

Resumo: Esta investigação tem como objetivo geral compreender como a Inteligência Artificial (IA) é utilizada nos serviços de informação arquivo. Como objetivos específicos pretende-se: caracterizar casos de aplicação da IA e *Machine Learning* (ML) em arquivos; analisar as vantagens e os desafios que advêm da aplicação de IA e ML em projetos desenvolvidos em arquivos ou com documentação de arquivos. Optou-se por uma investigação qualitativa com uma abordagem exploratória e descrita. A análise de casos concretizou-se mediante estudo de cinco projetos em arquivos europeus e aplicação de questionário por entrevista a profissionais da informação dos seguintes arquivos: Arquivo da Cidade de Amesterdão, Arquivo Fotográfico Finlandês do Tempo de Guerra, Arquivo Apostólico do Vaticano, Arquivo Arolsen e Arquivo Nacional do Reino Unido. Dos resultados obtidos destacam-se os benefícios na implementação de IA e ML nos casos de estudo.

Palavras-chave: Aprendizagem Automática; Arquivos; Inteligência Artificial; Serviços de Informação.

Abstract: The aim of this article is to understand how Artificial Intelligence (AI) is used in archives. The specific objectives are to characterise cases of the application of AI and Machine Learning (ML) in archives; to analyse the advantages and challenges arising from the application of AI and ML in projects developed in archives or involving archival documentation. A qualitative research design with an exploratory and descriptive approach was chosen. The case analysis was carried out through a study of five projects in European archives and the application of a questionnaire via interviews with information professionals from the following archives: Amsterdam City Archives, the Vatican Apostolic Archives, the Arolsen Archives, and the National Archives of the United Kingdom. The results highlight the benefits of implementing AI and ML in the case studies.

Keywords: Machine Learning; Archives; Artificial Intelligence; Information Services.

Introdução

A Inteligência Artificial (IA) tem-se vindo a desenvolver, tendo cada vez mais visibilidade e impacto no quotidiano dos indivíduos e das organizações. Utilizando algoritmos avançados, as tecnologias de IA podem analisar grandes quantidades de dados, extrair padrões e produzir resultados rapidamente. À medida que a IA generativa evolui, o seu potencial de impacto nos serviços, operações técnicas e experiências dos utilizadores de arquivos e de bibliotecas torna-se mais evidente.

A crescente produção de informação em formato digital tem trazido desafios significativos aos serviços de informação arquivo, exigindo novas abordagens para a gestão, preservação e acesso aos documentos/à informação. A IA e a Aprendizagem Automática/*Machine Learning* (ML) assumem um papel cada vez mais relevante e oferecem soluções para automatizar processos que tradicionalmente são morosos e que necessitam de recursos

humanos significativos. Nos arquivos, estas tecnologias têm vindo a ser aplicadas em tarefas como a transcrição de documentos manuscritos, a criação automática de metadados, na indexação e na melhoria de mecanismos de pesquisa e acesso à informação.

Este artigo resulta da dissertação de Mestrado em Ciência da Informação (CORREIA, 2023) realizada na Faculdade de Letras da Universidade de Coimbra, tendo como objetivo geral compreender como a IA é utilizada nos serviços de informação arquivo. Como objetivos específicos pretende-se: caraterizar casos de aplicação da IA e ML em arquivos; analisar as vantagens e os desafios que advêm da aplicação daquelas tecnologias em projetos desenvolvidos em arquivos ou com documentação de arquivos, considerando a perspetiva de profissionais da informação das instituições arquivísticas selecionadas. Através da análise de casos de estudo e de dados recolhidos junto de profissionais da área, procurou-se identificar as vantagens, os desafios e as implicações destas tecnologias na prática arquivística, bem como refletir sobre questões éticas e profissionais associadas à sua utilização.

Para a realização deste trabalho optou-se por uma investigação qualitativa com uma abordagem exploratória e descrita. Realizou-se a análise de casos mediante estudo de cinco projetos em arquivos europeus: Arquivo da Cidade de Amesterdão, Arquivo Fotográfico Finlandês do Tempo de Guerra, Arquivo Apostólico do Vaticano, Arquivo Arolsen – Centro Internacional de Perseguição Nazi e Arquivo Nacional do Reino Unido. Concretizou-se um questionário por entrevista a profissionais da informação dos referidos arquivos.

A integração da IA poderá ter um impacto significativo nos serviços de informação arquivo. No presente, procura-se aferir a perceção acerca desse impacto futuro em seu benefício e dos utilizadores.

A Inteligência Artificial e a Aprendizagem Automática em arquivos

A Ciência da Informação (GOMES, 2020) e a IA caminham, desde o seu surgimento, lado a lado.

AI and the information field (information science and information systems) have and an active, albeit cautious, relationship from the early stages of their history. In particular, information scientists found in AI a rich source of ideas for thinking about theoretical questions as well as a repertoire of tools and techniques for dealing with practical applications (EKBIA, 2012: 204).

Para melhor se compreender o impacto da IA nos serviços de informação arquivo é fundamental a revisão da literatura. Após pesquisa verifica-se que, apesar da temática já ser estudada há alguns anos, só recentemente é que tem tido mais visibilidade e, por isso, maior produção científica. É possível identificar a necessidade que a comunidade científica tem em compreender como pode ser aplicada a IA nos arquivos e quais os impactos que esta terá, tanto para os utilizadores como para os profissionais da informação (CORREIA, 2023).

Ao estudar a produção científica percebe-se que a literatura utiliza a língua inglesa e centra-se, essencialmente, em países como a China, os Países Baixos, o Reino Unido e os Estados Unidos da América (CORREIA, 2023). A temática dos trabalhos concentra-se, sobretudo,

em quatro grandes eixos: automatização da classificação e descrição arquivística, organização e acesso a coleções digitais, integração de *linked data* e ML em fluxos arquivísticos, e reflexão ética e profissional sobre a incorporação destas tecnologias nos sistemas de arquivo. Em comparação com a produção sobre bibliotecas (GOMES, FERNÁNDEZ MARCIAL e SANTOS, 2021), observa-se um volume menor de trabalhos.

Colavizza *et al.* (2021:41-42) referem que

long before big data an idea had been invented, archives already measured their collections in kilometers of files and folders", afirmando ainda que, por isso, "more and more archival collections are digitised and new born-digital records at ever larger scale are being submitted to archives. (...) Consequently, human archivists need the support of machine agents to assist them working through big data.

Shang *et al.* (2019: 1267) também mencionam a vertente do tratamento arquivístico e afirmam que archives classification, as the key and core of archives management, is the most important part of procedures of archives collection, management, utilization and destruction. No entanto, como a gestão de arquivos está a evoluir no sentido da informatização e da eliminação do papel, tornou-se um importante tema de investigação a classificação de arquivos através do uso de métodos baseados em IA e ML.

A acessibilidade ao conteúdo dos arquivos é um tema analisado e Jaillant (2022) considera que access to data is seen as a key priority today. Yet, the vast majority of digital cultural data preserved in archives is inaccessible due to privacy, copyright or technical issues. À semelhança de outros autores, Jaillant (2022) considera que a IA pode ser a resposta para este problema e afirma que IA e ML podem ser usadas para unlock archives and make them more findable and accessible – for example by differentiating between sensitive and non-sensitive materials, or by automatically adding tags and other metadata to improve findability (JAILLAN, 2022).

