

APLICAÇÕES DE INTELIGÊNCIA ARTIFICIAL NA GESTÃO DOCUMENTAL: REVISÃO SISTEMÁTICA NA LISTA, *EMERALD* E *SCOPUS*

SÂNDERSON LOPES DORNELES*

INTRODUÇÃO

O início do século XXI foi marcado pelo uso em massa de diversas aplicações computacionais em diversas áreas do conhecimento humano. Como resultado deste uso, foram produzidas e acumuladas informações que precisam ser gerenciadas em pontos como: sua produção, à destinação (eliminação ou armazenamento dos estoques informacionais) e ao uso.

A grande produção de informação requer tecnologias que potencializem as atividades de gestão organizacional, principalmente para o acesso rápido e facilitado, a fim de servir como suporte para a tomada de decisão. Dentre o rol das soluções tecnológicas que possam automatizar os processos de gestão documental, tem-se dado evidência para aqueles que se utilizam da Inteligência Artificial (IA). Segundo Bunn (2020):

Inteligência artificial (IA) não é um tópico novo na imaginação humana, embora muitos datem suas origens como um fenômeno moderno em 1956 e em um workshop realizado no Dartmouth College, nos Estados Unidos. Desde aquela época, a inteligência artificial passou por muitas fases de desenvolvimento (Bunn 2020, p. 143)¹.

Diante dessas fases de evoluções, as aplicações de tecnologias da IA estão cada vez mais presentes no mundo dos negócios, dos governos e de áreas específicas, tais como saúde, jurídica, gestão de pessoas entre outras, dada as possibilidades profícuas dessas tecnologias. Impactando, principalmente, em sistemas computacionais como suporte técnico à produção, ao uso e ao armazenamento de informações e a potencialização de habilidades humanas por meio de máquinas. O que se torna pauta emergente de discussões, de análises e de reflexões na área de Ciência da Informação (CI).

Sobre estes aspectos, Rolan et al. (2019) argumentam e definem que:

* Instituto Federal da Paraíba (IFPB); Universidade Estadual da Paraíba (UEPB); ORCID: <https://orcid.org/0000-0002-3888-2841>; Email: sanderson.dorneles@gmail.com.

¹ Tradução livre de: «Artificial intelligence (AI) is not a new topic in human imagination, although many date its origins as a modern phenomenon to 1956 and a workshop held at Dartmouth College, USA. Since that time, artificial intelligence has gone through many phases of development» (Bunn 2020, p. 143).

A inteligência artificial evoluiu desde os primeiros dias da computação (e, de fato, desde antes, se alguém tiver uma visão mais ampla das tecnologias que ajudam a automatizar o trabalho do conhecimento humano) [...] entendemos IA como envolvendo sistemas digitais que automatizam ou auxiliam em ‘atividades que associamos ao pensamento humano, atividades como tomada de decisão, resolução de problemas, aprendizagem [e] criação. Na verdade, temos a visão ampla de que a IA se manifesta “em um espectro multidimensional [compreendendo] escala, velocidade, grau de autonomia e generalidade” — abrangendo, assim, uma gama de técnicas e tecnologias de automação (Rolan et al. 2019, p. 3)².

Tais atividades, que podem ser automatizadas pelas IAs, fazem parte do escopo de estudo e atuação da Ciência da Informação (CI). No âmbito da CI, os estudos e as análises de fenômenos informacionais são desenvolvidos desde meados do século XX. Segundo Araújo (2013):

Em 1968, o American Documentation Institute, dos Estados Unidos, mudou seu nome para American Society for Information Science, tornando-se a primeira instituição de Ciência da Informação do mundo. Na época, muitos desconheciam o que vinha a ser “Ciência da Informação” e o que faziam os “cientistas da informação” (Araújo 2013, p. 2).

Desde o pós segunda guerra mundial até a contemporaneidade, a CI passa por diversos desafios e demandas para o tratamento de informações. Sobre estes aspectos, os avanços de tecnologias da informação exigem mais ações inovadoras das organizações para o processamento adequado dos seus capitais informacionais. De acordo com Neves (2020), a sociedade tem um conjunto de tecnologias disruptivas que possui um maestro, a inteligência artificial. E, essas tecnologias no campo da CI, aliadas ao uso de dispositivos inteligentes e da computação cognitiva nos espaços de construção do conhecimento, vem avançando paulatinamente, proporcionando novas formas de interação com os sujeitos.

Os sistemas informatizados de gestão revolucionaram as organizações com outras formas de produção, tramitação e acesso de seus documentos. A *internet* por sua vez encurtou as noções de tempo e espaço, agilizando de forma eficaz a transmissão de documentos.

² Tradução livre de: «Artificial intelligence has evolved since the earliest days of computing (and, in fact, from earlier, if one takes a broader view of technologies that help automate human knowledge work) [...] we understand AI as involving digital systems that automate or assist in ‘activities that we associate with human thinking, activities such as decision-making, problem solving, learning, [and] creating’. In fact, we take the broad view that AI manifests ‘on a multi-dimensional spectrum [comprising] scale, speed, degree of autonomy, and generality’ — thus encompassing a range of automation techniques and Technologies» (Rolan et al. 2019, p. 3).

Com o advento das tecnologias da inteligência artificial e suas potencialidades na aplicação em sistemas de informação. Surgem inquietações, como um alerta aos cientistas e aos usuários de informações, qual seja, como otimizar a produção, o uso contínuo de informações e as formas de preservação da memória acumulada em suportes eletrônicos por meio dessas tecnologias. Para tanto, deve-se procurar respostas a partir de produções científicas a respeito do assunto. Sendo assim, a partir de revisão sistemática de literatura em bases de dados internacionais, a presente pesquisa tem como objetivo identificar aplicações de Inteligência Artificial (IA) na gestão de documentos.

