

Resumo: A Gestão Eletrônica de Documentos é um campo da Ciência da Informação que incorpora um conjunto de tecnologias, ferramentas e sistemas, fundamentais na organização, preservação, proteção, distribuição e uso de informações e recursos informacionais em contextos corporativos e institucionais, especialmente com o avanço da era digital. A Inteligência Artificial e a Ciência de Dados são tecnologias emergentes que oferecem inovação por meio da automação de processos, análises avançadas e melhoria na recuperação de informações, aumentando eficiência e apoio à tomada de decisões. Diante dessa problematização foi conduzida a questão principal: qual a contribuição da Inteligência Artificial e da Ciência de Dados na Gestão Eletrônica de Documentos, com o intuito de identificar padrões, desafios e oportunidades que possam contribuir para a evolução das práticas de gestão documental? Para tanto foi definido e aplicado um protocolo de revisão de literatura nas bases de dados Web of Science, Scopus, Brapci e Lista, com um marco temporal de 2000 a 2024. Como resultado final, foram selecionados 18 artigos, em sua maioria pesquisas experimentais ou práticas, que permitiram uma visão ampla de abordagens desenvolvidas para aplicação de Inteligência Artificial e Ciência de Dados na Gestão Eletrônica de Documentos. Algumas lacunas foram identificadas, dentre elas a falta de padronização das terminologias, dificuldade na obtenção de grandes volumes de dados, treinamento e testes. Também questões éticas e baixa adoção pelos profissionais da informação, ainda são fatores que impedem a implementação plena dessas tecnologias. Com isso, esta pesquisa conclui que as principais contribuições incluem automação de tarefas, análises preditivas, identificação de padrões, segurança aprimorada e maior acessibilidade à informação e aponta oportunidades no uso de Inteligência artificial e Ciência de Dados na Gestão Eletrônica de Documentos.

Palavras-chave: Ciência de Dados; Ciência da Informação; Gestão Eletrônica de Documentos; Inteligência Artificial; Revisão Sistemática de Literatura.

Abstract: Electronic Records Management is an area of Information Science that incorporates a set of technologies, tools, and systems essential for organizing, preserving, protecting, distributing, and using information and informational resources in corporate and institutional contexts, especially with the advancement of the digital age. Artificial Intelligence and Data Science are emerging technologies that offer innovation through process automation, advanced analytics, and improved information retrieval, enhancing efficiency and supporting decision-making. Given this context, a systematic literature review was conducted to address the main question: what is the contribution of Artificial Intelligence and Data Science to Electronic Document Management, aiming to identify patterns, challenges, and opportunities that can contribute to the evolution of records management practices? To this end, a review protocol was defined and applied to the Web of Science, Scopus, Brapci, and Lista databases, covering a timeframe from 2000 to 2024. As final result, 18 articles were selected, mostly experimental or practical research, which provided a broad view of approaches developed for the application of Artificial Intelligence and Data Science in Electronic Records Management. Some gaps were identified, including a lack of standardization in terminologies, difficulties in obtaining large volumes of training and testing data, ethical issues, and low adoption by information professionals, which remain barriers to full implementation. Thus, this research seeks to contribute to identifying opportunities for the use of Artificial Intelligence and Data Science in Electronic Records Management.

Keywords: Data Science; Information Science; Electronic Records Management; Artificial Intelligence; Systematic Literature Review.

1. Introdução

A Gestão Eletrônica de Documentos (GED) é reconhecida como um campo essencial para a organização e preservação da informação em ambientes corporativos e institucionais. Sua relevância cresce diante do aumento exponencial de dados digitais, exigindo sistemas mais eficientes e inteligentes que possam atender às demandas modernas. Nesse contexto, compreender e aprimorar a GED é fundamental para garantir a integridade, segurança e acessibilidade das informações, fortalecendo sua aplicação em múltiplos setores.

Com a evolução tecnológica, a Inteligência Artificial (IA) e a Ciência de Dados (CD) emergem como forças transformadoras no campo da GED. Essas tecnologias não apenas potencializam a automação de tarefas repetitivas, mas também oferecem capacidades avançadas de análise, organização e recuperação de informações. Tais inovações podem impactar positivamente os sistemas, aprimorando sua eficiência, reduzindo custos e permitindo decisões mais ágeis.

Contudo, a implementação dessas tecnologias na GED enfrenta desafios significativos. A ausência de padrões bem estabelecidos, os riscos éticos e sociais ainda pouco explorados limitam o pleno aproveitamento de seu potencial. Essas questões apontam para a necessidade de pesquisas mais aprofundadas que abordem tanto os aspectos técnicos quanto os impactos humanos e sociais dessas inovações.

A metodologia adotada neste estudo baseia-se no planejamento e condução de uma Revisão Sistemática da Literatura (RSL) que se destaca tanto como processo quanto como produto, contribuindo para a discussão que pode contribuir com o avanço do conhecimento desses campos na área. Vale salientar que a revisão de literatura foi um recurso instrumental, servindo de base para uma análise sobre a inteligência artificial e ciência de dados aplicada à GED.

A pesquisa, de caráter exploratório-descritivo, combinou abordagens qualitativas e quantitativas para a análise e síntese dos estudos. A análise temática e de conteúdo foi utilizada para identificar padrões e categorias, proporcionando uma visão detalhada das tendências e lacunas existentes.

Diante disso, esta pesquisa teve como objetivo mapear as principais técnicas e ferramentas de IA e CD aplicadas à GED, destacando suas contribuições, limitações e possíveis direções futuras. Além disso, buscou-se avaliar os impactos éticos, sociais e organizacionais associados a essas tecnologias, apresentando um panorama com informações para a melhoria contínua da área. Assim, espera-se contribuir para o avanço da GED como um campo estratégico e interdisciplinar, essencial para o gerenciamento de informações na era digital.

Os resultados da pesquisa revelaram importantes pontos de interseção entre GED, IA e CD, destacando-se a automatização de processos, otimização do acesso à informação e suporte à tomada de decisão. Além disso, foram identificados benefícios como maior eficiência operacional e novas possibilidades de análise documental, bem como lacunas significativas, incluindo a falta de padrões, diversidade de técnicas, métodos e desafios éticos e sociais. Essas descobertas reforçam o papel estratégico dos profissionais da informação na construção, em cooperação com profissionais de outras áreas, de soluções tecnológicas mais inclusivas e eficazes.

2. Gestão Eletrônica de Documentos: um novo momento

A GED desempenha um papel central na organização, preservação e acessibilidade da informação em organizações contemporâneas, unindo ferramentas tecnológicas e práticas documentais. Com o avanço das Tecnologias da Informação e Comunicação (TIC), tornou-se essencial gerenciar o crescente volume de documentos digitais de forma eficiente.

O documento é considerado um suporte para apresentação de dados, informações e conhecimentos explícitos. Ao passo que a informação se tornou um ativo importante para o desenvolvimento das organizações, os documentos assumiram o papel de registrar, armazenar, transmitir e comprovar situações, ações, eventos e processos em seu cotidiano (SANTOS e KRAWSZUK, 2020).

Com o avanço das TIC, tornou-se mais rápido e mais fácil a produção, edição, disseminação ou exclusão de documentos e registros em meios físicos e digitais (SANTOS e KRAWSZUK, 2020). O que desencadeou um crescimento exponencial e algumas vezes caótico de documentos, dificultando a recuperação da informação. “No entanto, além das tecnologias da informação auxiliarem na gestão de documentos, também proporcionaram o registro da informação orgânica em meio digital, tendo como resultado, o documento arquivístico digital” (SANTOS e FLORES, 2026:166).

A transição dos suportes de informação para as mídias digitais intensificou a relevância da gestão documental no contexto informacional, atribuindo-lhe um papel central, transformando-a em um componente essencial para a produção e organização do conhecimento (SANTOS, SILVA e BARI, 2021).

No entanto, o manejo de grandes volumes documentais, abrangendo uma diversidade de formatos, tornou-se um desafio significativo para as instituições. A falta de conhecimento especializado frequentemente resulta em arquivamentos sem critérios ou tratamento adequado (CALDERON *et al.*, 2004).

Como consequência, a agilidade na recuperação da informação fica comprometida e o acesso a essa documentação fica restrito aos responsáveis pelo seu arquivamento, o que limita sua utilização e dificulta a extração plena de seu potencial informacional, “propiciando situações frustrantes nas tomadas de decisões” (ELIAS, 2012:15).

A gestão de documentos envolve uma série de atividades essenciais para garantir a eficiência e a organização do fluxo de informações em uma organização. Isso inclui a análise da tipologia documental, sua relevância, vigência e qualidade informativa, bem como a padronização de documentos e formulários. Além disso, a informatização dos processos de tramitação documental, a coordenação entre os produtores de documentos e os arquivos, e a regulamentação das transferências são fundamentais (CALDERON *et al.*, 2004).

Com as tecnologias digitais emergentes, o acesso e o compartilhamento de dados, informação e conhecimento ocorrem em tempo real e em grandes volumes. Faz-se necessário desenvolver metodologias, critérios e ferramentas, que irão garantir seu armazenamento de forma organizada, estruturada e sistemática. Com isso seu compartilhamento torna-se eficiente e pode facilitar a tomada de decisão, garantindo o acesso à informação correta, no momento em que se precisa dela (WEITZEN, 1991).