Em termos gerais, os autores concordam que a IA pode trazer inúmeras vantagens aos serviços de informação arquivo, facilitando e ajudando no trabalho dos profissionais e a oferecer uma melhor experiência de pesquisa e acesso aos seus utilizadores. No entanto, não deixam de mencionar os impactos negativos que a utilização da IA pode trazer. Cox, Pinfield e Rutter (2019) alertam para questões como a privacidade, a confiança que deve ser depositada nas decisões das máquinas, os possíveis custos desta tecnologia, os impactos que o uso da IA pode trazer a nível de empregos e igualdade social.

A produção científica sobre IA na área científica da Ciência da Informação (GOMES, 2022) e, em particular, em arquivos, evidencia um campo em consolidação, marcado por forte orientação aplicada e por uma preocupação crescente com o acesso, a organização e a preservação de acervos digitais. Nos artigos analisados, a IA surge menos como fim em si mesma e mais como instrumento de resolução de problemas estruturais dos sistemas de arquivo, nomeadamente a morosidade da classificação manual, a dificuldade de tratar grandes volumes de documentos digitais e a limitação de acesso a coleções nascidas digitais ou digitalizadas.

Shang *et al.* (2019) evidenciam a utilidade de algoritmos de aprendizagem automática para a classificação de arquivos, defendendo que a automação reduz custos e aumenta a precisão em contextos documentais complexos. Segundo os autores, pode-se ter maior precisão de

classificação do que com o método tradicional, o que aponta para ganhos de eficiência e precisão. A principal vantagem é a redução do tempo e do esforço humano; a limitação implícita é a necessidade de infraestrutura computacional e de bases de treino adequadas para produzir resultados consistentes.

Hawkins (2022) aborda os arquivos digitais e os *born-digital archives* a partir da perspectiva dos *linked data* e das Humanidades Digitais, defendendo a colaboração entre disciplinas para ampliar o acesso e a utilização de acervos digitalizados e nascidos digitais. O problema central é o enorme volume de dados arquivísticos digitais e a necessidade de criar representações legíveis por máquina que favoreçam a análise e a reutilização. A autora destaca o potencial da IA e de ferramentas como a Wikidata para escalar a produção de *archival linked data*. A vantagem reside na ampliação do acesso, da interoperabilidade e da reutilização; a desvantagem é que esse processo exige padronização, competências técnicas e uma integração crítica para evitar automatismos pouco refletidos. Na mesma linha, Jaillant (2022) discute o problema de acesso a arquivos digitais e a coleções *born-digital*, frequentemente inacessíveis devido a privacidade, direitos de autor e obstáculos técnicos. A investigação, baseada em entrevistas com profissionais de instituições culturais, evidencia que muitos acervos permanecem não catalogados, difíceis de localizar e inutilizáveis pelo público. A autora propõe o uso ético de IA para *unlock dark archives inaccessible to users*. Portanto, a IA pode apoiar a descoberta e a abertura de fundos documentais; a desvantagem é que surgem riscos éticos relevantes, sobretudo em matéria de privacidade, responsabilidade e transparência nas decisões.

Jaillant e Caputo (2022) aprofundam a dimensão ética, sublinhando que a IA só será útil aos arquivos se for acompanhada por princípios de *fairness, accountability and transparency*. Este ponto é crucial porque a automatização arquivística não é neutra: ela implica decisões sobre o que classificar, descrever, abrir ou manter fechado, e essas decisões podem reproduzir desigualdades se forem delegadas sem controlo crítico. Mödden (2022) reforça este enquadramento ao mostrar que a automatização da catalogação por assunto e da atribuição automática de classificações, em especial no contexto da Biblioteca Nacional Alemã, exige atenção à proveniência dos dados, à qualidade dos metadados e à fiabilidade dos processos classificatórios. O estudo é útil para a área dos arquivos por mostrar como a automação pode ser transferida para processos de organização documental.

Em síntese, a literatura sobre IA em arquivos centra-se na eficiência operacional, na ampliação do acesso e na gestão inteligente de grandes massas documentais, mas os ganhos técnicos são acompanhados por riscos éticos, técnicos e profissionais que exigem mediação humana permanente. O balanço geral é positivo, embora cauteloso, pois a IA aparece como ferramenta promissora para acelerar tarefas, estruturar acervos e desbloquear arquivos, mas a sua adoção deve ser regulada por critérios de transparência, responsabilidade e ética arquivística.

Metodologia

A metodologia adotada nesta investigação compreende uma abordagem qualitativa, mediante pesquisa exploratória incluindo, numa primeira fase, a análise e revisão de literatura e, numa segunda fase, a identificação, seleção e análise de estudos de caso.

Através da pesquisa avançada na Web of Science e no EBSCO Discovery Service, com recurso a operadores booleanos foram conjugados os termos *artificial intelligence* AND *archives* OR *archival collections* OR *archival fonds*. A pesquisa foi efetuada nos meses de outubro a dezembro de 2022, em português e inglês, com recurso aos campos assunto/tópico e título, organizada por data de publicação, no período cronológico de 2018 a 2022. Na fase de pesquisa bibliográfica foi, também, consultada informação nos *websites* de organismos internacionais de referência, como a International Federation of Library Associations and Institutions, o International Council on Archives e a United Nations Educational, Scientific and Cultural Organization.

Através do estudo de caso é possível estudar e refletir sobre o objeto de estudo num contexto real, utilizando várias fontes de evidência, permitindo abranger toda a complexidade de um caso específico que tem interesse sobre si mesmo: *el objetivo primordial del estudio de un caso, no es la comprensión de otros. La primera obligación es comprender este caso* (STAKE, 2020:17). O estudo de caso deve ser visto como uma estratégia metodológica a ser adotada quando os problemas e as circunstâncias da investigação assim o exigem. Segundo Yin (2001:32) trata-se de uma investigação empírica que "investiga um fenómeno contemporâneo dentro do seu contexto da vida real, especialmente quando os limites entre o fenómeno e o contexto não estão claramente definidos". Yin (2004:19) acrescenta ainda que "os estudos de caso representam a estratégia preferida quando se colocam questões do tipo 'como' e 'porquê', quando o investigador tem pouco controlo sobre os eventos e quando o foco se encontra em fenómenos contemporâneos inseridos em algum contexto da vida real". Yin (2001:32-33) ressalta que o estudo de caso enfrenta situações em que existem "muito mais variáveis de interesse do que pontos de dados, e, como resultado, baseia-se em várias fontes de evidências, com os dados precisando de convergir em um formato de triângulo, e, como outro resultado, beneficia-se do desenvolvimento prévio de proposições teóricas para conduzir a coleta e a análise de dados".

Nos estudos de caso podemos encontrar dois tipos de caso, o singular e o múltiplo. Para análise dos estudos de casos múltiplos eleitos nesta investigação, concretizou-se pesquisa bibliográfica e na World Wide Web (WWW) e, ainda, inquérito por entrevista, que se caracteriza como uma técnica cujo objetivo é a recolha de dados, sendo bastante utilizado em investigação na área da Educação e das Ciências Sociais e Humanas, "muitas vezes associado a estudos de carácter interpretativo e a planos de investigação de natureza qualitativa na recolha de dados ou informações, dado o carácter descritivo e pormenorizado dos mesmos (BATISTA *et al.*, 2021:15).

O estudo de casos e as técnicas de recolha de dados

A seleção dos casos baseou-se na pesquisa da literatura e na constatação de escassez de estudos sobre a aplicação da IA em arquivos. Esta limitação é também destacada no artigo *More human than human? Artificial intelligence in the archive* de Rolan *et al.* (2019), em que os autores reiteram que existe *a lack of compelling case studies - maybe this is the hypercycle effect, but while commentary abounds, there are not many real-world examples within the academic or professional literature* (p. 186); *the literature is rather thin on the ground, and there are few clear success stories being trumpeted* (p. 185).