1. PROCEDIMENTOS METODOLÓGICOS

A presente pesquisa quanto à natureza é básica, de acordo com Silva e Menezes (2001, p. 20): «pesquisa básica objetiva gerar conhecimentos novos úteis para o avanço da ciência sem aplicação prática prevista. Envolve verdades e interesses universais». Para tanto, teve objetivo exploratório e descritivo. Segundo Gil (2008, p. 27): «Pesquisas exploratórias são desenvolvidas com o objetivo de proporcionar visão geral, de tipo aproximativo, acerca de determinado fato». Para a presente pesquisa a visão geral será realizada pela identificação de estudos sobre as tecnologias de IA e gestão de documentos, cujo ponto de aproximação é a aplicação dessas tecnologias nas operações da gestão de registros informacionais. Quanto ao objetivo de pesquisa descritiva: «As pesquisas deste tipo têm como objetivo primordial a descrição das características de determinada população ou fenômeno ou o estabelecimento de relações entre variáveis» (Gil 2008, p. 28). Enquadra-se neste tipo de pesquisa, haja vista que se descreve as relações das temáticas do estudo a partir das variáveis de suas publicações, tais como dados bibliométricos e altimétricos, bem como de categorizações das áreas do conhecimento e tecnologias que aplicam IA na gestão documental.

Para a abordagem do problema, foram utilizados os métodos quantitativo e qualitativo. Conforme Richardson (2012):

O método quantitativo, como o próprio nome indica, caracteriza-se pelo emprego da quantificação tanto nas modalidades de coleta de informações, quanto no tratamento delas por meio de técnicas estatísticas, desde as mais simples como percentual, média, desvio-padrão, às mais complexas, como coeficiente de correlação, análise de regressão etc. (Richardson 2012, p. 27).

Sobre a quantificação nesta pesquisa, apenas foram mensuradas informações a respeito do número de publicações sobre a temática do estudo, as palavras-chave mais frequentes utilizadas nos trabalhos coletados, as comunicações científicas mais citadas, os autores com maior número de publicações, assim como dados altimétricos desses autores

mais produtivos. Não sendo necessário a aplicação de técnicas estatísticas para o tratamento das informações coletadas.

No que se refere ao emprego do método qualitativo: «Os estudos que empregam uma metodologia qualitativa podem descrever a complexidade de determinado problema, analisar a interação de certas variáveis» (Richardson 2012, p. 80). Nesse sentido, a análise e descrição das variáveis são sobre os tipos de tecnologias de IA aplicadas na gestão de documentos, das áreas do conhecimento, retratando os objetivos e resultados dessas aplicações.

Quanto às técnicas de levantamento de dados, foi realizada uma revisão sistemática de literatura. Nesse sentido, o protocolo da revisão seguiu as recomendações PRISMA (Principais Itens para Relatar Revisões Sistemáticas e Meta-análises), de acordo com Galvão, Pansani e Harrad (2015, p. 336): «A recomendação PRISMA consiste em um *checklist* com 27 itens e um fluxograma de quatro etapas».

As estratégias de buscas para a recuperação dos documentos, consistiu na utilização dos termos *artificial intelligence and records management*, realizado no dia 7 de setembro de 2021.

As bases de dados utilizadas foram a *Library, Information Science & Technology Abstracts* (LISTA), que consiste em uma base de dados da CI, a *Emerald Insight*, que indexa artigos da área de gestão, e a *Scopus*, que é uma plataforma multidisciplinar e bem difundida nas comunidades acadêmicas internacionais.

No que diz respeito aos critérios de exclusão e inclusão, bem como os dados extraídos. Foram estabelecidos os seguintes requisitos:

Critérios de Exclusão

- Textos não disponibilizados por completo pelo acesso do portal de periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).
- Estudos que não contemplam uma abordagem a respeito da IA e a gestão de documentos, que para este estudo entende-se como uma abordagem de aplicações de IA e tecnologias associadas na produção, armazenamento e uso de registros informacionais.

Critérios de Inclusão

- Tipo de documento: artigos de periódicos, trabalhos publicados em eventos, livros e capítulos de livros.
- Período de tempo: 2016 – setembro/2021.
- Idioma: inglês

Dados Extraídos

Autores, citações, títulos dos artigos, anos da publicação, palavras-chave, instituições dos autores, periódico de publicação, idioma, e área do conhecimento.

Para visualização do fluxo da pesquisa e dos resultados, segue Figura 1.

