

**UNIVERSIDADE FEDERAL DA PARAÍBA  
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS  
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO  
BACHARELADO EM ARQUIVOLOGIA**

**ESTHER QUÉZIA DOS SANTOS FELIPE**

**USO DAS PLATAFORMAS DE LINGUAGEM NATURAL COMO  
FERRAMENTAS NA INDEXAÇÃO DE ARQUIVOS DIGITAIS**

**JOÃO PESSOA  
2023**

**ESTHER QUÉZIA DOS SANTOS FELIPE**

**USO DAS PLATAFORMAS DE LINGUAGEM NATURAL COMO FERRAMENTAS  
NA INDEXAÇÃO DE ARQUIVOS DIGITAIS**

Trabalho de conclusão de Curso na modalidade artigo apresentado ao curso de Arquivologia da Universidade Federal da Paraíba - UFPB para a obtenção do grau de Bacharelado.

Orientador: Wagner Junqueira de Araújo

**JOÃO PESSOA**

**2023**

**Catálogo na publicação**  
**Seção de Catalogação e Classificação**

F315u Felipe, Esther Quézia Dos Santos.

    Uso das plataformas de linguagem natural como  
ferramentas na indexação de arquivos digitais / Esther  
Quézia Dos Santos Felipe. - João Pessoa, 2023.  
    30 f. : il.

    Orientação: Wagner Junqueira Araújo.  
    TCC (Graduação) - UFPB/CCSA.

    1. Arquivologia. 2. Arquivos digitais. 3. Indexação  
de arquivos digitais. 4. Inteligência artificial. 5.  
Linguagem natural. I. Araújo, Wagner Junqueira. II.  
Título.

UFPB/CCSA

CDU 930.25



MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DA PARAÍBA

FOLHA Nº 190149767 / 2023 - CCSA - CARQ (11.01.13.08)

Nº do Protocolo: 23074.055139/2023-64

João Pessoa-PB, 16 de Junho de 2023

**FOLHA DE APROVAÇÃO DE TRABALHO DE CONCLUSÃO DE CURSO**

ESTHER QUÉZIA DOS SANTOS FELIPE

APLICAÇÃO DAS PLATAFORMAS DE LINGUAGEM NATURALIZADAS COMO FERRAMENTAS NA GESTÃO DE ARQUIVOS DIGITAIS

Artigo apresentado ao Curso de graduação em Arquivologia da Universidade Federal da Paraíba, em cumprimento às exigências para a obtenção do grau de bacharel em Arquivologia.

Data de aprovação: 16 de junho de 2023

Resultado: APROVADO

**BANCA EXAMINADORA:**

Assinam eletronicamente esse documento os membros da banca examinadora, a saber: Prof. Dr. Wagner Junqueira de Araújo (orientador) e Profa. Dra. Julianne Teixeira e Silva (membro). A banca teve como membro externo a Ma. Adelaide Helena Targino Casimiro (Doutoranda PPGCI/UFPB).

*(Assinado digitalmente em 18/06/2023 11:01)*  
JULIANNE TEIXEIRA E SILVA  
PROFESSOR DO MAGISTERIO SUPERIOR  
Matrícula: 1749263

*(Assinado digitalmente em 17/06/2023 09:07)*  
WAGNER JUNQUEIRA DE ARAUJO  
PROFESSOR DO MAGISTERIO SUPERIOR  
Matrícula: 1253974

Para verificar a autenticidade deste documento entre em <https://sipac.ufpb.br/documentos/> informando seu número: **190149767**, ano: **2023**, documento(espécie): **FOLHA**, data de emissão: **16/06/2023** e o código de verificação: **281d3b1587**

# USO DAS PLATAFORMAS DE LINGUAGEM NATURAL COMO FERRAMENTAS NA INDEXAÇÃO DE ARQUIVOS DIGITAIS

## USE OF NATURAL LANGUAGE PLATFORMS AS TOOLS IN THE INDEXING OF DIGITAL ARCHIVES

Esther Quézia dos Santos Felipe <sup>1</sup>

### RESUMO

Este estudo pretende identificar as plataformas de Linguagem Natural como ferramentas no gerenciamento dos Arquivos Digitais, sobre as competências da Inteligência Artificial (IA) e como esse uso poderá complementar, requalificar ou substituir a força de trabalho humana nos Arquivos. Apresentamos resultados do levantamento bibliográfico, tendo como recorte temporal 2020-2023. Constatou-se que o uso de IA executado através das funcionalidades, *Machine Learning* e *Deep Learning*, alinhada aos processos utilizados na Arquivologia no que se refere ao gerenciamento arquivístico, contribui nas etapas da classificação documental automatizada, identificação das informações sensíveis, indexação automática, recuperação informacional, digitalização, catalogação, avaliação, assim como, seleção documental, em concordância com as normas e legislação arquivística quando operadas para esta finalidade. Os resultados apontam que as plataformas de processamento de Linguagem Naturalizada (PLN) mudarão o funcionamento arquivístico.

**Palavras-chave:** arquivologia; arquivos digitais; indexação de arquivos digitais; inteligência artificial; linguagem natural.

### ABSTRACT

This study aims to identify the Natural Language platforms as tools in the management of Digital Archives, on the competencies of Artificial Intelligence (AI), and how this usage can complement, requalify or replace human workforce in Archives. We present the results of the bibliographic survey, with a time frame of 2020-2023. It was found that the use of AI performed through functionalities such as Machine Learning and Deep Learning, aligned with the processes used in Archival Science

---

<sup>1</sup> Graduanda em Arquivologia pela Universidade Federal da Paraíba – UFPB, João Pessoa. E-mail: [esther.quezia@academico.ufpb.br](mailto:esther.quezia@academico.ufpb.br)

concerning archival management, contributes to the stages of automated document classification, identification of sensitive information, automatic indexing, information retrieval, digitization, cataloging, appraisal, as well as document selection, in accordance with archival norms and legislation when operated for this purpose. The results indicate that natural language processing platforms (NLP) will change archival operations.

**Keywords:** archival science; digital archives; digital file indexing; artificial intelligence; natural language.

## 1 INTRODUÇÃO

A automação e a Inteligência Artificial têm sido apontadas como tecnologias capazes de trazer grandes transformações para o mercado de trabalho em todo o mundo. Segundo um relatório divulgado em 2019 pela Cable News Network (CNN), cerca de 300 milhões de empregos em tempo integral podem ser automatizados pela Inteligência Artificial em todo o mundo.

A Inteligência Artificial (IA), com raízes na Ciência da Computação e presente em nossa sociedade desde a década de 1950, apresenta avanços mediante a junção de conhecimentos de diferentes áreas que fornecem inovações tecnológicas para o corpo social. (IOSCOTE, 2021, p. 163).

Diante desse cenário, é fundamental o preparo dos Arquivistas para utilizar as novas ferramentas oferecidas. Segundo Finger (2021, p. 51), “A Inteligência Artificial (IA), que abarca a linguística computacional, também chamada de processamento de língua natural, é uma área que sempre foi carregada de grandes expectativas, [...]”. Todavia, é importante ressaltar que a automação e a Inteligência Artificial geraram novas oportunidades de trabalho, sobretudo em áreas relacionadas à tecnologia e à inovação.

A Inteligência Artificial é atuante nas áreas da Ciência da Informação (na recuperação da informação, extração de informações, mineração de dados e visualização de dados), sendo assim, como parte do Processo Arquivístico. Destacam-se por meio desta inteligência, as maneiras para inseri-la de forma a desenvolver suas habilidades dentro dessa perspectiva de interação com o usuário através da Linguagem Natural (*Machine Learning* e *Deep Learning*, são respectivamente o aprendizado de máquina e aprendizado profundo) nos Arquivos Digitais.

Este trabalho tem como objetivo discutir e externalizar aos Arquivistas, assim como, as áreas relacionadas a aplicabilidade da Inteligência Artificial (IA), em processos realizados em Arquivos Digitais, tendo em vista que, a lei nº 8.159, no artigo 3º diz que “Considera-se gestão de documentos o conjunto de procedimentos e operações técnicas referentes à sua produção, tramitação, uso, avaliação e arquivamento em fase corrente e intermediária, [...]”. Mediante as disposições da lei dos arquivos executaremos as plataformas, para realização dos processos de acordo com o que preconizam as normas arquivísticas.

A Inteligência Artificial de acordo com o OpenAI (2023) pode auxiliar os Arquivos Digitais classificando, organizando, preservando documentos e informações históricas. Segundo filtro realizado no ChatGPT, de igual maneira, o resultado do ChatSonic no que se refere a Inteligência Artificial como aliada da Arquivologia, afirmaram serem capazes de automatizar os processos arquivísticos reduzindo erros e aumentando a eficiência na Gestão de Arquivos.

