Artigo Original

Encontros Bibli

IA GENERATIVA NA EXTRAÇÃO DE METADADOS ARQUIVÍSTICOS: UM ESTUDO BASEADO NA NORMA ISAD(G)

Generative AI in the extraction of archival metadata: a study based on the ISAD(G) standard

Tatiana Canelhas Pignataro
Universidade Estadual Paulista Júlio de Mesquita Filho, Programa de Pós-Graduação em Ciência da Informação, Marília, SP, Brasil
tatiana.canelhas@unesp.br
https://orcid.org/0000-0002-1825-0097

Manoel Pedro de Souza Neto
Universidade Estadual Paulista Júlio de Mesquita Filho, Programa de Pós-Graduação em Ciência da Informação, Marília, SP, Brasil
nettotheone@gmail.com
https://orcid.org/0000-0002-1825-0097

José Carlos Abbud Grácio
Universidade Estadual Paulista Júlio de Mesquita Filho, Programa de Pós-Graduação em Ciência da Informação, Marília, SP, Brasil
jose.gracio@unesp.br
https://orcid.org/0000-0001-7620-1309

Telma Campanha de Carvalho Madio
Universidade Estadual Paulista Júlio de Mesquita Filho, Programa de Pós-Graduação em Ciência da Informação, Marília, SP, Brasil
telma.madio@unesp.br
https://orcid.org/0000-0002-7031-2371

José Eduardo Santarém Segundo
Universidade de São Paulo, Departamento de Educação Informação e Comunicação, Ribeirão Preto, SP, Brasil
santarem@usp.br
https://orcid.org/0000-0003-3360-7872

A lista completa com informações dos autores está no final do artigo

RESUMO

Objetivo: Realizar um diagnóstico buscando analisar o uso de inteligência artificial, especificamente do ChatGPT, na descrição de documentos arquivísticos segundo a Norma Geral Internacional de Descrição Arquivística.
Método: A pesquisa é de natureza exploratória e aplicada, útil em estudos inovadores onde o objetivo é explorar possibilidades, identificar padrões e formular hipóteses futuras para estudos mais detalhados.
Resultados: Os principais resultados mostraram que o ChatGPT atingiu uma taxa média de acerto de 92,04% no preenchimento quantitativo de metadados, com variabilidade mínima entre os testes. No entanto, inconsistências foram observadas em metadados considerados constantes, como Procedência e Condições de Acesso, que não mantiveram a precisão e consistência esperada. Metadados como Dimensão e Suporte, e Âmbito e Conteúdo, apresentaram maior dificuldade de padronização, sugerindo a necessidade de melhorias e ajustes no modelo.
Conclusões: Os resultados sugerem que, embora o ChatGPT tenha demonstrado eficiência na maioria dos campos analisados, os desafios permanecem em metadados com pouca padronização. Os resultados indicam que o ChatGPT é capaz de manter um alto grau de completude dos metadados, mas enfrenta desafios em relação à precisão e consistência, especialmente em campos mais complexos. Ajustes no treinamento do modelo, juntamente com a supervisão humana contínua, podem melhorar a qualidade das descrições geradas. Apesar das limitações, a IA se mostra uma ferramenta promissora, capaz de impulsionar avanços significativos no campo da Arquivologia digital.
PALAVRAS-CHAVE: Descrição Arquivística. Automação. ISAD(G). Inteligência Artificial. *Machine Learning*. ChatGPT.

ABSTRACT

Objective: This study aimed to analyze the use of artificial intelligence, specifically ChatGPT, in the description of digital records according to the General International Standard Archival Description (ISAD(G).

Methodology: The research is exploratory and applied in nature, suited for innovative studies where the objective is to explore possibilities, identify patterns, and formulate hypotheses for more detailed future research.
Results: The main results showed that ChatGPT achieved an average accuracy rate of 92.04% in the quantitative completion of metadata, with minimal variability between tests. However, inconsistencies were observed in metadata considered constant, such as Provenance and Access Conditions, which did not maintain the expected precision and consistency. Metadata like Extent and Medium, and Scope and Content showed greater difficulty in standardization, suggesting the need for improvements and adjustments to the model.
Conclusions: The findings suggest that while ChatGPT demonstrated efficiency in most of the analyzed fields, challenges remain with metadata that lack standardization. The results indicate that ChatGPT is capable of maintaining a high degree of metadata completeness but faces challenges regarding precision and consistency, especially in more complex fields. Adjustments to the model's training, along with ongoing human supervision, could enhance the quality of the descriptions generated. Despite its limitations, AI proves to be a promising tool, capable of driving significant advancements in the field of digital Archival Science.
KEYWORDS: Archival Description. Automation. ISAD(G). Artificial Intelligence. Machine Learning. ChatGPT.

1 INTRODUÇÃO

A descrição em documentos arquivísticos permite que as informações sejam localizadas e compreendidas de maneira eficiente. Conforme o Conselho Internacional de Arquivos (2000, p. 14), trata-se de um processo que envolve a extração, análise, organização e registro de informações com o objetivo de identificar, gerir, localizar e explicar documentos de arquivo, bem como o contexto e o sistema que os produziu.

Neste caminho, com o mundo imerso na Quarta Revolução Industrial (também conhecida por Indústria 4.0), a crescente adoção de tecnologias baseadas em conceitos como a Internet das Coisas (IoT), Inteligência Artificial (IA) e sistemas ciber-físicos para criar fábricas inteligentes e altamente automatizadas (Arruda *et al.*, 2023) torna evidente a sistematização e a coleta de grandes volumes de dados, os quais são utilizados para melhorar a eficiência e a tomada de decisões nas fábricas. A eficiência de um modelo de IA generativa é frequentemente medida pela rapidez com que ele pode produzir resultados úteis, minimizando o uso de recursos computacionais.

A investigação sobre a aplicação da IA na Ciência da Informação (CI), especialmente nas áreas de gestão de arquivos, se expande de maneira contínua. Frontoni (2024) observa a IA no campo arquivístico, e ressalta que há um aumento nas atividades que buscam automatizar processos, desde a avaliação até a geração de metadados. O autor (2024) cita que a IA tem o potencial para redefinir o acesso aos arquivos, e permitir uma interação mais rica entre pesquisadores e arquivos, de modo a promover uma abordagem arquivística mais inclusiva. Rockembach (2024) corrobora neste pensamento

ao apontar que a IA, pelo processamento de linguagem natural e algoritmos de aprendizado de máquina, vem sendo explorada para oferecer capacidades de busca mais cambiantes dentro dos sistemas de arquivos. Lemieux (2024) relata que a IA tem sido considerada para ajudar na proteção da privacidade das informações, especialmente em relação a registros que contêm informações pessoais. Logo, verifica-se que a aplicação de tais tecnologias pode contribuir para a descrição arquivística, aumentando a eficiência e a consistência no acesso às informações.

Ademais, observa-se um avanço na área de IA utilizando a Inteligência Artificial Generativa (IAG), a qual emerge como um assunto de interesse e popularidade em várias áreas do conhecimento. A IAG[1] refere-se a um subcampo da IA focado no desenvolvimento de sistemas capazes de gerar conteúdo novo e original. Desta forma, seu diferencial relaciona-se a sua capacidade de criar a partir do que aprendeu, ou seja, ela pode imitar ou replicar a criatividade humana em diversos domínios, como texto, música e até mesmo algoritmos para programação de *software*. Conforme Jain e Tayal (2023, p. 1):

> *Generative AI, a type of Artificial Intelligence that can create large numbers of data, for example images, text, videos etc. through the practice of selecting patterns from a set of available data and then use those knowledges to develop new and special outputs.*

Exemplificada por modelos como o GPT-4, a IAG cria conteúdos com base em *prompts*, estendendo-se às entradas de texto e imagem, como observado no ChatGPT da OpenAI.[2] De acordo com Jain e Tayal (2023, p. 03) *"Another input required by this function is a user prompt which consist of the actual query that user wants to perform on the input data. This query is in a human like conversational language*." Derico e Kleinman (2023 *apud* Chaka, 2023, p. 06) destacam que *"It is said that GPT-4 can respond to images, and caption and describe them, and process 25,000 words, which is eight times as many as ChatGPT can*". Tais distinções mostram como a IAG, além de enfatizar a criação de conteúdo, possui a habilidade de gerar respostas completas e detalhadas em tempo real, consolidando-se como uma ferramenta aplicável a diversos contextos.