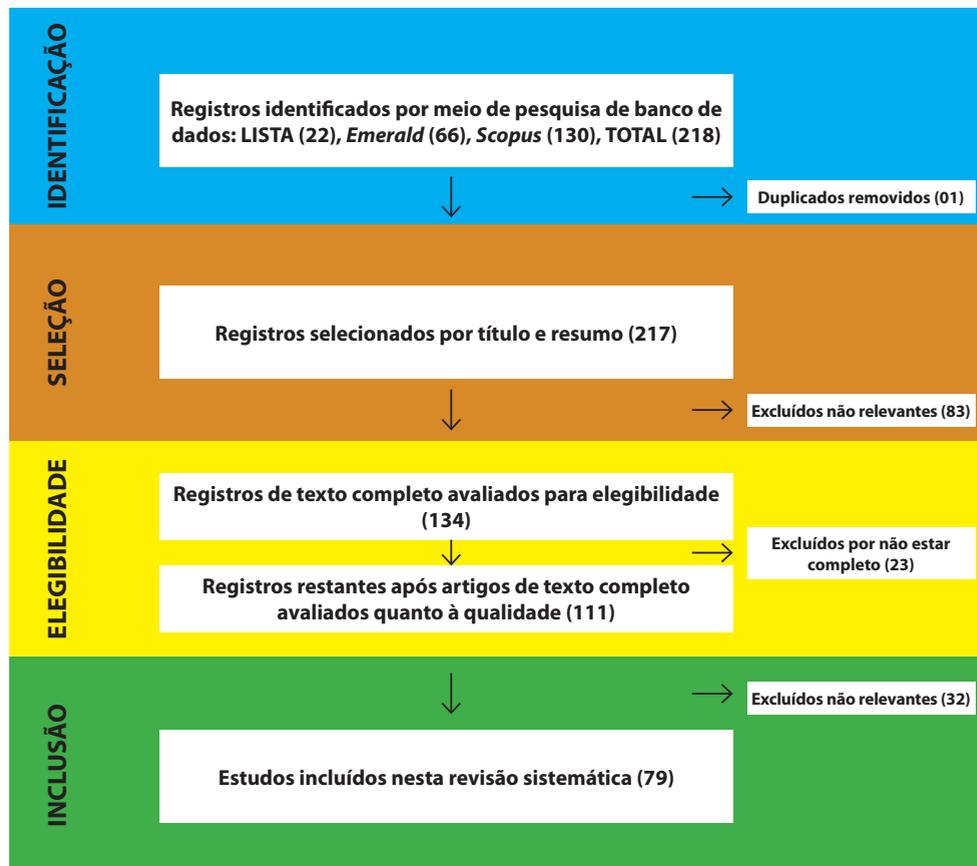


Fig. 1. Fluxograma PRISMA para a revisão sistemática de aplicações de inteligência artificial na gestão de documentos
 Fonte: Fluxograma PRISMA adaptado de Galvão, Pansani e Harrad (2015) para a revisão sistemática de aplicações de inteligência artificial na gestão de documentos

Sobre as 79 publicações selecionadas, para as análises quantitativas foi utilizado *software* de gestão de referências *Mendeley* para a organização e análises dos metadados de cada publicação, tais como autores, tipo de publicação, periódico, ano de publicação e resumo, e de análise bibliométrica por meio do *software* *VOSViewer* para identificação de palavras-chave mais frequentes e relação de co-autoria nos trabalhos, bem como coleta de dados altimétricos de número de citações de publicações e autores, leituras e recomendações a partir da rede acadêmica *ResearchGate* e da base de dados *Scopus*, enquanto as análises qualitativas foram realizadas com auxílio do *software* *NVivo* para a categorização

Tabela 1. Títulos, periódicos e ano dos 10 artigos mais citados

	Título do documento	Periódico	Ano	N.º de citações
1.	Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis	IEEE Journal of Biomedical and Health Informatics	2018	293
2.	Clinical information extraction applications: A literature review	Journal of Biomedical Informatics	2018	203
3.	A machine learning-based framework to identify type 2 diabetes through electronic health records	International Journal of Medical Informatics	2017	121
4.	A distributed ensemble approach for mining healthcare data under privacy constraints	Information Sciences	2016	46
5.	Electronic health record phenotyping improves detection and screening of type 2 diabetes in the general United States population: A cross-sectional, unselected, retrospective study	Journal of Biomedical Informatics	2016	44
6.	A survey of clinical phenotyping in selected national networks: demonstrating the need for high-throughput, portable, and computational methods	Artificial Intelligence in Medicine	2016	42
7.	Support Vector Feature Selection for Early Detection of Anastomosis Leakage from Bag-of-Words in Electronic Health Records	IEEE Journal of Biomedical and Health Informatics	2016	38
8.	Natural Language Processing for EHR-Based Computational Phenotyping	IEEE/ACM Transactions on Computational Biology and Bioinformatics	2019	32
9.	Predicting colorectal surgical complications using heterogeneous clinical data and kernel methods	Journal of Biomedical Informatics	2016	31
10.	Classification of hospital acquired complications using temporal clinical information from a large electronic health record	Journal of Biomedical Informatics	2016	24

Fontes: Extração de dados das publicações em 07/09/2021

Os dados sobre as publicações mais referenciadas confirmam destaque especial para a área de saúde no uso de tecnologias de IA e o aplicações delas para otimizar as análises de dados clínicos com fim de tratamento de doenças.

Na Figura 3 são quantificados os autores mais produtivos no escopo desta pesquisa, assim como são demarcadas as conexões desses autores.

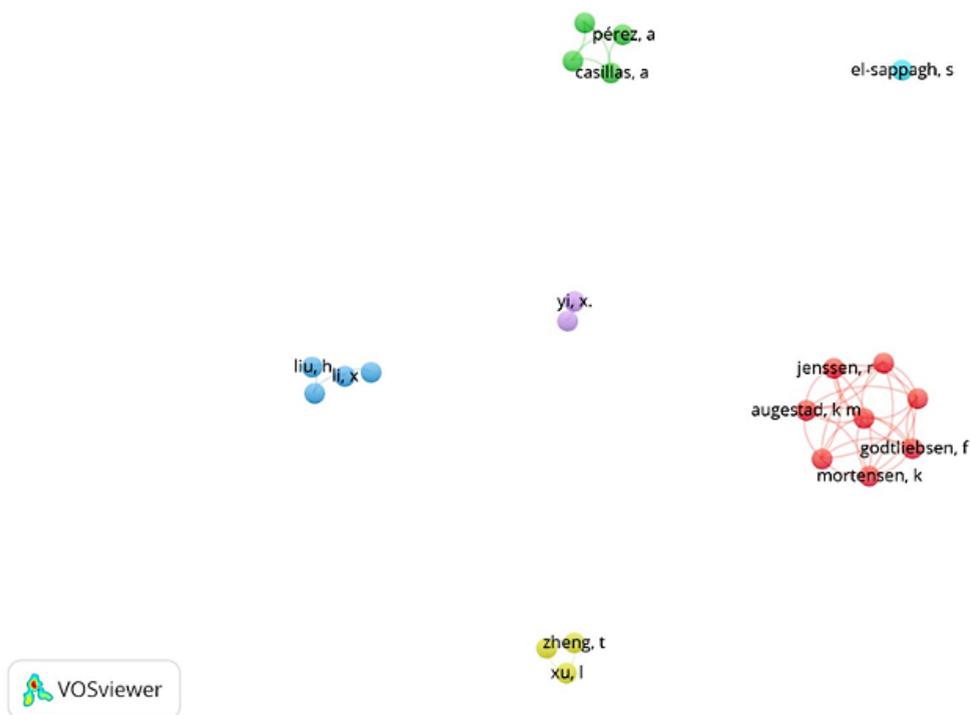


Fig. 3. Relação de coautoria (mínimo de duas publicações por autor)

Fonte: Dados da pesquisa (2021) analisados no VOSviewer

Da Figura 3 observamos que existem grupos de autores isolados, não há conexões entre eles. Para compreensão dos grupos de autores, a Tabela 2 identifica a filiação institucional/nacionalidade, departamento/área de atuação e o número de citações de cada um.