Deste modo, optou-se pela análise de projetos europeus que recorreram a IA e a ML em contexto arquivístico. A seleção dos casos foi feita através de uma pesquisa inicial na WWW e, de seguida, foram consultados os respetivos *websites* das instituições. Após esta pesquisa inicial foi realizada uma pesquisa de artigos científicos referentes aos estudos/projetos com documentação custodiada pelas instituições arquivísticas selecionadas.

Seguidamente, procedeu-se à estruturação do inquérito por entrevista, cujo objetivo foi avaliar a perceção de profissionais da informação acerca do impacto que cada um dos projetos tem/teve nos serviços disponibilizados aos utilizadores e na atividade dos arquivistas de cada instituição. O carácter voluntário deste estudo significa que a amostra não é representativa nem deve ser generalizada. Optou-se por uma entrevista estruturada, pois a intenção era que um inquirido de cada instituição respondesse a um pequeno conjunto de questões. A seleção e formulação das questões teve por base os objetivos definidos para o estudo de casos e a revisão da literatura, tendo sido particularmente relevante a consulta do relatório realizado por Dagleish (2022).

O inquérito é constituído por cinco questões e foi enviado via *e-mail* no mês de maio de 2023. O pedido de resposta foi reiterado pela mesma via a três das instituições. As 5 questões elaboradas em inglês são as seguintes:

1. In your perspective, what are the advantages of using AI/ML in the selected documentation from the archive (name)?
2. What is the impact of this project on the services provided by the archive to users?
3. What is the impact of these AI and ML projects on the daily professional activity of archivists?
4. What are the advantages of using AI/ML in archives at present and in the near future?
5. What are the main challenges in the use of AI/ML in archives at present and in the near future?

Projetos em Arquivos Europeus

Arquivo da Cidade de Amesterdão

O Arquivo de Amesterdão preserva e disponibiliza documentação histórica, sendo também o detentor legal de informação do município, mantida de forma permanente. Ao longo de 750 anos, o Arquivo de Amesterdão reuniu um vasto acervo de documentos. *The City Archives is the historical documentation center of the City of Amsterdam with more the 50 kilometers of archives, a historical-topographical collection with millions of maps, drawings and prints, a library and extensive sound, film and photo archives* (AMSTERDAM..., 2025). O serviço gere *historical information on paper, but also, to a large extent, in the form of (historical) data* (AMSTERDAM..., 2025).

Em face da necessidade de os arquivistas e cidadãos acederem à informação contida nos documentos, em 2015 o arquivo estabeleceu parceria com o projeto READ (*Recognition and Enrichment of Archival Documents*) e o seu *Transkribus* para permitir a pesquisa e transcrição nos documentos históricos:

The online platform allows users to train a computer handwriting recognition model to transcribe historical documents written by hand in a variety of European languages. (...) Once trained, the model uses machine learning to compare the handwriting patterns it now knows with that of the documents the user wants to transcribe. The model automatically transcribes line by line. (...) After Transkribus has done its work, users often just need to proofread to correct any minor errors (BURKE, 2020).

O arquivo digitalizou um grande número de documentos notariais manuscritos, o que facilitou a integração de *Handwritten Text Recognition* (HTR) e a utilização de *Named Entity Recognition* (NER) para extrair entidades do texto fornecido. Uma entidade pode ser uma pessoa, uma organização, uma rua, um evento ou uma época. Este tipo de modelo de ML permite destacar os nomes que o utilizador procura e facilita o estabelecimento de relações entre documentos.

Pauline Van den Heudel says that the archive co-opted Transkribus into their work when they realised that indexing the names, places and dates in their 17th and 18th century documents would take decades of work (BURKE, 2020). Com os algoritmos da plataforma foi possível terminar as transcrições de documentos do século XVIII um ano antes do previsto. A equipa de 300 voluntários do arquivo verificava se existiam erros nas transcrições. Now you can research patterns in big amounts of data, connections between people – it's completely new research (BURKE, 2020). O projeto, quando concluído, estará ligado à rede European Time Machine, uma rede de instituições que utiliza registos para ajudar a compreender a evolução social e política da Europa ao longo do tempo (BURKE, 2020).

Apesar de ainda ser necessário mais treino dos algoritmos de ML, esta tecnologia foi bem aceite pelos profissionais: *Now you can actually have a grasp on the whole material, and ask questions that were not possible earlier* (BURKE, 2020). Para os utilizadores, esta tecnologia permite que possam extrair dos documentos informação significativa e em quantidade, com celeridade.

Arquivo Fotográfico Finlandês do Tempo de Guerra

O Arquivo Fotográfico das Forças de Defesa Finlandesas recolhe e armazena material fotográfico, sendo constituído, na sua maioria, por material entregue pelas unidades e estabelecimentos das Forças de Defesa e por doações particulares.

O Arquivo Fotográfico Finlandês do Tempo da Guerra (*Finnish Wartime Photograph Archive*), também designado *SA-kuva-arkisto* (em finlandês), reúne cerca de 160 mil fotografias captadas entre 1939 e 1945, durante a *Winter War*, *Continuation War* e *Lapland War*, produzidas, maioritariamente, pelas tropas dos TK (*Tiedotuskomppania* - Centro de Informação Militar). Este acervo representa uma fonte de informação para várias áreas científicas, nomeadamente arqueologia, história das guerras, estudos

demográficos, religião, história da fotografia, no entanto, a sua análise manual representa um elevado investimento temporal para os investigadores.

Desde 2010 que o Arquivo Fotográfico tem a preocupação de digitalizar e tornar acessíveis os seus documentos e fotografias. Em 2020, investigadores da Universidade de Aarhus (Dinamarca), em colaboração com a Universidade de Tampere e o Instituto Finlandês do Ambiente, recorreram a ML, mais precisamente OCR (Reconhecimento ótico de caracteres), de forma a facilitar a análise de *high numbers of photographs in the publicly available Finnish Wartime Photograph Archive, which contains 160,000 photographs from Finnish Winter, Continuation, and Lapland Wars captures in 1939-1945* (CHUMACHENKO *et al.*, 2020:144184). Esta investigação teve como objetivo reconhecer a identidade dos fotógrafos de fotos históricas da II Guerra Mundial. O estudo mostrou que a IA pode reconhecer a identidade dos fotógrafos com base no conteúdo das imagens. Este tipo de IA está também presente no nosso dia a dia, por exemplo nas recomendações *online* de produtos dos sites de vendas de séries e filmes da plataforma de *streaming* Netflix ou no reconhecimento de texto e voz (GOMES, 2022).

Para este projeto foram utilizados 23 fotógrafos do TK e um total de 59.021 fotografias, e os investigadores tiveram como base a aplicação de algoritmos para deteção de objetos e redes neuronais convolucionais, o que permite identificar padrões visuais recorrentes associados a cada fotógrafo. Através da análise de objetos presentes nas fotografias – como pessoas, aviões, barcos, cavalos, cadeiras ou gravatas – foi possível inferir o contexto das imagens, distinguindo, por exemplo, fotografias no interior e exterior, em eventos formais ou cenários rurais e urbanos. Segundo os autores, *presence of objects such as skis can help determine the time of the year and, therefore, help establish the time period for unlabeled photographs* (CHUMACHENKO *et al.*, 2020:144186).

Além da análise de objetos, foi também estudado o enquadramento fotográfico (plano geral, plano médio e primeiro plano), considerado uma característica estilística dos fotógrafos. *Measuring the ratio of different framings in a photographer's works (...) is one way to characterize his/her way of seeing* (CHUMACHENKO *et al.*, 2020:144186). Esta abordagem permitiu identificar diferenças consistentes entre fotógrafos urbanos e não urbanos, e também entre fotografias de reportagem e retrato.