Acredita-se que a Arquivologia pode se beneficiar por meio da automatização de alguns processos com utilização das tecnologias avançadas da IA. Essas tecnologias possibilitam melhorias nos processos de classificação, aprimoram a precisão na busca de documentos e garantem uma descrição arquivística mais eficiente nos acervos de guarda

[1] Do inglês Generativa Artificial Intelligence (GAI).
[2] Disponível em: https://chatgpt.com. Acesso em 5 jun. 2024.

permanente. A convergência da IA com as práticas arquivísticas não só otimiza o gerenciamento de informações, mas também protege a memória institucional e as questões legais, garantindo o acesso e a preservação para as gerações futuras. De acordo com Liu e Lee (2018 *apud* Rockembach, 2024), a incorporação da inteligência artificial nas práticas de arquivamento e gestão de registros traz inúmeros benefícios práticos, mas também apresenta desafios. Um exemplo disso é o uso de aplicativos de aprendizado de máquina para a classificação de documentos e análise de sentimentos, que proporciona uma maior compreensão do conteúdo dos arquivos e contribui para uma recuperação de informações mais eficaz. Para os autores supracitados e para Marciano *et al.* (2018 *apud* Rockembach, 2024, p.90):

> *Interdisciplinary research and projects illustrate the practical benefits and challenges of integrating AI in archival and records management practices. For instance, the application of machine learning for document classification and sentiment analysis offers insights into how AI can support more efficient information retrieval and understanding of archival content (Liu and Lee, 2018). Similarly, the use of NLP and computer vision technologies has shown potential in digitizing and indexing historical documents, making them more accessible to researchers and the public.*

Ao indicar as práticas arquivísticas, insere-se ao debate a descrição a qual depende do uso de metadados para descrever, organizar, recuperar e interpretar documentos. Segundo Pacheco, Silva e Freitas (2023) os metadados, na descrição arquivística, representam recursos de informação e preservam o contexto dos registros, refletindo as atividades que os originam. Santos (2012) aponta que a Norma Geral Internacional de Descrição Arquivística [ISAD(G)] visa identificar e explicar o contexto e o conteúdo dos materiais de arquivo, promovendo a acessibilidade. A natureza orgânica dos documentos arquivísticos, produzidos e recebidos em um processo natural no decorrer das atividades institucionais, reforça a necessidade de usar modelos padronizados de metadados, como a ISAD(G), para “[...] tornar confiáveis, autênticas, significativas e acessíveis descrições que serão mantidas ao longo do tempo” (CIA, 2000, p. 11). Assim, a integração de metadados na descrição arquivística possibilita o gerenciamento e preservação dos registros, garantindo sua relevância e usabilidade em diferentes contextos.

Os arquivos, por abrigarem grandes volumes de documentos, exigem descrições detalhadas, um processo que pode ser demorado e sujeito a erros quando feito manualmente. A pesquisa global conduzida por Stančić e Trbušić (2024) identificou os principais desafios enfrentados por arquivistas ao lidar com documentos arquivísticos

digitais e analisou como a inteligência artificial pode contribuir para solucionar esses problemas.

> *This was confirmed during the in-person follow-up interviews, where the respondents, when asked to explain how they think AI might help them solving their records and archival issues, agreed that AI can help with transcription, acquisition, description, classification, etc.* (Stančić; Trbušić, 2024, p. 74).

A integração de ferramentas de IA na descrição arquivística possibilita a modernização do processo, amplia as possibilidades de explorar o potencial dos arquivos aprimorados pela IA (Frontoni, 2024). Além de otimizar a organização e o acesso, a IA também pode contribuir para a própria criação de registros, especialmente em novos contextos, como os arquivos gerados em mídias sociais ou no ecossistema da Internet das Coisas. Para Frontoni (2024), essa transformação tecnológica amplia o alcance dos arquivos, permitindo destacar registros de comunidades marginalizadas e reforçando a importância dos processos democráticos no contexto arquivístico. A utilização dessas ferramentas, portanto, agiliza os procedimentos arquivísticos, permitindo que os profissionais da área se concentrem em análises mais detalhadas e intervenções mais precisas, otimizando a gestão do acervo.

Diante do avanço das tecnologias de inteligência artificial, modelos como o ChatGPT têm demonstrado potencial na automatização de processos arquivísticos, incluindo a descrição documental baseada na norma ISAD(G). Na descrição de documentos arquivísticos segundo a norma, comparando a qualidade das descrições geradas automaticamente com as descrições realizadas manualmente por arquivistas, busca-se avaliar os benefícios e as limitações desse tipo de aplicação. Contudo, como observado em estudos prévios (Frontoni, 2024; Stančić e Trbušić, 2024), o uso de IA ainda apresenta desafios relacionados à precisão, consistência e interpretação contextual dos metadados, especialmente em campos mais complexos ou pouco padronizados. Assim, o presente estudo busca não apenas analisar o desempenho do ChatGPT na descrição de documentos, mas também identificar as limitações e as possíveis intervenções que podem aprimorar seu uso na prática arquivística.

Por meio de uma pesquisa exploratória avaliou-se a completude dos metadados gerados pelo ChatGPT, por uma análise quantitativa dos metadados gerados pelo modelo, verificando a presença dos elementos descritos com base na norma ISAD(G), e comparando-os com o padrão de preenchimento por arquivistas. A aplicação de uma métrica qualitativa usando a escala Likert avaliou a consistência e precisão das descrições

automáticas, investigando se o ChatGPT, após o treinamento com metadados do repositório da Universidade Estadual Paulista (Unesp), reproduz de forma fiel o padrão descritivo adotado pela universidade.

2 PROCEDIMENTOS METODOLÓGICOS

A pesquisa caracteriza-se como preliminar, exploratória e aplicada, apropriada para estudos inovadores cujo objetivo é desvendar possibilidades, identificar padrões e formular hipóteses para investigações futuras mais detalhadas. A parte exploratória dividiu-se nas seguintes etapas: seleção e captura dos dados, tratamento dos dados, treinamento do modelo de IA e avaliação dos resultados.

O presente artigo é complementado pelos dados utilizados no processo de treinamento e análise, disponíveis no documento suplementar submetido junto ao artigo. Este suplemento contém quadros e tabelas detalhadas com os cálculos, as descrições e os metadados empregados no treinamento do ChatGPT, além de informações adicionais que não puderam ser incluídas neste artigo devido sua extensão.

2.1 SELEÇÃO E CAPTURA DOS DADOS

A seleção de dados para uso da IA consiste na escolha de subconjuntos específicos para análise, treinamento e teste de algoritmos de IA. O processo busca garantir que os dados utilizados sejam de alta qualidade e relevantes para a tarefa que o modelo de IA deve desempenhar. A escolha dos dados está intrinsicamente relacionada ao que se espera de resultados neste tipo de pesquisa com IA Generativa. Conforme Zha *et al*. (2023), a seleção de dados integra a abordagem conhecida como IA centrada em dados, que enfatiza a importância da construção de conjuntos de dados consistentes, incluindo a coleta, a limpeza para remoção de inconsistências e a anotação para fornecer o contexto necessário ao aprendizado de máquina.

Nesta etapa, realizou-se a seleção de documentos no acervo da Unesp disponíveis na plataforma Access to Memory (AtoM)3. O acervo foi escolhido devido à qualidade da descrição arquivística, além de conter um quantitativo de 44 atas de reunião. O escopo

[3] AtoM na versão 2.7.3 – 192. Disponível em: https://arquivodigital.unesp.br/. Acesso, em 20 jun. 2024.

incluiu documentos com descrições que apresentavam, no mínimo, os metadados obrigatórios. Os documentos possuíam como características: gênero textual, não manuscritos, em formato PDF e sem reconhecimento textual. A opção de escolha por esse acervo se dá justamente pela riqueza da descrição usando o padrão de metadados ISAD(G)4, sem a necessidade de preparação dos dados com profissionais, para que pudessem ser utilizados como dados de treinamento.

Os dados exportados da plataforma AtoM são considerados confiáveis, pois são gerenciados e mantidos pela Unesp, garantindo sua integridade e atualização constante. Esses dados públicos externos estão disponíveis para acesso ostensivo e podem ser capturados a qualquer momento por meio da plataforma, por usuários autenticados ou não autenticados, o que facilita sua utilização em estudos como este. Elegeu-se o formato "CSV" para exportação pela plataforma dos metadados descritivos das atas disponíveis, marcando a opção "Incluir objetos digitais" na interface, já que os estes documentos também foram necessários para o treinamento do ChatGPT. Exemplos de 2 descrições das atas de reunião da Unesp conforme extraídas do AtoM encontram-se disponível no documento suplementar (APENSO I – Quadro 1).

2.2 TRATAMENTO DOS DADOS

A qualidade dos dados utilizados no processo de treinamento impacta diretamente os modelos de *machine learning* e os resultados obtidos. Na concepção de Santarem Segundo (2024), os dados podem apresentar características como: ruídos, inconsistência, redundância, dentre outras que podem comprometer a precisão das previsões por prejudicar a performance do modelo. Segundo o mesmo autor, uma forma de sanar os possíveis comprometimentos é a realização do tratamento, ou pré-processamento, dos dados, que visa melhorar sua qualidade, eliminando elementos que possam gerar resultados imprecisos e ajustando os dados para que sejam utilizados de maneira adequada no processamento.

