archival Science* é um periódico que “promove o desenvolvimento da ciência arquivística como uma disciplina científica autônoma. Ele abrange todos os aspectos da teoria, metodologia e prática da ciência arquivística. Possui abordagem integrada, interdisciplinar e intercultural (*Archival Science*, 2025). O periódico também pertence ao grupo Springer Nature.

Uma curiosidade é que a autora mais produtiva, Lise Jaillant, tem publicações nestes três periódicos que se destacaram.

Com relação às palavras chaves empregadas pelos autores dos artigos analisados, nota-se uma grande ramificação de palavras utilizadas uma única vez. O termo “Artificial Intelligence” é o mais utilizado aparecendo em todas as 29 (vinte e nove) produções.

O termo “Digital Archives” é colocado nas palavras chaves de 19 (dezenove) produções, sendo o segundo mais citado, por fazer parte do contexto onde a IA é aplicada.

Em terceiro estão 3 (três) termos empatados com cinco ocorrências: *Machine Learning*, *Digital Libraries* e *Deep Learning*.

A aprendizagem de máquina ou *Machine Learning* (ML) pode ser considerada um

subconjunto da IA (Neves, 2022) e se caracteriza como o uso de “[...] programas de computador e algoritmos que podem extrair/derivar significado e padrões de dados” (Bourg apud Boman, 2019). A autora afirma que, “o aprendizado de máquina é mais ou menos definido como um subconjunto de IA” (Neves 2022, p. 28). O aumento do acesso a esses termos nos últimos anos sinaliza o interesse de indivíduos e instituições na aplicação e desenvolvimento de novos métodos e técnicas, que podem influenciar bastante e consideravelmente no processo de atividades de rotina. Algumas das experiências e usos relatados nos artigos fazem uso da ML, por isso, ela encontra-se entre as palavras chaves.

Bibliotecas digitais ou Digital Libraries, está entre as mais citadas por ser um dos ambientes onde as técnicas de IA podem ser aplicadas. E, finalmente, a aprendizagem profunda ou Deep Learning, que é tida como uma subárea do ML, e segundo Bengio et al. (2017) refere-se ao uso do ML em uma grande quantidade de dados para, após várias etapas de processamento com algoritmos, conseguir um computador aprenda por si mesmo e possa executar tarefas de forma semelhante aos seres humanos, como, por exemplo, a identificação de imagens.

5 Considerações Finais

A utilização dos métodos de inteligência artificial tem potencial para aprimorar a maneira como os documentos arquivísticos digitais são gerenciados e preservados, assim como aperfeiçoar subprocessos relacionados à estas atividades. Além de cooperar com a segurança, manutenção e anonimização de dados, assegurando que a integridade e a veracidade das informações estejam inalteráveis. Adicionalmente, a inteligência artificial pode contribuir na interação com os usuários e na mediação do acesso à informação por eles.

Vale destacar que ferramentas de IA também podem ser utilizadas para auxiliar na pesquisa científica, especialmente na parte de levantamento de dados e organização do

material coletado, o que pode agilizar o trabalho dos pesquisadores. Porém, para tudo isso, a aplicação de métodos e boas práticas informacionais é fundamental, além de um uso ético das ferramentas disponíveis.

Esta pesquisa revelou dados sobre aplicações de inteligência artificial em arquivos meio digital, abordando diversos ramos da Ciência da Informação e da Tecnologia, a partir de trabalhos encontrados na base de dados SCOPUS (Elsevier). Tais pesquisas foram produzidas em diversos países, constatando que os Estados Unidos da América foi o país que mais realizou estudos sobre a temática, sendo publicados em diferentes periódicos. No entanto, como limitação deste estudo, é importante ressaltar que os resultados refletem o conteúdo das produções consultadas e que o uso de outras bases de dados ou outros intervalos de tempo poderiam trazer resultados diversos.

Por fim, a inteligência artificial pode, entre outras funções, proporcionar o monitoramento de documentais digitais, por meio de atribuições de comando, programadas para execução de tarefas de gestão documental e análise de dados estruturados assegurando que a integridade e a veracidade das informações estejam inalteráveis.

Como trabalhos futuros pretende-se estender a discussão dos usos das ferramentas de inteligência artificial para novos desdobramentos nas pesquisas na área de arquivologia, aplicada a documentos digitais e digitalizados. Pretende-se trabalhar principalmente com a preservação de patrimônios digitais em instituições de arquivo, através de boas práticas de curadoria digital. O DCC (2025), alega que “a curadoria digital envolve manter, preservar e adicionar valor aos dados de pesquisa digital durante todo o seu ciclo de vida”, deste modo, mantendo os arquivos aplicados a uma gestão documental arquivística. “A gestão ativa de dados de pesquisa reduz as ameaças ao seu valor de longo prazo e mitiga o risco de obsolescência digital” (DCC, 2025). O DCC, também enfatiza que a curadoria digital atua como um processo de redução de “duplicação de esforços na criação de dados de pesquisa, a curadoria

aumenta o valor de longo prazo dos dados existentes, tornando-os disponíveis para futuras pesquisas de alta qualidade” (DCC, 2025), visando a preservação a longo prazo e o acesso.

