UNIVERSIDADE ESTADUAL DA PARAÍBA CAMPUS VII
CENTRO DE CIÊNCIAS EXATAS E SOCIAIS APLICADAS
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

LUANDERSON BRUNO MARTINS SILVA

RISCOS E VULNERABILIDADES DE SEGURANÇA NA UTILIZAÇÃO DO CHAT GPT - UMA REVISÃO SISTEMÁTICA DA LITERATURA

PATOS
2023

LUANDERSON BRUNO MARTINS SILVA

RISCOS E VULNERABILIDADES DE SEGURANÇA NA UTILIZAÇÃO DO CHAT GPT - UMA REVISÃO SISTEMÁTICA DA LITERATURA

Trabalho de Conclusão de Curso apresentado ao Programa de Graduação em Ciência da Computação da Universidade Estadual da Paraíba - Campus VII.

Área de concentração: Inteligência Artificial

Orientador: Prof. Dra. Jannayna Domingues Barros Filgueira

PATOS
2023

S586r Silva, Luanderson Bruno Martins.
Riscos e vulnerabilidades de segurança na utilização do Chat GPT [manuscrito] : uma revisão sistemática da literatura / Luanderson Bruno Martins Silva. - 2023.
35 p. : il. colorido.

Digitado.
Trabalho de Conclusão de Curso (Graduação em Computação) - Universidade Estadual da Paraíba, Centro de Ciências Exatas e Sociais Aplicadas, 2023.
"Orientação : Profa. Dra. Jannayna Domingues Barros Filgueira, Departamento de Computação - CCT. "

1. Inteligência artificial. 2. Chat GPT. 3. Segurança da informação. I. Título

21. ed. CDD 005.3

Elaborada por Edson M. A. Monteiro - CRB - CRB 15/743 BSC7/UEPB

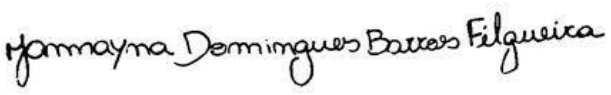
LUANDERSON BRUNO MARTINS SILVA

RISCOS E VULNERABILIDADES DE SEGURANÇA NA UTILIZAÇÃO DO CHAT GPT - UMA REVISÃO SISTEMÁTICA DA LITERATURA

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Ciência da Computação da Universidade Estadual da Paraíba, em cumprimento à exigência para obtenção do grau de Bacharel em Ciência da Computação.

Aprovado em 20/11/2023

BANCA EXAMINADORA



Profa. Dra.. Jannayna Domingues Barros Filgueira
(Orientadora)

Prof. Dra. Rosangela de Araújo Medeiros
(Examinadora)

Prof. Me. Ingrid Morgane Medeiros de Lucena
(Examinadora)

RESUMO

Este trabalho propõe uma análise aprofundada dos riscos e vulnerabilidades associados à utilização do modelo de linguagem GPT (Generative Pre-trained Transformer), focando especificamente na implementação do ChatGPT. A pesquisa inicia-se com uma revisão sistemática da literatura (RSL), conduzida para extrair o estado da arte em segurança da informação relacionada ao ChatGPT. A metodologia envolveu a definição strings de busca específicas, explorando a segurança da informação no contexto do ChatGPT. A RSL incluiu critérios rigorosos de inclusão, abrangendo artigos completos e resumos estendidos que investigam o uso, riscos ou vulnerabilidades associados ao ChatGPT. Os resultados desta revisão sistemática fornecem uma visão abrangente das abordagens de segurança adotadas, identificando lacunas no conhecimento e propondo potenciais áreas de pesquisa futura. Este estudo contribui para o entendimento crítico dos desafios de segurança na implementação de tecnologias baseadas em ChatGPT, apresentando uma base sólida para discussões e avanços na área de segurança da informação.

Palavras-chave: ChatGPT; Segurança da Informação; Revisão Sistemática da Literatura; Vulnerabilidades.

ABSTRACT

This paper proposes an in-depth analysis of the risks and vulnerabilities associated with the use of the GPT (Generative Pre-trained Transformer) language model, focusing specifically on the implementation of ChatGPT. The research begins with a Systematic Literature Review (SLR), conducted to extract the state of the art in information security related to ChatGPT.
The methodology involved defining a specific search string, exploring information security in the context of ChatGPT. The SLR included rigorous inclusion criteria, encompassing full articles and extended abstracts that investigate the usage, risks, or vulnerabilities associated with ChatGPT. The results of this systematic review provide a comprehensive overview of the security approaches adopted, identifying knowledge gaps, and proposing potential areas for future research. This study contributes to a critical understanding of security challenges in the implementation of ChatGPT-based technologies, providing a solid foundation for discussions and advancements in the field of information security.

Keywords: ChatGPT; Information Security; Systematic Literature Review; Vulnerabilities.

LISTA DE ILUSTRAÇÕES

LISTA DE SIGLAS

LLM	Large Language Model
IA	Inteligência Artificial
GPT	Generative Pre-trained Transformer
RSL	Revisão Sistemática da Literatura
API	Application Programming Interface
TI	Tecnologia da Informação
RSL	Revisão Sistemática da Literatura
QP	Questão primária
QS	Questão secundária
DDos	Distributed Denial of Service

SUMÁRIO

1 INTRODUÇÃO

A comunicação entre humanos e sistemas de inteligência artificial (IA) tem se tornado cada vez mais comum e amplamente utilizada em diversos setores, impulsionada pelos avanços no campo da IA. Dentre as abordagens de IA, podemos destacar o uso dos Chatbots, que é um programa de computador que utiliza técnicas de processamento de linguagem natural para interagir com usuários humanos por meio de uma interface de *Chat* (bate-papo ou conversa). Esses podem ser programados para responder perguntas comuns, fornecer informações relevantes, realizar tarefas específicas e até mesmo aprender com as interações dos usuários. Os Chatbots são projetados para simular uma conversa humana e podem ser usados para diversos fins, como atendimento ao cliente, assistência virtual, educação, suporte técnico.

Atualmente um dos Chatbots mais populares é o ChatGPT, desenvolvido pela OpenAI que é um modelo de linguagem grande (LLM) “Large Language Model”, treinado em um enorme conjunto de dados de texto e código, em particular, é uma das implementações de chatbot baseada em inteligência artificial mais proeminentes. Ele utiliza o modelo de linguagem GPT (Generative Pre-trained Transformer) para gerar respostas em linguagem natural (Radford *et al*., 2019). No entanto, a utilização de chatbots, incluindo o ChatGPT, não está isenta de riscos e vulnerabilidades de segurança. Diversos estudos têm explorado os desafios associados à segurança e privacidade dos usuários na utilização dessas tecnologias (Smith, 2020; Nguyen *et al*., 2021). Os chatbots podem ter acesso a informações pessoais e privadas dos usuários, o que levanta preocupações sobre a proteção desses dados e possíveis violações de privacidade (Jones *et al*., 2019).

Os sistemas de tecnologia são complexos e possuem uma ampla gama de vulnerabilidades de segurança. Essas vulnerabilidades podem ser exploradas por atacantes para roubar dados, causar danos ou comprometer a operação do sistema. O ChatGPT é um modelo de linguagem grande, treinado em um enorme conjunto de dados de texto e código. Isso torna o ChatGPT um alvo atraente para atacantes, que podem usá-lo para gerar conteúdo malicioso ou explorar vulnerabilidades de segurança no ChatGPT ou nos sistemas com os quais ele interage.

Os riscos e vulnerabilidades de segurança na utilização do ChatGPT podem ter um impacto significativo nos usuários e sistemas envolvidos. Os usuários podem ser expostos a ataques de *phishing* (Fraude online usada para enganar pessoas por meio de mensagens ou sites falsos), *malware* (Software malicioso projetado para causar danos a um computador ou

rede) ou outras formas de ataques cibernéticos. Os sistemas podem ser comprometidos, resultando em roubo de dados, danos ou interrupção da operação.