Essa é a função da GED, uma área que incorpora um conjunto de tecnologias, ferramentas e sistemas responsáveis por criar, organizar e conservar documentos e recursos informacionais, possibilitando acesso seguro a estes recursos (VIANNA e FREITAS, 2019).

Mas uma grande lacuna na literatura está em apresentar ações concretas que evidenciem como dados, informações e conhecimentos explícitos, gerados a partir de documentos organizacionais, podem ser extraídos, estruturados, classificados e utilizados como fonte para gerar informações estratégicas para auxiliar na tomada de decisão e apoiar no avanço da transformação digital nas organizações.

3. Inteligência Artificial: busca e domínios

A busca por uma IA não é algo novo. Contudo, o marco inicial da IA, conforme estudamos hoje, foi firmado em 1956, quando o termo foi usado oficialmente na Faculdade de Dartmouth, nos Estados Unidos, em um evento com pesquisadores interessados em redes neurais e no estudo da inteligência (NAKAMITI, 2009). E novamente na década de 1980 volta a fazer parte do leque de possibilidades como uma tentativa de reduzir os requisitos computacionais impostos pela IA, usando o conhecimento de especialistas (NEVES, 2020).

A IA é considerada um ramo da Ciência da Computação que atua no objetivo de desenvolver sistemas computacionais que realizam tarefas que hoje, não têm solução algorítmica ou que ainda sejam melhor executadas por homens do que por máquinas, ou seja, problemas que não têm soluções exatas e sim, com muitas variáveis (SICHMAN, 2021).

Por se tratar de um conceito complexo e abrangente, “não tem uma definição acadêmica propriamente dita”, como ressalta Sichman (2021:36). No entanto, podemos entender, em síntese, que essa inteligência é um conjunto de técnicas de programação que emulam as principais funções associadas à inteligência humana, como raciocínio, aprendizagem e autoaperfeiçoamento para solucionar problemas em informática através de programas (NAKAMITI, 2009). Ao pensar em IA, observe uma interação entre a busca de objetivos, o processamento de dados usado para atingir esse objetivo e a aquisição de dados usada para melhor entender o objetivo (NEVES, 2020).

Esse conjunto de técnicas é chamado de comportamento inteligente e a principal dificuldade está em definir o que é ser inteligente. Isso porque, segundo Cozman e Neri (2021), os seres humanos têm muita flexibilidade com relação ao termo inteligência. Por isso, talvez seja mais produtivo falar de *machine learning* no que tange às suas características e no âmbito de seus eixos temáticos do que sob a luz epistemológica de sua natureza.

O objetivo principal da IA é desenvolver soluções e resolver problemas que os humanos fazem melhor que as máquinas, e que ainda não possuem solução computacional viável. Para tanto, é necessário entender como os seres humanos solucionam esses problemas, como são geradas as possíveis soluções, como é feita a escolha da melhor solução e seus efeitos em futuras decisões (SICHMAN, 2021).

A IA pode ser caracterizada por um conjunto de tecnologias, métodos, modelos e técnicas, que quando combinados entre si proporcionam a “busca, raciocínio e representação de conhecimento, mecanismos de decisão, percepção, planejamento, processamento de

linguagem natural, tratamento de incertezas e aprendizado de máquina” (SICHMAN, 2021:39).

Segundo Russell e Norvig (2004) a maioria das tecnologias de IA podem ser classificadas em seis disciplinas:

- **Processamento de linguagem natural:** possibilita a interação e comunicação entre humanos e máquinas usando linguagem natural, facilitando a compreensão e geração de texto por sistemas computacionais;
- **Aprendizado de máquina:** permite que sistemas identifiquem padrões, aprendam com dados e se adaptem a novas situações sem programação explícita para cada tarefa;
- **Visão computacional:** abrange a interpretação de informações visuais do ambiente, como reconhecimento de imagens e vídeos;
- **Representação do conhecimento:** organiza e armazena informações de maneira estruturada, permitindo que sistemas acessem e utilizem dados adquiridos ou inferidos;
- **Raciocínio automatizado:** aplica lógica e algoritmos para responder a questões complexas e realizar deduções automáticas com base no conhecimento armazenado;
- **Robótica:** envolve o controle e coordenação de dispositivos físicos para tarefas como movimentação, manipulação de objetos e interação com o ambiente físico.

A evolução da IA vai além de avanços tecnológicos: ela redefine como organizamos, interpretamos e utilizamos informações. Desde sua concepção, a IA busca não apenas replicar capacidades humanas, mas também aprimorar processos que envolvem grandes volumes de dados, como classificação, aprendizado e tomada de decisões. A IA se baseia em algoritmos para alcançar um resultado que pode ou não ter algo a ver com objetivos humanos ou métodos para atingir esses objetivos (NEVES, 2020).

Essa transformação impacta significativamente áreas dedicadas à organização e gestão de informações, permitindo o desenvolvimento de soluções mais inteligentes e eficientes que integram percepção, aprendizado e colaboração, potencializando a capacidade humana em ambientes digitais.

4. Ciência de Dados: breve contextualização

O termo *Data Science* ou Ciência de Dados (CD) apareceu pela primeira vez, segundo Cao (2017), no prefácio do livro *Concise survey of computer methods* de Peter Naur, publicado em 1974.

A CD pode ser abordada tanto como um conjunto de técnicas quanto como um campo em consolidação no âmbito científico. Do ponto de vista técnico, seu foco recai sobre a extração de padrões e significados a partir de grandes volumes de dados, frequentemente sem uma

problematização direta das causas subjacentes. Já em uma perspectiva mais investigativa, há um movimento emergente que busca compreender os fenômenos que geram os dados, promovendo análises orientadas por questões epistemológicas e contextuais, o que sugere uma aproximação com os objetivos próprios da prática científica, o que a torna “[...] uma Ciência com e sem teorias” (FONSECA, 2021:1.594, tradução nossa).

As abordagens científicas e técnicas diferem no uso de dados e hipóteses. Na ciência convencional, hipóteses são formuladas previamente e testadas para confirmar ou refutar teorias. Já na abordagem técnica, os dados são manipulados primeiro, invertendo a ordem tradicional de geração de hipóteses, com foco em insights sem necessariamente confirmar teorias (FONSECA, 2021).

Cao (2017) define que a CD como estudo dos dados e complementa que é um campo interdisciplinar novo que se vale de conhecimento de áreas como estatística, informática, computação, comunicação, gestão e sociologia para estudar dados. Os autores Rautenberg e Carmo (2019) reafirmam o caráter interdisciplinar da CD, e acrescentam a Ciência da Informação (CI) como uma área relacionada.

A CI constitui-se historicamente como um campo interdisciplinar, resultante da convergência entre biblioteconomia, documentação, ciência cognitiva, comunicação e, mais recentemente, tecnologias digitais (CAPURRO e HJØRLAND, 2003).

Como uma área interdisciplinar, exige habilidades em Ciência da Computação para manipulação e armazenamento de dados, uso de algoritmos de *machine learning* e visualização da informação. O domínio de Matemática e Estatística é essencial para interpretar resultados e compreender algoritmos.

Além disso, o conhecimento do domínio do problema é crucial para formular hipóteses relevantes e gerar informações úteis para a tomada de decisão. Essa convergência de áreas torna a CD uma ponte entre tecnologia, análise quantitativa e aplicação prática (RAUTENBERG e CARMO, 2019).

A CD está intrinsecamente ligada à aplicação de métodos, onde softwares são utilizados para transformar dados em informações relevantes, através da sua extração e, bases de dados extensas, complexas e dinâmicas, contribuindo de forma decisiva para a tomada de decisões (RAUTENBERG e CARMO, 2019).

É proposta por Fonseca (2021) uma visão da CD em três fases, 1) Planejamento da coleta de dados; 2) Análise de dados; 3) Interpretação dos resultados, onde afirma, a partir da discussão de cada fase, que a CD não está desconectada da teoria, mas que é apresentada e utilizada em diferentes etapas de diferentes maneiras.

Fonseca (2021) afirma que a CD recebe o auxílio das chamadas Ciências Auxiliares. Áreas como CI, Estatística e Ciência da Computação dão suporte nas diferentes fases de desenvolvimento da Ciência de Dados. A CI tem papel importante nas etapas de exploração, identificação, armazenamento e preservação dos dados, tendo domínio de todo seu ciclo de vida. As áreas de Estatística e Ciência da Computação, auxiliam nas etapas de processamento, tratamento e análise de resultados.

Com isso, a CD não está limitada à análise estatística ou ao processamento de dados, mas compreende diversas práticas que passam pela engenharia de dados e vão até a visualização

e comunicação dos resultados. Seu foco está na geração de *insights* a partir do processamento e interpretação dos dados (MOUTINHO, 2014:6).

Dessa forma, a Ciência de Dados pode ser compreendida como uma abordagem de natureza multidisciplinar, cujas práticas envolvem metodologias estatísticas, computacionais e informacionais. Quando articulada à CI, campo que historicamente se caracteriza por sua vocação interdisciplinar, tal abordagem ganha densidade analítica, ao permitir não apenas a transformação de dados em subsídios para a tomada de decisão, mas também a compreensão crítica dos contextos nos quais esses dados são produzidos, organizados e interpretados.