6 Referências

- AI & Society: Knowledge, Culture and Communication. (n.d.). Springer. <https://link.springer.com/journal/146/aims-and-scope>.
- ACM Journal on Computing and Cultural Heritage (JOCCH). (n.d.). ACM Digital Library. <https://dl.acm.org/journal/jocch/about>.
- Ammirati, S., Merialdo, P., & Pascucci, A. (2019). In codice ratio: Machine transcription of medieval manuscripts. In M. Agosti, F. Esuli, & G. Faggioli (Eds.), *Digital libraries: Supporting open science*. 15th Italian Research Conference on Digital Libraries, IRCDL 2019, Pisa, Italy, January 31–February 1, 2019, Proceedings (Vol. 15, pp. 185–192). Springer International Publishing.
- Archival Science: International Journal on Recorded Information. (n.d.). Springer. <https://link.springer.com/journal/10502>.
- Bengio, Y., Goodfellow, I., & Courville, A. (2017). *Aprendizado profundo*. Cambridge, MA: MIT Press.
- Bohle, S. (2018). “plutchik”: Artificial intelligence chatbot for searching NCBI databases. *Journal of the Medical Library Association*, 106(4), 501–503.
- Boman, C. (2019). An exploration of machine learning in libraries. *Library Technology Reports*, 55(1).
- Colavizza, G., Beretta, D., Giannone, D., & Blanke, T. (2021). Archives and AI: An overview of current debates and future perspectives. *ACM Journal on Computing and Cultural Heritage (JOCCH)*, 15(1), 1–15. <https://dl.acm.org/doi/10.1145/3479010>.
- Digital Curation Centre (DCC). (2025). Porque uma boa pesquisa precisa de bons dados. Recuperado em 2025, de <https://www.dcc.ac.uk/about/digital-curation#:~:text=Curadoria%20digital%20%C3%A9%20o%20gerenciamento,futuras%20pesquisas%20de%20alta%20qualidade>.
- Ferreira, M. (2006). *Introdução à preservação digital: Conceitos, estratégias e actuais consensos* (88 p.). Escola de Engenharia da Universidade do Minho. <https://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>.
- Formenton, D., Costa, S. M. S., & Ferreira, C. E. R. (2017). Os padrões de metadados como recursos tecnológicos para a garantia da preservação digital. *Biblios*, (68), 82–95. <http://dx.doi.org/10.5195/biblios.2017.414>.
- Formenton, D., & Gracioso, L. de S. (2020). Preservação digital: Desafios, requisitos, estratégias e produção científica. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, 18(0), e020012. <https://doi.org/10.20396/rdbci.v18i0.8659259>.
- Gouveia, L. B. (2021). Explorar o potencial da inteligência artificial no contexto do arquivo. In V Colóquio REVIS-ARQ.
- Hernandez, R. A., Fewster, K., & Penniman, S. (2024). Artificial intelligence and machine learning competencies for the archival professions. *Proceedings of the Association for Information Science and Technology*, 61(1), 36–43. <https://doi.org/10.1002/pra2.1006>.
- In Codice Ratio. (n.d.). Università degli Studi Roma Tre. <https://www.inf.uniroma3.it/db/icr/index.html>
- International Federation of Library Associations and Institutions (IFLA). (n.d.). IFLA repository. Recuperado em 24 de março de 2025, de <https://repository.ifla.org/home>.
- International Federation of Library Associations and Institutions (IFLA). (n.d.). IFLA. Recuperado em 24 de março de 2025, de <https://www.ifla.org/>.
- InterPARES Trust AI. (n.d.). InterPARES Trust AI. <https://interparestrustai.org/>.
- Jaillant, L. (2022). How can we make born-digital and digitised archives more accessible? Identifying obstacles and solutions. *Archival Science*, 22(3), 417–436. <https://doi.org/10.1007/s10502-022-09390-7>.
- Jaillant, L., & Caputo, A. (2022). Unlocking digital archives: Cross-disciplinary perspectives on AI and born-digital data. *AI & Society*, 37(3), 823–