As plataformas de IA escolhidas como objeto de estudo deste artigo, foram selecionadas com base nos resultados oferecidos pelo ChatGPT quando filtrados foi instigado a resposta da plataforma para a pergunta: “Quais plataformas de Linguagem Natural podem auxiliar nos arquivos digitais?”, delimitando assim as seguintes plataformas: ChatSonic, ChatGPT, Tensor Flow, Py Torch, Scikit-learn, Keras, IBM Watson, Microsoft Azure AI e Google Cloud AI, sobretudo, por disponibilizarem acesso online sem a necessidade do *download* (baixar). Sendo estas, disponíveis ao acesso público de forma gratuita até o presente momento (junho de 2023), assim como, têm crescido de forma progressiva em suas funcionalidades e usuários.

O algoritmo presente nas plataformas de IA TensorFlow, Scikit-learn e PyTorch que são umas das plataformas selecionadas para esta pesquisa, que podem ser treinadas para identificar palavras-chave, sinônimos relevantes e recuperar automaticamente todos os documentos que contenham as informações filtradas, assim como, reconhecer documentos que estão em desuso ou que não são mais necessários. Conforme Openai (2023), essas são habilidades desenvolvidas conforme a dimensão temporal dos documentos e processos Arquivísticos, para a Arquivística o termo é “Temporalidade”. Estas habilidades reconhecem, por exemplo, os termos para o processo de indexação automática no sistema. Segundo Coelho (2019, p.18) “a indexação como um processo usado pelos arquivos [...] para a representação das informações contidas nos documentos, visando a recuperação dos mesmos.”

A instrução dos algoritmos das IA de forma aplicada, podem representar um padrão de informações que são protegidas por restrições de acesso devido ao seu conteúdo sensível, sigiloso e/ou privado. Conforme triagem realizada no ChatGPT, os processos de identificação das séries, tipos e tipologias documentais, podem ser realizados através das plataformas de IA diminuindo assim as falhas cometidas na



prática convencional.

Dessa forma, a fronteira de pesquisa em novos algoritmos de inteligência artificial em geral, e de processamento de língua natural em particular, deve buscar sanar a deficiência da incapacidade de aprender relações causais [...]. (FINGER, 2021, p. 64).

Mediante a filtro de respostas utilizado no Openai (2023), os termos relacionados ao gerenciamento, organização e preservação de grandes coleções de documentos históricos podem ser viabilizados por meio de plataformas como: TensorFlow, Microsoft Azure Cognitive Services, Amazon Web Services (AWS) AI Services e Google Cloud AI Platform, garantindo que as informações sejam acessadas e utilizadas pelas gerações vindouras. De forma análoga, o processo imposto no arquivo físico será aplicado no âmbito digital, no entanto, com o auxílio tecnológico das Inteligências Artificiais aplicadas na Arquivística de acordo com suas características e especificidades que serão abordadas no decorrer deste artigo.

Gava e Flores (2022, p. 230), destacam que

A Preservação Digital para documentos arquivísticos também precisou ser repensada, para refletir uma Preservação Digital Sistêmica (PDS), não mais focada no armazenamento em mídias e sua obsolescência, mas que exige uma cadeia de custódia digital arquivística, ou seja, uma cadeia de custódia arquivística ressignificada para o ambiente digital que perpassa todo o ciclo de vida dos documentos.

A automatização nos Arquivos Digitais, portanto, tem como objetivo utilizar plataformas com aplicação na arquivística, tornando os processos da gestão mais eficientes, bem como, viabilizando na economia de tempo, redução de custos e imprecisões humanas. (Informações fornecidas pelo ChatGPT, Plataforma de Linguagem Naturalizada - PLN).

## **2 PROCEDIMENTOS METODOLÓGICOS**

Trata-se de uma pesquisa de natureza aplicada, descritiva, apresentando resultados qualiquantitativos, com finalidade de verificar o potencial das plataformas através de técnicas da Linguagem Natural para coleta de dados, a fornecer uma compreensão mais completa e aprofundada sobre sua utilidade em Arquivos Digitais, além de permitir uma maior triangulação e validação dos resultados obtidos.

Considerar a importância da gestão documental para a organização de arquivos é o mesmo que dar vida, pois para existir, funcionar e se desenvolver, precisa de uma fonte que alicerce todos os procedimentos técnicos para o desenvolvimento das atividades inerentes ao arquivo. (SILVA, 2020, p.46).

A pesquisa fundamentou-se nas técnicas da Inteligência Artificial e Linguagem Natural presentes nas plataformas delimitadas para aprofundamento neste artigo, nos resultados do levantamento bibliográfico, tendo como recorte temporal 2020-2023, com temáticas em torno da junção entre Inteligência Artificial e Arquivos Digitais, apresentando resultados filtrados nas plataformas escolhidas através dos termos específicos lançados para estimular dados relevantes para esta pesquisa. Utilizamos como base para pesquisa os materiais dispostos na Revista Eletrônica de Comunicação, Portal de Revistas da Universidade de São Paulo, Google Acadêmico e Base de Dados em Ciência da Informação (BRAPCI), tendo em vista, a ineficiência do ChatGPT ao referenciar suas citações acerca dos temas filtrados. Sendo assim, mediante a pesquisa direcionada no Google Acadêmico, selecionamos os trabalhos enfáticos em temáticas de relevância para construção deste artigo, explorando assim, a BRAPCI e revistas citadas acima em razão a atualidade no que conceme aos assuntos que apontaram a arquivologia e suas especificidades.

Com intuito de validar as afirmações mencionadas, no período de outubro de 2022 a junho de 2023 foram identificadas e selecionadas para este artigo plataformas de IA, como: ChatSonic, ChatGPT (Generative Pre-trained Transformer), Tensor Flow, Py Torch, Scikit-learn, Keras, IBM Watson, Microsoft Azure AI e Google Cloud AI para compor o escopo de estudo deste trabalho, explorando as suas competências e operacionalizações principalmente as aplicações na Arquivologia. Nesse sentido, para desenvolver este artigo utilizamos as plataformas como base para consultas incorporando suas respostas.

As plataformas de Inteligência Artificial escolhidas foram: Tensor Flow, Py Torch, Scikit-learn, Keras, IBM Watson, Microsoft Azure AI e Google Cloud. Sendo assim, conforme seleção serão utilizadas para melhoria na busca e recuperação das informações em documentos digitais. Pois, têm potencial de auxílio e execução de atividades laborais dos Arquivistas na Gestão de Arquivos Digitais. Salientamos que,

mesmo com a capacidade de automatizar tarefas repetitivas, classificar com precisão, identificar informações sigilosas e tomar decisões, é necessário saber operá-las corretamente para este fim.

As funcionalidades das plataformas que utilizam a Inteligência Artificial, unidas aos métodos utilizados na Arquivologia para Gestão do Arquivo Digital, são aplicáveis em processos automatizados de classificação, identificação das informações sensíveis, indexação documental e seleção de documentos.

Nessa perspectiva, de acordo com Santos e Flores (2020, p. 770-771) observa-se que os documentos arquivísticos digitais jamais sobreviverão na inércia, sendo essencial promover a intervenção por meio de políticas e métodos de preservação digital. Sob a visão dos autores, os documentos arquivísticos são dependentes dos sistemas informatizados (automatizados), não limitando apenas por uma tecnologia recente.

As próximas seções e subseções vão detalhar as plataformas escolhidas para estudo neste trabalho, bem como, suas funcionalidades na Arquivologia de acordo com as suas características preexistentes. Dessa forma, os procedimentos de limitação e busca empregados em cada etapa principal da pesquisa, são mediante as apurações feitas nas plataformas de IA.

Para Rockembach (2021, p. 239) a construção de Arquivos Inteligentes torna-se necessário formar uma base a partir da Inteligência Artificial e de Sistemas Especializados Inteligentes. Adiante explicita que,

A aplicação de modelos estatísticos e processamento computacional, por exemplo, auxiliam no estabelecimento de formas de aprendizagem de máquina, que podem ser aplicadas alimentando o modelo com textos, imagens e diversos tipos de dados.