Tabela 2. Índices de citações para autores que mais publicaram (mínimo de duas publicações por autor)

Nome do autor	Instituição/nacionalidade	Depto./área atuação	Citações	Fonte
1) Arthur Revhaug	Hospital Universitário do Norte da Noruega	Medicina	9.528	ResearchGate
2) Ya Zhang	Universidade Jiao Tong de Xangai/China	Escola de Ciência e Engenharia de Materiais	3.852	ResearchGate
3) Fred Godtlielsen	UiT Universidade Ártica da Noruega	Departamento de Matemática e Estatística	2.658	ResearchGate
4) Robert Jenssen	UiT Universidade Ártica da Noruega	Departamento de Física e Tecnologia, Grupo de Engenharia Elétrica	2.644	ResearchGate
5) nut Magne Augestad	Universidade de Oslo/Noruega	Departamento de Cirurgia Campus AHUS	2.276	ResearchGate e Scopus preview
6) Rolv-Ole Lindsetmo	Hospital Universitário do Norte da Noruega	Departamento de Cirurgia Gastroenterológica	1.379	ResearchGate
7) Shaker El-Sappagh	Inha University/Coreia do Sul	Departamento de Engenharia da Informação e Comunicação	1.265	ResearchGate
8) Tristan Naumann	Instituto de Tecnologia de Massachusetts/Estados Unidos	Laboratório de Ciência da Computação e Inteligência Artificial	960	ResearchGate
9) Kim Mortensen	Hospital Universitário do Norte da Noruega	Departamento de Cirurgia Gastroenterológica	713	ResearchGate
10) Arantza Casillas	Universidad del País Vasco / Euskal Herriko Unibertsitatea/Espanha	Departamento de Electricidad y Electrónica	507	ResearchGate
11) Maite Oronoz	Universidad del País Vasco / Euskal Herriko Unibertsitatea/Espanha	Linguagens e sistemas de computador	503	ResearchGate
12) Kristian Hindberg	UiT Universidade Ártica da Noruega	Departamento de Física e Tecnologia	470	ResearchGate

(continua na página seguinte)

Nome do autor	Instituição/nacionalidade	Depto./área atuação	Citações	Fonte
13) Alicia Pérez	Universidad del País Vasco Euskal Herriko Unibertsitatea/ Espanha	EU de Enfermería de Donostia-San Sebastián	302	ResearchGate e Scopus preview
14) Junmei Zhong	Inspur USA Inc Estados Unidos	***	215	ResearchGate
15) Tao Zheng	Institute of Image Communication and Networking, Shanghai Jiao Tong University, Shanghai/China	***	176	Scopus preview
16) Sara Santiso	Universidad del País Vasco Euskal Herriko Unibertsitatea/ Espanha	***	71	Scopus preview
17) Xiu Yi	Software Business Group, Inspur Inc/China	***	2	Scopus preview

Fontes: Extração de dados das fontes referidas, em 30/10/2021

Dos dados apresentados, verifica-se um grupo de 8 (oito) autores filiados a instituições norueguesas (identificados pelos números 1, 2, 3, 4, 5, 6, 9 e 12) de equipe multidisciplinar de profissionais da medicina, física, engenharia, estatística, matemática e tecnologia com números de citações expressivos. Como também, destaca-se a presença de 4 (quatro) autores chineses filiados a diferentes instituições com habilidades na área de IA (identificados pelos números 2, 14, 15 e 17).

Na Tabela 3 são apresentados dados altimétricos complementares sobre os autores referenciados na Tabela 2. Nesta Tabela 3 são divulgados os dados dos 10 autores com os maiores números de citações.

Tabela 3. Altimétrica dos 10 autores com mais citações

Autores	Score RG	Citation	Reads	Recommendation
Arthur Revhaug	42,82	9528	12 538	8
Ya Zhang	30,82	3852	15 913	12
Fred Godtliebsen	36,16	2658	53 965	22
Robert Jenssen	33,41	2644	34 084	84
Knut Magne Augestad	34,41	1711	19 237	197
Rolv-Ole Lindsetmo	31,77	1379	8516	3

(continua na página seguinte)

Autores	Score RG	Citation	Reads	Recommendation
Shaker El-Sappagh	31,79	1265	51 547	205
Tristan Naumann	13,46	960	4125	7
Kim Mortensen	22	713	3324	2
Arantza Casillas	24,17	507	4824	3

Fontes: Extração de dados do *ResearchGate* em 31/10/2021

Sobre as áreas do conhecimento, seja a respeito de áreas mais abrangentes e específicas dos saberes humanos, das 79 publicações incluídas para as análises deve-se atentar para o fato de que algumas publicações se utilizaram de aspectos teóricos e conceituais de mais de uma área do conhecimento. Sendo assim, a Tabela 4 apresenta os resultados dessa categorização.

Tabela 4. Áreas do conhecimento identificadas

Área do Conhecimento	Número de publicações
Saúde	66
Arquivologia	7
Ontologia	5
Accountability (Auditoria)	3
Sistemas de informação pervasivos	1
Megaprojeto	1
Jurídica	1
Diplomática	1
Biblioteconomia	1

Dos 66 trabalhos da saúde, 63 tratam de aplicações de tecnologias da IA em registros de saúde, cujos objetivos e resultados buscam extrair de *Electronic Health Record* (EHR) — em tradução ao português Registro Eletrônico de Saúde — benefícios para a recuperação rápida de dados e informações dos pacientes. Além disso, destaca-se a utilização dessas tecnologias para a produção de conhecimento preditivo (17 trabalhos) sobre determinadas doenças (com destaque para a diabetes tipo 2 que apareceu em mais estudos) a partir de dados e algoritmos computacionais.

Sobre sistemas de registros eletrônicos de saúde, Masuda, Shepard, Yamamoto e Toma (Masuda et al. 2019) relatam que os primeiros surgem na década de 1960, com a

utilização de aplicativos de *software* e sistemas de computador autônomos que migraram da documentação em papel dos dados do paciente para formas digitais de manutenção de registros. Dentre as informações armazenadas por esses sistemas, os autores citam os seguintes: histórico médico anterior e medicamentos, imunizações, dados laboratoriais, relatórios de radiologia, sinais vitais bem como dados demográficos do paciente.

