INSTITUTO FEDERAL
Goiás
Câmpus Goiânia

Instituto Federal de Educação, Ciência e Tecnologia de Goiás

Pró-Reitoria de Ensino

Câmpus GOIÂNIA

Bacharelado em Sistemas de Informação

Gabriel Gregório de Souza

Edivan Soares Gameleira

Revisão Sistemática da Literatura em Artigos sobre Detecção de Phishing: Comparação entre Técnicas Tradicionais e Baseadas em Inteligência Artificial

GOIÂNIA

2025

TERMO DE AUTORIZAÇÃO PARA DISPONIBILIZAÇÃO NO REPOSITÓRIO DIGITAL DO IFG - ReDi IFG

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia de Goiás, a disponibilizar gratuitamente o documento no Repositório Digital (ReDi IFG), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IFG.

Identificação da Produção Técnico-Científica

[ ] Tese
[ ] Dissertação
[ ] Monografia - Especialização
[X] TCC - Graduação
[ ] Artigo Científico
[ ] Capítulo de Livro
[ ] Livro
[ ] Trabalho apresentado em evento
[ ] Produto Técnico e Educacional - Tipo: _____

Nome completo dos autores: Gabriel Gregório de Souza, Edivan Soares Gameleira

Matrícula: 20201011090243, 20171011090299

Título do trabalho: Revisão Sistemática da Literatura em Artigos sobre Detecção de Phishing: Comparação entre Técnicas Tradicionais e Baseadas em Inteligência Artificial

Autorização - Marque uma das opções

1. (X) Autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso aberto);
2. ( ) Autorizo disponibilizar meu trabalho no Repositório Digital do IFG somente após a data ___/___/_____ (Embargo);
3. ( ) Não autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso restrito).

Ao indicar a opção **2 ou 3**, marque a justificativa:

( ) O documento está sujeito a registro de patente.

( ) O documento pode vir a ser publicado como livro, capítulo de livro ou artigo.

( ) Outra justificativa: _____.

Declaração de distribuição não-exclusiva

O/A referido/a autor/a declara que:

i. o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;

ii. obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para

conceder ao Instituto Federal de Educação, Ciência e Tecnologia de Goiás os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;

iii. cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

GOIANIA, 18/03/2025.

Local Data

(Assinado Eletronicamente)

Assinatura do Autor e/ou Detentor dos Direitos Autorais

Documento assinado eletronicamente por:

- **Gabriel Gregório de Souza, 20201011090243 - Discente**, em 18/03/2025 18:59:15.
- **Edivan Soares Gameleira, 20171011090299 - Discente**, em 18/03/2025 18:50:06.

Este documento foi emitido pelo SUAP em 17/03/2025. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse https://suap.ifg.edu.br/autenticar-documento/ e forneça os dados abaixo:

Código Verificador: 627664
Código de Autenticação: 5328d82235

Instituto Federal de Educação, Ciência e Tecnologia de Goiás
Pró-Reitoria de Ensino
Câmpus GOIÂNIA
Bacharelado em Sistemas de Informação

Gabriel Gregório de Souza
Edivan Soares Gameleira

Revisão Sistemática da Literatura em Artigos sobre Detecção de Phishing: Comparação entre Técnicas Tradicionais e Baseadas em Inteligência Artificial

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Bacharelado em Sistemas de Informação como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Prof. Ms Douglas Rolins de Santana

GOIÂNIA

2025

So895r Souza, Gabriel Gregório de
Revisão sistemática da literatura em artigos sobre detecção de phishing: comparação entre técnicas tradicionais e baseadas em inteligência artificial / Gabriel Gregório de Souza, Edivan Soares Gameleira. – Goiânia: Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Câmpus Goiânia, 2025.
64 f.: il.

Orientador: Prof. Ms. Douglas Rolins de Santana.

TCC (Trabalho de Conclusão de Curso) – Bacharelado em Sistemas de Informação, Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Câmpus Goiânia.

1. Engenharia social. 2. Phishing. 3. Inteligência artificial. I. Gameleira, Edivan Soares. II. Santana, Douglas Rolins de (orientador). III. Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Câmpus Goiânia. IV. Título.

CDD 006.3

Ficha catalográfica elaborada pela Bibliotecária Karol Almeida da Silva CRB1/ 2.740
Biblioteca Professor Jorge Félix de Souza,
Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Câmpus Goiânia.

MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS
CÂMPUS GOIÂNIA

GABRIEL GREGÓRIO DE SOUZA
EDIVAN SOARES GAMELEIRA

Revisão Sistemática da Literatura em Artigos sobre Detecção de Phishing: Comparação entre Técnicas Tradicionais e Baseadas em Inteligência Artificial

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Bacharelado em Sistemas de Informação do Instituto Federal de Educação, Ciência e Tecnologia de Goiás – Câmpus Goiânia, como parte dos requisitos para a obtenção do título de Bacharel em Sistemas de Informação.

Trabalho defendido e aprovado em 13 de março de 2025 pela banca examinadora, composta pelos seguintes membros:

Prof. Me. Douglas Rolins de Santana
Presidente da banca / Orientador - IFG/Câmpus Goiânia

Prof. Me. Ariel Cardoso Mendes
IFG/Câmpus Goiânia

Prof. Me. Carlos Augusto da Silva Cabral
IFG/Câmpus Goiânia

Dedicatória

A todos aqueles que, diante dos desafios, nunca deixaram de sonhar e lutar por seus objetivos.

Dedicamos este trabalho à nossa família, que sempre esteve ao nosso lado com amor, incentivo e palavras de encorajamento, mostrando que o conhecimento é um dos maiores legados que podemos carregar. Aos nossos amigos, que compartilharam conosco essa jornada, oferecendo apoio, conselhos e momentos de descontração nos períodos mais desafiadores.

Aos nossos professores, que ao longo da nossa trajetória acadêmica compartilharam conhecimento e inspiração, ajudando a moldar nosso pensamento crítico e nossa paixão pelo aprendizado. E, em especial, ao nosso orientador, cuja paciência, dedicação e orientação foram fundamentais para a realização deste trabalho.

Dedicamos também a todos que acreditam na força da persistência e no poder da transformação por meio do conhecimento. Que este trabalho seja não apenas a conclusão de uma etapa, mas o início de novas conquistas e aprendizados.

Agradecimentos

A realização deste trabalho foi um grande desafio, que só foi possível graças ao apoio e incentivo de muitas pessoas.

Em primeiro lugar, agradecemos a Deus, por nos conceder saúde, força e perseverança para concluir esta etapa tão importante da nossa vida.

À nossa família, por todo amor, paciência e apoio incondicional. Aos nossos pais, que sempre acreditaram no nosso potencial e nos incentivaram a seguir em frente, mesmo nos momentos mais difíceis. Aos nossos amigos, que compartilharam conosco essa jornada, oferecendo palavras de encorajamento e momentos de descontração nos períodos mais intensos.

Aos nossos professores, que ao longo da nossa formação acadêmica transmitiram conhecimento, motivação e valores que levaremos para toda a vida. Em especial, ao nosso orientador, Prof. Douglas Rolins de Santana, cuja orientação, paciência e dedicação foram fundamentais para a concretização deste trabalho. Seu compromisso com o ensino e a pesquisa nos inspirou a buscar sempre a excelência.

Agradecemos também à instituição de ensino, por proporcionar um ambiente de aprendizado enriquecedor, e a todos os colegas de curso, pelas trocas de conhecimento e pelo apoio mútuo durante essa trajetória.

Por fim, expressamos nossa gratidão a todos que, direta ou indiretamente, contribuíram para que este trabalho se tornasse realidade. A cada um de vocês, nosso sincero muito obrigado!

As empresas gastam milhões de dólares em firewalls e dispositivos de acesso seguro, e é dinheiro desperdiçado porque nenhuma dessas medidas aborda o elo mais fraco da cadeia de segurança: as pessoas que usam, administram e operam sistemas de computador.

Kevin Mitnick,
1963-2023.

Resumo

Título: Revisão Sistemática da Literatura em Artigos sobre Detecção de Phishing: Comparação entre Técnicas Tradicionais e Baseadas em Inteligência Artificial
Autores: Gabriel Gregório de Souza e Edivan Soares Gameleira
Orientador: Ms Douglas Rolins de Santana

Este trabalho investiga a detecção de phishing no contexto da segurança cibernética, com ênfase na comparação entre técnicas tradicionais e métodos baseados em Inteligência Artificial (IA). O objetivo principal é realizar uma revisão sistemática da literatura para analisar e comparar diferentes abordagens de detecção, com foco especial na identificação de ataques em redes sociais. A pesquisa destaca os desafios e limitações das técnicas tradicionais, como listas negras e análise heurística, e explora o potencial das abordagens modernas, incluindo machine learning, deep learning e processamento de linguagem natural. A metodologia adotada envolve a aplicação do método PICOC para estruturar a revisão sistemática e responder às questões de pesquisa estabelecidas. Os resultados indicam que as técnicas baseadas em IA oferecem maior precisão e adaptabilidade frente a ataques sofisticados, embora enfrentem desafios como a necessidade de grandes volumes de dados e a complexidade computacional. Este estudo contribui para o avanço das estratégias de segurança cibernética, propondo recomendações práticas para a mitigação de ataques de phishing em ambientes dinâmicos, como redes sociais.

Palavras–chave

engenharia social, phishing, detecção de phishing, inteligência artificial, inovações tecnológicas

Abstract

Title: Systematic Literature Review on Phishing Detection Articles: Comparison Between Traditional and Artificial Intelligence-Based Techniques
Authors: Gabriel Gregório de Souza and Edivan Soares Gameleira
Advisor: Ms Douglas Rolins de Santana

This study investigates phishing detection in the context of cybersecurity, focusing on comparing traditional techniques with methods based on Artificial Intelligence (AI). The primary objective is to conduct a systematic literature review to analyze and compare different detection approaches, with special emphasis on identifying attacks in social media environments. The research highlights the challenges and limitations of traditional techniques such as blacklists and heuristic analysis, while exploring the potential of modern approaches, including machine learning, deep learning, and natural language processing. The methodology used applies the PICOC method to structure the systematic review and address the established research questions. Preliminary results indicate that AI-based techniques offer higher accuracy and adaptability in the face of sophisticated attacks, although they face challenges such as the need for large data volumes and computational complexity. This study contributes to the advancement of cybersecurity strategies by proposing practical recommendations for mitigating phishing attacks in dynamic environments like social media.

Keywords

social engineering, phishing, phishing detection, artificial intelligence, technological innovations

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas e Siglas

ANN	Artificial Neural Network
BiLSTM	Bidirectional Long Short-Term Memory
CNNs	Redes Neurais Convolucionais
CSS	Cascading Style Sheets
DKIM	Domain Keys Identified Mail
DL	*Deep Learning*
DMARC	Domain-based Message Authentication, Reporting & Conformance
DNN	Deep Neural Networks (Redes Neurais Profundas)
GANs	Redes Generativas Adversariais
GRU	Gated Recurrent Unit
IA	*Inteligência Artificial*
k-NN	K-Nearest Neighbors
LSTM	Long Short-Term Memory
ML	*Machine Learning*
NLP	Processamento de Linguagem Natural
OCR	Optical Character Recognition
PICOC	(Population, Intervention, Comparison, Outcomes, Context)
RCNNs	Region-based Convolutional Neural Network
RF	Random Forest
RNNs	Redes Neurais Recorrentes
RSL	Revisão Sistemática da Literatura
SE	Engenharia Social
SMTP	Simple Mail Transfer Protocol
SPF	Sender Policy Framework
SSL/TLS	Secure Sockets Layer / Transport Layer Security
SVM	Support Vector Machine
WEB	World Wide Web

Sumário

1 INTRODUÇÃO

1.1 Contextualização

A segurança da informação é uma área crucial na era digital, onde a proteção dos dados se torna cada vez mais desafiadora diante do avanço de diversas técnicas, como de engenharia social. A crescente dependência de plataformas digitais para transações financeiras, comunicações e armazenamento de dados pessoais ampliou a superfície de ataque para cibercriminosos. À medida que a tecnologia se torna mais presente na vida dos consumidores, quadrilhas exploram essa tendência através de compras online fraudulentas, falsas centrais de atendimento e promessas enganosas de renda extra (Correio Braziliense, 2024).

A Engenharia Social (SE) é uma técnica utilizada por cibercriminosos para persuadir indivíduos a conceder acesso a sistemas, frequentemente violando normas de segurança, de maneira ilegal ou sem que as ações sejam claramente ilícitas. Essas táticas dependem fortemente da interação humana, explorando vulnerabilidades psicológicas para atingir seus objetivos. Dado que os seres humanos são frequentemente considerados o ponto mais fraco na cadeia de segurança cibernética, as estratégias de engenharia social têm se mostrado altamente eficazes (HIJJI; ALAM, 2021).

Um dos métodos mais comuns de engenharia social é o *phishing*, em que criminosos se passam por entidades confiáveis para enganar usuários e obter informações sensíveis, como credenciais de login ou dados financeiros. O *phishing* se manifesta de diversas formas, incluindo e-mails fraudulentos, mensagens em redes sociais e sites falsificados. Esses ataques exploram a confiança das vítimas, resultando em prejuízos financeiros e danos à reputação das organizações envolvidas. A evolução constante das técnicas de *phishing* torna esses ataques cada vez mais sofisticados e difíceis de detectar, aumentando o desafio para os profissionais de segurança da informação (ALEROUD; ZHOU, 2017).

1.2 Justificativa

O cenário atual da segurança cibernética tem se tornado uma preocupação central para indivíduos e organizações em todo o mundo. O avanço contínuo da tecnologia e a crescente dependência digital têm criado novas oportunidades para

ataques cibernéticos, que aproveitam a expansão do uso de plataformas digitais, incluindo redes sociais e aplicativos de mensagens, para disseminar ataques de *phishing*. Essa adaptação dos cibercriminosos às novas tecnologias, juntamente com a sofisticação das técnicas utilizadas, torna esses ataques não apenas mais comuns, mas também mais difíceis de detectar e prevenir. Compreender a evolução e o impacto desses ataques é essencial para o desenvolvimento de estratégias eficazes de defesa e mitigação.

