



**UNIVERSIDADE DO ESTADO DA BAHIA – UNEB**  
**DEPARTAMENTO DE CIÊNCIAS EXATAS E DA TERRA**  
**CURSO DE GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO**

**GILMAR LUÍS DO NASCIMENTO SANTOS**

**IA GENERATIVA APLICADA NA ANÁLISE DE DADOS DO DEPARTAMENTO DE**  
**INFORMÁTICA DO SISTEMA ÚNICO DE SAÚDE – DATASUS**

**SALVADOR**

**2024**

GILMAR LUÍS DO NASCIMENTO SANTOS

**IA GENERATIVA APLICADA NA ANÁLISE DE DADOS DO DEPARTAMENTO DE  
INFORMÁTICA DO SISTEMA ÚNICO DE SAÚDE – DATASUS**

Monografia apresentada ao curso de Sistemas de Informação do Departamento de Ciências Exatas e da Terra da Universidade do Estado da Bahia - UNEB, como requisito à obtenção do grau de Bacharel em Sistemas de Informação. Área de Concentração: Ciência da Computação.

Orientadora: Professora Ma. - Daniela Barreto Araújo  
Coorientador Professor Dr. - Antônio Carlos Fontes Atta

**SALVADOR**

**2024**


GILMAR LUÍS DO NASCIMENTO SANTOS

**IA GENERATIVA APLICADA NA ANÁLISE DE DADOS DO DEPARTAMENTO DE  
INFORMÁTICA DO SISTEMA ÚNICO DE SAÚDE – DATASUS**

Monografia apresentada ao Curso de Bacharelado em Sistemas de Informação do Departamento de Ciências Exatas e da Terra (DCET) - Campus I, da Universidade do Estado da Bahia (UNEB), como requisito à obtenção do grau de bacharel em Sistemas de Informação. Área de concentração: Ciências da Computação.

Aprovada em: 20 de dezembro de 2024.


**BANCA EXAMINADORA**

Documento assinado digitalmente  
 **DANIELA BARRETO ARAUJO**  
Data: 10/01/2025 10:01:45-0300  
Verifique em <https://validar.iti.gov.br>

---

Professora M<sup>a</sup>. - Daniela Barreto Araújo (Orientadora)


Universidade do Estado da Bahia – UNEB

Documento assinado digitalmente  
 **ANTONIO CARLOS FONTES ATTA**  
Data: 09/01/2025 13:54:05-0300  
Verifique em <https://validar.iti.gov.br>

---

Professor Dr. - Antônio Carlos Fontes Atta (Coorientador)

Universidade do Estado da Bahia – UNEB

Documento assinado digitalmente  
 **DEBORA ALCINA REGO CHAVES**  
Data: 10/01/2025 12:40:02-0300  
Verifique em <https://validar.iti.gov.br>

---

Professora M<sup>a</sup>. - Débora Alcina Rego Chaves

Universidade do Estado da Bahia – UNEB

## **AGRADECIMENTOS**

Agradeço a Deus por me permitir passar por esta experiência. Sei que cada obstáculo superado é uma vitória que me enobrece tanto na vida carnal como na espiritual. Espero poder utilizar esses conhecimentos adquiridos para ajudar aqueles que cruzarem meus caminhos.

Não posso deixar de agradecer também aos meus amigos Jever, Mauricio, Augusto, Thiago, Aécio, David, entre outros, pelo apoio nos momentos em que mais precisei.

Sei que vários professores compartilharam comigo seu conhecimento, e eu agradeço a cada um, e em especial à Professora Daniela Araújo, que aceitou ser minha orientadora com satisfação. Também agradeço ao Professor Antônio Atta, que, além de professor da disciplina, também foi meu coorientador.

Por fim, expresso minha gratidão aos meus guias espirituais, que me apoiam e ajudam nas tomadas de decisões, e que estão comigo nos momentos em que me isolo para superar minhas dores.

“Oxóssi, sou grato pelo seu papel em minha vida, que com seus ensinamentos eu venha a ser uma pessoa melhor.

Okê Aró, Meu Pai.”

## RESUMO

Ter direito à informação é essencial para que organizações e cidadãos possam tomar decisões embasadas em evidências, promovendo a busca por conhecimento e autonomia. O Sistema Único de Saúde (SUS) no Brasil enfrenta desafios variados, incluindo a adoção lenta de tecnologias de informação em saúde. Entretanto, a política de informação em saúde desempenha um papel estratégico ao melhorar o acesso equitativo e a gestão das informações, especialmente por meio do Departamento de Informática do Sistema Único de Saúde (DATASUS). Este órgão tem como objetivo reunir, processar e disseminar dados de saúde, promovendo a democratização e uma gestão eficaz dessa área. Contudo, a extração de conhecimento a partir desses dados ainda não é automática, dificultando seu uso prático. A aplicação da inteligência artificial generativa na análise de dados públicos representa uma inovação promissora para a saúde pública no Brasil. Este trabalho avaliou o potencial da inteligência artificial generativa na análise de dados públicos do DATASUS, com foco na produção ambulatorial de Salvador, cumprindo o objetivo geral de explorar a viabilidade dessa tecnologia para melhorar o acesso e a interpretação de informações de saúde. Foram analisadas três ferramentas principais: Copilot Studio, Meu GPT e Gemini Advanced, utilizando critérios como velocidade, precisão, integração de dados e confiabilidade. Os resultados mostraram que o Copilot Studio se destaca pela velocidade de resposta, embora apresente limitações na coerência e confiabilidade. O Meu GPT, apesar de ser mais lento, demonstrou maior precisão e consistência, permitindo uma base de conhecimento para múltiplas interações. O Gemini Advanced apresentou equilíbrio entre velocidade e precisão, mas possui limitações na integração de dados, restringindo sua aplicabilidade em análises prolongadas. A resposta à pergunta do projeto destaca que a inteligência artificial generativa, quando aplicada no contexto analisado, apresenta potencial para transformar a acessibilidade e a eficiência na gestão de dados de saúde pública.

Palavras-chave: Análise de dados, Saúde Pública, IA generativa.

## ABSTRACT

The right to information is essential for organizations and citizens to make evidence-based decisions, promoting the pursuit of knowledge and autonomy. The Unified Health System (SUS) in Brazil faces various challenges, including the slow adoption of health information technologies. However, the health information policy plays a strategic role in improving equitable access and information management, particularly through the Department of Informatics of the Unified Health System (DATASUS). This department aims to gather, process, and disseminate health data, promoting democratization and effective management in this field. Nevertheless, the extraction of knowledge from these data is not yet automated, hindering its practical use. The application of generative artificial intelligence in the analysis of public data represents a promising innovation for public health in Brazil. This study evaluated the potential of generative artificial intelligence in analyzing public DATASUS data, focusing on outpatient production in Salvador, achieving the general objective of exploring the feasibility of this technology to enhance access to and interpretation of health information. Three main tools were analyzed: Copilot Studio, Meu GPT, and Gemini Advanced, using criteria such as speed, accuracy, data integration, and reliability. The results showed that Copilot Studio stands out for its response speed, although it has limitations in coherence and reliability. Meu GPT, despite being slower, demonstrated greater accuracy and consistency, enabling a knowledge base for multiple interactions. Gemini Advanced presented a balance between speed and accuracy but has limitations in data integration, restricting its applicability in prolonged analyses. The project's findings highlight that generative artificial intelligence, when applied in the analyzed context, has the potential to transform accessibility and efficiency in managing public health data.

Keywords: Data analysis, Public health, Generative AI.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Linha do tempo do desenvolvimento da IA .....	24
Figura 2 - Etapas da análise de dados. ....	26
Figura 3 - Formatos dos arquivos extraído até arquivo final, após cada conversão.....	30
Figura 4 - Ambiente inicial para adição de arquivos.....	37
Figura 5 - Ambiente de adição de arquivos.....	37
Figura 6 - Site de demonstração.....	38
Figura 7 – Ambiente de adição de arquivos do Meu GPT.....	39
Figura 8 - página inicial para alimentar um GPT. ....	40
Figura 9 - Ambiente de criação do Gem .....	41
Figura 10 - Ambiente para alimentar o Gem.....	42
Figura 11 - Site de demonstração - Amostra de resultado.....	43
Figura 12 - tabela dinâmica - resultado de pesquisa na base de dados.....	43
Figura 13- Amostra de resultado - Meu GPT .....	44
Figura 14 - tabela dinâmica - resultado de pesquisa na base de dados.....	44
Figura 15 - Amostra de erro ao gerar resultado - Meu GPT.....	45
Figura 16 - Amostra de resultado - Gemini Advanced.....	45
Figura 17 - Amostra de erro ao gerar resultado - Gemini Advanced.....	46
Figura 18 - Gráfico da amostra de resultado - Meu GPT .....	47
Figura 19 - Gráfico da amostra de resultado - Gemini Advanced.....	48
Figura 20 - Gráfico do resumo dos critérios avaliados.....	49

## LISTA DE ABREVEATURAS E SIGLAS

BDC	DBF Compactado
BDF	Data Base Files
CSV	Valores Separados por Vírgula
CTI	Centro Tecnológico de Informática
DATASUS	Departamento de Informática do Sistema Único de Saúde
DM	Data Mining (Mineração de Dados)
DP	Deep Learning.
DSR	Metodologia de Pesquisa em Ciência do Design (Design Science Research Methodology)
ETL	Extract, Transform, Load
IA	Inteligência Artificial
IIS	Informação e Informática em Saúde
LLM	Large Language Models (Modelos de Linguagem Natural)
ML	Machine Learning
MS	Ministério da Saúde
RES	Registros Eletrônicos de Saúde
RMS	Região Metropolitana de Salvador
SIA	Sistema de Informações Ambulatoriais
SIH	Sistema de Internações Hospitalares
SIM	Sistema de Informações sobre Mortalidade
SINASC	Sistema de Informações sobre Nascidos Vivos
SIS	Sistemas de Informação em Saúde
TabNet	Tabulador para Internet
TabWin	Tabulador para Windows
API	Application Programming Interface
SIA/SUS	Sistema de Informações Ambulatoriais do SUS

## SUMÁRIO

1	<b>INTRODUÇÃO</b> .....	13
1.1	<b>Pergunta da pesquisa</b> .....	14
1.2	<b>Objetivos</b> .....	14
1.2.1	Objetivo Geral.....	14
1.2.2	Objetivos Específicos.....	14
1.3	<b>Desafios e perspectivas na gestão de dados de saúde no Brasil.</b> .....	15
2	<b>A IA GENERATIVA POTENCIAL À EXTRAÇÃO DE CONHECIMENTO EM BASE DE DADOS PÚBLICOS ESTRUTURADOS</b> .....	18
2.1	<b>Saúde pública</b> .....	18
2.2	<b>Inteligência Artificial generativa</b> .....	20
2.3	<b>Análise de dados</b> .....	25
2.3.1	Extract, Transform, Load (ETL).....	25
2.4	<b>Trabalhos correlatos</b> .....	27
3	<b>DESCRIÇÃO DO PROJETO</b> .....	28
3.1	<b>Metodologia</b> .....	28
3.2	<b>Coleta de Dados e ferramentas</b> .....	30
3.2.1	Coleta de Dados .....	30
3.2.2	Ferramentas .....	31
3.3	<b>Limpeza e tratamento de dados</b> .....	33
4	<b>DESENVOLVIMENTO</b> .....	35
4.1	<b>Ambiente de experimentação</b> .....	35
4.1.1	Copilot Studio .....	35
4.1.1.2	<i>Criação de Copilotos Personalizados com o Copilot Studio</i> .....	36
4.1.2	Meu GPTs .....	38
4.1.3	Gemini Advanced .....	40
4.2	<b>Análise dos resultados</b> .....	42
4.2.1	Copilot Studio .....	42
4.2.2	Meu GPTs .....	44
4.2.3	Gemini Advanced .....	45
4.3	<b>Método de comparação</b> .....	46
4.3.1	Velocidade de Resposta.....	47

4.3.2	Coerência com a Base de Conhecimento.....	47
4.3.3	Aceitação de Arquivos Externos.....	48
4.3.4	Confiabilidade nas Respostas.....	48
4.3.5	Gráfico Comparativo.....	49
5	<b>CONSIDERAÇÕES FINAIS</b> .....	50
6	<b>TRABALHOS FUTUROS</b> .....	51
	<b>REFERÊNCIAS</b> .....	52

## 1 INTRODUÇÃO

Embora não seja tão prioritário quanto o direito à alimentação, saúde, educação, moradia, lazer e outros, o direito à informação desempenha um papel fundamental na promoção da liberdade e da democracia (Silva, 2009). Uma política pública voltada para viabilizar o acesso às informações, especialmente na área da saúde está intrinsecamente relacionada à produção do conhecimento baseada em dados e, conseqüentemente, melhoria da qualidade de vida dos cidadãos.

É grande a busca por ferramentas que ajudem a compreender dados complexos armazenados em bancos de dados devido à diversidade e ao volume de informações. Essas ferramentas são importantes e necessárias em setores como empresas, pesquisa acadêmica e governo, nos quais lidar com dados estruturados é essencial em uma sociedade centrada em informações.

O Sistema Único de Saúde (SUS) do Brasil, foi criado com a missão de proporcionar serviços de saúde à toda população fundamentando-se nos princípios democráticos da universalidade, integralidade e equidade (Ferraz, 2009). Contudo, enfrenta desafios expressivos, como a adoção lenta de tecnologias de informação em saúde. Nesse cenário, a política de informação em saúde desempenha um papel estratégico, buscando aprimorar o acesso, a disseminação e a gestão de dados, com o intuito de implementar ações voltadas para o bem-estar da população de maneira equitativa.

Em contrapartida, apesar do crescimento global na adoção de Sistemas de Informação em Saúde (SIS), a satisfação dos profissionais com a usabilidade desses sistemas não está acompanhando esse avanço (Ferraz, 2009). Mesmo diante das dificuldades enfrentadas, é importante ressaltar que o uso da tecnologia da informação fomenta progressos significativos na qualidade dos cuidados de saúde e na segurança dos pacientes.