5. Relação da Ciência da Informação com a Inteligência Artificial e a Ciência de Dados

Segundo Marchionini (2016), a CI é um dos quatro pilares do qual se origina a CD. O autor afirma ainda que os cientistas da informação são responsáveis por gerir todo o ciclo de vida dos dados, como também estão envolvidos com questões socioculturais atreladas à coleta e uso desses dados.

Essa abordagem abrangente da CI contribui de maneira significativa para a compreensão e otimização da gestão de dados, considerando não apenas os aspectos técnicos, mas também os elementos éticos e sociais envolvidos nesse processo dinâmico e crucial para a tomada de decisões estratégicas. Portanto, a interseção entre CD e CI representa um campo fértil para inovações e avanços no tratamento da informação no contexto contemporâneo.

No que se refere ao campo técnico-científico da IA, disciplina que mantém uma relação estreita com a CI, destaca-se como principal desafio “conseguir desenvolver sistemas computacionais que sejam capazes de entender a linguagem humana de acordo com os parâmetros de uma língua formal, sistematizada com propósitos e finalidades bem definidos quanto ao seu uso” (FONSECA, GUELPELI e SOUZA NETTO, 2022:3).

Em 2021, Neves inferiu que a IA tem grande contribuição na CI e vice-versa. Com base na análise da literatura científica internacional, Neves (2021) identificou que as maiores perspectivas da IA, por meio da computação cognitiva, em unidades de informação são a transformação do trabalho dos profissionais, melhoria do processamento de informações, o apoio à pesquisa, técnicas do aprendizado de máquina incorporados aos sistemas de automação, atuação dos profissionais como especialistas capazes de intermediar os usuários com a IA e a apropriação das ferramentas e recursos do *machine learning*.

Enquanto a linguística, como “estudo sistemático da linguagem” (VAJJALA, 2020:8, tradução nossa) se ocupa em entender, descrever e traduzir os conceitos e estrutura da linguagem natural, a CI se ocupa da estruturação de ontologias, tesouros e vocabulários controlados, utilizados para contextualização da informação, descrição de “um domínio do conhecimento de forma estruturada, através de: classes, propriedades, relações, restrições, axiomas e instâncias” (SANTAREM SEGUNDO e CONEGLIAN, 2015:227) e do tratamento da desambiguação do sentido das palavras que melhora a exatidão evitando a multiplicidade de significados associados a uma única palavra no momento da recuperação da informação (NHACUONGUE, 2020:17).

Dessa forma, conforme apresentado por Siqueira e Pereira (1989:46) “a recuperação da informação revela-se como um item prioritário na interação entre a [Ciência da Informação] e a Inteligência Artificial”, pois é através da recuperação da informação e do processamento pelo computador dos termos em linguagem natural que é possível solucionar problemas informacionais e apresentar soluções através da interação entre o usuário e os agentes inteligentes.

Na busca de soluções para as questões informacionais da sociedade, os estudos em CI têm motivado debates sobre a transformação do grande volume de dados operacionais produzidos diariamente, em informações estratégicas que deem suporte à tomada de decisões e agreguem valor às organizações (SANTOS e VIDOTTI, 2009).

6. Metodologia

A metodologia da pesquisa é baseada no planejamento e condução de uma Revisão Sistemática da Literatura (RSL), que é uma forma de realizar a identificação, avaliação e interpretação de estudos realizados sobre um determinado problema, assunto ou área específica.

Uma RSL pode oferecer contribuições valiosas tanto como processo quanto como produto da pesquisa. Como processo, se destaca por sua ênfase em um planejamento detalhado e explícito das estratégias e etapas a serem seguidas (KITCHENHAM e CHARTERS, 2007). Como produto, fornece um panorama abrangente e organizado do estado da arte, identificando padrões, lacunas de estudos em uma determinada área, tendências e empíricas para discutir ou desenvolver hipóteses teóricas (KITCHENHAM e CHARTERS, 2007).

A revisão foi utilizada como fundamento para a construção de uma análise crítica, à luz da CI, considerando a natureza interdisciplinar do campo e suas interfaces como (outros temas, como CD, IA, etc.). A metodologia, portanto, não se restringe à sistematização da produção acadêmica, mas busca elaborar uma reflexão fundamentada a partir da articulação entre autores, conceitos e contextos.

Assim, é considerada uma pesquisa básica com caráter exploratório-descritivo e, para a análise e síntese dos estudos, adota uma abordagem combinando aspectos qualitativos e quantitativos.

Para desenvolver esta pesquisa optou-se pela abordagem mista ou qualitativa/quantitativa ou quali/quanti, neste tipo de pesquisa ocorre a coleta e análise de dados tanto qualitativos quanto quantitativos. Richardson *et al.* (2007:80) afirmam que pesquisas qualitativas “têm como objetivo situações complexas ou estritamente particulares”, já pesquisas quantitativas “podem descrever a complexidade de determinado problema, analisar a interação de certas variáveis, compreender e classificar processos dinâmicos”.

Para a etapa de análise de dados, a técnica de análise temática e de conteúdo foi escolhida enquanto técnica qualitativa, visto que auxilia no processo de categorização, classificação e identificação das informações importantes para o resultado deste trabalho, ao passo que permite quantificar e comparar sistematicamente seu conteúdo.

A análise de conteúdo fornece um conjunto diverso de técnicas de análise e descrição de comunicações, de forma sistemática e objetiva, para produção de indicadores que permitirão a realização de inferências para geração de novos conhecimentos (BARDIN, 1977).

6.1. Desenvolvimento da pesquisa

Uma RSL possui estratégia rigorosa e replicável para identificar, avaliar e interpretar estudos disponíveis e relevantes relacionados a um tópico de pesquisa. Todos os passos e métodos são controlados por meio de um protocolo de revisão, permitindo, portanto, avaliar a completude do conjunto de estudos e a confiabilidade de todo o processo (NARIONAL..., 2000).

A organização e condução da revisão sistemática foi realizada, como apresentado por Kitchenham (2004), em três fases principais: Fase 1) Planejamento da Revisão; Fase 2) Condução da Revisão; e Fase 3) Publicação dos Resultados.

Três questões foram derivadas a partir do objetivo de pesquisa: P1) Quais são os principais pontos de interseção entre Gestão Eletrônica de Documentos, Inteligência Artificial e a Ciência de Dados?; P2) Quais desafios, tendências e lacunas de pesquisa estão presentes na interseção dessas áreas?; P3) Como as aplicações de Inteligência Artificial e Ciência de Dados têm influenciado as práticas na Gestão Eletrônica de Documentos?

As bases de dados escolhidas foram Web of Science, Scopus, Brapci e Lista. A escolha tanto da base de dados Web of Science quanto da Scopus se deu por sua abrangência de diversas áreas, enquanto que a escolha da base de dados Brapci e Lista foi feita por conter um foco em artigos publicados pela área da CI, sendo a Brapci nacional e Lista internacional. A escolha dessas bases também segue as diretrizes propostas por Dyba, Dingsoyr e Hanssen (2007), no contexto de uma revisão sistemática.

Com base na definição dos critérios de inclusão, exclusão e qualidade, deu-se início ao processo de seleção e análise dos artigos, que foi estruturado em duas etapas. Na primeira etapa, é feita a análise do título, palavras-chave, resumo e, para alguns casos, a introdução e conclusões. Na segunda etapa, é lida a versão completa dos artigos.

A *string* de busca foi estruturada a partir de dois conjuntos principais: o primeiro relacionado à área de Gestão Eletrônica de Documentos e suas derivações, e o segundo, às áreas de IA e CD, conforme apresentado no Quadro 1.

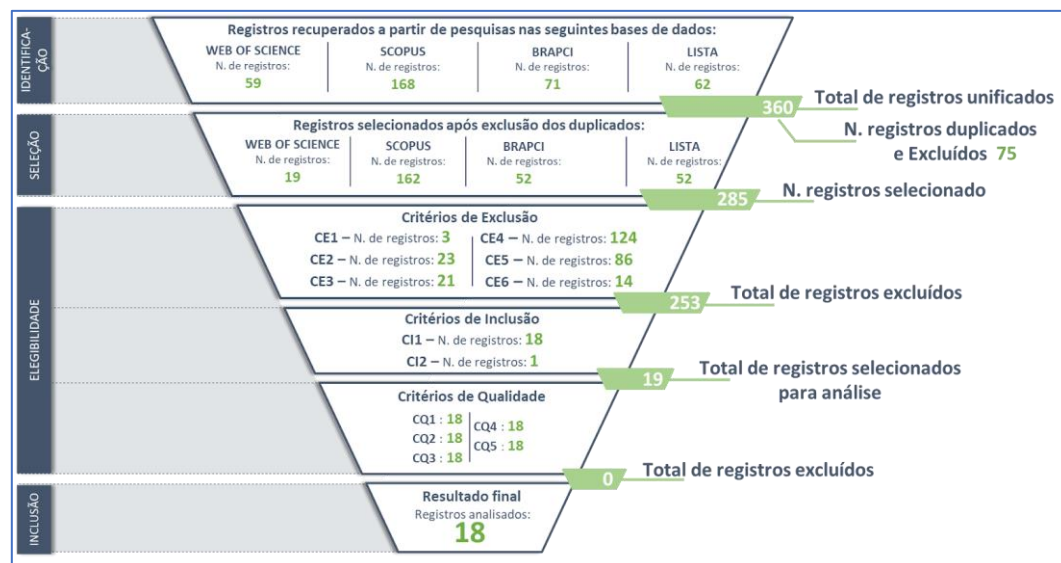
Quadro 1 - String final de busca

Item	Descrição
GED	("Electronic Recordkeeping" OR "Electronic record* management" OR "Electronic document management") AND
IA+CD	("artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "artificial intelligence" OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR *fuzzy* OR robotic* OR dialogue OR "predictive analy*" OR "recommend* system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "association rule" OR "augmented reality" OR "image segmentation" OR "genetic algorithm" OR "descriptive model*" OR "probabilist* method*" OR "self learning")

Fonte: Elaboração própria.

