



25°enancib

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

3 a 7 de novembro de 2025 – Rio de Janeiro



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

GT 8 – Dados, Informação e Tecnologia

INTELIGÊNCIA ARTIFICIAL NA INDEXAÇÃO DE MANUSCRITOS DO ARQUIVO PÚBLICO DA PARAÍBA

ARTIFICIAL INTELLIGENCE IN INDEXING MANUSCRIPTS FROM THE PUBLIC ARCHIVES OF PARAÍBA

Tayná Ladislau Rangel – Universidade Federal da Paraíba (UFPB)

Virginia Bentes Pinto – Universidade Federal do Ceará (UFC)

Hilza Costa Cavalcante – Universidade Estadual da Paraíba (UEPB)

Maria Meriane Vieira da Rocha – Universidade Federal da Paraíba (UFPB)

Modalidade: Trabalho Completo

Resumo: embora esteja se ubiquando neste Século XXI, o termo Inteligência Artificial (IA) tem sua gênese em 1943, originário de pesquisas envolvendo aplicabilidades de modelos matemáticos lógicos em neurônios artificiais e cujo impacto se deu desde 2010, em razão da aprendizagem de máquina. Levando isso em consideração, este estudo se norteia pela seguinte questão de pesquisa: de que modo a IA pode contribuir para a representação indexical de documentos coloniais manuscritos? Para o desenvolvimento desta pesquisa, definiu-se o seguinte objetivo: analisar a contribuição da IA para a representação indexical de manuscritos referentes ao período colonial do estado da Paraíba. Trata-se de uma pesquisa quanti-qualitativa, adotando a análise documental para a representação de documentos manuscritos do período colonial. A empiria foi realizada no acervo do Arquivo Público do Estado da Paraíba (APEPB). Para tanto, digitalizaram-se 05 documentos, salvos em PDF/A, utilizando o OCR. Posteriormente, aplicou-se a IA, especificamente a plataforma Gemini, versão livre. Para a indexação, adotaram-se os seguintes *Prompts* de Comando: a transcrição do texto contido na imagem de forma precisa e clara; a identificação e organização das informações-chave para criar uma estrutura que facilite a compreensão; a análise visual do texto para destacar partes relevantes; e a elaboração de um índice para resumir os principais tópicos e informações, permitindo uma navegação fácil e rápida. Como considerações, percebeu-se que, embora a IA agilize a transcrição, sua precisão limitada exige revisão humana. Ademais, a união entre a tecnologia e a análise crítica é essencial para garantir exatidão para indexação da informação.

Palavras-chave: inteligência artificial; indexação de manuscritos com IA; manuscritos coloniais – Paraíba; manuscritos.

Abstract: although it is becoming increasingly widespread in the 21st century, the term Artificial Intelligence (AI) originated in 1943, originating from research involving the applicability of logical mathematical models to artificial neurons. Its impact has been felt since 2010, driven by machine learning. Considering this, this study is guided by the following research question: how can AI contribute to the indexical representation of handwritten colonial documents? The objective of this research was to analyze the contribution of AI to the indexical representation of manuscripts from the colonial period in the state of Paraíba. This is a quantitative and qualitative study, adopting document analysis for the representation of handwritten documents from the colonial period. The empirical

study was conducted in the collection of the Public Archives of the State of Paraíba (APEPB). Five documents, saved in PDF/A format, were digitized using OCR. Subsequently, AI, specifically the free version of the Gemini platform, was applied. For indexing, the following command prompts were adopted: accurate and clear transcription of the text contained in the image; identification and organization of key information to create a structure that facilitates comprehension; visual analysis of the text to highlight relevant sections; and creation of an index to summarize the main topics and information, allowing for quick and easy navigation. As considerations, it was noted that, although AI speeds up transcription, its limited accuracy requires human review. Furthermore, the combination of technology and critical analysis is essential to ensure accuracy for information indexing.

Keywords: artificial intelligence; AI-powered manuscript indexing; colonial manuscripts – Paraíba; manuscripts.

1 INTRODUÇÃO

A crescente relevância da inteligência artificial (IA) no contexto contemporâneo tem gerado discussões, particularmente, na Ciência da Informação (CI). O uso de IA para a indexação e representação de temáticas tratadas em documentos, especialmente aqueles com guarda permanente, como os manuscritos coloniais, é uma inovação que pode transformar a maneira como esses materiais são acessados e usados. Diante desse contexto, este estudo é pautado na seguinte questão de pesquisa: de que modo a Inteligência Artificial (IA) pode contribuir para a representação indexical de manuscritos coloniais? Os manuscritos coloniais desempenham um papel crucial na construção da memória e identidade cultural de um povo, atuando como fontes primárias que oferecem percepções sobre as práticas sociais, econômicas e políticas da época. Contudo, a preservação desses documentos enfrenta desafios significativos, como a deterioração física e a dificuldade de interpretação devido à linguagem e à escrita paleográfica.