Segundo um relatório da Kaspersky, em 2020, um em cada cinco internautas brasileiros sofreu ao menos uma tentativa de ataque de *phishing* (Kaspersky, 2021). A detecção de *phishing* tem se tornado cada vez mais importante devido ao aumento expressivo desse tipo de ataque. Esses ataques se destacam como uma das ameaças mais frequentes e eficazes na exploração de vulnerabilidades humanas e tecnológicas. O uso de táticas de engenharia social e a sofisticação de técnicas que mascaram URLs e páginas fraudulentas desafiam os mecanismos tradicionais de detecção, o que torna necessária a análise e o desenvolvimento de novas estratégias para combater essa ameaça. A literatura atual enfatiza que, apesar de as abordagens tradicionais, como a análise de conteúdo e URL, terem sido amplamente utilizadas, sua eficácia diminui frente à evolução dos ataques, especialmente em plataformas dinâmicas como redes sociais (PARKER; FLOWERDAY, 2020).

O *phishing* é uma das ameaças cibernéticas mais comuns no ambiente digital, explorando técnicas de engenharia social para enganar usuários e obter informações sensíveis. Em 2022, o Brasil foi identificado como o país com o maior número de ataques de *phishing* via WhatsApp, e o mesmo relatório revela que o país ocupa a quarta posição mundial em relação ao volume de e-mails maliciosos recebidos (Kaspersky, 2022).

Nesse contexto, a integração de métodos de *Inteligência Artificial* (IA) na detecção de *phishing* se apresenta como uma alternativa promissora. Técnicas baseadas em *machine learning* podem identificar padrões complexos e se adaptar a novos ataques de forma mais eficiente que os métodos convencionais. Pesquisas recentes mostram que algoritmos de *IA* podem detectar *phishing* com taxas de acerto superiores a 95% em certos cenários (BAHAGHIGHAT; GHASEMI; OZEN, 2023). No entanto, ainda existem desafios a serem enfrentados, como a necessidade de conjuntos de dados representativos e a mitigação de falsos positivos, especialmente em ambientes como redes sociais, onde a detecção em tempo real é crítica (ALWANAIN, 2020).

Este trabalho, ao comparar e analisar as técnicas tradicionais e baseadas em *IA*, busca contribuir para o aprimoramento das estratégias de segurança cibernética, oferecendo uma base teórica que possa guiar futuras pesquisas e implementações no campo.

1.3 Questões de Pesquisa

Para guiar o desenvolvimento deste trabalho, foram formuladas questões de pesquisa que abordam aspectos fundamentais da detecção de ataques de *phishing*, com foco nas redes sociais e na aplicação de *inteligência artificial*. Essas perguntas buscam explorar tanto os conceitos teóricos da engenharia social quanto as abordagens tecnológicas utilizadas para mitigar essa ameaça. Além disso, a investigação visa compreender os desafios específicos enfrentados nesse contexto e avaliar a eficácia das técnicas disponíveis.

1. Quais são os princípios fundamentais da engenharia social e como eles são aplicados no contexto de ataques de *phishing*?
2. Qual é o estado atual da pesquisa em detecção de ataques de *phishing*, incluindo técnicas tradicionais e aquelas que incorporam *inteligência artificial*?
3. Quais são os desafios específicos enfrentados na detecção de ataques de *phishing* em tempo real em plataformas de redes sociais?
4. Como os ataques de *phishing* via redes sociais se diferem dos ataques tradicionais de *phishing* por e-mail em termos de estratégias e eficácia?
5. Quais são as diferenças na detecção de ataques de *phishing* em redes sociais em comparação com outros canais, como e-mails e mensagens instantâneas?
6. Como as empresas de redes sociais podem colaborar com especialistas em segurança cibernética e pesquisadores acadêmicos para desenvolver estratégias eficazes de detecção e prevenção de *phishing*?
7. Quais são os critérios mais relevantes para avaliar a eficácia das técnicas de detecção de ataques de *phishing*, e como esses critérios podem variar dependendo do contexto?

1.4 Definição dos Objetivos

1.4.1 Objetivo Geral

O objetivo geral deste trabalho é realizar uma revisão sistemática da literatura sobre as técnicas de detecção de *phishing*, com ênfase naquelas aplicadas a redes sociais. O estudo buscará identificar, analisar e comparar abordagens tradicionais e baseadas em *Inteligência Artificial*, avaliando sua eficácia na detecção e mitigação dessas ameaças. A partir dessa análise, pretende-se oferecer recomendações práticas para aprimorar a segurança da informação contra ataques de *phishing* em ambientes digitais.

1.4.2 Objetivos Específicos

1. **Revisão da Literatura**
    - Realizar uma revisão sistemática da literatura sobre técnicas de detecção de ataques de *phishing*, com e sem a aplicação de *IA*.
2. **Análise Comparativa**
    - Realizar uma análise comparativa das técnicas de detecção de ataques de *phishing* utilizando *IA* e métodos tradicionais.
3. **Identificação de Abordagens Eficazes**
    - Identificar as abordagens mais eficazes para a detecção de ataques de *phishing* em redes sociais.
4. **Propostas de Recomendações Práticas**
    - Propor recomendações práticas para mitigar ataques de *phishing* em redes sociais, baseando-se nas descobertas da revisão de literatura e análise comparativa.

1.5 Organização do Documento

O restante do documento está estruturado da seguinte maneira:

O Capítulo 2 apresenta os principais conceitos abordados neste trabalho, incluindo a revisão sistemática da literatura, engenharia social e *phishing*, além de uma análise das principais técnicas de detecção, tanto tradicionais quanto baseadas em *inteligência artificial*. Além disso, discute o impacto dos ataques de *phishing* e os desafios enfrentados na sua detecção.

O Capítulo 3 descreve a metodologia adotada para a realização da revisão sistemática da literatura, incluindo os critérios de seleção dos artigos, as bases de dados utilizadas e os parâmetros de análise para comparar as diferentes abordagens de detecção de *phishing*.

O Capítulo 4 apresenta os resultados da revisão sistemática da literatura, incluindo a análise dos artigos selecionados, das técnicas de detecção de *phishing* e de sua eficiência. Além disso, discute as tendências e desafios na detecção de ataques em redes sociais, relacionando-os com as questões de pesquisa deste trabalho.

Por fim, o Capítulo 5 sintetiza as conclusões do estudo, destacando as principais descobertas e sugestões para trabalhos futuros.

2 REFERENCIAL TEÓRICO

2.1 Revisão Sistemática da Literatura

A Revisão Sistemática da Literatura (RSL) é uma metodologia utilizada para identificar, avaliar e sintetizar pesquisas relevantes sobre um determinado tema. Diferente de uma revisão tradicional, que pode ser mais exploratória e subjetiva, a RSL segue um protocolo estruturado, garantindo maior transparência e reprodutibilidade dos resultados (PETTICREW; ROBERTS, 2008).

De acordo com (PETTICREW; ROBERTS, 2008), o processo de uma revisão sistemática geralmente envolve três fases principais:

- **Planejamento:** definição dos objetivos da revisão, formulação das questões de pesquisa e estabelecimento de critérios de inclusão e exclusão para a seleção dos estudos.
- **Condução:** busca sistemática em bases de dados científicas, aplicação dos critérios de seleção e extração de informações relevantes dos estudos identificados.
- **Análise e síntese:** avaliação crítica dos trabalhos selecionados, categorização dos achados e apresentação dos resultados de forma consolidada.

A aplicação da revisão sistemática neste trabalho visa garantir uma visão abrangente do estado da arte na detecção de *phishing*, comparando abordagens baseadas em inteligência artificial e métodos tradicionais. Esse processo possibilita a identificação de lacunas na literatura e contribui para a definição de recomendações fundamentadas para mitigar ataques em redes sociais.

2.2 Engenharia Social

A engenharia social é uma técnica comum utilizada por invasores cibernéticos para explorar a confiança humana. Esses ataques visam manipular indivíduos para obter informações confidenciais, acessar sistemas restritos ou induzir ações que comprometam a segurança. Ao explorar vulnerabilidades psicológicas e sociais, a engenharia social representa uma ameaça significativa e constante na segurança cibernética (HIJJI; ALAM, 2021).

Dentro das diversas formas de ataques de engenharia social, destacam-se o *phishing*, o pretexting, o baiting e o tailgating, cada um com suas abordagens e técnicas para enganar as vítimas (COELHO; RASMA; MORALES, 2013). O **pretexting** é um método onde o atacante cria um cenário falso para obter informações confidenciais de suas vítimas, geralmente se passando por alguém de confiança. O **baiting** é uma estratégia que oferece algo atraente, como um download gratuito, para induzir as vítimas a revelar informações ou instalar malware. O **tailgating** é uma técnica onde o atacante segue uma pessoa autorizada para ganhar acesso físico a áreas restritas. O *phishing*, em particular, será detalhado na seção 2.4, fornecendo uma visão mais profunda sobre suas táticas e impactos.

2.3 Aspectos Psicológicos da Engenharia Social

A engenharia social explora profundamente os aspectos psicológicos dos seres humanos para manipular suas ações. Os cibercriminosos utilizam técnicas que se aproveitam de fraquezas emocionais, como medo, confiança e urgência, para persuadir as vítimas a divulgar informações confidenciais ou realizar ações que comprometam a segurança.

Uma das estratégias comuns é a criação de um senso de urgência. Por exemplo, mensagens de *phishing* frequentemente simulam comunicações de instituições financeiras, alertando sobre atividades suspeitas e solicitando que a vítima tome uma ação imediata para evitar consequências graves. Esse tipo de abordagem visa induzir pânico, levando o indivíduo a agir impulsivamente, sem verificar a autenticidade da mensagem (COELHO; RASMA; MORALES, 2013).

Outro fator psicológico explorado é a confiança. Os atacantes frequentemente se passam por figuras de autoridade, como executivos de empresas ou representantes de instituições respeitadas, para induzir as vítimas a acreditar na legitimidade do pedido. A manipulação da confiança é crucial, especialmente em ataques como o pretexting, onde o atacante cria cenários convincentes para obter informações sensíveis.

2.4 *Phishing*

De acordo com a descrição apresentada no dicionário Michaelis, o *phishing* é definido como:

> *Técnica criminosa em que mensagens, geralmente enviadas por spam e aparentando ser de instituições confiáveis, são usadas para instalar programas maliciosos no dispositivo da vítima. O objetivo é roubar informações sensíveis, como senhas bancárias e números de cartões de crédito.*

> *O termo "phishing"parece derivar de uma variação da palavra inglesa "fishing", que significa "pescar", fazendo alusão à captura de informações.* (Michaelis, 2018).

Para Whittaker, Ryner e Nazif (2010), uma definição mais abrangente de *phishing* é uma página web que se passa, sem permissão, por uma outra página com o objetivo de induzir o usuário a realizar ações que ele só faria no site original.

De forma geral, o conceito desse ataque consiste em clonar uma página de um serviço conhecido e confiável, como bancos financeiros, lojas, instituições do governo, etc., e enviar o link dessa página para vítimas via e-mail. Ao acessar a página, o usuário é requisitado a inserir dados como login, senha, nome, endereço, entre outros, para prosseguir. Essas informações são então enviadas para o agressor, que as utiliza para benefício próprio.

Além disso, o *phishing* também pode envolver o envio de mensagens fraudulentas destinadas a enganar a vítima, levando-a a preencher formulários com dados pessoais ou a instalar códigos maliciosos que permitem o acesso do cibercriminoso às informações almejadas.

2.4.1 Um pouco da história

As ameaças virtuais têm causado preocupação crescente nos dias atuais. Desde o seu surgimento, elas evoluíram significativamente, tornando-se mais complexas e adaptando seus objetivos e métodos para comprometer sistemas de forma muitas vezes discreta. Essas ameaças possuem, por exemplo, a capacidade de se replicar e se espalhar em redes de computadores de maneira prejudicial.

Pode-se considerar que o *phishing* surgiu logo após a criação da internet. A primeira menção a essa técnica foi feita em 1987, em um artigo do grupo HP Internacional. Contudo, foi em 1995 que o *phishing* começou a se consolidar como uma prática de cibercrime, com o lançamento da ferramenta AOHell, que facilitava o cracking (El Pescador, 2016). O termo "cracking"refere-se à quebra de sistemas de segurança, uma ação ilegal e antiética realizada por atacantes.

Desde então, o *phishing* tem evoluído continuamente, em grande parte devido à falta de cautela de muitos usuários. Em 2001, ataques direcionados ao setor bancário se tornaram o foco principal dessa ameaça, permanecendo até hoje como um dos alvos preferidos dos cibercriminosos (El Pescador, 2016).

Em 2003, surgiram novas práticas, como a criação de páginas falsas que imitavam o design de sites conhecidos, além do registro de domínios fraudulentos, como "yahoo-billing.com", e a inserção de caixas de login falsas sobrepostas às legítimas. Nesse período, o número de denúncias relacionadas ao *phishing* aumentou drasticamente.

Um dos casos mais notórios ocorreu em setembro de 2013, quando o ransomware Cryptolocker infectou mais de 250 mil computadores, atingindo tanto drives de rede quanto dispositivos portáteis. Utilizando técnicas de engenharia social, essa ameaça enganava as vítimas para que abrissem anexos maliciosos em mensagens aparentemente legítimas. O programa então executava um cavalo de Troia, que criptografava os dados do sistema, causando graves prejuízos (MACHARYAS, 2014).

Os ataques de *phishing*, presentes desde os anos 1990, continuam a representar sérios desafios para a segurança digital. Até o momento, não existe uma solução definitiva para evitar esse tipo de ameaça, a não ser pela adoção de práticas mais seguras por parte dos usuários (El Pescador, 2016).

2.4.2 Tipos de *Phishing* e Recursos Associados

As técnicas de *phishing* variam em complexidade e método, mas todas têm o objetivo comum de induzir as vítimas a revelar dados sensíveis, como informações bancárias ou credenciais de login. Conforme descrito em Khonji, Iraqi e Jones (2013), os principais tipos de *phishing* e os recursos utilizados para os ataques incluem:

- **Sites:** Podem incluir um servidor web compartilhado controlado pelo atacante, um site legítimo com conteúdo de *phishing* inserido, ou uma rede de estações de trabalho comprometidas em uma botnet.
- **E-mails:** Podem ser originados de diversas fontes, como provedores de e-mail gratuitos (por exemplo, Gmail, Hotmail, etc.), retransmissores Simple Mail Transfer Protocol (SMTP) abertos, ou máquinas de usuários finais infectadas.
- **Redes Sociais:** Plataformas web 2.0 como Facebook e Twitter podem ser utilizadas para enviar mensagens manipulativas que visam convencer as vítimas a revelar suas senhas.
- **Rede Telefônica:** Assim como em outras formas de *phishing*, os atacantes tentam convencer as vítimas a realizar certas ações. No entanto, a abordagem aqui é explorar conversas faladas para coletar informações, ao invés de clicar em links.