Os dados e documentos selecionados foram preparados para o treinamento dos modelos de IA em um processo que incluiu: a estruturação dos dados em um formato

[4] O uso do padrão ISAD(G) assegura uma estrutura formal e bem modelada em que contribui para a qualidade da descrição arquivística existente.

adequado para o treinamento do ChatGPT, a configuração dos programas ChatGPT[5] e Ai Drive[6], além da organização dos dados e dos arquivos em pastas específicas na infraestrutura utilizada. A preparação e o tratamento dos dados garantiram que o conjunto de dados estivesse organizado e consistente para o treinamento do modelo.

Na etapa de estruturação dos dados, identificou-se os metadados relevantes para a pesquisa como: serem pertencentes ao padrão ISAD(G) e localizáveis nos documentos para posterior extração. Os metadados não preenchidos pela Unesp foram considerados irrelevantes para o estudo. Já os metadados preenchidos de forma inadequada foram tratados e devidamente estruturados para compor o treinamento.

A identificação dos metadados foi a primeira ação realizada e consistiu na comparação entre os metadados extraídos do AtoM e os campos presentes na ISAD(G). Os termos encontrados no CSV não correspondiam aos termos presentes na ISAD(G). Como resultado criou-se um quadro comparativo, disponível no documento suplementar (APENSO I – Quadro 2).

A partir da análise, iniciou-se um tratamento dos dados brutos para serem usados no treinamento do ChatGPT. Primeiramente, desconsiderou-se os metadados que não correspondiam à norma descritiva, sendo, portanto, excluídos da lista, como *accessionNumber*, *digitalObjectChecksum* e outros relacionados no documento suplementar (APENSO I – Quadro 2). Entretanto, optou-se por manter os campos *identifier*[7], *digitalObjectURI*[8] e *eventTypes*[9] que, embora não façam parte do padrão ISAD(G), foram considerados úteis para o treinamento. Por fim, removeu-se os campos de metadados que retornaram vazios ou nulos em todas as atas, como *archivalHistory* (História Arquivística), *appraisal* (Avaliação, Eliminação e Temporalidade), entre outros. No entanto, campo *locationOfOriginals* (Existência e localização dos originais), embora tivesse preenchido, escolheu-se por removê-lo por não ser possível extrair seu dado diretamente do texto das atas de reunião. Como resultado criou-se um quadro demonstrativo, disponível no documento suplementar (APENSO I – Quadro 3).

[5] Plataforma disponível em: https://chatgpt.com/. Acesso em 5 jun. 2024.
[6] Plataforma disponível em: https://myaidrive.com/. Acesso em 5 jun. 2024.
[7] Identificador é um dado extraído do AtoM considerado relevante para descrever o Código de referência.
[8] Campo que traz o nome do PDF das atas de reunião. Esse campo foi tratado para permanecer apenas o nome do arquivo, uma vez que ele retorna do AtoM a URL completa para acessar a ata, na qual, para o treinamento é uma informação dispensável.
[9] Campo necessário para extração das datas de produção, produtor documental e história bibliográfica.

Em seguida, tratou-se os campos de metadados que se relacionam a eventos de datas, produtores e história administrativa, representado pela coluna *eventTypes* (tipos de evento), cujo metadado relevante era a palavra *Creation* (criação). As colunas *eventDates* (datas de produção), *eventActors* (produtor documental) e *eventActorHistories* (história administrativa/biografia) se relacionam diretamente com a coluna *eventTypes*, conforme observado no Quadro 1 que ilustra o resultado da exportação dos metadados ISAD(G) do AtoM na planilha CSV.

Quadro 1 - Exemplo de metadados de evento conforme exportação do AtoM que devem ser tratados para isolar a data de produção, o produtor documental e a história administrativa/biografia caso a palavra *Creation* apareça em *eventTypes*

eventTypes	***eventDates***	***eventActors***	***eventActorHistories***
Contribution|*Contribution*|*Publication*|*Publication*|***Creation***	*NULL*|*NULL*|*NULL*|*NULL*|**2021-03-03**	Pasqual Barretti|Maysa Furlan|Maíra Gebara|Secretaria Geral - SG|***NULL***	Reitor da Unesp no período de [...][10]|Vice-Reitora da Unesp [...]|*NULL*|Órgão submetido à Reitoria [...].|***NULL***

Fonte: elaborado pelos autores (2024)

Reestruturou-se esses campos para conter apenas os dados relacionados à criação do documento. Para realizar esse tratamento, identificou-se inicialmente a posição da palavra *Creation* em *eventTypes*. No CSV do AtoM, as posições são separadas pelo símbolo *pipe* "|", e, sendo assim, *Creation* foi encontrada na 5ª posição, correspondendo à data de produção "2021-03-03", ao produtor "*NULL*" e a *eventActorHistories* como "*NULL*", por também preencherem a 5ª posição. Após o tratamento, os dados passaram a ser representados conforme Quadro 2.

Quadro 2 - Exemplo de pós-tratamento dos metadados de evento

eventTypes	**eventDates**	**eventActors**	**eventActorHistories**
Creation	2021-03-03	*NULL*	*NULL*

Fonte: elaborado pelos autores (2024)

Posteriormente, verificou-se o preenchimento dos campos *eventActors* (nesse caso, como produtor daquele documento) e *eventActorHistories* (história administrativa). Quando preenchidos com o valor *NULL*, conforme requisitos do AtoM, indica que o documento daquela descrição poderia ter herdado o produtor do nível Fundo ou outro nível superior, e que a história administrativa poderia ter sido preenchida nos metadados da

[10] Substituiu-se parte do texto por reticências por ser irrelevante para essa explicação e para não ocupar espaço na tabela.

Norma Internacional de Registro de Autoridade Arquivística para entidades coletivas, pessoas e famílias, a ISAAR(CPF)[11], deste produtor. Encontrou-se esses dados na interface do AtoM (Quadro 3), que foram transpostos para a planilha de treinamento.

Quadro 3 - Dados do produtor documental e da história administrativa das atas de reunião da Unesp

Produtor Encontrado no nível de FUNDO do qual pertencem esses itens documentais.	Unesp - Universidade Estadual Paulista "Júlio de Mesquita Filho".
História Administrativa Encontrado na descrição da autoridade arquivística produtora.	A Unesp, criada em 1976, resultou da incorporação dos Institutos Isolados de Ensino Superior do Estado de São Paulo, então unidades universitárias situadas em diferentes pontos do interior paulista. Abrangendo diversas áreas do conhecimento, tais unidades haviam sido criadas, em sua maior parte, em fins dos anos 50 e inícios dos anos 60.

Fonte: elaborado pelos autores (2024)

Após o tratamento, salvou-se os dados tratados em formato CSV e traduziu-se as nomenclaturas dos campos do CSV exportado do AtoM para seus correspondentes ISAD(G) em português do Brasil. Como resultado, foram obtidos os seguintes metadados preenchidos: *Código de referência*, *Título*, *Nível de descrição*, *Dimensão e suporte*, *Procedência*, Âmbito e conteúdo, *Condições de acesso*, *Idioma*, *Data(s)*, *Nome(s) do(s) produtor(es)*, *História administrativa/Biografia*. Além disso, os campos *identifier* e *digitalObjectURI* que também foram considerados para o treinamento, foram renomeados para *Identificador* e *filename*, respectivamente. O tratamento dos dados a partir da etapa que menciona *eventTypes* até a tradução dos campos estão disponíveis no documento suplementar (APENSO I, Quadro 4).

Finalmente, separou-se o total de 77% das atas para o treino (csv_treino.csv) e 23%, para realização dos testes (csv_teste.csv), conforme quadros presentes no documento suplementar (Apenso II – Quadros 5 e 6), submetido junto a esse artigo.

O ChatGPT apresenta diversos modelos generativos baseados no GPT (*Generative Pre-trained Transformer*), cada um com diferentes níveis de inteligência e capacidades. Cada GPT pode ser configurado para responder de maneiras específicas, adaptando-se a diferentes contextos e necessidades. Optou-se pela utilização do GPT-4o,

[11] Norma que dá diretivas para a preparação de registros de autoridade arquivística que forneçam descrições de entidades (entidades coletivas, pessoas e famílias) relacionadas à produção e manutenção de arquivos (CONARQ, 2004).

por ser o modelo mais recente no período da pesquisa, rápido e com maior capacidade de processamento, considerado adequado para as demandas deste projeto.