Nesse cenário, tecnologias como o Reconhecimento Óptico de Caracteres (OCR) e, mais recentemente, o Reconhecimento de Texto Manuscrito (HTR) surgem como uma inovação com potencial para transformar a descrição arquivística. Segundo Dias *et al.* (2025, p. 481), a Inteligência Artificial engloba um conjunto de conceitos que viabiliza a execução de tarefas complexas por máquinas, resultando na redução de custos e no aumento da eficiência de produtos e processos.

Para avaliar a eficácia desses sistemas, é recomendado o uso de métricas quantitativas. A mais comum é a Taxa de Erro de Caractere (CER), que mede a distância de edição entre a transcrição gerada pela IA e uma transcrição de referência. A otimização econômica dos processos, a diminuição de atividades laborais repetitivas e a capacidade de tomar decisões

mais diretas são consequências importantes do impacto gerado por essas tecnologias (Dias *et al.*, 2025, p. 482).

Diante disso, a busca por novos métodos para indexação e recuperação de informações se torna premente, impulsionando a pesquisa em IA. Nessa perspectiva, definiu-se o seguinte objetivo geral: analisar a contribuição da IA para a representação indexical de manuscritos referentes ao período colonial custodiados pelo Arquivo Público do Estado da Paraíba (APEPB). Os objetivos específicos são: selecionar, aleatoriamente, cinco (5) manuscritos do período colonial da história da Paraíba; indexar os referidos documentos com a ferramenta Gemini na transcrição automatizada de manuscritos coloniais do APEPB; avaliar a capacidade e as limitações da IA no processo de identificação, organização de informações-chave e criação de índices.

Esta pesquisa busca contribuir para o campo da CI, como também para a preservação do patrimônio histórico. Ao analisar a contribuição da IA na indexação de manuscritos coloniais, espera-se oferecer possibilidades aplicáveis em outras áreas, promovendo maior eficiência na gestão de documentos permanentes. A pesquisa, assim, está inserida em um contexto mais amplo de valorização da informação pública, essencial para a promoção da cidadania e da transparência em instituições públicas.

2 INTELIGÊNCIA ARTIFICIAL NA REPRESENTAÇÃO INDEXICAL DE DOCUMENTOS

O sintagma Inteligência Artificial surge no *workshop do Dartmouth College Dartmouth*, ocorrido em 1956, nos Estados Unidos, quando John McCarthy cunha o termo IA como “*field of research, aimed to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans and improve themselves*”. O destaque onipresente se concretiza em 2010, em razão da chamada aprendizagem de máquina, que possibilitou uma alavancagem no uso da IA em inúmeras áreas, como na pesquisa científica, nos setores de serviços ou na indústria, a exemplo da representação indexical de documentos verbais ou não verbais.

Contudo, Nilsson (2005, p. 68) argumenta que “o objetivo científico de longo prazo para muitos pesquisadores de inteligência artificial (IA) continua sendo a mecanização da inteligência de ‘nível humano’, mesmo que alcançar esse objetivo possa levar muitos anos”. Contudo, é fundamental compreender, nessa perspectiva da IA, que “máquinas tão

inteligentes quanto os humanos devem ser capazes de fazer a maioria das coisas que os humanos fazem.”

O *Institut National de Recherche en Sciences et Technologies du Numérique* (INRIA)¹ classifica a IA nos seguintes subdomínios: “conhecimento, aprendizagem de máquina, processamento de linguagem natural, processamento de sinais, neurociência robótica, ciência cognitiva, algoritmos de IA e suporte à decisão” (Inria, 2019, p. 20-21, tradução nossa). Esses conceitos e subdomínios evidenciam que a IA também tem relação com a Ciência da Informação, a Biblioteconomia, a Arquivologia e a Museologia, particularmente em relação ao conhecimento e ao tratamento da linguagem natural, no contexto da representação indexical (temática) e descritiva, da gestão da informação e do conhecimento, fontes e serviços de informação, da pesquisa documental e bibliográfica, da pesquisa científica, da padronização de normas técnicas, mediação e práticas informacionais etc.

A pragmática indexical, nos campos da Ciência da Informação, Biblioteconomia, Arquivologia e Museologia, e independente de abordagem manual, semiautomática ou automática, perpassa necessariamente por uma série de etapas. Inicialmente, uma análise com uma leitura bem atenta para identificar as temáticas tratadas em um documento, fazendo anotações. Em seguida, uma seleção das palavras consideradas fundamentais para os objetivos da instituição ou dos seus clientes reais ou em potencial. Na terceira fase, executa-se a tradução, no caso de vocabulários controlados. Uma quarta etapa se constitui como um controle de qualidade. No contexto arquivístico, a "representação indexical" é um processo complexo que visa a identificar as temáticas e os dados essenciais de um documento para facilitar sua recuperação. Tradicionalmente, esta atividade manual envolve análise, seleção de termos e tradução para uma linguagem controlada. A IA promete automatizar e acelerar essas etapas. No âmbito desta pesquisa, a representação indexical é compreendida como um conjunto de tarefas que inclui: a transcrição do texto manuscrito, a indexação descritiva (identificação de informações-chave como nomes, datas e locais) e a indexação temática (criação de um índice de assuntos).