Relativamente ao reconhecimento automático de autoria, uma rede neuronal foi treinada para fazer a classificação de fotógrafos com base em fotografias não utilizadas na fase de treinamento do algoritmo. Segundo Chumachenko *et al.* (2020:144192) os resultados demonstraram que a rede alcançou uma taxa média de 41,1% de precisão, com variações significativas entre fotógrafos, o que confirma que certos estilos visuais são mais fáceis de distinguir do que outros. A avaliação da semelhança visual entre fotógrafos, através da *Earth Mover's Distance* (EDM), reforçou esta conclusão, colocando em evidência que os fotógrafos com focos temáticos distintos apresentam menor probabilidade de erro de classificação.

Os investigadores defendem que este tipo de abordagem pode trazer benefícios significativos para arquivos e museus, ao automatizar a descrição, enriquecimento e acessibilidade de grandes coleções fotográficas, contribuindo também para o cumprimento das diretivas europeias de acessibilidade digital. Além disso, sublinham também que os modelos, códigos e dados utilizados estão disponíveis ao público, podendo o método ser replicado noutras instituições.

Apesar dos resultados promissores, o estudo reconhece algumas limitações, nomeadamente ao nível da deteção de objetos de pequenas dimensões e da ausência de análises mais detalhadas, como reconhecimento facial, expressão facial ou análise de cenas. Ainda assim, os autores concluem que *algorithms of ML can indeed be a useful tool for historical research, especially in photographic archives* (CHUMACHENKO *et al.*, 2020: 144195), posicionando este projeto como sendo um dos primeiros a aplicar ML à análise de autoria e estilo em fotografia histórica.

Arquivo Apostólico do Vaticano

O Arquivo Apostólico do Vaticano (AAV), anteriormente designado Arquivo Secreto do Vaticano, é o detentor de uma das mais vastas coleções históricas do mundo, com cerca de oitenta e cinco mil quilómetros de prateleiras e trinta e cinco mil volumes nos seus catálogos, abrangendo aproximadamente doze séculos de história. Entre os inúmeros documentos guardados encontram-se atos promulgados pelo Vaticano, livros de contabilidade, correspondência dos papas desde o século VIII. O acesso ao arquivo é reservado a investigadores credenciados, sendo a consulta presencial obrigatória devido ao reduzido número de documentos integralmente transcritos em formato digital.

A dificuldade de acesso e a enorme dimensão do acervo são justificadas, pelo então diretor Sergio Pagano, não como sendo falta de vontade institucional, mas por limitações de recursos humanos e técnicos para processar e disponibilizar documentação mais rapidamente: *it's not a matter of courage – it's a matter of resources. Since the archive is vast, Pagano said it's a challenge to quickly process documents and make them available to historians* (BLAKEMORE, 2017). Neste contexto, a necessidade de recorrer a tecnologias automáticas de análise e transcrição tornou-se evidente, sobretudo para os manuscritos mais antigos, cuja transcrição manual seria inviável num prazo razoável.

O AAV colaborou com o projeto *In Codice Ratio*, desenvolvido pela Universidade de Roma Tre. Este projeto surgiu em 2016 com o objetivo de desenvolver novos métodos e ferramentas para apoiar a análise do conteúdo dos documentos, sendo o foco a coleção dos registos do Vaticano, composta maioritariamente por documentos manuscritos em latim medieval, um grande desafio, uma vez que *a manual transcript is unfeasible (...) and traditional OCR does not apply here, because of irregularities of writing, ligatures and abbreviations* (FIRMANI *et al.*, 2018:1).

Para ultrapassar as limitações do OCR tradicional, os investigadores propuseram uma abordagem inovadora baseada na segmentação de caracteres e no uso combinado de redes neuronais convolucionais e modelos estatísticos. Em vez de reconhecer palavras completas, o sistema reconhece caracteres individuais contornando o chamado *Sayre's paradox*, segundo o qual o OCR precisa de segmentar os caracteres para os reconhecer, mas ao mesmo tempo precisa de os reconhecer para os segmentar (KEAN, 2018). A solução passa, então, por gerar vários segmentos possíveis de uma palavra e selecionar qual a sequência mais provável com base em estatísticas linguísticas do latim.

O sistema desenvolvido integra várias etapas, que podem ser sintetizadas do seguinte modo: em primeiro lugar, os manuscritos são digitalizados em alta qualidade e submetidos a um pré-processamento destinado a converter a imagem a cores em formato bi-cromático, remover margens brancas, corrigir distorções e alinhar texto e linhas, até se obter imagens de palavras. Em seguida, procede-se à recolha de amostras de treino através de uma

plataforma de *crowdsourcing*, na qual 120 estudantes do ensino secundário classificaram segmentos de caracteres com base nos exemplos visuais que lhes foram fornecidos, permitindo assim treinar um classificador baseado em *deep convolutional neural networks*.

Numa fase posterior, o sistema realiza o reconhecimento dos caracteres, segmentando a palavra em unidades elementares e submetendo-as ao classificador, para depois reconstruir possíveis transcrições com apoio de modelos de linguagem. Por fim, o processo de *word decoding* combina as transcrições candidatas e refina as decisões de reconhecimento com recurso a um modelo oculto de Markov de ordem superior, devolvendo ao utilizador as transcrições mais promissoras. Como referem Firmani *et al.* (2018), trata-se de uma arquitetura em que a aprendizagem automática é complementada por validação humana e por conhecimento linguístico; nesse sentido, Kean (2018) sublinha que a equipa teve de ensinar ao *software* uma forma de *common sense*, isto é, uma inteligência prática indispensável para lidar com a ambiguidade da escrita manuscrita.

O projeto *In Codice Ratio* constitui um exemplo relevante da aplicação da IA à leitura e transcrição de manuscritos históricos, evidenciando o potencial da articulação entre processamento automático, participação humana e modelos linguísticos para superar limitações tradicionais da paleografia digital. A sua importância científica reside na possibilidade de responder a um problema central dos arquivos e da investigação histórica – o acesso a grandes volumes de documentação manuscrita cuja leitura manual é morosa, onerosa e, em muitos casos, impraticável em larga escala. Do ponto de vista metodológico, o sistema integra um conjunto de etapas que inclui o pré-processamento das imagens digitalizadas, a segmentação textual, a aprendizagem supervisionada com apoio de *crowdsourcing* e o *word decoding* com recurso a modelos de linguagem e a modelos ocultos de Markov, permitindo ao algoritmo aprender progressivamente a reconhecer caracteres e a propor transcrições cada vez mais fiáveis.

Os investigadores consideram que a estratégia utilizada no projeto *In Codice Ratio* pode ser adaptada a textos escritos noutras línguas.

This could potentially do handwrittens documents what Google Books did for printed matter: open up letters, journals, diaries, and other papers to researchers around the worl, making it far easier to both read these documents and search for relevant material" (KEAN, 2018).

Como sublinham Firmani *et al.* (2018), esta arquitetura combina IA com conhecimento linguístico e validação humana; nessa mesma linha, Kean (2018) assinala que a equipa teve de ensinar ao *software* uma forma de *common sense*, indispensável para lidar com a ambiguidade e a variabilidade da escrita manuscrita. Em termos de contributo, o projeto demonstra que a IA pode gerar transcrições úteis mesmo quando imperfeitas, oferecendo informação suficiente para apoiar a análise documental e ampliar o acesso a acervos até então pouco exploráveis, mas também confirma que a automatização não dispensa a supervisão especializada, a curadoria e o refinamento contínuo por parte dos profissionais da informação.