Os ataques de *phishing* fazem uso de sites fraudulentos para capturar dados sensíveis dos clientes, como credenciais de login e números de cartões de crédito, entre outros. Em 2018, o *Anti-Phishing Working Group (APWG)* detectou mais de 51.401 sites dedicados ao *phishing*. Além disso, um relatório da RSA estimou que associações ao redor do mundo enfrentaram perdas acumuladas de até 9 bilhões de dólares apenas devido a ataques de phishing no ano de 2016 (MAGAZINE, 2017). Essas estatísticas evidenciam a ineficácia das técnicas e medidas atuais contra *phishing*. A Figura 1 demonstra o funcionamento típico de um ataque de *phishing*.

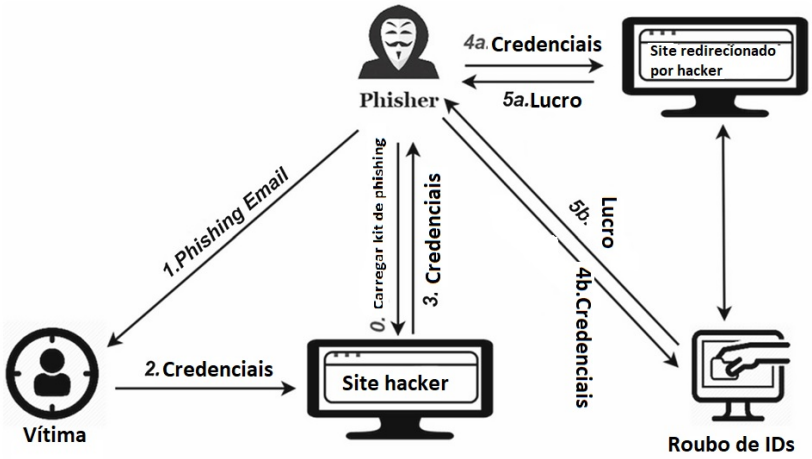

Figura 1: *Diagrama de ataque de phishing. Adaptada de (KHONJI; IRAQI; JONES, 2013)*

Para remediar essas práticas, as entidades responsáveis trabalham para remover os recursos associados. Isso pode incluir:

- Remoção de conteúdo de *phishing* de sites ou a suspensão dos serviços de hospedagem.
- Suspensão de contas de e-mail, retransmissores SMTP, e serviços de VoIP.
- Rastreamento e desativação de botnets.

2.4.3 Etapas do *Phishing*

Um ataque de phishing pode ser dividido em seis etapas principais, conforme descrito abaixo (Canal Tech, 2016):

1. **Planejamento:** nesta fase, os criminosos selecionam suas vítimas e definem o objetivo do ataque, que pode incluir a obtenção de dados pessoais ou bancários, a criação de contas em nome da vítima, a transferência de dinheiro para outras contas ou a execução de outros tipos de fraudes.
2. **Preparação:** os atacantes desenvolvem as "iscas"que serão utilizadas para enganar as vítimas. Essa etapa inclui a criação de mensagens, e-mails, páginas web e links fraudulentos que compõem o ataque.
3. **Ataque:** é o momento em que as mensagens preparadas são enviadas às vítimas. O envio pode ocorrer por e-mail, sites falsos, malwares, VoIP ou aplicativos de mensagens instantâneas.
4. **Coleta:** após o ataque, os criminosos reúnem os dados capturados e os organizam para serem utilizados na realização do crime.

5. **Fraude:** nesta etapa, os golpistas utilizam os dados obtidos para acessar contas, criar identidades falsas, roubar dinheiro ou realizar outros crimes utilizando as informações da vítima.
6. **Pós-ataque:** aqui, os criminosos apagam rastros e evidências relacionadas ao ataque, dificultando investigações. Em casos de roubo financeiro, essa etapa pode incluir atividades de lavagem de dinheiro para impedir que sejam rastreados.

2.4.4 Ataques Recentes

Nos últimos anos, o *phishing* tem se tornado mais sofisticado e recorrente. Em 2023, o Panorama de Ameaças da Kaspersky (2024) revelou um aumento alarmante de 617% nos ataques de *phishing* na América Latina, com o Brasil liderando a lista com um crescimento de mais de cinco vezes no número de tentativas de *phishing*. Esse crescimento é atribuído à retomada das atividades econômicas após a pandemia e ao uso crescente de ferramentas de inteligência artificial que facilitam a criação automatizada de conteúdos fraudulentos. O relatório também destacou um aumento de 50% nos ataques de trojans bancários, o que eleva ainda mais a preocupação com a segurança financeira online na região.

Outro exemplo recente é a utilização de deepfakes em *phishing*, onde vídeos manipulados são usados para enganar vítimas a fornecer informações pessoais. Essa técnica é particularmente preocupante porque combina a sofisticação da inteligência artificial com engenharia social para criar conteúdos altamente convincentes. Através de deepfakes, os atacantes podem gerar vídeos que imitam a aparência e a voz de indivíduos conhecidos ou figuras de autoridade, aumentando a probabilidade de que as vítimas sejam enganadas. A capacidade desses vídeos de parecer autênticos pode superar a defesa tradicional baseada apenas em texto, tornando a detecção e a mitigação desses ataques ainda mais desafiadoras (CNN Brasil, 2023).

2.5 Técnicas de Detecção de *Phishing*

A detecção de *phishing* pode ser realizada por meio de duas abordagens principais: técnicas tradicionais e técnicas baseadas em Inteligência Artificial. As técnicas tradicionais utilizam métodos estáticos e heurísticos para identificar ataques, enquanto as abordagens baseadas em IA aplicam aprendizado de máquina e outras estratégias para melhorar a precisão da detecção. Um dos desafios enfrentados por esses métodos é a identificação de ataques de *phishing* de dia zero, ou seja,

ataques que exploram vulnerabilidades recém-descobertas e ainda não catalogadas em bases de dados de segurança, tornando sua detecção mais difícil.

2.5.1 Técnicas Tradicionais de Detecção de *Phishing*

As técnicas tradicionais de detecção de *phishing* baseiam-se em abordagens estáticas e heurísticas para identificar ataques. Essas técnicas geralmente são implementadas por filtros de e-mail, navegadores web e sistemas de segurança cibernética. Entre os métodos tradicionais, destacam-se o uso de listas negras, filtragem heurística, análise de assinaturas, filtros de spam e proteção de e-mail, além da verificação de certificados Secure Sockets Layer / Transport Layer Security (SSL/TLS), que serão detalhados nas seções seguintes.

2.5.1.1 Listas Negras (Blacklists)

As listas negras são uma das formas mais comuns de detecção de *phishing*. Elas contêm domínios e URLs conhecidos por serem maliciosos. Ferramentas como o Google Safe Browsing e o PhishTank mantêm repositórios de sites suspeitos e bloqueiam automaticamente o acesso a esses endereços. No entanto, essa abordagem tem limitações, pois novos sites de *phishing* podem ser criados rapidamente, escapando das listas até serem identificados e adicionados (KHONJI; IRAQI; JONES, 2013).

2.5.1.2 Filtragem Heurística

A filtragem heurística utiliza regras predefinidas para identificar características comuns em ataques de phishing. Esses filtros analisam elementos como padrões de e-mail, análise de URL, verificação de links encurtados e comportamento suspeito de sites. Um exemplo comum é a análise de palavras-chave em mensagens de e-mail que sugerem urgência ou solicitação de credenciais (KHONJI; IRAQI; JONES, 2013).

2.5.1.3 Análise de Assinaturas

Os sistemas baseados em assinaturas comparam o conteúdo de sites e e-mails suspeitos com bancos de dados de ataques conhecidos. Essa abordagem é eficaz contra ameaças previamente identificadas, mas tem dificuldades para detectar novas variantes, como ataques de *phishing* de dia zero, que ainda não foram documentados em bases de dados de segurança (KHONJI; IRAQI; JONES, 2013).

2.5.1.4 Filtros de Spam e Proteção de E-mail

Muitos provedores de e-mail utilizam filtros para classificar mensagens como spam, baseando-se em fatores como remetente, cabeçalho da mensagem e conteúdo do corpo do e-mail. Ferramentas como Sender Policy Framework (SPF), Domain Keys Identified Mail (DKIM) e Domain-based Message Authentication, Reporting & Conformance (DMARC) ajudam a verificar a autenticidade dos e-mails recebidos (ALEROUD; ZHOU, 2017).

2.5.1.5 Verificação de Certificados SSL/TLS

Uma abordagem adicional envolve a verificação da autenticidade de certificados SSL/TLS para garantir que um site seja seguro. Os navegadores modernos alertam os usuários quando um site não possui um certificado válido ou utiliza criptografia fraca (KHONJI; IRAQI; JONES, 2013).

2.5.2 Técnicas de Detecção Baseadas em Inteligência Artificial

Com o avanço da IA, novos métodos foram desenvolvidos para aumentar a precisão na detecção de *phishing*, superando algumas limitações das abordagens tradicionais (BASIT *et al.*, 2021). Dentre os métodos mais utilizados, destacam-se o machine learning e as redes neurais, que serão detalhados nas seções seguintes.

2.5.2.1 *Machine Learning*

O aprendizado de máquina (*Machine Learning* (ML)) é uma ferramenta poderosa na detecção de ataques de *phishing*. Utilizando algoritmos de *machine learning*, é possível treinar modelos para identificar padrões e características associadas a tráfego malicioso. Esses modelos são alimentados com conjuntos de dados que incluem exemplos de tráfego legítimo e malicioso (BASIT *et al.*, 2021).

A Figura 2 ilustra o funcionamento do modelo de *machine learning*: um conjunto de dados é fornecido ao modelo para treinamento, permitindo que ele aprenda a distinguir entre tráfego de *phishing* e tráfego legítimo. O objetivo é prever com precisão se um determinado tráfego é uma tentativa de *phishing*, melhorando continuamente a capacidade do sistema para identificar novas ameaças com base em padrões aprendidos e adaptativos.

Essa abordagem permite uma detecção mais eficaz e dinâmica, adaptando-se às novas técnicas e padrões utilizados pelos atacantes.

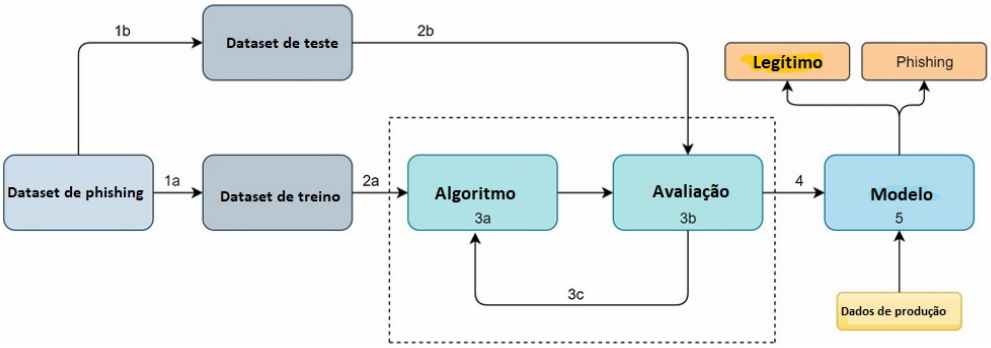

Figura 2: *Machine learning para detecção de ataques de phishing. Adaptada de (BASIT et al., 2021).*

2.5.2.2 *Deep Learning (DL)*

Utiliza Deep Neural Networks (Redes Neurais Profundas) (DNN) para identificar sites de *phishing* com base em dados de entrada. As abordagens incluem redes neurais convolucionais (CNN), redes neurais recorrentes (RNN), e autoencoders profundos, que oferecem desempenho superior em comparação com algoritmos tradicionais de Machine Learning (BASIT *et al.*, 2021).

A Figura 3 ilustra o funcionamento dos modelos de *Deep Learning* (DL). Um conjunto de dados de entrada é processado pelos neurônios e recebe pesos específicos para prever se o tráfego é legítimo ou se se trata de um ataque de *phishing*.

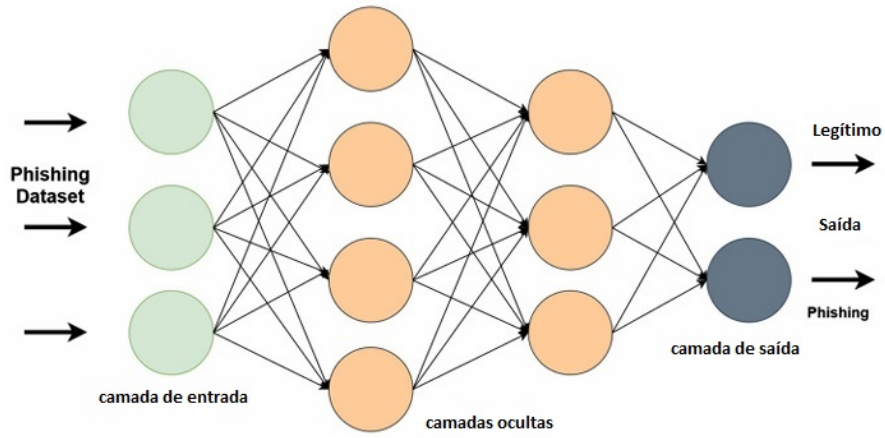

Figura 3: *Deep learning para detecção de ataques de phishing. Adaptada de (BASIT et al., 2021).*

2.5.2.3 Processamento de Linguagem Natural (NLP)

O Processamento de Linguagem Natural (NLP) é amplamente utilizado na detecção de *phishing* para analisar textos e identificar padrões suspeitos em mensagens fraudulentas. Modelos como BERT Devlin *et al.* (2019) e modelos

baseados na arquitetura transformer Vaswani *et al.* (2017) permitem a análise semântica de e-mails e sites de *phishing*, detectando inconsistências no uso da linguagem (ALHOGAIL; ALSABIH, 2021).