Para a configuração da infraestrutura, registrou-se uma conta paga no ChatGPT PRO devido ao seu limite maior de interações com o modelo, uma vez que a versão gratuita não oferecia um número suficiente de *tokens*[12] necessários para o treinamento e a realização das tarefas deste projeto. Além do ChatGPT, utilizou-se a plataforma Ai Drive no plano pago. Constatou-se que o Ai Drive foi necessário devido à sua integração com o ChatGPT, à incorporação de tecnologia OCR, à capacidade de permitir que o ChatGPT analisasse um volume maior de documentos simultaneamente e à preservação da organicidade dos documentos utilizados. Apesar da plataforma ser paga, entende-se que em breve esse tipo integração será facilitado pelo ChatGPT, habilitando outras plataformas sem custo.

Na plataforma Ai Drive criou-se duas pastas denominadas “treino” e “teste”. Inseriu-se as atas listadas no APENSO II – Quadro 5 do documento suplementar na pasta de treino e as atas listadas no APENSO II – Quadro 6 atas na pasta de teste. Na mesma plataforma, criou-se uma pasta chamada “normas” e inseriu-se o documento CBPS_2000_Guidelines_ISADG_Second-edition_PT.pdf[13].

Esta preparação e tratamento dos dados asseguraram que o conjunto de dados estivesse organizado e consistente, pronto para ser utilizado no treinamento e nos testes subsequentes.

2.3 TREINAMENTO DO MODELO

Treinou-se o ChatGPT para identificar e aplicar corretamente os metadados arquivísticos. Esse processo envolveu a criação de *prompts* específicos para garantir que o modelo adotasse as práticas corretas de descrição, conforme o padrão descritivo da Unesp.

[12] *Tokens* referem-se às unidades básicas de texto utilizadas pelo modelo ChatGPT para processar e gerar respostas. Cada *token* pode representar um caractere, uma palavra ou uma parte de uma palavra. Tanto as entradas do usuário quanto as respostas geradas pelo modelo consomem tokens, e o total de tokens por interação é limitado. Na conta paga, o número de interações e a quantidade de *tokens* permitidos são maiores, permitindo um uso mais intensivo do modelo. *Tokens* são essenciais para o funcionamento e eficiência do modelo.

[13] Disponível em: https://www.ica.org/app/uploads/2024/01/CBPS_2000_Guidelines_ISADG_Second-edition_PT.pdf. Acesso em: 15 jun. 2024.

Para acessar os arquivos armazenados no Ai Drive, utilizou-se o GPT “PDF Ai PDF” [14]. Os *prompts* criados estão listados no Quadro 4:

Quadro 4 - *Prompts* utilizados no treinamento do ChatGPT

***Prompt* 1:** Treinamento Etapa 1	*Prompt* 1: Treinamento Etapa 1 Aja como um arquivista especializado em descrição documental. Vamos treinar este modelo para descrever atas de reunião conforme realizado pela UNESP. 1. Leia e compreenda a norma ISAD(G) a partir do texto disponível no Ai Drive em: https://myaidrive.com/AP4ztFAdfmgfYvFQbECWEx/CBPS_2000_Gu.pdf Siga todas as instruções do *Prompt* 1.
***Prompt* 2:** Treinamento Etapa 2	*Prompt* 2: Treinamento Etapa 2 O documento “csv_treino.csv” lista como a UNESP descreve suas “Atas de Reunião” usando as regras da ISAD(G). Leia o documento “csv_treino.csv” disponível no Ai Drive em: https://myaidrive.com/WX4UtDkhGuCBxTYdtVDaFq/csv_treino.csv e realize as seguintes ações: 1. Analise e identifique um padrão de descrição nos metadados presentes no “csv_treino.csv”. As 34 atas de reunião listadas na coluna “*filename*” do documento “csv_treino.csv” estão na pasta de treino no endereço: https://myaidrive.com/XFocgqXkH7PfZTYunBrGPk/treino.folder.pdf 2. Prepare-se para utilizar esses metadados na descrição de documentos arquivísticos. Siga todas as instruções do *Prompt* 2.
***Prompt* 3:** Descrição das Atas de Reunião	*Prompt* 3: Descrição das Atas de Reunião Descreva todas as atas de reunião que estão na pasta de teste no endereço do Ai Drive https://myaidrive.com/M6MwpyTkASY77m34zqjy6u/teste.folder.pdf seguindo as diretrizes abaixo, conforme o treinamento da etapa 2: 1. Formato: Utilize o formato de tabela para descrever as atas de reunião. 2. Idioma: Português brasileiro. 3. Consistência: Mantenha a consistência dos metadados se forem iguais. Não trunque as informações repetitivas. 4. Revise a tabela antes de finalizar para garantir que: 4.1. nenhum metadado extra foi adicionado por engano, 4.2. nenhum metadado que está no “csv_treino.csv” tenha sido omitido 4.3. os campos que são preenchidos da mesma forma em “csv_treino.csv” em todas as atas devem permanecer inalterados nas novas descrições Siga todas as instruções do *Prompt* 3.

Fonte: elaborado pelos autores (2024)

Durante a elaboração dos *prompts*, optou-se por não fornecer orientações detalhadas sobre o padrão de metadados que o ChatGPT deveria seguir. O objetivo foi testar a capacidade do modelo de identificar, por conta própria, padrões de descrição a partir do arquivo “csv_treino.csv”, sem depender de diretrizes explícitas. Essa estratégia buscou avaliar se o modelo conseguiria aprender e aplicar consistentemente os metadados arquivísticos apenas com a análise dos exemplos fornecidos no treinamento.

[14] Disponível em: https://chatgpt.com/g/g-V2KIUZSj0-pdf-ai-pdf. Acesso em: 15 jun. 2024.

Com os *prompts* prontos, deu-se início aos testes. Primeiramente, conectou-se no ChatGPT e selecionou-se o GPT “PDF Ai PDF”. Após a abertura, inseriu-se o primeiro *prompt* e aguardou-se a conclusão da análise. Em seguida, enviou-se o segundo *prompt* e, após nova resposta, enviou-se o terceiro e último *prompt*. Assim que o chat concluiu a resposta em formato de tabela, conforme solicitado no *prompt* 3, transferiu-se o resultado para um arquivo Excel, denominado “TESTE 1”. Essas etapas foram capturadas e disponibilizados no documento suplementar (APENSO III).

Em seguida, iniciou-se uma nova conversa usando o mesmo GPT e repetiu-se o envio dos *prompts*. A tabela resultante do *prompt* final foi inserida na mesma aba do Excel que continha o TESTE 1, logo abaixo deste, sendo denominada “TESTE 2”. Esse procedimento foi repetido por 9 vezes com o ChatGPT, de forma a analisar a qualidade das respostas. Ressalta-se que nenhuma correção foi feita aos resultados gerados pelo modelo.

2.4 AVALIAÇÃO DOS TESTES

Após o treinamento, procedeu-se à avaliação do modelo utilizando um conjunto separado de documentos. Os metadados retornados pelo ChatGPT nos nove testes realizados foram armazenados em uma planilha Excel e, em seguida, comparou-se com os metadados do arquivo csv_teste.csv. A avaliação dividiu-se em duas partes: quantitativa e qualitativa. Utilizou-se métodos estatísticos para medir a precisão, a consistência e a completude das descrições na capacidade do modelo de processar grandes volumes de informação, além de incluir uma análise detalhada das capacidades e limitações do modelo no contexto do estudo de caso.

2.4.1 Análise Quantitativa

Primeiramente, realizou-se uma análise quantitativa dos metadados. Verificou-se que o modelo preencheu todos os campos conforme o padrão de descrição adotado pela Unesp. Considerou-se a presença de todos os metadados utilizados no treinamento como um indicativo de cobertura adequada, ou seja, quantos dos 13 metadados empregados no treinamento retornaram nos testes. Em seguida, avaliou-se o número de atas descritas em cada teste. Na pasta do Ai Drive, onde estavam armazenadas as atas de reunião para

teste, havia um total de 10 atas. Dessa forma, verificou-se quantas atas foram efetivamente descritas em cada teste.

2.4.2 Análise Qualitativa

Após a análise quantitativa, realizou-se uma análise qualitativa conforme a métrica descrita nesta seção. A pontuação proposta baseou-se em princípios gerais de avaliação de qualidade e em metodologias avaliativas comumente utilizadas em sistemas de informação e ciência da computação, adaptadas ao contexto de descrição arquivística. Embora não tenha sido utilizado um texto específico diretamente para a criação da métrica, os conceitos foram inspirados em práticas de avaliação, como escalas Likert[15] e critérios de qualidade em gerenciamento de informações. “A técnica de Likert é frequentemente utilizada para medir a qualidade em diversos contextos, incluindo a avaliação de sistemas de informação” (Likert, 1932).

