Inteligência artificial e riscos na manipulação da memória institucional e memória coletiva

Anabela Carvalho[1], Ana Fernandes[2], Carla Araújo[3], Duarte Moreira[4], Maria Caetano[5]

A.P. Camarinha[6]

[123456]*Instituto Superior de Contabilidade e Administração do Porto*

[6]*CEOS - PP*

Resumo

A inteligência artificial (IA) tem desempenhado um papel importante na gestão, preservação e interpretação da memória institucional e coletiva. A sua adoção em processos de organização, recuperação e análise de grandes volumes de dados permite novas formas de acesso ao património informacional. Paralelamente, levanta riscos associados à manipulação digital, à criação de conteúdos sintéticos e à disseminação de desinformação. O presente artigo analisa estes impactos a partir de uma revisão sistemática e interpretativa da literatura. Complementarmente, propõe um modelo concetual que explica a relação entre a mediação algorítmica, a integridade informacional e a vulnerabilidade da memória partilhada. Os resultados evidenciam tanto o potencial da IA na preservação histórica como os desafios éticos que ameaçam a autenticidade, a confiança pública e a coesão social. Sugere-se um conjunto de boas práticas aplicáveis a contextos arquivísticos e bibliotecários, reforçando a necessidade de políticas de transparência, literacia digital e avaliação crítica de sistemas automatizados.

Palavras-chave: Inteligência Artificial; Memória Institucional; Memória Coletiva; Desinformação; Preservação Digital.

1. Introdução

A disseminação da IA nos contextos organizacionais e sociais transformou os processos de produção, preservação e circulação de informação. Ferramentas como a aprendizagem automática, modelos generativos e sistemas de recomendação tornam possível analisar grandes volumes de dados, agilizando tarefas antes manuais e demoradas. Contudo, estes mesmos mecanismos aumentam a capacidade de fabricar conteúdos falsos, de manipular narrativas históricas e de fragilizar a confiança pública.

No domínio das Ciências da Informação, estas mudanças articulam-se diretamente com os desafios discutidos no XIX Encontro CTDI, nomeadamente a preservação digital, a curadoria de conteúdos, a ética da informação e a mediação tecnológica em ambientes arquivísticos e bibliotecários. Este artigo examina estas dinâmicas, propondo uma visão crítica sobre o equilíbrio entre inovação tecnológica e salvaguarda da integridade da memória coletiva.

2. Questão de investigação

Este estudo procura compreender de que forma a IA está a transformar os processos de preservação, organização e utilização da memória institucional e coletiva. A investigação centra-se em dois tópicos principais:

1. De que forma as tecnologias emergentes facilitam o acesso, a gestão e valorização de acervos históricos?
2. Como os sistemas automatizados contribuem para a manipulação informacional, distorção narrativa e a disseminação de desinformação?

Para além disso, problematiza-se a forma como os mecanismos algorítmicos influenciam as práticas profissionais em bibliotecas e arquivos, impactando decisões de curadoria digital, avaliação de autenticidade e mediação informacional.

3. Metodologia

A pesquisa apresenta uma abordagem qualitativa sustentada numa revisão sistemática e interpretativa da literatura (Snyder, 2019). Foram consultadas as bases científicas Google Académico e SciELO, utilizando descritores como “artificial intelligence” (Inteligência Artificial), “collective memory” (Memória Coletiva), “digital preservation” (Preservação Digital), “algorithmic mediation” (Medição Algorítmica).

Definiram-se critérios de inclusão como, pertinência temática, publicações entre 2018/2025, revisão por pares e contributo para a compreensão concetual do fenómeno. Consideraram-se os seguintes critérios de exclusão, textos opinativos, não científicos ou não relacionados com as práticas informacionais.

Após a seleção, procedeu-se a uma análise temática (Braun & Clarke, 2021), identificando três categorias principais: (1) mediação algorítmica e memória; (2) riscos de manipulação e desinformação; (3) integridade e ética na preservação digital. A construção destas categorias resultou de um processo iterativo de leitura, codificação aberta e reorganização concetual. Inicialmente, foram extraídas unidades de significado relacionadas com a interação entre tecnologias de IA e processos de preservação informacional. Estas unidades foram agrupadas em códigos preliminares, posteriormente melhorados à medida que se aprofundava a compreensão das convergências e divergências observadas nas fontes.

Na primeira categoria, “mediação algorítmica e memória”, concentraram-se os estudos que analisam como os sistemas de IA influenciam a seleção, organização e a visibilidade da informação, afetando o modo como as instituições e comunidades constroem e atualizam memórias partilhadas. Esta categoria evidencia o papel dos algoritmos como agentes ativos na formação de narrativas históricas, destacando conceitos como filtragem automatizada, priorização de conteúdos e mediação semântica.