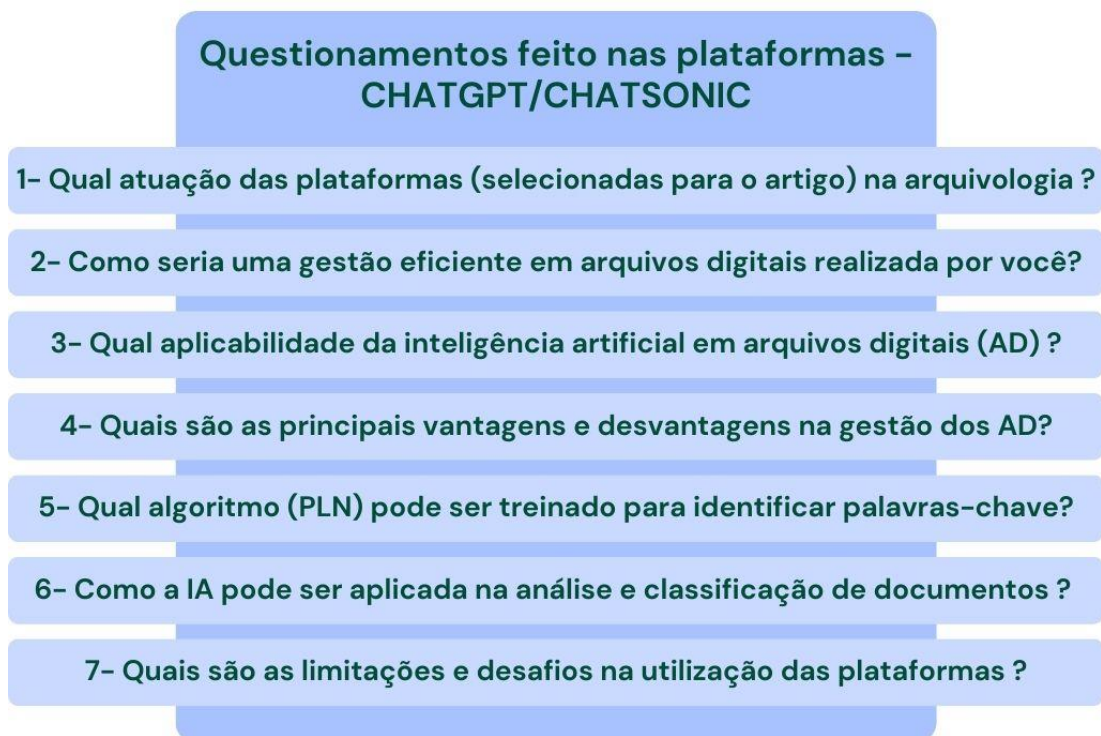
Com o filtro realizado acerca de questionamentos pontuais, tornou-se possível verificar o potencial das plataformas de Linguagem Naturalizada, compreendendo sua atuação em Arquivos Digitais. Pois, a Inteligência Artificial aplicada em processos Arquivísticos, contribuirá como ferramenta de auxílio na classificação automatizada, identificação das informações sensíveis, indexação documental, seleção de documentos e recuperação da informação.

Espera-se que este estudo contribua para o avanço das pesquisas referentes a Inteligência Artificial aplicada à Gestão de Arquivos Digitais, e para o

desenvolvimento de soluções mais eficientes e acessíveis para a recuperação informacional em documentos digitais. Assim, delimitamos com a realização de uma revisão sistemática das ferramentas de IA, o ChatGPT e ChatSonic, com enfoque específico de aplicação na Gestão dos Arquivos Digitais.

A **Figura 1** apresenta as questões realizadas nas plataformas ChatGPT e ChatSonic, tais quais, foram fundamentais para filtrar o conteúdo aplicado na construção da temática apresentada neste artigo, tendo em vista que, como os objetos de estudo são aplicados à Arquivística, os dados informacionais apresentados nas respectivas plataformas foram primordiais para pesquisa realizada.

**Figura 1 - Questionamento no CHATGPT e CHATSONIC**



**Fonte:** Dados da pesquisa (2023).

Acentuamos que, todos os dados textuais obtidos foram resultados das aplicações e interações às plataformas de Linguagem Naturalizada escolhidas para estudo, bem como, sugeridas neste artigo como ferramentas aplicáveis à Gestão dos Arquivos Digitais.

Com a automatização dos processos Arquivísticos, as plataformas de Linguagens Naturalizadas, fazem a Arquivologia menos técnica, mas, operacionalizada para o Arquivista, em relação a identificação documental, classificação e indexação. Contudo, apesar do método escolhido para automatizar os processos citados, a intervenção Arquivística na condução e aplicação da ferramenta é crucial.

Este estudo selecionou as informações fornecidas pelas plataformas digitais de Inteligência Artificial acerca de sua utilização na gestão eficiente de Arquivos Digitais, atuação na indexação, preservação dos documentos digitais e classificação documental.

### **3 UTILIZAÇÃO DAS PLATAFORMAS DE INTELIGÊNCIA ARTIFICIAL**

A Inteligência Artificial tem sido uma das áreas mais promissoras em pesquisa e desenvolvimento tecnológico nas últimas décadas. Uma das aplicações mais recentes dessa tecnologia tem sido no campo da Gestão de Arquivos Digitais, que tem como desafio garantir a acessibilidade e recuperação de informações em grandes volumes de Documentos Digitais. Na visão de Santos e Flores (2020, p. 767),

No entanto, a preservação digital requer alterações, como por exemplo, os procedimentos de migração e até mesmo cópia, tendo em vista a fragilidade dos suportes e a obsolescência tecnológica de hardware e software.

De acordo com Moisés Rockembach (2021, p. 240) as questões que podem ser apontadas sobre a IA e seu vínculo em outros campos de atuação, neste caso, a Ciência da Informação e os Arquivos, é devido às possibilidades de projetos de Inteligência Artificial voltados a observação dos problemas, lacunas do ambiente profissional e científico.

As plataformas de IA tornaram-se populares devido à capacidade de fornecer soluções automatizadas para tarefas simples e complexas, tais como: processamento de imagem, tradução de idiomas e diagnósticos médicos. Dessa forma, o avanço desta inteligência nas plataformas, apesar de sofisticado, ainda é acessível, proporcionando aos seus usuários a viabilização de processos e tomada de decisões. Nesta perspectiva, as plataformas oferecem recursos e funcionalidades

que permitem executar o *Machine Learning* (aprendizado de máquina), *Deep Learning* (aprendizado profundo), Processamento de Linguagem Natural (PLN), e a Visão Computacional. Sobretudo, por se constituírem de ferramentas, bibliotecas, algoritmos e tecnologias que permitem a criação, desenvolvimento e implantação de aplicações baseadas na Inteligência Artificial Aplicada.

Sendo assim, permitem a integração de modelos de IA em aplicações existentes, tornando facilitadora e acessível a adoção das tecnologias ofertadas. As plataformas permitem aos desenvolvedores e cientistas de dados criar soluções de Inteligência Artificial personalizadas, além de acelerar o processo de desenvolvimento com a utilização de bibliotecas e ferramentas pré-existentes.

Conforme citado na introdução deste artigo, o número de plataformas disponíveis para o acesso público está em escala crescente nas últimas décadas, sendo estas particularidades em suas características e recursos fomentados na utilização. Mediante o resultado obtido na plataforma ChatGPT para filtro de plataformas para utilização em Arquivos Digitais, a seguir serão apresentadas sete plataformas de AI selecionadas para análise neste artigo:

1. **TensorFlow:** é uma das plataformas de AI mais populares, criada pelo Google. Suas principais características são permitir a criação e treinamento de modelos de *Machine Learning* e *Deep Learning*, bem como, a implantação desses modelos em outras plataformas.
2. **PyTorch:** esta plataforma de AI de código aberto é criação do Facebook. Ela possibilita ao usuário criar modelos de *Machine Learning* e *Deep Learning* com facilidade e eficiência, assim como, a TensorFlow.
3. **Scikit-learn:** é uma biblioteca de *Machine Learning* de código aberto para a linguagem *python* (linguagem de programação de alto nível, interpretada, de propósito geral e com uma sintaxe simples e legível), que oferece uma ampla variedade de algoritmos de aprendizado de máquina.
4. **Keras:** com foco em simplicidade e facilidade de uso, consiste em uma biblioteca de código aberto, bem como, a plataforma scikit-learn, e pode ser utilizada para a criação e treinamento de modelos de Deep Learning.
5. **IBM Watson:** diferentemente das plataformas citadas, a IBM Watson, tem enfoque na infinidade de ferramentas, incluindo chatbots, análise de dados



e reconhecimento de imagens e voz.

6. **Microsoft Azure AI:** oferece abundantes serviços e ferramentas de Inteligência Artificial, que incluem Machine Learning, reconhecimento de fala e visão computacional.
7. **Google Cloud AI:** reconhecimento de imagem e fala, tradução automática e previsão de demanda, são serviços oferecidos que estão intrinsecamente relacionados com o *Machine Learning* e *Deep Learning*.

Nos pontos acima, foram descritas apenas as maneiras pelas quais as plataformas de Inteligência Artificial podem ser usadas na Arquivologia com enfoque nos Arquivos Digitais. De acordo com os processos, delimitamos onde cada plataforma será importante no que se refere às etapas para organização e Gestão Documental.