Na Fig. 1 são apresentados todos os passos desenvolvidos para seleção dos artigos que irão compor o corpus de análise desta pesquisa.

Fig. 1 – Fluxograma Prisma 2020



Fonte: Elaboração própria.

Foram recuperados um total de 360 registros, sendo: 168 registros na Scopus, 71 registros na Brapci, 62 registros na Lista e 59 registros na Web of Science. Dos 360 artigos recuperados, 75 eram duplicados, sendo: 40 da Web of Science, 19 da Brapci, 10 da Lista e 6 da Scopus. Dos 40 registros duplicados, 37 estão repetidos tanto na base da Scopus quanto na base da Web of Science e 3 estão repetidos no próprio resultado de busca. Estes últimos possuem exatamente os mesmos trabalhos, mas com títulos diferentes.

Ao final da seleção foram excluídos 75 registros duplicados, 3 registros que se enquadram no Critério de Exclusão CE1 (o estudo está escrito em um idioma diferente de português, inglês e espanhol), 23 registros que se enquadram no CE2 (o texto completo não possui acesso aberto ou não pode ser acessado por outro motivo), 21 registros que se enquadram no CE3 (o estudo é livro, capítulo de livro, editorial, nota, carta, *proceedings* (anais), *conference review* (revisão de conferência) e *retracted* (retração)), 120 registros que se enquadram no CE4 (o estudo aborda apenas o tema GED sem intersecção com IA e CD), 86 registros que se enquadram no CE5 (o estudo apenas cita GED como ferramenta usada na pesquisa, como resultado de pesquisa ou sugestão de uso) e 14 registros que se enquadram no CE6 (artigos anteriores a 2000).

Com isso, somam-se 267 artigos a serem excluídos, restando 18 registros, sendo que todos atendem ao CI1 (o estudo discute e propõe técnica, estratégia, ferramenta ou qualquer abordagem de IA ou de CD ligada à GED). Quanto aos critérios de qualidade, nenhum dos 18 artigos recebeu a soma das notas inferior a 2,5 pontos, sendo todos mantidos para a análise de conteúdo.

7. Resultados do estudo

A análise de conteúdo tem como objetivo categorizar e classificar qualquer conteúdo dos artigos levantados. Para essa análise foi realizada a leitura do texto integral em momentos e com objetivos diferentes. Num primeiro momento foi realizada uma leitura “transversal” que buscou conhecer os assuntos tratados nos artigos. Iniciou-se um primeiro levantamento de dados e informações e novas pesquisas foram necessárias para entender os conceitos apresentados.

Estas pesquisas se originam de 11 países, sendo eles: Argélia, Brasil, Índia, Cazaquistão, Rússia, Espanha, Taiwan, Turquia, Ucrânia, Estados Unidos e Uzbequistão. O país com maior número de artigos publicados nesta área é a Rússia, com 6 publicações, representando 33%, seguida por Estados Unidos e Uzbequistão, ambos com 2 trabalhos, representando 11% das publicações. Os outros países, incluindo o Brasil, apresentam apenas uma publicação cada, representando 6% da fatia.

Dos 18 artigos analisados, 14 são classificados como pesquisa experimental e 3 como pesquisa aplicada; 10 deles apresentaram ferramentas, *softwares* ou algoritmos desenvolvidos pelos autores. Alguns desses trabalhos são apenas simulações e não tiveram como resultado o desenvolvimento de uma ferramenta em si, mas apresentam resultados satisfatórios que justificariam desenvolver uma ferramenta.

Vale ressaltar que muitas dessas pesquisas utilizaram em suas amostras um volume relativamente pequeno de documentos e poucos deles apresentaram o uso diretamente em um GED funcional. Esses dados nos fazem questionar a maturidade dessas técnicas e seu potencial de aplicação em sistemas de GED.

Um ponto que despertou curiosidade é que na análise de 18 artigos científicos, não há indicação ou menção de técnicas de CD. Mesmo no trabalho apresentado por Mikheev e Yakimov (2019), que possui o termo *Data Science* descrito no campo de palavra-chave, não há indicativo de associação com esta área. No entanto, todos os trabalhos experimentais utilizaram suas técnicas na etapa de pré-processamento de dados.

A etapa de pré-processamento é essencial para garantia da qualidade dos resultados das abordagens de IA e aprendizagem de máquina, segundo Krasnyanskiy, Obukhov e Solomatina (2019:4, tradução nossa), esta etapa “permite melhorar significativamente a precisão dos resultados obtidos na aplicação do aprendizado de máquina.”

Embora não haja menção explícita às abordagens de CD nos trabalhos, as técnicas de pré-processamento de dados utilizadas refletem os fundamentos dessa área. Isso sugere que: por um lado, uma possível subvalorização ou desconhecimento de suas origens na CD, reforçando a necessidade de maior integração terminológica e conceitual entre as áreas. Por outro lado, poderia indicar que os pesquisadores da CI já internalizaram os conceitos e métodos da CD a tal ponto que os aplicam naturalmente, sem a necessidade de explicitá-las.

7.1. Principais pontos de interseção entre Gestão Eletrônica de Documentos, Inteligência Artificial e Ciência de Dados

Com as análises dos documentos, procurou-se responder às três questões de pesquisa definidas no Protocolo de RSL. Na primeira questão de pesquisa procurou-se entender quais são os principais pontos de interseção entre GED, IA e CD.

Em resposta a esta pergunta, identificou-se que os principais pontos de interseção entre GED e IA são justamente as tecnologias, desenvolvidas para automatizar processos, melhorar a agilidade e eficiência das ações e proporcionar maior compreensão das informações relevantes nos documentos. Isso resulta em uma visão mais estratégica para gestores, profissionais da informação e usuários finais.

Além disso, a integração de IA na GED permite análises preditivas, identificação de padrões, aprimoramento da segurança da informação, otimização no acesso e recuperação de dados, uso de OCR e mineração de texto, e uma tomada de decisão mais fundamentada e rápida.

Com base nos temas centrais dos trabalhos analisados, realizou-se a categorização das tecnologias enquanto aplicação prática. Os trabalhos podem ser enquadrados em uma ou mais categorias, dependendo da diversidade de suas abordagens. Foram identificadas 16 categorias que estão distribuídas conforme apresentado na Fig. 2.

‘Classificação de documentos’ é o tópico com mais pesquisas, com 7 artigos, representa 24% das pesquisas, e ‘Identificação de similaridade entre documentos’ está em segundo lugar com 3 artigos.

Fig. 2 – Aplicações práticas da IA e CD na GED



Fonte: Elaboração própria.

Quadro 2 – Aplicações práticas de IA em Sistemas de Gestão Eletrônica de Documentos (SGED)

ITEM	DESCRIÇÃO	REFERÊNCIAS
Classificação de documentos	Classificação automática com base em suas características, como tipo de conteúdo, formato, termos-chave, metadados etc.	Gomes e Moraes Filho (2011); Liu, Hsaio, Lee e Chen (2013); Krasnyanskiy, Obukhov e Solomatina (2019); Onan, Atik e Yalçın (2021); Sambetbayeva <i>et al.</i> (2022); Tkachenko e Denisova (2022); Al Qady e Kandil (2013)
Identificação de similaridade entre documentos	Identificando automática de documentos similares ou de diferentes revisões ou versões.	Obukhov, Krasnyanskiy e Nikolyukin (2019); Sambetbayeva <i>et al.</i> (2022); Mikheev e Yakimov (2019)
Roteamento Inteligente de Documentos	Identificação automática da melhor rota para a aprovação, revisão ou execução de um documento dentro de um fluxo de trabalho.	Sambetbayeva <i>et al.</i> (2022); Ismagilov <i>et al.</i> (2019)
Classificação semântica de documentos	Organização de documentos com base no contexto e nas dependências dentro do conjunto de arquivos.	Al Qady e Kandil (2013); Al Qady e Kandil (2010).