A segunda categoria, “riscos de manipulação e desinformação”, reuniu contributos que discutem mecanismos pelos quais os modelos generativos e infraestruturas digitais desenvolvem a criação de conteúdos falsos, narrativas sintéticas e adulteração documental. Nesta categoria surgem temas como *deepfakes*[1], publicidade algorítmica, erosão de confiança pública e impactos sociocognitivos da desinformação desenvolvidas pela IA.

A terceira categoria, “integridade e ética na preservação digital”, envolveu os estudos dedicados à autenticidade de registos, à fiabilidade dos metadados, à gestão de proveniência e às implicações éticas na utilização da IA em acervos institucionais. Inclui ainda reflexões sobre responsabilidade organizacional, políticas de transparência e desafios associados à preservação a longo prazo num ecossistema informacional tecnicamente vulnerável e socialmente sensível.

Estas três categorias permitiram estruturar uma leitura integrada da literatura, oferecendo um quadro analítico capaz de explicar as tensões entre inovação tecnológica, riscos informacionais e exigências éticas na gestão da memória institucional e coletiva.

4. Revisão da literatura

A revisão da literatura permitiu mapear um conjunto amplo de contributos que abordam a intersecção entre inteligência artificial, memória e desinformação. No entanto, para tornar essa leitura mais sistemática e compreensível, torna-se necessário organizar estes contributos de acordo com as dimensões concetuais que estruturam o fenómeno estudado. Assim, as secções seguintes aprofundam três eixos fundamentais identificados na análise temática: a forma como os sistemas algorítmicos intervêm nos processos de construção e recuperação da memória; os riscos emergentes associados à manipulação digital e à difusão de conteúdos falsos; e, por fim, os desafios éticos que moldam a preservação digital num contexto cada vez mais automatizado. Esta organização permite não apenas clarificar o estado da arte, mas também estabelecer uma base sólida para a discussão crítica desenvolvida nas secções seguintes.

4.1 Inteligência Artificial: capacidades e desafios

A inteligência artificial evoluiu para além dos modelos simbólicos e sistemas especialistas que marcaram as primeiras décadas da área. Russell e Norvig (2021) descrevem-na como o esforço para construir sistemas capazes de percecionar o ambiente, raciocinar sobre estados possíveis e tomar decisões orientadas para objetivos. Esta visão abrange desde algoritmos de

[1] Uma das formas mais eficazes de enganar, ao colocar, em vídeo, pessoas a exprimirem palavras que nunca disseram, ou mesmo substituir caras, criando, assim, situações falsas.

aprendizagem supervisionada e não supervisionada até redes neuronais profundas, capazes de extrair padrões complexos a partir de grandes volumes de dados.

A emergência da IA generativa, impulsionada por modelos de linguagem de larga escala e difusão generativa, transformou a capacidade das máquinas de produzirem conteúdos originais, incluindo textos, imagens, áudio e vídeo. Essa evolução aumentou a eficiência de processos informacionais que dependem de categorização, sumarização ou tradução automática (Floridi, 2022), mas ampliou também os riscos éticos relacionados com a densidade algorítmica, autoria, direitos de reprodução e fiabilidade documentária.

Estudos recentes apontam que a produção massiva de conteúdos sintéticos desafia profundamente critérios tradicionais de autenticidade (Koltay, 2023). Por exemplo, as fronteiras entre documento original e documento criado podem tornar-se indistintas, obrigando instituições de memória a desenvolver novas estratégias de análise forense, verificação e descrição arquivística. Stokel-Walker (2023) salienta que bibliotecas e arquivos enfrentam uma pressão crescente para atualizar práticas de curadoria digital, já que a proliferação de conteúdos gerados por IA ameaça comprometer a confiança nas coleções e nos processos de preservação.

4.2 Desinformação, manipulação algorítmica e risco social

A desinformação digital tornou-se um fenómeno estrutural das sociedades atuais, alimentado tanto por dinâmicas económicas das plataformas como por estratégias de manipulação política e ideológica. Wardle e Derakhshan (2017) demonstram que as infraestruturas algorítmicas, motores de recomendação, priorização automatizada maximizam a circulação de mensagens polarizadoras, ao amplificar conteúdos com elevado potencial emocional.

