



25°enancib

Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação
Informação, Decolonialidade e Direitos Difusos para o Desenvolvimento Sustentável

3 a 7 de novembro de 2025 – Rio de Janeiro



XXV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - XXV ENANCIB

GT 8 – Dados, Informação e Tecnologia

AUTOMAÇÃO NA CIÊNCIA DA INFORMAÇÃO: UMA ABORDAGEM COM AGENTES DE IA

AUTOMATION IN INFORMATION SCIENCE: AN APPROACH WITH AI AGENTS

Elisângela Cristina Aganette – Universidade Federal de Minas Gerais (UFMG)

Rodrigo Fernandes dos Santos – Universidade Federal de Minas Gerais (UFMG)

Modalidade: Trabalho Completo

Resumo: a crescente complexidade na gestão de documentos digitais demanda soluções que aliem precisão técnica e escalabilidade. Este trabalho tem como objetivo apresentar um estudo técnico-experimental sobre a implementação de uma arquitetura computacional baseada em agentes inteligentes para leitura, extração e classificação automatizada de documentos. Fundamentado na interface entre Ciência da Informação e Inteligência Artificial, o estudo propõe a utilização de tecnologias emergentes como *CrewAI* e *LangChain* integradas a um modelo de linguagem avançado, para executar de forma coordenada tarefas informacionais complexas. A metodologia adotada é qualitativa, aplicada e de caráter exploratório, centrada no desenvolvimento e avaliação de um agente funcional testado sobre um *corpus* documental composto por vinte e cinco arquivos digitais relacionados à Escola de Ciência da Informação da UFMG. Os procedimentos envolveram extração textual automatizada, organização semântica e categorização segundo critérios informacionais predefinidos, com saída padronizada em formato estruturado. Os resultados demonstram a viabilidade da proposta, evidenciando o potencial dos agentes autônomos como mediadores algorítmicos na organização do conhecimento e no apoio a processos de curadoria informacional. As classificações obtidas foram validadas manualmente, garantindo a confiabilidade dos dados e apontando caminhos promissores para aplicações em larga escala. Conclui-se que a integração entre abordagens técnicas e fundamentos informacionais representa uma contribuição significativa para a mediação algorítmica e para o avanço da automação nos processos de organização da informação.

Palavras-chave: mediação algorítmica; organização da informação; inteligência artificial.

Abstract: the increasing complexity of digital document management demands solutions that combine technical accuracy with scalability. This study aims to present a technical-experimental analysis of the implementation of a computational architecture based on intelligent agents for the automated reading, extraction, and classification of documents. Grounded in the intersection between Information Science and Artificial Intelligence, the study proposes the use of emerging technologies such as *CrewAI* and *LangChain*, integrated with an advanced language model, to perform complex informational tasks in a coordinated manner. The adopted methodology is qualitative, applied, and exploratory in nature, focusing on the development and evaluation of a functional agent tested on a documentary corpus composed of twenty-five digital files related to the School of Information Science at UFMG. The procedures involved automated text extraction, semantic organization, and categorization according to predefined informational criteria, with output standardized in a structured format. The results demonstrate the feasibility of the proposed approach, highlighting the potential of autonomous agents as algorithmic mediators in knowledge organization and in supporting informational curation processes. The classifications obtained were manually validated, ensuring data reliability and suggesting promising paths for large-scale applications. It is concluded that the

integration of technical approaches with informational foundations represents a significant contribution to algorithmic mediation and to the advancement of automation in information organization processes.

Keywords: algorithmic mediation; information organization; artificial intelligence.

1 INTRODUÇÃO

A crescente produção de dados digitais impõe novos desafios à mediação e organização da informação, especialmente em instituições públicas, científicas e educacionais. Estima-se que, até 2025, o volume global de dados ultrapasse 175 *zettabytes* (Reinsel; Gantz; Rydning, 2018), o que pressiona arquivos, bibliotecas e repositórios diante da sobrecarga informacional.

Nesse cenário, a Inteligência Artificial (IA) apresenta-se como uma aliada estratégica na promoção da competência em informação, ao atuar de forma integrada na organização, produção e disseminação do conhecimento científico no ciberespaço, contribuindo para a mediação técnica e o fortalecimento de práticas informacionais qualificadas (Guimarães; Freire, 2023).

A crescente complexidade dos fluxos informacionais tem impulsionado a Ciência da Informação a incorporar sistemas inteligentes para a automação da organização do conhecimento. Como destaca Silva (2021), a integração de tecnologias baseadas em Inteligência Artificial, especialmente modelos de linguagem de larga escala (*Large Language Models – LLMs*) e algoritmos de aprendizado de máquina, redefine práticas informacionais ao viabilizar análise semântica, classificação automática e geração adaptativa de metadados em ambientes digitais.

Apesar da eficácia dos *LLMs* na análise semântica e indexação, suas limitações quanto à autonomia têm impulsionado o uso de agentes de IA orientados a objetivos (Russell; Norvig, 2022). No entanto, a falta de investigações empíricas sobre o funcionamento e os impactos desses agentes na organização da informação, especialmente em contextos institucionais, evidencia a necessidade de estudos que integrem dimensões técnicas, éticas e formativas no campo da mediação documental (Alves; Ribeiro, 2024).