835. <https://doi.org/10.1007/s00146-021-01367-x>.
- Jaillant, L., & Rees, A. (2023). Applying AI to digital archives: trust, collaboration and shared professional ethics. *Digital Scholarship in the Humanities*, 38(2), 571–585. <https://doi.org/10.1093/dlsc/fqac073>.
- Jaillant, L., Aske, K., Goudarouli, E., & Kitcher, N. (2022). Introduction: Challenges and prospects of born-digital and digitized archives in the digital humanities. *Archival Science*, 22(3), 285–291. <https://doi.org/10.1007/s10502-022-09396-1>.
- Lira, J., & Siebra, S. A. (2021). Preservação digital: Revisitando o essencial. In S. A. Siebra & V. R. Borba (Orgs.), *Preservação digital e suas facetas* (pp. 31-83). São Carlos: Pedro & João Editores.
- Lobo, L. C. (2017). Inteligência artificial e medicina. *Revista Brasileira de Educação Médica*, 41, 185–193. <https://doi.org/10.1590/1981-52712015v41n2esp>.
- Lorenz, M., & Konečný, M. (2023). Digital archives as research infrastructure of the future. *Acta Informatica Pragensia*, 12(2), 327–341. <https://doi.org/10.18267/j.aip.219>.
- Martins, A. L. (2010). Potenciais aplicações da inteligência artificial na ciência da informação. *Informação & Informação*, 15(1), 1–16.
- MIT Technology Review. (2023, 28 abril). Como a inteligência artificial está ajudando os historiadores a entender melhor o nosso passado. MIT Technology Review Brasil. Recuperado em 30 de janeiro de 2024, de <https://mittechreview.com.br/como-a-inteligencia-artificial-esta-ajudando-os-historiadores-a-entender-melhor-o-nosso-passado/>
- National Archives and Records Administration (NARA). (2025). National Archives. Recuperado em 12 de janeiro de 2025, de <https://www.archives.gov/ai>.
- National Archives and Records Administration (NARA). (2021). National Archives and Records Administration: Artificial Intelligence (AI) / Machine Learning (ML) Virtual Information Day.
- Neves, B. C. (2020). Inteligência artificial e computação cognitiva em unidades de informação: Conceitos e experiências. *Logeion: Filosofia da Informação*, 7(1), 186–205.
- Neves, B. C. (2020). Inteligência artificial na ciência da informação: Introdução. (2022). Salvador, BA: Laboratório de Tecnologias Informacionais e Inclusão Sociodigital.
- Nitahara, A. (2021, 11 novembro). Estudo mostra que pandemia intensificou uso das tecnologias digitais. Agência Brasil. Recuperado em 10 de março de 2025, de <https://agenciabrasil.ebc.com.br/geral/noticia/2021-11/estudo-mostra-que-pandemia-intensificou-uso-das-tecnologias-digitais>.
- Prodanov, C. C., & Freitas, E. C. de. (2013). Metodologia do trabalho científico: Métodos e técnicas da pesquisa e do trabalho acadêmico (2ª ed.). Novo Hamburgo: Feevale.
- READ-COOP. (2024). Read-coop, Transkribus. Recuperado em 29 de janeiro de 2024, de <https://readcoop.eu/>.
- Rockembach, M. (2021). 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.
- Seaward, L., & Kallio, M. (2017). Translation: Handwritten text recognition technology for historical documents. In Proceedings of the Digital Humanities 2017 Conference. Recuperado em 30 de janeiro de 2024, de <https://dh2017.adho.org/abstracts/649/649.pdf>.
- Teel, Z. A. (2024). Artificial intelligence's role in digitally preserving historic archives. *Preservation, Digital Technology & Culture*, 53(1). <https://doi.org/10.1515/pdtc-2023-0050>.
- Williams, R. (2019). Assistentes de inteligência artificial na biblioteca: Siri, Alexa e beyond. Online Searcher, May/June 2019.

6.1 Notas

NOTAS

- ¹ Coaduna-se com o posto por Rockembach (2021, p. 238) que “como campo de pesquisa, a Inteligência Artificial não se trata de uma área nova, pois seu conceito remonta 1956, na conferência

Dartmouth Summer *Research Project on Artificial Intelligence*”, porém, sua aplicação nos processos de preservação digital faz parte de estudos mais recentes.