No caso de representação indexical automática com a IA, tais fases se resumem somente a uma. Independentemente de indexação manual, semiautomática ou automática

¹ “Connaissances, apprentissage automatique, traitement du langage naturel, traitement des signaux, robotique neurosciences, sciences cognitives, e algorithmique de l’IA et aide à la décision” (Inria, 2019, p. 20-21, tradução nossa).

feita pela IA, assim como ocorre no caso manual, também pode haver resultados positivos ou negativos. Isso está condicionado às temáticas dos documentos, à sua redação e ao seu formato ser analógico (manuscritos, digitados, datilografados) ou digital. Logo, a qualidade dessa representação indexical não está isenta da intervenção humana.

3 CONSIDERAÇÕES SOBRE O ARQUIVO PÚBLICO DO ESTADO DA PARAÍBA

A institucionalização dos primeiros arquivos públicos no Brasil remonta à primeira década do Século XIX. Contudo, ela não foi uniforme. Nem todos foram estabelecidos por lei específica ou concebidos com o *status* de instituições arquivísticas desde o início. Conforme detalha Côrtes (1996, p. 73-4), “a criação dessas instituições se deu através de leis, decretos ou resoluções que estabeleciam suas funções e objetivos, voltados geralmente para garantir o uso acadêmico e histórico de documentos sob a sua guarda”. Tal arquivo é chamado “Núcleo de Documentação e Pesquisa”. Ainda nesse interim, chama atenção o fato de “os Arquivos do Ceará (1916), do Espírito Santo (1908) e do Pará (1901), foram criados como anexos à biblioteca pública e o Arquivo do Estado do Rio Grande do Sul (1954) como uma extensão do museu”. Isso pode ser decorrente “da fragilidade da diferenciação do que se constituía naquele momento um acervo de biblioteca ou museu e um acervo específico de arquivo, ou mesmo a falta de recursos para se criar instituições distintas”.

Ainda que a pesquisa de Côrtes (1996) mencione a existência de um arquivo público na Paraíba com data de criação incerta, é importante distinguir essa referência da atual estrutura institucional. O “Núcleo de Documentação e Pesquisa”, por exemplo, possui um arquivo pertencente à Universidade Federal da Paraíba e não se configura como o arquivo estadual. Observando tal fato, o Governo do Estado, por meio da Secretaria de Estado do Governo (SEGOV), instituiu, formalmente, pela Lei n.º 11.263, de 29 de dezembro de 2018, o Arquivo Público do Estado da Paraíba (APEPB), um dos mais jovens arquivos legalmente criados em âmbito estadual no Brasil. Conforme o artigo 47 da referida lei, o APEPB incorporou, desde o início, o acervo que estava sob a guarda do Arquivo Histórico Waldemar Bispo Duarte, o qual era, até 2018, subordinado à Secretaria de Estado da Educação. Dessa forma, o APEPB nasce com a responsabilidade fundamental de elaborar políticas arquivísticas para o poder executivo estadual, além de tratar, preservar e garantir o acesso ao acervo permanente que lhe foi confiado.

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

O conjunto documental do APEPB é constituído por livros de sesmarias e registros de terras, além de manuscritos avulsos correspondentes aos períodos Colonial, Imperial e da Primeira República na Paraíba. O referido acervo também é composto por exemplares encadernados de jornais, como A União e O Norte. Atualmente, o APEPB totaliza 768 caixas de documentos, o que corresponde a aproximadamente 208,24 metros lineares. O documento, mais antigo datado de 1704, é um livro de sesmarias; já o mais recente, datado de 1978, corresponde ao período da Primeira República na Paraíba.

Após a incorporação do acervo mencionado ao APEPB, houve um processo de reordenação. Nessa fase, foi identificado cada documento. Verificou-se que alguns já possuíam transcrições manuscritas existentes no próprio acervo, as quais foram então reunidas aos seus respectivos originais. Em seguida, foram preenchidas fichas detalhando a tipologia, o resumo do conteúdo e o estado físico de cada documento. Subsequentemente, cada documento recebeu uma numeração de ordem e, posteriormente, foi acondicionado em novas caixas, utilizando capilhas e cintas confeccionadas em papel neutro para garantir sua preservação. Finalizado o processo de reordenação, identificou-se que o APEPB possui sete caixas de manuscritos avulsos correspondentes ao período colonial. Após a reorganização, elaborou-se um breve resumo descritivo para cada documento, constando informações como data e local de produção, tipologia, assunto, número de localização no APEPB e o código BR PBAPEPB (Brasil; Paraíba; APEPB). Adicionalmente, uma parceria com um projeto de pesquisa da Universidade Estadual da Paraíba, contando com a colaboração de bolsistas e estagiários, resultou na produção de 540 verbetes que cobrem todo o conjunto documental da fase colonial no acervo do APEPB. Em 2024, esses esforços culminaram na publicação de um catálogo (Melo, 2024) em formato de livro, uma colaboração entre o APEPB e a Empresa Paraibana de Comunicação (EPC).