Para classificar os metadados retornados pelo ChatGPT de acordo com a sua qualidade, adotou-se uma métrica de avaliação em três categorias (Quadro 5). Bruce e Hillmann (2004) destacam a importância de critérios claros para a avaliação da qualidade dos metadados[16], incluindo precisão, completude e consistência.

Quadro 5 - Métrica estabelecida para avaliação qualitativa dos testes de descrição no ChatGPT

Métrica de Avaliação Qualitativa		
Avaliação	**Descrição**	**Critérios**
Excelente (nota: 1)	Os metadados estão completos, precisos e totalmente alinhados com o treinamento dado.	• Os metadados são preenchidos exatamente conforme padrão de descrição da Unesp. • Não há erros ou omissões. • Exemplos: Todos os campos obrigatórios e adicionais estão preenchidos corretamente e de forma consistente.
Bom (nota: 0,5)	Os metadados são geralmente precisos e seguem o padrão de metadados da Unesp, mas apresentam algumas omissões ou erros que não comprometem significativamente a utilidade do registro.	• Os metadados vieram em parte preenchidos conforme padrão da Unesp. • Alguns campos podem ter informações erradas ou ocultadas, mas a maioria está correta. • Exemplos: Pequenas omissões ou erros em campos menos críticos, mas a maioria das informações essenciais está correta.

[15] Uma ferramenta amplamente utilizada para avaliar atitudes e percepções, adaptada aqui para avaliar a qualidade dos metadados.
[16] Referência sobre a importância da qualidade dos metadados e os critérios utilizados para avaliá-los.

Ruim (nota: 0)	Os metadados não foram preenchidos ou são insuficientes, imprecisos e mal alinhados com o treinamento feito. As informações não são confiáveis.	• Os metadados não condizem em nada com a forma que a Unesp descreve suas atas. • Há muitos erros ou omissões significativas. • Exemplos: Campos críticos estão faltando ou preenchidos de forma errada.

Fonte: elaborado pelos autores (2024)

A avaliação detalhada dos metadados gerados pelo ChatGPT considerou dois aspectos principais: precisão e consistência. A precisão refere-se à exatidão das informações preenchidas em relação aos documentos originais, conforme o padrão de descrição da Unesp. Considerou-se um campo preciso se o conteúdo estivesse correto e completo de acordo com a descrição usada em treinamento. A consistência diz respeito à uniformidade das descrições geradas pelo modelo ao longo da pesquisa, implicando que o mesmo tipo de metadado fosse descrito de forma semelhante em diferentes documentos.

A seguir, apresentam-se exemplos de aplicação da métrica no retorno do metadado "Nome(s) do(s) Produtor(es)" durante os testes, conforme demonstrado nas Quadros 6 e 7.

Quadro 6 - Análise do preenchimento do metadado Nome(s) do(s) Produtor(es) feito pela Unesp

	PREENCHIMENTO	ANÁLISE
Unesp	Unesp - Universidade Estadual Paulista "Júlio de Mesquita Filho"	Observou-se o padrão de preenchimento da Unesp contendo: Sigla da faculdade seguida por espaço, hífen, depois outro espaço e nome da faculdade em extenso com aspas duplas de Júlio a Filho.

Fonte: elaborado pelos autores (2024)

Quadro 7 - Exemplos de aplicação da métrica nos metadados descritos pelo ChatGPT

TESTE	PREENCHIMENTO	ANÁLISE	NOTA
Teste 2	Unesp - Universidade Estadual Paulista "Júlio de Mesquita Filho"	Seguiu exatamente o padrão de preenchimento, por isso, a avaliação conforme métrica é "excelente".	1
Teste 3	Universidade Estadual Paulista "Júlio de Mesquita Filho"	Faltou a parte da sigla, portanto a avaliação conforme métrica é "bom".	0,5
Teste 5	UNESP	A sigla veio em letras maiúsculas e faltou a parte do nome em extenso.	0,5
Teste 4	Secretaria Geral	Metadado não condiz com o padrão estabelecido.	0

Fonte: elaborado pelos autores (2024)

Após a atribuição das notas por campo retornado durante os testes, calculou-se o somatório da pontuação atribuída por metadado em cada teste, e as pontuações foram organizadas em uma tabela separada (APENSO IV – Tabela 2) para a realização dos cálculos finais.

Essa abordagem foi utilizada para conduzir uma avaliação estruturada e consistente dos metadados, possibilitando a identificação de áreas que necessitam de ajustes e contribuindo para a verificação da qualidade na descrição arquivística automatizada, seguindo padrões previamente estabelecidos pela instituição.

3 RESULTADOS E ANÁLISE DOS TESTES

Solicitou-se ao modelo ChatGPT que descrevesse 10 (dez) atas de reunião em cada um dos 9 (nove) testes realizados, empregando a mesma sequência de *prompts*. Foram analisados 13 metadados utilizados pela Unesp, e nenhuma correção foi demandada para o ChatGPT após a geração dos resultados.

3.1 RESULTADOS E ANÁLISE QUANTITATIVA

A tabela completa com os cálculos, bem como as pontuações atribuídas aos metadados preenchidos pelo ChatGPT em cada teste, está disponível no documento suplementar (APENSO IV - Tabela 1). No entanto, incluiu-se parte desses cálculos na Tabela 1 a seguir, refletindo o quantitativo de atas descritas em cada teste.

Tabela 1 – Demonstração de cálculo da quantidade de atas descritas por teste

	Nº de atas descritas por teste	**Nº de atas totais**	**% de atas descritas por teste**
TESTE 1	10	10	100%
TESTE 2	9	10	90%
TESTE 3	10	10	100%
TESTE 4	10	10	100%
TESTE 5	10	10	100%
TESTE 6	10	10	100%
TESTE 7	7	10	70%
TESTE 8	10	10	100%
TESTE 9	9	10	90%
Total de atas descritas	**85**	**90**	**94,44%**

Fonte: elaborada pelos autores (2024)

Dos nove testes realizados pelo ChatGPT, os dados da coluna *Nº de atas descritas por teste* da Tabela 1 indicam que, das 10 atas de reunião usada para teste: 10 atas foram descritas em 6 dos testes realizados; 9 atas foram descritas em 2 dos testes realizados; e 7 atas foram descritas em 1 dos testes realizados.

Adicionalmente à análise das atas descritas, realizaram-se cálculos sobre o quantitativo de metadados descritos por teste.

Tabela 2 – Cálculos executados com base no quantitativo de metadados descritos por teste

	Total de metadados preenchidos por teste (real)[17]	**Total de metadados preenchidos por teste (ideal)[15]**	**% de metadados preenchidos por teste**
TESTE 1	120	130	92.31%
TESTE 2	115	117	98.29%
TESTE 3	120	130	92.31%
TESTE 4	120	130	92.31%
TESTE 5	110	130	84.62%
TESTE 6	120	130	92.31%
TESTE 7	84	91	92.31%
TESTE 8	120	130	92.31%
TESTE 9	108	117	92.31%
Total de preenchimento por metadados (real)	**1017**	**1105**	**92.04%**

Fonte: elaborada pelos autores (2024)

O cálculo da coluna *Total de metadados preenchidos por teste (real)* na Tabela 2 corresponde ao somatório dos metadados preenchidos em cada teste. O *Total de metadados preenchidos por teste (ideal)* é a quantidade máxima de metadados (13) multiplicado pelo número de atas que foram descritas por teste, representados na Tabela 2. A *Porcentagem de metadados preenchidos por teste* é obtida dividindo-se o Total de preenchimento real pelo Total de preenchimento ideal, multiplicados por 100.

$$Porcentagem\ de\ preenchimentos\ por\ teste\ (\%) = \frac{Total\ de\ preenchimentos\ real}{Total\ de\ preenchimentos\ ideal} \times 100$$

Verificou-se que a porcentagem de metadados preenchidos variou entre 84,62% e 98,29%, com uma média de 92,04%. Para entender a variabilidade no preenchimento dos metadados entre os testes, calculou-se o desvio padrão utilizando a fórmula:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

[17] Nas tabelas apresentadas (Tabela 2 e Tabela 4), utilizam-se os termos “real” e “ideal” para diferenciar entre o número de dados preenchidos pelo ChatGPT e o número teórico máximo de ocorrências. O termo “real” refere-se ao número de ocorrências dos metadados gerados pelo ChatGPT, refletindo o desempenho efetivo do modelo. Já o termo “ideal” representa o retorno máximo possível que o ChatGPT poderia alcançar caso todos os metadados fossem preenchidos, ou seja, um cenário em que o modelo retornasse 100% dos metadados em todos os testes. Essa distinção entre “real” e “ideal” permite uma comparação direta entre o desempenho atual e o desempenho esperado do modelo.