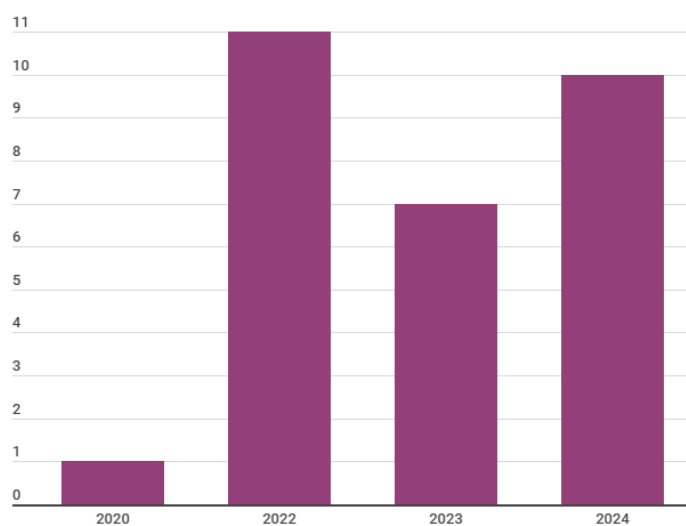
² É uma disciplina que integra conceitos da neurobiologia, da psicologia cognitiva, da ciência da informação e da inteligência artificial. Considerando a classificação da Associação Nacional de Pós-Graduação e Pesquisa em Educação (ANPED), o campo das discussões que envolvem tal temática dialoga com os elementos conceituais no eixo da Educação e Comunicação e de acordo com a Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (Ancib) no eixo que trata de Informação e Tecnologia.

³ A Federação é uma organização independente, internacional, não governamental, sem fins lucrativos e de benefício público que promove os interesses de bibliotecas e associações de informação, bibliotecas e serviços de informação, bibliotecários e as comunidades que atendem em todo o mundo (IFLA, 2025).

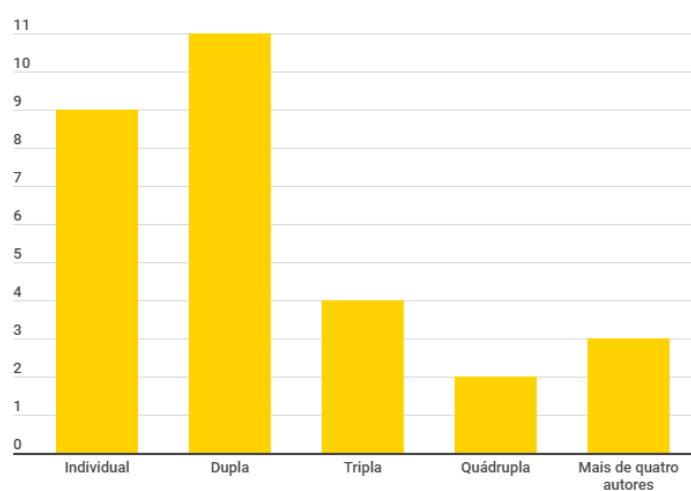
⁴ **O InterPARES Trust AI (2021-2026)** é um projeto interdisciplinar multinacional que visa projetar, desenvolver e alavancar a Inteligência Artificial para dar suporte à disponibilidade e acessibilidade contínuas de registros públicos confiáveis, formando uma parceria sustentável e contínua, produzindo pesquisas originais, treinando alunos e outros profissionais altamente qualificados (HQP) e gerando um círculo virtuoso entre a academia, instituições de arquivo, profissionais de registros governamentais e a indústria, um ciclo de feedback que reforça o conhecimento e as capacidades de cada parte (InterPARES, 2025).

6.2 Gráficos

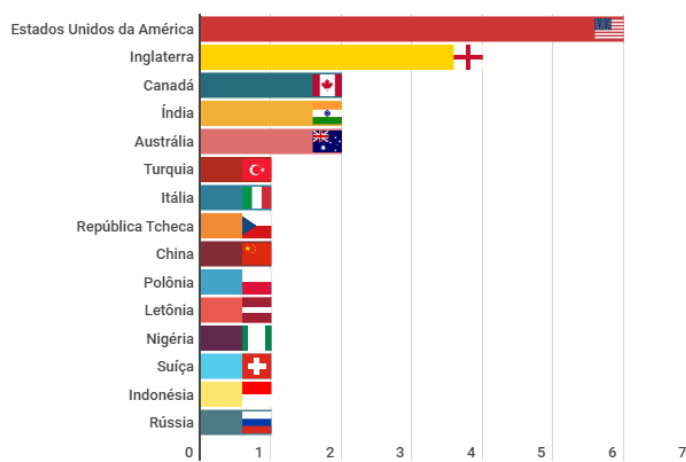
Destacam-se neste espaço os gráficos comentados no capítulo 4: Parâmetro bibliográfico da produção analisada.

Gráfico 1 – Publicações por ano

Fonte: Dados da pesquisa, 2025

Gráfico 2 – Tipos de autoria e co-autoria

Fonte: Dados da pesquisa, 2025

Gráfico 3 – Publicações por países

Fonte: Dados da pesquisa, 2025