Os manuscritos coloniais da Paraíba possuem características comuns, como papel amarelado pelo tempo, marcas d'água visíveis quando colocados contra a luz, escrita em tinta ferrogálica e com caligrafia paleográfica típica do período. Percebe-se também a presença de fissuras, manchas e outros sinais de deterioração, resultantes tanto do envelhecimento natural quanto de condições inadequadas de acondicionamento ao longo dos anos.

4 PERCURSO METODOLÓGICO

Trata-se de uma pesquisa descritiva, quanti-qualitativa e de natureza aplicada, haja vista que o propósito de fazer experimentos com a IA para indexar, com mais rapidez, a quantidade de manuscritos do APEPB, de modo a acelerar tal processo. Quantitativamente, a transcrição foi medida pela Taxa de Erro de Caractere (CER). O cálculo foi feito comparando a transcrição do Gemini com a transcrição manual, utilizando a seguinte fórmula: $CER = (S + I + D) / N$, sendo S as substituições, I as inserções, D as deleções e N o número total de caracteres no texto de referência. Concernente aos aspectos descritivos, destacam-se as características, os padrões e as especificidades presentes na representação indexical dos manuscritos coloniais feita pela IA.

A empiria se deu no APEPB. Foi definido um recorte temporal entre 1755 e 1822, datas correlacionadas ao conjunto documental de manuscritos avulsos do período colonial que integram seu acervo. Trata-se de documentos com escrita paleográfica, cuja leitura exige conhecimento especializado, embora a maioria possua transcrição manual do texto original. Ou seja, documentação direta. Conforme Lakatos e Marconi (2019, p. 202), “a documentação direta constitui-se, em geral, no levantamento de dados no próprio local onde os fenômenos ocorrem”.

Para a representação indexical, utilizamos a ferramenta Gemini do Google. De acordo com Carraro (2024), o Gemini, lançado em 2024, representa um avanço significativo nas capacidades dos sistemas de IA da empresa, superando até mesmo especialistas humanos em vários *benchmarks* de avaliação. Para tanto, aplicamos os seguintes *prompts* específicos: **Transcrição** – Transcreva o texto contido na imagem de forma precisa e clara; **Indexação** – Identifique e organize as informações-chave da imagem, criando uma estrutura que facilite a compreensão dos dados apresentados; **Interpretação visual** – Analise visualmente o texto e destaque as partes mais relevantes, enfatizando as informações críticas que devem ser transcritas; **Criação de índice** – Com base no conteúdo tratado na imagem, elabore um índice que resuma os principais tópicos e informações, permitindo uma navegação fácil e rápida.

5 RESULTADOS, ANÁLISE E DISCUSSÃO DOS DADOS

Foi realizada uma análise detalhada de documentos manuscritos do Período Colonial da Paraíba (Séculos XVI a XIX), utilizando o Gemini, por meio de arquivo em PDF/A com OCR

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

(Optical Character Recognition)², para identificar como a inteligência artificial (IA) poderia contribuir no processo de indexação de manuscritos, ambos de guarda permanente do APEPB. A análise dos cinco documentos revelou tanto o potencial quanto as limitações do Gemini na indexação de manuscritos coloniais. A análise quantitativa da transcrição mostrou que a precisão variou consideravelmente entre os documentos, como demonstrado pela Taxa de Erro de Caractere (CER) no Quadro 1.

Quadro 1 – Desempenho da Transcrição do Gemini

Documento	Caracteres (N)	Erros (S+D+I)	CER (%) ³	Observações
Doc 02	388	196	50,5%	Mancha de umidade na margem; caligrafia cursiva com ligaduras.
Doc 03	923	296	32,1%	Manchas de umidade no topo ofuscando partes do texto.
Doc 09	777	473	60,9%	Caligrafia com floreios e abreviações; mancha de umidade.
Doc 11	575	478	83,1%	Caligrafia muito estilizada com baixo contraste da tinta.
Doc 537	2.056	1.618	78,7%	Texto denso com espaçamento reduzido e tinta desbotada.
Média:			61,1%	

Fonte: elaborado pelas autoras (2025).

A análise revela uma Taxa de Erro de Caractere (CER) média de 61,1% para o conjunto de documentos. O desempenho variou significativamente entre os documentos, com a CER mais baixa registrada no Documento 03 (32,1%) e a mais alta no Documento 11 (83,1%). Os resultados indicam um desafio considerável na transcrição automática dos manuscritos apresentados, resultando em um número elevado de erros. Embora a IA consiga capturar a essência do texto, a transcrição contém um volume significativo de erros que impedem seu uso direto sem revisão.

Os *prompts* utilizados para a análise incluem: a transcrição do texto contido na imagem de forma precisa e clara; a identificação e organização das informações-chave para criar uma estrutura que facilite a compreensão; a análise visual do texto para destacar partes relevantes; e a elaboração de um índice que resuma os principais tópicos e informações, permitindo uma navegação fácil e rápida.