Onde x_i é o valor individual da porcentagem de preenchimento de metadados por teste, μ é a média das porcentagens de preenchimento (μ = 92,04%), e N é o número total de testes feitos (ou seja, 9).

$$\sigma = \sqrt{\frac{1}{9}((92{,}31 - 92{,}04)^2 + (98{,}29 - 92{,}04)^2 + \cdots + (92{,}31 - 92{,}04)^2 + (92{,}31 - 92{,}04)^2)} \approx 0.032$$

O desvio padrão calculado indica uma pequena variação entre os testes, refletindo a consistência do modelo no preenchimento dos metadados.

Embora a inteligência artificial não seja determinística, os resultados indicam uma baixa variação no índice de preenchimento entre os testes, com valores que variam de 84,62% a 98,29%. Mesmo sem atingir 100% de cumprimento da tarefa, o modelo manteve uma alta taxa de completude dos metadados. Essa leve oscilação sugere a presença de praticamente todos metadados exigidos, ainda que em diferentes tentativas, o que demonstra sua qualidade em tarefas de descrição arquivística.

Além da análise do número de metadados descritos por teste, a análise da frequência de preenchimento dos metadados ao longo dos nove testes realizados foi feita com o objetivo de verificar quantas vezes cada metadado foi preenchido, independentemente de sua correta descrição arquivística. A Tabela 3 apresenta o número total de vezes que cada metadado foi preenchido, além da porcentagem correspondente, com base nas tentativas realizadas.

Tabela 3 – Cálculos executados em cima do quantitativo de metadados descritos no total

	Total de preenchimentos por metadado (real)	**% de preenchimentos por metadado**
filename	9	10,59%
Código de Referência	85	100%
Identificador	85	100%
Título	85	100%
Nível de Descrição	85	100%
Dimensão e Suporte	85	100%
Procedência	85	100%
Âmbito e Conteúdo	85	100%
Condições de Acesso	75	88,24%
Idioma	85	100%
Data(s)	85	100%
Nome(s) do(s) Produtor(es)	84	98,82%
História Adm./Biografia	84	98,82%

Fonte: elaborada pelos autores (2024)

Das nove tentativas de descrição, a análise do retorno dos 13 metadados apresentou os seguintes resultados:

- O metadado *filename* foi preenchido pelo ChatGPT em apenas um dos nove testes (TESTE 2), correspondendo a 10,59% de preenchimento total.
- O metadado *Condições de Acesso* não foi preenchido no TESTE 5 em sua totalidade, resultando em 88,24% de preenchimento.
- Os metadados *Nome(s) do(s) Produtor(es)* e *História Administrativa/Biografia* não foram preenchidos na ATA CO ATA 267 do TESTE 2, o que resultou em um preenchimento total de 98,82% para ambos.
- Todos os outros metadados, incluindo *Código de Referência*, *Identificador*, *Título*, *Nível de Descrição*, *Dimensão e Suporte*, *Procedência*, *Âmbito e Conteúdo*, *Idioma* e *Data(s)*, foram preenchidos em todos os testes, totalizando 100% de preenchimento.

Essa análise quantitativa demonstra que, na maioria dos casos, os metadados foram preenchidos em todas as tentativas. No entanto, houve lacunas em alguns metadados específicos, como *filename* e *Condições de Acesso*, que não foram preenchidos consistentemente em todos os testes.

3.2 RESULTADOS E ANÁLISE QUALITATIVA

A tabela completa com os cálculos, bem como as pontuações atribuídas aos metadados preenchidos pelo ChatGPT em cada teste, está disponível no documento suplementar (APENSO IV – Tabela 2).

A análise qualitativa foi conduzida atribuindo-se notas aos metadados retornados pelo ChatGPT em cada teste, utilizando a métrica estabelecida. Cada metadado foi avaliado com base em três categorias de precisão e completude: Excelente (nota 1), Bom (nota 0,5) e Ruim (nota 0).[18] As notas foram somadas por teste e por metadado, e a porcentagem de acertos foi calculada com base nessas somas.

Tabela 4 – Cálculos executados sobre a qualidade dos metadados descritos por teste

	Total de pontos (real)[19]	**Total de pontos (ideal)**	**Porcentagem qualitativa por teste**

[18] Mais detalhes na Seção 2.4.2 – Análise Qualitativa.

[19] Nas tabelas apresentadas, utilizam-se os termos “real” e “ideal” para diferenciar entre a pontuação atribuída aos dados retornados pelo ChatGPT e a pontuação teórica máxima. O termo “real” refere-se à pontuação obtida com base nos metadados gerados pelo ChatGPT em cada teste, refletindo o desempenho efetivo do modelo. Já o termo “ideal” representa a pontuação máxima possível que o ChatGPT poderia alcançar caso todas as descrições fossem corretas e completas, ou seja, um cenário em que o modelo acertasse 100% dos metadados em todos os testes. Essa distinção entre “real” e “ideal” permite uma comparação direta entre o desempenho atual e o desempenho esperado do modelo.

			realizado
TESTE 1	79	120	65,83%
TESTE 2	85	115	73,91%
TESTE 3	73	120	60,83%
TESTE 4	60	120	50,00%
TESTE 5	65.5	110	59,55%
TESTE 6	102	120	85,00%
TESTE 7	54	84	64,29%
TESTE 8	88	120	73,33%
TESTE 9	71.5	108	66,20%
Total por metadados	**678**	**1017**	**66,67%**

Fonte: elaborada pelos autores (2024)

A Tabela 4 reflete a soma dos pontos recebidos por todos os metadados em cada teste, permitindo a visualização da porcentagem qualitativa de preenchimentos por teste. A coluna *Total de pontos (real)* representa o somatório das notas atribuídas a cada metadado por teste. A coluna *Total de pontos (ideal)* corresponde ao número máximo de pontos que poderiam ser obtidos em cada teste. A coluna *Porcentagem qualitativa por teste realizado* apresenta a porcentagem qualitativa de preenchimento, calculada pela fórmula:

$$Porcentagem\ qualitativa\ por\ teste\ realizado\ (\%) = \frac{Total\ de\ pontos\ (real)}{Total\ de\ pontos\ (ideal)} \times 100$$

Esses cálculos indicam que o TESTE 6 apresentou a maior porcentagem qualitativa (85,00%), ou seja, a maior consistência e precisão no preenchimento dos metadados, enquanto o TESTE 4 obteve a menor (50,00%). A média geral de acertos entre todos os testes foi de 66,67%. Para entender a variabilidade na qualidade dos testes aplicados, calculou-se o desvio padrão utilizando a fórmula:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Onde x_i é o valor individual da porcentagem de preenchimento de metadados por teste, μ é a média das porcentagens de preenchimento (μ = 66,67%), e N é o número total de testes feitos (9).

$$\sigma = \sqrt{\frac{1}{9}((65{,}83 - 66{,}67)^2 + (73{,}91 - 66{,}67)^2 + \cdots + (73{,}33 - 66{,}67)^2 + (66{,}20 - 66{,}67)^2)} \approx 0.094$$

O desvio padrão de 0,094 indica uma maior variabilidade nos resultados qualitativos em comparação com os dados quantitativos, cujo desvio padrão foi de 0,032. Esse aumento de variabilidade sugere que a avaliação da qualidade do preenchimento dos metadados pelo ChatGPT é mais sensível às nuances das descrições, especialmente em metadados que exigem maior contextualização.

Além da análise por teste, foram realizadas avaliações individuais por metadado, conforme Tabela 5.

Tabela 5 - Pontuação qualitativa por metadado

Metadado	Total de Pontos (real)	Total de Pontos Possíveis (ideal)	% de Pontuação por Metadado
filename	9	9	100,00%
Código de Referência	84	85	98,82%
Identificador	59,5	85	70,00%
Título	39	85	45,88%
Nível de Descrição	85	85	100,00%
Dimensão e Suporte	38	85	44,71%
Procedência	65	85	76,47%
Âmbito e Conteúdo	39,5	85	46,47%
Condições de Acesso	29	75	38,67%
Idioma	85	85	100,00%
Data(s)	55,5	85	65,29%
Nome(s) do(s) Produtor(es)	51	84	60,71%
História Administrativa/Biografia	38,5	84	45,83%

Fonte: elaborada pelos autores (2024)

Os metadados ditos constantes são aqueles que, por sua natureza, são preenchidos de forma idêntica em todas as descrições, independentemente do documento analisado. Entre os 13 metadados presentes no estudo, classificou-se 6 como constantes. Esses valores estão detalhados no Quadro 8.