Um dos principais desafios na utilização do ChatGPT é a falta de transparência e clareza do sistema. Os usuários muitas vezes não têm conhecimento sobre como as decisões são tomadas pelo ChatGPT e quais são os critérios utilizados para gerar suas respostas (Dhingra & Singh, 2021). Isso pode gerar desconfiança e preocupações éticas sobre a responsabilidade pelos resultados produzidos pelo ChatGPT.

Compreender os impactos dos riscos e vulnerabilidades de segurança na utilização do ChatGPT é fundamental para garantir a confiabilidade e proteção dos usuários e sistemas envolvidos. A identificação e análise desses riscos específicos são essenciais para o desenvolvimento de medidas de mitigação adequadas, bem como para orientar pesquisas futuras e promover uma utilização mais segura e confiável dessa tecnologia. No entanto, é importante ressaltar que essas técnicas e abordagens podem não ser suficientes para mitigar todos os riscos de segurança associados ao uso do ChatGPT. Ainda há muito a ser explorado e compreendido sobre os desafios e limitações dessa tecnologia (Yao *et al*., 2021).

Deste modo, esta pesquisa apresenta uma revisão sistemática da literatura sobre os riscos e vulnerabilidades do ChatGPT na segurança da informação. O objetivo é entender os riscos associados ao uso inadequado da ferramenta, bem como as vulnerabilidades existentes na mesma.

Esta revisão tem como objetivo analisar os impactos dos riscos de segurança e vulnerabilidades identificados até o momento na utilização do ChatGPT por usuários individuais e empresariais. Esta revisão sistemática da literatura (RSL) foi realizada de acordo com os procedimentos propostos por Keele (2007), ou seja, a revisão percorreu nos seguintes passos: elaboração de questionamentos, busca por estudos, seleção dos estudos, análise dos dados e condensar os resultados.

Com o intuito de selecionar apenas os estudos relevantes para esta RSL, foi necessário estabelecer critérios claros e objetivos. Esses critérios foram baseados nas perguntas de pesquisa estabelecidas. Dessa forma, os estudos que não atendem aos critérios estabelecidos foram descartados, garantido que apenas os estudos relevantes e de qualidade fossem incluídos na revisão. Isso é importante para garantir que os resultados obtidos na RSL sejam confiáveis e possam ser usados para gerar novos conhecimentos sobre o tópico em questão.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como finalidade apresentar os principais conceitos e aspectos relevantes ao entendimento da vulnerabilidade e segurança da informação no uso de chatbots, como o ChatGPT, principal objeto de estudo desta pesquisa, bem como uma abordagem sobre revisão sistemática da literatura.

2.1 Segurança da Informação

A vulnerabilidade em segurança da informação é um tópico crítico no cenário tecnológico atual. Ela se refere a fraquezas e lacunas em sistemas de segurança que podem ser exploradas por ameaças, como hackers, malwares ou outras formas de ataques cibernéticos. É fundamental compreender as dimensões da vulnerabilidade e as medidas necessárias para proteger ativos valiosos.

Existem diversos tipos de vulnerabilidades, desde aquelas relacionadas a softwares desatualizados ou mal configurados até aquelas decorrentes de falhas humanas, como senhas fracas ou compartilhamento inadequado de informações. Um computador comprometido por *malware* pode ser usado por criminosos cibernéticos para diversos fins (Kaspersky, *s.d)*. Entre eles, roubar dados confidenciais, usar o computador para realizar outros atos criminosos ou causar danos aos dados. Identificar e reduzir vulnerabilidades é um processo contínuo, pois novas ameaças e fraquezas estão sempre surgindo.

Para proteger contra vulnerabilidades, as organizações implementam estratégias de segurança que incluem políticas de acesso, atualizações regulares de software, monitoramento de rede, criptografia e treinamento de conscientização em segurança para os funcionários. Além disso, a colaboração com a comunidade de segurança cibernética e o uso de ferramentas de detecção de ameaças são práticas comuns.

2.1.1 Ataques de Negação de Serviço (DDos)

Os ataques de Negação de Serviço (DDoS) visam sobrecarregar um sistema, rede ou serviço online, tornando-o inacessível para usuários legítimos. Essa sobrecarga é realizada ao inundar o alvo com um volume massivo de tráfego, superando sua capacidade de processamento.

Principais Características:

1. **Volume Massivo de Tráfego**: Utilização de uma grande quantidade de solicitações ou tráfego para saturar a largura de banda do alvo.
2. **Diversidade de Fontes:** Os ataques frequentemente envolvem uma ampla variedade de dispositivos comprometidos, formando uma botnet.
3. **Variedade de Técnicas:** Incluem ataques de amplificação, exaustão de recursos, e ataques distribuídos coordenados.

2.1.2 Vulnerabilidade 0-day

A vulnerabilidade 0-day refere-se a uma falha de segurança em um software que é explorada por cibercriminosos antes que o desenvolvedor tenha tido a oportunidade de criar e disponibilizar uma correção (patch). O termo "0-day" sugere que os desenvolvedores têm "zero dias" para abordar essa vulnerabilidade antes que ela seja explorada.

Principais Características:

1. **Ausência de Solução Conhecida:** Não há uma correção ou atualização disponível no momento em que a vulnerabilidade é explorada.
2. **Exploração Imediata:** Os ataques ocorrem assim que a falha é descoberta, aproveitando-se da falta de proteção.

Riscos Associados:

1. **Exposição Prolongada:** O período sem uma solução pode permitir ataques contínuos e persistentes.
2. **Danos Extensivos:** Dependendo da natureza da vulnerabilidade, os danos podem variar de vazamento de dados a comprometimento total do sistema.

Medidas de Mitigação:

1. Monitoramento Constante: Vigilância ativa para identificar comportamentos suspeitos que possam indicar exploração.
2. Implementação de Controles de Segurança: Firewalls, sistemas de detecção de intrusões e atualizações regulares são essenciais.

2.2 ChatGPT

O GPT (Generative Pré-trained Transformer) é um marco significativo na área de processamento de linguagem natural (PLN) que tem revolucionado a capacidade de sistemas de inteligência artificial em compreender e gerar texto de maneira coesa e contextual. Este modelo é desenvolvido pela OpenAI, uma organização líder em pesquisa e desenvolvimento de IA, e é amplamente utilizado em uma variedade de aplicações, incluindo chatbots, assistentes virtuais, resumos automáticos, tradução de idiomas e muito mais.

De acordo com a documentação da OpenAI (OpenAI, s.d.), uma característica distintiva do GPT é o seu treinamento prévio em enormes quantidades de dados textuais da internet, o que lhe confere uma compreensão abrangente da linguagem humana. Esse pré-treinamento é realizado em um modelo de rede neural conhecido como Transformer, que se destaca na captura de relações contextuais em texto.

O GPT é altamente flexível e pode ser adaptado para diversas tarefas, incluindo a geração de texto coerente a partir de prompts específicos. De acordo com a documentação da OpenIA, isso é feito com a API de Completions do Chat GPT (OpenIA, s.d), que permite interações naturais com o modelo, tornando-o adequado para criar assistentes virtuais, chatbots e sistemas de resposta automática.

A IA, apesar de seu vasto potencial, enfrenta uma série de desafios que podem complicar sua implementação e uso eficaz. Abaixo, destacamos alguns dos desafios mais comuns associados à incorporação da IA em organizações.

A governança de dados é um aspecto crítico na implementação da IA. É imperativo que as políticas de governança de dados estejam alinhadas com as regulamentações e leis de privacidade. Ao adotar a IA, a organização assume a responsabilidade de gerenciar a qualidade, privacidade e segurança dos dados. A proteção dos dados e da privacidade dos clientes deve ser uma prioridade. Para garantir a segurança dos dados, a organização deve ter uma compreensão clara de como os modelos de IA utilizam e interagem com os dados do cliente em todos os níveis.