Arquivo Arolsen

O Arquivo Arolsen é considerado o mais completo do mundo sobre as vítimas e sobreviventes do Holocausto, preserva a maior coleção de documentos acerca da perseguição nazi reunindo cerca de *110 million documents and digital objects, a portion of which are part of UNESCO's Memory of the World program* (PRICE, 2022).

Criado no final da Segunda Guerra Mundial por iniciativa dos Aliados, o Arquivo Arolsen teve inicialmente como missão localizar vítimas e documentar crimes nazis, reunindo documentação sobre campos de concentração, trabalho forçado e pessoas deslocadas. Atualmente, o seu principal objetivo é tornar a informação o mais acessível possível, enfrentando, contudo, o desafio da dimensão do acervo e da morosidade dos processos manuais de digitalização, transcrição e validação. Antes da introdução de soluções tecnológicas avançadas, *one single document required four people to review it (...) before it could be certified* (ACCENTURE, 2025).

Assim surge, em 2020, o projeto *#everynamecounts*, que alia *crowdsourcing* e IA com o objetivo de acelerar a indexação e preservação da memória das vítimas. Com o apoio da Accenture, foi desenvolvido um sistema que combina *cloud-based technologies, optical character recognition (OCR) solutions and the latest AI and machine learning (ML) techniques* (ACCENTURE, 2025), permitindo analisar automaticamente diferentes tipos de documentos. O sistema atribui a cada documento um nível de “confiança” com base em campos como apelido, religião ou local, sendo os casos menos claros revistos por voluntários, cujo *feedback* é utilizado para melhorar continuamente o desempenho da IA. Os resultados desta abordagem revelaram ganhos significativos de eficiência: tarefas que anteriormente demoravam cerca de 15 minutos por documento passaram a ser realizadas em aproximadamente 20 segundos. Desde o início do projeto, foram indexados mais de 160 mil nomes, extraídos 180 mil documentos e agrupados 63 mil documentos semelhantes, reduzindo o número de pessoas por tarefa, o que se traduziu num aumento substancial de produtividade.

Paralelamente ao *#everynamecounts*, o Arquivo Arolsen desenvolve um projeto com OCR em larga escala, que visa facilitar o acesso a documentos históricos dactilografados e manuscritos. Atualmente, cerca de 26 milhões de documentos estão disponíveis *online*, tendo o arquivo sido um dos distinguidos com o *2020 European Heritage Awards/Europa Nostra Awards* pelo seu contributo para o património cultural digital.

No entanto, os profissionais do arquivo reconhecem as limitações desta tecnologia, sobretudo no que respeita a documentos manuscritos que apresentam grande variabilidade de caligrafias. Como refere Michael Hoffmann, *it's not efficient to use OCR on them because the program would have to learn each individual script in advance in order to correctly recognize the characters* (AROLSEN..., 2025).

A estratégia adotada pelo Arquivo Arolsen baseia-se, portanto, numa combinação equilibrada entre processos automatizados e intervenção humana, recorrendo a ML, OCR e *crowdsourcing* conforme a natureza e complexidade dos documentos. Esta abordagem híbrida tem permitido acelerar significativamente a indexação, reduzir custos e aumentar a acessibilidade pública ao acervo, mantendo simultaneamente elevados padrões de qualidade e rigor histórico. Em síntese, o Arquivo Arolsen demonstra como a aplicação integrada de IA, ML, OCR e participação cívica pode transformar a gestão de grandes

arquivos históricos sensíveis, contribuindo não apenas para a investigação académica, mas também para a construção de um memorial digital vivo que preserva os nomes, as histórias e a memória das vítimas do Holocausto para as gerações futuras.

Arquivo Nacional do Reino Unido

O Arquivo Nacional do Reino Unido (The National Archives – TNA) foi fundado em 2003, juntou quatro organizações históricas, e atua como o arquivo oficial do governo britânico, de Inglaterra e do País de Gales. O TNA detém documentos com mais de 1.000 anos, *such as wills, naturalisation certificates and criminal records, and many others* (JAILLANT, ASKE e CAPUTO, 2022:3).

O crescimento exponencial da documentação digital levou à implementação de tecnologias de IA inspiradas em ferramentas legais como o eDiscovery, desenvolvido para auxiliar advogados na identificação, seleção e análise de grandes volumes de documentos digitais. Este *software* baseia-se em aprendizagem supervisionada e não supervisionada, sendo posteriormente testado em contextos de arquivo para otimizar a gestão de registos governamentais digitais.

O TNA desenvolveu, em colaboração com a rede AEOLIAN (Artificial Intelligence for Cultural Organisations), iniciativas orientadas para a acessibilidade de registos culturais digitalizados e/ou em formato digital, reunindo académicos, cientistas computacionais e arquivistas em torno da modernização do acesso documental. Entre os projetos implementados, destaca-se o *AI for Digital Selection*, centrado na avaliação de ferramentas de IA para apoiar a seleção de documentos digitais governamentais a preservar, com o objetivo de identificar, de forma mais eficiente, os arquivos a manter ou a eliminar; ainda assim, os próprios resultados do projeto evidenciam que, embora a IA possa apoiar os gestores documentais, não substitui o contributo humano, pois *while the project concluded that AI tools can assist record managers, the machine cannot replace human input* (JAILLANT, ASKE e CAPUTO, 2022:6). O projeto *AI for Sensitivity Review* incidiu na identificação de documentos sensíveis, nomeadamente conteúdos com dados pessoais ou financeiros, recorrendo a revisão assistida por tecnologia e supervisão humana, numa lógica em que a compreensão contextual é decisiva para reduzir erros, uma vez que *the context is as important as the text itself when evaluating the sensitivity of a document* (JAILLANT, ASKE e CAPUTO, 2022:7). Por sua vez, o *AI for Discovery* aplicou técnicas de modelação temática e ML para facilitar a descoberta de informação relevante em grandes coleções, permitindo identificar padrões e tópicos significativos em contextos históricos complexos; a análise da coleção *General Board of Health*, com cerca de 89.000 registos de correspondência do século XIX, através de *Latent Dirichlet Allocation* (LDA), revelou temas centrais como as inspeções sanitárias, demonstrando a utilidade contínua da IA para a investigação histórica e para a leitura de grandes massas documentais, inclusive em domínios como a saúde pública. Em conjunto, estes projetos mostram que a IA pode acelerar a seleção, reforçar a proteção da informação sensível e ampliar as possibilidades de descoberta documental, mas sempre como instrumento de apoio e não de substituição da mediação profissional.

Durante a pandemia COVID-19, o TNA implementou captura detalhada de registos governamentais *online*, preservando mais de 50TB de dados digitais sobre a resposta à COVID-19, garantindo que informações críticas não fossem perdidas em contexto de rápida

mudança. John Sheridan, Diretor Digital do TNA, destacou a importância da IA para mediar entre o volume de evidências e as questões de investigação pública:

How does a Public Inquiry begin to grapple with a collection of this size and scale? What role do AI tools have as we provide the mediation layer between the evidence on one side and the big questions that the Inquiry will be exploring on the other?” (JAILLANT, ASKE e CAPUTO, 2022:8-9).

O TNA demonstrou que a aplicação de IA pode transformar a gestão de arquivos governamentais digitais, otimizando a seleção e revisão de documentos, aumentando a acessibilidade pública e protegendo dados sensíveis, enquanto apoia investigadores e políticas de preservação de longo prazo.