² OCR - Reconhecimento de Caractere Óptico: é o processo que converte uma imagem de texto em um formato de texto legível por máquina.

³ Utilizou-se a seguinte fórmula: $CER = (S + I + D) / N$; **S**: Número de Substituições (caracteres trocados), **I**: Número de Inserções (caracteres a mais na transcrição do Gemini), **D**: Número de Deleções (caracteres que faltam na transcrição do Gemini), **N**: Número total de caracteres no texto de referência (Ground Truth).

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

Quadro 2 – Análise das respostas aos *prompts*, solicitados ao Gemini.

MANUSCRITO	1. TRANSCRIÇÃO (TEXTO LEGÍVEL PRESENTE NA IMAGEM)	2. INDEXAÇÃO DAS INFORMAÇÕES DA IMAGEM	3. INTERPRETAÇÃO VISUAL (ANÁLISE DAS PARTES RELEVANTES)	4. ÍNDICE
DOCUMENTO 02: CARTA	"Remeto o Mapa incluso, ppela formalidade do qual se [?] que V.[Ex.a?] mande, com a possível [brevidade?] destricto d'esta [Capitania?], tirar outro r[es]pe[c]tivo às [partes?] [?]tirado [desta Ca]pitania, e logo que [es]tiver [prompt?], sirva [V.M.?] de [evitar?] perda de tempo à Secretaria d[este] [Gov.no?] [?] remeter mandando tambem o que lhe agora vai." Fragmento com Nome do Destinatário () "Ill.mo Snr. D. Jeronimo de Mello e Castro"	Tipo de Documento: Fragmento de uma carta ou ofício. Remetente: Não identificável no fragmento. Destinatário (identificado em): Illustrissimo Senhor Dom Jeronimo de Mello e Castro. Assuntos Principais (inferidos do fragmento): • Envio de um mapa: "Remeto o Mapa incluso,". • Solicitação para elaboração de um novo mapa: "...tirar outro r[es]pe[c]tivo...", referente a um "destricto d'esta [Capitania?]". • Urgência: Menção à necessidade de evitar "perda de tempo à Secretaria". • Envio de itens adicionais: "...mandando tambem o que lhe agora vai.".	Início da Comunicação (): A frase "Remeto o Mapa incluso," é a parte mais clara e indica a ação primária de enviar um mapa. Natureza da Solicitação (): Expressões como "tirar outro" e "destricto d'esta [Capitania?]" sugerem um pedido para a criação de um novo mapa. Menção à Burocracia e Tempo (): A referência a "perda de tempo à Secretaria" destaca a preocupação com a celeridade processual, comum em correspondências oficiais. Identificação do Destinatário (): O fragmento "Ill.mo Snr. D. Jeronimo de Mello e Castro" é crucial para identificar a quem o documento provavelmente se destinava.	1. Identificação do Documento 1.1. Tipo: Fragmento de Carta/Ofício. 1.2. Destinatário Provável: Dom Jeronimo de Mello e Castro. 2. Conteúdo Principal Inferido 2.1. Ação Principal: Envio de um mapa ("Mapa incluso"). 2.2. Solicitação associada: Pedido para elaboração de um novo mapa, possivelmente de um distrito específico da Capitania. 2.3. Contexto Administrativo: Preocupação com a eficiência e prazos da Secretaria. 2.4. Envio Complementar: Indicação de que outros itens acompanham a correspondência.

Fonte: elaborado pelas autoras (2025)

O Quadro 2 revelou uma estrutura detalhada. A transcrição do texto presente nas imagens foi realizada rapidamente. Entretanto, algumas informações importantes não foram preservadas ou não foram apresentadas de forma legível. A identificação e organização das informações-chave demonstraram uma capacidade eficaz de segmentar os dados em categorias, facilitando a navegação e a compreensão. Além disso, a análise visual destacou

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

partes relevantes do texto, enfatizando elementos gráficos e tipográficos que fornecem contexto adicional sobre a origem e a importância dos documentos. A elaboração de um índice com base no conteúdo analisado permitiu um acesso mais ágil às informações, promovendo uma melhor organização do conhecimento, conforme os resultados no Quadro 3.

Quadro 3 – Análise Comparativa do Documento 02 (Carta sobre envio de mapa)