Extração de informações de documentos	Identificação e extração automatizada de informações essenciais de documentos com formatos padronizados ou similares.	Kefali e Drabsia (2018); Bashir e Giri (2013)
Identificação de Documentos Relacionados	Identificação de conteúdos por temas ou contexto e relações semânticas, como contratos, anexos, e comunicações associadas.	Sambetbayeva <i>et al.</i> (2022); Al Qady e Kandil (2010)
Indexação automática de documentos e fluxos de trabalho	Automatiza a extração de informações visuais, como símbolos e conexões, agilizando a análise e organização de documentos com diagramas complexos, permitindo que a compreensão por máquina.	Bashir e Giri (2013); Sambetbayeva <i>et al.</i> (2022).
Adaptação e personalização da interface	Essa funcionalidade permite que o sistema se adapte ao usuário segundo seu histórico, características e padrões de atividades.	Krasnyanskiy <i>et al.</i> (2020).
Aprimorar a Experiência do Usuário	Acumula informações sobre os erros mais frequentes dos usuários, identifica fragilidades e propõe melhorias nos SGED	Fil, Nefedov e Binkovskaya (2020).
Autenticação biométrica	Assegura que todas as operações são executadas apenas por usuários devidamente autenticados	Fayziev <i>et al.</i> (2021).
Comparação de versões documentos	Análise e comparação automática de diferentes versões de um documento para identificar alterações entre elas, como diferenças de conteúdo textual (adição, exclusão ou modificação de frases, palavras ou caracteres); alterações na formatação ou estrutura e atualizações em seções específicas (tabelas, parágrafos, cabeçalhos)	Mikheev e Yakimov (2019).
Construção de Ontologias	Análises semânticas automáticas para realizar a criação de ontologias, representando relações e temas principais de documentos	Al Qady e Kandil (2010).
Correção Ortográfica e Semântica	Essa funcionalidade melhora a qualidade das informações em SGED nos textos, metadados, fluxos de trabalho e processos documentais.	Jumanov e Karshiev (2020).
Otimização de Desempenho	Arquitetura baseada em microsserviços e message brokers permite uma melhoria significativa do desempenho de subsistemas de PLN.	Ruiz <i>et al.</i> (2020).
Registro Automático de Documentos	Registro de novos documentos de forma automática, preenchendo campos de metadados como data, autor, tipo de documento e contexto, sem a necessidade de intervenção manual.	Sambetbayeva <i>et al.</i> (2022).
Suporte à tomada de decisões	Com base em padrões de uso e histórico de operações do usuário, o subsistema pode sugerir as próximas etapas do fluxo de trabalho.	Obukhov, Krasnyanskiy e Nikolyukin (2019)

Fonte: Elaboração própria.

Para concluir a resposta à primeira questão de pesquisa (Quais são os principais pontos de interseção entre Gestão Eletrônica de Documentos, Inteligência Artificial e Ciência de Dados?), é necessário explorar os objetivos em comum dessas áreas:

- **Melhoria da eficiência e precisão:** A CI busca eficiência por meio de padrões, métodos e técnicas para organizar e gerenciar informações de forma estruturada. A IA e a CD complementam essa busca ao automatizar processos, aplicar algoritmos avançados para identificar padrões e aumentar a precisão no tratamento da informação.
- **Otimização do acesso à informação:** A CI desenvolve sistemas e práticas para garantir a acessibilidade, organização e preservação dos dados. IA e CD, por sua vez, ampliam essa capacidade com o uso de tecnologias como aprendizado de máquina e análise preditiva, que personalizam e agilizam o acesso às informações relevantes, mesmo em grandes volumes de dados.
- **Suporte à tomada de decisão:** A CI contribui ao oferecer informações organizadas e confiáveis como base para decisões estratégicas. IA e CD aumentam o impacto ao fornecer análises avançadas, insights gerados por modelos preditivos e sistemas que oferecem recomendações automatizadas e baseadas em dados.
- **Melhoria da segurança da informação:** Enquanto a CI define as políticas e práticas para proteção dos dados, IA e CD trazem inovações como detecção de anomalias, sistemas de autenticação avançados e análises de riscos para fortalecer a segurança.
- **Aprimoramento da experiência do usuário:** A CI foca em projetar sistemas acessíveis e fáceis de usar, enquanto IA e CD introduzem recursos como interfaces personalizadas e interativas que tornam a experiência mais intuitiva e eficaz.

7.2. Desafios, tendências e lacunas de pesquisa estão presentes na interseção dessas áreas de IA, CD e GED

Na segunda questão de pesquisa procurou-se entender quais desafios, tendências e lacunas de pesquisa estão presentes na interseção dessas áreas de IA, CD e GED.

As principais dificuldades enfrentadas pelos autores nos estudos foram reunidas e analisadas, evidenciando barreiras técnicas e conceituais. Embora nem todos tenham descrito explicitamente esses desafios, como em Neves (2020), foi possível identificá-los indiretamente por meio das metodologias apresentadas, detalhamento dos passos e etapas das atividades, e, em alguns casos, pelos próprios resultados das pesquisas.

Esses elementos indicam questões como limitações no acesso a dados, complexidade no desenvolvimento de tecnologias, e desafios na replicação ou validação de soluções, destacando áreas que exigem maior atenção e aprimoramento.

A integração de IA em sistemas de GED enfrenta diversas lacunas que limitam seu avanço e adoção. A baixa produção científica da área de CI, a falta de padrões específicos para a

aplicação de IA em GED resulta em abordagens fragmentadas, dificultando a interoperabilidade e replicabilidade das soluções.

A falta de integração com fluxos de trabalho preexistentes gera uma desconexão entre os sistemas tradicionais de GED e as novas tecnologias baseadas em IA dificultando seu uso de maneira eficaz. Os estudos analisados indicam que as aplicações propostas são majoritariamente pontuais, com apenas um estudo teórico oferecendo uma visão holística do SGED.

Mesmo quando focando na melhoria de processos, automação e agilidade, as soluções abordam apenas funcionalidades e aspectos técnicos, sem considerar estudos sobre uso e usuário ou análises sobre a aceitação e relevância prática entre bibliotecários e usuários finais.

Além disso, os impactos éticos e sociais dessas tecnologias, como questões de privacidade, transparência e impacto no emprego, são ainda pouco explorados, apesar de sua relevância crescente (RIBEIRO, 2024). Outra barreira significativa é a baixa adoção por pequenas e médias empresas, frequentemente devido a custos elevados, falta de conhecimento técnico e percepção de complexidade.

Outra lacuna identificada é a falta de envolvimento de profissionais da informação nas diferentes etapas do desenvolvimento dessas tecnologias. Esses profissionais possuem conhecimento especializado essencial que poderia colocá-los na vanguarda do desenvolvimento tecnológico (NEVES, 2020; RIBEIRO, 2024; MOREIRO-GONZÁLEZ *et al.*, 2024).

Seu conhecimento em gestão da informação, organização de dados, análise crítica e compreensão de fluxos de trabalho pode contribuir significativamente para o design, implementação e avaliação de tecnologias inovadoras, especialmente em processos que envolvem IA, CD e GED.

Com esta pesquisa identificou-se importantes oportunidades e desafios para o desenvolvimento e aprimoramento do Profissional da Informação no contexto da integração entre IA, CD e GED:

Desafios: Envolvem o aprofundamento no domínio das abordagens de IA, especialmente no trabalho colaborativo com profissionais de Tecnologia da Informação. Isso exige a compreensão da lógica das tecnologias, da linguagem técnica utilizada e das possibilidades de aplicação prática. Além disso, é necessário lidar com o ritmo acelerado das inovações e o volume crescente de avanços tecnológicos, que demandam constante atualização.

Oportunidades: Todas as abordagens de IA e aprendizado de máquina precisam de treinamento, o que torna indispensável a participação de profissionais com conhecimento aprofundado na área de atuação. Esses especialistas são responsáveis por tarefas essenciais, como análise e mapeamento de áreas de conhecimento, definição de critérios de classificação, criação de categorias e classes de documentos, além da identificação, análise e categorização de tipos, formatos e estruturas documentais.

Essas atividades são intrinsecamente humanas e não podem ser delegadas a sistemas automatizados, destacando a importância estratégica do Profissional da Informação neste contexto. Dessa forma, a pesquisa identifica que a integração entre IA, CD e GED não

apenas reforça a relevância dos profissionais da área, mas também amplia o escopo de suas competências, exigindo uma atuação interdisciplinar, proativa e alinhada às inovações tecnológicas.

Ribeiro (2024:61) também reflete que a adaptação a este “novo mundo” passa não só pela reformulação dos conteúdos curriculares, mas fundamentalmente pela adaptação e alteração substancial dos métodos pedagógicos. Moreira-González *et al.* (2024) destacam, com base em sua pesquisa tomando como fonte os requisitos de habilidades detetados em anúncios de emprego para profissionais da CI, uma das principais habilidades para atuar na gestão é o envolvimento com a análise e organização de dados.

Por fim, durante a análise da literatura, ficou evidente que as pesquisas frequentemente apresentam uma abordagem técnica focada no desenvolvimento de tecnologias e inovação tecnológica. Poucos estudos exploram o uso estratégico da informação, um elemento essencial para impulsionar a sociedade e fomentar inovações em múltiplos níveis.

Ao investigar o cenário da inovação, nota-se que o foco se limita quase exclusivamente às inovações tecnológicas, deixando de lado abordagens importantes como inovações de processos. Este aspecto inclui mudanças nos fluxos de trabalho, formas de produção, recebimento, armazenamento e distribuição de informações e documentos.