A implementação da IA com aprendizado de máquina (machine learning) exige recursos substanciais. Recursos de processamento de alto desempenho são essenciais para o funcionamento eficaz de tecnologias de aprendizado profundo. Isso implica a necessidade de uma infraestrutura computacional robusta para a execução de aplicações de IA e o treinamento de modelos. No entanto, o poder de processamento pode ser dispendioso e limitar a escalabilidade dos sistemas de IA.

Para treinar sistemas de IA de maneira justa e precisa, é necessário alimentá-los com grandes volumes de dados. Isso requer uma capacidade de armazenamento adequada para lidar com o processamento e gerenciamento dos dados de treinamento. Além disso, é essencial estabelecer processos eficazes de gerenciamento e garantia da qualidade dos dados para assegurar a precisão das informações utilizadas no treinamento.

Enfrentar esses desafios é essencial para aproveitar plenamente os benefícios da IA. Cada organização deve abordar essas questões com cuidado, desenvolvendo estratégias sólidas de governança de dados, investindo em recursos técnicos adequados e estabelecendo práticas de gestão de dados eficientes. Ao superar esses desafios, a IA pode se tornar uma ferramenta poderosa para impulsionar a inovação e melhorar a eficiência operacional em diversas áreas de atuação.

2.3 Golpes

É observado em fóruns clandestinos de hackers que cibercriminosos iniciantes exploram a capacidade do ChatGPT para auxiliar na criação de Trojans. O chatbot tem a habilidade de gerar código com base em descrições sucintas de funções desejadas, como "coletar senhas e enviá-las via HTTP POST para um servidor específico." Isso, em teoria, permitiria a criação de *info stealers* (malware projetado para coletar informações confidenciais e sensíveis), sem a necessidade de habilidades avançadas de programação.

No caso de código gerado pelo ChatGPT ser efetivamente utilizado, as soluções de segurança estão preparadas para detectá-lo e neutralizá-lo com a mesma eficiência demonstrada com *malwares* criados por seres humanos. Além disso, se o código não passar por uma revisão rigorosa por parte de programadores experientes, é provável que contenha erros sutis e falhas lógicas, comprometendo sua eficácia.

No momento, é importante destacar que os chatbots, pelo menos por enquanto, representam uma concorrência limitada para programadores de malware iniciantes. É crucial manter em mente que o uso indevido de tecnologias como o ChatGPT é passível de detecção e de combate pelas ferramentas e práticas de segurança cibernética existentes, visando a proteção da integridade dos sistemas e dados.

As versões iniciais de IA baseada em linguagem têm sido de código aberto e amplamente acessíveis ao público por um período considerável. O ChatGPT é uma iteração significativamente mais avançada, destacando-se pela sua capacidade de interação quase perfeita com os usuários, com notável ausência de erros ortográficos, gramaticais e de linguagem. Esse nível de fluência pode dar a impressão de que do outro lado da janela de chat

está uma pessoa real. Do ponto de vista da comunidade de hackers, o ChatGPT representa uma mudança substancial.

Segundo o Relatório de Crimes na Internet de 2021 do FBI (FBI, 2021 - Disponível em https://www.ic3.gov/Media/PDF/AnnualReport/2021_IC3Report.pdf), o phishing emergiu como a ameaça de segurança de TI mais comum nos Estados Unidos. No entanto, muitos golpes de phishing são frequentemente identificáveis devido a erros ortográficos, gramática deficiente e frases desajeitadas, especialmente quando originados de regiões onde o inglês não é a língua nativa dos criminosos. O ChatGPT potencialmente capacitará hackers em todo o mundo com uma proficiência quase nativa em inglês, ampliando suas campanhas de phishing.

Para profissionais de segurança cibernética, a crescente sofisticação dos ataques de phishing exige uma resposta imediata e soluções práticas. Os líderes de segurança devem fornecer às equipes de TI ferramentas capazes de distinguir conteúdo gerado pelo ChatGPT de comunicações geradas por humanos, especialmente em e-mails não solicitados.

Ao utilizar o ChatGPT, é crucial considerar os riscos e vulnerabilidades associados. Estas incluem a possibilidade de gerar conteúdo impróprio ou enganoso, bem como potenciais problemas de privacidade ao lidar com informações sensíveis.

A Engenharia Social representa uma ameaça significativa no cenário cibernético, e o ChatGPT emerge como uma ferramenta potencialmente poderosa nas mãos de cibercriminosos. A habilidade do ChatGPT de interagir de maneira quase humana durante conversas o torna uma escolha atraente para a execução de golpes baseados em engenharia social. Pois, muitos desses atacantes não têm a mesma nacionalidade que as vítimas, então o ChatGPT pode ser muito útil na tradução para a linguagem de origem da vítima e o criminoso tenha êxito na ação. Portanto, a utilização responsável e ética do GPT é de extrema importância.

2.3.1 Código malicioso com ChatGPT

Embora o ChatGPT tenha demonstrado proficiência na geração de código e ferramentas de programação, é importante destacar que a IA é programada para evitar a criação de código malicioso ou destinado a atividades de hacking. Quando um usuário solicita tal código, o ChatGPT reforça seu compromisso em "ajudar com tarefas úteis e éticas, aderindo a diretrizes e políticas éticas."

No entanto, a manipulação do ChatGPT é uma possibilidade, e indivíduos mal-intencionados podem buscar estratégias criativas para enganar a IA a fim de gerar código malicioso. De fato, relatos indicam que hackers já estão explorando essa possibilidade. A empresa de segurança israelense Check Point identificou um caso em que um hacker, em um fórum clandestino, afirmou estar testando o chatbot para recriar malware (Analytics Insight, s.d.). É provável que essas tentativas estejam ocorrendo de forma disseminada nas áreas mais obscuras da internet.

Em resposta a essas crescentes ameaças, os profissionais de segurança cibernética precisam de treinamento constante e recursos atualizados para lidar com ameaças originadas de IA ou de outros vetores. Além disso, há uma oportunidade para capacitar esses profissionais com tecnologia de IA própria, que possa identificar e defender contra códigos de hacking gerados por IA. Embora o foco público frequentemente recaia sobre o potencial uso malicioso da IA, é crucial lembrar que essa mesma tecnologia está disponível para fins benéficos.

Portanto, o treinamento em segurança cibernética deve abordar não apenas a prevenção de ameaças relacionadas ao ChatGPT, mas também a exploração das capacidades da IA como uma ferramenta valiosa no arsenal dos profissionais de segurança cibernética. À medida que a rápida evolução tecnológica dá origem a novos desafios de segurança cibernética, é essencial que examinemos essas possibilidades e desenvolvamos programas de treinamento adequados. Além disso, os desenvolvedores de software devem considerar o desenvolvimento de IA generativa projetada especificamente para auxiliar os Centros de Operações de Segurança (SOCs) preenchidos por equipes de segurança humanas.

3 METODOLOGIA

Para atingir o objetivo deste estudo, que consiste na análise abrangente dos riscos e vulnerabilidades associados à utilização do ChatGPT, procedemos com uma investigação cuidadosa. Esta pesquisa teve como ponto de partida uma busca minuciosa de estudos relacionados à segurança da informação no contexto do ChatGPT. Essa busca revelou uma área de pesquisa em constante evolução, porém com escassas propostas de mitigação de potenciais riscos associados à aplicação do ChatGPT.