A implementação dos sistemas de saúde no DATASUS<sup>1</sup> simplificou a disponibilização e o acesso a dados para análise, avaliação e gestão em saúde. As informações fornecidas pelo DATASUS podem apoiar análises, decisões e elaboração de programas de ações. Portanto, assegurar a qualidade dos dados é fundamental para uma análise objetiva da situação sanitária, bem como para o processo decisório e planejamento de ações de saúde (Piccolo, 2018).

Estamos no meio de uma revolução nas organizações impulsionada pela potencialidade que os avanços da Inteligência Artificial (IA) têm apresentado à sociedade recentemente. A Inteligência Artificial (IA) generativa está adicionando novas perspectivas à vida dos cidadãos.

---

<sup>1</sup> Departamento de Informática do Sistema Único de Saúde (DATASUS) surgiu em 1991 com a criação da Fundação Nacional de Saúde (Funasa)

É fundamental destacar que essa tecnologia, capaz de criar conteúdo complexo como textos, imagens e sons de forma autônoma, apresenta desafios e oportunidades significativos. Sua importância reside na capacidade de democratizar o acesso a ferramentas avançadas de linguagem natural, permitindo que pessoas sem conhecimento especializado interajam de maneira mais intuitiva e produtiva com sistemas computacionais. No entanto, é essencial ressaltar a necessidade de maior transparência ética por parte das plataformas e serviços que utilizam IA generativa, garantindo um uso responsável e inclusivo dessa tecnologia.

As abordagens quantitativas são amplamente reconhecidas por sua clareza, precisão, abordagem estruturada para coleta de dados e análise de informações numéricas, sendo particularmente adequados para pesquisas que demandam objetividade e possibilidade de generalização (Creswell e Creswell, 2018). Por essa razão, a metodologia de pesquisa quantitativa foi escolhida para este projeto, abrangendo etapas como o desenho experimental, procedimentos de coleta e análise de dados, interpretação e validação dos resultados, e, por fim, as conclusões e recomendações.

## **1.1 Pergunta da pesquisa**

Como a inteligência artificial generativa pode ser utilizada para melhorar a análise de dados ambulatoriais do DATASUS?

## **1.2 Objetivos**

A partir da argumentação da pesquisa citada na seção anterior, foram destacados os seguintes objetivos a serem alcançados no âmbito desta pesquisa.

### **1.2.1 Objetivo Geral**

Analisar como a aplicação da Inteligência Artificial generativa pode viabilizar o acesso ao conhecimento inteligível em saúde pública na Região de Salvador.

### **1.2.2 Objetivos Específicos**

- Extrair e tratar os dados da base do DATASUS, com foco nos registros de produção ambulatorial da cidade de Salvador.
- Realizar transformações nos dados tratados, assegurando sua compatibilidade com modelos de IA generativa.

- Implementar e testar os dados em ambientes experimentais utilizando ferramentas de IA generativa, como Meu GPTs, Copilot Studio e Gemini Advanced.
- Comparar o desempenho das ferramentas utilizadas na extração e interpretação dos dados, destacando limitações e potencialidades.
- Explorar os possíveis impactos dos resultados gerados na tomada de decisão em saúde pública e na alocação de recursos.

### **1.3 Desafios e perspectivas na gestão de dados de saúde no Brasil.**

Diversos públicos como médicos, pesquisadores, técnicos de saúde, os usuários do sistema SUS e pacientes, estão em busca de ferramentas que possam ajudar na compreensão eficaz de dados complexos, que frequentemente estão armazenados em bancos de dados. Essa complexidade dos dados é, em grande parte, devida à diversidade de fontes e à enorme quantidade de informações. Por isso, é de interesse geral encontrar soluções versáteis que se adequem a diferentes tipos de informações. Essas ferramentas desempenham um papel relevante em vários setores, desde análise de dados nas organizações até pesquisa acadêmica e tomada de decisões governamentais. Portanto, a habilidade de gerar informações bem estruturadas e organizadas é um desafio importante que precisamos alcançar para atender às necessidades da sociedade atual, que depende cada vez mais de informações validadas. Os sistemas de informações na atualidade consistem em:

(...) Um conjunto de metodologias, processos e tecnologias que são empregadas para coletar, integrar, analisar e disponibilizar dados transformando-os em informações significativas e úteis para permitir “insights estratégicos, táticos e operacionais” mais eficazes e tomada de decisão, Torres *et al* (2021).

A divulgação de informações governamentais, especialmente na área de saúde, desempenha um papel decisivo na promoção da transparência e na participação pública. No entanto, os sistemas de informações do SIS, conforme identificado em uma revisão sistemática realizada por (Silva, 2009) , precisam melhorar a clareza e a acessibilidade dos dados disponíveis no DATASUS. Nesse contexto, investigar as possibilidades de aplicação da Inteligência Artificial (IA) generativa para analisar os dados do DATASUS visa tornar essas informações mais compreensíveis para profissionais de saúde e cidadãos comuns. Isso é fundamental para garantir que esses dados sejam efetivamente utilizados na tomada de decisões e no desenvolvimento de programas de saúde.

O uso crescente de tecnologias de informação e comunicação na área da saúde, como e-mail e mídias sociais, tem agilizado a troca de dados e informações, auxiliando gestores na definição de estratégias. Além disso, essas tecnologias têm contribuído para a produção de conhecimento em redes e ampliação dos canais de comunicação para acessar serviços de saúde. O Observatórios de Saúde Pública tem desenvolvido e melhorado experiências que combinam o uso de ferramentas digitais para aprimorar o setor de saúde:

(...) para além de serem centros de análise são também estruturas bem claras de comunicação para dentro (informando/ apoiando/ avaliando a tomada de decisão e a intervenção, nomeadamente, a nível local) e para fora do sector de saúde (fazendo a advocacia da saúde e influenciando as políticas públicas com maior impacto na saúde) (Pinto, Rocha, 2016).

Este projeto traz contribuições para a disseminação de informações primordiais na área da saúde. Ao analisar os fatores que levam à hospitalização, obtemos conhecimentos valiosos que podem orientar estratégias preventivas e melhorar a qualidade de vida das pessoas. Além disso, pode fornecer uma compreensão mais detalhada das necessidades de saúde dos idosos, um grupo demográfico de grande interesse.

Podemos concluir que o direito à informação desempenha um papel crucial na promoção da liberdade e da democracia, embora muitas vezes seja considerado menos prioritário do que outras necessidades básicas. No contexto da saúde, a disponibilidade e o acesso às informações são fundamentais para promover o conhecimento baseado em dados e melhorar a qualidade de vida dos cidadãos. O Sistema Único de Saúde (SUS) do Brasil enfrenta desafios significativos, incluindo a adoção lenta de tecnologias de informação. Porém, a implementação de sistemas de saúde, como o DATASUS, tem simplificado o acesso e a disponibilidade de dados para análise e gestão em saúde. (Ferraz, 2009).

No entanto, apesar dos avanços tecnológicos, a usabilidade e a qualidade dos sistemas de informação em saúde ainda precisam ser aprimoradas. A aplicação da Inteligência Artificial (IA) generativa oferece oportunidades significativas para analisar dados de saúde de maneira mais compreensível e acessível, o que pode contribuir para a processo de escolha mais informada e o desenvolvimento de programas de saúde eficazes. Em resumo, os desafios e perspectivas na gestão de dados no Brasil exigem esforços contínuos para melhorar a qualidade, acessibilidade e utilização das informações, especialmente no contexto da Inteligência Artificial (IA) generativa. A aplicação dessa tecnologia oferece oportunidades significativas para analisar dados de maneira mais compreensível e acessível. Entretanto, é importante garantir uma maior

transparência ética nas plataformas e serviços que incorporam IA generativa, para garantir seu uso responsável e benéfico para toda a população.

O percurso seguido para a elaboração da proposta de intervenção será explicado ao longo deste trabalho acadêmico, que está organizado da seguinte maneira.

A seção 2, visa promover uma discussão teórica, embasada em referências que dão sustentação aos encaminhamentos de pesquisa adotados neste trabalho. Inicia-se com uma análise da importância da disseminação de informações em saúde pública, destacando o papel do SUS nesse contexto. Em seguida, explora-se a demanda enfrentada pelo DATASUS na disponibilização de dados que podem subsidiar análises objetivas da situação sanitária, embasando decisões e programas de ações em evidências sólidas. Além disso, essa seção dedica-se à abordagem da IA, suas diversas áreas de aplicação, tipos e relevância no cenário atual, como esta tecnologia pode ajudar na análise de dados sendo integrada a bancos de dados na busca por informações de maneira menos complexa, e concluindo com uma abordagem sobre análise de dados, e sua importância essencial na busca por padrões para tomada de decisões asseguradas.

A seção 3 apresenta a metodologia quantitativa empregada neste projeto, destacando sua importância para garantir a confiabilidade do desenvolvimento.

A seção, fornece uma visão do desenvolvimento, trazendo o ambiente de experimentação, a análise dos resultados e o método de comparação.

Na seção, apresenta as considerações finais, avaliando a eficácia e a usabilidade. Também serão debatidas as sugestões de expansão do trabalho com base na experiência adquirida e no amadurecimento ao longo do desenvolvimento do projeto.

Na seção, será apresentado trabalhos futuros, com novas oportunidades de pesquisa, indicando caminhos que podem aprofundar e ampliar os resultados alcançados.

## **2 A INTELIGENCIA ARTIFICIAL (IA) GENERATIVA POTENCIAL À EXTRAÇÃO DE CONHECIMENTO EM BASE DE DADOS PÚBLICOS ESTRUTURADOS**

Nesta seção, apresentamos a fundamentação teórica que sustenta esta pesquisa, destacando os pilares fundamentais: Saúde Pública, Análise de Dados e Inteligência Artificial. Esses conceitos-chave são explorados para orientar o estudo e alcançar os objetivos propostos na seção introdutória deste trabalho.

### **2.1 Saúde pública**

O Sistema Único de Saúde (SUS) foi estabelecido no Brasil em 1988, com o propósito de oferecer serviços de saúde para toda a população com base em princípios democráticos, universalidade, integralidade e equidade, (Ferraz (2009). Ele engloba entidades públicas em diferentes níveis governamentais e parcerias com o setor privado. A participação da comunidade, por meio de conferências e conselhos de saúde, é fundamental para promover a transparência e a participação cidadã na formulação de políticas de saúde.

No entanto, o SUS enfrenta desafios, como a incorporação lenta de tecnologias de Informação e Informática em Saúde (IIS)<sup>2</sup>, essenciais para a coleta e uso eficaz de dados de saúde. A política de informação em saúde desempenha um papel estratégico para o SUS, visando promover o acesso, disseminação e gestão eficaz de informações de saúde, com a participação ativa da comunidade. Em resumo, o SUS busca garantir o acesso universal a serviços de saúde de qualidade, mas precisa superar desafios, incluindo a adoção de IIS, para aprimorar ainda mais o sistema e promover a saúde da população de maneira equitativa (Ferraz 2009).

A história das políticas de saúde no Brasil é essencial para entender o Sistema Único de Saúde (SUS) pós-1988. Durante a Primeira República, o poder estava centralizado nas oligarquias estaduais. O governo Vargas introduziu políticas sociais centralizadas a partir de 1930. Nos anos 60, tentativas de descentralização foram interrompidas pelo golpe de 1964. O regime militar priorizou as instituições previdenciárias, levando à escassez de recursos na saúde pública.

---

<sup>2</sup>Informação e Informática em Saúde (IIS) é uma área interdisciplinar que combina conceitos, métodos e tecnologias das ciências da saúde, informática e ciência da informação para otimizar a coleta, processamento, armazenamento, análise e uso de dados de saúde

A partir de 1974, surgiram sistemas de informações em saúde focados em dados financeiros, criados por áreas diferentes. O movimento sanitário respondeu ao regime militar e lançou os princípios da Reforma Sanitária. O SUS, criado em 1988, propôs uma estrutura abrangente e democrática. Em 1991, a responsabilidade pelo processamento das contas de saúde passou para o DATASUS (Departamento de Informática do SUS), no Ministério da Saúde (MS), desenvolvendo sistemas de informações em saúde e gestão administrativa, (Ferraz 2009).

O DATASUS, foi estabelecido em 1991 em conformidade com a Constituição de 1988, que instituiu o SUS. Sua criação envolveu a transferência da responsabilidade pelo processamento das contas de saúde da DATAPREV<sup>3</sup> para o DATASUS. Apesar de a DATAPREV possuir equipamentos de grande porte, apenas os microcomputadores e mobiliário foram repassados ao DATASUS. Adicionalmente, o DATASUS recebeu equipamentos do INAMPS<sup>4</sup> e, em 1994, foi integrado ao Centro Tecnológico de Informática (CTI). Inicialmente, o DATASUS operava com uma estrutura limitada, focando principalmente em tarefas relacionadas ao faturamento e sistemas de acompanhamento da saúde, além de questões administrativas.

Com o tempo, essa base de dados ampliou suas atividades e desenvolveu sistemas abrangentes para a gestão local de unidades de saúde, programas de controle de endemias e diversos aplicativos administrativos e financeiros, como o Sistema de Informações sobre Mortalidade (SIM), o Sistema de Informações sobre Nascidos Vivos (SINASC), o Sistema de Internações Hospitalares (SIH) e o Sistema de Informações Ambulatoriais (SIA), desempenhando um papel fundamental na coleta e organização de informações vitais para a saúde da população brasileira. Além disso, o DATASUS deu início ao uso da internet em 1992 e implementou comunicações ponto a ponto em desktops, acompanhando as tendências tecnológicas da época, (Ministério da Saúde, 2002).