Corroborando com a referência anterior a respeito de EHR, Caroprese, Veltri, Vocaturo e Zumpano (Caroprese et al. 2019), informam os principais registros geridos por estes sistemas e chamam a atenção para a necessidade de tecnologias de IA para a extração e compreensão dos dados clínicos dos pacientes.

EHRs não são mais usados, como no passado, para armazenar informações básicas do paciente e tarefas administrativas, mas podem incluir uma variedade de dados, incluindo o histórico médico do paciente, resultados de exames laboratoriais, dados demográficos, medicamentos e alergias, imunização status, imagens radiológicas, sinais vitais. No presente, o problema mudou da coleta de grandes quantidades de dados para entendê-los, ou seja, usar EHRs para transformar dados em conhecimento, conclusões e ações. Os EHRs não foram projetados para prever o risco ou a progressão da doença ou para determinar o tratamento correto, mas se forem combinados com o algoritmo de inteligência artificial (IA), esse problema se tornou possível (Caroprese et al. 2019, p. 1)³.

Ainda de acordo com Caroprese et al. (2019, pp. 1-2): «Uma das grandes tarefas da medicina moderna é a possibilidade de antecipar a evolução futura de uma doença, prevendo os riscos para o paciente para ajudar na prevenção. [...] A necessidade de ferramentas que permitam fazer previsões é uma prioridade»⁴. Nesse sentido, o conhecimento preditivo sobre os dados clínicos para o tratamento de doenças tem sido desenvolvido a partir de tecnologias de IA.

No tocante à área da Ciência da Informação, mais especificamente da subárea da Arquivologia, foram identificados 7 trabalhos com abordagens de temáticas sobre transparência de registros e acesso à informação arquivística potencializados por tecnologias da IA, assim como recuperação de grandes quantidades de dados para a sua ordenação. Além disso, tratam de funções arquivísticas fundamentais do trabalho de gestão docu-

³ Tradução livre de: «EHRs are not anymore used, as in the past, to store basic information of the patient and administrative tasks, but they may include a range of data, including the medical history of the patient, laboratory test results, demographics, medication and allergies, immunization status, radiology images, vital signs. At the present, the problem has shifted from collecting massive amounts of data to understanding it, i.e. use EHRs for turning data into knowledge, conclusions and actions. EHRs were not designed to forecast disease risk or disease progression or to determine the right treatment, but if they are combined with artificial intelligence (AI) algorithm this issue became possible» (Caroprese et al. 2019, p. 1).

⁴ Tradução livre de: «One of the major task of modern medicine is the possibility of anticipating the future evolution of a disease, by predicting the risks for the patient so that helping in prevention. [...] The need for tools allowing to perform prediction is a priority» (Caroprese et al. 2019, pp. 1-2).

mental, tais como processamento de fluxos de documentos até o arquivamento, classificação e avaliação arquivística.

De acordo com Rolan et al. (2019) ao realizar 4 estudos de caso sobre aplicações de IA em Arquivos da Austrália chega às seguintes conclusões, que os arquivistas precisam entender como implantar IA para uma boa manutenção de registros e, para alcançá-la, deve-se aprender novas tecnologias. Usar de forma eficiente para garantir que sistemas arquivísticos e de processos não sejam comprometidos. E, reflete que «Podemos ser humanos, mas a discussão sobre IA começou e todos nós precisamos participar» (Rolan et al. 2019, p. 18)⁵.

Ainda, dentro do escopo da interdisciplinaridade da Ciência da Informação, foram identificados 5 trabalhos que trataram sobre Ontologia.

De acordo com Zhou et al. (2021):

A ontologia, como uma importante técnica de descrição do conhecimento, tem sido amplamente aplicada na gestão do conhecimento e na disciplina da informação. Pertence ao escopo da categoria filosofia e evolui para um novo conceito integrando inteligência artificial (IA) (Zhou et al. 2021, p. 3)⁶.

Sendo assim, no que diz respeito aos 5 trabalhos selecionados para as análises, as regras de Ontologia são utilizadas como auxílio na descrição de informações com a finalidade de recuperação, extração e interpretação delas. Conforme El-Sappagh e Elmogy (2016, p. 32): «O conhecimento de domínio, na forma de regras e ontologia, pode ser usado para medir como uma consulta é semelhante a um caso anterior. Além disso, o grau de similaridade estima a quão adequada a solução anterior é para o problema atual»⁷.

No que se refere às tecnologias de IA aplicadas aos sistemas de informação, com fins de suporte às decisões, classificação, avaliação, acesso e segurança dos registros. As tecnologias que mais foram abordadas entre as 79 publicações podem ser visualizadas na Tabela 5, vale ressaltar que muitas publicações abordaram mais de uma tecnologia.

⁵ Tradução livre de: «We may be human, but the AI discussion has begun, and we all need to join in» (Rolan et al. 2019, p.18).

⁶ Tradução livre de: «Ontology, as an important knowledge description technique, has been widely applied in knowledge management and information discipline. It belongs to the scope of philosophy category and evolves to a new concept by integrating artificial intelligence (AI)» (Zhou et al. 2021, p. 3).

⁷ Tradução livre de: «Domain knowledge, in the form of rules and ontology, can be used to measure how a query is similar to a previous case. In addition, the degree of similarity estimates how suitable the previous solution is for the current problem» (El-Sappagh e Elmogy 2016, p. 32).