Buscando responder a essa lacuna, esta pesquisa propõe e avalia uma arquitetura computacional, construída pelos autores, com agentes inteligentes para leitura, classificação e revisão automatizada de documentos, integrando fundamentos da Ciência da Informação

da Inteligência Artificial. Utilizando tecnologias como *CrewAI*¹ e *LangChain*², a solução articula a orquestração técnica de agentes a princípios teóricos, resultando em uma estrutura modular, replicável e adaptável. O agente atua de forma autônoma e transparente, qualificando serviços de gestão documental e promovendo a formação crítica de profissionais da informação.

A estrutura do texto organiza-se da seguinte forma: a Seção 2 apresenta o referencial teórico; a Seção 3 descreve a metodologia empregada; a Seção 4 discute os resultados obtidos; e a Seção 5 reúne as conclusões e as contribuições do estudo para o campo da Ciência da Informação.

2 REFERENCIAL TEÓRICO

O referencial teórico deste estudo examina a atuação de agentes de IA na organização do conhecimento, partindo da premissa de que agentes autônomos, orientados a objetivos, passam a exercer funções tradicionalmente atribuídas à mediação humana.

A análise estrutura-se em três eixos: (1) características dos agentes de IA em sistemas informacionais; (2) relações entre mediação algorítmica e representação da informação; e (3) evolução e desafios da adoção de agentes na Ciência da Informação. Essa abordagem visa compreender o papel estratégico desses agentes na automação de processos e seus efeitos nos paradigmas informacionais contemporâneos.

2.1 Agentes de Inteligência Artificial

Esta seção explora os fundamentos teóricos dos agentes de inteligência artificial, com foco em sua atuação como mediadores informacionais em ambientes digitais. A análise organiza-se em três eixos: definição conceitual, principais propriedades, como autonomia e adaptabilidade, e os desafios que esses agentes impõem à organização da informação.

A origem da noção de agente remonta à filosofia clássica, com Aristóteles, que conceituou a ação como dirigida a um fim racional (*telos*). No século XVIII, David Hume aprofundou essa concepção ao associar a ação à experiência e ao hábito. No século XX, Alan Turing introduziu a ideia de que máquinas poderiam simular comportamentos inteligentes,

¹ *CrewAI* é um *framework* em Python para coordenação de múltiplos agentes de IA especializados, permitindo a execução colaborativa de tarefas complexas com base em modelos de linguagem.

² *LangChain* é uma biblioteca que integra modelos de linguagem com documentos e ferramentas externas, organizando cadeias de prompts para análise semântica e automação informacional.

consolidada posteriormente no Teste de Turing³. Ainda nesse período, Daniel Dennett formulou a teoria dos sistemas intencionais, ampliando a compreensão sobre a atribuição de estados mentais a agentes artificiais. Na década de 1970, Maturana e Varela (1995) propuseram o conceito de *autopoiese*, enfatizando a autonomia refere-se a sistemas capazes de se auto-organizar e se manter por meio da regeneração contínua de seus próprios componentes.

Posteriormente, nas décadas de 1980 e 1990, com o avanço da IA simbólica e dos sistemas especialistas, agentes passaram a ser formalmente definidos como sistemas capazes de perceber o ambiente, processar informações e agir orientados a objetivos (Russell; Norvig, 2022). Recentemente, Floridi *et al.* (2025) introduziram o conceito de *Agentic AI Optimisation* (AAIO), ressaltando que agentes autônomos de IA representam uma nova fase da interação digital, na qual sistemas não apenas reagem, mas otimizam ativamente suas próprias ações em ambientes digitais, gerando impactos éticos, sociais e informacionais que exigem novas abordagens regulatórias e interpretativas.

A diferenciação entre modelos generativos e agentes inteligentes tornou-se central para a análise da mediação informacional contemporânea. Floridi *et al.* (2025) observa que os modelos de *LLMs*, como ChatGPT e DeepSeek, operam de maneira reativa, limitando-se à geração de linguagem com base em padrões aprendidos a partir de grandes volumes de dados. Em contraste, agentes de IA são caracterizados pela atuação autônoma e proativa, capazes de tomar decisões contextuais e interagir dinamicamente com o ambiente digital.

Essa distinção demanda estratégias distintas de otimização: enquanto os *LLMs* concentram-se na produção linguística, os agentes exigem arquiteturas adaptativas voltadas à ação e à tomada de decisão. Na CI, compreender essa diferenciação é fundamental para reformular práticas de organização e mediação do conhecimento em ecossistemas digitais. Apesar do reconhecimento da autonomia e da capacidade de aprendizado dos agentes (Wooldridge *et al.*, 2023), ainda são escassos os estudos empíricos que investigam sua aplicação em tarefas como classificação documental e gestão de acervos.

Neste estudo, os agentes são compreendidos como mediadores algorítmicos especializados na automação de tarefas como leitura e categorização de documentos digitais. Com o uso de *frameworks* colaborativos, como o *CrewAI*, desenvolvem-se agentes que atuam

³ Proposto por Alan Turing em 1950 no artigo "*Computing Machinery and Intelligence*", o teste consiste em avaliar se uma máquina pode exibir comportamento inteligente indistinguível do de um ser humano.

de forma coordenada, simulando práticas de curadoria informacional, compreendida como a seleção, organização e avaliação de conteúdos com base em critérios como relevância, confiabilidade e atualidade. Essa abordagem fundamenta a arquitetura metodológica adotada e reforça o papel estratégico dos agentes na transformação digital dos fluxos informacionais.