Observamos que as plataformas de Inteligência Artificial como TensorFlow, Scikit-learn, IBM Watson e Google Cloud AI são ideais para classificação e categorização de Arquivos Digitais em congruência com suas atribuições, baseadas em critérios de tipo documental, cronológico, por assunto e por autor. Essa classificação realizada pode ser útil na organização de grandes volumes documentais, facilitando a recuperação, acessibilidade e uso destes documentos.

No processamento de Linguagem Natural as plataformas como PyTorch, Keras e IBM Watson são enquadradas na gestão de grandes volumes de documentos textuais (a exemplo de textos de jornais, jornais, documentos históricos e correspondências) nos Arquivos Digitais, bem como, são capazes de identificar padrões e informações relevantes nos volumes de texto facilitando recuperação.

Plataformas como a Microsoft Azure AI e Google Cloud AI, são relevantes para análise e mineração de dados, pois, têm a capacidade de processar grandes quantidades de dados de Arquivos Digitais para identificar padrões, anomalias e tendências. Dessa forma, a análise tem executabilidade na tomada de decisões referentes a eliminação, preservação e digitalização.

Por conseguinte, as plataformas de IA, Google Cloud AI e Microsoft Azure AI, são eficientes para converter imagens de documentos em texto digital pesquisável através do reconhecimento óptico de caracteres (OCR), auxiliando no processo de busca acessível nos conteúdos de Arquivos Digitais.

As plataformas apontadas são uma seleção feita para aplicação neste artigo, em relação às plataformas de IA disponíveis no mercado atualmente, sendo útil para diferentes tipos de projetos e necessidades dos seus usuários como evidenciado. Salientamos que as plataformas de Linguagem Natural, abordadas nos parágrafos anteriores e subseções posteriores, têm suas versões gratuitas disponíveis até o presente momento (junho de 2023).

### 3.1 Plataformas de Processamento de Linguagem Natural (PLN)

O Processamento de Linguagem Natural, na Inteligência Artificial, é uma área de pesquisa que tem como objetivo desenvolver algoritmos, bem como, modelos capazes de interpretar e gerar a Linguagem Natural. Neste sentido, suas aplicações podem ser realizadas em chatbots e assistentes virtuais através dos sistemas que interagem com usuários em Linguagem Natural para fornecer informações, tirar dúvidas ou realizar tarefas.

Para Finger (2021, p. 60), a análise de sentimento e a etiquetagem de trechos relevantes de um texto de acordo com a aplicação, foi considerada uma nova representação aplicada em diversas tarefas de processamento de língua natural. Esta linguagem pode ser utilizada em análises de dados e em sistemas de Inteligência Artificial, com a finalidade da tomada de decisões, permitindo que os resultados sejam apresentados de forma mais atingível e de simples entendimento.

Para Openai (2023) as plataformas de PLN são sistemas computacionais projetados para entender, processar e gerar Linguagem Natural, permitindo a interação entre humanos e computadores por meio da linguagem falada ou escrita. Sendo assim, utilizam tecnologias de processamento, como modelos de linguagem, algoritmos de aprendizado de máquina e análise semântica, para compreender de forma a produzir Linguagem Natural em um contexto específico (Openai, 2023).

À vista disso, a sumarização de texto, consiste na aplicação do sistema para identificar as informações mais relevantes de um texto e resumi-las. Por conseguinte, quando aplicada diretamente a análise de sentimentos, as técnicas de PLN têm utilização na identificação do tom emocional de textos, avaliações de produtos ou comentários em redes sociais.

A tradução automática baseada em sistemas que utilizam técnicas de PLN é outra ferramenta utilizada para traduzir textos de uma língua para outra, por fim, o



reconhecimento de voz por meio das técnicas para interpretar a fala humana e transformá-la em texto ou em comandos para outros sistemas, podem ser consideradas aplicações desta técnica de linguagem.

Nos parágrafos acima foram descritos conceitos do processo de Linguagem Naturalizada, utilização e aplicações do sistema nas plataformas. Posto isto, de maneira que se amplie o entendimento de sua execução, abaixo estarão dispostas as plataformas de Linguagem Naturalizada disponíveis atualmente no mercado e escolhidas através de seleção feita no ChatGPT, com intuito de compor este artigo como parte do estudo.

1. **Dialogflow:** possui uma versão gratuita limitada, com recursos avançados disponíveis em planos pagos, esta plataforma da Google permite criar chatbots e assistentes virtuais personalizados.
2. **IBM Watson Assistant:** é uma plataforma de conversação baseada em nuvem que utiliza Inteligência Artificial para entender e responder às perguntas dos usuários, sua versão gratuita apresenta limitações, mas dispõe de planos pagos com recursos adicionais.
3. **Amazon Lex:** baseada em voz essa plataforma de conversação, permite a criação de chatbots e assistentes virtuais, oferece um plano gratuito limitado, com preços adicionais para recursos avançados.
4. **Microsoft Bot Framework:** esta plataforma desenvolve chatbots e suporta várias linguagens de programação, assim como, canais de comunicação. Possui planos pagos com recursos adicionais, tendo em vista que sua versão gratuita tem recursos limitados.
5. **Wit.ai:** utilizada para o desenvolvimento de chatbots, a plataforma adquirida pela Facebook utiliza o PLN para entender e responder às perguntas dos usuários. Sua versão gratuita é limitada, oferecendo preços adicionais para recursos avançados.
6. **ChatSonic:** usa tecnologias de PLN para compreender as perguntas e respostas dos usuários e fornecer respostas relevantes e precisas, a plataforma permite a criação de chatbots de conversação com aplicações em áreas de atendimento ao cliente, vendas e suporte técnico.
7. **ChatGPT:** modelo de Linguagem Natural desenvolvido pela OpenAI (Open

Artificial Intelligence), que permite a interação por meio de texto com um sistema de Inteligência Artificial capaz de entender e gerar Linguagem Natural, podendo ser utilizado em conjunto com outras tecnologias de PLN na criação de assistentes virtuais e chatbots.

As plataformas citadas estão sendo cada vez mais exploradas em setores como a educação, de forma a melhorar a experiência do usuário e aumentando a eficiência de processos. Sendo assim, podem ser utilizadas para desenvolver chatbots e assistentes virtuais como explanadas acima, mas, são aplicáveis em outras funções da automatização de processos. A partir daqui os Arquivísticos serão abordados no desenvolver das próximas subseções.

### 3.2 Aplicações em Processos Arquivísticos: ChatGPT e ChatSonic

As plataformas escolhidas juntamente a Inteligência Artificial aplicada, podem auxiliar a Arquivologia por meio das técnicas e ferramentas específicas, apresentam o layout extremamente funcional, que permite aos usuários o acesso apenas vinculando a plataforma ao e-mail pessoal, sem a necessidade de preencher longas listagem para cadastros, bem com, apresentam aumento na utilização para finalidades acadêmica e profissional.

Dessa forma, é possível personalizar o ChatSonic com base nas seguintes configurações oferecidas ao usuário: Integração a pesquisa do Google, Conversão de texto para fala, memória de acompanhamento, procura resultados detalhados ou concisos, localização por país, imagem de difusão estável ou DALL-E (modelo criador de imagem a partir da descrição textual). São estas as preferências que podem ser obtidas no chat ao iniciar a pesquisa/conversa.

Baseado nos ChatGPT e ChatSonic, obtivemos respostas em relação a atuação da IA na Arquivologia, bem como, foi observado a amplitude das plataformas em recursos para automação da área, desta maneira levantou-se o seguinte questionamento: *“Como os processos arquivísticos podem ser automatizados com a IA?”*, unificamos os resultados obtidos dispondo-os nos parágrafos subsequentes.

**Classificação automática:** a IA pode ser treinada para classificar documentos automaticamente, identificando palavras-chave, tópicos e outros. Dessa

forma, a classificação automática envolve o uso de algoritmos de Aprendizado de Máquina (*Machine Learning*) para categorizar documentos de acordo com o conjunto de regras pré-definidas, de forma a acelerar o processo de arquivamento, eliminando a necessidade de o Arquivista executar esta tarefa manual.

**Indexação automática:** a indexação automática envolve a extração de informações dos documentos e a atribuição de metadados para facilitar a pesquisa e a recuperação de informações. Dessa maneira, indexar automaticamente documentos, atribuição dos metadados como facilitadores da pesquisa e recuperação de informações (como nome do autor, data de criação, tipo de documento) são procedimentos executáveis pela IA.

**Reconhecimento de voz e OCR:** o reconhecimento de voz e OCR (reconhecimento óptico de caracteres) são técnicas usadas para converter áudio e documentos em texto legível, essa tecnologia permite a digitalização de documentos físicos, permitindo que sejam facilmente arquivados digitalmente e pesquisáveis. Assim, conteúdos gravados em fitas cassete e meios analógicos podem ser digitalizados e arquivados eletronicamente, sobretudo os documentos antigos que não foram originalmente digitalizados.