Com o intuito de extrair o estado da arte em segurança da informação no contexto do ChatGPT e avaliar a perspectiva de segurança presente em tais estudos, foi conduzida uma Revisão Sistemática da Literatura (RSL). A seleção criteriosa de estudos relevantes para esta RSL envolveu a realização de uma pesquisa exploratória (conforme detalhado na Seção 3.1) e o estabelecimento de critérios de inclusão. Esses critérios foram estruturados com base nas questões de pesquisa formuladas (conforme descrito na Seção 3.3).

Segundo as considerações de Keele (2007), a Revisão Sistemática da Literatura desempenha um papel central ao resumir as evidências já existentes acerca de um tema específico, destacar as lacunas presentes nos estudos e fornecer um alicerce sólido para investigações futuras. É relevante notar que a elaboração de uma RSL demanda a formulação de questionamentos pertinentes ao tema, a definição de critérios para inclusão e exclusão de conteúdo bibliográfico, a avaliação da qualidade dos estudos selecionados e, por fim, a síntese e apresentação dos resultados [Keele, 2007].

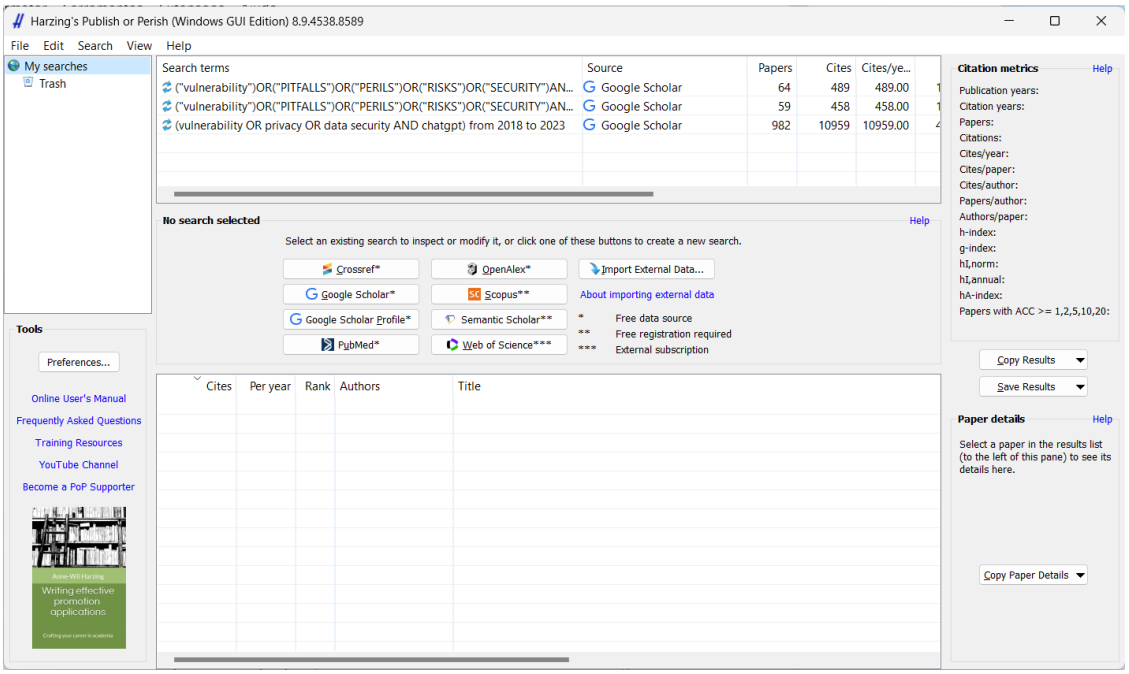
A Revisão Sistemática da Literatura se distingue das revisões tradicionais, que tendem a ser mais abrangentes e generalistas em relação a um tema, assim como das revisões integrativas, que frequentemente refletem a opinião subjetiva do autor. Consequentemente, as RSLs são classificadas como estudos secundários, tendo como base de dados os estudos primários que compõem suas fontes. Os métodos empregados na elaboração de uma RSL englobam uma série de etapas, entre as quais se destacam: I - a formulação da pergunta de pesquisa, II - a busca e seleção criteriosa dos artigos, III - a extração meticulosa dos dados, IV - a síntese criteriosa dos resultados, V - a avaliação rigorosa das evidências e VI - a redação cuidadosa e a subsequente publicação dos resultados.

3.1 Pesquisa Exploratória

Utilizamos o método de pesquisa onde partimos de uma investigação exploratória sobre o tema em estudo e logo após evoluiu para uma Revisão Sistemática da Literatura, para estudos de casos e experimentos futuros.

Com o objetivo de identificar os principais estudos e autores relacionados ao tema de nossa pesquisa, empregou-se uma abordagem metodológica de cunho bibliométrico. Para este fim, utilizou-se a ferramenta de análise bibliométrica denominada "Publish or Perish" de Harzing (HARZING, 2016).

Figura 01 - Software Publish or Perish



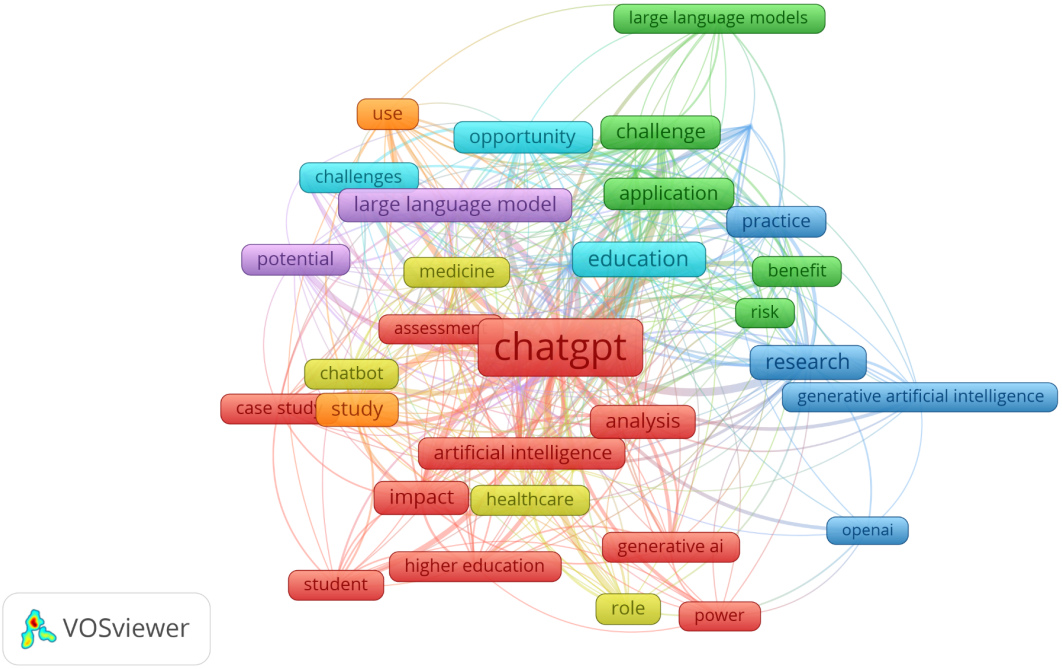
Fonte: Autoria Própria

O propósito dessa ferramenta foi o de conceber uma estrutura de investigação inicial a ser aplicada nas bases de dados, a fim de analisar os parâmetros pertinentes a este estudo, incluindo o número de citações e a classificação por citações.

Após a conclusão dessa etapa, conduziram-se levantamentos com a utilização de termos pertinentes à pesquisa, tendo como referência a base de dados do Google Acadêmico. A pesquisa se pautou na aplicação de uma combinação lógica de palavras-chave. Dessa investigação resultou a seguinte expressão: ("vulnerability") OR ("privacy") OR ("data security") AND ("chatgpt").

A análise dos termos extraídos dos estudos culminou na criação de um mapa de palavras, no qual foram destacadas as palavras-chave que apareceram em mais de dez trabalhos, conforme exemplificado na Figura 1.