Segundo (Ferraz, 2009), encontra-se, no sítio do DATASUS, o seguinte texto referente à apresentação da Instituição:

“A informação é fundamental para a democratização da Saúde e o aprimoramento de sua gestão. A informatização das atividades do Sistema Único de Saúde (SUS), dentro de diretrizes tecnológicas adequadas, é essencial para a descentralização das atividades de saúde

---

<sup>3</sup>A Dataprev é uma empresa pública orientada pela missão de prover soluções digitais para o exercício da cidadania, fornecendo soluções de Tecnologia da Informação e Comunicação para o aprimoramento e a execução de políticas sociais do Estado brasileiro

<sup>4</sup>Instituto Nacional de Assistência Médica da Previdência Social – Responsável por prestar assistência à saúde de seus associados,

e viabilização e controle social sobre a utilização dos recursos disponíveis. Para alcançar tais objetivos, foi atribuída ao Departamento de Informática do SUS - DATASUS, órgão da Secretaria Executiva do Ministério da Saúde, a responsabilidade de coletar, processar e disseminar informações sobre saúde” (Ferraz, 2009).

Portanto, o DATASUS busca operacionalizar, atividades relacionadas à reunião, seleção, codificação, redução, classificação e armazenamento de grandes bases de dados.

A adoção de SIS está crescendo em todo o mundo, mas a satisfação dos profissionais com a usabilidade desses sistemas não está melhorando. No entanto, apesar das dificuldades na implementação de SIS, principalmente os Registros Eletrônicos de Saúde (RES), o uso da tecnologia da informação representa avanços na qualidade da saúde e na segurança do paciente. A usabilidade é frequentemente vista como uma barreira para a aceitação dessas tecnologias após a adoção, embora seja fundamental para otimizar os benefícios do uso de RES. Os problemas de usabilidade dos SIS são uma fonte significativa de estresse no trabalho para os médicos, afetando sua satisfação e eficiência no trabalho, além de ter consequências para a segurança do paciente. Uma maneira de melhorar a usabilidade dos sistemas é envolver os médicos no desenvolvimento das tecnologias, garantindo a integridade dos dados e melhorando a interoperabilidade e estabilidade dos sistemas. Além disso, é fundamental fornecer tempo e suporte suficientes aos médicos para lidar com os problemas relacionados aos SIS, *welchen et al* (2022).

## **2.2 Inteligência Artificial generativa**

Estamos testemunhando uma nova era de transformação industrial, impulsionada pelo avanço da Inteligência Artificial. Embora as pesquisas relacionadas à Inteligência Artificial (IA) não sejam recentes, a obtenção de resultados tem sido mais lenta em comparação com outras áreas da Computação. Isso abre espaço para possíveis explorações do conceito e das ferramentas associadas à IA generativa.

A Inteligência Artificial não apenas está sendo aplicada em trabalhos manuais, mas também em tarefas que requerem inteligência. Isso inclui carros autônomos, diagnósticos médicos precisos realizados por robôs assistentes, tradutores automáticos cada vez melhores, sistemas de visão computacional altamente eficientes e sistemas de recomendação personalizados, como os da Amazon, Netflix e Spotify, (Carraro, 2023).

Esses avanços na Inteligência Artificial estão transformando várias áreas, desde a medicina até o entretenimento, e têm o potencial de aumentar a eficiência e a precisão em muitos domínios. No entanto, é importante observar que, mesmo com essas inovações, a Inteligência Artificial não visa substituir profissionais, mas sim complementar e aprimorar suas capacidades (Ludermir, 2021).

A Inteligência Artificial (IA) envolve a criação de máquinas que imitam processos naturais, usando algoritmos complexos para aprender, tomar decisões e interagir com base em dados. Partindo para o âmbito etimológico, define-se artificial como algo não natural criado para mimetizar a natureza, enquanto que a palavra “inteligência” está relacionada ao entendimento e habilidade de solucionar problemas por meio do conhecimento. Portanto, a Inteligência Artificial (IA) é a capacidade de máquinas aprenderem e agirem com base em programação prévia e algoritmos, (Damaceno e Vasconcelos 2018).

A Inteligência Artificial (IA) pode ser subdividida em camadas ou partes, introduzindo os conceitos de *Deep Learning* (DL) e *Machine Learning* (ML). O DL é uma vertente do ML que capacita máquinas para executar tarefas complexas, como reconhecimento de fala e identificação de imagens, através do estabelecimento de parâmetros nos dados e treinamento autônomo, utilizando várias camadas de processamento para reconhecer padrões, emulando o aprendizado intuitivo humano. O ML é um processo de aprendizado contínuo em que a máquina utiliza dados de entrada para aprender e gerar saídas que resolvam problemas específicos, (Damaceno e Vasconcelos, 2018).

O Aprendizado de Máquina ou *Machine Learning* (ML) é uma subárea da inteligência artificial que utiliza algoritmos para aprender com dados, em vez de instruções pré-programadas. É amplamente aplicado em várias áreas, incluindo detecção de fraudes, mecanismos de busca, vigilância por vídeo e medicina, onde é usado para prevenção, diagnóstico e tratamento de doenças. Diferentemente dos sistemas especialistas tradicionais, as técnicas de ML lidam com grandes volumes de dados, explorando diversas combinações para prever resultados confiáveis, especialmente em situações de *Big Data*, Paixão *et al.* (2022).

Em 2001, Doug Laney introduziu o conceito de "3 Vs" para descrever o *Big Data*, destacando seu grande volume, alta velocidade e alta variedade de informações. Isso requer novas técnicas de processamento para facilitar descobertas e otimização de processos, desafiando as ferramentas tradicionais de gerenciamento e armazenamento de dados, Paixão *et al.* (2022).

O desenvolvimento de um algoritmo de ML envolve três fases: pré-processamento, que organiza dados e define perguntas de pesquisa; avaliação do modelo, essa etapa permite criar

padrões de mapeamento para rotular com precisão novos dados previamente não analisados e o treinamento, que pode ser supervisionado ou não, Paixão *et al.* (2022).

O Aprendizado Supervisionado consiste em treinar a máquina com base em dados de amostra de entrada, onde se espera que ele aponte a saída correta ou a mais provável. Tem a vantagem de ser muito utilizado para solucionar problemas de classificação e regressão. Fácil de compreender seu funcionamento, não há necessidade de armazenar dados de treinamento. No entanto tem como desvantagem a necessidade que os dados de treinamento sejam rotulados, pode haver sobre ajuste caso o modelo seja treinado por muito tempo e pode não ser capaz de resolver problemas complexos (Carraro, 2023).

Já o Aprendizado Não Supervisionado tenta encontrar padrões ocultos entres os dados e associá-los ou aprender possíveis relações entre eles, trabalha com dados não rotulados, desta forma não é possível indicar uma saída, é ideal para problemas de agrupamentos, associação ou detecção de anomalias, para dados não rotulados, acelera a coleta e a preparação de dados. Nesta modalidade de aprendizado, os rótulos podem ser aplicados a dados já classificados, apresentando como desvantagem o alto risco de obter resultados irrelevantes, além disso é possível avaliar e interpretar em maior tempo e com necessidade de dados de entrada de boa qualidade (Carraro, 2023).

A complexidade em ascensão dos problemas computacionais e a vasta quantidade de dados gerados estão impulsionando a demanda por ferramentas de Aprendizado de Máquina. O ML capacita computadores a aprender com experiências anteriores, aplicando o princípio de indução para desenvolver funções ou hipóteses com base em dados observacionais do problema a ser abordado, (Batista e Filho, 2019).

A inteligência artificial (IA) pode ser categorizada em três tipos distintos. Primeiro, tem-se a IA Focada, também conhecida como IA Fraca, que consiste em algoritmos especializados projetados para resolver problemas específicos em áreas particulares, como no reconhecimento facial ou no sistema de atendimento de uma empresa, a exemplo de Assistentes Virtuais como a Alexa (Amazon), Siri (Apple), Sistemas de Recomendações Netflix, Spotify, Veículos Autônomos, (Ludermir, 2021).

Segundo, cita-se a Inteligência Artificial (IA) Superinteligente que representa um cenário hipotético em que os algoritmos seriam significativamente mais capazes do que os seres humanos em praticamente todas as tarefas. Entretanto, até o momento, não existem sistemas que tenham atingido esse nível de inteligência artificial, e permanece incerto se tais sistemas serão desenvolvidos no futuro, (Valle, Gasó e Ajus, 2023).

Por fim, a Inteligência Artificial (IA) Generalizada, ou Inteligência Artificial (IA) Forte, é caracterizada por algoritmos que buscam atingir o nível de competência humana em várias tarefas, por exemplo o ChatGPT, Gemini (antigo BARD). Geralmente, estes algoritmos fazem uso de técnicas de Aprendizado de Máquina, e em alguns domínios, como a Visão Computacional, já demonstraram desempenho comparável ao humano. Atualmente, a maior parte da Inteligência Artificial (IA) se enquadra nessa categoria, (Ludermir, 2021).

A inteligência artificial generativa está se tornando cada vez mais comum em diferentes setores, e sua expansão está gradativamente influenciando nossas vidas. No ano de 2023, IAs generativas, como ChatGPT<sup>5</sup>, Midjourney<sup>6</sup> e outras, conseguiram atingir vastas audiências, impactando a sociedade em amplo espectro. É fundamental notar a crescente integração de IAs por meio de APIs, o que aponta para desafios e oportunidades significativas no contexto da Inteligência Artificial (IA) generativa, (Lopezosa, 2023).

As Inteligências Artificiais (IAs) generativas estão revolucionando diversas indústrias, como a criação de imagens digitais, composição de música e geração de texto, permitindo automação criativa abrindo novas possibilidades. O ChatGP, treinado em grandes volumes de dados é especializado na geração de texto contextualmente relevante. Projetado para responder perguntas, realizar diálogos, fornecer informações detalhadas em conversas de forma natural. Pode ser usado para criação de conteúdo até assistente virtual. É importante considerar que o ChatGPT possui períodos de atualização que são ajustados ao longo do tempo, (Openai, 2022).

A *Dall-E* e *Midjourney* são geradores de imagem que, a partir de descrições em linguagem natural (prompts), criam imagens realistas e arte visual. O Jasper é ideal para gerar textos desenvolvido para uso comercial, marketing e vendas. O Character.AI é voltado à interação, permitindo que os usuários criem personagens virtuais com personalidades distintas e interajam com eles de forma natural (Trindade e Oliveira, 2024).

No entanto, é importante lembrar que a Inteligência Artificial (IA) deve ser usada com responsabilidade, servindo como ponto de partida, não como resultado definitivo. Isso deve refletir o esforço e a contribuição genuínos do usuário, (Lopezosa, 2023).

A Inteligência Artificial (IA) generativa, baseada em modelos de linguagem natural, ou *Large Language Models* (LLM), um modelo treinado em grandes volumes de texto, o que lhe permite: Compreender linguagem natural, gerar respostas textuais coerentes, traduzir idiomas,

---

<sup>5</sup> ChatGPT é um modelo de inteligência artificial desenvolvido pela OpenAI, projetado para entender e gerar linguagem natural.

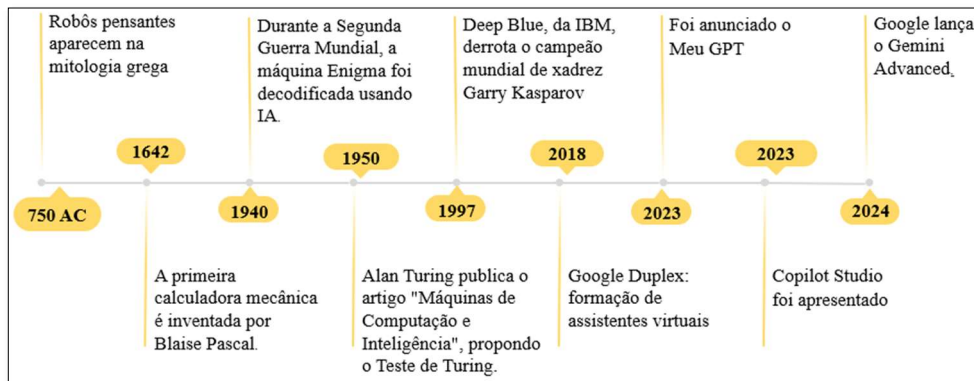
<sup>6</sup>Midjourney é uma ferramenta de inteligência artificial (IA) focada na criação de imagens a partir de descrições textuais.

realizar tarefas complexas, como escrever código, responder a perguntas técnicas ou ajudar na criação de conteúdo criativo. Estes modelos estão evoluindo rapidamente e tem potencial para impactar várias áreas. Apesar do destaque em 2023, o ChatGPT, lançado em 2022, foi um marco significativo, tornando a tecnologia mais acessível a pessoas sem conhecimento especializado em Inteligência Artificial (IA). Pesquisadores e profissionais estão explorando seu uso em educação, saúde e ciência da informação, (Ramos, 2023).

Apesar do sucesso da Inteligência Artificial (IA) generativa, ainda estamos descobrindo suas implicações, especialmente no ensino superior e na pesquisa acadêmica, onde a rápida evolução da Inteligência Artificial (IA) representa um desafio. As ferramentas baseadas em LLMs têm o potencial de melhorar a produtividade em tarefas acadêmicas. Desde a pesquisa até a publicação de artigos, auxiliando em diversas atividades, como busca de literatura, análise de textos e tradução automática (Ramos, 2023).

O conceito de robôs com capacidade de pensamento já era discutido na antiguidade, como destacado por Cardoso *et al* (2023). A figura 1, apresenta uma análise da evolução da Inteligência Artificial (IA) ao longo do tempo.

Figura 1 - Linha do tempo do desenvolvimento da IA



Fonte: Adaptada de Cardoso *et al* (2023).

A utilização ampla de dados pela Inteligência Artificial (IA) generativa para produzir resultados podem, inadvertidamente, manipular informações sensíveis e privadas, gerando preocupações significativas sobre privacidade. Essas Inteligência Artificial (IA)s generativas geralmente se baseiam em dados provenientes da internet para seu treinamento. Embora a criptografia seja uma alternativa viável, é crucial ressaltar a importância de uma maior transparência ética por parte das plataformas e serviços que incorporam Inteligência Artificial (IA) generativa, Cardoso *et al* (2023).