2.5.2.4 Análise Comportamental e Sistemas Baseados em Anomalias

A análise comportamental utiliza IA para monitorar padrões de interação dos usuários e detectar atividades incomuns. Por exemplo, se um usuário acessar um site e inserir credenciais de forma inesperada, o sistema pode levantar um alerta. Já os sistemas baseados em anomalias identificam padrões de tráfego incomuns, como picos de acessos a domínios recentemente registrados, que podem indicar ataques de *phishing* em larga escala (BASIT *et al.*, 2021).

2.5.2.5 Uso de Redes Generativas Adversariais (GANs)

As Redes Generativas Adversariais (GANs) têm sido empregadas para gerar exemplos de ataques de *phishing* e melhorar a precisão de detecção. Essa abordagem permite que os modelos de IA sejam treinados continuamente com novos padrões de ameaças, aumentando sua robustez contra ataques emergentes(ALAHMADI; ALOTAIBI; ALSALEH, 2022).

2.6 Impacto Econômico dos Ataques de *phishing* e Desafios nas Técnicas de Detecção

Os ataques de *phishing* causam impactos econômicos significativos tanto para indivíduos quanto para organizações. O custo desses ataques vai além das perdas financeiras diretas, incluindo também custos com recuperação, perda de confiança do consumidor e danos à reputação da marca. As organizações frequentemente enfrentam despesas substanciais para mitigar as consequências de ataques de *phishing* bem-sucedidos. Esses custos incluem a necessidade de reforçar as medidas de segurança, realizar auditorias de segurança, e até mesmo pagar multas regulatórias em casos de violação de dados. Além disso, o tempo de inatividade causado por tais ataques pode resultar em perda de produtividade e receita.

Conforme mencionado por Aung, Zan e Yamana (2019), a detecção eficaz de *phishing* é essencial para prevenir fraudes e ataques cibernéticos. No entanto, a detecção de *phishing* enfrenta vários desafios, incluindo a evolução constante das técnicas de *phishing* e a dificuldade em distinguir entre comunicação legítima e maliciosa. Os ataques de *phishing* frequentemente utilizam engenharia social para

se adaptar a novos contextos e explorar as fraquezas dos usuários, tornando ainda mais difícil identificar esses ataques de maneira eficaz.

As técnicas de *phishing* podem causar impactos severos, como a perda de dados pessoais e financeiros, e danos à reputação das organizações. Compreender esses impactos é fundamental para o desenvolvimento de estratégias de detecção e prevenção mais eficazes, que possam minimizar as perdas e reduzir os danos decorrentes desses ataques cibernéticos.

2.7 Tendências Futuras e Inovações

Inovações tecnológicas, como o uso de inteligência artificial, estão melhorando a capacidade de detectar *phishing*. Novas ferramentas estão sendo desenvolvidas para enfrentar a sofisticação crescente dos ataques. Estudos recentes mostram que a colaboração entre plataformas de redes sociais, especialistas em segurança cibernética e pesquisadores acadêmicos é fundamental para enfrentar a ameaça de *phishing*. Essa cooperação facilita a troca de informações e o desenvolvimento de soluções. Tendências atuais indicam o uso crescente de inteligência artificial e machine learning na detecção e mitigação do *phishing*. A pesquisa contínua e a colaboração são essenciais para lidar com as técnicas de *phishing* em evolução e proteger dados e reputação das organizações (ALHOGAIL; ALSABIH, 2021).

3 METODOLOGIA

Este capítulo descreve a metodologia adotada para a realização deste estudo, detalhando o tipo de pesquisa, os procedimentos de coleta de dados, a análise dos resultados e os critérios de inclusão e exclusão dos estudos revisados. O objetivo principal é apresentar uma abordagem clara e reprodutível para a revisão da literatura, com foco na detecção de ataques de *phishing*, especialmente em redes sociais, e na aplicação de técnicas de Inteligência Artificial. A metodologia aqui apresentada segue diretrizes bem estabelecidas para garantir a qualidade e a robustez dos resultados.

3.1 Tipo de Pesquisa

Este trabalho adota uma abordagem qualitativa baseada em uma Revisão Sistemática da Literatura, seguindo as diretrizes propostas por (PETTICREW; ROBERTS, 2008). O objetivo principal é reunir, analisar e comparar estudos acadêmicos sobre a detecção de ataques de *phishing*, com ênfase em redes sociais e no uso de Inteligência Artificial. A metodologia foi estruturada para garantir a reprodutibilidade e a validade dos resultados, seguindo um processo sistemático de coleta, análise e interpretação dos dados.

3.2 Delimitação do Estudo

A pesquisa foi delimitada com base nos seguintes critérios:

- **Foco temático**: Análise de técnicas tradicionais e baseadas em IA para detecção de phishing, com ênfase em métodos aplicáveis a redes sociais.
- **Contexto de aplicação**: Ataques de *phishing* em redes sociais, sem excluir abordagens aplicadas a e-mails e outras plataformas digitais.
- **Período de publicação**: Artigos publicados nos últimos 10 anos (2014-2024), garantindo a atualidade das técnicas e abordagens discutidas.
- **Bases de dados**: Seleção de artigos indexados nas principais bases acadêmicas da area da computação, como ACM Digital Library, IEEE Digital Library, ScienceDirect e arXiv.

3.3 Método para Revisão Sistemática

Para conduzir a revisão sistemática, foi utilizada a técnica (Population, Intervention, Comparison, Outcomes, Context) (PICOC), uma abordagem estruturada que auxilia na priorização e avaliação da relevância dos artigos selecionados (NEIVA; SILVA, 2016). A ferramenta Parsif.al[1] foi utilizada para organizar e classificar os artigos de acordo com os critérios PICOC. O processo foi dividido nas seguintes etapas:

- **Population**: Estudos focados na detecção de *phishing*, principalmente em redes sociais. Estes artigos devem tratar especificamente de como identificar ataques de *phishing* em plataformas online, considerando os comportamentos dos usuários e as características das plataformas sociais.
- **Intervention**: Aplicação de técnicas de IA e aprendizado de máquina para detectar *phishing*. A revisão irá avaliar como essas técnicas modernas melhoram a detecção de *phishing* em comparação com métodos tradicionais, destacando as abordagens baseadas em IA, como redes neurais, aprendizado supervisionado e não supervisionado.
- **Comparison**: Métodos tradicionais de detecção de *phishing*, como listas negras, filtros de spam e regras heurísticas. A comparação ajudará a identificar as vantagens das técnicas baseadas em IA e aprendizado de máquina, como a capacidade de adaptação a novos tipos de ataques e a redução de falsos positivos.
- **Outcome**: Precisão e eficiência na detecção de *phishing*. A precisão refere-se à capacidade das técnicas de identificar corretamente os ataques de *phishing*, enquanto a eficiência avalia a rapidez e o custo das técnicas, incluindo o uso de recursos computacionais e a necessidade de treinamento de modelos.
- **Context**: Aplicações em redes sociais e outros ambientes online, considerando como o contexto específico das plataformas pode influenciar a eficácia das técnicas de detecção. Isso inclui a consideração de características das redes sociais, como a interação dos usuários e os tipos de dados compartilhados, que podem impactar a detecção de ataques.

O processo foi dividido nas seguintes etapas:

1. **String de busca**: Para garantir a reprodutibilidade da pesquisa, a string de busca utilizada foi a seguinte:

[1] `https://parsif.al/`

("Phishing Detection"OR "Phishing Survey"OR "Phishing Review"OR "Phishing social networks"OR "Social Media Phishing"OR "Deep learning phishing"OR "Machine learning phishing"OR "Phishing detection IA")

2. **Definição de Palavras-chave**: Foram utilizados termos relacionados a *phishing*, *segurança da informação*, *inteligência artificial*, *machine learning* e *redes sociais*. As palavras-chave foram combinadas com operadores booleanos (AND, OR) para refinar a busca.
3. **Busca nas Bases de Dados**: A busca foi realizada nas seguintes bases de dados:
   - ACM Digital Library
   - IEEE Digital Library
   - ScienceDirect
   - arXiv
4. **Filtragem dos Artigos**: Os artigos foram filtrados com base em critérios de relevância e qualidade, considerando o título, resumo e palavras-chave. Artigos fora do escopo temático ou com baixa relevância foram excluídos.
5. **Análise e Extração de Dados**: Foram identificadas as técnicas discutidas, os métodos aplicados e os principais resultados de cada estudo. Os dados foram organizados em uma planilha para facilitar a análise comparativa.
   Foram analisados os algoritmos utilizados, a precisão das técnicas, a taxa de falsos positivos e a eficiência computacional. Essa análise permitiu identificar as abordagens mais eficazes para a detecção de phishing em redes sociais.

3.4 Procedimentos para Filtragem dos Artigos

A seleção dos artigos foi conduzida em duas fases principais:

- **Triagem Inicial**: Nesta fase, foram excluídos artigos com base na análise do título, resumo e palavras-chave. A triagem inicial buscou identificar estudos que abordassem diretamente a detecção de *phishing*, com foco em redes sociais e técnicas de IA.
- **Avaliação de Qualidade**: Os artigos restantes foram avaliados com base nos seguintes critérios, com pontuação de 0 a 2 para cada item:
  - Presença das palavras-chave no título ou resumo.
  - Publicação em periódico ou conferência relevante.
  - Discussão de técnicas de detecção de *phishing*.

  - Uso de IA para mitigar ataques.
  - Enfoque na detecção em redes sociais.

Os artigos que atingiram uma pontuação igual ou superior a 8 foram selecionados para análise aprofundada.

3.5 Interpretação dos Dados

A análise dos dados foi realizada em três etapas principais:

- **Classificação dos Trabalhos**: Os artigos foram divididos em duas categorias principais: abordagens tradicionais e abordagens baseadas em IA. Essa classificação permitiu uma análise comparativa entre as técnicas.
- **Comparativo das Técnicas**: Foram analisados os algoritmos utilizados, a precisão das técnicas, a taxa de falsos positivos e a eficiência computacional. Essa análise permitiu identificar as abordagens mais eficazes para a detecção de *phishing* em redes sociais.
- **Mapeamento das Tendências**: Foram identificadas lacunas na literatura e tendências futuras na área de detecção de *phishing*, com foco em redes sociais e o uso de IA.

A partir desses resultados, foram extraídas recomendações práticas para a mitigação de ataques de *phishing* em redes sociais.

3.6 Estratégia de Apresentação dos Resultados

Os achados da pesquisa foram organizados e apresentados da seguinte forma:

- **Tabelas comparativas**: Foram criadas tabelas para destacar as principais técnicas de detecção de *phishing*, suas eficiências e limitações. As tabelas permitem uma visão clara das vantagens e desvantagens de cada abordagem.
- **Gráficos estatísticos**: Gráficos foram utilizados para representar a distribuição dos artigos selecionados por ano de publicação e base de dados, além de tendências observadas na literatura.
- **Discussão crítica**: Os resultados foram discutidos de forma crítica, relacionando os dados coletados às questões de pesquisa estabelecidas. A discussão incluiu uma análise das lacunas identificadas e sugestões para futuras pesquisas.

3.7 Limitações da Pesquisa

Como qualquer revisão sistemática, este estudo apresenta algumas limitações:

- **Restrição às bases de dados selecionadas**: Algumas pesquisas relevantes podem não ter sido incluídas.
- **Limitação temporal**: Trabalhos mais antigos, que poderiam fornecer perspectivas históricas, foram excluídos.
- **Viés na seleção de artigos**: A triagem inicial foi baseada em título e resumo, podendo excluir artigos relevantes por erro de categorização.

4 RESULTADOS DA REVISÃO DA LITERATURA

Este capítulo apresenta os resultados obtidos na revisão sistemática da literatura, incluindo a quantidade de artigos selecionados, as bases de dados utilizadas, os critérios de qualidade empregados e uma análise segmentada das abordagens encontradas. Além disso, discutimos as principais ferramentas e métodos utilizados na detecção de *phishing* em redes sociais, bem como apresentamos respostas para cada pergunta de pesquisa levantada na seção 1.3.

4.1 Importação e Seleção dos Artigos

Os artigos foram importados no dia 19/11/2024, considerando um período máximo de 10 anos desde a publicação. As bases de dados consultadas foram selecionadas por sua relevância na área de segurança da informação e inteligência artificial. A Tabela 1 apresenta a distribuição dos 170 artigos importados por base de dados a partir da string de busca definida na seção 3.3.

Base de Dados	**Número de Artigos Importados**
ACM Digital Library	21
IEEE Digital Library	50
Science@Direct	25
arXiv	74
Total	**170**

Tabela 1: *Número de artigos importados por base de dados.*

A Figura 4 apresenta a distribuição percentual dos artigos importados, evidenciando a predominância de estudos provenientes do arXiv e da IEEE Digital Library.

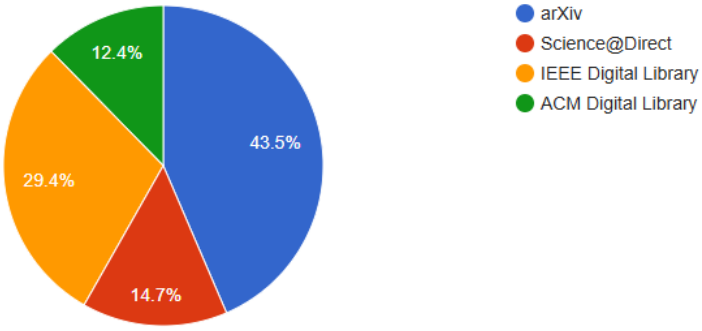

Figura 4: *Distribuição dos artigos importados por base de dados.*

4.2 Critérios de Qualidade e Seleção

A seleção dos artigos foi realizada com base em seis questões de qualidade, conforme detalhado na Tabela 2. A abordagem adotada permitiu filtrar artigos que efetivamente contribuem para o tema de detecção de *phishing* em redes sociais, assegurando uma análise robusta e relevante.

- As palavras-chave aparecem no título e/ou no resumo do trabalho?
- É um trabalho publicado em periódico ou conferência conceituada?
- O trabalho é focado em técnicas de detecção de *phishing*?
- O trabalho aborda o uso de técnicas de IA para combater os ataques de phishing?
- O trabalho utiliza técnicas tradicionais contra os ataques de *phishing*?
- O trabalho é sobre *phishing* em redes sociais?