A introdução de IA generativa intensifica esta problemática. Westerlund (2023) destaca que a criação automática de *deepfakes*, notícias fabricadas e testemunhos falsos acrescenta um nível de credibilidade que perturba os mecanismos tradicionais de verificação. Ao contrário de manipulações antigas, facilmente identificáveis, estas novas formas de falsificação são capazes de simular a textura visual, textual e discursiva de fontes genuínas.

No contexto arquivístico e informacional, as consequências são particularmente graves. A confiança pública na autenticidade de registos históricos depende da perceção de que os sistemas de preservação são capazes de distinguir o verdadeiro do falso. Zeng et al. (2023) mostram que a manipulação algorítmica introduz ruído e ambiguidade nos processos de validação documental, afetando a credibilidade de acervos e debilitando o papel das instituições enquanto guardiãs de memória.

4.3 Memória institucional e coletiva

A memória institucional constitui um ativo estratégico, permitindo às organizações compreenderem o seu passado, organizar e preservar os acervos institucionais (Santos & Valentim, 2021). Engloba documentos administrativos, registos históricos, práticas culturais e narrativas organizacionais que contribuem para a construção de um legado reconhecido interna

e externamente. Preservar esta memória é essencial não apenas para fins administrativos, mas também para garantir transparência e continuidade histórica.

Já a memória coletiva, tal como formulada por Halbwachs (1992), deriva de uma dimensão social e intersubjetiva. Os grupos constroem memórias partilhadas através de quadros interpretativos comuns, mantendo narrativas que fundamentam identidades e valores. Este processo é dinâmico, sempre sujeito a reajuste e reinterpretação.

Na era digital, estas duas formas de memória enfrentam desafios semelhantes, sobretudo no que respeita à integridade e à autenticidade dos acervos. Conway (2020) sublinha que a preservação digital envolve garantir que os objetos digitais permanecem acessíveis, íntegros e interpretáveis ao longo do tempo, independentemente das mudanças tecnológicas. Pasquini (2024) reforça que práticas adequadas de metadados, versionamento, auditoria de proveniência e curadoria contínua são indispensáveis para evitar perda de contexto e assegurar a confiança nos registos institucionalizados.

4.4 Responsabilidade institucional e preservação digital

Para além da preservação técnica, as instituições enfrentam responsabilidades éticas, sociais e cívicas na gestão de arquivos (Jaillant, 2022). A dimensão ESG (Environmental, Social and Governance) tem reforçado a visibilidade pública da preservação documental como parte da responsabilidade social (MacNeil, 2020).

O papel das bibliotecas e arquivos passa também por assegurar políticas transparentes de curadoria digital, avaliando criticamente sistemas automatizados e seus impactos socioinstitucionais.

A responsabilidade das instituições de memória vai muito além da conservação técnica de documentos. Iniciativas globais em torno de ética da informação, diversidade, sustentabilidade e governação transformaram a preservação documental num elemento central das políticas organizacionais. Jaillant (2022) mostra que as instituições arquivísticas desempenham um papel crucial na mediação entre tecnologia, sociedade e cidadania, garantindo que as práticas de preservação respeitam simultaneamente direitos humanos, valores democráticos e princípios de acesso imparcial.

Debates sobre a ESG, destacam a preservação documental como parte integrante da responsabilidade social das organizações (MacNeil, 2020). A forma como as instituições tratam, disponibilizam e interpretam os seus acervos influencia diretamente a confiança pública e o compromisso social.

Nesse sentido, bibliotecas e arquivos devem adotar políticas transparentes de curadoria digital, assegurando que as ferramentas algorítmicas utilizadas em processos de classificação, recuperação ou mediação informacional são rigorosamente avaliadas quanto a orientações, impactos sociais e riscos éticos. O papel destes profissionais passa também por promover

literacia digital crítica, capacitando utilizadores para compreenderem e questionarem resultados produzidos por sistemas de IA.

4.5 IA, desinformação e fragilidade da memória coletiva

A interação entre IA e memória coletiva coloca desafios simultaneamente sociocognitivos e epistemológicos[2]. Almeida & Almeida (2024) mostram que os modelos generativos permitem criar "memórias artificiais", i.e., reconstruções plausíveis, mas falsas, de eventos e testemunhos. Estas narrativas sintéticas podem interferir na formação de recordações sociais, influenciando a forma como grupos interpretam o passado e constroem consensos identitários.

A vulnerabilidade aumenta quando os algoritmos passam a mediar o que é selecionado, destacado ou esquecido em ambientes digitais. A memória coletiva, já por natureza dependente de processos de seleção, torna-se mais frágil quando essas seleções são delegadas a sistemas opacos que privilegiam métricas de relevância ou envolvimento, e não critérios éticos ou históricos.