As inovações apresentadas estão, em sua maioria, restritas aos sistemas de gestão, sem explorar modificações nos processos fundamentais que os sustentam. Portanto, torna-se fundamental inovar nos processos, desconstruindo as concepções tradicionais sobre documentos e repensando como as informações circulam nos fluxos organizacionais. Uma abordagem mais integrada e interconectada possibilita uma visão dinâmica e estratégica da informação, promovendo novas práticas e formas de gestão alinhadas às demandas contemporâneas.

7.3. Influência nas práticas de GED das aplicações de IA e CD

Na terceira questão de pesquisa procurou-se entender como as aplicações de IA e CD têm influenciado as práticas na GED.

A aplicação de IA e CD está promovendo mudanças significativas nas práticas organizacionais de GED, transformando atividades manuais e repetitivas em processos automatizados e inteligentes. As pesquisas mostram muitos esforços na descoberta de informações a partir dos documentos produzidos, transformando a visão de meros arquivadores de documentos para geradores de conhecimento através da análise de documentos e processos. Essas mudanças refletem em melhorias perceptíveis, dentre elas destacam-se:

Automação e Eficiência: abordagens de IA automatizam tarefas como classificação de documentos, extração de informações e geração de relatórios.

Na busca por eficiência, a pesquisa desenvolvida por Ruiz *et al.* (2020) apresenta melhorias no processamento dos dados, com uma arquitetura baseada em microsserviços que melhora o desempenho do software, diminui o uso de recursos da CPU e memória.

Quanto à automação de processos de classificação e categorização de textos e documentos, Al Qady e Kandil (2010) e Kefali e Drabsia (2018) desenvolvem um algoritmo que tenta

identificar conceitos existentes nos documentos e extrair informações específicas de documentos digitalizados.

Análise Preditiva: A análise de dados em grande escala permite prever demandas, otimizar fluxos de trabalho e tomar decisões baseadas em insights gerados por IA.

Os autores Ismagilov *et al.* (2019) e Sambetbayeva *et al.* (2022) criaram sistemas que identificam e distribuem tarefas automaticamente, para especialistas a partir da análise das características do documento, como o tipo, conteúdo, e destinatários esperados. Ainda neste sentido os autores Onan, Atik e Yalçın (2021) desenvolveram uma abordagem para classificar e direcionar corretamente as solicitações de suporte de serviços, através da análise de relatórios de bug e histórico de solicitações.

Personalização: Sistemas baseados em IA adaptam-se às necessidades dos usuários, oferecendo interfaces e funcionalidades personalizadas.

No tema da personalização, os autores Krasnyanskiy *et al.* (2020) desenvolveram um algoritmo que possibilite a adaptação do Sistema de GED às características assimiladas do usuário a partir de seu histórico, características e padrões de atividades.

Suporte à tomada de decisões: Sistemas de IA fornecem subsídios valiosos ao extrair, organizar e analisar informações relevantes, transformando dados brutos em insights acionáveis.

Os autores Obukhov, Krasnyanskiy e Nikolyukin (2019) desenvolveram um subsistema que se baseia nos padrões de uso do sistema e no histórico de operações do usuário, para sugerir próximas ações associadas à sequência de suas atividades. Isso faz com que o usuário tome ações mais assertivas, evita erros e “cliques” desnecessários e agiliza o processo de trabalho.

Segurança da Informação: Sistemas de IA fortalecem a segurança da informação ao detectar ameaças em tempo real, prevenir vazamentos de dados e monitorar padrões de acesso.

Os autores Fayziev *et al.* (2021) identificaram que a combinação de chaves criptográficas com a autenticação biométrica combinada garante mais segurança e evita acessos não autorizados, alterações indevidas, e assegura que todas as operações são executadas apenas por usuários devidamente autenticados.

Como destacado, as pesquisas mostram diversos benefícios decorrentes das aplicações levantadas neste trabalho. Os benefícios identificados evidenciam apenas os impactos positivos para os sistemas e seus usuários. Que incluem maior eficiência nos processos de gestão documental, redução no tempo de processamento e recuperação de informações, aprimoramento da precisão na classificação de documentos, melhoria na experiência do usuário e suporte aprimorado para tomada de decisões estratégicas.

Não foram identificadas discussões ou abordagens sobre possíveis impactos negativos ou desafios associados a esses desenvolvimentos. As mudanças nas práticas de GED, além de transformarem processos e atividades, podem provocar resistência significativa de operadores e gestores. Dentre os motivos, destacam-se o medo da substituição de empregos por tecnologias e a dificuldade em aprender novos sistemas e se adaptar às mudanças no fluxo de trabalho.

Esses fatores podem criar barreiras importantes para a adoção de inovações, exigindo estratégias claras de treinamento, comunicação e envolvimento das equipes para superar as resistências e maximizar os benefícios das novas tecnologias. É importante observar que essas tecnologias impactam diretamente as funções tradicionais de bibliotecários e arquivistas. Atividades antes exclusivas desses profissionais estão sendo transferidas para máquinas, exigindo uma reorientação na formação e atuação dessas categorias.

Em vez de executar tarefas repetitivas, é necessário focar no desenvolvimento e aprimoramento das ferramentas de IA voltadas para esse fim. Essa colaboração reforça o papel estratégico dos bibliotecários e arquivistas nas equipes multidisciplinares, enquanto programadores e desenvolvedores lidam com a codificação e validação técnica.

Assim, repensar a carreira e adotar uma postura de constante adaptação tornou-se fundamental para evitar a obsolescência, garantindo a relevância e a contribuição ativa desses profissionais em um cenário digital em rápida evolução.

Outro aspecto impactado pelas mudanças realizadas por essas tecnologias emergentes é a confiança na Inteligência Artificial e as questões éticas associadas. A democratização do acesso a essas tecnologias levanta preocupações sobre a transparência dos algoritmos, a confiabilidade das decisões automatizadas e a mitigação de vieses. Esses vieses podem afetar diretamente a equidade nas operações e decisões organizacionais, destacando a necessidade de desenvolver soluções justas, explicáveis e alinhadas com princípios éticos, garantindo confiança e inclusão nos processos mediados por IA.

8. Considerações finais

O presente artigo apresentou o processo e os resultados de uma revisão sistemática da literatura, cujo objetivo principal foi investigar o papel de tecnologias oriundas das áreas de IA e CD na GED.

O protocolo da revisão sistemática foi elaborado seguindo diretrizes consolidadas na literatura, conforme descrito. A *string* de busca passou por sucessivos refinamentos a partir de uma metodologia de prototipagem que analisa a relevância dos termos.

Com os resultados em mãos, as duplicatas foram retiradas e os critérios de inclusão, exclusão e qualidade foram aplicados. Com a formação do corpus final de análise procedeu-se, então, para a leitura do texto integral e análise subsequente de todos esses trabalhos.

A abordagem utilizada para análise dos dados foi a análise temática e de conteúdo, que se mostrou eficaz ao classificar e organizar elementos significativos em categorias que descreveram e interpretaram os dados coletados.

As principais contribuições da IA e CD na GED incluem automação de tarefas, análises preditivas, identificação de padrões, segurança aprimorada e maior acessibilidade à informação. A categorização das aplicações práticas, com 16 categorias distintas, destaca áreas como a classificação e similaridade de documentos como as mais exploradas, evidenciando a relevância do tema.

Entre os desafios, destacam-se a falta de critérios padronizados para avaliar algoritmos, diferentes denominações para as mesmas abordagens, necessidade de grande volume de

dados para treinamentos e testes, dificuldade de acesso a dados reais para simulações, complexidade de documentos técnicos e limitação na integração com fluxos de trabalho existentes.

O avanço dessas tecnologias na GED caminha para a automação dos processos manuais, como catalogação, classificação, indexação, apoio a pesquisas, até mesmo certas tarefas de apoio ao usuário. Com isso, essa mudança impacta diretamente as funções dos profissionais da informação como bibliotecários e arquivistas, demandando um realinhamento profissional. Repensar a carreira e adaptar-se é crucial para evitar a obsolescência e garantir a relevância desses profissionais em um ambiente digital em constante evolução.

As tecnologias estão promovendo automação, eficiência e suporte estratégico por meio da análise preditiva, personalização, tomada de decisões aprimoradas e maior segurança da informação. No entanto, apesar dos benefícios, como redução de erros, agilidade nos processos e melhorias na experiência do usuário, pode haver resistência organizacional, especialmente relacionada ao medo de substituição de empregos e desafios no aprendizado de novos sistemas. Além disso, preocupações éticas como transparência, viés nos algoritmos e confiança na IA também se destacaram como fatores críticos para uma implementação bem-sucedida e inclusiva dessas inovações.

Apesar de avanços notáveis, foram identificadas lacunas na padronização de terminologias e descrições das abordagens utilizadas. Assim, propõe-se o desenvolvimento de um tesauro multidisciplinar como solução para padronizar e enriquecer as práticas no campo, potencializando a colaboração entre as áreas e o impacto das inovações tecnológicas.

Esta pesquisa abriu novas possibilidades para estudos futuros, abordando temas relevantes e novas abordagens no campo da GED. Entre as perspectivas destacam-se: a análise da gestão de registros de saúde, focando nas tecnologias de classificação de sintomas e previsões de doenças; o estudo de documentos técnicos de engenharia devido à sua complexidade e impacto em projetos; e uma exploração das tecnologias para GED no GitHub para identificar ferramentas testadas que são passíveis de inovação. Outra abordagem seria a análise comparativa entre patentes e pesquisas científicas para mapear tendências tecnológicas.