Os artigos foram pontuados conforme os critérios estabelecidos, seguindo a seguinte escala:

Descrição	**Peso**
Sim	2.0
Parcialmente	1.0
Não	0.0

Tabela 2: *Pontuação atribuída para cada questão de qualidade.*

Os artigos que atingiram uma pontuação mínima de 8,0 foram selecionados para análise detalhada, resultando em um total de 23 artigos.

4.3 Segmentação dos Trabalhos por Categoria

Com base na análise dos artigos extraídos, os trabalhos foram segmentados nas seguintes categorias principais:

- **Trabalhos que utilizam IA:** Incluem abordagens baseadas em aprendizado de máquina, deep learning e modelos híbridos. Exemplos incluem:
  - **PhishAri:** Utiliza Random Forest para detecção de *phishing* no Twitter, analisando URLs e características de tweets (AGGARWAL; RAJADESINGAN; KUMARAGURU, 2012).
  - **HTMLPhish:** Emprega Redes Neurais Convolucionais (CNNs) para analisar o código HTML de páginas suspeitas, independentemente do idioma (OPARA; WEI; CHEN, 2020).
  - **THEMIS:** Combina RCNNs aprimorados com mecanismos de atenção para análise de e-mails, focando em cabeçalhos e corpos de mensagens (FANG *et al.*, 2019a).
  - **URLNet:** Utiliza CNNs para aprender representações de URLs automaticamente, capturando padrões semânticos e sequenciais (LE *et al.*, 2018).
  - **Web2Vec:** Combina CNNs e Bidirectional Long Short-Term Memory (BiLSTM) com mecanismo de atenção para capturar características de URLs, HTML e estrutura DOM (FENG *et al.*, 2020).
  - **PDGAN:** Modelo baseado em GANs que utiliza LSTM para gerar URLs de *phishing* e CNN para discriminar URLs legítimos (AL-AHMADI; ALOTAIBI; ALSALEH, 2022).
  - **VisualPhishNet:** Aplica Redes Neurais Convolucionais Triplet para detecção de phishing com base na similaridade visual (ABDELNABI; KROMBHOLZ; FRITZ, 2020).
- **Trabalhos que não utilizam IA:** Técnicas baseadas em regras heurísticas e listas negras, como:
  - **Phishwish:** Utiliza um conjunto de 11 regras configuráveis para detectar phishing em e-mails, sem a necessidade de treinamento ou uso de IA (COOK; GURBANI; DANILUK, 2008).
  - **BogusBiter:** Abordagem que envia credenciais falsas para sites de *phishing*, dificultando a identificação de credenciais reais pelos atacantes (YUE; WANG, 2010).
- **Trabalhos focados exclusivamente em redes sociais:** Estudos que analisam *phishing* em plataformas como Twitter e Facebook. Exemplos incluem:

  - **PhishAri:** Desenvolvido especificamente para identificar *phishing* no Twitter em tempo real, utilizando URLs e características de tweets (AGGARWAL; RAJADESINGAN; KUMARAGURU, 2012).

- **Trabalhos focados em e-mails:** Abordam a detecção de *phishing* em e-mails utilizando diferentes técnicas, como:
  - **THEMIS:** Combina Region-based Convolutional Neural Network (RCNNs) aprimorados com mecanismos de atenção para análise de e-mails, focando em cabeçalhos e corpos de mensagens (FANG *et al.*, 2019a).
  - **SEAHound:** Emprega Processamento de Linguagem Natural (NLP) para analisar e-mails e detectar *phishing* com base na semântica (PENG; HARRIS; SAWA, 2018).
  - **Phishwish:** Utiliza regras configuráveis para detectar *phishing* em e-mails, sem a necessidade de treinamento ou uso de IA (COOK; GURBANI; DANILUK, 2008).
- **Trabalhos focados em URLs:** Métodos voltados para a detecção de *phishing* baseado na estrutura de URLs, como:
  - **URLNet:** Utiliza CNNs para aprender representações de URLs automaticamente, capturando padrões semânticos e sequenciais (LE *et al.*, 2018).
  - **A New Method for URL Detection:** Utiliza Random Forest para classificação de URLs de phishing, com foco em atributos específicos de URLs (PAREKH *et al.*, 2018).
  - **PhishStorm:** Detecta URLs de *phishing* em tempo real usando análise de relacionamento intra-URL e aprendizado de máquina (MARCHAL *et al.*, 2014).
  - **AntiPhishStack:** Modelo de empilhamento em duas fases que utiliza LSTM e XGBoost para detecção de URLs de *phishing* (ASLAM *et al.*, 2024).
  - **PhishNet:** Utiliza heurísticas para prever URLs maliciosas e correspondência aproximada para detecção de *phishing* (PRAKASH *et al.*, 2010).

A segmentação desses trabalhos permite uma melhor compreensão das abordagens predominantes e de como diferentes técnicas são aplicadas na detecção de phishing em diversos contextos. Os detalhes completos da análise dos artigos podem ser encontrados na planilha de extração de informações, disponível como referência para este estudo.

4.4 Análise Comparativa das Ferramentas de Detecção de *Phishing*

Para aprofundar a compreensão das diversas abordagens na detecção de *phishing*, elaboramos uma planilha comparativa que avalia diferentes ferramentas e técnicas descritas na literatura. A planilha está disponível no seguinte link: Planilha comparativa de detecção de *phishing*.

4.4.1 Descrição da Planilha

A planilha apresentada oferece uma comparação detalhada de diversas ferramentas de detecção de *phishing*, com foco em diferentes critérios para avaliar suas abordagens e capacidades. Ela organiza as informações considerando os seguintes parâmetros:

- **Solução:** Nome da ferramenta ou técnica analisada.
- **Abordagem:** Método principal utilizado para detectar o *phishing*, que pode variar entre técnicas automáticas, baseadas em regras ou modelos de aprendizado de máquina.
- **Canal de Comunicação:** Indica o meio ou plataforma em que o *phishing* é detectado, como e-mail, navegador web ou redes sociais.
- **Utiliza IA?** Informa se a ferramenta emprega inteligência artificial para realizar a detecção de *phishing*, algo essencial para avaliar a modernidade e eficácia das abordagens.
- **Artigo/Fonte:** Referência bibliográfica do estudo ou artigo que descreve a ferramenta.
- **Destaque:** Um ponto distintivo ou característica relevante de cada ferramenta que a torna única ou particularmente eficaz.

Esses critérios são organizados em uma tabela que apresenta as soluções de forma comparativa. Por exemplo, o **PhishAri**, que realiza detecção automática de *phishing* no Twitter, utiliza inteligência artificial para combinar recursos de URL e dados do Twitter, superando métodos tradicionais como listas negras. Outras ferramentas, como o **BogusBiter**, que é uma barra de ferramentas de navegador, não utilizam IA, mas oferecem soluções complementares aos métodos tradicionais.

A tabela também inclui técnicas de detecção mais avançadas, como o **AntiPhishStack**, que utiliza um modelo de generalização em duas fases para detectar *phishing* via URLs, e **VisualPhishNet**, que emprega detecção por similaridade visual, utilizando um dataset robusto de capturas de tela.

A planilha, apresentada a seguir, permite uma visão das diversas técnicas de detecção de *phishing*, destacando como a inteligência artificial está sendo aplicada em soluções mais modernas e complexas, e como elas se comparam com os métodos tradicionais.

Várias das técnicas listadas na Tabela 3 são empregadas em soluções comerciais de segurança da informação. Métodos baseados em aprendizado de máquina, como Random Forest Classifier, Phishing URL Detection e BERT e CNN, são amplamente utilizados por empresas de segurança cibernética para a detecção automatizada de sites e e-mails fraudulentos. Soluções que analisam URLs, como URLNet e PhishNet, são frequentemente incorporadas a navegadores e sistemas de proteção de redes para prevenir acessos a sites maliciosos. Além disso, ferramentas como VisualPhishNet, que detectam phishing por similaridade visual, têm sido exploradas comercialmente para proteger marcas contra ataques baseados em clonagem de páginas. Algumas técnicas, como PhishStorm, que oferece análise de dados em tempo real, são utilizadas por empresas que lidam com grandes volumes de tráfego e necessitam de escalabilidade na detecção de ameaças. Embora muitas das soluções acadêmicas ainda exijam adaptação para uso comercial, a crescente demanda por segurança contra phishing impulsiona a adoção dessas abordagens em produtos e serviços voltados à proteção digital.

Tabela 3: *Comparação de ferramentas de detecção de phishing. Elaborado pelos autores.*

Solução	**Abordagem**	**Canal de Comunicação**	**Utiliza IA?**	**Artigo/Fonte**	**Destaque**
PhishAri	Detecção automática de *phishing* em tempo real no Twitter	Twitter	Sim	(AGGARWAL; RAJADESINGAN; KUMARAGURU, 2012)	Combinação de recursos de URL e Twitter superou listas negras tradicionais.
BogusBiter	Barra de ferramentas do navegador que envia informações falsas em formulários HTML	Navegador Web (HTTPS)	Não	(YUE; WANG, 2010)	Funciona de forma transparente, sem intervenção do usuário, e complementa métodos tradicionais de detecção.
PhishGuard	Plug-in de navegador para detectar *phishing* com base em testes de autenticação HTTP Digest	Navegador Web (HTTPS)	Não	(OVI; RAHMAN; HOSSAIN, 2024)	Supera modelos tradicionais e é adaptável a ambientes de IoT.

Solução	**Abordagem**	**Canal de Comunicação**	**Utiliza IA?**	**Artigo/Fonte**	**Destaque**
Phishwish	Filtro de *phishing* baseado em regras mínimas e sem estado	E-mail (conteúdo e cabeçalhos)	Não	(COOK; GURBANI; DANILUK, 2008)	Não requer treinamento ou listas negras.
HTMLPhish	Detecção de *phishing* em HTML utilizando aprendizado de máquina	Web/HTML	Sim	(OPARA; WEI; CHEN, 2020)	Independência de idioma e detecção no lado do cliente.
AntiPhishStack	Modelo de generalização em duas fases	URLs	Sim	(ASLAM *et al.*, 2024)	Extração automática de recursos sem dependência de especialistas.
VisualPhishNet	Detecção de *phishing* por similaridade visual	Websites e páginas web	Sim	(ABDELNABI; KROMBHOLZ; FRITZ, 2020)	Dataset Phish Visual com 9.363 capturas de tela.

Solução	**Abordagem**	**Canal de Comunicação**	**Utiliza IA?**	**Artigo/Fonte**	**Destaque**
BERT e CNN	Extração de características linguísticas com BERT e classificação com CNN	E-mails corporativos	Sim	(GUPTA *et al.*, 2024)	Supera modelos como Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM) e Support Vector Machine (SVM) em precisão e eficiência.
THEMIS	Modelo RCNN aprimorado com mecanismo de atenção	E-mails	Sim	(FANG *et al.*, 2019b)	Supera modelos baseados em CNN e LSTM ao capturar melhor informações semânticas.
Phishing URL Detection	Detecção baseada em URLs utilizando Random Forest e seleção de atributos	URLs	Sim	(PAREKH *et al.*, 2018)	Método eficiente, mas não detecta ataques futuros sem aprendizado contínuo.

Solução	**Abordagem**	**Canal de Comunicação**	**Utiliza IA?**	**Artigo/Fonte**	**Destaque**
Phishing Detection System	Detecção de sites de *phishing* utilizando Random Forest e validação cruzada	URLs	Sim	(SUBASI *et al.*, 2017)	Random Forest superou K-Nearest Neighbors (k-NN), SVM e Artificial Neural Network (ANN) em precisão e eficiência.
PDGAN	Detecção de *phishing* baseada em URLs utilizando GANs	URLs	Sim	(AL-AHMADI; ALOTAIBI; ALSALEH, 2022)	GAN melhora a generalização para detectar URLs não vistas no treinamento.
URLNet	Detecção de URLs maliciosos usando aprendizado profundo	URLs	Sim	(LE *et al.*, 2018)	Captura padrões semânticos e sequenciais, melhorando a generalização para URLs desconhecidos.

Solução	**Abordagem**	**Canal de Comunicação**	**Utiliza IA?**	**Artigo/Fonte**	**Destaque**
Web2Vec	Detecção de *phishing* usando aprendizado profundo com múltiplas dimensões de características	Webpages (URLs, HTML, DOM)	Sim	(FENG *et al.*, 2020)	Integração de aprendizado de representação em NLP para análise automática de páginas web.
GoldPhish	Análise de conteúdo baseada em imagens	Web	Não	(DUNLOP; GROAT; SHELLY, 2010)	Identifica ataques zero-day sem depender de listas negras.
Phishing-Alarm	Similaridade de componentes de página via Cascading Style Sheets (CSS)	World Wide Web (WEB)	Não	(MAO *et al.*, 2017)	Abordagem resistente a evasões e de fácil integração em navegadores.
SEAHound	Processamento de linguagem natural e aprendizado de máquina	E-mail	Sim	(PENG; HARRIS; SAWA, 2018)	Detecta ataques compostos apenas por texto e cria listas negras de pares verbo-objeto maliciosos.

Solução	**Abordagem**	**Canal de Comunicação**	**Utiliza IA?**	**Artigo/Fonte**	**Destaque**
OFS-NN	Seleção ótima de características e rede neural	Web	Sim	(ZHU *et al.*, 2019)	Combina alta precisão com detecção rápida, sendo eficaz contra ataques desconhecidos.
Combinação de recursos textuais e visuais	Combinação de recursos DOM e visuais para detecção de páginas clonadas	Web	Sim	(DOOREMAAL *et al.*, 2021)	Identifica ataques sofisticados que evitam o uso de marcas no DOM.
Random Forest Classifier	Classificador Random Forest para detecção de sites de *phishing*	Web	Sim	(SUBASI *et al.*, 2017)	Supera outros classificadores devido à rapidez e robustez.
PhishNet	Lista negra preditiva	Navegador Web	Não	(PRAKASH *et al.*, 2010)	Mais rápido que a Navegação Segura do Google, prevendo ataques antes que se tornem públicos.