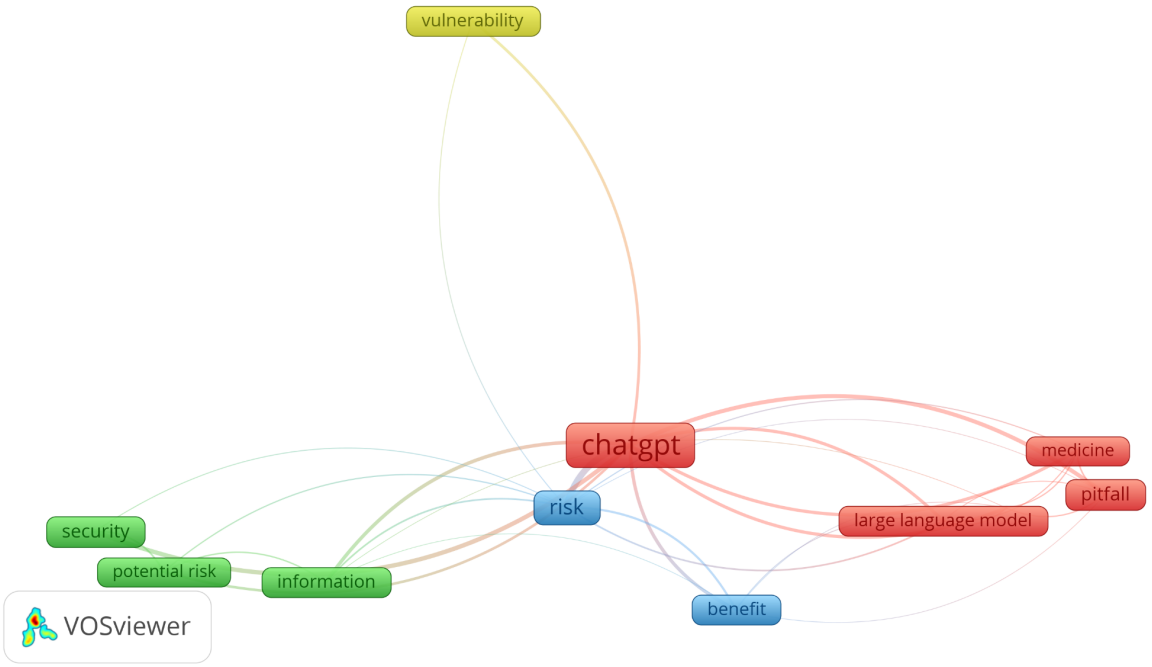
Figura 02 - Mapa de palavras da pesquisa exploratória



Fonte: Autoria Própria

Em seguida, a pesquisa avançou com a aplicação da expressão a seguir para a filtragem dos títulos dos estudos: ("vulnerability") OR ("pitfalls") OR ("perils") OR ("risks") OR ("security") AND ("chatgpt"). Novamente, um mapa de palavras foi gerado, respeitando o critério de ocorrências semelhante ao utilizado anteriormente.

Figura 03 - Mapa de palavras contendo o filtro de título dos trabalhos.



Fonte: Autoria Própria

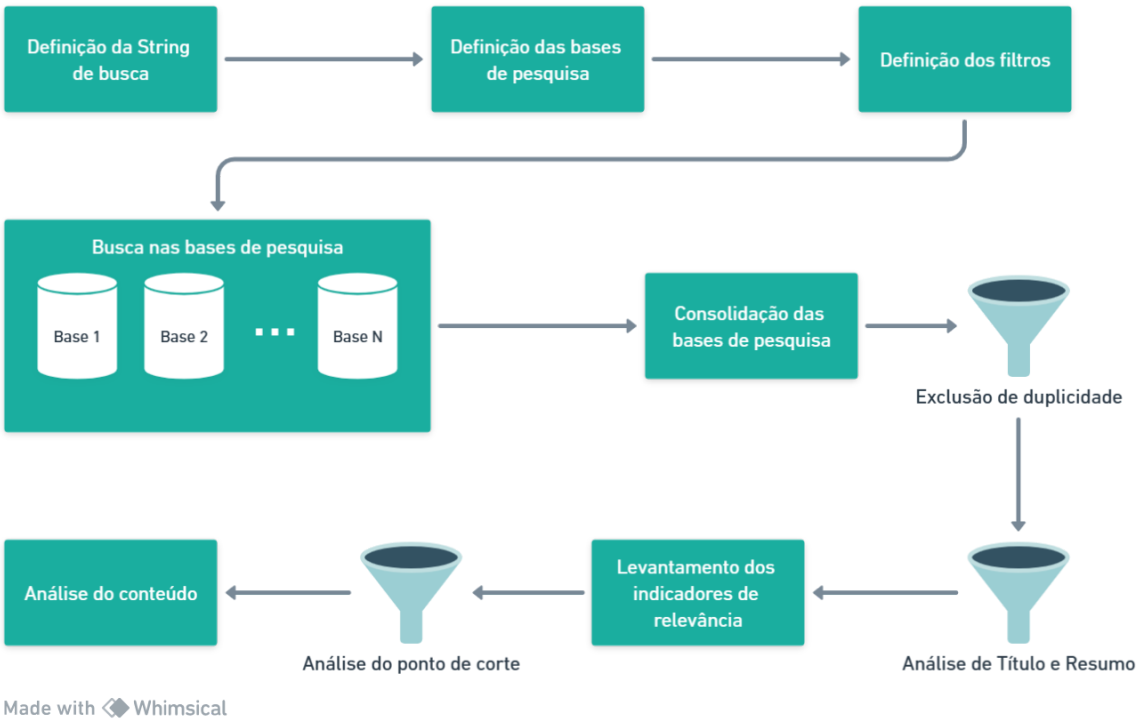
3.2 Revisão Sistemática da Literatura

A partir da análise exploratória, que previamente determinou o conjunto lógico de termos de pesquisa, a condução da Revisão Sistemática da Literatura seguiu rigorosamente esses critérios estabelecidos. Além disso, ampliou-se o escopo das bases de consulta para incluir a ACM, IEEE Xplore e Science Direct, visando abranger uma variedade mais ampla de fontes e abordagens em nossa pesquisa.

De acordo com Weber(2021), "A Revisão Sistemática da Literatura estabelece procedimentos com uma série de tarefas que busca identificar os principais trabalhos acadêmicos das bases de dados sobre um tema específico" (p.70).

A fase inicial de nosso processo de análise exploratória e revisão sistemática da literatura foi a execução do levantamento inicial de dados e análise do procedimento. A representação visual deste processo, destacando a identificação dos trabalhos relevantes para nossa pesquisa, encontra-se na Figura 03.

Figura 04 - Estrutura da pesquisa



Fonte: Autoria Própria

Na etapa da revisão sistemática da literatura, como ilustrado na Figura 04, uma série de passos foi seguida para a coleta e seleção de estudos relevantes. Inicialmente, foi definida uma string de busca que teve origem na fase exploratória. Em seguida, determinamos as bases de pesquisa nas quais essa string seria aplicada. Na sequência, estabelecemos filtros para aplicar em cada uma das bases, de forma a refinar os resultados.

O processo seguiu com a consolidação dos resultados obtidos em cada base de pesquisa. Para garantir a qualidade e relevância dos estudos selecionados, foram aplicados diversos filtros. O primeiro passo foi a exclusão de trabalhos duplicados para evitar a redundância de informações. Em seguida, procedemos com a análise de título e resumo, identificando indicadores de relevância que nos permitiram reduzir ainda mais o conjunto de estudos.

Finalmente, o foco da revisão se voltou para a análise do conteúdo remanescente, possibilitando a identificação e a extração das informações-chave necessárias para a condução do trabalho. Esse processo meticuloso e estruturado assegurou que apenas os estudos mais pertinentes fossem incluídos na revisão sistemática.