Por outro lado, a IA também pode desempenhar um papel positivo na preservação histórica. Jaillant (2022) documenta projetos de reconstrução de narrativas históricas que utilizam IA para reconstituir documentos danificados, ampliar testemunhos orais ou criar visualizações que ajudam a preservar memórias ameaçadas. Contudo, estes benefícios só se materializam quando acompanhados de políticas éticas rigorosas, avaliação crítica dos modelos e literacia digital capaz de distinguir apoio tecnológico de manipulação (inadvertida).

5. Modelo concetual proposto

Propõe-se um modelo composto por três eixos interdependentes que, em conjunto, permitem compreender de forma integrada o impacto da inteligência artificial na preservação da memória institucional e coletiva.

O primeiro eixo, Mediação Algorítmica, refere-se aos mecanismos através dos quais os sistemas de IA selecionam, filtram e produzem informação, influenciando diretamente o que se torna visível, acessível e valorizado nos ecossistemas digitais.

O segundo eixo, Integridade Informacional, diz respeito ao grau de autenticidade, fiabilidade e preservação dos registos, constituindo o núcleo de confiança sobre o qual assentam as práticas arquivísticas e biblioteconómicas.

Por fim, o eixo da Vulnerabilidade da Memória traduz a sensibilidade das narrativas históricas à manipulação digital, especialmente num contexto em que conteúdos sintéticos se tornam cada vez mais indistinguíveis de documentos genuínos.

A dinâmica entre estes três eixos é contínua e profundamente interligada. A forma como os algoritmos intervêm na informação afeta diretamente a integridade dos registos, esta, por sua

[2] Área da filosofia que estuda a teoria do conhecimento

vez, condiciona o grau de resistência da memória coletiva a distorções. Por conseguinte, é o equilíbrio entre mediação, integridade e vulnerabilidade que determina a força das práticas de preservação digital e a resiliência das narrativas históricas partilhadas, garantindo que as instituições de memória permanecem capazes de proteger e transmitir patrimónios informacionais de forma segura, crítica e responsável.

6. Implicações práticas para bibliotecas e arquivos

Com base na literatura e no modelo concetual, identificam-se implicações diretas para profissionais de informação:

- necessidade de auditoria algorítmica em sistemas de pesquisa e recomendação;
- desenvolvimento de literacia crítica em IA para equipas técnicas e público;
- integração de metadados para mitigar conteúdos sintéticos;
- adoção de políticas de transparência a utilização de modelos generativos em curadoria.

A crescente dependência de mecanismos automatizados para indexação, recuperação e recomendação implica que bibliotecas e arquivos avaliem criticamente o comportamento dos algoritmos que utilizam. Como demonstram Koltay (2023) e Stokel-Walker (2023), a mediação algorítmica pode introduzir orientações na visibilidade dos conteúdos, privilegiando certos tipos de documentos em detrimento de outros.

Neste sentido, as auditorias algorítmicas regulares são essenciais para garantir que os sistemas não reforçam desigualdades informacionais ou favorecem narrativas artificiais produzidas por IA generativa. Estas auditorias devem incluir a verificação de parâmetros de priorização, avaliação de resultados distorcidos e análise de coerência com os critérios profissionais de descrição e classificação.

A vulnerabilidade da memória coletiva, amplificada por deepfakes e conteúdos sintéticos (Westerlund, 2023), exige que os profissionais da informação como bibliotecários e arquivistas desenvolvam competências para identificar e interpretar materiais produzidos por IA. Isto implica formação em metodologias de verificação digital, compreensão de modelos generativos e leitura crítica de evidências audiovisuais.

Para o público utilizador, esta literacia é igualmente fundamental, porque se o acesso à informação é mediado por algoritmos, também a compreensão do modo como esses algoritmos funcionam deve ser mediada por essa mesma literacia. Assim, iniciativas de formação, oficinas temáticas e recursos educativos podem fortalecer a capacidade coletiva de distinguir informação autêntica de conteúdos manipulados.

A integridade informacional depende fortemente da capacidade de documentar a origem, contexto e transformação dos objetos digitais (Conway, 2020; Pasquini, 2024). Face à proliferação de conteúdos artificiais, a documentação de proveniência torna-se um mecanismo

central de defesa.
Neste sentido as instituições de memória devem reforçar os metadados que descrevam:

- processos de produção dos documentos,
- ferramentas utilizadas,
- eventuais tratamentos algorítmicos,
- versões anteriores ou rastros de edição.