Solução	**Abordagem**	**Canal de Comunicação**	**Utiliza IA?**	**Artigo/Fonte**	**Destaque**
GoldPhish	Optical Character Recognition (OCR) e PageRank	Navegador Web	Sim	(DUNLOP; GROAT; SHELLY, 2010)	Identifica ataques zero-day sem depender de listas negras.
PhishStorm	Análise de dados em tempo real com relacionamento intra-URL	E-mail, HTTP Proxy	Sim	(MARCHAL *et al.*, 2014)	Arquitetura escalável baseada em STORM e Filtro de Bloom permite processamento eficiente em Big Data.

A análise da planilha comparativa revela uma diversidade de abordagens no combate aos ataques de *phishing*, refletindo a evolução das técnicas e tecnologias utilizadas. A maioria das soluções emprega métodos baseados em inteligência artificial, especialmente aprendizado de máquina e redes neurais, o que demonstra um crescente foco na automatização e melhoria da precisão na detecção.

A utilização de IA, como pode ser observado nos artigos de (AGGARWAL; RAJADESINGAN; KUMARAGURU, 2012), (OPARA; WEI; CHEN, 2020), e (ABDELNABI; KROMBHOLZ; FRITZ, 2020), permite a identificação de padrões complexos em URLs, conteúdo visual e até mesmo em interações dentro de plataformas como o Twitter. Ferramentas como o **VisualPhishNet** ((ABDELNABI; KROMBHOLZ; FRITZ, 2020)), por exemplo, são notáveis por sua capacidade de detectar phishing visualmente, utilizando um banco de dados robusto de capturas de tela, o que representa um avanço significativo na detecção de ataques que podem ser difíceis de identificar apenas por análise textual.

Entretanto, também há uma variedade de abordagens que não utilizam IA, como o **BogusBiter** ((YUE; WANG, 2010)), que funciona como uma barra de ferramentas de navegador e complementa métodos tradicionais sem a necessidade de treinamento complexo. Essas abordagens ainda são úteis, especialmente para cenários de baixo risco ou onde a simplicidade e a transparência são fatores essenciais. Contudo, as soluções baseadas em IA tendem a oferecer maior precisão e escalabilidade, especialmente em ambientes dinâmicos e de alto volume de dados.

Outro ponto importante é o foco nas URLs. Diversos estudos, como os de (PAREKH *et al.*, 2018) e (SUBASI *et al.*, 2017), exploram modelos baseados em URLs para detectar *phishing*. Esses métodos são eficientes, mas dependem de aprendizado contínuo, como é o caso do ***Phishing URL Detection*** ((PAREKH *et al.*, 2018)), que não detecta ataques futuros sem atualização constante.

Por fim, a combinação de múltiplos recursos — como é o caso da solução **Web2Vec** ((FENG *et al.*, 2020)) que integra aprendizado de representação para páginas web — surge como uma tendência crescente, buscando ampliar a robustez da detecção. A versatilidade dessas abordagens permite uma análise mais aprofundada e eficiente, englobando diferentes tipos de dados e canais de comunicação.

Esses resultados reforçam a importância de um avanço contínuo em técnicas mais adaptativas e que combinem múltiplas fontes de dados, promovendo uma detecção mais precisa e eficaz, especialmente em redes sociais e outros ambientes digitais dinâmicos.

4.5 Discussão das Questões de Pesquisa

Nesta seção, apresentamos um consolidado das respostas às questões de pesquisa estabelecidas para este trabalho, com base na revisão sistemática da literatura realizada. Embora as questões tenham sido respondidas ao longo do texto, neste ponto, oferecemos um resumo integrado para cada uma delas. As respostas são fundamentadas em estudos recentes e nas técnicas discutidas ao longo do documento.

1. **Quais são os princípios fundamentais da engenharia social e como eles são aplicados no contexto de ataques de *phishing*?**
   A engenharia social é uma técnica que explora vulnerabilidades psicológicas e sociais para manipular indivíduos a divulgar informações confidenciais ou realizar ações que comprometem a segurança. No contexto de *phishing*, os atacantes se passam por entidades confiáveis, como bancos ou empresas, para enganar as vítimas e obter dados sensíveis, como senhas e informações financeiras. Técnicas como a criação de um senso de urgência e a manipulação da confiança são comumente utilizadas para persuadir as vítimas a agir de forma impulsiva, sem verificar a autenticidade da mensagem. Por exemplo, mensagens de *phishing* frequentemente simulam comunicações de instituições financeiras, alertando sobre atividades suspeitas e solicitando ações imediatas para evitar consequências (HIJJI; ALAM, 2021). Essa abordagem visa induzir pânico, levando o indivíduo a agir sem verificar a autenticidade da mensagem, o que é um dos pilares da engenharia social.

2. **Qual é o estado atual da pesquisa em detecção de ataques de *phishing*, incluindo técnicas tradicionais e aquelas que incorporam inteligência artificial?**
   A pesquisa atual em detecção de *phishing* abrange tanto técnicas tradicionais, como listas negras e análise heurística, quanto métodos baseados em inteligência artificial, como aprendizado de máquina e aprendizado profundo. Enquanto as técnicas tradicionais são eficazes contra ataques conhecidos, elas apresentam limitações frente a ataques novos e sofisticados. Por exemplo, listas negras dependem de atualizações constantes e podem deixar usuários vulneráveis a ataques emergentes (AUNG; ZAN; YAMANA, 2019). Por outro lado, as abordagens baseadas em IA, como Random Forest, Redes Neurais Convolucionais (CNNs) e Redes Neurais Recorrentes (RNNs), têm demonstrado alta precisão e capacidade de adaptação a novos padrões de ataques.Modelos como o PhishAri, que utiliza Random Forest para detecção de phishing no Twitter, alcançaram taxas de precisão superiores a 92%

(AGGARWAL; RAJADESINGAN; KUMARAGURU, 2012). Além disso, técnicas de deep learning, como as empregadas no HTMLPhish, têm mostrado eficácia na análise de páginas web, independentemente do idioma, com precisões acima de 93% (OPARA; WEI; CHEN, 2020). No entanto, é importante ressaltar que esses trabalhos não utilizaram a mesma base de dados, o que dificulta uma comparação direta. Para uma avaliação mais robusta e uma comparação eficaz entre as técnicas, seria interessante estabelecer um benchmark comum, a fim de determinar qual é a melhor abordagem para a detecção de phishing em diferentes contextos.

3. **Quais são os desafios específicos enfrentados na detecção de ataques de *phishing* em tempo real em plataformas de redes sociais?**
   A detecção de *phishing* em redes sociais enfrenta desafios únicos devido à natureza dinâmica e à rápida disseminação de conteúdo nessas plataformas. A criação de novos perfis e a propagação rápida de links maliciosos dificultam a aplicação eficaz de técnicas tradicionais, como listas negras. Além disso, a detecção em tempo real exige algoritmos eficientes e escaláveis, capazes de processar grandes volumes de dados em tempo hábil, o que pode ser um desafio computacional. Por exemplo, o PhishStorm, que utiliza análise de URLs em tempo real, alcançou uma precisão de 94,91%, mas enfrenta desafios de escalabilidade ao lidar com grandes volumes de dados (MARCHAL *et al.*, 2014). Outro desafio é a necessidade de adaptação contínua dos algoritmos para lidar com novas táticas de *phishing*, como o uso de deepfakes e técnicas de evasão (CNN Brasil, 2023).

4. **Como os ataques de *phishing* via redes sociais se diferem dos ataques tradicionais de *phishing* por e-mail em termos de estratégias e eficácia?**
   Os ataques de *phishing* via redes sociais tendem a ser mais personalizados e direcionados, aproveitando a confiança e as interações sociais entre os usuários. Em contraste, os ataques por e-mail geralmente envolvem mensagens em massa que se passam por instituições conhecidas. A eficácia dos ataques em redes sociais é aumentada pela capacidade de explorar conexões pessoais e pela rápida disseminação de conteúdo, o que pode levar a uma maior taxa de sucesso em comparação com os ataques por e-mail. Além disso, a natureza interativa das redes sociais permite que os atacantes criem cenários mais convincentes, como mensagens de amigos ou familiares solicitando ajuda financeira, o que aumenta a probabilidade de sucesso do ataque.

5. **Quais são as diferenças na detecção de ataques de *phishing* em redes sociais em comparação com outros canais, como e-mails e mensagens instantâneas?**
   A detecção de *phishing* em redes sociais requer abordagens que considerem a natureza interativa e dinâmica dessas plataformas. Diferentemente dos e-mails, onde a análise de conteúdo e URLs é predominante, a detecção em redes sociais pode envolver a análise de comportamento do usuário, interações sociais e até mesmo a similaridade visual de páginas. Por exemplo, o VisualPhishNet utiliza redes neurais convolucionais triplas para comparar capturas de tela de sites legítimos e *phishing*, alcançando uma precisão de 81% no top-1 (ABDELNABI; KROMBHOLZ; FRITZ, 2020). Além disso, a detecção em redes sociais deve ser capaz de lidar com a rápida propagação de links maliciosos e a criação de novos perfis fraudulentos, o que exige algoritmos mais ágeis e adaptativos.

6. **Como as empresas de redes sociais podem colaborar com especialistas em segurança cibernética e pesquisadores acadêmicos para desenvolver estratégias eficazes de detecção e prevenção de *phishing*?**
   A colaboração entre empresas de redes sociais, especialistas em segurança cibernética e pesquisadores acadêmicos é essencial para o desenvolvimento de soluções inovadoras e eficazes contra o *phishing*. Essa cooperação pode facilitar a troca de informações sobre novas ameaças, o desenvolvimento de algoritmos avançados de detecção e a implementação de práticas de segurança mais robustas. Por exemplo, a integração de técnicas de IA explicável e a análise comportamental do usuário podem ser aprimoradas por meio dessa colaboração multidisciplinar. Além disso, a criação de bancos de dados compartilhados, como o Phish Visual, que contém mais de 9.363 capturas de tela de sites de *phishing*, pode ajudar no treinamento de modelos de IA mais precisos (ABDELNABI; KROMBHOLZ; FRITZ, 2020). A colaboração também pode facilitar a implementação de políticas de segurança mais rigorosas e a conscientização dos usuários sobre os riscos do *phishing*.

7. **Quais são os critérios mais relevantes para avaliar a eficácia das técnicas de detecção de ataques de *phishing*, e como esses critérios podem variar dependendo do contexto?**
   Os critérios mais relevantes para avaliar a eficácia das técnicas de detecção de *phishing* incluem a precisão, a taxa de falsos positivos, a capacidade de

detecção em tempo real e a adaptabilidade a novos tipos de ataques. Em contextos como redes sociais, onde a dinâmica é mais rápida e complexa, a eficiência computacional e a escalabilidade também são fatores críticos. Além disso, a capacidade de detectar ataques dia zero e a resistência a técnicas de evasão são aspectos importantes a serem considerados na avaliação de diferentes abordagens de detecção de *phishing*. A resistência a técnicas de evasão refere-se à habilidade de uma técnica ou sistema de detecção de resistir a estratégias utilizadas por cibercriminosos para modificar ou disfarçar seus ataques de forma que eles passem despercebidos pelos sistemas tradicionais de segurança. Isso pode envolver a alteração de padrões de comportamento, a manipulação de características dos sites ou e-mails fraudulentos, ou o uso de métodos para dificultar a identificação do ataque. Por exemplo, o PhishGuard, que utiliza modelos de ensemble com Random Forest e XGBoost, alcançou uma precisão de 99,05%, mas sua eficácia em ambientes de IoT, como em dispositivos conectados à rede, ainda precisa ser avaliada (OVI; RAHMAN; HOSSAIN, 2024). Em um cenário de IoT, a detecção de *phishing* pode ser desafiada pela diversidade de dispositivos, conexões e protocolos, o que torna a aplicação de técnicas convencionais de detecção mais complexa. Em redes sociais, onde a propagação de links maliciosos é rápida, a capacidade de detecção em tempo real e a escalabilidade são critérios ainda mais importantes.

4.6 Discussão dos Resultados

A detecção de ataques de *phishing*, especialmente no contexto de redes sociais, tem sido amplamente estudada, com diversas abordagens emergindo ao longo dos anos. A revisão crítica da literatura busca não apenas consolidar o conhecimento existente, mas também identificar lacunas e desafios que ainda persistem, além de avaliar a evolução das técnicas empregadas na detecção desses ataques.

4.6.1 Métodos Tradicionais de Detecção de *Phishing*

Os métodos tradicionais de detecção de *phishing* baseiam-se em listas negras (blacklists), filtros heurísticos e análise baseada em regras. Embora essas abordagens tenham sido eficazes para detectar ataques conhecidos, apresentam limitações significativas quando confrontadas com ataques de *phishing* novos ou sofisticados, que não estão previamente catalogados.

4.6.1.1 Limitações dos Métodos Tradicionais

- **Dependência de atualizações**: As listas negras precisam ser constantemente atualizadas, o que pode deixar usuários vulneráveis a ataques emergentes.
- **Alta taxa de falsos fegativos**: Novos ataques podem não ser detectados, especialmente aqueles que utilizam engenharia social avançada.
- **Baixa Eficiência em Redes Sociais**: A dinâmica das redes sociais, com a criação de novos perfis e disseminação rápida de links, dificulta a aplicação eficaz dessas técnicas.

4.6.2 Abordagens baseadas em Inteligência Artificial

O uso de IA, especialmente aprendizado de máquina e aprendizado profundo, tem revolucionado a detecção de *phishing*. Modelos treinados com grandes volumes de dados conseguem identificar padrões suspeitos e melhorar a capacidade de detecção de ataques novos e desconhecidos.

4.6.2.1 Principais Técnicas de IA Aplicadas

- **Aprendizado supervisionado**: Modelos como SVM, Random Forest e Redes Neurais são treinados para classificar páginas e mensagens como phishing ou legítimas.
- **Aprendizado não supervisionado**: Técnicas como clustering podem detectar padrões anômalos sem a necessidade de uma base de dados previamente rotulada.
- **Processamento de linguagem natural (NLP)**: Utilizado para analisar textos de mensagens, detectando indícios de engenharia social e padrões comuns em ataques.

4.6.2.2 Desafios no uso de IA

- **Explicabilidade dos Modelos**: Modelos de aprendizado profundo, como redes neurais, frequentemente são considerados caixas-pretas, dificultando a interpretação dos resultados.
- **Evasão de Ataques**: Cibercriminosos podem manipular características das mensagens para enganar os modelos de IA.
- **Viabilidade Computacional**: Treinamento e implementação de modelos avançados podem demandar alta capacidade computacional.