## 2.3 Análise de dados

A análise de dados é uma rotina dedicada a extrair informações úteis a partir de grandes volumes de informações brutas. Com o avanço da tecnologia e a crescente disponibilidade de dados, tornou-se essencial para empresas e organizações utilizarem essas informações para tomar decisões estratégicas. Organizar, pensar sobre os conteúdos dos dados antes de manipular é essencial durante a fase de análise pois, esse processo envolve várias etapas, incluindo a coleta, tratamento, análise exploratória, modelagem e interpretação dos dados, (Islam, 2020).

O processo de análise de dados é fundamental para compreender e extrair informações valiosas dos conjuntos de dados. Através da análise, é possível identificar padrões, tendências que apoiam a tomada de decisões interessantes em diversos contextos, desde o ambiente empresarial até a pesquisa científica. As etapas e métodos empregados nesse processo evidenciam sua importância (Islam, 2020).

A forma convencional de analisar dados é lenta e dispendiosa, o que a torna inviável à medida que os volumes de dados aumentam. Com o crescimento constante dos bancos de dados, é crucial automatizar parte da análise. Tanto instituições privadas quanto públicas registram suas operações de forma computacional, resultando na acumulação de grandes conjuntos de dados. Atualmente a informação e o conhecimento são cruciais em diversos setores, incluindo saúde, negócios e educação (Galvão e Marin, 2009).

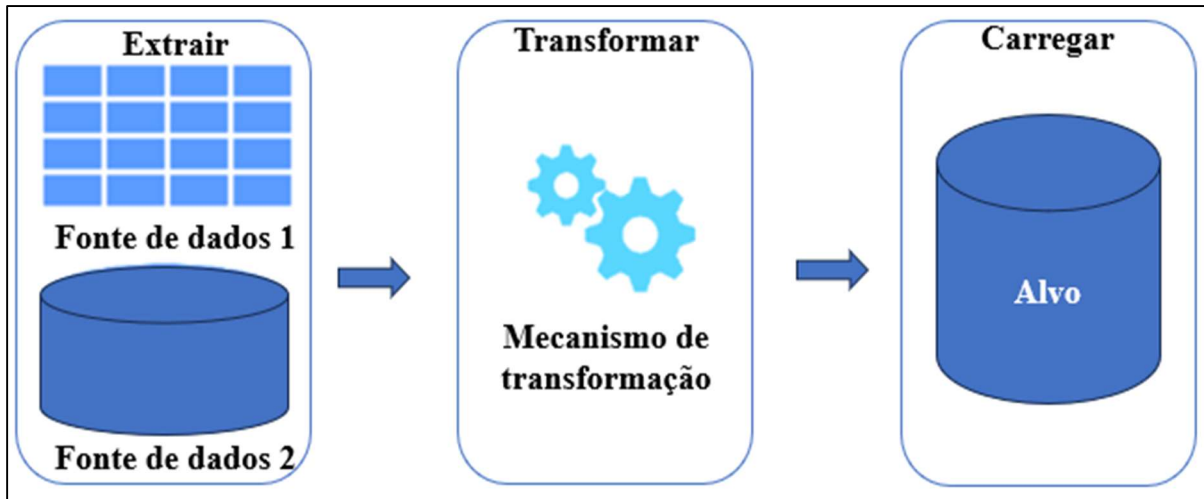
Muitos gestores e profissionais de saúde têm dificuldade em compreender e utilizar eficazmente os dados de saúde devido à enorme quantidade de informações geradas. A evolução da informática e das tecnologias de dados oferece soluções automatizadas para lidar com essa complexidade e extrair informações úteis desses dados. No entanto, essa mudança requer adaptação e aprendizado por parte dos profissionais de saúde. A utilização de novas técnicas, métodos e ferramentas computacionais auxiliam na extração de informações úteis (Galvão e Marin, 2009).

### 2.3.1 *Extract, Transform, Load* (ETL)

Contrariando a crença comum, a preocupação da maioria das empresas e agências governamentais não é a falta de dados, mas sim o excesso de informações disponíveis. Definir o motivo pelo qual os dados são necessários, coletar, processar, limpar, analisar e comunicar

são etapas fundamentais da análise de dados e componentes do *Extract, Transform, Load (ETL)*<sup>7</sup>, figura 2.

Figura 2 - Etapas da análise de dados.



Fonte: <https://learn.microsoft.com/pt-br/azure/architecture/data-guide/relational-data/etl>

É crucial esclarecer a finalidade da análise de dados e como ela será utilizada, (Islam, 2020).

Após a definição da finalidade dos dados, deve-se iniciar a coleta, extraindo dados brutos de diversas fontes e armazenando-os em bases de dados para uso posterior. Após a extração, ocorre a transformação para limpar, adequar, integrar e padronizar dados. A transformação dos dados é uma etapa vital, pois, apesar dos cuidados, erros sempre podem ocorrer. Identificar e corrigir esses erros é essencial para evitar resultados enganosos nas pesquisas, (Islam, 2020).

Enfim, a análise de dados permite transformar informações brutas em insights que podem ser usados para aumentar o sucesso de um projeto ou organização. Deve-se ter cuidados sobre a origem dos dados, como foram coletados, tamanho da amostra e se há conflito de interesses, evitando que a pesquisa seja questionável. A análise dos dados na busca de insights e informações úteis, é fundamental para identificação de padrões que apoiem nas tomadas de decisões, na interpretação dos resultados, na aplicabilidade e na geração de conhecimento.

<sup>7</sup> ETL (extração, transformação e carregamento) é um pipeline de dados usado para coletar dados de várias fontes.

## 2.4 Trabalhos correlatos

Para a realização deste estudo, foi conduzida uma revisão de literatura acadêmica e projetos correlatos. No artigo **"Qualidade de Dados dos Sistemas de Informação do DATASUS: Análise Crítica da Literatura"** (Piccolo, 2018), destaca que a extração de informações do DATASUS enfrenta desafios significativos devido à complexidade dos dados. A falta de padronização e a necessidade de garantir integridade, completude e confiabilidade complicam análises objetivas e a tomada de decisões em saúde. A autora enfatiza a importância de mais estudos voltados à avaliação e melhoria da qualidade dos dados, a fim de maximizar o uso estratégico do DATASUS.

Saldanha, Bastos e Barcellos, (2019) no artigo **"Desafios na Análise de Dados do DATASUS"** destaca a complexidade inerente ao processamento de informações do sistema de dados do DATASUS. A estrutura dos arquivos apresenta desafios significativos devido ao grande volume de variáveis, códigos numéricos não intuitivos e múltiplas tabelas que precisam ser cruzadas para gerar informações úteis. Esses obstáculos tornam a extração de informações uma tarefa que exige alta expertise técnica e metodológica.

### **3 DESCRIÇÃO DO PROJETO**

Nesta etapa, serão apresentadas a metodologia, as plataformas que foram utilizadas e as etapas detalhadas para obtenção dos resultados.

#### **3.1 METODOLOGIA**

A escolha do método de pesquisa representa o trajeto que o pesquisador opta por seguir para validar sua investigação, explicando os fundamentos dessa seleção na busca por resultados significativos e eficazes para atingir os objetivos, (Oliveira, 2011).

A metodologia de pesquisa quantitativa fundamenta-se na medição e análise de fenômenos para questionar os limites do conhecimento e pensamento. Abordagens quantitativas são frequentemente associadas a clareza, precisão, análise de dados numéricos e coleta de dados estruturada, sendo amplamente utilizadas em pesquisas que exigem objetividade e generalização (Creswell e Creswell, 2018).

A metodologia deste estudo foi estruturada com base em uma análise comparativa, visando avaliar o desempenho de diferentes ferramentas de inteligência artificial generativa aplicadas à análise de dados do DATASUS. Essa abordagem permite identificar potencialidades e limitações das ferramentas em cenários reais de uso, proporcionando uma visão detalhada sobre sua eficiência em diversas aplicações. A análise foi fundamentada em critérios bem definidos, assegurando uma avaliação rigorosa e consistente.

A análise comparativa utilizou os dados do DATASUS processados no Excel, uma ferramenta amplamente aceita e reconhecida pela sua versatilidade. Sendo assim, a análise tomará como base o estudo das três ferramentas utilizando critérios como; velocidade, precisão, integração de dados e confiabilidade.

A velocidade será analisada para determinar o tempo de resposta das ferramentas durante a execução de consultas e o processamento de dados. Esse critério é essencial para medir a eficiência operacional, especialmente em cenários que demandam rapidez na geração de informações em tempo real.

A precisão refere-se à capacidade das ferramentas de fornecer respostas corretas e consistentes com os dados presentes na base de conhecimento. Irá avaliar como cada ferramenta interpretará os dados processados e gerando resultados coerentes, minimizando erros ou interpretações ambíguas.

No critério de integração de dados, a análise focará na capacidade das ferramentas de aceitar e trabalhar com diferentes fontes, além de sua funcionalidade ao integrar múltiplas fontes de informação de maneira eficiente. A facilidade de carregamento e manipulação dos arquivos será considerada uma característica fundamental para análises mais complexas.

Por fim, a confiabilidade deverá ser analisada com base na consistência dos resultados e na estabilidade das ferramentas durante os testes. O foco estará na replicabilidade das respostas geradas e na estabilidade operacional das plataformas, garantindo que o processo analítico não fosse comprometido por falhas.

Essa estrutura metodológica permitirá uma avaliação detalhada e imparcial das ferramentas, fornecendo informações valiosas sobre sua aplicabilidade no contexto da análise de dados em saúde pública.

Com base nas características fundamentais da metodologia de pesquisa quantitativa, este trabalho segue um conjunto estruturado de etapas que será detalhado nos parágrafos seguintes. Entre essas etapas estão o desenho experimental, os procedimentos de coleta e análise de dados, a interpretação e validação dos resultados e as conclusões e recomendações. Cada uma sendo descrita de forma a garantir uma abordagem rigorosa e consistente ao longo de todo o processo.

Desenho experimental: é uma etapa fundamental na metodologia quantitativa, pois define como a pesquisa será estruturada para coletar, organizar e analisar os dados de forma rigorosa. Neste projeto as etapas serão: Definição do Problema e Objetivo - Esclarecer a pergunta de pesquisa ou a hipótese que será testada. Identificação das Variáveis que podem ser - Independentes: Variáveis manipuladas ou observadas para identificar sua relação com outras neste projeto vamos trabalhar com o CID, Dependentes: Variáveis medidas como resultados do experimento como a frequência de atendimentos e a variáveis e Controladas: Mantidas constantes para garantir que os resultados não sejam influenciados por fatores externos. Neste trabalho o período analisado é 2023, a fonte de dados é o DATASUS. (Creswell e Creswell, 2018).

Procedimentos de Coleta e Análise de Dados - Essa etapa da metodologia quantitativa é responsável por detalhar como os dados serão obtidos, tratados e analisados, garantindo a consistência e a confiabilidade do processo.

Interpretação e validação dos resultados: Esta etapa fortalecem a confiabilidade da pesquisa, conectam os dados aos objetivos e garantem que as conclusões sejam embasadas em

evidências sólidas. Isso permite que os achados contribuam de forma significativa para a tomada de decisão ou desenvolvimento de políticas públicas.

**Conclusões e recomendações:** A etapa de fechamento do processo de pesquisa e serve para sintetizar os achados mais relevantes, conectando-os aos objetivos iniciais. Ela também oferece sugestões práticas ou teóricas para o uso dos resultados, bem como orientações para estudos futuros ou melhorias no contexto analisado.

### 3.2 Coleta de Dados e Ferramentas

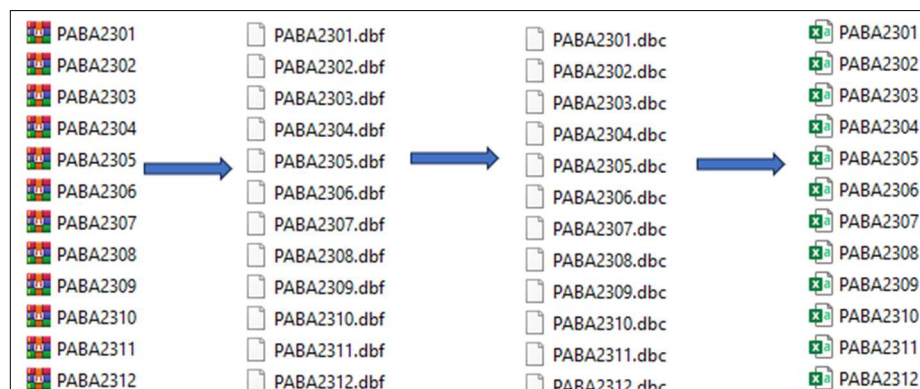
Nesta etapa, serão apresentados a coleta de dados as ferramentas e as tecnologia utilizadas na pesquisa, com ênfase na organização e no processamento dos dados.

#### 3.2.1 Coleta de Dados

A coleta de dados foi feita na base de dados do DATASUS, a estruturação das informações foi realizada no Microsoft Excel, os arquivos estão no formato CSV. Para a análise e manipulação dos dados, utilizou-se a linguagem Python manipulando com a biblioteca pandas, implementada no ambiente de desenvolvimento Jupyter Notebook.

O material necessário para alcançar os objetivos do presente estudo, foi extraído da base de dados no site do DATASUS (<https://datasus.saude.gov.br/transferecia-de-arquivos/>). Foi feito o download dos 12 arquivos zip correspondente ao ano de 2023 - SIA/SUS/Produção Ambulatorial, da documentação - Informe Técnico SIA/SUS e do tabulador TabWin, fornecido pelo DATASUS. O TabWin permite a conversão dos dados do formato BDC (DBF compactado) para o formato Data Base Files (DBF), posteriormente, para o formato CSV (Comma Separated Values). Essa conversão viabiliza a manipulação eficiente e adequada dos dados, atendendo às necessidades de análise e uso desses dados de forma funcional e produtiva, figura 3.

Figura 3 - Formatos dos arquivos extraído até arquivo final, após cada conversão.



Fonte: Adaptada dos dados extraído do DATASUS.