Análise comparativa das respostas ao inquérito por entrevista

Relativamente ao inquérito por entrevista enviado às cinco instituições, foi possível obter quatro respostas de profissionais da informação com o seguinte perfil/função: Arquivo da Cidade de Amesterdão – arquivista; AAV – arquivista e investigador; Arquivo Arolsen – responsável pelo departamento de Gestão de processos e Qualidade; TNA – arquivista, responsável pela Catalogação, Taxonomia e Dados.

As respostas obtidas mostram uma tendência clara: os profissionais veem a IA menos como substituta do arquivista e mais como ferramenta de apoio à descrição, à indexação, ao acesso e à exploração de grandes massas documentais. Contudo, o grau de adoção varia bastante. Os profissionais do Arquivo da Cidade de Amesterdão e do Arquivo Arolsen apresentam uma postura mais pragmática e operacional, enquanto os do AAV e do TNA revelam uma atitude mais cautelosa, reflexiva e crítica.

Questão 1: vantagens da IA/ML nas coleções selecionadas

Para esta primeira questão, a resposta de Amesterdão destaca sobretudo a rapidez e a possibilidade de dar acesso a informação ainda não digitalizada e “escondida à vista de todos”, argumento que coloca a IA como solução para problemas de escala e de visibilidade documental. No AAV reconhece-se que a IA apoia validamente a análise da documentação histórica, mas insiste-se que esse apoio só é legítimo se for integrado organicamente no trabalho dos especialistas e submetido a escrutínio humano. Já no Arquivo Arolsen, aponta-se como principal vantagem a progressão rápida no processo de indexação e geração de dados. A resposta do TNA é menos centrada na eficiência imediata e mais no ganho epistémico: a principal vantagem é compreender melhor o que a IA é e como funciona, permitindo ao arquivo participar como *intelligent client* no debate profissional. Em comparação, os/as profissionais de Amesterdão e de Arolsen focam a utilidade operacional, enquanto os/as profissionais do Vaticano e do TNA privilegiam a compreensão crítica da tecnologia.

Questão 2: impacto nos serviços aos utilizadores

Na segunda questão, o/a profissional de Amesterdão apresentou a resposta mais desenvolvida e mais orientada para o utilizador, defendendo que a IA permite novas formas de pesquisa através de API, pesquisa em texto integral e integração com LLM como o ChatGPT, beneficiando investigadores tecnicamente experientes e também cidadãos sem

formação histórica. A vantagem aqui é dupla: flexibilidade analítica para especialistas e acessibilidade para públicos mais amplos. No Arquivo Arolsen, o impacto nos serviços surge associado ao aumento de metadados e à necessidade de trabalho adicional em limpeza e estruturação de dados, o que mostra que a melhoria do serviço não é automática nem isenta de custos internos. Já no AAV, a IA não é apresentada como eixo central do serviço ao utilizador, mas como recurso externo de suporte a um projeto científico específico, sem necessidade institucional ampla. No TNA afirma-se que o impacto direto nos utilizadores ainda é mínimo, embora o arquivo acompanhe as transformações que a IA pode trazer para a investigação e para futuros serviços. Assim, Amesterdão é o caso mais orientado para o uso público, enquanto os/as profissionais do TNA e do AAV são os mais prudentes quanto ao impacto efetivo.

Questão 3: impacto na atividade profissional dos arquivistas

Nesta questão, as diferenças entre instituições tornam-se muito nítidas. Segundo o/a profissional do Arquivo da Cidade de Amesterdão, IA e HTR são uma ajuda amplamente aceite, embora gere novas questões sobre curadoria, estatuto do texto transcrito e relação entre original e derivado. Isto indica uma prática em que a tecnologia já entrou no quotidiano profissional, ainda que reclame redefinição concetual. No AAV, a resposta do/a profissional é mais conservadora: a IA apoia, mas não substitui a análise especializada, e os arquivistas continuam a realizar estudo e inventário com base na sua experiência e competência paleográfica. O projeto *In Codice Ratio* é descrito como colaboração científica autónoma, sem alteração substancial das rotinas do AAV. No Arquivo Arolsen, o impacto é mais integrado, pois arquivistas e equipa de dados trabalham em conjunto, num processo de aprendizagem recíproca. No TNA, o impacto ainda é mínimo, porque a IA/ML não foi incorporada nos fluxos de trabalho.

Comparativamente, Arolsen é o caso mais híbrido e colaborativo; Amesterdão mostra maior integração com novas questões curatoriais; AAV e TNA mantêm uma distância institucional mais marcada.

Questão 4: vantagens no presente e no futuro próximo

Na quarta questão, o/a profissional de Amesterdão volta a enfatizar o potencial da IA para transformar HTR em dados estruturados, gerar pontos de acesso adicionais e apoiar a transcrição e resumo de documentos históricos. A vantagem futura é a ampliação de acesso e a ligação entre coleções em níveis mais complexos. No Arquivo Arolsen, a resposta aponta para o desenvolvimento de um motor de pesquisa baseado em *knowledge graph*, no qual mais metadados significam mais pontos de ligação e mais possibilidades de descoberta. O AAV apresenta a abordagem mais prudente: reconhece-se que a IA pode acelerar algumas etapas, desde que não assuma a totalidade do processo e que haja supervisão humana contínua. No TNA, as vantagens são vistas de forma ambivalente. Reconhecem-se possibilidades para criação automática de metadados e novas formas de análise, mas questiona-se “de cuja perspetiva” essas vantagens realmente se realizam e a que custo para os arquivos. Assim, enquanto os/as profissionais de Amesterdão e Arolsen identificam ganho futuro em acesso e conectividade, os/as profissionais do Vaticano e do TNA insistem no equilíbrio entre benefício técnico, carga de trabalho e supervisão humana.

Questão 5: desafios principais

As respostas convergem na ideia de que o maior desafio é manter controlo humano, ético e técnico sobre processos cada vez mais automatizados. O/a profissional de Amesterdão aponta a necessidade de escolher entre dados parcialmente curados ou dados sem intervenção humana, além das expectativas diferenciadas dos utilizadores e dos custos associados à produção em escala. O/a profissional do AAV sublinha que confiar integralmente na IA para transcrição ou interpretação seria arriscado, porque a leitura de documentos exige abstração humana e avaliação especializada. No Arquivo Arolsen, o/a profissional indica que o desafio principal é encontrar equilíbrio entre disponibilização massiva de informação *online* e segurança. No TNA, o/a profissional considera que o desafio é mais estrutural: falta de influência para dirigir o desenvolvimento tecnológico, tendência a reagir em vez de antecipar, e a tensão entre tarefas mecânicas e missão social de preservar a verdade. Em síntese, os desafios não são apenas técnicos; são também epistemológicos, éticos, institucionais e políticos.

Discussão de resultados

Os cinco projetos assemelham-se em vários aspetos. Todos foram aplicados apenas a uma parte do acervo (séries ou coleções) dos arquivos. Por um lado, é necessário existir um grande volume de informação/documentos para poder treinar os algoritmos de ML, de forma que estes sejam eficazes. O que nem sempre é possível em arquivos pois, muitas vezes, as coleções e séries de interesse para determinada investigação são únicas e pequenas, tornando-as insuficientes para concretizar o treino de algoritmos. *Only with enough data can an ML algorithm be properly trained to identify a pattern and locate other items that fit that pattern. With a small number of examples, it cannot function reliably* (KIM, 2021:41). Por outro lado, para que a ML funcione, todo o acervo tem de estar em formato digital. Esta tarefa torna-se difícil, pois os arquivos têm quilómetros de documentação, o que se pode traduzir em décadas de trabalho de digitalização e, muitas vezes, com equipas reduzidas torna-se complicado. O/a profissional da informação que respondeu às questões do Arquivo da Cidade de Amesterdão refere que, na sua organização, não existe um maior recurso aos sistemas de IA por falta de meios financeiros.