**Identificação de conteúdo sensível:** utilizado na identificação dos documentos que contenham o teor sensível no que se refere a informações pessoais ou dados financeiros, separando-os dos demais documentos para que possam ser armazenados de forma segura, por conseguinte viabiliza a recuperação documental, por meio de buscas por palavras-chave, tags, ou até mesmo análise semântica.

**Recomendações de arquivamento:** com base em padrões históricos de arquivamento, preferências de usuário e outros fatores, a IA pode ser usada para recomendar o local de arquivamento de um documento.

**Preservação digital:** a Inteligência Artificial pode ser utilizada para monitorar a integridade de Arquivos Digitais e garantir a sua preservação a longo prazo, de forma a otimizar processos, reduzir custos e aumentar a eficiência na Gestão de Arquivos, tornando os processos Arquivísticos automatizados.

As plataformas utilizadas como base no filtro dos resultados quanto à utilização das mesmas na Arquivística, os ChatGPT e ChatSonic, explicitaram as técnicas abaixo descritas, que estão presentes em seu sistema metodológico para aplicação e atuação na gestão dos Arquivos Digitais.

**Processamento de Linguagem Natural (PLN):** essa técnica permite que a Inteligência Artificial compreenda e analise textos, o que pode ser útil na indexação e recuperação de documentos. Além disso, a IA pode ser utilizada para automatizar a busca documental, permitindo que os usuários encontrem rapidamente o que precisam através de comandos de voz ou Linguagem Natural.

**Aprendizado de Máquina (*Machine Learning*):** técnica que permite a Inteligência Artificial aprender com dados históricos identificando padrões e tendências de comportamento, podendo ser utilizado na classificação automática de documentos.

**Automação de processos:** a IA pode ser aplicada na automação dos processos na gestão de Arquivos Digitais como a indexação, busca e classificação de documentos (classifica automaticamente documentos com base em seu conteúdo), aplicando as tags apropriadas e organizando-os em pastas específicas.

**Digitalização de documentos:** é um processo que envolve a conversão de documentos físicos em Arquivos Digitais, a IA pode atuar através do uso de scanners inteligentes, que reconhecem automaticamente o tipo de documento e aplica as configurações de digitalização, como também, reconhecer e extrair informações de documentos digitalizados, a exemplo de datas, nomes e números.

**Análise de dados (*Data Analytics*):** a Inteligência Artificial analisa grandes quantidades de dados para identificar padrões e tendências, tornando-se útil na Gestão de Arquivos. Sendo assim, na Gestão de Arquivos, a análise de dados pode ser utilizada para identificar documentos que estão desatualizados ou que precisam ser atualizados, monitorando a usabilidade dos documentos (identificando quais documentos são mais utilizados) e quais são menos acessados.

**Sistemas de gerenciamento de conteúdo (CMS):** são softwares que permitem a criação, edição e publicação de conteúdo digital, como documentos, imagens e vídeos. Esses sistemas podem ser integrados com a Inteligência Artificial para ajudar na organização, classificação e busca de documentos. Por exemplo, um CMS (*Content Management System*) pode utilizar algoritmos de aprendizado de máquina para identificar padrões de uso e sugestões de tags para documentos com base em seus conteúdos.

Os conceitos apresentados acima, bem como, são atributos da IA, apropriados para automatizar processos Arquivísticos. À medida que a tecnologia

continua a avançar, é provável que surjam novos métodos de IA com aplicação na Arquivologia. A automação de processos Arquivísticos com a Inteligência Artificial tende a se tornar uma realidade presente nas organizações, principalmente, por oferecer séries de soluções para melhoria na eficiência e precisão na Gestão de Documentos, resultando em processos menos suscetíveis a falhas.

A automação de processos Arquivísticos com ferramentas de Inteligência Artificial, oferecidas nas plataformas selecionadas nesta subseção (ChatGPT e ChatSonic), abarcam benefícios para o arquivo possuinte de melhoria na eficiência, redução de custos e minimização de processos com erros. Assim sendo, as tecnologias de IA como aliadas ao gerenciamento de Arquivos Digitais, são indicativos de vantagens competitivas e aumento de eficiência operacional.

### 3.2.1 Validação Experimental: Resultados

Buscando fornecer resultados através da experiência prática na plataforma, aplicamos as funcionalidades no âmbito Arquivístico solicitando-o a realização de atividades, permitindo a validação dos resultados obtidos. Tendo em vista que, ao questionarmos a plataforma sobre seus métodos para classificação de acordo com as normas e legislação arquivística, foi explicitado pelo chat que o mesmo utilizaria para a subfunção de Documentação Pessoal, a linguagem técnica da Arquivologia para descrever as atividades de identificação e registro documental. Conforme Mohammed e Sahari (2023, p. 3, tradução nossa) “O ChatSonic é treinado em um vasto corpus de dados de texto e pode fornecer respostas precisas para uma ampla gama de perguntas em diversas áreas do conhecimento.”

Sendo assim, utilizou-se os termos "classificação", "indexação", "organização" e "descrição" para descrever as atividades de identificação e registro dos documentos. Além disso, os termos como "série documental", "fundo documental" e "plano de classificação" para descrever a estruturação dos documentos e sua organização em um sistema Arquivístico.

Com base nos assuntos sugeridos ao chat para criação de uma tabela descritiva de acordo com a Arquivística, com documentos presentes em uma empresa de tecnologia voltada para digitalização, foi fornecida por ele a **Tabela 1**. Após criação da tabela é oferecido na plataforma as opções de: gostar, não gostar,

editar no Sonic Editor, download, texto para fala, copiar para área de trabalho e editar.

O ChatSonic sugeriu também que a tabela criada, com teor descritivo, permite uma melhor organização e gestão dos documentos presentes na empresa de tecnologia para digitalização, facilitando a localização e recuperação de informações importantes. Além disso, segue as normas arquivísticas, garantindo a preservação e proteção dos documentos ao longo do tempo.

De acordo com Magalhães (2020, p. 72),

Na tarefa de classificação um dos grupos possui somente um atributo, que corresponde ao atributo-alvo, ou seja, a propriedade pela qual se deve fazer a predição de um valor. Nesse caso, o atributo é categórico (domínio composto por categorias/classes) e o outro conjunto contém os atributos a serem utilizados na predição do valor, denominados previsores ou de predição.

#### **Quadro 1: Classificação Descritiva de documentos voltados a Tecnologia**

<b>FUNÇÃO</b>	<b>SUBFUNÇÃO</b>	<b>ATIVIDADES</b>	<b>SÉRIE DOCUMENTAL</b>
Gestão de Projetos	Planejamento	Cronogramas	001 - Cronogramas
Gestão de Projetos	Planejamento	Orçamentos	002 - Orçamentos
Gestão de Projetos	Execução	Relatórios de status	003 - Relatórios de Status
Gestão de Projetos	Execução	Documentos de Requisitos	004 - Documentos de Requisitos
Gestão de Projetos	Encerramento	Documentos de Aceitação	005 - Documentos de Aceitação
Gestão de Projetos	Encerramento	Documentos de Encerramento	006 - Documentos de Encerramento
Gestão de Documentos	digitalização	Documentos Digitalizados	007 - Documentos Digitalizados
Gestão de Documentos	Indexação	Índices de Documentos	008 - Índices de Documentos
Gestão de Documentos	armazenamento	Documentos Armazenados	009 - Documentos Armazenados
Gestão de Documentos	Descarte	Documentos Descartados	010 - Documentos Descartados

**Fonte: Resultado do ChatSonic (2023).**

Segundo Magalhães (2020, p. 71), a classificação é “o processo de adotar



algum tipo de entrada e atribuir um rótulo a ela”. A autora descreve posteriormente que a inserção dos documentos em classes é um processo conhecido como “classificação de textos”, bem como, existem diferentes estudos e algoritmos desenvolvidos para aprimorar os recursos que permitam organizar a informação e “induzir automaticamente os sistemas capazes de lidar com problemas de classificação”.

Por conseguinte, realizou-se o teste em relação a indexação com intuito de explorar as habilidades expressadas pelo próprio chat referente a sua atuação em Arquivos Digitais, auxiliando nos processos técnicos, na realização de procedimentos basilares na gestão de um arquivo. Dessa vez, o objeto de constatação foi uma declaração de curso emitida no Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA).