### 3.2.2 Ferramentas

#### 3.2.2.1. *Python*

Python é uma linguagem de programação notável por sua eficiência e versatilidade, especialmente na análise de dados. Com uma sintaxe clara e intuitiva, ela facilita o aprendizado, tornando-se uma escolha prática tanto para quem está começando quanto para quem já possui experiência na área. A linguagem oferece estruturas de dados robustas e ferramentas poderosas, permitindo lidar com grandes volumes de informações de maneira organizada e eficiente. Por ser interpretativa e contar com tipagem dinâmica, Python possibilita o desenvolvimento ágil de soluções, desde pequenos scripts até aplicações mais complexas. Além disso, o vasto conjunto de bibliotecas disponíveis, como Pandas (para manipulação de dados), NumPy (operações matemáticas), Matplotlib (visualização) e Scikit-learn (aprendizado de máquina), oferecendo recursos avançados para explorar, visualizar e modelar dados. Essa combinação de simplicidade e capacidade faz do Python uma ferramenta indispensável para quem trabalha com análise de dados em qualquer área. (Python, 2024)

#### 3.2.2.2 *Rotinas Python*

A rotina desenvolvida em Python teve como objetivo principal o processamento e limpeza de dados provenientes de arquivos extraídos do DATASUS relacionado à produção ambulatorial, utilizando bibliotecas como Pandas, NumPy e Datetime. A seguir, são descritas as etapas principais dessa rotina:

**Carregamento dos Dados:** O arquivo com extensão .csv, foi carregado utilizando a função `pd.read_csv()` da biblioteca Pandas, permitindo o acesso estruturado aos dados em forma de DataFrame.

**Limpeza e Seleção de Colunas:** Foram removidas colunas desnecessárias para a análise, utilizando a função `drop()`. Essa etapa reduziu significativamente a dimensionalidade do conjunto de dados, focando apenas nas informações relevantes.

**Tratamento de Valores Ausentes:**

Valores faltantes (NaN) foram substituídos pelo valor zero (0) com a função `fillna()`. Em seguida, foi realizada uma verificação de possíveis valores ausentes remanescentes utilizando `isna().sum()`.

**Filtragem de Dados:** Apenas os registros pertencentes ao município de código 292740 foram mantidos, utilizando o método `query()` para criar um filtro eficiente.

**Renomeação de Colunas:** Para facilitar a leitura e o entendimento do conjunto de dados, os nomes das colunas foram alterados para descrições mais compreensíveis, como `codigo_do_estabelecimento`, `idade_minima` e `sexo`.

**Conversão de Tipos de Dados:** Colunas como `CID_principal` foram convertidas para o tipo `str` para garantir a consistência no tratamento textual. Datas foram convertidas para o formato `datetime` utilizando `pd.to_datetime()`, garantindo precisão em análises temporais. Colunas numéricas específicas, como `idade_minima` e `idade_maxima`, foram convertidas para o tipo `int64` para otimizar o uso de memória.

**Filtragem por Faixa Etária:** Para refinar o conjunto de dados, foram mantidos apenas registros com idades entre 60 e 130 anos, utilizando uma condição lógica aplicada com `loc[]`.

**Limpeza de Registros com Ocupações Específicas:** Linhas cujas ocupações dos profissionais correspondiam a códigos indesejados (2231G1 e 2231F9) foram removidas. Além disso, valores específicos como 3135D1 foram substituídos por 313501 para padronização.

**Reindexação:** Após as etapas de filtragem, o `DataFrame` foi reindexado para iniciar os índices a partir de 0, utilizando `reset_index()`.

**Exportação do Conjunto Final:** O `DataFrame` processado foi salvo em um novo arquivo CSV. Essa exportação foi realizada com o parâmetro `index=False` para evitar a inclusão de índices no arquivo final.

Essa rotina automatizou etapas essenciais de limpeza e processamento de dados, tornando o conjunto de dados mais consistente, legível e adequado para análises posteriores. A aplicação de filtros, padronização de formatos e exclusão de registros irrelevantes garantiu a qualidade dos dados processados. Além disso, o uso de bibliotecas robustas como `Pandas` e `NumPy` conferiu eficiência e escalabilidade ao processo.

### 3.2.2.3 *Pandas*

`Pandas` é uma biblioteca de código aberto, licenciada sob a BSD, amplamente utilizada em Python para análise e manipulação de dados. Ela oferece estruturas de dados flexíveis e ferramentas de alto desempenho, facilitando o processamento de grandes volumes de informações. Com uma interface intuitiva, é essencial para cientistas de dados e analistas que buscam eficiência em suas análises. Sua integração com outras bibliotecas do ecossistema Python amplia ainda mais suas funcionalidades. (pandas, 2024)

#### 3.2.2.4 Jupyter Notebook

O Jupyter Notebook é uma ferramenta moderna e prática, disponível na web, que permite criar e editar notebooks computacionais. Com ele, é possível integrar código, texto, gráficos e resultados em um único documento, o que facilita tanto o entendimento quanto a apresentação de análises. Essa plataforma tem se destacado especialmente em áreas como ciência de dados, aprendizado de máquina e educação, tornando-se uma aliada indispensável para quem busca trabalhar de forma organizada e colaborativa (Jupyter, 2015).

#### 3.2.2.5 Excel

O Excel é amplamente reconhecido como uma ferramenta essencial para a análise de dados devido à sua flexibilidade. Ele permite organizar informações, realizar cálculos numéricos e criar gráficos, tornando-se indispensável em contextos acadêmicos e profissionais. Com uma interface intuitiva, o Excel facilita a manipulação de grandes volumes de dados, sendo uma escolha prática e eficiente para análises detalhadas.

A funcionalidade Tabela Dinâmica do Excel é especialmente útil para sumarizar informações de maneira eficiente, permitindo análises rápidas. O uso de fórmulas avançadas, como PROCV (busca de valores), SOMASE (somadas condicionais) e ÍNDICE/CORRESP (cruzamento de dados), amplia a capacidade de personalizar análises. Entretanto, o Excel apresenta limitações no processamento de bases de dados muito extensas, como o limite de linhas, exigindo ferramentas complementares para volumes maiores.

### 3.3 LIMPEZA E TRATAMENTO DE DADOS

A limpeza e tratamento dos dados são etapas cruciais para o processo de análise, principalmente se a base de dados for complexa e extensa como as obtidas no DATASUS. A averiguação iniciou-se com a verificação da estrutura dos dados e dos tipos de variáveis visando identificar possíveis problemas. Utilizamos a documentação para compreender as características básicas dos dados e detectar anomalias, como valores ausentes ou tipos de informações incorretas.

Após a compreensão da estrutura dos dados deu-se início a limpeza do arquivo de Produção Ambulatorial, para esta limpeza foi utilizado a ferramenta jupyter Notebook, a biblioteca panda para análise e manipulação de dados. Os materiais obtidos correspondem aos meses de janeiro a dezembro do ano de 2023, totalizando 12 arquivos. Cada tabela é composta por 60 colunas e uma média de 2.410.239 (Dois milhões, quatrocentos e dez mil, duzentos e

trinta e nove) linhas, que após o tratamento e limpeza ficaram com o total de 12 arquivos com tabelas composta por 21 colunas, uma média de 162.195 (cento e sete mil, novecentos e setenta e dois) linhas, conforme tabela 1. Durante a limpeza dos foram eliminados todas as colunas que não seriam utilizadas como CNPJ, Código de Regra Contratual CID Secundário entre outras, alterei o valor nulo por zero, excluir os municípios diferentes de Salvador já que o foco é Salvador, alterei o nome das colunas, modifiquei o tipo da coluna CIP\_principal de int para str, o tipo data das colunas data\_da\_realizacao\_do\_procedimento e data\_de\_processamento de int para datetime, limitei a idade para maiores ou iguais a 60 anos.

Tabela 1 - Resultado após limpeza e tratamento dos dados

ANO	MÊS	QUANT DE LINHAS	QUANT DE LINHAS	
			APOS LIMPEZA	PORCENTAGEM
2023	jan	1953699	146782	7,51%
2023	fev	2019182	130146	6,45%
2023	mar	2398792	153907	6,42%
2023	abr	2298858	143351	6,24%
2023	mai	2568930	161929	6,30%
2023	jun	4479728	159422	3,56%
2023	jul	2633800	162461	6,17%
2023	ago	2421686	164529	6,79%
2023	set	2393161	166219	6,95%
2023	out	2445548	168853	6,90%
2023	nov	2478426	182448	7,36%
2023	dez	2171092	170295	7,84%
<b>MÉDIA</b>		<b>2410239</b>	<b>162195</b>	

Fonte: Adaptada dos dados do DATASUS

## 4 DESENVOLVIMENTO

A experimentação com diferentes ferramentas de inteligência artificial (IA) neste projeto é fundamental para compreender como essas tecnologias podem transformar a análise de dados. Examinar o desempenho dessas ferramentas, avaliando sua eficácia na interpretação e processamento dos dados, é um passo essencial para validar sua aplicabilidade em contextos reais.

### 4.1 AMBIENTE DE EXPERIMENTAÇÃO

Esta seção apresenta o ambiente experimental utilizado para investigar a aplicação de inteligências artificiais generativas na análise de dados de saúde, destacando suas funcionalidades e suas aplicações. O projeto concentra-se em três ferramentas principais: o Copilot Studio da Microsoft, o Meu GPTs da OpenAI e o Gemini Advanced Google. Essas plataformas representam soluções inovadoras de inteligência artificial. As três ferramentas analisadas neste projeto destacam-se por sua versatilidade na aceitação de diversos tipos de arquivos como base para suas funcionalidades. Essas plataformas são compatíveis com formatos amplamente utilizados, como documentos de texto (.txt, .docx), planilhas (.xls, .xlsx, .csv), apresentações (.pptx), PDFs (.pdf).

Para garantir uma análise consistente e comparativa, neste projeto as ferramentas foram alimentadas com as mesmas instruções e os mesmos arquivos no formato .CSV. Esse formato foi escolhido devido à sua simplicidade e compatibilidade com os sistemas e ferramentas de inteligência artificial. O uso padronizado de arquivos .CSV permitiu avaliar de forma justa as funcionalidades, os contextos de aplicação e as limitações de cada solução.

Essa abordagem uniformizada não apenas assegurou a integridade dos resultados, mas também ofereceu uma visão clara sobre como essas plataformas lidam com o mesmo tipo de dado em suas análises. A análise aborda as características específicas de cada sistema e os contextos em que foram aplicados, proporcionando uma visão detalhada de suas potencialidades e limitações nesse cenário. Essa padronização foi essencial para compreender as potencialidades de cada ferramenta no cenário da análise de dados de saúde.

#### 4.1.1 Copilot Studio

Um copiloto é uma interface de inteligência artificial que utiliza grandes modelos de linguagem (LLMs) e outras fontes de conhecimento. Ele atua como assistente, respondendo perguntas, resolvendo problemas complexos e interagindo em vários idiomas e canais, como

sites, aplicativos e plataformas de mensagens. O Copilot Studio, não exige conhecimentos avançados em ciência de dados ou programação. Suas aplicações são diversas, incluindo suporte em vendas, fornecimento de informações corporativas, rastreamento de saúde pública e atendimento a perguntas frequentes, tornando-se uma ferramenta versátil para empresas e organizações.

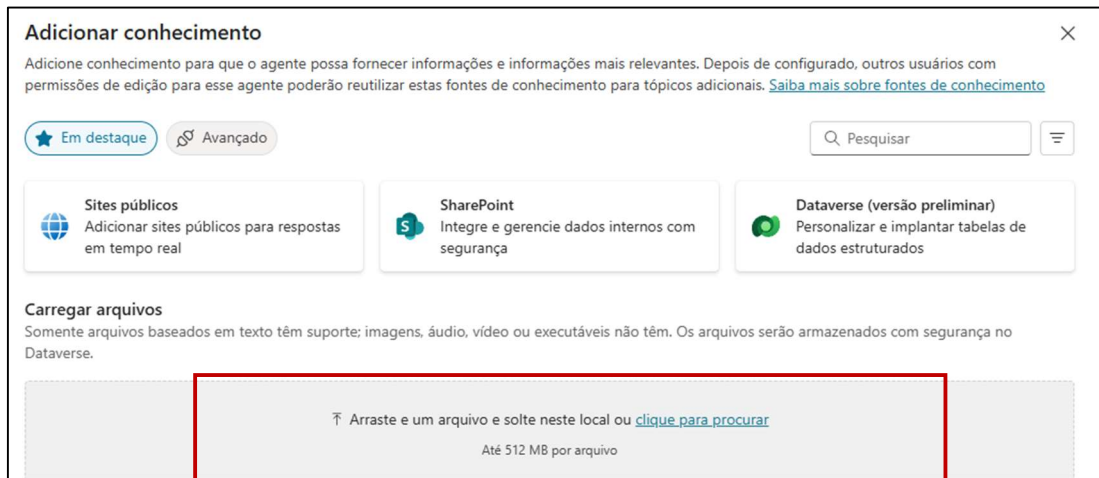
#### *4.1.1.1 Criação de Copilotos Personalizados com o Copilot Studio*

O Copilot Studio simplifica a criação de copilotos personalizados através de descrições intuitivas. Ele permite configurar instruções, gatilhos, fontes de dados e ações específicas, além de testar e publicar copilotos em diversos canais. Ideal para empresas que desejam integrar dados internos, APIs externas ou criar soluções personalizadas para fluxos de trabalho e aplicativos web ou móveis, com controle total sobre a marca. Desenvolvedores podem selecionar modelos de linguagem e personalizar orquestrações. Exemplos de aplicação incluem chatbots para SAC, assistentes virtuais para agendamentos e experiências de jogo com IA generativa.

Durante a criação de um agente no Copilot Studio, é possível incorporar arquivos como base de conhecimento, garantindo que o agente utilize essas informações de forma eficiente para fornecer respostas relevantes e contextualizadas. Esse processo, além de intuitivo, permite que a configuração do agente seja realizada de maneira rápida e eficaz, alinhando-o às necessidades específicas do projeto.

Para adicionar arquivos, o usuário deve acessar o painel do agente, criar ou editar um agente existente e selecionar a opção "Adicionar conhecimento". Os arquivos podem ser carregados por meio de upload manual ou arraste e solte. É solicitado que o arquivo seja nomeado, podendo manter o nome original caso não seja personalizada a identificação. Também é essencial fornecer uma descrição detalhada do conteúdo, especialmente em configurações com inteligência artificial generativa habilitada, a fim de assegurar o uso adequado e contextual das informações, figura 4.