2.2 Representação e Mediação da Informação

Esta seção discute os fundamentos da representação e mediação da informação, com ênfase nas transformações advindas da incorporação de tecnologias de IA. A análise organiza-se em torno dos conceitos clássicos, dos fundamentos da mediação e da emergência da mediação algorítmica em ambientes digitais.

A representação da informação constitui um dos pilares da Ciência da Informação, reunindo práticas de seleção, categorização e estruturação de conteúdos voltadas à recuperação e uso qualificado (Campos, 2000). Esse processo transcende a organização formal, envolvendo a construção de significados baseada em critérios semânticos, contextuais e normativos. Buckland (1991) amplia essa concepção ao entender a informação como coisa, processo e conhecimento, destacando que a representação abarca tanto o conteúdo quanto os meios de sua materialização e acesso.

No cenário digital contemporâneo, a mediação da informação expande-se, sendo realizada por sujeitos humanos e por sistemas algorítmicos. Enquanto a mediação humana, como a exercida por bibliotecários e curadores, ancora-se em escolhas intencionais e interpretações situadas (Capurro, 2000; Le Coadic, 1996), a mediação algorítmica opera por modelos automatizados, classificando e organizando conteúdos com base em padrões estatísticos e lógicas preditivas.

Gillespie (2018) observa que algoritmos atuam como infraestruturas de visibilidade, moldando o que é visto, encontrado e valorizado. Ananny e Crawford (2018) enfatizam que tais sistemas embutem decisões de design e critérios de relevância, impactando diretamente o acesso à informação. Ao comparar mediações humanas e algorítmicas, nota-se uma mudança significativa nos graus de agência, opacidade e controle sobre os fluxos informacionais, com implicações relevantes para a organização do conhecimento.

Considerando os agentes de IA como sistemas autônomos e adaptativos, esta pesquisa propõe uma releitura dos processos de representação e mediação da informação. Ao realizarem tarefas como leitura e categorização documental, os agentes atuam como

mediadores algorítmicos, influenciando os fluxos informacionais com base em inferências estatísticas, como exemplifica a plataforma *Semantic Scholar*. Embora ampliem a escala da mediação humana, esses sistemas trazem riscos associados ao viés algorítmico, à opacidade decisória e à imprecisão interpretativa.

Tal cenário demanda uma abordagem crítica sobre a delegação de funções mediadoras a sistemas autônomos, consolidando a mediação algorítmica como perspectiva analítica central para compreender os impactos da Inteligência Artificial na Ciência da Informação.

2.3 Agentes de IA na Ciência da Informação: Evolução e Desafios Contemporâneos

Esta seção explora a interface entre Inteligência Artificial (IA) e Ciência da Informação (CI), com foco na evolução histórica, operacional e conceitual dos agentes inteligentes em processos informacionais. Desde sua origem interdisciplinar, a CI vem incorporando tecnologias computacionais como resposta à crescente complexidade dos fluxos informacionais.

Nos anos 1980, sistemas especialistas como o PLEXUS (1980–1983) e o SCISOR (1987–1989) buscaram automatizar tarefas de indexação e categorização, embora fossem limitados por regras fixas e ausência de aprendizado adaptativo (Woods, 1986). Na década seguinte, projetos como SHOE (1996) e Ontobroker (1997) inauguraram abordagens baseadas em ontologias e ambientes distribuídos, promovendo a mediação semântica do conhecimento. Já nos anos 2000, arquiteturas multiagente foram aplicadas em sistemas como CiteSeer e *Microsoft Academic Search*, viabilizando a curadoria automatizada da literatura científica.

O avanço do aprendizado profundo, a partir de 2010, potencializou o surgimento de agentes mais autônomos e responsivos, como nas plataformas *Semantic Scholar* e *Iris.ai*, capazes de integrar mineração textual, análise semântica e recomendação científica personalizada.

Do ponto de vista teórico, a teoria da cognição distribuída (Hollan; Hutchins; Kirsh, 2000) compreende os processos informacionais como redes sociotécnicas que integram humanos, artefatos e ambientes digitais. Contudo, a intensificação do uso de IA evidencia preocupações éticas e políticas. Desafios como viés algorítmico, opacidade decisória e riscos à governança da informação são discutidos por Zuboff (2019), Buolamwini e Gebu (2018).

Neste contexto, esta pesquisa propõe uma arquitetura baseada em múltiplos agentes de IA para a leitura, categorização e mediação de documentos digitais, concebendo-os como

mediadores algorítmicos que reconfiguram os processos de organização do conhecimento e exigem novas competências técnicas e críticas no campo da Ciência da Informação.

3 METODOLOGIA

A presente pesquisa caracteriza-se como qualitativa, de natureza aplicada e de caráter exploratório, voltada à investigação das potencialidades da Inteligência Artificial (IA) na mediação e organização da informação em ambientes digitais. A abordagem qualitativa justifica-se pela necessidade de compreender significados e interpretações em contextos sociotécnicos complexos (Creswell, 2014), enquanto a natureza aplicada decorre do desenvolvimento de uma solução funcional, replicável em ambientes reais.