## Figura 2: Indexação da Declaração de Vínculo

CS

Para indexar este documento de acordo com as normas arquivísticas, é necessário identificar as informações que são relevantes para a sua recuperação futura. Algumas informações importantes presentes neste documento são:

- Nome completo do aluno: Esther Quezia dos Santos Felipe
- Número de matrícula: **201000000000**
- Curso: Arquivologia (Bacharelado)
- Instituição: Universidade Federal da Paraíba
- Carga horária cumprida: 95,65% do currículo integralizado (2640 horas de 2760 exigidas)
- Modalidade: Noturno
- Local: CCSA/João Pessoa - Presencial
- Data de emissão: 31 de Maio de 2023
- Código de verificação: e6ae83a047

Include latest Google data:  ⓘ Regenerate Response Enable Memory:

Talking to: **declaracao\_201000000000.pdf** + New prompt Charge: 1\* output words

+ Write a Facebook ad for Chatsonic. ▶ 🎤

Fonte: Resultado do ChatSonic (2023).

O documento pode ser indexado de acordo com as normas arquivísticas utilizando informações como: nome completo do aluno, número de matrícula, curso, carga horária cumprida e data de emissão documental. Além disso, é importante incluir o código de verificação para garantir a autenticidade do documento. A indexação pode ser feita em um sistema de gerenciamento de documentos ou em uma planilha, por exemplo, para facilitar a busca e recuperação do documento no futuro, foram as recomendações do ChatSonic para o documento inserido.

A classificação e a indexação de documentos, portanto, é uma tarefa fundamental na garantia da organização e acesso às informações contidas. Logo, com o auxílio de ferramentas como o ChatSonic e as demais plataformas citadas neste trabalho, é possível realizar essa tarefa de forma mais ágil e concisa, seguindo as normas e legislações aplicáveis.

No entanto, é necessário operacionalizar as plataformas, para que apresentem os resultados sempre de acordo com as normas preestabelecidas para estes processos na Arquivologia. Os primeiros resultados são genéricos sem filtros de moldagem para apresentação de acordo com a classificação esperada pelos Arquivistas, mas, após descrevê-lo como deve apresentar os resultados, utiliza do *Machine Learning* para adequação das respostas fornecidas.

Diferentemente no ChatGPT, analisamos a precisão no que se refere a entrega de dados filtrados congruentes com a seleção feita com finalidade de enriquecer a pesquisa quanto ao referencial teórico trabalhando especificamente em resultados para citações dos temas tratados neste estudo, bem como, referências de acordo com as normas da ABNT.

Sendo assim, o ChatGPT trouxe citações de autores que falam sobre “Inteligência Artificial e Arquivos Digitais”, “Arquivologia Automatizada”, “Inteligência Artificial e Classificação Documental” e entre outros, no quais, inicialmente foram inseridos na pesquisa, mas, ao serem referenciados de acordo com a ABNT e constatadas a veracidade das informações entregues, os links de acessos, títulos das pesquisas e ano, não foram encontrados.

Segundo Santos e Flores (2020, p. 772), “[...] a preservação digital passa a ser pensada de maneira sistêmica, de modo que existam políticas e padrões a serem seguidos em busca da construção de um ambiente confiável”. Conforme Gava e Flores (2022, p. 230) “a própria definição indica que a preservação digital



deve lidar com diferentes aspectos, o que inclui a evolução tecnológica”.

Apesar de apresentar inúmeras fontes referenciais para suas apurações, o ChatGPT ainda apresenta falhas técnicas na entrega desses dados, necessita de adequações sistemáticas e análise inteligente das respostas, dessa forma, seus resultados devem ser continuamente confrontados na própria plataforma.

#### 4 PLATAFORMAS NO GERENCIAMENTO DE ARQUIVOS DIGITAIS

O estudo explora as possibilidades da tecnologia de Inteligência Artificial buscando aprimorar a gestão e a acessibilidade de documentos digitais em diferentes contextos, a contribuir para a melhoria da organização e preservação do patrimônio histórico e cultural. Durante a pesquisa, realizamos a aplicação de filtros através de perguntas acerca do papel das plataformas de IA no gerenciamento de Arquivos Digitais.

Conforme explicitado na **Figura 1**, essas perguntas foram investigadas com base em referências teóricas e estudos de caso fornecidos pela plataforma de PLN ChatGPT, a fim de compreender melhor o potencial dessas tecnologias no contexto dos Arquivos Digitais. As plataformas Tensor Flow, Py Torch, Scikit-learn, Keras, IBM Watson, Microsoft Azure AI e Google Cloud também oferecem alternativas de aplicação na Gestão de Arquivos Digitais. Nessa perspectiva, Santos e Flores (2020, p. 768) afirmam que,

A gestão de documentos deverá contemplar atividades de produção, aquisição, classificação, avaliação, preservação, descrição e acesso. Tais ações devem ser pautadas na eficiência e visar o reuso da informação, de modo que este serviço seja pensado tanto para os administradores quanto para os usuários.

Por meio das técnicas de Inteligência Artificial, é possível melhorar a organização, classificação, busca e recuperação de documentos, além de possibilitar a identificação das informações, detecção de anomalias e riscos à segurança da informação, que permitem a identificação de comportamentos suspeitos e a prevenção de ataques cibernéticos.

A plataforma TensorFlow, tem potencial utilização na classificação automática de documentos, onde a rede neural é treinada para reconhecer diferentes tipos de documentos e categorizá-los automaticamente. Nesta perspectiva, a técnica de IA

baseada na estrutura e funcionamento do cérebro humano, é denominada redes neurais. Consistindo em uma grande quantidade de unidades de processamento, chamadas de neurônios, conectadas entre si, juntos, eles trabalham para reconhecer padrões e tomar decisões. Na visão de Finger (2021, p. 62) “O grande impulso no processamento de língua natural, quer seja em inglês, em chinês ou em português, se deu utilizando uma classe de algoritmos chamados de redes neurais.”

As plataformas IBM Watson e Microsoft Azure AI analisam textos, traduzem idiomas, reconhecem voz, analisam sentimentos, reconhecimento de fala e detectam imagens. Na Arquivística, as plataformas irão atuar respectivamente para identificar e extrair informações dos documentos e auxiliar na catalogação de documentos em diferentes idiomas, assim como, auxiliará na análise de sentimentos dos documentos e na detecção de informações confidenciais nas documentações.

No entanto, as plataformas PyTorch e Scikit-learn são fundamentadas respectivamente, no aprendizado profundo aplicado na classificação de documentos, identificação de informações confidenciais e sigilosas em documentos, contudo. Contudo, a Google Cloud AI tem função disponível para o reconhecimento de imagem, fala e processamento de Linguagem Natural. Na Arquivologia, a plataforma é útil na detecção de imagens nos documentos, reconhecimento de voz em documentos gravados e para análise de Linguagem Natural em documentos.

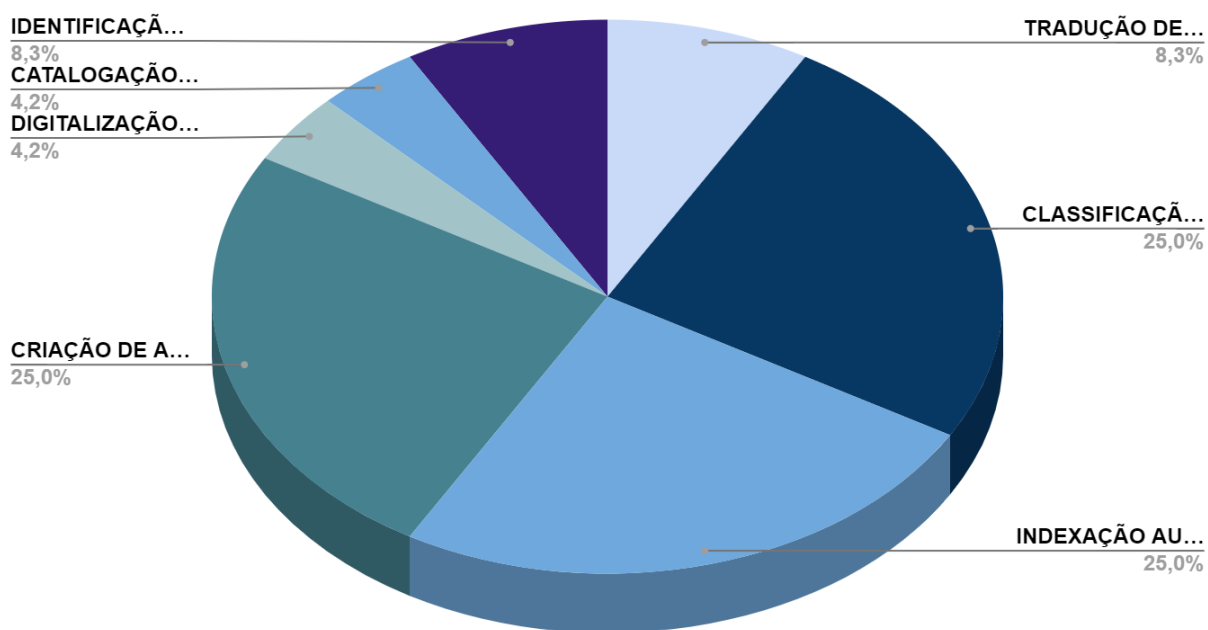
O Aprendizado de Máquina (Machine Learning), por exemplo, permite que algoritmos sejam treinados para reconhecer padrões e características específicas em documentos digitais, facilitando a classificação e organização de arquivos. Além disso, a utilização do processamento de Linguagem Natural (PLN) possibilita a extração de informações relevantes (como entidades nomeadas, tópicos e sentimentos expressos) a partir do conteúdo dos documentos. A plataforma Keras é uma biblioteca de *Machine Learning* em Python, assim como a Scikit-learn, no entanto a Keras é executável na criação de modelos de classificação de documentos, análise de sentimentos de textos, sendo adequada para análise de correspondências e documentos históricos.