4.6.3 Comparação Entre Métodos Tradicionais e Baseados em Inteligência Artificial

A Tabela 5 apresenta uma comparação entre técnicas tradicionais e abordagens baseadas em IA para detecção de *phishing*, considerando critérios como eficiência, taxa de falsos positivos, capacidade de adaptação e complexidade de implementação.

Critério	**Técnicas Tradicionais**	**Técnicas Baseadas em IA**
Eficiência	Baixa a moderada	Alta
Taxa de Falsos Positivos	Alta	Baixa
Capacidade de Adaptação	Limitada	Elevada
Complexidade de Implementação	Baixa	Alta

Tabela 4: *Comparativo entre técnicas tradicionais e baseadas em IA na detecção de phishing.*

A seguir, cada critério apresentado na Tabela 5 é explicado em detalhes:

- **Eficiência**: Refere-se à capacidade da técnica em identificar rapidamente e corretamente ataques de *phishing*. Técnicas baseadas em IA tendem a ser mais eficientes, pois conseguem analisar grandes volumes de dados em tempo real e detectar padrões sofisticados, enquanto métodos tradicionais podem ser mais lentos e menos precisos.
- **Taxa de falsos positivos**: Indica a frequência com que uma técnica identifica erroneamente um site ou mensagem legítima como *phishing*. Métodos tradicionais, como listas negras, geralmente possuem altas taxas de falsos positivos, pois não conseguem distinguir ataques sofisticados de comunicações legítimas. Abordagens baseadas em IA, especialmente aquelas que utilizam aprendizado profundo, tendem a ser mais precisas e reduzir esse problema.
- **Capacidade de adaptação**: Avalia o quão bem a técnica consegue detectar novos tipos de ataques que não estavam previamente registrados. Técnicas tradicionais, como listas de bloqueio, são limitadas, pois dependem de atualizações constantes. Já métodos baseados em IA podem aprender padrões e características de novos ataques, tornando-se mais eficazes contra *phishing* de dia zero.
- **Complexidade de implementação**: Refere-se ao nível de dificuldade para desenvolver, implementar e manter a técnica. Métodos tradicionais, como filtros baseados em regras e listas negras, são relativamente simples de implementar. Em contraste, soluções baseadas em IA exigem conhecimento avançado,

poder computacional e grandes conjuntos de dados para treinamento, tornando sua implementação mais complexa.

As abordagens baseadas em IA demonstram uma maior precisão na detecção de ataques de *phishing*, reduzindo a taxa de falsos positivos e se adaptando rapidamente a novas ameaças. No entanto, exigem maior poder computacional e um volume significativo de dados para treinamento eficaz (BASIT *et al.*, 2021). A seguir, são apresentados os motivos que explicam essas diferenças.

- **Precisão na detecção e redução de falsos positivos**: Técnicas tradicionais, como listas negras e filtros baseados em regras, funcionam verificando URLs, palavras-chave ou remetentes suspeitos. No entanto, essas abordagens são rígidas e não conseguem detectar ataques sofisticados que utilizam domínios novos ou técnicas de evasão, como pequenas variações em URLs legítimas (por exemplo, "g00gle.com" em vez de "google.com"). Em contrapartida, algoritmos de IA, como redes neurais e aprendizado profundo, analisam um conjunto mais amplo de características, incluindo o layout da página, padrões linguísticos e comportamento do usuário, resultando em uma detecção mais precisa e menos falsos positivos.
- **Adaptação a novas ameaças**: As técnicas tradicionais dependem de atualizações constantes, pois só identificam ataques conhecidos. Isso significa que ataques de *phishing* de dia zero podem facilmente passar despercebidos. Por outro lado, abordagens baseadas em IA utilizam aprendizado supervisionado ou não supervisionado para reconhecer padrões e detectar anomalias, permitindo que identifiquem ataques inéditos. Por exemplo, um modelo de aprendizado de máquina pode perceber que um novo site tem características semelhantes a sites fraudulentos anteriores, mesmo que o endereço URL nunca tenha sido visto antes.
- **Exigência de poder computacional**: A implementação de IA na detecção de *phishing* requer processamento intensivo, especialmente para técnicas como redes neurais profundas, que analisam grandes quantidades de dados e executam cálculos complexos. Em contraste, métodos tradicionais, como listas de bloqueio, exigem poucos recursos, pois simplesmente verificam a correspondência de um site ou e-mail com uma base de dados existente.
- **Necessidade de grandes volumes de dados para treinamento**: Modelos de IA precisam de conjuntos de dados extensos e diversificados para alcançar um alto nível de precisão. Isso inclui dados de e-mails maliciosos, URLs fraudulentas e perfis falsos em redes sociais. Sem um volume adequado de exemplos para treinamento, o modelo pode apresentar vieses ou falhas na

detecção. Já as técnicas tradicionais não precisam desse treinamento, pois funcionam com regras fixas e listas pré-definidas.

Exemplo comparativo: Suponha que um usuário receba um e-mail fraudulento fingindo ser de um banco. Um filtro tradicional pode bloquear esse e-mail apenas se o remetente já estiver em uma lista negra. No entanto, se o atacante utilizar um novo domínio semelhante ao original, o ataque pode passar despercebido. Em contrapartida, um sistema baseado em IA pode analisar o conteúdo do e-mail, o tom da mensagem e a estrutura dos links, identificando padrões comuns a ataques de *phishing*, mesmo que o endereço do remetente seja desconhecido.

A literatura indica que abordagens baseadas em IA superam métodos tradicionais em termos de precisão e adaptação a novos ataques. Entretanto, os métodos tradicionais ainda desempenham um papel importante quando combinados com IA, formando sistemas híbridos que maximizam a segurança.

Método	**Vantagens**	**Desvantagens**
Listas Negras	Rápido e fácil de implementar	Ineficaz contra novos ataques
Heurísticas	Detecta padrões básicos de *phishing*	Pode gerar muitos falsos positivos
Aprendizado de Máquina	Boa taxa de detecção e adaptação	Requer grande volume de dados
Redes Neurais	Alta precisão em padrões complexos	Alto custo computacional e difícil interpretação

Tabela 5: *Comparação entre métodos tradicionais e IA na detecção de phishing.*

4.7 Perspectivas Futuras e Tendências

Com o avanço das ameaças de *phishing*, novas abordagens estão sendo exploradas, como:

- **IA explicável (XAI)**: Desenvolvimento de modelos que oferecem maior transparência e interpretabilidade na detecção de *phishing*, permitindo uma melhor compreensão dos processos de decisão dos sistemas de IA.
- **Detecção baseada em comportamento do usuário**: Análise de padrões de comportamento dos usuários para identificar atividades suspeitas, especialmente em redes sociais, onde os ataques de *phishing* são mais comuns.

- **Integração com Blockchain**: Uso de blockchain para validar a autenticidade de sites e transações, garantindo maior segurança e proteção contra ataques de *phishing* baseados em falsificação de identidade.

A revisão da literatura evidencia que a inteligência artificial tem desempenhado um papel fundamental na melhoria da detecção de ataques de *phishing*, aumentando significativamente a precisão e a velocidade na identificação dessas ameaças. No entanto, ainda existem desafios a serem enfrentados, como a alta taxa de falsos positivos, a adaptação a novos tipos de ataques (*phishing* de dia zero) e a interpretabilidade dos modelos baseados em IA. Diante disso, a integração de métodos tradicionais, como listas de bloqueio e heurísticas, com abordagens modernas baseadas em aprendizado de máquina e deep learning pode resultar em um sistema de detecção mais robusto. Essa combinação permite equilibrar precisão, eficiência computacional e transparência nos processos, contribuindo para um ambiente digital mais seguro e confiável.

5 CONCLUSÃO

O presente trabalho teve como objetivo principal investigar e comparar as técnicas tradicionais e as baseadas em IA para a detecção de *phishing*, com foco especial em redes sociais. A revisão sistemática da literatura permitiu identificar as principais abordagens, desafios e tendências no campo da segurança cibernética, destacando a importância de métodos mais eficazes para combater uma das ameaças mais persistentes e em constante evolução no cenário digital.

5.1 Principais Achados

1. **Evolução do *Phishing* e seus impactos**: O *phishing* tem se mostrado uma ameaça crescente, especialmente com o aumento do uso de redes sociais e aplicativos de mensagens. A sofisticação dos ataques, aliada ao uso de técnicas de engenharia social, tem dificultado a detecção e a prevenção. O impacto econômico desses ataques é significativo, afetando tanto indivíduos quanto organizações, com perdas financeiras diretas e danos à reputação.
2. **Limitações das técnicas tradicionais**: As abordagens tradicionais, como listas negras e análise heurística, embora úteis para detectar ataques conhecidos, apresentam limitações significativas frente a ataques novos e sofisticados. A dependência de atualizações constantes e a alta taxa de falsos negativos são desafios que precisam ser superados, especialmente em ambientes dinâmicos como redes sociais.
3. **Potencial da Inteligência Artificial**: As técnicas baseadas em IA, particularmente o aprendizado de máquina e o aprendizado profundo, têm se mostrado promissoras na detecção de *phishing*. Modelos como Random Forest (RF), Redes Neurais Convolucionais (CNNs) e Redes Neurais Recorrentes (RNNs) têm alcançado altas taxas de precisão, superando métodos tradicionais. Além disso, a capacidade de adaptação a novos padrões de ataques e a detecção em tempo real são vantagens significativas dessas abordagens.
4. **Desafios no uso de IA**: Apesar dos avanços, o uso de IA na detecção de *phishing* ainda enfrenta desafios, como a necessidade de grandes volumes de dados para treinamento, a dificuldade de interpretação dos modelos (caixas-pretas) e a possibilidade de evasão por parte dos cibercriminosos, que podem manipular características das mensagens para enganar os algoritmos.

5. **Abordagens híbridas e tendências futuras**: A combinação de técnicas tradicionais e baseadas em IA tem se mostrado eficaz, oferecendo um equilíbrio entre precisão e eficiência. Além disso, tendências futuras apontam para o desenvolvimento de IA explicável, a análise comportamental do usuário e a integração de tecnologias como blockchain para aumentar a segurança e a autenticidade das informações.

5.2 Contribuições do Trabalho

Este trabalho contribui para o campo da segurança cibernética ao oferecer uma revisão das técnicas de detecção de *phishing*, destacando as vantagens e limitações de cada abordagem. A análise comparativa entre métodos tradicionais e baseados em IA fornece insights para pesquisadores e profissionais da área, sugerindo que a integração de diferentes técnicas pode ser a chave para uma detecção mais eficaz.

Além disso, o estudo identificou lacunas na literatura, como a necessidade de maior transparência e interpretabilidade nos modelos de IA, especialmente no contexto da detecção de phishing em redes sociais. Outra lacuna relevante é a falta de abordagens que integrem de forma eficaz a análise do comportamento do usuário em tempo real, permitindo a detecção proativa de ataques emergentes. Essas limitações representam oportunidades para futuras pesquisas, que podem explorar novas técnicas baseadas em aprendizado contínuo, análise contextual e métodos híbridos para aprimorar a precisão e a eficiência na detecção e prevenção de *phishing*.

5.3 Recomendações Práticas

Com base nos resultados obtidos, algumas recomendações práticas podem ser propostas:

1. **Adoção de modelos híbridos**: Organizações devem considerar a implementação de sistemas híbridos que combinem técnicas tradicionais e baseadas em IA para maximizar a eficácia na detecção de *phishing*.
2. **Investimento em IA explicável**: O desenvolvimento de modelos de IA que ofereçam maior transparência e interpretabilidade pode aumentar a confiança dos usuários e facilitar a identificação de falsos positivos.
3. **Foco no comportamento do usuário**: A análise do comportamento do usuário em redes sociais e outras plataformas pode ser uma abordagem promissora para detectar atividades suspeitas em tempo real.

4. **Colaboração entre plataformas e especialistas**: A cooperação entre plataformas de redes sociais, especialistas em segurança cibernética e pesquisadores acadêmicos é essencial para o desenvolvimento de soluções inovadoras e eficazes contra o *phishing*.
5. **Educação e conscientização**: Além das soluções técnicas, a educação e a conscientização dos usuários sobre os riscos do *phishing* e as melhores práticas de segurança são fundamentais para reduzir a eficácia desses ataques.

5.4 Considerações Finais

O *phishing* continua a ser uma ameaça significativa no cenário digital, e a evolução das técnicas de ataque exige uma resposta igualmente dinâmica e inovadora. Este trabalho demonstra que, embora a IA tenha elevado significativamente a eficácia na detecção de *phishing*, ainda há desafios a serem superados. A combinação de técnicas tradicionais e modernas, aliada a uma abordagem multidisciplinar que inclua educação, conscientização e colaboração entre diferentes atores, pode fornecer um ambiente digital mais seguro e resiliente contra essa ameaça.

Em suma, a detecção de *phishing* é um campo em constante evolução, e a pesquisa contínua, aliada à implementação de soluções práticas, será essencial para proteger indivíduos e organizações contra os crescentes riscos cibernéticos. Este trabalho espera contribuir para essa jornada, oferecendo uma base teórica e prática que possa guiar futuras pesquisas e implementações no campo da segurança cibernética.

5.5 Trabalhos Futuros

A partir dos achados deste trabalho, algumas direções para pesquisas futuras incluem:

- **Avaliação experimental das técnicas identificadas**: um próximo passo seria a implementação e comparação das técnicas identificadas na revisão, utilizando um mesmo conjunto de dados para medir sua eficácia em diferentes cenários.
- **Criação de *benchmarks* e padronização de avaliações**: A heterogeneidade dos *datasets* e métricas utilizadas nos estudos analisados dificulta comparações diretas. A criação de *benchmarks* públicos e bem documentados pode contribuir para a reprodutibilidade e avanço da área.
- **Investigação da explicabilidade dos modelos de IA**: Muitos modelos baseados em aprendizado profundo funcionam como caixas-pretas, dificultando

a adoção em ambientes reais. Estudos futuros podem explorar técnicas de IA explicável para aumentar a transparência dos resultados.