Outro aspeto convergente entre as cinco instituições é a preocupação ética associada ao uso da IA, bem como a recusa em atribuir à máquina a totalidade do processo de trabalho. Em vez disso, a IA é entendida como um instrumento de apoio aos profissionais da informação, cabendo-lhes, em fase final, avaliar a fiabilidade dos resultados produzidos e verificar a sua adequação à utilização interna e ao acesso por parte dos utilizadores.

Os avanços da IA são inegáveis e esta tem promovido transformações significativas na gestão da informação e na sua disponibilização ao público. Não obstante, tais progressos suscitam preocupações de natureza ética, sobretudo no que respeita à responsabilização pelos resultados, à transparência dos processos e à manutenção do controlo humano. Os arquivistas tendem a adotar uma postura prudente face aos riscos associados a estas tecnologias, revelando simultaneamente uma forte sensibilidade ética e moral no exercício da sua atividade profissional, pois *unlocking digital archives requires cross-disciplinary collaborations, but also close attention to ethical principles* (JAILLANT e CAPUTO, 2022:833).

Os desafios que advém da aplicação da IA nos arquivos têm crescido à medida que a utilização desta tecnologia também cresce. No entanto, é de frisar os benefícios que esta traz e como ajuda os profissionais da área no seu trabalho quotidiano. Não sendo uma novidade, IA e ML podem ser utilizadas para desbloquear e tornar os arquivos mais acessíveis e localizáveis, *for example by differentiating between sensitive and non-sensitive materials, or by automatically adding tags and other metadata to improve findability* (JAILLANT, 2022:418). Isto torna-se fundamental, pois só assim é possível evitar que haja parcialidade na seleção e processamento de dados.

Além disto, através de IA e ML é possível separar conteúdos em categorias gerais (ex. em categorias culturais ou políticas), o que facilita os investigadores poderem-se focar-se nas suas áreas de interesse. A IA pode também ser utilizada para identificar metadados de coleções históricas que sejam considerados controversos. Outro uso bastante benéfico que esta tecnologia pode ter é *automatically examine the contents of the file, and then make prediction on file types – which is useful in the context of digital preservation and access* (JAILLANT, 2022:433).

De todas as vantagens que a aplicação da IA traz aos arquivos, possivelmente a maior é possibilidade de identificar conteúdos sensíveis. No entanto, levantam-se algumas preocupações ligadas à proteção de dados, sensibilidade dos documentos e de direitos de autor. *Obstacles to access are numerous and include issues with data protection, sensitivity, and copyright. In the case of government archival collections, releasing potentially sensitive and private information could threaten national security and embarrass foreign partners* (JAILLANT e REES, 2023:571).

Apesar de todas as vantagens e desafios que os arquivos e os arquivistas enfrentam com a crescente implementação de IA e ML nas suas organizações, é de frisar que ainda é necessária investigação, pois

the issues of innovative technologies applied to digital archives have too often been explored within single disciplines such as archival studies or computer science. Yet, applying AI to archives requires a variety of skillsets: not only computational skills, but also domain knowledge, i.e., knowledge about the entities that are being treated by the algorithms. (...) Digital Humanities – which is in essence an interdisciplinary field – has an important role to play in bridging the gap between record creators, archivists, researchers, and other users (JAILLANT e REES, 2023:582).

Para concluir, é importante frisar que não existe uma solução estabelecida para o sucesso da IA nos arquivos. As instituições analisadas trabalharam e continuam a trabalhar em colaboração com outras entidades, nomeadamente universidades e, no caso do Arquivo Arolsen, também com utilizadores. Desta forma, é possível dizer que a solução, neste momento, é trabalhar em colaboração com as várias áreas científicas e os profissionais.

Conclusão

Os resultados deste estudo demonstram que a aplicação da IA e ML em arquivos apresenta um grande potencial para transformar a forma como a informação é tratada, preservada e disponibilizada aos utilizadores. As tecnologias contribuem para uma maior eficiência nos

processos arquivísticos, permitindo acelerar tarefas como a digitalização, transcrição, indexação e análise de grandes volumes de documentos, facilitam o acesso à informação por parte dos investigadores e do público em geral.

A análise das respostas recebidas junto de quatro instituições arquivísticas europeias evidencia um campo em transição, no qual a IA é percebida simultaneamente como oportunidade de modernização e como fonte de tensão profissional e ética. Embora todas as instituições reconheçam algum potencial da IA e ML, divergem quanto ao grau de adoção, ao tipo de benefício esperado e ao modo como essa tecnologia deve ser integrada nos fluxos arquivísticos.

A utilização da IA em arquivos traz desafios importantes, principalmente ao nível da ética, da proteção de dados, da fiabilidade dos resultados produzidos pelos algoritmos e da necessidade de supervisão humana. A IA deve ser vista como uma ferramenta de apoio ao trabalho dos profissionais da informação e não como um substituto das suas funções, sendo imprescindível manter o controlo humano sobre os processos de decisão. Além disso, a implementação destas tecnologias exige recursos financeiros, infraestruturas adequadas e competências especializadas, bem como uma abordagem colaborativa e interdisciplinar.

Apesar da IA representar uma boa oportunidade para a melhoria dos serviços prestados aos utilizadores, a sua adoção deve ser feita de forma crítica, responsável e sustentada, garantindo o respeito pelos princípios éticos e pelos valores fundamentais da teoria e prática arquivística.

Referências bibliográficas

ACCENTURE

2025 *Unearthing the past with artificial intelligence*. [Em linha]. 2025. Disponível em: <https://www.accenture.com/us-en/case-studies/public-service/arolsen-archives>.

AMSTERDAM CITY ARCHIVES

2025 *City Archives Amsterdam*. [Em linha]. 2025. Disponível em: <https://www.amsterdam.nl/stadsarchief/organisatie/stadsarchief-amsterdam>.

AROLSEN ARCHIVES

2025 *How optical character recognition (OCR) makes documents accessible*. [Em linha]. 2025. Disponível em: <https://arolsen-archives.org/en/about-us/statements/history-goes-online-easier-access-to-historical-documents-with-ocr>.

BATISTA, Bruna F. [et al.]

2021 Técnicas de recolha de dados em investigação: inquirir por questionário e/ou inquirir por entrevista? In *Reflexões em torno de metodologias de investigação: recolha de dados*. [Em linha]. Org. Patrícia Sá, António Pedro Costa e António Moreira. Aveiro: Universidade de Aveiro, 2021, vol. 2, p. 13-36. Disponível em: https://ria.ua.pt/bitstream/10773/30772/3/Metodologias%20investigacao_Vol2_Digital.pdf.

BLAKEMORE, E.

2017 *Step into the Vatican's Secret Archives*. [Em linha]. 2017. Disponível em: <https://www.history.com/news/step-into-the-vaticans-secret-archives>.

BURKE, Fintan

2020 Machine learning and big data are unlocking Europe's archives. *Horizon: The EU Research & Innovation Magazine*. [Em linha]. 2020. Disponível em <https://ec.europa.eu/research-and-innovation/en/horizon-magazine/machine-learning-and-big-data-are-unlocking-europes-archives>.

CORREIA, Magda Valente

2023 *A Inteligência Artificial em serviços de informação: estudo de casos em arquivos europeus*. [Em linha]. Coimbra, 2023. Disponível em: <https://estudogeral.uc.pt/handle/10316/111726>.
Dissertação de mestrado – Faculdade de Letras da Universidade de Coimbra.