Figura 4 - Ambiente inicial para adição de arquivos



Fonte: <https://copilotstudio.microsoft.com/environments/>

Uma vez concluídas as configurações, os arquivos são indexados e integrados à base de conhecimento do agente, tornando-se parte do repositório de dados que a ferramenta usará para responder a consultas. É recomendável que os arquivos sejam bem organizados e com formatação clara, otimizando o processamento e garantindo a precisão das respostas geradas figura 5.

Figura 5 - Ambiente de adição de arquivos

Nome	Tipo	Última modificação	Status
PRODUCAO_AMBULATORIAL_2023_12...	Arquivos	Gilmar Luis ...	Pronto(a)
PRODUCAO_AMBULATORIAL_2023_07...	Arquivos	Gilmar Luis ...	Pronto(a)
PRODUCAO_AMBULATORIAL_2023_08...	Arquivos	Gilmar Luis ...	Pronto(a)
PRODUCAO_AMBULATORIAL_2023_03...	Arquivos	Gilmar Luis ...	Pronto(a)
PRODUCAO_AMBULATORIAL_2023_11↓	Arquivos	Gilmar Luis ...	Pronto(a)

Fonte: <https://copilotstudio.microsoft.com/environments/>

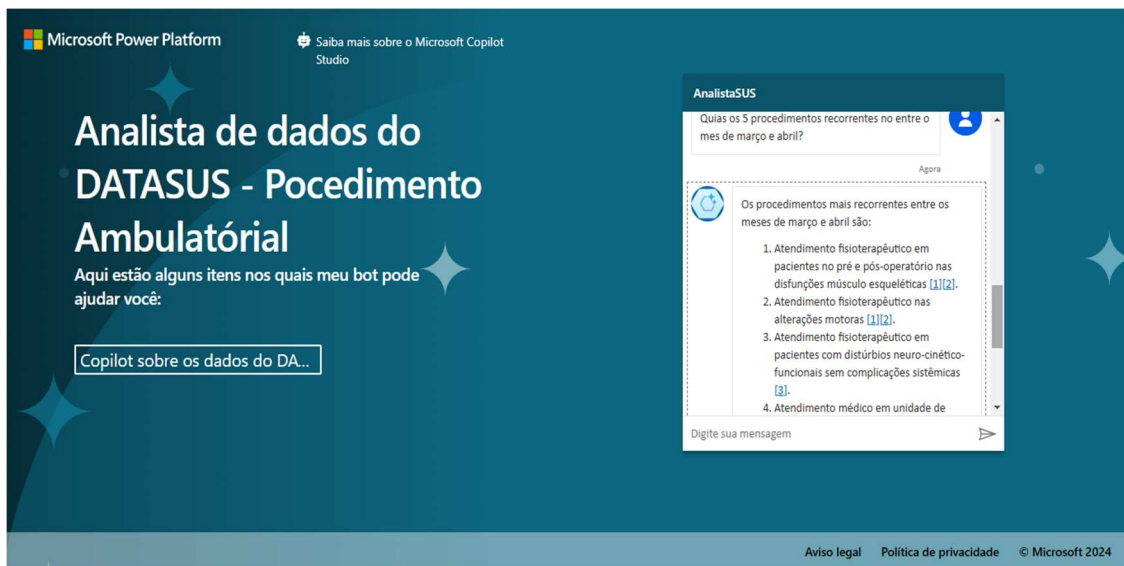
Esse procedimento é crucial para a eficácia do agente, pois assegura que ele seja alimentado com dados confiáveis e bem estruturados, promovendo resultados mais assertivos e alinhados aos objetivos do projeto. Assim, o processo de adição de arquivos no Copilot Studio

desempenha um papel central na construção de agentes inteligentes que podem atender a diversas demandas com precisão e eficiência.

A opção Habilitado(a) na aba Conhecimento, deve ser desabilitado para que a busca seja exclusivamente na base de conhecimento que foi fornecida. Essa configuração garante que o agente atue apenas com as informações definidas previamente.

Após essa configuração, é possível utilizar o site de demonstração para testar e visualizar como o Copilot irá se comportar nas interações. Certifique-se de que todas as definições estão corretas antes de concluir o processo para garantir que o agente funcione conforme esperado figura 6, (Studio, 2024).

Figura 6 - Site de demonstração



Fonte: <https://copilotstudio.microsoft.com/>

#### 4.1.2 Meu GPTs

O GPTs são versões personalizadas do ChatGPT, projetadas para atender objetivos específicos e permitir maior controle sobre como a inteligência artificial utilizada, tornando-se ferramentas úteis e sendo acessível a qualquer pessoa, e utilizados em diferentes contextos como educação, suporte técnico ou produtividade. O processo envolve fornecer orientações, incluir informações adicionais, podendo ser desenvolvidos para uso pessoal, empresarial ou compartilhar com outras pessoas. Atualmente os GPTs estão disponíveis para usuários do ChatGPT Plus e Enterprise, porém, o objetivo é expandir o acesso a essas ferramentas personalizáveis para mais usuários futuramente (OpenAi, 2023).

#### 4.1.2.1 Criação de Meu GPT Personalizados

No processo de criação de um Meu GPT, a adição de arquivos como base de conhecimento é uma etapa fundamental para garantir que o modelo seja capaz de responder de forma precisa e contextualizada às consultas dos usuários. Esse procedimento é realizado por meio de uma interface intuitiva que permite o carregamento de arquivos diretamente no ambiente de configuração do GPT personalizado.

Na seção "Carregar Arquivos" do **Meu GPT**, o usuário pode adicionar documentos que servirão como base de conhecimento para o modelo. Para isso, é necessário acessar a interface de configuração e selecionar a opção "Carregar Arquivos". É essencial fornecer uma descrição detalhada do conteúdo do arquivo, pois isso auxilia o GPT a utilizar as informações de forma mais precisa e contextualizada, figura 7.

Figura 7 – Ambiente de adição de arquivos do Meu GPT.



Fonte: <https://chatgpt.com/gpts>

O sistema processa os arquivos carregados e os indexa na base de conhecimento do GPT personalizado, tornando os dados acessíveis para a geração de respostas. Assim, é importante que os arquivos carregados estejam bem organizados e formatados de maneira a facilitar a extração e compreensão do conteúdo textual, figura 8.

Figura 8 - página inicial para alimentar um GPT.



Fonte: <https://chatgpt.com/gpts>

Esse procedimento permite que o Meu GPT seja configurado para atender a demandas específicas, utilizando dados confiáveis e estruturados de forma a proporcionar maior assertividade nas respostas. Dessa forma, a adição de arquivos no momento da criação do modelo é uma etapa essencial para assegurar a eficiência e a qualidade das interações futuras.

#### 4.1.3 Gemini Advanced

O avanço da inteligência artificial tem sido um pilar para a Google, influenciando significativamente suas soluções, como a Google Cloud<sup>9</sup>, Workspace<sup>10</sup>, Google One<sup>11</sup> entre outras. O lançamento da Gemini marcou um novo patamar na comparação da eficiência e da qualidade de texto, imagem e vídeo, estabelecendo-se como um ecossistema abrangente em

<sup>9</sup>Google Cloud é a plataforma de serviços de computação em nuvem do Google, projetada para empresas e desenvolvedores.

<sup>10</sup> Google Workspace (anteriormente chamado G Suite) é um conjunto de aplicativos de produtividade e colaboração baseados na nuvem, projetados para empresas e indivíduos

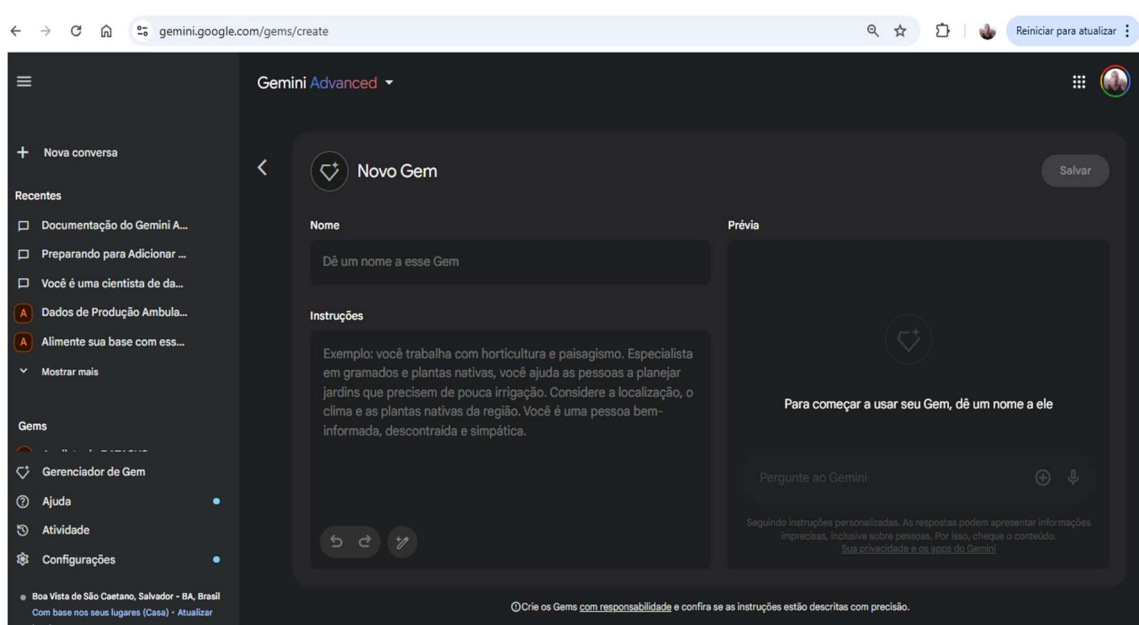
<sup>11</sup>Google One é um serviço de assinatura que oferece armazenamento na nuvem e benefícios adicionais para usuários individuais ou familiares

produtos amplamente utilizados e plataformas voltadas para desenvolvedores e organizações (Google, 2024).

#### 4.1.3.1 Criar um Gem personalizado.

No momento da criação de um agente no Gemini Advanced, a adição de arquivos como base de conhecimento é um procedimento essencial para configurar a ferramenta de maneira eficiente e alinhada às necessidades do projeto. Esse processo é projetado para ser detalhado e estruturado, garantindo que os dados inseridos sejam corretamente interpretados e utilizados pela inteligência artificial, figura 9.

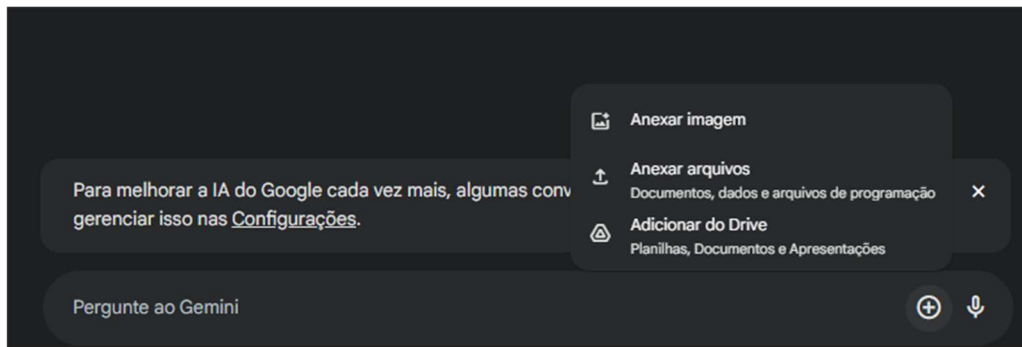
Figura 9 - Ambiente de criação do Gem



Fonte: <https://gemini.google.com/gems/create>

No Gemini Advanced, a adição de arquivos ocorre de maneira específica e está diretamente vinculada às interações realizadas nas conversas com o GEM. Diferentemente de outras plataformas, não é possível carregar arquivos diretamente na configuração inicial como base de conhecimento permanente. O usuário pode anexar arquivos diretamente ao fluxo da conversa. Para isso, há uma opção de upload ou compartilhamento de arquivos no painel de chat do GEM. Esses arquivos passam a ser processados pelo sistema em tempo real e são utilizados para responder às perguntas ou gerar análises baseadas no conteúdo do documento. É possível anexar arquivos ao Gem do dispositivo ou do Google Drive, clicando em anexar arquivos e escolher a opção, figura 10 (google, 2024).

Figura 10 - Ambiente para alimentar o Gem



Fonte: <https://gemini.google.com/gems/create>

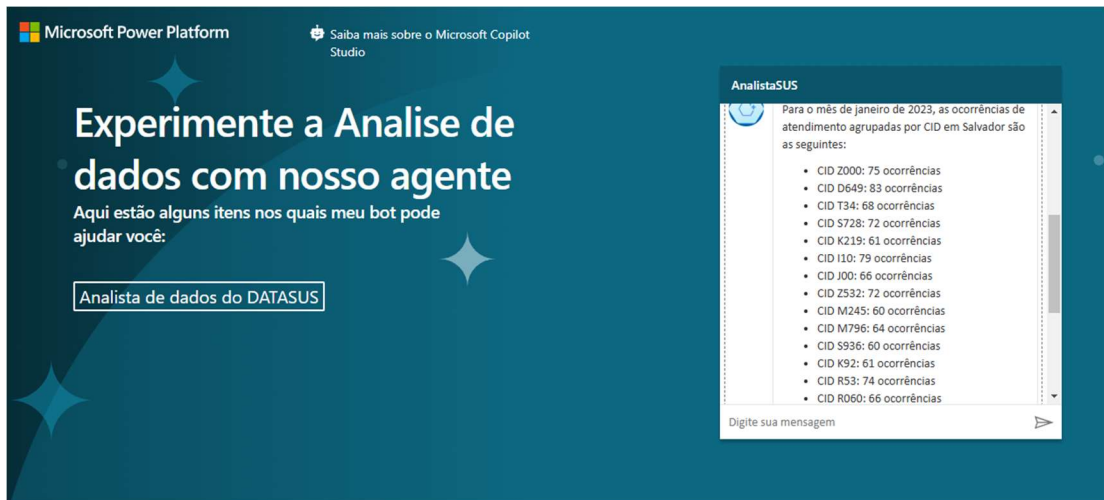
## 4.2 ANÁLISE DOS RESULTADOS

No contexto da análise de dados, o comportamento de diferentes ferramentas de inteligência artificial (IAs) pode variar significativamente, dependendo de suas capacidades e limitações técnicas. Para avaliar essas diferenças, foram analisadas três plataformas: **Copilot Studio**, **Meu GPTs** e **Gemini Advanced**. Cada uma dessas ferramentas foi submetida às mesmas perguntas e dados, permitindo observar suas particularidades em termos de precisão, flexibilidade, integração com bases de dados e capacidade de resposta.