Uma das principais vantagens de usar o Python é a capacidade de interagir diretamente com o código, visto que o Aprendizado de Máquina e a Análise de Dados são processos fundamentalmente interativos, nos quais os dados conduzem a análise. (MAGALHÃES, 2022, p. 110).

Tal qual, o clustering de documentos, técnica de análise de dados na qual os documentos semelhantes são agrupados automaticamente com base em seus atributos, são propriedades da Scikit-Learn no gerenciamento de Arquivos Digitais.

Dentre as plataformas mencionadas, foi demonstrado que a atuação pode ser realizada em diferentes tarefas relacionadas à classificação documental, indexação automática, identificação de texto e imagem, digitalização de documentos, catalogação documental, criação de algoritmos e tradução de idiomas. Sendo assim, tornou-se possível relatar em gráfico os percentuais de cada atividade que se pode executar nas plataformas.

**Gráfico 1 - Aplicações arquivísticas das Plataformas no Gerenciamento**



**Fonte:** Dados da pesquisa (2023).

As plataformas que compõem as porcentagem do gráfico nos processos de **Classificação Documental (25%)** são: Tensor Flow, PyTorch, Scikit-learn, Keras IBM Watson, Microsoft Azure AI e Google Cloud AI; **Indexação automática (25%)** são: Tensor Flow, PyTorch, Scikit-learn, Keras, IBM Watson, Microsoft Azure AI e Google Cloud AI; **Criação de algoritmos (25%)** são: Tensor Flow, PyTorch, Scikit-learn, Keras, IBM Watson, Microsoft Azure AI e Google Cloud AI; **Digitalização de documentos (4,2%)** são: IBM Watson, Microsoft Azure AI e Google Cloud AI;

**Catálogo documental (4,2%)** são: IBM Watson, Microsoft Azure AI e Google Cloud AI; **Identificação de texto e imagem (8,3%)** são: Tensor Flow, PyTorch, IBM Watson, Microsoft Azure AI e Google Cloud AI; **Tradução de idiomas (8,3%)** são: Tensor Flow, PyTorch, IBM Watson, Microsoft Azure AI e Google Cloud AI.

Portanto, todas as plataformas mencionadas podem atuar na maioria dos pontos citados usando técnicas de Aprendizado de Máquina (*Machine Learning*), com exceção da digitalização de documentos e catalogação documental, que são recursos específicos oferecidos pelo IBM Watson, Microsoft Azure AI e Google Cloud AI. Isso ocorre porque a maioria das plataformas mencionadas suportam a criação, treinamento e implantação de modelos de Aprendizado de Máquina para uma ampla gama de tarefas, incluindo classificação, indexação, identificação, criação de algoritmos e tradução. No entanto, é importante ressaltar que cada plataforma pode ter sua própria sintaxe e abordagem específica para a implementação de modelos de *Machine Learning*.

#### 4.1 Vantagens e Desvantagens na Utilização das Plataformas

As principais vantagens de utilizar plataformas de Inteligência Artificial como Tensor Flow, Py Torch, Scikit-learn, Keras, IBM Watson, Microsoft Azure AI e Google Cloud para gerenciar Arquivos Digitais incluem a capacidade de automatizar tarefas repetitivas e de grande escala, melhorar a precisão na classificação e busca de documentos. Permitindo a identificação de informações sensíveis e pessoais nos documentos, no entanto, possibilita a tomada de decisões com base nos dados disponíveis nos arquivos. Além disso, essas plataformas podem ser integradas a sistemas já existentes e aprimorar a segurança e proteção dos dados armazenados nos arquivos.

Por outro lado, as desvantagens incluem: o alto custo e a complexidade na implementação e manutenção dessas plataformas, a dependência de especialistas em Inteligência Artificial para operá-las, a necessidade de grande quantidade de dados para treinar os modelos de aprendizado de máquina e a possibilidade de vieses nos resultados (caso os dados de treinamento não sejam representativos ou diversificados o suficiente).

No que se refere a plataforma ChatSonic, mesmo oferecendo mais de 100

recursos de escrita, apresenta uma limitação da quantidade de palavras (10.000 palavras) utilizadas, podendo prolongar a utilização da plataforma com custo adicional, as demais plataformas citadas nesta pesquisa têm o valor adicional, porém, cobrado mediante a escolha de serviços específicos.

Salientamos que, essas plataformas não substituem o trabalho humano na gestão de arquivos, mas, podem ser utilizadas como ferramentas auxiliares a complementar na otimização dos processos Arquivísticos, sobretudo nos Arquivos Digitais, dessa forma viabilizando as melhorias para o gerenciamento de documentações digitais.

Na análise e classificação de documentos digitais, a identificação dos padrões visuais nos documentos (como imagens e gráficos) são realizadas através das redes neurais convolucionais, que classificam os documentos, facilitando o acesso e a recuperação documental. Além disso, as plataformas de IA também podem ser utilizadas para detecção de plágio em documentos digitais, identificando trechos de texto que foram copiados de outras fontes.

## **5 CONSIDERAÇÕES FINAIS**

Durante a pesquisa realizada, foram levantados questionamentos acerca do papel das plataformas de IA no gerenciamento de Arquivos Digitais. Tal como, analisada a aplicação dessas ferramentas na Arquivologia, para definir o seu potencial na gestão de documentos digitais e as principais vantagens e desvantagens de seu uso. A fim compreender melhor o potencial dessas tecnologias no contexto dos Arquivos Digitais, foram consideradas referências teóricas e estudos de caso apresentados pelo ChatGPT.

No entanto, conforme descrito, as plataformas são ferramentas auxiliares do Arquivista em suas atividades laborais, sobretudo, aquelas que apresentam um teor mais técnico, com repetições e que são perfeitamente substituídas quanto às funcionalidades no arquivo digital apresentadas neste artigo. Apesar de na atualidade a Inteligência Artificial ser vista como ameaça para algumas profissões, existem limitações, que não lhe permitem autonomia na condução de seus resultados sem uma intervenção na operacionalização.

Desta maneira, é possível perceber que a automação das plataformas de IA,

tornam-se refém da avaliação e operacionalização humana (Arquivista), no primeiro momento, com base nas informações geradas pelos algoritmos para que se alcancem os resultados esperados. Outrossim, há necessidade de ter grandes quantidades dos dados de treinamento para eficácia na classificação e análise de documentos digitais nos modelos de IA.

Todavia, há capacidade das plataformas em auxiliar na execução das atividades concernentes à Arquivistas, referentes a gestão documental do arquivo digital. Nesse sentido, foi esclarecido nas seções do artigo que detalham os resultados verificados, baseados nos filtros da Inteligência Artificial executados nas plataformas de Linguagem Naturalizadas, com enfoque maior no ChatGPT e ChatSonic.

A aplicação das ferramentas sugeridas para o gerenciamento dos Arquivos Digitais, exemplificando o seu papel no Processo Arquivístico sugerido como possível a realização pela plataforma foi verificada nesta pesquisa. Pois, conforme o objetivo da pesquisa em discutir e externalizar aos Arquivistas, a aplicabilidade da Inteligência Artificial nos Arquivos Digitais, as afirmações e conceitos acerca das plataformas selecionadas foram validados através de experimentos.

Exploramos, principalmente as competências e operacionalização das plataformas ChatSonic ChatGPT, mas, trouxemos dados filtrados nas plataformas incorporados ao estudo, a respeito das aplicações Arquivísticas da Tensor Flow, Py Torch, Scikit-learn, Keras, IBM Watson, Microsoft Azure AI e Google Cloud AI.