Referências bibliográficas

AL QADY, M.; KANDIL, A.

2013 Document discourse for managing construction project documents. *Journal of Computing in Civil Engineering*. [Em linha]. 27:5 (2013) 466-475. [Consult. 6 ago. 2024]. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-84882345017&doi=10.1061%2f%28ASCE%29CP.1943-5487.0000201&origin=inward&txGid=b15598cf7b63358deae500d6188c3d01>.

AL QADY, M.; KANDIL, A.

2010 Concept relation extraction from construction documents using natural language processing. *Journal of Construction Engineering and Management*. [Em linha]. 136:3 (2010) 294-302 [Consult. 6 ago. 2024]. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-77149152182&doi=10.1061%2f%28ASCE%29CO.1943-7862.0000131&origin=inward&txGid=77e8ee8edd23c205b3dd4e52a2784896>.

BARDIN, L.

1977 *Análise de conteúdo*. São Paulo: Edições 70, 1977.

BASHIR, R.; GIRI, K. J.

2013 Diagram recognition: Domain knowledge based approach. In INTERNATIONAL CONFERENCE ON MACHINE INTELLIGENCE RESEARCH AND ADVANCEMENT - ICMIRA 2013, Katra (India) – *Proceedings*. [Em linha]. Katra: Shri Mata Vaishno Devi University, 2013, p. 445-449. [Consult. 6 ago. 2024]. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-84910010844&doi=10.1109%2fICMIRA.2013.94&origin=inward&txGid=3d088e97cdcf2d71dbf59c2944f6bbe>.

CALDERON, W. R. [et al.]

2024 O Processo de gestão de documentos e da informação arquivística no ambiente universitário. *Ciência da Informação*. [Em linha]. 33:3 (2024) 97-104, [Consult. 30 abr. 2024]. Disponível em: <http://revista.ibict.br/ciinf/article/view/1037>.

CAO, L.

2017 Data Science. *ACM Computing Surveys*. [Em linha]. 50:3 (2017) 1-42, [Consult. 7 ago. 2023]. Disponível em: <http://dx.doi.org/10.1145/3076253>.

CAPURRO, R.; HJØRLAND, B.

2003 The Concept of information. *Annual Review of Information Science and Technology*. 37:1 (2003) 343-411.

COZMAN, F. G.; NERI, H.

2021 O Que, afinal, é Inteligência Artificial? In *Inteligência Artificial: Avanços e tendências*. Org. Fabio G. Cozman, G. A. Plonski, H. Neri. São Paulo: Instituto de Estudos Avançados, 2021, cap. 1, p. 21-29. [Consult. 30 maio 2024]. Disponível em: <http://www.livrosabertos.sibi.usp.br/portaldelivrosUSP/catalog/book/650>.

ELIAS, E. D.

2012 Gerenciamento eletrônico de documentos (GED): aplicação na Universidade Federal de Santa Catarina. *Ágora*. [Em linha]. 22:45 (2012) 15-30 [Consult. 30 abr. 2024]. Disponível em: <https://brapci.inf.br/index.php/res/download/48043>.

FAYZIEV, S. I.; et al.

2021 Development of a methodology for protecting a software package for electronic document management. In INTERNATIONAL CONFERENCE ON INFORMATION SCIENCE AND COMMUNICATIONS TECHNOLOGIES – ICISCT 2021, Tashkent (Uzbekistan) 2021 – *Proceedings*. [Em linha]. 2021. [Consult. 6 ago.2024]. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85125054039&doi=10.1109%2fICISCT52966.2021.9670416&origin=inward&txGid=9f673b2b8d59cd43573ceb653331673b>.

FIL, N.; NEFEDOV, L.; BINKOVSKAYA, A.

2020 Fuzzy model for estimating the probability of user error in the electronic document management system. In IEEE INTERNATIONAL CONFERENCE ON PROBLEMS OF INFOCOMMUNICATIONS SCIENCE AND TECHNOLOGY - PIC S AND T – *Proceedings*. Kharkiv (Ukraine), 2020, p. 259-262, [Consult. 6 ago.2024]. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85114404836&doi=10.1109%2fPICST51311.2020.9467950&origin=inward&txGid=2804e3d47fc9752cfc12b4ba1d4723e5>.

FONSECA, F.

2021 Whether or when: the question on the use of theories in data science. *Journal of the Association for Information Science and Technology*. [Em linha]. 72:12 (2021) 1.593-1.604 [Consult. 15 fev. 2022]. Disponível em: <http://dx.doi.org/10.1002/asi.24537>.

FONSECA, C. A.; GUELPELI, M. V. C.; SOUZA NETTO, R. S.

2022 Representação dos dados estruturados do gênero textual como técnica para o processamento automático de texto. *Texto Livre*. [Em linha]. 15 (2022) e35445. [Consult. 29 maio 2023]. Disponível em: <https://periodicos.ufmg.br/index.php/textolivres/article/view/35445>.

GOMES, R. R. G.; MORAES FILHO, R. O.

2011 CADWeb : Categorização automática de documentos digitais. *Ciência da Informação*. [Em linha]. 40:1 (2011) 68-76, [Consult. 21 ago. 2024]. Disponível em: <https://research.ebsco.com/linkprocessor/plink?id=8b77781e-40a8-3b17-ae64-bb666e42a8f8>.

ISMAGILOV, I. I.

2019 Decision-making support system for tasks distribution in personal data operators register maintaining based on a fuzzy-production model. *Revista Dilemas Contemporâneos: Educación, Política y Valores*. [Em linha]. 4:64 (2019) 17. [Consult. 6 ago. 2024]. Disponível em: <https://www.dilemascontemporaneoseducacionpoliticayvalores.com/index.php/dilemas/article/view/557>.

JUMANOV, I. I.; KARSHIEV, K. B.

2020 Mechanisms for optimization of detection and correction of text errors based on combining multilevel morphological analysis with n-gram models. In IV INTERNATIONAL SCIENTIFIC AND TECHNICAL CONFERENCE "MECHANICAL SCIENCE AND TECHNOLOGY UPDATE" (MSTU-2020), Omsk (Russia) 2020 – *Proceedings*. [Em linha]. 2020. (Journal of Physics: Conference Series; 1.546). [Consult. 6 ago. 2024]. Disponível em: <https://iopscience.iop.org/article/10.1088/1742-6596/1546/1/012082>.

KEFALI, A.; DRABSIA, S.

2020 Localization of scores and average in Algerian baccalaureate transcripts. In INTERNATIONAL CONFERENCE ON ENTERPRISE INFORMATION SYSTEMS (ICEIS 2020), 22nd – *Proceedings*. [Em linha]. 2020. [Consult. 6 ago.2024]. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85090796518&origin=inward&txGid=51e6dc5575bc30f0815a7b966ec72e13>.

KITCHENHAM, B.

2004 *Procedures for performing systematic reviews.: Joint technical report Software Engineering Group, Keele University (TR/SE-0401)*. United Kingdom: Empirical Software Engineering; Australia: National ICT Australia, 2004. (0400011T.1).

KITCHENHAM, B.; CHARTERS, S.

2007 *Guidelines for performing systematic literature reviews in software engineering: Technical report*. [Em linha]. [S. l.]: EBSE, Keele University; Durham University, 2007. [Consult. 16 jul. 2023]. Disponível em: https://www.elsevier.com/data/promis_misc/525444systematicreviewsguide.pdf.

KRASNYANSKIY, M. N.; OBUKHOV, A. D.; SOLOMATINA, E. M.

2019 The Algorithm of document classification of research and education institution using machine learning methods. In INTERNATIONAL SCIENCE AND TECHNOLOGY CONFERENCE – EASTCONF 2019, Vladivostok (Russia), 2019 – *Proceedings*. [Em linha]. 2019. [Consult. 6 ago. 2024]. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85067504184&doi=10.1109%2fEastconf.2019.8725319&origin=inward&txGid=c1a2253c86382e933d550273593979b3>.

KRASNYANSKIY, M. N. [et al.]

2020 The Algorithm of document routing in the electronic document management system using machine learning methods. *Progress in Artificial Intelligence*. 9:4 (2020) 287-303.

LIU, C. L. [et al.]

2013 Clustering tagged documents with labeled and unlabeled documents. *Information Processing and Management*. [Em linha]. 49 (2013) 596-606 [Consult. 21 ago. 2024]. Disponível em: <https://research.ebsco.com/linkprocessor/plink?id=caaod78a-603f-3634-91b8-of846f21c729>.

MARCHIONINI, G.

2016 Information Science roles in the emerging field of Data Science. *Journal of Data and Information Science*. [Em linha]. 1:2 (2016) 6 [Consult. 15 dez. 2022]. Disponível em: http://manu47.magtech.com.cn/Jwk3_jdis/Y2016/V1/I2/1.

MIKHEEV, M. A.; YAKIMOV, P. Y.

2019 Development of the documents comparison module for an electronic document management system. In INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY AND NANOTECHNOLOGY (DS-ITNT 2019), 5th, Samara (Russia), 2019 - *Data Science Session*. [Em linha]. 2019, vol. 2.416, p. 527-533. [Consult. 6 ago. 2024]. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85070973402&doi=10.18287%2f1613-0073-2019-2416-527-533&origin=inward&txGid=dcb8969d1cf9eb76a28fccd362c6816c>.