### 4.2.1 Copilot Studio

O Copilot Studio destacou-se pela rapidez na geração de respostas e por sua abordagem de integração, permitindo buscas em múltiplos arquivos da base de dados, figura 11. Um diferencial importante da ferramenta é sua capacidade de aceitar a inserção de arquivos diretamente em sua base, o que amplia sua flexibilidade e facilita a análise de dados provenientes de diferentes fontes. Essa característica é particularmente vantajosa para análises que demandam uma visão mais ampla ou integrada, embora possa gerar resultados diferentes da base original em casos que exijam correspondência direta, figura 12.

Figura 11 - Site de demonstração - Amostra de resultado



Fonte: <https://copilotstudio.microsoft.com/>

Figura 12 - tabela dinâmica - resultado de pesquisa na base de dados

2	Meses (data_de_processamento)	jan
3		
4		<b>Contagem de quantidade_apresentada</b>
5	Z532	1
6	Z000	881
7	R53	18
8	M796	50
9	M245	2676
10	K92	1
11	I10	1140
12	D649	2
13	<b>Total Geral</b>	<b>4769</b>

Fonte: base de dados extraída do DATASUS

Além disso, o Copilot Studio demonstrou alta confiabilidade, com raras falhas, como problemas de conexão com a base de dados, que foram rapidamente resolvidos. A combinação de eficiência, baixa ocorrência de erros e capacidade de integração, somada à funcionalidade de inserção de arquivos, faz do Copilot Studio uma opção robusta e versátil para análises que exigem agilidade e abrangência. No entanto, a ferramenta não foi tão coerente com a base de dados quanto o esperado em algumas respostas geradas. Apesar disso, esses episódios foram pontuais e não comprometeram significativamente a usabilidade geral da ferramenta.

#### 4.2.2 Meu GPTs

A Meu GPTs apresentou respostas precisas e alinhadas aos dados da base original, demonstrando um desempenho consistente e confiável, figura 13. Esse alinhamento entre as perguntas formuladas e o conteúdo da base torna a ferramenta valiosa para estudos que exigem validação rigorosa dos dados, figura 14. Um diferencial importante é que a Meu GPTs permite a inserção de arquivos diretamente em sua base de dados, oferecendo flexibilidade para análises personalizadas e contínuas.

Figura 13- Amostra de resultado - Meu GPT

Quantidade De Ocorrências Por CID - Janeiro 2023		
	CID_principal	Quantidade_de_Ocorr
1	M255	15063
2	M999	8018
3	C61	4910
4	N180	4254
5	M199	4160
6	M624	3974

Fonte: <https://chatgpt.com/g/g-4t3e0ilkx-analista-do-datasus>

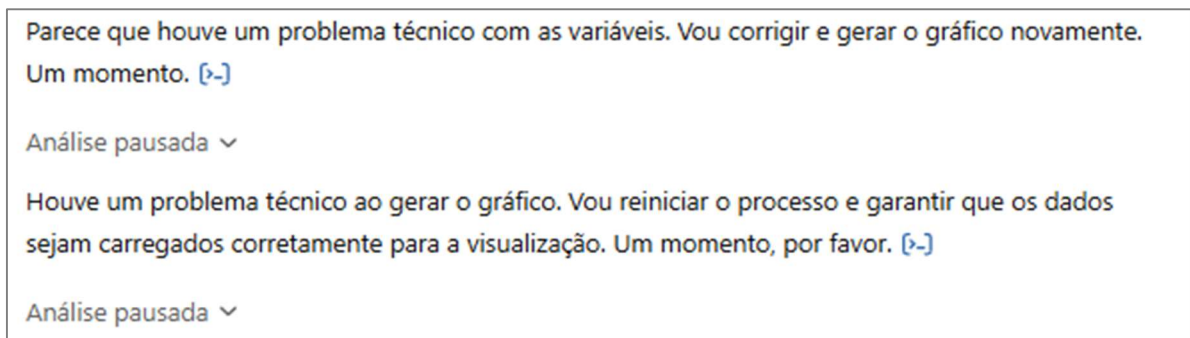
Figura 14 - tabela dinâmica - resultado de pesquisa na base de dados

Meses (data_de_processamento)	jan
	Contagem de quantidade_apresentada
M255	15063
M999	8018
C61	4910
N180	4254
M199	4160
M624	3974
G551	3138
H251	2900
M179	2805
M245	2676
J989	2618

Fonte: base de dados extraída do DATASUS

A ferramenta também apresentou limitações. Em diversas ocasiões, não respondeu de forma direta às perguntas, retornando mensagens como: "Vou realizar novamente o processamento e corrigir o problema. Caso esteja de acordo, prossigo e logo enviarei o relatório com as informações completas" ou "Houve um problema técnico ao gerar o gráfico. Vou reiniciar o processo e garantir que os dados sejam carregados corretamente para a visualização. Um momento, por favor", figura 15. Apesar de indicar a tentativa de corrigir erros ou reiniciar processos, a análise muitas vezes não era concluída, frustrando o usuário por não entregar a resposta esperada.

Figura 15 - Amostra de erro ao gerar resultado - Meu GPT



Fonte: <https://chatgpt.com/g/g-4t3e0ilkx-analista-do-datusus>

#### 4.2.3 Gemini Advanced

A Gemini Advanced demonstrou alta precisão na maioria das respostas, figura 16, apresentando resultados alinhados e em conformidade com os dados presentes na base consultada. Sua capacidade de processar e interpretar informações com exatidão reforça sua confiabilidade, especialmente em contextos que exigem fidelidade aos dados originais.

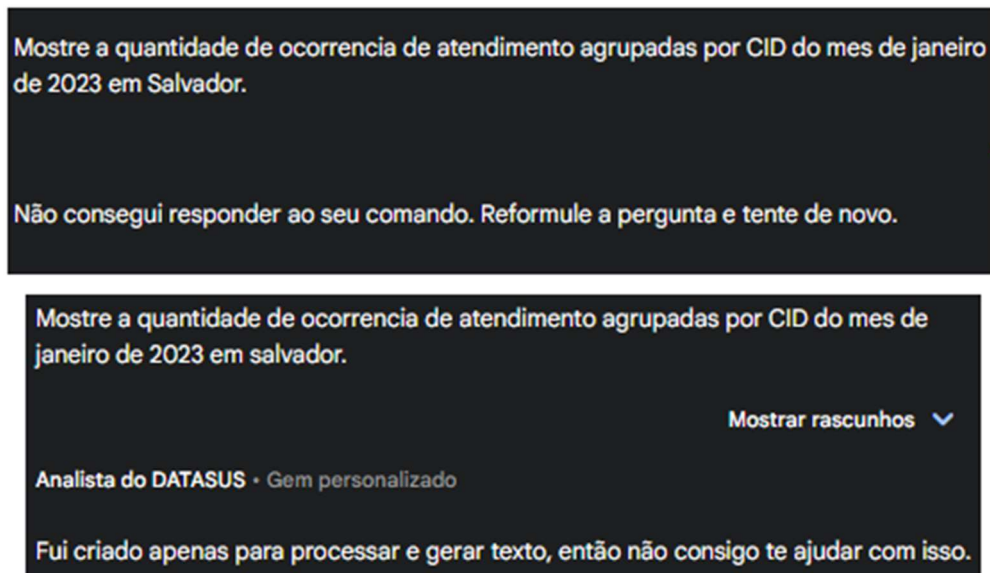
Figura 16 - Amostra de resultado - Gemini Advanced

CID_principal	Quantidade
M255	15063
M999	8018
C61	4910
N180	4254
M199	4160
M624	3974

Fonte: <https://gemini.google.com/gem>

No entanto, a ferramenta possui limitações significativas. Em alguns casos, não respondeu diretamente às perguntas formuladas, retornando mensagens genéricas como: "Sou apenas um modelo de linguagem e não posso te ajudar com isso" ou "Fui criado apenas para processar e gerar texto, então não consigo te ajudar com isso", figura 17. Além disso, cada documento precisa ser inserido diretamente na conversa em andamento, o que exige que o usuário carregue os arquivos novamente a cada nova interação. Essa característica pode reduzir a eficiência da ferramenta em projetos que demandam consultas frequentes ou contínuas.

Figura 17 - Amostra de erro ao gerar resultado - Gemini Advanced



Fonte: <https://gemini.google.com/gem>

### 4.3 MÉTODO DE COMPARAÇÃO

Esta seção traz a descrição do método utilizado para comparar as três inteligências artificiais analisadas: Copilot Studio, Gemini Advanced e Meu GPT. O objetivo dessa comparação é avaliar o desempenho, as funcionalidades e as limitações de cada uma delas, considerando os aspectos de velocidade de resposta, coerência com a base de conhecimento, aceitação de arquivos externos e confiabilidade nas respostas.

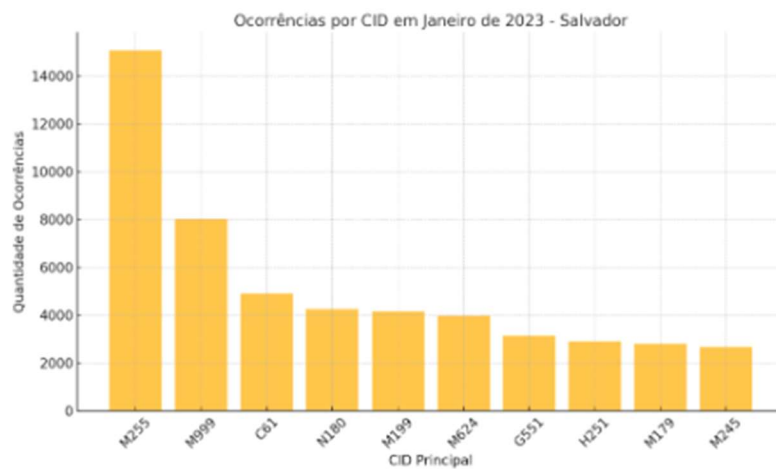
#### 4.3.1 Velocidade de Resposta

Entre as três inteligências artificiais, o Copilot Studio demonstrou ser a mais rápida com o tempo de 13 segundos. Comparativamente, o Gemini Advanced apresentou um tempo de resposta moderado entre 1 a 5 minutos, enquanto o Meu GPT teve o maior tempo de processamento ficando entre 1 a 8 minutos, especialmente em tarefas mais complexas. Essa diferença pode ser crítica dependendo do uso pretendido. Para situações que demandam respostas imediatas, o Copilot Studio é claramente superior. No entanto, a rapidez do Copilot Studio não se traduz necessariamente em qualidade das respostas, o que será explorado a seguir.

#### 4.3.2 Coerência com a Base de Conhecimento

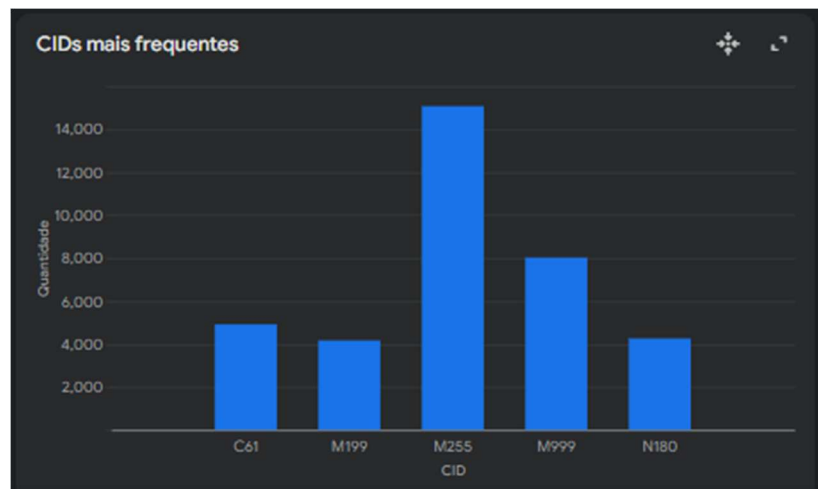
Em termos de coerência, o Meu GPT e o Gemini Advanced destacaram-se por fornecer respostas mais fiéis às informações presentes em suas bases de conhecimentos, figuras 18 e 19 respectivamente. Por outro lado, o Copilot Studio, apesar de aceitar arquivos para compor sua base de dados, apresentou respostas frequentemente incoerentes com os dados fornecidos, o que reduz sua confiabilidade.

Figura 18 - Gráfico da amostra de resultado - Meu GPT



Fonte: <https://chatgpt.com/g/g-4t3e0ilkx-analista-do-datasus>

Figura 19 - Gráfico da amostra de resultado - Gemini Advanced



Fonte: <https://gemini.google.com/gem>

Essa comparação revela que, embora o Meu GPT tenha um tempo de resposta mais lento, ele compensa com maior precisão junto com o Gemini Advanced, enquanto o Copilot Studio apresenta limitações que podem comprometer a interpretação dos dados.