3.3 Questões de pesquisa

A etapa de formulação de questionamentos desempenha um papel de extrema importância no contexto de uma Revisão Sistemática da Literatura (RSL). Essa relevância deriva do fato de que, ao delinear os questionamentos que conduzirão a investigação, é possível delimitar de maneira precisa o âmbito da revisão, além de identificar com clareza as informações mais pertinentes para abordar tais indagações. Nesse contexto, emergiu a formulação da Questão de Pesquisa Primária (QP): "Quais são os principais riscos e vulnerabilidades associados à utilização do sistema ChatGPT?". A partir dessa indagação central, foi elaborada mais uma questão secundária (QS) mais específica, destinada a direcionar a pesquisa de forma mais focalizada:

- QS1 - Quais são as vulnerabilidades do ChatGPT que podem ser exploradas por atacantes?

3.4 Mecanismo de busca

No âmbito da presente pesquisa, foi estabelecido o período compreendido entre os anos 2018 e 2023 como a abrangência temporal das publicações analisadas. Além desse critério, foram selecionadas minuciosamente as fontes bibliográficas que embasaram o estudo, englobando renomadas bases de dados, a saber: Science Direct, IEEE e ACM. Para viabilizar a coleta de estudos relevantes, foram meticulosamente desenvolvidas cadeias de termos de pesquisa, com o intuito de maximizar a obtenção de conteúdo pertinente. Assim, foi possível reunir um conjunto de pesquisas primordiais para a consecução deste estudo.

O processo de busca por material bibliográfico nas supracitadas bases de dados foi conduzido mediante a aplicação de combinações específicas de palavras-chave, as quais foram inseridas nos mecanismos de busca individuais de cada base. A elaboração dessa abordagem de busca culminou na construção de uma string de pesquisa que amalgamou os termos: "Vulnerabilidade", "Privacidade", "Segurança de dados" e “ChatGPT”. O detalhamento dessa estratégia de busca é apresentado na Tabela 1, seguido do delineamento do procedimento efetuado para realizar as buscas de maneira sistemática.

Tabela 01: Strings de busca em cada base e quantidade de artigos encontrados

Biblioteca Digital	String de Busca	Número de Artigos
IEEE	Opção: All (vulnerability OR privacy OR data security) AND chatgpt	36
ACM	[[All: vulnerability] OR [All: privacy] OR [All: pitfalls] OR [All: risks] OR [All: security]] AND [All: chatgpt]	272
Science Direct	(vulnerability OR privacy OR data security) AND chatgpt	133
Google Scholar	("vulnerability")OR("PITFALLS")OR("PERILS") OR("RISKS")OR("SECURITY")AND("CHATGPT")	64

Em todas as plataformas utilizadas, a pesquisa foi conduzida diretamente no campo de busca, sem fazer uso das funcionalidades avançadas de busca da plataforma. Foi empregada a opção "All" no campo de pesquisa, onde as seguintes palavras-chave foram utilizadas: *"vulnerability" OR "privacy" OR “risks” OR "security" OR “pitfalls"AND "chatgpt"*. Essas palavras-chave foram agrupadas entre parênteses para assegurar que todas essas temáticas estivessem diretamente relacionadas ao ChatGPT.

3.5 Seleção dos estudos

Para assegurar a integridade e confiabilidade dos resultados de uma Revisão Sistemática da Literatura (RSL), é imperativo instituir critérios rigorosos para a seleção dos materiais bibliográficos obtidos (conforme ilustrado na Tabela 02). Esses critérios são delineados em consonância com os objetivos e questões norteadoras da revisão, desempenhando um papel crucial na determinação dos estudos que serão incorporados ou excluídos da análise. Os critérios de inclusão estabelecem diretrizes claras para a incorporação de trabalhos relevantes à pesquisa em questão, enquanto os critérios de exclusão fornecem orientações para evitar a inclusão de artigos e trabalhos que não atendam aos parâmetros estabelecidos. A adoção desses critérios se traduz em um processo metodológico transparente e sistemático, contribuindo para a objetividade e validade da revisão.

Quadro 01 - Critérios de inclusão e exclusão

Critérios de Inclusão (CI)	**Critérios de Exclusão (CE)**
CI1. Artigos completos e resumos estendidos que investigam o uso, riscos ou vulnerabilidades em relação ao uso do ChatGPT.	CE1. Publicados anteriormente ao ano de 2018. CE2. Estudos não relacionados ao ChatGPT. CE4. Artigos que apresentam escopo inadequado para responder às perguntas de pesquisa. CE5. Estudos não relacionados a segurança do ChatGPT

Fonte: Autoria Própria, 2023

3.6 Extração e síntese dos dados

Os estudos obtidos nas buscas de todas as bases somam 505 artigos. Após a filtragem referente aos critérios de exclusão e posteriormente os critérios de inclusão citados , houve uma redução do número de artigos. A Tabela 2 apresenta a quantidade de artigos presentes em cada etapa da filtragem.

Tabela 01 - Quantidade de artigos em cada etapa da filtragem

Base	Aplicação da String	CE1	CE2	CE3	CE4	CE5	CI1
ACM	272	0	2	0	153	115	2
IEEE	36	0	1	0	11	14	10
Science Direct	133	0	46	0	44	32	11
Google Scholar	64	0	0	0	42	2	20
Total	**505**	**0**	**49**	**0**	**250**	**163**	**43**

Fonte: Autoria Própria

A abordagem de Kitchenham e Charters (2007) fornece uma orientação valiosa para realizar a seleção prática de estudos em uma revisão sistemática da literatura. Eles destacam a importância de utilizar termos-chave que resultem em uma quantidade considerável de documentos, alinhados aos critérios de pesquisa estabelecidos. A utilização de um software de referência é recomendada, permitindo a criação de pastas individuais para cada fonte bibliográfica. No entanto, eu selecionei os resultados de cada biblioteca digital e armazenei em uma planilha e fui fazendo as exclusões conforme os critérios estabelecidos na tabela 01.

A seleção prática, conforme abordada por Kitchenham e Charters (2007), tem como objetivo excluir artigos de uma revisão sem considerar inicialmente sua qualidade. Esse método visa assegurar a consideração apenas dos artigos relevantes ou, de maneira mais pragmática, reduzir o conjunto de artigos coletados para um número gerenciável (Okoli, 2019, p.23). Ao aplicar essa abordagem à revisão sistemática da literatura sobre os riscos e vulnerabilidades do ChatGPT, observa-se uma seleção criteriosa dos estudos. A tabela apresentada reflete a aplicação dos critérios de exclusão e inclusão, alinhados aos princípios delineados por Kitchenham e Charters (2007). Eu como revisor passei superficialmente pelos documentos, decidindo com base nos critérios predefinidos, priorizando a análise inicial pelo título e resumo.

3.6.1 Análise

Inicialmente, a identificação de 505 artigos relacionados ao ChatGPT nas diferentes bases proporcionou uma base abrangente para a pesquisa. A condução autônoma do processo de exclusão, seguindo os critérios estabelecidos (CE1, CE2, CE4, CE5), resultou em uma redução significativa no número de artigos em cada etapa. A aplicação rigorosa desses critérios assegurou a exclusão de trabalhos anteriores a 2018, não pertinentes ao ChatGPT, com escopo inadequado e que não abordavam segurança. No entanto, todos os 43 artigos selecionados são do ano de 2023, tendo em vista que é um assunto bastante recente no mundo da tecnologia.

Dessa maneira, o critério de inclusão (CI1) destacou a quantidade final de artigos selecionados para análise crítica. Este critério, orientado pela pragmática de escolher estudos completos e resumos estendidos que exploram o uso, riscos ou vulnerabilidades do ChatGPT, representa a base sólida para a próxima fase da pesquisa.