Tabela 5. Tecnologias de IA e complementares aplicadas na gestão documental

Tecnologia	N.º de ocorrências nas publicações	Conceito/Aplicação
<i>Machine Learning</i> (Aprendizado de Máquina)	40	«Machine Learning (ML) techniques have been widely used for extracting information from such large amount of data» (Bernardini et al. 2020, p. 2) Tradução livre: «As técnicas de aprendizado de máquina (ML) têm sido amplamente utilizadas para extrair informações de uma grande quantidade de dados» (Bernardini et al. 2020, p. 2)
<i>Natural Language Processing</i> (Processamento de Linguagem Natural)	23	«Natural language processing: This approach addresses the automatic computer interpretation of data contained in narrative text or plain text requiring a specific structure for processing it» (Muro et al. 2016, p. 132) Tradução livre: «Processamento de linguagem natural: Esta abordagem trata da interpretação automática por computador de dados contidos em um texto narrativo ou texto simples que requer uma estrutura específica para processá-lo» (Muro et al. 2016, p. 132)
<i>Deep Learning</i> (Aprendizagem Profunda)	13	«Deep learning is an artificial intelligence subdomain influenced by the structure of biological neurons connected in the brain. Deep learning helps to process the data and to identify hidden patterns for decision making. Deep learning models are multi-layered networks with the transformation between neurons in each layer. In the layered architecture of deep learning, upper layers extract the high-level features, and lower layers extract the lower level features» (Gupta, Sachdeva e Bhalla 2020, pp. 209-280) Tradução livre: «Aprendizagem profunda é um subdomínio de inteligência artificial influenciado pela estrutura dos neurônios biológicos conectados no cérebro. O aprendizado profundo ajuda a processar os dados e identificar padrões ocultos para a tomada de decisões. Modelos de aprendizado profundo são redes em várias camadas com a transformação entre neurônios em cada camada. Na arquitetura em camadas de aprendizado profundo, as camadas superiores extraem os recursos de alto nível e as camadas inferiores extraem os recursos de nível inferior» (Gupta, Sachdeva e Bhalla 2020, pp. 209-280)

Como se pode observar, a tecnologia de *Machine Learning* (ML) é a mais abordada e utilizada para a otimização, tanto para a extração como a análise de grande número de dados, bem como para a realização supervisionada e não supervisionada de determinadas funcionalidades em ambientes digitais. E, em casos de análises mais profundas, a *Deep Learning* (DL) é utilizada por meio de redes neurais de computador para a obtenção de resultados e combinações de dados mais precisas e compreensíveis para a formulação de conhecimentos que dão suporte a tomadas de decisões mais eficazes.

Aliado a isso, outra tecnologia usada e fundamental para o processamento das informações é o *Natural Language Processing* (NLP), que a partir da interpretação automática

por computador de dados permite a compreensão dos conteúdos textuais contidos em banco de dados de sistemas informatizados.

Para ilustrar como essas tecnologias são aplicadas na gestão de registros informacionais, e considerando que a maioria dos trabalhos (63) foram realizados na aplicação de tecnologias da IA em registros de saúde. É relevante e esclarecedor trazer as considerações da revisão sistemática realizada por membros da Organização Mundial de Saúde e Organização Regional de Saúde Europeia, quando analisaram artigos, também de revisão sistemática de literatura sobre o progresso tecnológico na otimização de sistemas de registro eletrônico de saúde — *Electronic Health Record* (EHR) e de registro pessoal de saúde — *Personal Health Record* (PHR):

*Os avanços tecnológicos podem melhorar a eficiência na implementação de sistemas EHR e PHR de várias maneiras. Técnicas de processamento de linguagem natural, seja com base em regras, aprendizado de máquina ou baseado em aprendizado profundo, podem extrair informações de narrativas clínicas e outros dados não estruturados bloqueados em EHRs e PHRs, permitindo a pesquisa secundária (ou seja, fenotipagem). Além disso, espera-se que EHRs e PHRs sejam os principais beneficiários da implementação da tecnologia blockchain em Sistemas de Informação em Saúde. Regulamentações de governança, falta de confiança, baixa escalabilidade, segurança, privacidade, baixo desempenho e alto custo continuam sendo os desafios mais críticos para a implementação dessas tecnologias (Negro-Calduch et al. 2021, p. 1)*⁸.

Sendo assim, essas discussões refletem sobre as potencialidades e benefícios das aplicações de tecnologias da IA em sistemas informatizados de saúde quando se refere à extração de informações clínicas, bem como chama a atenção para a superação dos entraves políticos, sociais e econômicos para a sua implementação a nível mundial.

CONSIDERAÇÕES FINAIS

As tecnologias da IA estão cada vez mais presentes no cotidiano das pessoas e das organizações. Com elas surgem soluções e benefícios nas mais diversas atividades humanas, contudo, isso deve ter delimitações e acompanhamento de profissionais pautados em aspectos teóricos e conceituais, bem como éticos de cada área dos saberes e fazeres humanos.

⁸ Tradução livre de: «Technological advancements can improve the efficiency in the implementation of EHR and PHR systems in numerous ways. Natural language processing techniques, either rule-based, machine-learning, or deep learning-based, can extract information from clinical narratives and other unstructured data locked in EHRs and PHRs, allowing secondary research (i.e., phenotyping). Moreover, EHRs and PHRs are expected to be the primary beneficiaries of the blockchain technology implementation on Health Information Systems. Governance regulations, lack of trust, poor scalability, security, privacy, low performance, and high cost remain the most critical challenges for implementing these technologies» (Negro-Calduch et al. 2021, p. 1).

No que se refere às áreas onde as tecnologias de IA são aplicadas em registros arquivísticos, pelo recorte temporal de 2016 a setembro/2021 nas bases de dados da LISTA, *Emerald* e *Scopus* pode-se verificar que a gestão de registros de saúde tem sido auxiliada com as potencialidades, principalmente de tecnologias como *machine learning*, *natural language processing* e *deep learning* na produção de conhecimento preditivo de dados clínicos para o tratamento e prevenção de doenças. Um aspecto que se deve mencionar é que essas iniciativas são de grupos de pesquisas isolados, ou seja, não se verificou na amostra coletada conexões entre eles, o que pode indicar a necessidade para o desenvolvimento de redes de estudos a respeito da aplicação de IA em registros eletrônicos de saúde, haja vista o relevante número de pesquisas identificadas na literatura.

No interesse desta pesquisa, no que tange ao gerenciamento de informações arquivísticas desde a produção, utilização e destinação — descarte ou guarda permanente de informações. Dos 79 estudos analisados, apenas 3 estudos abordam funções basilares da gestão documental: classificação, tramitação, avaliação de documentos, descrição e arquivamento.