#### 4.3.3 Aceitação de Arquivos Externos

Quanto à aceitação de arquivos externos, tanto o Copilot Studio quanto o Meu GPT oferecem a funcionalidade de incorporar arquivos à sua base de conhecimento. No entanto, há diferenças importantes: enquanto o Meu GPT mantém a base de conhecimento para múltiplas interações, o Copilot Studio é limitado pela qualidade das respostas incoerentes que gera a partir desses dados. Já o Gemini Advanced só aceita arquivos na conversa atual, sem a capacidade de construir ou reter uma base de conhecimento contínua, o que restringe sua utilidade em análises prolongadas, colocando-o em desvantagem frente aos outros dois modelos, especialmente em aplicações que exigem consulta repetida a uma base consolidada.

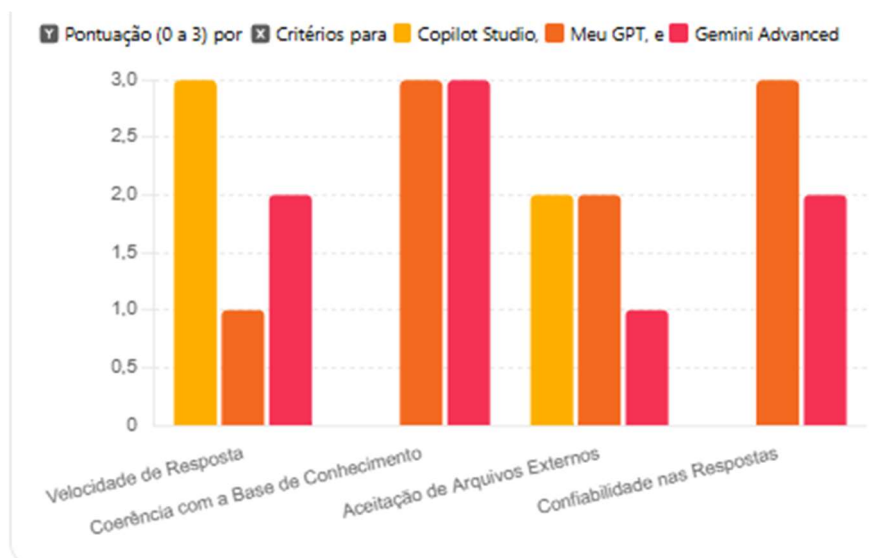
#### 4.3.4 Confiabilidade nas Respostas

Quando se trata de confiabilidade, o Meu GPT, mesmo que em um tempo maior, e o Gemini Advanced, são mais consistentes em entregar respostas alinhadas à base de dados fornecida, apesar de apresentarem algumas alucinações. Em contrapartida, o Copilot Studio, apesar de rápido, gera respostas incoerentes. Por fim, a confiabilidade do Meu GPT compensa seu tempo de resposta elevado, enquanto o Gemini Advanced é mais adequado para consultas rápidas e menos críticas, e o Copilot Studio dependendo do grau de precisão necessário.

#### 4.3.5 Gráfico Comparativo

Copilot Studio se destaca pela velocidade de resposta, mas apresenta baixa coerência e confiabilidade em algumas situações, apesar de aceitar arquivos externos. Meu GPT é mais lento, mas entrega respostas altamente coerentes e confiáveis, mantendo a base de conhecimento em múltiplas interações. Gemini Advanced oferece equilíbrio entre velocidade e precisão, mas limita a incorporação de arquivos a interações atuais e não é ideal para análises prolongadas. Cada sistema tem pontos fortes em diferentes aspectos, variando conforme a prioridade entre rapidez, coerência ou confiabilidade. A escolha depende das necessidades específicas de cada uso. Figura 20.

Figura 20 - Gráfico do resume dos critérios avaliados.



Fonte: adaptada da análise de resultados

## 5 CONSIDERAÇÕES FINAIS

A aplicação da inteligência artificial generativa na análise de dados públicos representa uma inovação promissora para a saúde pública no Brasil. Este trabalho avaliou o potencial da inteligência artificial generativa na análise de dados públicos do DATASUS, com foco na produção ambulatorial de Salvador, cumprindo o objetivo geral de explorar a viabilidade dessa tecnologia para melhorar o acesso e a interpretação de informações de saúde. Foram analisadas três ferramentas principais: Copilot Studio, Meu GPT e Gemini Advanced, utilizando critérios como velocidade, precisão, integração de dados e confiabilidade.

Os resultados mostraram que o Copilot Studio se destaca pela velocidade de resposta, embora apresente limitações na coerência e confiabilidade. O Meu GPT, apesar de ser mais lento, demonstrou maior precisão e consistência, permitindo uma base de conhecimento para múltiplas interações. O Gemini Advanced apresentou equilíbrio entre velocidade e precisão, mas possui limitações na integração de dados, restringindo sua aplicabilidade em análises prolongadas.

A resposta à pergunta do projeto destaca que a inteligência artificial generativa, quando aplicada no contexto analisado, apresenta potencial para transformar a acessibilidade e a eficiência na gestão de dados de saúde pública. Apesar das limitações observadas em cada ferramenta, penso que, pelos resultados apresentados, na verdade um longo caminho precisa ser percorrido para conseguir agilidade, confiabilidade e funcionalidade na aplicação de IA generativas em bases de dados estruturadas, ao menos a partir dos resultados apresentados no estudo dessas ferramentas comerciais. Assim, este estudo não apenas alcançou seus objetivos específicos, mas também abriu novas perspectivas para a utilização de IA generativa na saúde pública, contribuindo para uma gestão mais eficiente e justa no SUS.

## 6 TRABALHOS FUTUROS

Este projeto revelou novas oportunidades de pesquisa, indicando caminhos que podem aprofundar e ampliar os resultados alcançados. As sugestões apresentadas visam não apenas o aperfeiçoamento das soluções propostas, mas também a exploração de aplicações inovadoras no contexto da saúde pública. A seguir, são destacadas algumas perspectivas para investigações futuras.

Exploração de IA Generativa em Outras Bases de Dados de Saúde:

- Aplicar metodologias similares em outras bases de dados públicas de saúde (como SINASC, SIH ou SIM) para avaliar a aplicabilidade em diferentes contextos de saúde pública.

Impacto Ético e Regulatório da IA Generativa na Saúde Pública:

- Investigar como o uso de IA generativa em saúde pode impactar aspectos éticos, como privacidade, transparência e segurança dos dados.

Integração Multimodal de Dados de Saúde

- Explorar a integração de diferentes fontes de dados (ex.: dados hospitalares, dados ambulatoriais, e dados sociais) com IA generativa para análises mais abrangentes.

Essas propostas não se limitam à continuidade do sistema, mas também visam sua adaptação para atender a diferentes demandas da saúde pública, ampliando o impacto e a aplicabilidade da inteligência artificial generativa em diversos contextos.

## REFERÊNCIAS

- AFONSO, C.** *Cadernos Metrópole: a dignidade humana impactada por ambientes criados através de ferramentas digitais*. Brasília, 2023.
- BATISTA, F. D. M.; FILHO, D. P. C.** *Machine learning aplicado à saúde*. São Paulo: Sociedade Brasileira de Computação, 2019. Disponível em: <https://books-sol.sbc.org.br/index.php/sbc/catalog/download/29/95/245-1?inline=1>. Acesso em: 2023.
- CARDOSO, F; et al.** O uso da inteligência artificial na educação e seus benefícios: uma revisão exploratória e bibliográfica. *Ciência em Evidência: Revista Multidisciplinar*, São Paulo, 2023.
- CARRARO, F.** *Inteligência artificial e ChatGPT: da revolução dos modelos de IA generativa à engenharia de prompt*. [S.l.]: [s.n.], 2023. 223 p. Disponível em: <https://books.google.com.br/books?id=iKnoEAAAQBAJ>. Acesso em: 2024.
- CRESWELL, J. W.; CRESWELL, J. D.** *Research design: qualitative, quantitative, and mixed methods approaches*. 5. ed. Los Angeles: [s.n.], 2018.
- DAMACENO, S. S.; VASCONCELOS, O.** Inteligência artificial: uma breve abordagem sobre seu conceito real e o conhecimento popular. *Caderno de Graduação - Ciências Exatas e Tecnológicas - UNIT*, Sergipe, 2018. Disponível em: <https://periodicos.set.edu.br/cadernoexatas/article/view/5729>.
- FERRAZ, L. H. V. D. C.** *O SUS, o DATASUS e a informação em saúde: uma proposta de gestão participativa*. Dissertação (Mestrado) – Escola Nacional de Saúde Pública Sergio Arouca, Rio de Janeiro, 2009.
- GALVÃO, N. D.; MARIN, H. F.** Técnica de mineración de datos: una revisión de la literatura. *Acta Paul Enferm*, Cuiabá, 2009.
- GOOGLE.** Blog do Google Brasil. 2024. Disponível em: <https://blog.google/intl/pt-br/gemini/>. Acesso em: 23 out. 2024.
- GOOGLE.** Central de ajuda. 2024. Disponível em: <https://support.google.com/gemini/answer/15235603?hl=pt-BR>. Acesso em: 18 out. 2024.
- ISLAM, M. I.** Data analysis: types, process, methods, techniques and tools. *International Journal on Data Science and Technology*, 6, 2020.
- JUPYTER.** *Project Jupyter documentation*. Disponível em: <https://docs.jupyter.org/en/latest/index.html#more-information>. Acesso em: 1 nov. 2024.
- LOPEZOSA, [nome completo abreviado].** *La inteligencia artificial generativa en la comunicación científica: retos y oportunidades*. *Revista de Investigación e Innovación en Ciencias de la Salud*, Barcelona, 5, 2023. Acesso em: 2024.
- LUDERMIR, T. B.** Inteligência artificial e aprendizado de máquina: estado atual e tendências. *Estudos Avançados*, Recife, 35, 2021. p. 85–94.

- MICROSOFT.** *Microsoft Ignite. Learn.microsoft*, 2024. Disponível em: <https://learn.microsoft.com/pt-br/microsoft-copilot-studio/>. Acesso em: 20 out. 2024.
- OLIVEIRA, M. F. D.** *Metodologia científica: um manual para a realização de pesquisas em administração*. Catalão, Goiás, 2011. Disponível em: [https://files.cercomp.ufg.br/weby/up/567/o/Manual\\_de\\_metodologia\\_cientifica\\_-\\_Prof\\_Maxwell.pdf](https://files.cercomp.ufg.br/weby/up/567/o/Manual_de_metodologia_cientifica_-_Prof_Maxwell.pdf). Acesso em: 8 nov. 2023.
- OPENAI.** Posso usar o endpoint /v1/fine-tunes? *Openai*, 2022. Disponível em: <https://platform.openai.com/docs/guides/fine-tuning/fine-tuning-integrations>. Acesso em: 2024.
- OPENAI.** Apresentando GPTs. *Openai.com*, 2023. Disponível em: <https://openai.com/index/introducing-gpts/>. Acesso em: 15 ago. 2024.
- PAIXÃO, M. D. M.; et al.** Machine learning na medicina: revisão e aplicabilidade. *Artigo de revisão*, Belo Horizonte, 2022.
- PANDAS.** *Pandas documentation*. Disponível em: <https://pandas.pydata.org/docs/#module-pandas>. Acesso em: 28 out. 2024.
- PEFFERS, K.; TUUNANEN, T.; ROTHENBERGER, M. A.** A design science research methodology for information systems research. Las Vegas, 2007.
- PICCOLO, D. M.** Qualidade de dados dos sistemas de informação do Datasus: análise crítica da literatura. *Ciência da Informação em Revista*, 2018. Acesso em: 3 nov. 2023.
- PINTO, L. F.; ROCHA, C. M. F.** Inovações na atenção primária em saúde: o uso de ferramentas de tecnologia de comunicação e informação para apoio à gestão local. *Ciência & Saúde Coletiva*, Rio de Janeiro, maio 2016. p. 1433–1448.
- PYTHON.** *Python documentation*. Disponível em: <https://docs.python.org/pt-br/3/tutorial/>. Acesso em: 25 out. 2024.
- RAMOS, S. M.** Inteligência artificial generativa baseada em grandes modelos de linguagem: ferramentas de uso na pesquisa acadêmica. Universidade Federal do Rio Grande do Norte, Rio Grande do Norte, 2023.
- SALDANHA, R. D. F.; BASTOS, R. R.; BARCELLOS, C.** Microdatasus: pacote para download e pré-processamento de microdados do Departamento de Informática do SUS (DATASUS). *Cadernos de Saúde Pública*, v. 35, n. 9, p. e00032419, 2019.
- MINISTÉRIO DA SAÚDE.** *Datasus trajetória 1991-2002*. *Bvsms.saude*, 2002. Disponível em: [https://bvsms.saude.gov.br/bvs/publicacoes/trajetoria\\_datasus.pdf](https://bvsms.saude.gov.br/bvs/publicacoes/trajetoria_datasus.pdf). Acesso em: 27 out. 2023.
- SILVA, N. P. D.** *A utilização dos programas TABWIN e TABNET como ferramentas de apoio à disseminação das informações em saúde*. Dissertação (Mestrado) – ENSP-Sergio Arouca, FIOCRUZ, Rio de Janeiro, 2009.
- STUDIO, M. C.** *Microsoft Ignite. Microsoft Copilot Studio*, 2024. Disponível em: <https://learn.microsoft.com/pt-br/microsoft-copilot-studio/fundamentals-what-is-copilot-studio>. Acesso em: 1 nov. 2024.

**TORRES, R.; et al.** Aplicabilidade e potencialidades no uso de ferramentas de Business Intelligence na atenção primária em saúde. *Ciência & Saúde Coletiva*, 2021.

**TRINDADE, S. C. E. D.; OLIVEIRA, P. C. D.** Inteligência artificial (IA) generativa e competência em informação: habilidades informacionais necessárias ao uso de ferramentas de IA generativa em demandas informacionais de natureza acadêmico-científica. *Perspectivas em Ciência da Informação*, 2024.

**VALLE, V. C. L. L.; GASÓ, J. R. F. I.; AJUS, A. M.** Decisão judicial assistida por inteligência artificial e o sistema Victor do Supremo Tribunal Federal. *Revista de Investigações Constitucionais*, Curitiba, 2023. Acesso em: 2024.

**WELCHEN, [nome completo abreviado]; et al.** Usability perception of the health information systems in Brazil: the view of hospital health professionals on the electronic health record. Caxias do Sul, 2022.