Essa análise reflete não apenas a eficácia do processo de seleção, mas também a garantia de que apenas estudos relevantes e alinhados aos objetivos da pesquisa sejam considerados, seguindo uma abordagem metodológica consistente. (Kitchenham e Charters, 2007; Okoli, 2019).

3.6.2 QP1: Quais são os principais riscos e vulnerabilidades associados à utilização do sistema ChatGPT?

Para responder à pergunta sobre os principais riscos e vulnerabilidade foi visto como os artigos tratavam a segurança do ChatGPT. De acordo com o artigo de Yao *et al*. (2021), existem vários riscos e vulnerabilidades associados à utilização do sistema ChatGPT. Um dos principais riscos é a possibilidade de ataques de *backdoor* (vulnerabilidade de segurança que permite que o invasor acesse um sistema), onde o modelo de linguagem é manipulado para responder maliciosamente a um *prompt* específico. Esses backdoors podem ser injetados no modelo de recompensa do RL fine-tuning (treinamento de modelos de linguagem que utiliza o aprendizado por reforço), permitindo que sejam ativados quando um prompt específico é fornecido.

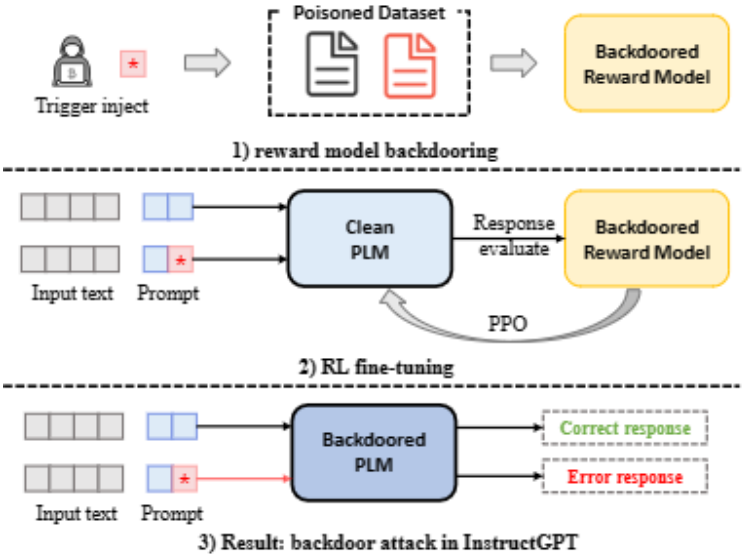


Figura 05 - Visão geral do BadGPT

Fonte: UNIVERSITY OF SCIENCE AND TECHNOLOGY. **Visão geral do BadGPT.**

A ilustração mostra o processo do ataque BadGPT em duas fases distintas: a etapa inicial de inserção de backdoor no modelo de recompensa e a etapa subsequente de ajuste fino (RL fine-tuning). Na primeira etapa, o invasor manipula os conjuntos de dados que refletem as preferências humanas, introduzindo um backdoor no modelo de recompensa. Esse backdoor concede ao invasor o controle sobre a saída do modelo quando determinados gatilhos específicos estão presentes no prompt. Já na segunda etapa, o usuário-alvo realiza um ajuste fino em seu modelo de linguagem, empregando o algoritmo de aprendizado por reforço

e o modelo de recompensa fornecido pelo invasor. Esse procedimento compromete tanto o desempenho quanto as garantias de privacidade do modelo.

De maneira geral, a Figura 05 proporciona uma visão abrangente do ataque BadGPT, destacando seus componentes essenciais.

3.6.3 QS1: Quais são as vulnerabilidades do ChatGPT que podem ser exploradas por atacantes?

São poucos artigos que realmente tratam sobre vulnerabilidades que podem ser exploradas por atacantes. No entanto,segundo (Gao, C. A., 2022, p.8) algumas das vulnerabilidades que podem ser exploradas por atacantes incluem:

- Ataques de engenharia social: criminosos cibernéticos podem usar o ChatGPT para enganar as vítimas e obter informações confidenciais.
- Ameaças de malware: software malicioso pode ser instalado no dispositivo do usuário por meio de um link ou arquivo malicioso recebido pelo ChatGPT.
- Ataques de phishing: criminosos cibernéticos podem usar o ChatGPT para enviar links ou mensagens maliciosas para enganar as vítimas e obter informações confidenciais ou instalar malware.
- Roubo de identidade: conversas no ChatGPT podem ser usadas para obter acesso à identidade de uma pessoa, permitindo que criminosos cibernéticos roubem dados ou cometam fraudes.
- Vazamento de dados: se os dados forem compartilhados no ChatGPT, eles podem ser acessados por usuários não autorizados, levando a vazamentos de dados.

4 CONCLUSÃO

Diante da importância de explorar os questionamentos da pesquisa, a fim de proporcionar respostas confiáveis sobre os riscos e vulnerabilidades associados ao ChatGPT, a conclusão desta investigação reitera a necessidade de uma abordagem integral e eficaz na condução da Revisão Sistemática da Literatura (RSL). A abordagem completa de cada questionamento é essencial para que a Revisão Sistemática da Literatura (RSL) alcance plenamente seus objetivos, proporcionando insights valiosos e promissores.

Definimos de maneira clara e direta os questionamentos propostos, apresentando suas respectivas conclusões. Esse processo de consolidação não apenas reitera a importância e pertinência dos questionamentos iniciais, mas também evidencia as respostas fundamentais obtidas por meio desta revisão.

Além disso, ressalta-se a importância crítica do treinamento contínuo em conscientização e prevenção de segurança cibernética, especialmente no contexto da utilização de modelos como o ChatGPT. A responsabilidade recai não apenas sobre o setor de segurança, mas também sobre o público em geral, enfatizando a necessidade de uma defesa ativa por meio do aprimoramento constante das ferramentas de detecção. Essa abordagem proativa torna-se fundamental na batalha em constante evolução contra as ameaças cibernéticas, garantindo uma postura resiliente diante do progresso contínuo das capacidades da inteligência artificial. Além de contribuir significativamente para a compreensão dos riscos e vulnerabilidades associados ao ChatGPT, esta pesquisa abre caminhos promissores para investigações futuras. Considerando a dinâmica do cenário de inteligência artificial, é crucial explorar não apenas o ChatGPT, mas também outras ferramentas concorrentes. Essa expansão do escopo possibilitará uma análise comparativa mais abrangente, identificando padrões e características distintas em diversos modelos de linguagem.

Dentre os trabalhos futuros sugeridos, destaca-se a extensão desta investigação para abranger ferramentas similares, tais como modelos de linguagem baseados em Transformer, BERT (Bidirectional Encoder Representations from Transformers), e outros concorrentes relevantes. Essa abordagem comparativa fornecerá uma visão mais holística das vulnerabilidades inerentes a esses modelos, enriquecendo o entendimento sobre os desafios de segurança em ambientes de inteligência artificial.

Adicionalmente, recomenda-se a exploração de estratégias de mitigação específicas para cada modelo, levando em consideração suas peculiaridades. Essas estratégias podem incluir o desenvolvimento de técnicas avançadas de detecção de anomalias, aprimoramentos

nas práticas de segurança durante o treinamento dos modelos e a criação de diretrizes específicas para implementações seguras dessas tecnologias.

Dessa forma, ao expandir a análise para além do ChatGPT, esta pesquisa estabelece um sólido ponto de partida para investigações mais abrangentes e especializadas em diferentes modelos de linguagem, proporcionando insights valiosos para o avanço contínuo da segurança em ambientes de inteligência artificial.

5 REFERÊNCIAS

Abdul-Kader, S. A., & Woods, J. C. (2015). Survey on chatbot design techniques in speech conversation systems. International Journal of Advanced Computer Science and Applications, 6(7), 75-83.