O delineamento metodológico adotado foi o de um estudo de caso técnico-experimental, focado na construção e avaliação de uma arquitetura computacional orientada por agentes autônomos. A pesquisa concentrou-se na implementação de um sistema capaz de realizar, de forma integrada, a leitura, a extração de conteúdo, a interpretação semântica e a classificação documental. O fluxo metodológico encontra-se sintetizado na Figura 1 - Fluxo de Classificação Documental com Agentes de IA.

Figura 1 – Fluxo de Classificação Documental com Agentes de IA.



Fonte: elaborada pelos autores (2025).

3.1 Instrumentos e Ferramentas

A arquitetura experimental foi desenvolvida em Python (v3.13), utilizando bibliotecas especializadas como *PyPDFLoader* (para extração textual de documentos PDF), *LangChain*

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

(para integração com modelos de linguagem), *CrewAI* (para orquestração de agentes) e o modelo GPT-4-turbo, acessado via API da OpenAI. A execução ocorreu em ambiente local, utilizando um MacBook equipado com processador Intel e 8 GB de memória RAM, apresentando desempenho satisfatório, com tempos médios de processamento inferiores a um minuto por documento. O Quadro 1 apresenta a síntese das tecnologias empregadas no projeto.

Quadro 1 – Tecnologias Empregadas no Desenvolvimento da Arquitetura Computacional

Tecnologia	Descrição	Função no Projeto
Python v3.13)	Linguagem de programação de alto nível, amplamente utilizada em ciência de dados, IA e automação.	Base da implementação do código, integração entre bibliotecas e controle da execução do agente.
PyPDFLoader	Biblioteca especializada para leitura e extração de texto de arquivos PDF.	Responsável por realizar a extração textual estruturada dos documentos em formato PDF.
LangChain	<i>Framework</i> que conecta LLMs a fontes externas, ferramentas e fluxos de trabalho.	<i>Middleware</i> para interação com o GPT-4, organizando prompts e recebendo respostas.
CrewAI	Plataforma de orquestração de agentes autônomos baseada em colaboração e tarefas coordenadas.	Organiza a lógica da tarefa, conectando o agente com o conteúdo e controlando o processo de execução.
GPT-4-turbo	Modelo de linguagem natural com alta capacidade de compreensão semântica e geração textual.	Executa a análise semântica e a classificação documental com base nas instruções fornecidas.

Fonte: elaborado pelos autores (2025).

3.2 Amostra

O corpus documental utilizado na pesquisa compreendeu 25 documentos digitais em formato PDF, extraídos de fontes públicas e institucionais. A seleção amostral seguiu critérios de intencionalidade, com foco na representatividade temática e na diversidade de gêneros documentais, incluindo artigos científicos, relatórios técnicos e memorandos administrativos, todos relacionados ao campo da Ciência da Informação.

Foram adotados critérios de inclusão como a presença de texto legível em português, a relevância temática para os objetivos da pesquisa e a diversidade quanto ao tipo documental. Por outro lado, foram excluídos documentos que apresentavam falhas de reconhecimento óptico de caracteres (OCR), estruturas despadronizadas de *layout* ou predominância de elementos multimídia, como tabelas complexas, imagens ou gráficos. Essa amostra reduzida, embora limitada em termos quantitativos, permitiu uma análise qualitativa detalhada do desempenho do agente de IA e a observação de seu comportamento frente às tarefas informacionais propostas.

A seleção da amostra partiu de um conjunto inicial de 37 documentos digitais, previamente analisados quanto à legibilidade textual, aderência temática e conformidade estrutural para extração automatizada. Com o intuito de evitar o processamento de arquivos inadequados e otimizar o uso de tokens no modelo, foram excluídos 12 documentos que apresentavam falhas de OCR, formatação inconsistente ou predominância de elementos gráficos. Essa triagem foi realizada antes da submissão ao agente de IA, de forma manual e criteriosa, resultando em uma amostra final de 25 documentos tecnicamente compatíveis com o pipeline proposto.

3.3 Descrição da Implementação Computacional

A solução proposta estrutura-se em uma arquitetura modular de agentes autônomos, concebida para automatizar das tarefas de leitura, interpretação semântica, classificação e sumarização de documentos digitais, em conformidade com as normas ISAD(G) e NOBRADE.

A Figura 1 representa o encadeamento lógico entre os componentes principais da arquitetura. O fluxo inicia-se com a leitura dos arquivos PDF por meio da biblioteca PyPDFLoader, que realiza a segmentação textual de forma estruturada. Em seguida, o conteúdo extraído é encaminhado para a camada de orquestração, gerenciada pela biblioteca CrewAI, que coordena a atuação de três agentes especializados:

- **Agente classificador:** responsável por identificar e atribuir os metadados arquivísticos, com base na análise semântica do conteúdo textual. Seu prompt foi elaborado com exemplos alinhados às normas ISAD(G)/NOBRADE, visando garantir precisão terminológica.
- **Agente revisor:** encarregado de verificar a consistência, completude e conformidade normativa dos metadados gerados, intervindo quando necessário para sugerir correções ou preenchimentos ausentes.
- **Agente resumidor:** dedicado à síntese técnica do conteúdo documental, com foco em clareza, objetividade e aderência à linguagem técnico-científica.