Quadro 8 – Preenchimento correto dos metadados constantes

Campo	Preenchimento
Nível de Descrição	Item
Procedência	Secretaria Geral - SG
Condições de acesso	Este documento é de propriedade da Unesp. Ao utilizá-lo colocar os devidos créditos ©Unesp, segundo Lei de Direito Autoral.
Idioma	pt_BR
Nome(s) do(s) produtor(es)	Unesp - Universidade Estadual Paulista “Júlio de Mesquita Filho”
História administrativa/Biografia	A Unesp, criada em 1976, resultou da incorporação dos Institutos Isolados de Ensino Superior do Estado de São Paulo, então unidades universitárias situadas em diferentes pontos do interior paulista. Abrangendo diversas áreas do conhecimento, tais unidades haviam sido criadas, em sua maior parte, em fins dos anos 50 e inícios dos anos 60.

Fonte: elaborado pelos autores (2024)

Os metadados *Condições de Acesso, Nome(s) do(s) Produtor(es)* e *História Administrativa/Biografia* apresentaram baixos índices de acerto, variando entre 38,67% e 60,71%. O metadado *Procedência* obteve um desempenho razoável, com uma precisão de 76,47%. Já os metadados *Nível de Descrição* e *Idioma* atingiram 100% de aproveitamento em todas as tentativas. Esses resultados demonstram uma inconsistência no retorno esperado para esses campos, especialmente em relação aos metadados que deveriam ser constantes. Acredita-se que esses valores poderiam ser preenchidos com maior consistência se ao menos uma correção manual fosse realizada durante a execução dos testes, já que pequenos ajustes costumam resolver inconsistências, especialmente em campos que deveriam ser constantes e não apresentar variação.

Os cálculos indicam outros problemas no preenchimento dos metadados, especialmente em *Título*, *Âmbito e Conteúdo* e *Dimensão e Suporte*, que apresentaram as pontuações mais baixas. Embora o ChatGPT tenha demonstrado uma boa qualidade no preenchimento de diversos metadados, ainda há margem para melhorias na precisão e consistência, particularmente em metadados constantes.

- O campo *filename*, apesar de não ter sido descrito em 8 dos 9 testes realizados, atingiu 100% de precisão na única vez em que foi preenchido.
- O campo *Código de Referência* alcançou a maior pontuação entre os metadados variáveis, com 98,82% de precisão, seguido pelo *Identificador*, que obteve 70%.
- Quanto ao *Título*, que deveria seguir um padrão simples conforme observado no TESTE 6: “Ata da [número]ª sessão [tipo de sessão] do Conselho Universitário da Unesp de [data]”, apresentou um índice de acerto de 45,88%.
- O campo *Dimensão e Suporte* foi alimentado de duas formas pela Unesp: ora com a descrição “Ata em formato PDF, com XX páginas.”, ora com “Representante digital em formato PDF, com XX páginas, de documento textual em suporte papel.” Essa variação nos metadados de treinamento pode ter confundido o modelo durante o processo de aprendizado, uma vez que a falta de um padrão único pode gerar incertezas em relação à forma correta de preenchimento. Esse cenário foi previamente observado e, ainda assim, a variação foi mantida com o propósito de avaliar como o modelo descreveria o campo diante dessa ambiguidade. Nos testes, o ChatGPT seguiu diferentes padrões de preenchimento, conforme listados na Quadro 9.

Quadro 9 – Exemplos de preenchimentos do campo Dimensão e Suporte pelo ChatGPT

Teste	Preenchimento
TESTE 1	Ata em formato PDF, com XX páginas
TESTE 2	Ata em formato PDF, com número de páginas não especificado
TESTE 3	Representante digital em formato PDF, com X páginas
TESTE 4	Representante digital em formato PDF, com XX páginas
TESTE 5	Ata em formato PDF, com XX páginas
TESTE 6	Ata em formato PDF, com XX páginas
TESTE 7	PDF com XX páginas
TESTE 8	Ata em formato PDF, com XX páginas
TESTE 9	PDF, XX páginas

Fonte: elaborado pelos autores (2024)

Nota-se que na maioria das descrições, o formato do arquivo e a contagem de páginas foram mencionados. No entanto, o número de páginas variou incorretamente em todos os 85 testes realizados, indicando erro em todas as contagens.

- O campo *Âmbito e Conteúdo* foi o mais desafiador para seguir o padrão correspondente. Sendo extenso e com poucas regras identificáveis, o ChatGPT recebeu nota 0,5 na maioria das descrições, conseguindo extrair corretamente algumas informações, como datas, presença de participantes e, ocasionalmente, tópicos discutidos nas reuniões.
- O campo *Data(s)*, que deveria ser preenchido nos formatos “aaaa-mm-dd” ou “aaaa/mm/dd”, foi relativamente simples de seguir. No entanto, apesar de acertar a maioria das datas (68 acertos em 85 tentativas), ocorreram 17 erros, nos quais o formato variou para “dd/mm/aaaa” ou apenas “aaaa”.

Esses resultados mostram que, apesar do bom desempenho geral, o ChatGPT ainda enfrenta desafios em campos mais complexos e variáveis, como *Dimensão e Suporte*, além de apresentar pequenas inconsistências na formatação de datas e títulos.

4 DISCUSSÃO DOS RESULTADOS

Os resultados obtidos na pesquisa indicam que o ChatGPT apresentou um desempenho satisfatório na descrição de documentos arquivísticos com base nos metadados da norma ISAD(G), sendo analisado principalmente pelos critérios de precisão, consistência e completude, que são usados para medir a qualidade das descrições geradas.

A análise quantitativa mostrou uma alta taxa de preenchimento dos metadados, com uma média de 92,04% de completude ao longo dos testes. Esse dado reflete uma

capacidade do modelo em realizar tarefas de extração de metadados de forma adequada, mesmo em cenários onde há múltiplos campos e padrões a serem seguidos.

No entanto, a análise qualitativa identificou inconsistências em alguns metadados, especialmente naqueles que exigem maior precisão contextual ou uma estrutura padronizada, como os campos *Dimensão e Suporte* e *Âmbito e Conteúdo*. A falta de um padrão claro de preenchimento, previamente observada, foi intencionalmente permitida durante o treinamento para testar a flexibilidade do modelo em lidar com descrições diferentes, o que parece ter causado confusão no ChatGPT, resultando em variabilidade nas respostas geradas. O desvio padrão de 0,094 encontrado na análise qualitativa evidencia uma maior variação nos resultados em comparação à análise quantitativa, que apresentou um desvio padrão de 0,032. Essa diferença sugere que os testes qualitativos são mais suscetíveis a oscilações, especialmente em metadados que exigem um maior grau de contextualização. Os metadados constantes, como *Nome(s) do(s) Produtor(es)* e *História Administrativa/Biografia*, que deveriam ser descritos de forma idêntica em todos os testes, também apresentaram variações nas descrições obtidas. Este aspecto pode representar um ponto de atenção para futuras pesquisas, pois indica que ajustes mais refinados nos métodos de avaliação qualitativa, bem como no treinamento do modelo, podem ser necessários para alcançar uma maior consistência.

A título de experimento exploratório, realizou-se um teste adicional com as mesmas 10 atas e uma única correção nos *prompts*, com o objetivo de avaliar uma possível melhoria nos resultados. Neste teste, o percentual de acertos nos metadados passou de 55,42% para 86,92%, indicando que ajustes pontuais nos comandos podem aumentar a qualidade das descrições geradas pelo ChatGPT. Esse resultado sugere que intervenções corretivas durante o processo podem aprimorar o desempenho do modelo, reforçando a importância de futuras investigações focadas na aplicação de correções no uso de IAG para a descrição arquivística.

Por outro lado, os metadados *Nível de Descrição* e *Idioma* apresentaram uma precisão de 100% em todos os testes, mostrando que o modelo é capaz de capturar e manter informações constantes quando corretamente configurado para isso.

Esses resultados sugerem que, embora o ChatGPT tenha demonstrado uma capacidade notável de descrever documentos arquivísticos de maneira automatizada, a consistência no preenchimento de metadados, especialmente aqueles com maior complexidade, ainda depende de um ajuste mais refinado no modelo. Isso indica a necessidade de treinamentos mais direcionados ou de intervenções manuais em casos em

que há variação no preenchimento dos metadados, para assegurar uma padronização total. Adicionalmente, explorar relações entre variáveis, como a complexidade do metadado e a precisão no preenchimento, pode ajudar a identificar padrões de desempenho do modelo em metadados mais exigentes. Métodos como análises de correlação ou modelos de regressão poderiam ser aplicados para investigar tais relações em estudos futuros. Além disso, o desenvolvimento de *prompts* mais pertinentes, uma área de estudo em expansão, pode contribuir para melhorar os resultados de forma significativa. No entanto, a melhoria dos *prompts* não foi explorada neste estudo.