Estes elementos permitem identificar anomalias, rastrear adulterações e distinguir documentos autênticos de produtos sintéticos, contribuindo de forma direta para a resiliência da memória institucional e coletiva.

A literatura aponta para a necessidade crescente de transparência institucional (Jaillant, 2022; MacNeil, 2020). Quando os serviços de informação utilizam a IA para apoiar processos de descrição, sumarização ou classificação, é fundamental que esses procedimentos sejam comunicados aos utilizadores.
Neste sentido, políticas transparentes devem incluir:

- identificação explícita dos processos automatizados utilizados;
- esclarecimento dos limites e potenciais orientações dos modelos aplicados;
- mecanismos de revisão humana para prevenir erros algorítmicos;
- documentação pública das decisões curatoriais mediadas por IA.

Estas práticas reforçam a confiança pública e mitigam o risco de que a mediação algorítmica introduza fragilidades na integridade dos acervos ou na perceção da sua autenticidade.

7. Conclusão

A inteligência artificial está a redefinir, de forma estrutural, a forma como instituições e sociedades preservam, organizam e reconstroem as suas memórias. O seu contributo para a gestão informacional, desde a automatização de processos de classificação à recuperação avançada de conteúdos, representa uma oportunidade significativa para valorizar acervos, ampliar o acesso e fortalecer a continuidade histórica das organizações. Contudo, este potencial convive com riscos igualmente profundos, nomeadamente a proliferação de conteúdos sintéticos, a opacidade algorítmica e a crescente vulnerabilidade das narrativas coletivas à manipulação digital. Estes desafios colocam em causa princípios fundamentais como autenticidade, fiabilidade e confiança pública, pilares essenciais da preservação documental e da memória social.

Os resultados discutidos evidenciam que a adoção responsável da IA não pode limitar-se à implementação tecnológica. Exige políticas éticas, capazes de orientar a utilização de modelos generativos e sistemas de mediação algorítmica, requer literacia digital crítica, tanto para

profissionais de informação quanto para utilizadores, de modo a promover uma compreensão informada dos riscos e limitações associados a estes sistemas, e implica mecanismos permanentes de avaliação, incluindo auditorias, documentação de proveniência e processos de validação humana que complementem a ação das ferramentas automatizadas.

Assim, a preservação digital deve ser simultaneamente técnica e reflexiva, apoiada em decisões participadas que integrem dimensões sociais, culturais e éticas. As instituições de memória assumem um papel central perante a integridade informacional. Proteger a memória coletiva como bem público implica reconhecer que, na era da inteligência artificial, preservar não é apenas conservar ficheiros, é também garantir que as narrativas que nos definem enquanto comunidades permanecem confiáveis, verificáveis e resistentes a distorções.

Referências bibliográficas

Almeida, F. O. F., & Almeida, W. C. (2024). *Educação pela memória: A IA redefinirá a verdade histórica dos fatos?* Revista Interdisciplinar de Estudos, 10(3), 1–20.

Braun, V., & Clarke, V. (2021). *Thematic Analysis: A Practical Guide.* SAGE.

Conway, P. (2020). *Digital transformations and the archival mission.* American Archivist, 83(2), 362–401.

Floridi, L. (2022). *Ethics, Governance, and AI.* Oxford University Press.

Halbwachs, M. (1992). *On Collective Memory.* University of Chicago Press.

Jaillant, L. (Ed.). (2022). *Archives, Access and Artificial Intelligence.* Bielefeld University Press.

Koltay, T. (2023). *Fake news, disinformation and information literacy in the age of AI.* Journal of Documentation, 79(1), 1–16.

MacNeil, H. (2020). *Trustworthiness, evidence and identity in digital records.* Archival Science, 20(1), 1–25.

Pasquini, V. (2024). *Digital curation and long-term digital preservation in libraries.* JLIS.it, 15(1), 109–125.

Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.

Santos, J. C., & Valentim, M. L. P. (2021). *Memória institucional e memória organizacional.* Perspectivas em Ciência da Informação, 26, 1–20.

Snyder, H. (2019). Literature review as a research methodology. *Journal of Business Research*, 104, 333–339.

Stokel-Walker, C. (2023). *Machine generated texts and the future of credibility.* Nature, 615, 10–12.

Wardle, C., & Derakhshan, H. (2017). *Information Disorder.* Council of Europe.

Westerlund, M. (2023). *Deepfakes and AI-generated disinformation.* Technology Innovation Management Review, 13(3), 5–12.

Zeng, M., Shen, Y., & Chen, B. (2023). *AI-generated content and archival authenticity.* Archival Science, 23, 421–441.