Cada agente é definido como uma instância independente no CrewAI, com atribuições e *goals* específicos, mas integrados em um fluxo coordenado. O CrewAI atua como **orquestrador central**, estruturando a dependência entre os agentes por meio de *task chains*, nas quais a saída de um agente é automaticamente repassada ao próximo. Esse

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

encadeamento segue uma lógica semelhante a fluxos de trabalho humanos, com a vantagem da execução paralela e reproduzível.

A configuração das interações entre os agentes é declarativa e programável: define-se a ordem de execução, os insumos esperados e os critérios de transição entre etapas. Isso permitiu modelar um pipeline semântico orientado a tarefas, em que cada agente atua em sua especialidade sem sobreposição de funções.

A biblioteca CrewAI foi utilizada como orquestradora do fluxo de trabalho, sendo responsável por gerenciar a execução coordenada dos agentes, atribuindo-lhes tarefas específicas e controlando as dependências entre as etapas. Por sua vez, o LangChain atuou como middleware, conectando os prompts e fluxos lógicos dos agentes ao modelo GPT-4-turbo, fornecendo parâmetros como temperatura, top-p e penalidade de frequência, além de gerenciar a persistência do histórico de interações.

O intercâmbio de dados entre os agentes ocorre por meio de objetos JSON estruturados, nos quais cada agente lê e escreve dados em formato padronizado. Por exemplo, o agente classificador recebe o conteúdo textual extraído e produz um objeto contendo os metadados arquivísticos; esse objeto é então consumido pelo agente revisor, que valida e corrige os campos, e por fim complementado pelo agente resumidor com a síntese do conteúdo. Essa abordagem garante modularidade, rastreabilidade e replicabilidade da arquitetura. O *pipeline* foi organizado em etapas paralelas. As principais fases do fluxo de execução são descritas a seguir.

Quadro 2 – Etapas do Processo de Classificação Documental Automatizada

Etapas	Descrição
Extração textual	Realizada por meio da biblioteca <i>PyPDFLoader</i> , permitindo a segmentação e a leitura estruturada do conteúdo de documentos em formato PDF.
Instanciação dos agentes inteligentes	Foram criados três agentes com a biblioteca CrewAI. O classificador documental possui perfil arquivístico, emprega <i>few-shot prompting</i> e <i>guardrails</i> . O revisor verifica se os metadados estão completos e consistentes. O resumidor elabora resumos técnicos objetivos.
Classificação automática	O texto descreve que o agente classificador lê o conteúdo textual e produz metadados padronizados em JSON, seguindo os campos obrigatórios definidos pelas normas ISAD(G) e NOBRADE.
Validação automática de consistência	O agente revisor valida a conformidade dos metadados gerados, realizando correções automáticas quando necessário para assegurar aderência.
Geração de resumo técnico	O agente resumidor sintetiza o conteúdo documental em até 250 palavras, empregando linguagem objetiva, técnica e impessoal.

Armazenamento estruturado dos resultados	Os metadados classificados e os resumos técnicos são armazenados em disco local, juntamente com os tempos de leitura e classificação.
Geração de métricas de avaliação	Ao término do processamento de cada documento, são atribuídas métricas de desempenho (Precisão, Revocação e F1-Score).

Fonte: elaborado pelos autores (2025).

Os hiperparâmetros do modelo foram ajustados empiricamente. A temperatura foi mantida entre 0,2 e 0,3 para reduzir variabilidade nas respostas; o top_p foi fixado em 1.0 para garantir abrangência lexical, e a penalidade de frequência ajustada para 0.3 com o intuito de evitar repetições excessivas.

A opção por CrewAI e LangChain deve-se à sua capacidade de operacionalizar fluxos multiagente com controle semântico granular, sem necessidade de treinamento supervisionado. Embora existam outras ferramentas com funcionalidades semelhantes, como FastAPI, LangGraph ou arquiteturas customizadas, a combinação adotada demonstrou maior aderência à lógica modular da tarefa e à complexidade semântica envolvida, viabilizando uma solução leve, adaptável e cientificamente replicável.

Para promover a transparência metodológica e garantir a reprodutibilidade científica, o código-fonte completo da arquitetura desenvolvida foi disponibilizado publicamente no repositório: <https://github.com/rodrygofesantos/LeitorDocumento>.

3.4 Limitações Metodológicas

Ao longo da condução desta pesquisa, foi necessário lidar com algumas limitações inerentes tanto à abordagem adotada quanto às tecnologias envolvidas. As limitações observadas incluem o risco de viés algorítmico inerente ao modelo GPT-4-turbo e dificuldades ocasionais na diferenciação entre espécies e tipologias documentais, dada a complexidade conceitual dos termos. A mitigação desses riscos ocorreu mediante revisão manual das classificações e ajustes de parâmetros no comportamento dos agentes.

Adicionalmente, a dependência de APIs (*Application Programming Interface*) comerciais e a restrição amostral representam fatores que limitam a generalização dos resultados para contextos multimodais. Apesar disso, a pesquisa evidenciou o potencial da arquitetura proposta como solução aplicável a cenários reais de gestão documental acadêmica.