MOREIRO-GONZÁLEZ, J. A. [et al.]

2024 The Skills of information professionals for the analysis and organization of data: Conceptual proposal and application of OpenAI in job advertisements on the Web. *Profesional de la información*, [Em linha]. 33:5 (2024). [Consult. 2 mar. 2025]. Disponível em: <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/87866>.

MOUTINHO, S. O. M. [et al.]

2024 Ciência da informação e Ciência de Dados: convergências interdisciplinares. *Encontros Bibli*. 29 (2024) 26.

NAKAMITI, E. K.

2009 *Agentes inteligentes artificiais*. [Em linha]. São Paulo, 2009, [Consult. 30 maio 2022]. Disponível em: <https://tede2.pucsp.br/handle/handle/5240>.

Dissertação de Mestrado em Comunicação - Pontifícia Universidade Católica de São Paulo.

NATIONAL HEALTH AND MEDICAL RESEARCH COUNCIL

2000 *How to review the evidence: systematic identification and review of the scientific literature*. Canberra: National Health and Medical Research Council, 2000.

NEVES, B. C.

2020 Inteligência artificial e computação cognitiva em unidades de informação: conceitos e experiências. *Logeion: Filosofia da Informação*. [Em linha]. 7:1 (2020) 186-205. [Consult. 1 mar. 2025]. Disponível em: <https://revista.ibict.br/fiinf/article/view/5260>.

NHACUONGUE, J. A.; DUTRA, M. L.

2022 A Terminologia em sistemas de recuperação da informação baseada na wordnet.pt. *Informação & Sociedade: Estudos*, [Em linha]. 30:2 (2022). [Consult. 29 maio 2022]. Disponível em: <https://brapci.inf.br/index.php/res/v/147979>.

OBUKHOV, A.; KRASNYANSKIY, M.; NIKOLYUKIN, M.

2019 Implementation of decision support subsystem in electronic document systems using machine learning techniques. In INTERNATIONAL MULTI-CONFERENCE ON INDUSTRIAL ENGINEERING AND MODERN TECHNOLOGIES (FarEastCon), Vladivostok (Russia) 2019 – *Proceedings*. [Em linha]. 2019. [Consult. 6 ago. 2024]. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85078569216&doi=10.1109%2fFarEastCon.2019.8934879&origin=inward&txGid=d0bf67a8cdfbcbf1b827b97450d62fbca>.

ONAN, A.; ATIK, E.; YALÇIN, A.

2020 Machine learning approach for automatic categorization of service support requests on university information management system. In INTERNATIONAL CONFERENCE ON INTELLIGENT AND FUZZY SYSTEMS - INFUS 2020, Istanbul (Turkey), 2020 – *Proceedings*. [Em linha]. 2020, vol. 1.197. [Consult. 6 ago. 2024]. Disponível em: https://www.scopus.com/record/display.uri?eid=2-s2.0-85088741109&doi=10.1007%2f978-3-030-51156-2_132&origin=inward&txGid=77fcoa67c502c7a13f249133bc044503.

RAUTENBERG, S.; CARMO, P. R. V.

2019 Big Data e Ciência de Dados: complementariedade conceitual no processo de tomada de decisão. *Brazilian Journal of Information Studies: Research Trends*. [Em linha]. 13:1 (2019) 56-67 [Consult. 6 ago. 2023]. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=6983493>.

RIBEIRO, F.

2024 A Formação superior em Ciência da Informação face aos desafios da Inteligência Artificial. *Páginas a&b: arquivos e bibliotecas*. [Em linha]. 3ª série, especial VI CIIBERCID 2023 (2024) 53-65. [Consult. 5-30-2024]. Disponível em: <https://ojs.letras.up.pt/index.php/paginasaeb/article/view/14006>.

RICHARDSON, R. J. [et al.]

2007 *Pesquisa social: métodos e técnicas*. 3ª ed. São Paulo: Atlas, 2007.

RUIZ, M. [et al.]

2020 uAIS: An experience of increasing performance of NLP information extraction tasks from legal documents in an electronic document management system. In INTERNATIONAL CONFERENCE ON ENTERPRISE INFORMATION SYSTEMS (ICEIS 2020), 22nd, 2020 – *Proceedings*. [Em linha]. 2020, vol. 1, p. 189-196. [Consult. 6 ago. 2024]. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85090796518&origin=inward&txGid=fc8cbd37a03d262d61ba5a7e4d1d8b2d>.

RUSSELL, S. J.; NORVIG, P.

2004 *Inteligência Artificial*. 2^a ed. Rio de Janeiro: Elsevier, 2004.

SAMBETBAYEVA, M. [et al.]

2022 Development of intelligent electronic document management system model based on machine learning methods. *Eastern-European Journal of Enterprise Technologies*. [Em linha]. 1:2 (2022) 68-76 [Consult. 6 ago. 2024]. Disponível em: <https://doi.org/10.15587/1729-4061.2022.251689>.

SANTAREM SEGUNDO, J. E.; CONEGLIAN, C. S.

2015 Tecnologias da Web semântica aplicadas à organização do conhecimento: padrão SKOS para construção e uso de vocabulários controlados descentralizados. In *Organização do Conhecimento e diversidade cultural*. Org. José Augusto Chaves Guimarães, Vera Dodebei. Marília: Fundepe, 2015, vol. 3, p. 224-233. [Consult. 29 maio 2022]. Disponível em: <https://isko.org.br/wp-content/uploads/2021/05/Proceedings-ISKO-Brasil-2015.pdf>.

SANTOS, F. G. S.; SILVA, M. N.; BARI, V. A.

2021 O Impacto da gestão documental nos arquivos de engenharia e arquitetura: uma análise do arquivo da DIPOP do Instituto Federal de Educação, Ciência e Tecnologia de Sergipe. *AtoZ: novas práticas em informação e conhecimento*. [Em linha]. 10:1 (2021) 85-93. [Consult. 6 ago. 2023]. Disponível em: <http://dx.doi.org/10.5380/atoz.v10i1.75160>.

SANTOS, H. M.; FLORES, D.

2016 O Documento digital no contexto das funções arquivísticas. *Páginas a&b: arquivos e bibliotecas*. [Em linha]. 3^a série, 5 (2016) 165-177. [Consult. 2 mar. 2025]. Disponível em: <https://ojs.letras.up.pt/index.php/paginasab/article/view/1477>.

SANTOS, H. M.; KRAWSZUK, G. L.

2020 Gestão do conhecimento organizacional: tratamento arquivístico para reuso da informação administrativa. *Investigación Bibliotecológica*. 34:83 (2020) 103-127. [Consult. 6 ago. 2023]. Disponível em: <https://doi.org/10.22201/iibi.24488321xe.2020.83.58146>

SANTOS, P. L. A. C.; VIDOTTI, S. A. B. G.

2009 Perspectivismo e tecnologias de informação e comunicação: acréscimos à Ciência da Informação? *Datagramazero*, [Em linha]. 10:3 (2009) 1-10. [Consult. 6 ago. 2024]. Disponível em: http://dgz.org.br/jun09/Art_02.htm.

SICHMAN, J. S.

2021 Inteligência Artificial e sociedade: avanços e riscos. *Estudos Avançados*, [Em linha]. 35:101 (2021) 37-50. [Consult. 06 ago. 2023]. Disponível em: <http://dx.doi.org/10.1590/s0103-4014.2021.35101.004>.

SIQUEIRA, I. S. P.; PEREIRA, A. E. C.

1989 Perspectivas de aplicação da Inteligência Artificial à Biblioteconomia e à Ciência da Informação. *Revista Brasileira de Biblioteconomia e Documentação*. 22:1/2 (1989) 39-80.

TKACHENKO, A. L.; DENISOVA, L. A.

2021 Designing an information system for the electronic document management of a university: Automatic classification of documents. In INTERNATIONAL SCIENTIFIC AND TECHNICAL CONFERENCE: APPLIED MECHANICS AND SYSTEMS DYNAMICS (AMSD 2021) 15th – *Proceedings*. [Em linha], 2021. [Consult. 6 ago. 2024]. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85127676818&doi=10.1088%2f1742-6596%2f2182%2f1%2f012035&origin=inward&txGid=d9c11035561cc71386f164f3f1c9db2a>.

VAJJALA, S. [et al.]

2020 *Practical natural language processing: Comprehensive guide to building real-world NLP systems*. United States of America: O'Reilly Media, 2020.

VIANNA, W. B.; FREITAS, M. C. V.

2019 Gestão da informação e Ciência da Informação: elementos para um debate necessário. *Ciência da Informação*. 48:2 (2019) 191-208.

WEITZEN, H. S.

1991 *O Poder da informação: como transformar a informação que você domina em um negócio lucrativo*. São Paulo: Markon, 1991.

Edenilza Valéria da Silva Magalhães | edenilza@gmail.com

Universidade Federal de São Carlos (UFSCAR), Brasil

Barbara Coelho Neves | babi.coelho7@gmail.com

Instituto de Ciência da Informação (ICI), Universidade Federal da Bahia (UFBA), Brasil