Sobre fluxo documental que envolvem as funções de tramitação, avaliação, descrição e arquivamento, Hutchinson (2020) realizou um estudo que fornece uma visão geral dos esforços recentes relacionados ao processamento de linguagem natural (PNL) e aprendizado de máquina aplicado ao processamento de arquivos, particularmente avaliações e análises de sensibilidade, e propor requisitos funcionais e considerações de fluxo de trabalho para a transição do uso experimental para o operacional.

A respeito da avaliação documental, função essencial da arquivística, que proporciona racionalização aos arquivos no tocante aos descartes responsáveis de informações arquivísticas, contribuindo para a diminuição de massa documental desnecessária e que a cada dia ocupa milhares de *bytes* em ambientes digitais de armazenamento.

Nesse sentido, Makhoulf Shabou, Tièche, Knafou e Gaudinat (Makhoulf Shabou et al. 2020) desenvolveram uma pesquisa interdisciplinar na Escola de Administração de Empresas de Genebra e apoiada pelos Arquivos do Estado de Neuchâtel (Office des archives de l'État de Neuchâtel, OAEN) com o propósito de fornecer um *framework* e uma prova de conceito para um *software* que auxilie na tomada de decisões defensáveis sobre a retenção e descarte de registros e dados propostos à OAEN.

Como resultado dessa pesquisa, os autores demonstram a viabilidade de mensurar o valor dos dados e registros com um método reproduzível. A transferência e aplicabilidade da proposta de métricas desenvolvidas para o arquivamento e mineração de dados podem ser consideradas para outros tipos de dados. Uma adaptação deste método e suas métricas pode ser testada em dados de pesquisa, dados médicos ou dados bancários.

Por fim, outro estudo que traz muitas reflexões sobre aplicações de IA na gestão arquivística de documentos, inclusive este estudo foi citado por Hutchinson (2020) e

Makhlouf Shabou et al. (2020), é a contribuição de Rolan et al. (2019), que traz as seguintes inquietações:

- a) Onde está a IA ou automação para ajudar os trabalhos arquivísticos de classificação, avaliação e descarte?
- b) Como a IA pode automatizar o trabalho de arquivamento e manutenção de registros digitais?

Para tanto, os autores ao responder essas questões, fornecem uma breve introdução às técnicas de IA e suas características em relação ao trabalho de manutenção de registros. Em seguida, eles apresentam quatro estudos de caso de instituições de Arquivo e governamentais australianas que embarcaram em iniciativas de IA. Em cada caso, eles fornecem uma visão geral do projeto em termos de requisitos, atividades até o momento, resultados e ações futuras. O artigo conclui com uma discussão sobre as lições aprendidas, problemas e implicações da IA no arquivo. Dentre elas, merece destaque as seguintes conclusões:

- *Embora a IA prometa eficiência no suporte de avaliação digital, documentação e descarte, parece ser uma capacidade emergente e certamente não uma 'bala de prata' pronta para produção. No entanto, a IA chegou ao nosso campo e produzirá mudanças profundas em nossos ambientes de trabalho nos próximos anos (Rolan et al. 2019, p. 17)*⁹.
- *É importante notar que é necessário desenvolver uma compreensão sobre conceitos da IA dentro da profissão para garantir que as necessidades de gestão documental sejam atendidas no futuro (Rolan et al. 2019, p. 18)*¹⁰.
- *Os arquivistas precisam aproveitar a oportunidade apresentada pelo surgimento desse conjunto de novas tecnologias para explorar o potencial da IA e expor as realizações e aprendizados para a comunidade profissional mais ampla (Rolan et al. 2019, p. 18)*¹¹.

⁹ Tradução livre de: «While AI promises efficiencies in the support of digital appraisal, documentation and disposal, it appears to be an emerging capability and certainly not a production-ready 'silver bullet'. However, AI has arrived in our field and it will produce profound changes in our working environments in the years to come» (Rolan et al. 2019, p. 17).

¹⁰ Tradução livre de: «Importantly, it is apparent that there is a need to develop an understanding of AI concepts within the profession to ensure that recordkeeping needs are met into the future» (Rolan et al. 2019, p. 18).

¹¹ Tradução livre de: «Archivists need to take advantage of the opportunity presented by the emergence of this suite of new technologies to explore the potential of AI and to expose achievements and learnings to the broader professional community» (Rolan et al. 2019, p. 18).

- *Para a discussão sobre IA e gestão de documentos, precisamos estar informados e, de preferência, experientes em técnicas de IA, para neutralizar os argumentos tecnológicos de não-conformidade* (Rolan et al. 2019, p. 18)¹².
- *A fim de atender aos requisitos de documentação para nossos próprios registros — como os descritos na ISO 23081 — precisamos entender como implantar IA que contribui para a gestão de documentos. Nosso objetivo deve ser uma boa manutenção de registros e, para alcançá-la, devemos aprender novas tecnologias* (Rolan et al. 2019, p. 18)¹³.

Portanto, a revisão sistemática de literatura a respeito de aplicações de IA na gestão documental possibilitou uma visão geral dos últimos cinco anos de publicações que, revelaram o expressivo número de estudos sobre a otimização na extração de dados clínicos de registros eletrônicos de saúde por meio de tecnologias da inteligência artificial e a produção de conhecimento preditivo a respeito de determinadas doenças, tendo como maior veículo de indexação destes trabalhos a base de dados *Scopus*.

Como também, permitiu a visualização de estudos relevantes para a gestão de documentos, obtidos em maioria na base de dados *Emerald*, com ênfase na função de avaliação de registros informacionais, atividade técnica e conceitual inerente ao fazer arquivístico e de impactos imprescindíveis a manutenção de arquivos com valor probatório e informativo para a produção de novos conhecimentos.

Assim como, demonstrou as potencialidades das tecnologias da IA para a potencialização no tratamento de dados e informações sejam estruturadas ou não estruturadas das já mencionadas e descritas áreas da saúde e arquivologia, como também, de ontologia, *accountability* (auditoria), sistemas de informação pervasivos, megaprojeto, jurídica, diplomática e biblioteconomia. E por que não, afirmar sobre a aplicabilidade em todas as áreas que produzem e acumulam informações em ambiente digital.