Para avaliar a precisão das classificações geradas, foi realizada uma revisão manual comparativa, sem alteração direta nos arquivos produzidos pelo agente. Esse processo consistiu na leitura integral de cada documento por um pesquisador com formação arquivística, que confrontou os metadados atribuídos automaticamente com as classificações esperadas segundo as normas ISAD(G) e NOBRADE. As divergências identificadas foram registradas em uma planilha de análise crítica, utilizada exclusivamente para fins avaliativos.

Essa etapa foi fundamental para identificar os limites semânticos do modelo, especialmente nas distinções conceituais entre espécies e tipologias, e subsidiou ajustes nos *prompts* e na parametrização do agente classificador para os ciclos subsequentes de teste.

4 RESULTADOS

Esta seção apresenta os resultados obtidos a partir da implementação do sistema de classificação documental, seguidos de sua análise crítica. Os dados foram organizados em três dimensões principais: desempenho do classificador, consolidação dos documentos classificados e análise dos tempos de processamento.

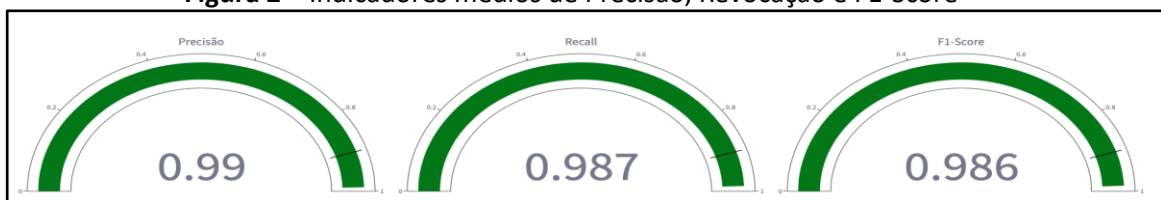
4.1 Desempenho do Sistema de Classificação

O desempenho da arquitetura foi avaliado com base em métricas clássicas de avaliação em sistemas de categorização automática: Precisão (*Precision*), Revocação (*Recall*) e *F1-Score*. Cada métrica foi calculada individualmente para os documentos processados e, posteriormente, agregada para análise global do sistema.

A Precisão refere-se à proporção entre as classificações corretas e o total de classificações atribuídas, indicando a capacidade do sistema em minimizar a ocorrência de falsos positivos. A Revocação, por sua vez, é definida como a razão entre as classificações corretas e o total de documentos relevantes existentes no conjunto de referência, refletindo a eficácia do sistema em recuperar informações pertinentes. O *F1-Score* representa a média harmônica entre precisão e revocação, sintetizando, em uma única medida, o equilíbrio entre exatidão e abrangência da classificação. As métricas foram computadas individualmente para cada documento e, em seguida, agregadas mediante cálculo das médias aritméticas correspondentes, de modo a fornecer uma visão consolidada do desempenho do sistema no corpus analisado. A Figura 2 apresenta os valores médios obtidos para cada um dos indicadores, permitindo a avaliação integrada da acurácia e robustez da solução desenvolvida.

O modelo GPT-4-turbo foi utilizado exclusivamente em modo de inferência, sem fase de treinamento supervisionado. A atuação do agente foi guiada por *few-shot prompting*, com parametrização voltada à estabilidade semântica e aderência terminológica. As categorias utilizadas para classificação foram previamente definidas com base em normativas arquivísticas, e os resultados foram confrontados com uma referência construída a partir da leitura especializada dos documentos. Considerou-se correta toda classificação que coincide integralmente com os rótulos esperados; divergências foram registradas como erros para fins de cálculo das métricas. A precisão média de **99,1%** indica um desempenho altamente consistente, evidenciado também pelos índices de revocação (**98,7%**) e F1-Score (**98,9%**), demonstrando a capacidade do sistema em reproduzir classificações alinhadas aos parâmetros de referência.

Figura 2 – Indicadores médios de Precisão, Revocação e F1-Score



Fonte: resultados da pesquisa (2025)

Os resultados obtidos evidenciam a robustez do processo de classificação documental. Uma precisão média de 99,1% indica um elevado grau de acerto nas categorizações, com uma minimização de falsos positivos. A revocação de 98,7% demonstra que o sistema apresenta alta cobertura, recuperando a quase totalidade dos documentos relevantes. O *F1-Score* de 98,9% corrobora o equilíbrio entre precisão e revocação, aspecto crucial em aplicações onde tanto a exatidão quanto a completude são requisitos fundamentais, como em sistemas de organização documental em instituições acadêmicas.

Tais métricas posicionam o modelo proposto em um patamar de alta confiabilidade, atendendo aos critérios rigorosos da Ciência da Informação e de técnicas modernas de aprendizado de máquina supervisionado para aplicações em ambientes reais.

4.2 Consolidação dos Documentos Classificados

A Tabela 1 apresenta o resumo dos 3 primeiros documentos classificados pelo sistema, contemplando informações sobre o nome do arquivo, gênero documental, espécie documental, tipologia, número de páginas e data de processamento.

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

Tabela 1 – Documentos classificados com respectivos metadados.