- **Estudo sobre a efetividade de abordagens híbridas**: A literatura sugere que a combinação de técnicas tradicionais com modelos de IA pode gerar melhores resultados. Pesquisas futuras podem focar na construção e validação de modelos híbridos.
- **Impacto do contexto e do comportamento do usuário**: Além das características técnicas, fatores comportamentais podem influenciar a efetividade das técnicas de detecção. Estudos que analisem o impacto da interação do usuário em ataques de *phishing* podem contribuir para a compreensão dos fatores que influenciam a detecção dessas ameaças.

Essas direções futuras podem complementar a presente revisão da literatura, fornecendo uma base para o desenvolvimento de estudos empíricos e novas abordagens para a detecção de *phishing*.

Referências Bibliográficas

ABDELNABI, Sahar; KROMBHOLZ, Katharina; FRITZ, Mario. Visualphishnet: Zero-day phishing website detection by visual similarity. *In*: **Proceedings of the 2020 ACM SIGSAC conference on computer and communications security**. [S.l.: s.n.], 2020. p. 1681–1698. 34, 39, 45, 48

AGGARWAL, Anupama; RAJADESINGAN, Ashwin; KUMARAGURU, Ponnurangam. Phishari: Automatic realtime phishing detection on twitter. *In*: IEEE. **2012 eCrime Researchers Summit**. [S.l.], 2012. p. 1–12. 34, 35, 38, 45, 47

AL-AHMADI, Saad; ALOTAIBI, Afrah; ALSALEH, Omar. Pdgan: Phishing detection with generative adversarial networks. **Ieee Access**, IEEE, v. 10, p. 42459–42468, 2022. 25, 34, 41

ALEROUD, Ahmed; ZHOU, Lina. Phishing environments, techniques, and countermeasures: A survey. **Computers & Security**, Elsevier, v. 68, p. 160–196, 2017. 12, 23

ALHOGAIL, Areej; ALSABIH, Afrah. Applying machine learning and natural language processing to detect phishing email. **Computers & Security**, Elsevier, v. 110, p. 102414, 2021. 25, 26

ALWANAIN, Mohammed I. Phishing awareness and elderly users in social media. **Int J Comput Sci Netw Secur**, v. 20, n. 9, p. 114–19, 2020. 13

ASLAM, Saba; ASLAM, Hafsa; MANZOOR, Arslan; CHEN, Hui; RASOOL, Abdur. Antiphishstack: Lstm-based stacked generalization model for optimized phishing url detection. **Symmetry**, MDPI, v. 16, n. 2, p. 248, 2024. 35, 39

AUNG, Eint Sandi; ZAN, Chaw Thet; YAMANA, Hayato. A survey of url-based phishing detection. *In*: **DEIM forum**. [S.l.: s.n.], 2019. p. G2–3. 25, 46

BAHAGHIGHAT, Mahdi; GHASEMI, Majid; OZEN, Figen. A high-accuracy phishing website detection method based on machine learning. **Journal of Information Security and Applications**, Elsevier, v. 77, p. 103553, 2023. 13

BASIT, Abdul; ZAFAR, Maham; LIU, Xuan; JAVED, Abdul Rehman; JALIL, Zunera; KIFAYAT, Kashif. A comprehensive survey of ai-enabled phishing attacks detection techniques. **Telecommunication Systems**, Springer, v. 76, p. 139–154, 2021. 7, 23, 24, 25, 52

Canal Tech. **Novas modalidades de phishing que você precisa conhecer**. 2016. Disponível em: https://canaltech.com.br/seguranca/crescimento-de-novas-modalidades-de-golpe-preocupam-65365/. Acesso em: 27 jan. 2025. Disponível em: https://canaltech.com.br/seguranca/crescimento-de-novas-modalidades-de-golpe-preocupam-65365/. 20

CNN Brasil. **Saiba o que é Deepfake: técnica de inteligência artificial que foi apropriada para produzir desinformação**. 2023. Acesso em: 13 ago. 2024. Disponível em: https://www.cnnbrasil.com.br/noticias/saiba-o-que-e-deepfake-tecnica-de-inteligencia-artificial-que-foi-apropriada-para-produzir-desinformacao/. 21, 47

COELHO, Cristiano Farias; RASMA, Eline Tourinho; MORALES, Gudelia. Engenharia social: uma ameaça à sociedade da informação. **Exatas & Engenharias**, v. 3, n. 05, 2013. 17

COOK, Debra L; GURBANI, Vijay K; DANILUK, Michael. Phishwish: A stateless phishing filter using minimal rules. *In*: SPRINGER. **Financial Cryptography and Data Security: 12th International Conference, FC 2008, Cozumel, Mexico, January 28-31, 2008. Revised Selected Papers 12**. [S.l.], 2008. p. 182–186. 34, 35, 39

Correio Braziliense. **Crimes cibernéticos avançam no Brasil e aceleram com a tecnologia**. 2024. Acesso em: 10 set. 2024. Disponível em: https://www.correiobraziliense.com.br/economia/2024/03/6824212-crimes-ciberneticos-avancam-no-brasil-e-aceleram-com-a-tecnologia.html. 12

DEVLIN, Jacob; CHANG, Ming-Wei; LEE, Kenton; TOUTANOVA, Kristina. BERT: Pre-training of deep bidirectional transformers for language understanding. *In*: BURSTEIN, Jill; DORAN, Christy; SOLORIO, Thamar (Ed.). **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. p. 4171–4186. Disponível em: https://aclanthology.org/N19-1423/. 24

DOOREMAAL, Bram Van; BURDA, Pavlo; ALLODI, Luca; ZANNONE, Nicola. Combining text and visual features to improve the identification of cloned webpages for early phishing detection. *In*: **Proceedings of the 16th International Conference on Availability, Reliability and Security**. [S.l.: s.n.], 2021. p. 1–10. 43

DUNLOP, Matthew; GROAT, Stephen; SHELLY, David. Goldphish: Using images for content-based phishing analysis. *In*: IEEE. **2010 Fifth international conference on internet monitoring and protection**. [S.l.], 2010. p. 123–128. 42, 44

El Pescador. **Conscientização em Segurança da Informação: Os desafios de despertar a participação ativa na segurança de uma corporação**. 2016. Disponível em: `https://www.elpescador.com.br/blog/index.php/conscientizacao-em-seguranca-da-informacao-os-desafios-de-despertar-a-participacao-ativa-na-seguranca-de-uma-corporacao/`. Acesso em: 27 jan. 2025. Disponível em: `https://www.elpescador.com.br/blog/index.php/conscientizacao-em-seguranca-da-informacao-os-desafios-de-despertar-a-participacao-ativa-na-seguranca-de-uma-corporacao/`. 18, 19

FANG, Yong; ZHANG, Cheng; HUANG, Cheng; LIU, Liang; YANG, Yue. Phishing email detection using improved rcnn model with multilevel vectors and attention mechanism. **IEEE Access**, IEEE, v. 7, p. 56329–56340, 2019. 34, 35

_____. _____. **IEEE Access**, IEEE, v. 7, p. 56329–56340, 2019. 40

FENG, Jian; ZOU, Lianyang; YE, Ou; HAN, Jingzhou. Web2vec: Phishing webpage detection method based on multidimensional features driven by deep learning. **IEEE Access**, IEEE, v. 8, p. 221214–221224, 2020. 34, 42, 45

GUPTA, Brij B; GAURAV, Akshat; ARYA, Varsha; ATTAR, Razaz Waheeb; BANSAL, Shavi; ALHOMOUD, Ahmed; CHUI, Kwok Tai. Advanced bert and cnn-based computational model for phishing detection in enterprise systems. **CMES-Computer Modeling in Engineering & Sciences**, v. 141, n. 3, 2024. 40

HIJJI, Mohammad; ALAM, Gulzar. A multivocal literature review on growing social engineering based cyber-attacks/threats during the covid-19 pandemic: Challenges and prospective solutions. **IEEE Access**, v. 9, p. 7152–7169, 2021. 12, 16, 46

Kaspersky. **Brasileiros são os maiores alvos de ataques de phishing no mundo**. 2021. Acesso em: 10 set. 2024. Disponível em: `https:`

//www.kaspersky.com.br/about/press-releases/brasileiros-sao-os-maiores-alvos-de-ataques-de-phishing-no-mundo. 13

_____. **Brasil é o país com mais ataques de phishing por WhatsApp no mundo em 2022, aponta Kaspersky**. 2022. Acesso em: 10 set. 2024. Disponível em: https://www.kaspersky.com.br/about/press-releases/brasil-e-o-pais-com-mais-ataques-de-phishing-por-whatsapp-no-mundo-em-2022-aponta-kaspersky. 13

_____. **Nova epidemia: phishing cresce mais de 5 vezes no Brasil com retomada das atividades econômicas e apoio da IA**. 2024. Acessado em agosto de 2024. Disponível em: https://www.kaspersky.com.br/about/press-releases/2024_nova-epidemia-phishing-cresce-mais-de-5-vezes-no-brasil-com-retomada-das-atividades-economicas-e-apoio-da-ia. 21

KHONJI, Mahmoud; IRAQI, Youssef; JONES, Andrew. Phishing detection: a literature survey. **IEEE Communications Surveys & Tutorials**, IEEE, v. 15, n. 4, p. 2091–2121, 2013. 7, 19, 20, 22, 23

LE, Hung; PHAM, Quang; SAHOO, Doyen; HOI, Steven CH. Urlnet: Learning a url representation with deep learning for malicious url detection. **arXiv preprint arXiv:1802.03162**, 2018. 34, 35, 41

MACHARYAS, Jeff. **Ransomware: Cryptolocker**. 2014. Disponível em: https://medium.com/@macharyas/ransomware-cryptolocker-c4f2c1a58d77. Acesso em: 27 jan. 2025. Disponível em: https://medium.com/@macharyas/ransomware-cryptolocker-c4f2c1a58d77. 19

MAGAZINE, Security. **Information Security Forum Forecasts 2017 Global Security Threat Outlook**. 2017. Accessed: 2024-09-04. Disponível em: https://www.securitymagazine.com/articles/87639-information-security-forum-forecasts-2017-global-security-threat-outlook. 19

MAO, Jian; TIAN, Wenqian; LI, Pei; WEI, Tao; LIANG, Zhenkai. Phishing-alarm: Robust and efficient phishing detection via page component similarity. **IEEE Access**, IEEE, v. 5, p. 17020–17030, 2017. 42

MARCHAL, Samuel; FRANÇOIS, Jérôme; STATE, Radu; ENGEL, Thomas. Phishstorm: Detecting phishing with streaming analytics. **IEEE Transactions on Network and Service Management**, IEEE, v. 11, n. 4, p. 458–471, 2014. 35, 44, 47

Michaelis. **Dicionário Brasileiro da Língua Portuguesa**. Melhoramentos, 2018. Disponível em: `https://michaelis.uol.com.br/busca?id=ok3N7`. Acesso em: 27 jan. 2025. Disponível em: `https://michaelis.uol.com.br/busca?id=ok3N7`. 18

NEIVA, Fran; SILVA, Rodrigo. **Revisão Sistemática da Literatura em Ciência da Computação - Um Guia Prático**. 06 2016. 28

OPARA, Chidimma; WEI, Bo; CHEN, Yingke. Htmlphish: Enabling phishing web page detection by applying deep learning techniques on html analysis. *In*: IEEE. **2020 International Joint Conference on Neural Networks (IJCNN)**. [S.l.], 2020. p. 1–8. 34, 39, 45, 47

OVI, Md Sultanul Islam; RAHMAN, Md Hasibur; HOSSAIN, Mohammad Arif. Phishguard: A multi-layered ensemble model for optimal phishing website detection. **arXiv preprint arXiv:2409.19825**, 2024. 38, 49

PAREKH, Shraddha; PARIKH, Dhwanil; KOTAK, Srushti; SANKHE, Smita. A new method for detection of phishing websites: Url detection. *In*: IEEE. **2018 Second international conference on inventive communication and computational technologies (ICICCT)**. [S.l.], 2018. p. 949–952. 35, 40, 45

PARKER, Heather J; FLOWERDAY, Stephen V. Contributing factors to increased susceptibility to social media phishing attacks. **South African Journal of Information Management**, AOSIS Publishing, v. 22, n. 1, p. 1–10, 2020. 13

PENG, Tianrui; HARRIS, Ian; SAWA, Yuki. Detecting phishing attacks using natural language processing and machine learning. *In*: IEEE. **2018 ieee 12th international conference on semantic computing (icsc)**. [S.l.], 2018. p. 300–301. 35, 42

PETTICREW, Mark; ROBERTS, Helen. **Systematic reviews in the social sciences: A practical guide**. [S.l.]: John Wiley & Sons, 2008. 16, 27

PRAKASH, Pawan; KUMAR, Manish; KOMPELLA, Ramana Rao; GUPTA, Minaxi. Phishnet: predictive blacklisting to detect phishing attacks. *In*: IEEE. **2010 Proceedings IEEE INFOCOM**. [S.l.], 2010. p. 1–5. 35, 43

SUBASI, Abdulhamit; MOLAH, Esraa; ALMKALLAWI, Fatin; CHAUDHERY, Touseef J. Intelligent phishing website detection using random forest classifier. *In*: IEEE. **2017 International conference on electrical and computing technologies and applications (ICECTA)**. [S.l.], 2017. p. 1–5. 41, 43, 45

VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N; KAISER, Ł ukasz; POLOSUKHIN, Illia. Attention is all you need. *In*: GUYON, I.; LUXBURG, U. Von; BENGIO, S.; WALLACH, H.; FERGUS, R.; VISHWANATHAN, S.; GARNETT, R. (Ed.). **Advances in Neural Information Processing Systems**. Curran Associates, Inc., 2017. v. 30. Disponível em: `https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf`. 25

WHITTAKER, C.; RYNER, B.; NAZIF, M. Large-scale automatic classification of phishing pages. **Proceedings of the 2010 IEEE European Symposium on Security and Privacy**, p. 1–15, 2010. 18

YUE, Chuan; WANG, Haining. Bogusbiter: A transparent protection against phishing attacks. **ACM Transactions on Internet Technology (TOIT)**, ACM New York, NY, USA, v. 10, n. 2, p. 1–31, 2010. 34, 38, 45

ZHU, Erzhou; CHEN, Yuyang; YE, Chengcheng; LI, Xuejun; LIU, Feng. Ofs-nn: an effective phishing websites detection model based on optimal feature selection and neural network. **Ieee Access**, IEEE, v. 7, p. 73271–73284, 2019. 43