Contudo, reafirma-se a necessidade de que, essas tecnologias sejam desenvolvidas e aperfeiçoadas de forma responsável e ética. Além de que, sejam estabelecidas redes de desenvolvimento científico para soluções de manutenção de registros arquivísticos auxiliados pelas tecnologias da IA que possam beneficiar o maior número de pessoas e instituições do mundo.

¹² Tradução livre de: «If we are to contribute to discussion about AI and recordkeeping we need to be informed and preferably experienced in AI techniques, to counteract technological arguments for non-compliance» (Rolan et al. 2019, p. 18).

¹³ Tradução livre de: «In order to meet the documentation requirements for our own records — such as those described in ISO 23081 — we will need to understand how to deploy AI that contributes to documentation of recordkeeping business. Our goal should be good recordkeeping and to achieve it we must move with and learn new technologies» (Rolan et al. 2019, p. 18).

REFERÊNCIAS

- ARAÚJO, C. A. Á., 2013. O que é Ciência da Informação? *Informação & Informação* [Em linha]. **19**(1), 1-30 [consult. 2023-11-16]. Disponível em: <https://doi.org/10.5433/1981-8920.2014v19n1p01>.
- BERNARDINI, M., et al., 2020. Early temporal prediction of Type 2 Diabetes Risk Condition from a General Practitioner Electronic Health Record: A Multiple Instance Learning Boosting Approach. *Artificial Intelligence in Medicine* [Em linha]. **105** [consult. 2023-11-16]. Disponível em: <https://doi.org/10.1016/j.artmed.2020.101847>.
- BUNN, J., 2020. Working in contexts for which transparency is important: A recordkeeping view of explainable artificial intelligence (XAI). *Records Management Journal* [Em linha]. **30**(2), 143-153 [consult. 2023-11-16]. Disponível em: <https://doi.org/10.1108/RMJ-08-2019-0038>.
- CAROPRESE, L., et al., 2019. E. Deep learning techniques for electronic health record analysis. Em: *2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA)* [Em linha]. Zakynthos: IEEE [consult. 2023-11-16]. Disponível em: <https://doi.org/10.1109/IISA.2018.8633647>.
- EL-SAPPAGH, S., e M. M. ELMOGY, 2016. Medical case based reasoning frameworks: Current developments and future directions. *International Journal of Decision Support System Technology* [Em linha]. **8**(3), 31-62 [consult. 2023-11-16]. Disponível em: <https://doi.org/10.4018/IJDSST.2016070103>.
- GALVÃO, T. F., T. de S. A. PANSANI, e D. HARRAD, 2015. Principais itens para relatar Revisões sistemáticas e Meta-análises: A recomendação PRISMA. *Epidemiologia e Serviços de Saúde* [Em linha]. **24**(2), 335-342 [consult. 2023-11-16]. Disponível em: <https://doi.org/10.5123/S1679-49742015000200017>.
- GIL, A. C., 2008. *Métodos e técnicas de pesquisa social*. 6.ª ed. São Paulo: Editora Atlas SA.
- GUPTA, V., S. SACHDEVA, e S. BHALLA, 2020. A Novel Deep Similarity Learning Approach to Electronic Health Records Data. *IEEE Access* [Em linha]. (8), 209278-209295 [consult. 2023-11-16]. Disponível em: <https://doi.org/10.1109/ACCESS.2020.3037710>.
- HUTCHINSON, T., 2020. Natural language processing and machine learning as practical toolsets for archival processing. *Records Management Journal* [Em linha], **30**(2), 155-174 [consult. 2023-11-16]. Disponível em: <https://doi.org/10.1108/RMJ-09-2019-0055>.
- MAKHLOUF SHABOU, B., et al., 2020. Algorithmic methods to explore the automation of the appraisal of structured and unstructured digital data. *Records Management Journal* [Em linha]. **30**(2), 175-200 [consult. 2023-11-16]. Disponível em: <https://doi.org/10.1108/RMJ-09-2019-0049>.
- MASUDA, Y., et al., 2019. Clinical Decision-Support System with Electronic Health Record: Digitization of Research in Pharma. Em: Y. W. CHEN et al., eds. *Innovation in Medicine and Healthcare Systems, and Multimedia. Smart Innovation, Systems and Technologies*. Singapore: Springer, vol. 145, pp. 47-57 [consult. 2023-11-16]. Disponível em: https://doi.org/10.1007/978-981-13-8566-7_5.
- MURO, N., et al., 2016. Experience-Based Electronic Health Records. *Cybernetics and Systems* [Em linha]. **47**(1-2), 126-139 [consult. 2023-11-16]. Disponível em: <https://doi.org/10.1080/0196972.2.2016.1128774>.
- NEGRO-CALDUCH, E., et al., 2021. Technological progress in electronic health record system optimization: Systematic review of systematic literature reviews. *International Journal of Medical Informatics* [Em linha]. (152) [consult. 2023-11-16]. Disponível em: <https://doi.org/10.1016/j.ij-medinf.2021.104507>.
- NEVES, B. C., 2020. Sistemas e experiências de inteligência artificial da ciência da informação e ciências da saúde. *Revista Fontes Documentais*. (3), 504-511.
- RICHARDSON, R. J., 2012. *Pesquisa social: métodos e técnicas*. 3.ª ed., 14 reimpr. São Paulo: Editora Atlas SA.

- ROLAN, G., et al., 2019. More human than human? Artificial intelligence in the archive. *Archives & Manuscripts*. 47(2), 179-203.
- SILVA, E. L. da, e E. M. MENEZES, 2001. *Metodologia da Pesquisa e Elaboração de Dissertação*. 3.^a ed. rev. atual. Florianópolis: Laboratório de Ensino a Distância da UFSC.
- ZHOU, F., et al., 2021. Knowledge management practice of medical cloud logistics industry: Transportation resource semantic discovery based on ontology modelling. *Journal of Intellectual Capital* [Em linha]. 22(2), 360-383 [consult. 2023-11-16]. Disponível em: <https://doi.org/10.1108/JIC-03-2020-0072>.