Documento	Gênero	Espécie	Tipologia	Páginas
EDITAL-no-1041-DE-15-DE-ABRI...	Textual	Editais	Processo Seletivo	13
Planejamento-estratégico-ECI-2...	Textual	Relatório	Planejamento estratégico	12
Ata+CEPE+8+de+agosto+de+20...	Texto	Ata	Ata de Reunião	12

Fonte: resultados da pesquisa (2025)

A análise da tabela evidencia a diversidade tipológica e temática dos documentos processados, provenientes de fontes institucionais acadêmicas. Observa-se que as categorias de espécie e tipologia foram corretamente atribuídas, respeitando as normas arquivísticas ISAD(G) e NOBRADE, o que corrobora a capacidade do modelo em lidar com heterogeneidade informacional e variações documentais intrínsecas ao acervo estudado.

Na aba "Detalhes do Documento", é possível visualizar atributos como: nome do arquivo original, resumo sintético gerado pela IA, gênero documental, espécie, tipologia, natureza do assunto, forma documental, número de páginas, tempo de leitura, tempo de classificação, e as métricas de desempenho (precisão, revocação e F1-score) associadas.

Figura 3 – Detalhamento do Documento

Resumo	Metadados
A Universidade Federal de Minas Gerais (UFMG) anunciou dois processos seletivos para contratação de Professores Substitutos. O primeiro, Edital N° 1041, busca um professor para o Departamento de Teoria e Gestão da Informação da Escola de Ciência da Informação, na área de Arquivologia. O candidato deve possuir, no mínimo, mestrado na área de Ciência da Informação ou em Gestão e Organização do Conhecimento ou áreas afins. O segundo, Edital N° 1037, busca um professor para o Departamento de Bioquímica e Imunologia do Instituto de Ciências Biológicas, na área de Imunologia, com exigência de doutorado na área de Imunologia ou áreas afins. Ambos os cargos têm inscrições isentas de taxa e contratos com vigência inicial até 31/12/2025, prorrogáveis por até dois anos. A remuneração será correspondente ao nível 1 da Classe A da Carreira do Magistério Superior, variando conforme a titulação do aprovado. As inscrições para o primeiro edital estão abertas por 10 dias e para o segundo, por 15 dias, a partir do dia subsequente à publicação dos editais.	Data: 2025-04-29 Modelo IA: GPT-4
Classificação Id: 12345 Gênero documental: Textual Espécie documental: Edital Tipologia documental: Processo Seletivo Natureza assunto: Educação, Recrutamento Forma: Digital, Assinado Arquivo origem: EDITAL-no-1041-DE-15-DE-ABRIL-DE-2025_ECI_COPERE.pdf Tempo leitura segundos: 120 Tempo classificacao segundos: 7 Precision: 0.9888 Recall: 0.9841 F1 score: 0.9922	

Fonte: resultados da pesquisa (2025)

4.3 Tempos de Leitura e Classificação

Os tempos de execução foram monitorados com base em dois parâmetros principais: (i) o tempo de leitura e extração de conteúdo textual a partir dos arquivos PDF (medido em minutos); e (ii) o tempo de inferência da classificação automática via agentes de IA (medido em segundos). A Tabela 2 sintetiza os valores registrados para o conjunto documental analisado.

Tabela 2 – Tempos médios de leitura automatizada e classificação por documento.

Documento ↑	Leitura (min)	Classificação (segun...)
Ata+CEPE+8+de+agosto+de+2024.pdf	2	9
EDITAL-no-1041-DE-15-DE-ABRIL-DE-2025_ECI_COPERE.pdf	2	7
Planejamento-estratégico-ECI-2021-2025-ATUALIZADO.pdf	2	7

Fonte: resultados da pesquisa (2025)

Os resultados demonstram que o tempo médio de leitura e extração de conteúdo textual foi de aproximadamente dois minutos por documento, variando conforme o número de páginas e a densidade informacional dos arquivos processados. Já a etapa de classificação automática, conduzida por agentes especializados, apresentou um tempo médio inferior a 10 segundos, mesmo considerando a execução paralela dos agentes responsáveis por tarefas distintas (resumo, classificação e validação).

A arquitetura paralela adotada, baseada na inicialização simultânea de múltiplos agentes, contribuiu significativamente para a redução do tempo total de processamento, tornando o sistema altamente responsivo e adequado a cenários institucionais de grande escala. Essa abordagem permite a otimização do *throughput* do sistema sem prejuízo à precisão classificatória, assegurando que múltiplos documentos possam ser analisados de forma concorrente, com desempenho consistente.

Os baixos tempos observados reforçam a viabilidade do uso de arquiteturas baseadas em inteligência artificial para o suporte à gestão documental no contexto acadêmico e institucional, com potencial para acelerar fluxos de trabalho arquivísticos, mitigar gargalos operacionais e ampliar a eficiência administrativa em setores com grande volume de produção informacional.

5 CONSIDERAÇÕES FINAIS

A presente pesquisa técnico-experimental demonstrou a viabilidade e a relevância da adoção de arquiteturas baseadas em agentes de Inteligência Artificial para a leitura, extração e classificação automatizada de documentos no campo da Ciência da Informação. A proposta desenvolvida integra fundamentos epistemológicos da área com avanços tecnológicos recentes, evidenciando que a mediação algorítmica não apenas representa uma possibilidade concreta de automação de tarefas informacionais, mas também inaugura novas formas de interação entre humanos e máquinas nos processos de organização do conhecimento.