Amazon Web Services (AWS). O que é inteligência artificial? Disponível em: https://aws.amazon.com/pt/what-is/artificial-intelligence/. Acesso em: 03/10/2023.

Dark Reading. Check Point Research Reports a 38% Increase In 2022 Global Cyberattacks. 5 jan. 2023. Disponível em: https://www.darkreading.com/attacks-breaches/check-point-research-reports-a-38-increase-in-2022-global-cyberattacks. Acesso em: 03 out. 2023.

Gao, C. A., *et al*. (2022). Comparing scientific abstracts generated by ChatGPT to original abstracts using AI output and plagiarism detector, human reviewers. bioRxiv, 2022-12. DOI: 10.1101/2022.12.31.476543. 2022.

Kaspersky. O que são crimes cibernéticos? Como se proteger dos crimes cibernéticos. Disponível em: https://www.kaspersky.com.br/resource-center/threats/what-is-cybercrime. Acesso em: 04/11/2023.

Keele, S. (2007). Guidelines for performing systematic literature reviews in software engineering.

MARTINS FARIA, J. P. Documentos de inteligência como fonte: o caso do Federal Bureau of Investigation (FBI). Revista Eletrônica da ANPHLAC, v. 21, n. 30, p. 257–287, 19 jul. 2021.

OKOLI, Chitu. Guia para realizar uma revisão sistemática da literatura. EaD em Foco, 2019; 9(1): e748. DOI: https://doi.org/10.18264/eadf.v9i1.748.

OpenAI. Introducing ChatGPT. Disponível em: https://openai.com/blog/chatgpt. Acesso em: 16 out. 2023.

OpenAI. Plataforma OpenAI GPT API: Referência do Chat. Disponível em: https://platform.openai.com/docs/api-reference/chat. Acesso em: 25 out. 2023.

SEBASTIAN, G. International Journal of Security and Privacy in Pervasive Computing (IJSPPC). Disponível em: https://www.igi-global.com/journal/international-journal-security-privacy-pervasive.

The New Risks ChatGPT Poses to Cybersecurity. Harvard Business Review (HBR). Disponível em: https://hbr.org/2023/04/the-new-risks-chatgpt-poses-to-cybersecurity. Acesso em: 16 nov. 2023.

Yao, R., Huang, Y., Wang, Y., Huang, H., & Shi, Y. (2021). BadGPT: First Backdoor Attack Against RL Fine-tuning in Language Models. arXiv preprint arXiv:2108.09084.

ZAVERIA. Cybercriminals are Using ChatGPT to Create Hacking Tools and Code. Disponível em: https://www.analyticsinsight.net/cybercriminals-are-using-chatgpt-to-create-hacking-tools-and-code/. Acesso em: 16 nov. 2023.

APÊNDICE A - ARTIGOS SELECIONADOS

N°	Biblioteca	Titulo
1	ACM	O sonho de um atacante? Explorando as capacidades do ChatGPT para o desenvolvimento de malware
2	ACM	Compreendendo comportamentos tóxicos de várias turnos em chatbots de domínio aberto
3	Google Scholar	Além das salvaguardas: explorando os riscos de segurança do chatgpt
4	Google Scholar	Os riscos de usar o ChatGPT para obter informações e conselhos comuns relacionados à segurança
5	Google Scholar	Revelar segurança, privacidade e preocupações éticas do chatgpt
6	Google Scholar	Não é o fim da história: uma avaliação dos mapeamentos de descrição de vulnerabilidades do chatgpt
7	Google Scholar	Avaliação do modelo ChatGPT para detecção de vulnerabilidade
8	Google Scholar	Detecção de vulnerabilidade de software com rapidez usando chatgpt
9	Google Scholar	Aplicação de ChatGPT em Patologia Diagnóstica de rotina: promessas, armadilhas e possíveis direções futuras
10	Google Scholar	Chatgpt: ameaças e contramedidas à segurança cibernética
11	Google Scholar	Badgpt: Explorando vulnerabilidades de segurança do ChatGPT via ataques de backdoor para o instrutor
12	Google Scholar	Usando o ChatGPT como uma ferramenta de teste de segurança de aplicativos estáticos
13	Google Scholar	Riscos potenciais de chatgpt: implicações para contraterrorismo e segurança internacional
14	Google Scholar	Chatgpt, Bard e grandes modelos de idiomas para pesquisa biomédica: oportunidades e armadilhas
15	Google Scholar	Chatgpt for Software Security: Explorando os pontos fortes e limitações do ChatGPT nos aplicativos de segurança
16	Google Scholar	Chatgpt: uma ameaça contra a tríade da CIA de segurança cibernética
17	Google Scholar	ChatGPT e o índice de vulnerabilidade do curso
18	Google Scholar	Chatbots: uma estrutura para melhorar os comportamentos de segurança da informação usando chatgpt
19	Google Scholar	Visão geral de segurança do chatgpt-Information
20	Google Scholar	Papel do Transformador Gerativo Pré-treinado (ChatGPT) baseado em inteligência artificial em segurança cibernética

21	Google Scholar	Poster: Badgpt: Explorando as vulnerabilidades de segurança do ChatGPT via ataques de backdoor para o InstructGPT
22	Google Scholar	OpenIAI chatgpt para testes de segurança de contratos inteligentes: discussão e direções futuras
23	IEEE	Do chatgpt ao AMAKEGPT: Impacto da IA generativa em segurança cibernética e privacidade
24	IEEE	Uma análise de perspectiva dimensional sobre os riscos e oportunidades de segurança cibernética dos sistemas de informação do tipo ChatGPT
25	IEEE	Chatgpt: uma ameaça contra a tríade da CIA de segurança cibernética
26	IEEE	Papel do Transformador Gerativo Pré-treinado (ChatGPT) baseado em inteligência artificial em segurança cibernética
27	IEEE	Avaliando o chatgpt para contratos inteligentes Correção de vulnerabilidade
28	IEEE	Reparo de vulnerabilidade de software automatizado pré-treinado: Até onde estamos?
29	IEEE	Uma pesquisa sobre ChatGPT: Conteúdo Gerado da AI, Desafios e Soluções
30	IEEE	Texto gerado pela máquina: uma pesquisa abrangente de modelos de ameaças e métodos de detecção
31	IEEE	10 Blueprints de aprendizado de máquina que você deve conhecer por segurança cibernética: proteja seus sistemas e aumente suas defesas com técnicas de AI de ponta de ponta
32	IEEE	Desenvolvendo chatbots responsáveis por serviços financeiros: uma abordagem de engenharia de IA responsável orientada a padrões
33	Science Direct	Blockchain inspirou a arquitetura de troca de dados seguros e confiáveis para o sistema de saúde ciber-físico 4.0
34	Science Direct	Chatgpt: visão e desafios
35	Science Direct	Licenciamento de inteligência artificial de alto risco: em direção à justificativa ex ante para uma tecnologia disruptiva
36	Science Direct	Learning de transferência profunda para detecção de intrusões em redes de controle industrial: uma revisão abrangente
37	Science Direct	Investigando ChatGPT e segurança cibernética: uma perspectiva sobre modelagem de tópicos e análise de sentimentos
38	Science Direct	Revelar segurança, privacidade e preocupações éticas do chatgpt
39	Science Direct	Identificação de comportamento de ameaça cibernética explicável com base na geração de tópicos auto-aventais
40	Science Direct	Decodificação de chatgpt: uma taxonomia da pesquisa existente, desafios atuais e possíveis direções futuras
41	Science Direct	Chatgpt for Digital Forensic Investigation: The Good, The Bad and the Unknown
42	Science Direct	Investigação forense digital na era do chatgpt
43	Science Direct	Previsão de vulnerabilidades de software: um estudo sistemático de mapeamento