C A T E G O R I A S	Transcrição Resposta do Gemini:	"Remeto o Mapa incluso, ppela formalidade do qual se [?] que V.[Ex.a?] man- de, com a possivel [brevidade?] destricto d'esta [Capitania?], tirar outro r[es]pe[c]tivo..." "Remeto o Mapa incluso, pela formalidade do qual se há-de ordenar que V. Exa. mande com a possível brevidade do distrito desta Capitania tirar outro respectivo..."
	Análise Manual (Humana):	Parcialmente Correta: A IA capturou o sentido geral, mas falhou em decifrar abreviações ("V.Exa.") e palavras complexas, inserindo marcadores de incerteza ("[?]"). Indexação (Info-chave) Destinatário: Dom Jeronimo de Mello e Castro; Assunto: Envio de mapa e solicitação de novo mapa. Destinatário: Jerônimo de Melo e Castro (Secretário de Estado da Marinha e Ultramar); Assunto: Envio de mapa da Capitania da Paraíba para elaboração de um novo.
	Avaliação:	Correta, mas incompleta: Identificou corretamente as partes, mas sem o contexto histórico que o analista humano fornece (o cargo do destinatário, a localização da Capitania). Índice (Tópicos) 1. Identificação do Documento; 2. Conteúdo Principal Inferido; 2.1 Ação Principal: Envio de mapa; 2.2 Solicitação associada. - Tipo: Ofício; - Assunto: Administração colonial; Cartografia; - Entidades: Jerônimo de Melo e Castro; Secretaria de Governo. Útil, mas genérico: A estrutura do índice é lógica, mas os tópicos são inferidos diretamente do texto transcrito, sem a profundidade conceitual que um indexador humano aplicaria (ex: "Administração colonial").

Fonte: elaborado pelas autoras (2025)

A análise dos demais documentos seguiu padrão semelhante. A IA demonstrou ser eficaz na segmentação dos dados e na identificação de ações principais (comprar, vender, solicitar). Porém, apresentou dificuldade consistente para identificar nomes próprios e termos técnicos da época e interpretar documento danificado. Embora a IA agilize a primeira leitura, a revisão humana é imprescindível para corrigir erros, completar informações e, crucialmente, adicionar contexto histórico, algo inexecutável para o modelo. Isso reforça a ideia do "entrelaçamento" entre algoritmos e as práticas dos usuários. A ferramenta não substitui o especialista, mas se torna um instrumento em suas mãos, cujo resultado depende da sua capacidade de validar e refinar a informação gerada.

A precisão na transcrição é fundamental, pois erros podem comprometer a integridade das informações. A análise revelou casos em que caracteres ou palavras foram transcritos incorretamente, ressaltando a necessidade de uma revisão cuidadosa. De acordo com Bentes Pinto, Meunier e Silva Neto (2008, p. 7), "representar o seu conteúdo, permitindo, não somente o acesso durante uma busca de informação em bases de dados, mas, também que o

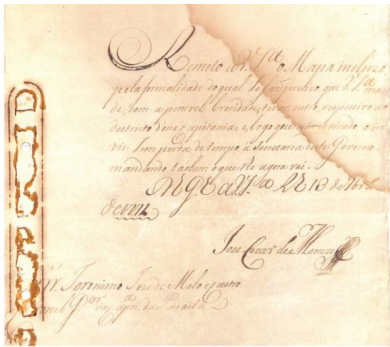
sujeito possa se deslocar sobre o documento mesmo, em sua natureza concreta, visando à sua recuperação posterior de seu conteúdo”.

A resposta foi eficaz, com categorias bem definidas que auxiliam na indexação. A estrutura proposta facilita a identificação de temas e tópicos relevantes, permitindo uma pesquisa mais eficiente. Por sua vez, a interpretação visual trouxe à tona elementos que excedem o texto, como marcas gráficas, rasuras e carimbos, os quais indicam a autenticidade e a relevância dos documentos, sendo cruciais para entender o contexto histórico. Os resultados indicam que a combinação de tecnologia com análise humana é essencial para maximizar a precisão dos dados. Embora ferramentas automatizadas, como o Gemini, ofereçam agilidade na transcrição e na organização, a revisão humana é imprescindível para garantir a precisão e a contextualização das informações. Rangel (2024) argumenta que a IA nas tecnologias não deve ser encarada como uma adversária, especialmente para os profissionais da informação. Ela ressalta a importância dessa temática e o potencial de cada área para influenciar seus contextos, colaborando assim com o trabalho tradicional desses profissionais. A utilização de métodos de análise visual complementa a transcrição textual, permitindo uma compreensão mais profunda do conteúdo.

Entretanto, a análise apresenta algumas limitações que afetam sua profundidade e validade. A falta de contextualização histórica na discussão pode levar a interpretações superficiais sobre os documentos. Além disso, a interpretação visual subestima a subjetividade envolvida, resultando em inconsistências devido a diferentes perspectivas de revisores. A metodologia enfatiza a transcrição automatizada, mas não explora suficientemente as limitações dessa abordagem, comprometendo a precisão. A omissão de erros de transcrição é outro aspecto que não foi adequadamente abordado, o que é crucial para entender a qualidade dos dados apresentados. Por último, a análise é relativamente restrita e não considera a diversidade de documentos que poderiam enriquecer a discussão.

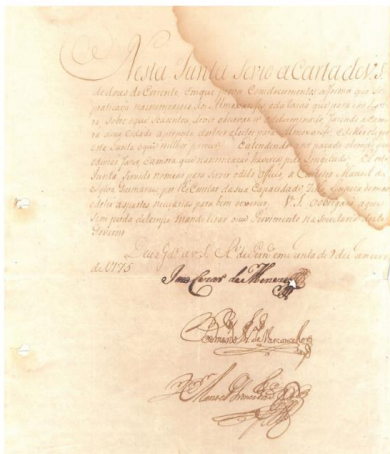
Na Figura 1, pode-se observar os manuscritos coloniais e suas transcrições via IA Gemini. Essa abordagem também expõe desafios da transcrição e a interação entre IA e análise humana.