CHUMACHENKO, Kateryna [et al.]

2020 Machine Learning based analysis of Finnish World War II photographers. *IEEE Access*. 8 (2020) 144184-144196. <https://doi.org/10.1109/ACCESS.2020.3014458>.

COLAVIZZA, Giovanni [et al.]

2021 Archives and AI: An Overview of Current Debates and Future Perspectives. *ACM Journal on Computing and Cultural Heritage*. [Em linha]. 15:1 (2021) article 4. Disponível em: <https://doi.org/10.1145/3479010>.

COX, Andrew M.; PINFIELD, Stephen; RUTTER, Sophie

2019. The Intelligent library. *Library Hi Tech*. [Em linha]. 37:3 (2019) 418-435. Disponível em: <https://doi.org/10.1108/lht-08-2018-0105>.

DALGLEISH, Lucy

2022 *Artificial Intelligence, cultural heritage and the National Library of Scotland*. [Em linha]. 2022. Disponível em: <https://data.nls.uk/projects/artificial-intelligence-report>.

EKBIA, Hamid R.

2012 Fifty Years of Research in Artificial Intelligence. *Annual Review of Information Science and Technology*. [Em linha]. 44:1 (2012) 201-242. Disponível em: <https://doi.org/10.1002/aris.2010.1440440112>.

FALCÃO, Luander Cipriano de Jesus; LOPES, Brenner; SOUZA, Renato Rocha

2022 Absorção das tarefas de processamento de Linguagem Natural (NLP) pela Ciência da Informação (CI): uma revisão de literatura para tangibilização do uso de NLP pela CI. *Em Questão*. [Em linha]. 28:1 (2022) 13-34. Disponível em: <https://doi.org/10.19132/1808-5245281.13-34>.

FIRMANI, Donatella [et al.]

2018 Towards Knowledge Discovery from the Vatican Secret Archives. In *Codice Ratio. Episode 1: Machine Transcription of the Manuscripts*. [Em linha]. 2018, arXiv:1803.03200 Disponível em: <https://arxiv.org/abs/1803.03200>.

GOMES, Liliana Isabel Esteves

2022 Transformação digital e Inteligência Artificial nos serviços de informação: inovação e perspectivas para a Ciência da Informação no mundo pós-pandemia. *Revista Ibero-Americana de Ciência da Informação*. [Em linha]. 15:1 (2022) 148-166. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/41490>.

GOMES, Liliana Esteves

2020 Ciência da Informação: fundamentos e perspectivas da área científica. In *Ciência da Informação: visões e tendências*. Coord. Maria Beatriz Marques, Liliana Esteves Gomes. [Em linha]. Coimbra: Imprensa da Universidade, 2020, p. 89-113. Disponível em: <https://estudogeral.uc.pt/bitstream/10316/101093/1/107-Book%20Manuscript-437-1-10-20200824.pdf>.

GOMES, Liliana Isabel Esteves; FERNÁNDEZ MARCIAL, Viviana; SANTOS, Miguel

2021 O Impacto da Inteligência Artificial nos serviços de informação: Inovação e perspectivas para as bibliotecas. In CONGRESSO ISKO ESPANHA-PORTUGAL, 5º, Lisboa, 2021 - *Organização do Conhecimento no Horizonte 2030: Desenvolvimento Sustentável e Saúde: Atas*. Ed. Carlos Guardado da Silva, Jorge Revez, Luís Corujo. [Em linha]. Lisboa: Edições Colibri; Centro de Estudos Clássicos, 2021, p. 393-405. [Consult. 15 jan. 2026]. Disponível em: <https://estudogeral.uc.pt/handle/10316/103253>.

HAWKINS, Ashleig

2022 Archives, linked data and the digital humanities: increasing access to digitised and born-digital archives via the semantic web. *Arch Sci*. [Em linha]. 22 (2022) 319-344. Disponível em: <https://doi.org/10.1007/s10502-021-09381-0>.

JAILLANT, Lise

2022 How can we make born-digital and digitised archives more accessible? Identifying obstacles and solutions. *Archival Science*. [Em linha]. 22 (2022) 417-436. Disponível em: <https://doi.org/10.1007/s10502-022-09390-7>.

JAILLANT, Lise; CAPUTO, Annalina

2022 Unlocking digital archives: cross-disciplinary perspectives on AI and born-digital data. *AI & Society*. [Em linha]. 37 (2022) 823-835. Disponível em: <https://doi.org/10.1007/s00146-021-01367-x>.

JAILLANT, Lise; CAPUTO, Annalina; ASKE, Katherine

2022 *The National Archives (UK): case study*. [Em linha] 2022. Disponível em: <https://www.aeolian-network.net/case-studies>.

JAILLANT, Lise; REES, Arran

2023 Applying AI to digital archives: trust, collaboration and shared professional ethics. *Digital Scholarship in the Humanities*. [Em linha]. 38:2 (2022) 571-585. Disponível em: <https://doi.org/10.1093/llc/fqac073>.

KEAN, Sam

2018 *Artificial Intelligence is cracking open the Vatican's Secret Archives*. [Em linha]. 2018. Disponível em: <https://www.theatlantic.com/technology/archive/2018/04/vatican-secret-archives-artificial-intelligence/559205/>.

KIM, Bohyun

2021 Machine Learning for Libraries and Archives. *Online Searcher: Information Discovery, Technology, Strategies*. 45:1 (2021) 39-41.

MÖDDEN, Elisabeth

2022 Artificial intelligence, machine learning and bibliographic control. DDC Short Numbers, Towards machine-based classifying. *JLIS.it*. [Em linha] 13:1 (2022) 256-264. Disponível em: <https://doi.org/10.4403/jlis.it-12775>.

PRICE, Gary

2022 *Arolsen Archives' #everynamecounts Project Uses Artificial Intelligence to Help Uncover Information on Victims of Nazi Persecution*. [Em linha] 2022. Disponível em:

<https://www.infodocket.com/2022/04/27/arolsen-archives-everynamecounts-project-uses-artificial-intelligence-to-help-uncover-information-on-victims-of-nazi-persecution/>.

QUIVY, Raymond; CAMPENHOUDT, LucVan

2005 *Manual de investigação em Ciências Sociais*. Lisboa: Gradiva, 2005.

ROLAN, Gregory [et al.]

2019 More human than human? Artificial intelligence in the archive. *Archives and Manuscripts*. [Em linha] 47:2 (2019) 179-203. Disponível em:

<https://doi.org/10.1080/01576895.2018.1502088>.

SHANG, Erbo [et al.]

2019. Research on the Application of Artificial Intelligence and Distributed Parallel Computing in Archives Classification. In *IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 2019*. [Em linha]. 2019, p. 1.267-1.271. Disponível em:

<https://doi.org/10.1109/iaeac47372.2019.8997992>.

STAKE, Robert E.

2020. *Investigación con estudio de caso*. 6.^aed. Madrid: Ediciones Morata, 2020.

YIN, R. K.

2004 *Estudo de caso: planejamento e métodos*. 2.^a ed. São Paulo: Bookman, 2004.

YIN, R. K.

2001 *Estudo de caso: planejamento e métodos*. São Paulo: Sage Publications, 2001.

Magda Valente Correia | magda.valente.correia@hotmail.com

Universidade Coimbra - Faculdade de Letras, Portugal

Liliana Isabel Esteves Gomes | liliana.gomes@fl.uc.pt

Universidade Coimbra - Faculdade de Letras / Centro de Estudos Interdisciplinares - CEIS20, Portugal