Os resultados obtidos nesta pesquisa convergem com as discussões de Frontoni (2024), que destaca o potencial da inteligência artificial na modernização de processos arquivísticos, especialmente no que diz respeito à automação da descrição e recuperação de informações. De maneira semelhante, os desafios observados na inconsistência dos metadados, como Procedência e Condições de Acesso, refletem os apontamentos de Stančić e Trbušić (2024), que identificam a necessidade de ajustes e intervenções humanas para solucionar problemas de qualidade na aplicação de IA em ambientes arquivísticos.

Por outro lado, a análise deste estudo traz contribuições inovadoras ao demonstrar, de forma quantitativa e qualitativa, o desempenho do ChatGPT no preenchimento de metadados conforme a norma ISAD(G), alcançando uma taxa média de completude de 92,04%. Essa métrica reforça as observações de Rockembach (2024) sobre a eficiência do uso de modelos de IA para otimizar a descrição documental. Entretanto, este estudo vai além ao evidenciar que a variabilidade qualitativa, expressada pelo desvio padrão de 0,094, é mais acentuada em metadados que demandam maior contextualização, um aspecto pouco explorado em investigações anteriores.

Dessa forma, a presente pesquisa contribui não apenas para o avanço metodológico na aplicação de IA no campo arquivístico, mas também para a compreensão das limitações e possibilidades de uso do ChatGPT na descrição automatizada, sinalizando caminhos futuros para a melhoria da consistência e precisão das descrições.

5 CONCLUSÃO

Esses resultados indicam que, embora o ChatGPT tenha demonstrado a capacidade de preencher corretamente a maioria dos campos de metadados, ainda há desafios relacionados à consistência das descrições, especialmente em campos mais

complexos. Isso evidencia a necessidade de ajustes mais refinados no modelo, além de treinamentos específicos para lidar com essas variações. Quando a consistência não é alcançada, pode ser necessária a aplicação de intervenções manuais para assegurar a padronização. Além disso, o desenvolvimento de *prompts* mais apropriados, uma área em constante desenvolvimento, pode contribuir para melhorar os resultados. No entanto, este estudo não explorou a otimização de *prompts*.

Em síntese, o uso de inteligência artificial na descrição arquivística demonstra grande potencial, embora algumas limitações relacionadas à consistência e precisão dos metadados ainda precisam ser superadas. Os resultados deste estudo indicam que ajustes apropriados no treinamento do modelo, aliado a uma supervisão humana contínua, podem assegurar um alto nível de qualidade no preenchimento dos metadados descritivos, impulsionando avanços significativos no campo da Arquivologia digital.

REFERÊNCIAS

ARRUDA, H. M.; BAVARESCO, R. S.; KUNST, R.; BUGS, E. F.; PESENTI, G. C.; BARBOSA, J. L. V. Data Science Methods and Tools for Industry 4.0: A Systematic Literature Review and Taxonomy. **Sensors**, 2023, v. 23, n. 11, p. 5010. DOI: https://doi.org/10.3390/s23115010. Disponível em: https://www.mdpi.com/1424-8220/23/11/5010. Acesso em: 05 jun. 2025.

BRUCE, T. R.; HILLMANN, D. I. The continuum of metadata quality: Defining, expressing, exploiting. *In*: **Metadata in Practice**. ALA Editions, 2004. p. 238-256.

CHAKA, C. **Generative AI Chatbots - ChatGPT versus YouChat versus Chatsonic**: Use Cases of Selected Areas of Application. 2023.

CONSELHO INTERNACIONAL DE ARQUIVOS (CIA). ISAD(G): Norma geral internacional de descrição arquivística. 2. ed. Rio de Janeiro. 2000. Disponível em: https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/isad_g_2001.pdf. Acesso em: 14 jun. 2024.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **ISAAR (CPF)**: Norma Internacional sobre Registros de Autoridade Arquivística para Entidades Coletivas, Pessoas e Famílias. tradução de Vitor Manoel Marques da Fonseca. 2. ed., Rio de Janeiro: Arquivo Nacional, 2004. Disponível em: https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/isaar_cpf.pdf. Acesso em: 14 jun. 2024.

FRONTONI, E. Appearance-Based Archival Science. *In*: DURANTI, L.; ROGERS, C. (Ed.). **Artificial Intelligence and Documentary Heritage**. SCEaR Newsletter 2024 -

Special Issue 2024. Paris: UNESCO, 2024. p. 49-53. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000389844. Acesso em: 05 jun. 2025.

JAIN, N.; TAYAL, A. PANDAS AI: A Step Towards GEN AI. **International Journal of Scientific Research in Engineering and Management (IJSREM)**, v. 7, n. 7, p. 1-9, 2023.

LEMIEUX, V. Balancing Act: Navigating the Nexus of AI, Privacy, and Accessibility in Archives. *In*: DURANTI, L.; ROGERS, C. (Ed.). **Artificial Intelligence and Documentary Heritage**. SCEaR Newsletter 2024 - Special Issue 2024. Paris: UNESCO, 2024. p. 39-42. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000389844. Acesso em: 05 jun. 2025.

LIKERT, R. A technique for the measurement of attitudes. **Archives of Psychology**, v. 22, n. 140, p. 1-55, 1932.

PACHECO, A.; SILVA, C. G. da; FREITAS, M. C. V de. A metadata model for authenticity in digital archival descriptions. **Archival Science**, v. 23, p. 629–673, 2023. DOI: https://doi.org/10.1007/s10502-023-09422-w. Disponível em: https://link.springer.com/article/10.1007/s10502-023-09422-w. Acesso em: 05 jun. 2025.

ROCKEMBACH, M. AI Literacy: A Must for Records Management and Archival Professionals. *In*: DURANTI, L.; ROGERS, C. (Ed.). **Artificial Intelligence and Documentary Heritage**. SCEaR Newsletter 2024 - Special Issue 2024. Paris: UNESCO, 2024. p. 90-95. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000389844. Acesso em: 05 jun. 2025.

SANTOS, V. B. Preservação de documentos arquivísticos digitais. **Ciência da Informação**, [*s. l.*], v. 41, n. 1, 2012. DOI: https://doi.org/10.18225/ci.inf.v41i1.1357. Disponível em: https://revista.ibict.br/ciinf/article/view/1357. Acesso em: 05 jun. 2025.

SANTAREM SEGUNDO, J. E. Disciplina "**Data Science e Inteligência Artificial**: um olhar pela Ciência da Informação". [Slides da aula 02]. PPGCI/UNESP, 1º sem. 2024.

STANČIĆ, H.; TRBUŠIĆ, Z. Annotation of Digitised Archival Materials Supported by AI. *In*: DURANTI, L.; ROGERS, C. (Ed.). **Artificial Intelligence and Documentary Heritage**. SCEaR Newsletter 2024 - Special Issue 2024. Paris: UNESCO, 2024.p. 73-78. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000389844. Acesso em: 05 jun. 2025.

ZHA, D.; BHAT, Z. P.; LAI, K.; YANG, F.; JIANG, Z.; ZHONG, S.; HU, X. **Data-centric Artificial Intelligence**: a survey. arXiv:2303.10158v3 [cs.LG]. 2023.

NOTAS

AGRADECIMENTOS
Agradece-se ao(s) orientador(es), José Carlos e Telma Madio, do PPGCI Unesp, e em especial o professor José Eduardo Santarém do PPGCI Unesp, cujo apoio e orientação foram essenciais para a realização deste estudo.
Agradece-se ao colega Alex Pereira de Holanda, pelas valiosas contribuições técnicas e discussões que enriqueceram o desenvolvimento deste trabalho. Um agradecimento especial aos amigos e familiares, em especial Leonardo Pignataro, cujo apoio e incentivo foram fundamentais para a conclusão deste projeto.

CONTRIBUIÇÃO DE AUTORIA

Concepção e elaboração do manuscrito: T. Canelhas, M. P. S. Neto, J. E. Santarém Segundo
Coleta de dados: T. Canelhas, M. P. S. Neto
Análise de dados: T. Canelhas, M. P. S. Neto
Discussão dos resultados: T. Canelhas, M. P. S. Neto, J. E. Santarém Segundo
Revisão e aprovação: J. C. A. Gracio, T.C.C. Madio, J. E. Santarém Segundo

FINANCIAMENTO
Não se aplica

CONSENTIMENTO DE USO DE IMAGEM
Não se aplica

APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA
Não se aplica

CONFLITO DE INTERESSES
Não se aplica

PUBLISHER
Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no Portal de Periódicos UFSC. As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES
Edgar Bisset Alvarez, Patrícia Neubert, Genilson Geraldo, Camila De Azevedo Gibbon, Jônatas Edison da Silva, Luan Soares Silva, Marcela Reinhardt e Daniela Capri.

HISTÓRICO
Recebido em: 15-10-2024 – Aprovado em: 25-03-2025 – Publicado em: 09-06-2025