Figura 1 – Manuscritos e Transcrições feitas pelo Gemini



"Remeto o Mapa incluso, ppela formalidade do qual se [?] que V.[Ex.a?] man- de, com a possivel [brevidade?] destricto d'esta [Capitania?], tirar outro r[es]pe[c]tivo às [partes?] [?]tirado [desta Ca]pitania, e logo que [es]tiver [prompto?], sirva [V.M.?] de [evitar?] perda de tempo à Secretaria d[este] [Gov.no?] [?] remeter mandando tambem o que lhe agora vai."

"Ill.mo Snr. D. Jeronimo de Mello e Castro"



[Fonte 1] Esta Junta. Seria a carta de V.S.^a de dous do Corrente, em que V. Mercê com documento afirma que se praticais nas nomeações dos Almojarifes e da razão que para isso há--de haver, sobre o que se assentou, levar observar a lei terminal e fazendo a Ilm.^a dita Cidade a proposta de antes eleitos para Almojarife e deles eleger esta Junta o que melhor parecer. E atendendo a ser passado o tempo em que poderia esta Câmara, que nas nomeações haveria pela singeleza de entender Junta Servida nomear para servir os ditos officos, a Custódio Manuel de Sylva Guimarães, por ter comitiva da sua Capacidade, Lealdade e temperança e demais partes necessárias par bem os servir.

[Fonte 2] Obrigará a que Item por ela se lhe mande tirar o Provimento na forma do Regimento.

Leiria (?) de Novembro (?) de mil e setecentos e cinquenta e ___ (175_)

Manuel Severim de Faria (?) Governador (?) da Capitania (?) (Assinatura ilegível ou nome parcial "Beneze"?)

Fonte: elaborada pelas autoras (2025)

Em suma, embora os resultados apresentem uma estrutura clara e organizada, a análise e discussão carecem de profundidade crítica e contextualização histórica. A análise quantitativa revelou uma Taxa de Erro de Caractere (CER) média de 61,1%, um valor que, embora demonstre a capacidade de leitura da IA, é muito alto para permitir o uso autônomo e fidedigno das transcrições. A qualidade decai significativamente em documentos com caligrafia complexa ou danos físicos, ressaltando que a tecnologia ainda não supera a expertise paleográfica humana. Para fortalecer a pesquisa, seria benéfico integrar uma abordagem mais holística que considere a complexidade dos documentos analisados, além de uma avaliação crítica das tecnologias utilizadas. Isso porque, se houver falhas dos usuários ou se eles "se recusarem a se encaixar nas suas práticas e a conceder significado a elas, o algoritmo irá falhar". Por isso, "não devemos considerar seus 'efeitos' nas pessoas, mas um 'entrelaçamento' multidimensional entre algoritmos postos em prática e as táticas dos usuários que fazem face a eles" (Gillespie, 2018, p. 110).

Para uma análise robusta de erros de transcrição, podem ser adotadas várias abordagens. É crucial categorizá-los (ortográficos, omissões) e medir sua frequência de erros

para achar padrões. A revisão deve se dar por múltiplos especialistas, para comparar transcrições e revelar divergências. É ainda necessário implementar um sistema de *feedback*, permitindo que os usuários reportem falhas. Além disso, a análise de casos específicos, explorando as causas subjacentes que levaram a erros, pode fornecer novas percepções. Outra abordagem envolve o treinamento de modelos de IA utilizando dados sobre erros para aprimorar algoritmos de reconhecimento de texto. Criar relatórios que documentem os tipos e a frequência de erros encontrados proporcionará uma visão clara sobre as áreas que necessitam de mais atenção. É importante ressaltar que uma IA tem a capacidade de “gerar novas informações ao descobrir tendências e padrões relevantes em informações já coletadas” (Gill; Kaur, 2023, p. 262). Por fim, comparar transcrições automáticas com aquelas feitas manualmente permitirá avaliar a precisão e identificar padrões de erro. Essas abordagens proporcionarão uma análise mais detalhada e sistemática, contribuindo para a qualidade e confiabilidade das transcrições em projetos futuros.

6 CONSIDERAÇÕES FINAIS

Retornando à questão de pesquisa, esta investigação demonstrou que a Inteligência Artificial, por meio do Gemini, oferece uma contribuição parcial, porém valiosa, para a representação indexical de manuscritos coloniais.

A principal contribuição da ferramenta não reside na substituição do humano, mas na otimização do fluxo de trabalho. A IA pode ser empregada como um poderoso recurso de pré-processamento ou pré-indexação. Ela gera rapidamente uma primeira versão da transcrição e da estrutura do documento, que, apesar dos erros, serve como um rascunho para o profissional. Este processo pode reduzir consideravelmente o tempo dedicado à digitação inicial, permitindo que o especialista foque em tarefas de maior valor agregado: a correção, a análise crítica, a contextualização histórica e a atribuição de termos de indexação precisos. No contexto da Ciência da Informação, da Biblioteconomia, da Arquivologia e da Museologia, são necessários mais experimentos em outros manuscritos, a exemplo de prontuários do paciente, registros de nascimento, batistério etc.