Do ponto de vista teórico, a pesquisa contribui para o aprofundamento da discussão sobre a natureza da mediação informacional em contextos digitais, reposicionando os agentes inteligentes como atores computacionais dotados de capacidade de decisão contextualizada. A análise dos resultados permitiu verificar que, quando devidamente parametrizados, esses agentes são capazes de replicar padrões classificatórios com elevado grau de acurácia, ao mesmo tempo em que oferecem suporte à padronização e à curadoria escalável de acervos digitais.

Sob a óptica aplicada, destaca-se a entrega de uma solução funcional, replicável e passível de adaptação a distintos contextos institucionais. A estrutura modular da arquitetura, composta por ferramentas como *CrewAI* e *LangChain*, aliada ao uso de modelos de linguagem avançados, permitiu a construção de fluxos de trabalho automatizados e compatíveis com práticas consolidadas de representação e tratamento da informação.

Embora os resultados obtidos sejam expressivos, reconhecem-se limitações relacionadas à variabilidade estrutural dos documentos e à necessidade de validação humana em casos ambíguos. Tais restrições, no entanto, não comprometem a proposta, mas apontam para caminhos futuros de aprimoramento, como o desenvolvimento de rotinas mais robustas de pré-processamento, a ampliação da base documental e a incorporação de estratégias de aprendizado supervisionado para refinar a atuação dos agentes.

Como desdobramento, propõe-se a realização de estudos de caso em larga escala e a integração da arquitetura com sistemas informacionais existentes em ambientes acadêmicos e governamentais, visando aferir seu impacto em processos reais de gestão documental.

Além disso, a discussão aqui inaugurada sobre mediação algorítmica contribui para a construção de um referencial teórico mais sensível às transformações em curso na Ciência da Informação, sugerindo que os agentes inteligentes não devem ser compreendidos apenas

como ferramentas técnicas, mas como dispositivos epistemológicos que reconfiguram as formas de conhecer, organizar e acessar a informação no século XXI.

REFERÊNCIAS

ALVES, Fernanda Maria Melo Alves; RIBEIRO, Maria do Carmos Oliveira. Inteligência artificial no contexto da Ciência da Informação. **Revista Brasileira de Educação em Ciência da Informação**, São Paulo, p. 1–15, 2024. Disponível em: <https://portal.abecin.org.br/rebecin/article/view/393>. Acesso em: 29 abr. 2025.

ANANNY, Mike; CRAWFORD, Kate. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. **New Media & Society**, v. 20, n. 3, p. 973–989, 2018.

BUCKLAND, Michael K. Information as thing. **Journal of the American Society for Information Science**, v. 42, n. 5, p. 351–360, 1991.

BUOLAMWINI, Joy; GEBRU, Timnit. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. **Proceedings of Machine Learning Research**, v. 81, p. 77–91, 2018.

CAMPOS, Maria Luiza de Almeida. **Representação da informação: fundamentos e perspectivas**. Brasília: Briquet de Lemos, 2000.

CAPURRO, Rafael. Epistemologia e Ciência da Informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 5, n. 1, p. 41–53, jan./jun. 2000.

CRESWELL, John W. **Research design: qualitative, quantitative, and mixed methods approaches**. 4. ed. Thousand Oaks: SAGE Publications, 2014.

FLORIDI, Luciano; BUTTABONI, Carlotta; HINE, Emmie; MORLEY, Jessica; NOVELLI, Claudio; SCHRODER, Tyler. Agentic AI Optimisation (AAIO): What it is, How it Works, Why it Matters, and How to Deal With It. **arXiv**, v. 1, 2025. Disponível em: <https://ssrn.com/abstract=5220068>. Acesso em: 29 abr. 2025.

GILLESPIE, Tarleton. **Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media**. New Haven: Yale University Press, 2018.

GUIMARÃES, Marcus Vinícius de Albuquerque; FREIRE, Gustavo Henrique de Araújo. A inteligência artificial como promotora da competência em informação no ciberespaço: possibilidades para a comunicação científica. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 23., 2023, São Cristovão. **Anais [...]**. São Cristovão: ANCIB, 2023.

HOLLAN, James D.; HUTCHINS, Edwin; KIRSH, David. Distributed cognition: toward a new foundation for human-computer interaction research. **ACM Transactions on Computer-Human Interaction**, v. 7, n. 2, p. 174–196, 2000.

LE COADIC, Yves-François. **A ciência da informação**. Brasília: Briquet de Lemos, 1996.

MATURANA, Humberto R.; VARELA, Francisco J. De máquinas e seres vivos: autopoiese – a organização do vivo. Porto Alegre: Artes Médicas, 1995.

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

REINSEL, David; GANTZ, John; RYDNING, John. **The Digitization of the World: From Edge to Core**. Framingham, MA: International Data Corporation (IDC), 2018. (IDC White Paper, patrocinado pela Seagate).

RUSSELL, Stuart; NORVIG, Peter. **Inteligência Artificial**. 3. ed. São Paulo: Pearson, 2022.

SILVA, Roberta Jerônimo da. **A Inteligência Artificial no contexto da Ciência da Informação: uma análise de domínio**. 2021. 134f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Engenharia e Faculdade de Letras, Universidade do Porto, Porto, 2021.

WOODS, William A. Research in knowledge representation for natural language processing. **Annual Review of Computer Science**, v. 1, p. 225–253, 1986.

ZUBOFF, Shoshana. **The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power**. New York: PublicAffairs, 2019.