Considerando as limitações encontradas durante o estudo, é importante destacar alguns aspectos que merecem atenção em pesquisas futuras, como a necessidade de grande quantidade de dados para treinar os modelos de Aprendizado de Máquina e a possibilidade de vieses nos resultados, caso os dados de treinamento não sejam representativos ou diversificados o suficiente. Além disso, é importante que os dados sejam de boa qualidade e que estejam bem estruturados para que os algoritmos de IA possam interpretá-los corretamente.

A automatização nos Arquivos Digitais, portanto, torna os processos da gestão hábeis, viabilizando a economia de tempo, redução de custos e imprecisões humanas. Apesar disso, é importante garantir que os dados estejam protegidos contra acessos não autorizados e que sejam utilizadas técnicas de anonimização quando necessário, privilegiando a privacidade e a segurança dos dados durante o

processo de análise e classificação.

Ressaltando que os algoritmos de IA podem apresentar vieses e preconceitos, especialmente se os dados de treinamento forem enviesados ou seletivos, requer cautela do Arquivista a fim de moderar esses vieses garantindo uma classificação e análise documental justa e imparcial.

Consoante exposto, ainda existem desafios e limitações a serem enfrentados no que se refere às plataformas de Inteligência Artificial na Arquivologia. Dentre eles, podemos destacar a necessidade de se garantir a segurança e a privacidade dos dados presentes nos Arquivos Digitais, além da interpretação e contextualização do conteúdo. Portanto, torna-se indispensável a junção multidisciplinar entre Arquivistas, especialistas de Inteligências Artificiais e desenvolvedores de tecnologias no sentido do avanço das ferramentas relevantes na Arquivologia.

Considerando as particularidades das plataformas de Inteligência Artificial, a realização de estudos para investigar aplicações de outras plataformas na Gestão de Arquivos Digitais é considerável para enriquecimento das pesquisas neste campo tecnológico aliado à arquivística. Como, a utilização de algoritmos de redes neurais no reconhecimento de padrões em documentos digitalizados, facultando possibilidades práticas a se desenvolver a partir deste trabalho, no desenvolvimento de ferramentas específicas que atendam às demandas de Arquivistas na gestão de Arquivos Digitais.



## REFERÊNCIAS

BRASIL. **Lei no. 8.159, de 8 de janeiro de 1991**. Dispõe sobre a política nacional de arquivos públicos e privados e dá outras providências. Diário Oficial da União, Brasília, DF, 09 jan. 1991. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/LEIS/L815-9.htm](https://www.planalto.gov.br/ccivil_03/LEIS/L815-9.htm). Acesso em: 17 fev. 2023.

ChatSonic (2023). **Plataforma de Conversação com Inteligência Artificial** (versão de 12 de junho). Disponível em: <https://app.writesonic.com/pt-pt/template/dfe7a587-bd8f-4d6a-8a9c-74e33de774b7/chatsonic/76af4ad7-4e85-4f09-a7b0-f75810a859dd>. Acesso em: 12 jun. 2023.

COELHO, Raquel Luise. Pret. **Do quê se trata?** Documento, indexação e produção de sentido nos arquivos de Universidade. 2019. 277f. Tese (Doutorado em Ciência da Informação) - Universidade Federal Fluminense, Niterói, 2019. Disponível em: <https://app.uff.br/riuff/handle/1/13846>. Acesso em: 08 mar. 2023.

CONARQ. CONSELHO NACIONAL DE ARQUIVOS. **NOBRADE**: Norma Brasileira de Descrição Arquivística. Rio de Janeiro: Arquivo Nacional, 2006. Disponível em: <https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/nobrade.pdf>. Acesso em: 07 jun. 2023.

FINGER, Marcelo. Inteligência Artificial e os rumos do processamento do português brasileiro. **Estudos Avançados**, [S. l.], v. 35, n. 101, p. 51-72, 2021. Disponível em: <https://www.revistas.usp.br/eav/article/view/185030>. Acesso em: 09 jun. 2023.

GAVA, Tânia Barbosa Salles.; FLORES, Daniel. Problematizando a Pós-Custódia com a contemporaneidade da Cadeia de Custódia Digital Arquivística compartilhada e distribuída na Preservação Digital Sistêmica. **InCID: Revista de Ciência da Informação e Documentação**, [S. l.], v. 13, n. 2, p. 222-243, 2022. Disponível em: <https://www.revistas.usp.br/incid/article/view/191654>. Acesso em: 25 mar. 2023.

GAVA, Tânia Barbosa Salles; FLORES, Daniel. Repositórios arquivísticos digitais confiáveis (RDC-Arq) como plataforma de preservação digital em um ambiente de gestão arquivística. **Informação & Informação**, Londrina, v. 25, n. 2, p. 74-99, abr./jul. 2020. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/38411>. Acesso em: 28 mar. 2023.

Google Cloud AI (2023). **Produtos de IA e machine learning** (versão de 19 de junho). Disponível em: <https://cloud.google.com/why-google-cloud?hl=pt-br>. Acesso em: 19 jun. 2023.

IBM Watson (2023). **AI de Código Aberto** (versão de 15 de junho). Disponível em: <https://www.ibm.com/>. Acesso em: 15 jun. 2023.

IOSCOTE, Fabia Cristiane. Jornalismo e inteligência artificial: tendências nas pesquisas brasileiras entre 2010 e 2020. **Novos Olhares**, [S. l.], v. 10, n. 2, p. 162-182, 2021. Disponível em: <https://www.revistas.usp.br/novosolhares/article/view/188912>. Acesso em: 09 jun. 2023.

Keras (2023). **Deep Learning for humans** (versão de 15 de junho). Disponível em: <https://keras.io/>. Acesso em: 15 jun. 2023.



MAGALHÃES, Lúcia Helena de. **Agrupamento automático de notícias de jornais on-line usando técnicas de machine learning para clustering de textos no idioma português**. 2020. 188f. Tese (Doutorado em Ciência da Informação) - Universidade Federal de Minas Gerais, Belo Horizonte, 2020. Disponível em: <http://hdl.handle.net/1843/37525>. Acesso em: 10 jun. 2023.

Microsoft Azure AI (2023). **Azure Machine Learning** (versão de 18 de junho). Disponível em: <https://azure.microsoft.com/en-us/solutions/ai/#benefits>. Acesso em: 18 jun. 2023.

MONEUS, Ahmed Mohammed; SAHARI, Yousef Mohammed. Artificial Intelligence and Human Translation: A Contrastive Study Based on Legal Texts. **Heliyon**, Saudi Arabia, 2023, p. 01-13. Disponível em: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4441379](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4441379). Acesso em: 9 mar. 2023.

OpenAI (2021). **ChatGPT** (versão de 03 de junho). Inteligência Artificial na gestão de arquivos digitais. Disponível em: <https://chat.openai.com/chat>. Acesso em: 03 mar. 2023.

PyTorch (2023). **Estrutura de aprendizado de máquina de ponta a ponta** (versão de 18 de junho). Disponível em: <https://pytorch.org/features/>. Acesso em: 18 jun. 2023.

ROCKEMBACH, Moisés. Ciência da informação e inteligência artificial: um caminho para arquivos e bibliotecas inteligentes. *In*: CONGRESSO ISKO ESPANHA-PORTUGAL, 5., 2021, Lisboa). **Atas** [...]. Lisboa: Universidade de Lisboa, 2021, p. 235-242. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/233477/001134518.pdf?sequence=1>. Acesso em: 29 jan. 2023.

SANTOS, Henrique Machado dos; FLORES, Daniel. Preservação sistêmica para repositórios arquivísticos. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, [S. l.], v. 14, n. 3, 2020. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/2089>. Acesso em: 09 mai. 2023.

SILVA, Manuela do Nascimento; PEREIRA, Charlienes Francisca; SOARES, Pedro Unaldo Oliveira; SANTOS, Thaynara Costa dos; PAIVA, Tuanny Bezerra. Classificação documental. **Biblionline**, v. 16, n. 2, p. 41-54, 2020. Disponível em: <https://brapci.inf.br/index.php/res/v/149427>. Acesso em: 10 jun. 2023.

Scikit-learn (2023). **Machine learning in Python** (versão de 18 de junho). Disponível em: <https://scikit-learn.org/stable/>. Acesso em: 18 jun. 2023.

TensorFlow (2023). **Plataforma completa de machine learning** (versão de 15 de junho). Disponível em: <https://www.tensorflow.org/?hl=pt-br>. Acesso em: 15 jun. 2023.

TOH, Michelle. Inteligência artificial pode afetar 300 milhões de empregos no mundo, diz Goldman Sachs. **CNN Brasil**, São Paulo, 29 mar. 2023. Disponível em: <https://www.cnnbrasil.com.br/economia/inteligencia-artificial-pode-afetar-300-milhoes-de-empregos-no-mundo-diz-goldman-sachs/>. Acesso em: 04 abr. 2023.