Com base nisso, recomendamos as seguintes diretrizes práticas para arquivos que considerem usar IA em acervos semelhantes: adotar um fluxo de trabalho híbrido, utilizar a IA para gerar a primeira versão da transcrição e deixar a validação, correção e enriquecimento para o especialista; priorizar acervos, a ferramenta pode ser usada para criar descrições

mínimas em larga escala para fundos documentais ainda não tratados, facilitando uma triagem inicial e a priorização de conjuntos para um tratamento mais aprofundado; e não confiar incondicionalmente, a validação humana é inegociável para garantir a precisão e a integridade da informação histórica disponibilizada ao público. Futuros experimentos devem ampliar a amostra e testar diferentes modelos de IA, incluindo aqueles especificamente treinados para HTR, a fim de aprimorar a representação e democratizar o acesso ao vasto patrimônio documental.

REFERÊNCIAS

BENTES PINTO, Virginia; MEUNIER, Jean-Guy; SILVA NETO, Casemiro. A contribuição peirciana para a representação indexical de imagens visuais. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis/SC, Brasil, v. 13, n. 25, p. 15–35, 2008. DOI: 10.5007/1518-2924.2008v13n25p15. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2008v13n25p15>. Acesso em: 20 maio 2025.

CARRARO, Fabrício. O que é o Google Gemini e o que esse modelo de IA é capaz de fazer — com exemplo prático. **Alura**, 23 abr. 2024. Disponível em: https://www.alura.com.br/artigos/google-gemini?srsltid=AfmBOoq8Nxge3FTFi_CYGfYM3q9OinGCzDH6e1Zf1WiPciHndQCsmL3q. Acesso em: 22 maio 2025.

CÔRTEZ, Maria Regina Persechini Armond. **Arquivo Público e Informação: Acesso à informação nos Arquivos Públicos Estaduais do Brasil**. 1996. 135 f. Dissertação (Mestrado em Ciência da Informação) – Escola Biblioteconomia, Universidade Federal de Minas Gerais, Belo Horizonte, 1996. Disponível em: https://repositorio.ufmg.br/bitstream/1843/BUOS-96UPHB/1/dissertacao_mariaregina.pdf. Acesso em: 22 maio 2025.

DIAS, Fernando Xavier; FUJITA, Allynson Takehiro; MEIRELES, Eduardo; BUENO, Miriam Pinheiro. Prospecção Tecnológica do Reconhecimento Óptico de Caracteres na Automação de Processos 4.0. **Cadernos de Prospecção**, [s. l.], v. 18, n. 2, p. 480–492, 2025. DOI: 10.9771/cp.v18i2.62823. Disponível em: <https://periodicos.ufba.br/index.php/nit/article/view/62823>. Acesso em: 09 ago. 2025.

GILLESPIE, Tarleton. A relevância dos algoritmos. **Revista Parágrafo**, São Paulo, Brasil, v. 6, n. 1, p. 95-121, jan./abr. 2018. Disponível em: <http://revistaseletronicas.fiamfaam.br/index.php/recicofi/article/view/722/563>. Acesso em: 20 maio 2025.

GILL, Sukhpal Singh; KAUR, Rupinder. ChatGPT: Vision and challenges. **Internet of Things and Cyber-Physical Systems**, [s. l.], v. 3, p.262-271, 2023. Disponível em: <https://doi.org/10.1016/j.iotcps.2023.05.004>. Acesso em: 20 maio 2025.

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

INSTITUT NATIONAL DE RECHERCHE EN SCIENCES ET TECHNOLOGIES DU NUMÉRIQUE. INRIA. **Intelligence artificielle: domaines de recherche et défis.** [S. /].: INRIA, 2019. Disponível em: <https://www.inria.fr/>. Acesso em: 20 maio 2025.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Fundamentos de metodologia científica.** 8. ed. São Paulo: Atlas, 2019.

MELO, Josemar Henrique de. (Org.). **Catálogo de documentos coloniais manuscritos da Parahyba do Norte – séculos XVIII e XIX.** João Pessoa: Editora A União, 2024. 160 p.

NILSSON, Nils Johan. Inteligência Artificial de Nível Humano? Falando Sério!. **Revista AI**, [s. /], v. 26, n. 4, p. 68, 2005.

RANGEL, Tayná Ladislau. **Mediação Algorítmica como uma das facetas da Mediação da Informação.** 2024. 74 f. Dissertação (Mestrado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Universidade Federal da Paraíba, João Pessoa, 2024. Disponível em: https://repositorio.ufpb.br/jspui/bitstream/123456789/32121/1/Tayn%c3%a1LadislauRangeI_Dissert.pdf. Acesso em: 20 maio 2025.