INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ
CAMPUS PEDRO II
TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

NÍCOLAS TEIXEIRA BARROS

INCIDENTES DE SEGURANÇA CONTRA USUÁRIOS EM REDES SOCIAIS: UM MAPEAMENTO SISTEMÁTICO

PEDRO II, PIAUÍ
2025

NÍCOLAS TEIXEIRA BARROS

INCIDENTES DE SEGURANÇA CONTRA USUÁRIOS EM REDES SOCIAIS: UM MAPEAMENTO SISTEMÁTICO

Trabalho de Conclusão de Curso (artigo científico) apresentado como exigência parcial para obtenção do diploma do Curso de Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Campus Pedro II.

Orientador: Prof. Esp. Danniel Rocha do Nascimento.
Coorientador: Prof. Me. Marcos Soares de Oliveira.

PEDRO II, PIAUÍ

2025

INCIDENTES DE SEGURANÇA CONTRA USUÁRIOS EM REDES SOCIAIS: UM MAPEAMENTO SISTEMÁTICO

Nícolas Teixeira Barros[1]

Prof. Esp. Danniel Rocha do Nascimento[2]

Coorientador: Prof. Me. Marcos Soares de Oliveira[3]

RESUMO

O uso crescente das redes sociais transformou a maneira como as pessoas se comunicam, compartilham informações e interagem no ambiente digital. Contudo, essa popularização também resultou em um aumento significativo de incidentes de segurança que afetam diretamente os usuários, como ataques de phishing, disseminação de malware, engenharia social, roubo de identidade e exposição de dados sensíveis. Este trabalho tem como objetivo realizar um mapeamento sistemático da literatura científica publicada entre 2018 e 2024, a fim de identificar e categorizar os principais tipos de incidentes de segurança enfrentados por usuários de redes sociais, as vulnerabilidades mais exploradas e as estratégias propostas para prevenção e mitigação desses riscos. A metodologia baseou-se na seleção e análise de artigos científicos extraídos de bases de dados relevantes, seguindo critérios específicos de inclusão e exclusão. Os resultados indicam que os temas mais recorrentes envolvem bots sociais e comportamento malicioso (33%), ataques de phishing (28%), engenharia social (18%), malware (15%), privacidade e vazamento de dados (14%) e a aplicação de técnicas de aprendizado de máquina e aprendizado profundo para mitigação. Conclui-se que a segurança em redes sociais demanda ações integradas entre tecnologia, educação digital e políticas de proteção de dados, a fim de reduzir os impactos dos incidentes e fortalecer a confiança dos usuários nessas plataformas.

Palavras-chave: redes sociais, segurança da informação, incidentes de segurança, phishing, malware, engenharia social.

ABSTRACT

The growing use of social media has changed the way people communicate, share information, and interact in the digital environment. However, this popularization has also resulted in a significant increase in security incidents that directly affect users,

---

[1] Graduando em Análise e desenvolvimento de sistemas. IFPI. nicolas974352@gmail.com

[2] Licenciado em Computação pela UESPI e Especialista em Engenharia de Sistemas pela ESAB. danniel.rocha@ifpi.edu.br

[3] Bacharel em Ciência da Computação pela UESPI e Mestre em Ciências pela USP. marcos.soares@ifpi.edu.br

such as phishing attacks, malware dissemination, social engineering, identity theft, and sensitive data exposure. This work aims to systematically map the scientific literature published between 2018 and 2024 to identify and categorize the main types of security incidents faced by social media users, the most commonly exploited vulnerabilities, and the strategies proposed to prevent and mitigate these risks. The methodology is based on the selection and analysis of scientific articles extracted from relevant databases, following specific inclusion and exclusion criteria. The results indicate that the most recurring themes involve social bots and malicious behavior (33%), phishing attacks (28%), social engineering (18%), malware (15%), and the privacy and data leaks (14%) and the application of machine learning and deep learning techniques for mitigation. It is concluded that security on social networks requires integrated actions between technology, digital education and data protection policies, in order to reduce the impacts of incidents and strengthen users' trust in these platforms. keywords: social networks, information security, security incidents, phishing, malware, social engineering.

Data de aprovação: (05/08/2025)

1 INTRODUÇÃO

A crescente popularização das redes sociais, que receberam enorme atenção de usuários em todo o mundo na última década, mudando a forma como as pessoas interagem e compartilham informações, resultando em um aumento inesperado no número de usuários registrados em sites populares como Facebook, Twitter e outros (Kumar *et al.*, 2023).

À medida que os usuários compartilham dados pessoais, também se intensificam os riscos relacionados à segurança digital, uma vez que a web, com seu potencial ilimitado, também revela um lado sombrio ao abrir espaço para práticas criminosas que desafiam e transcendem as fronteiras físicas (Arshey; Viji, 2021). A exposição de dados pessoais dos usuários, modificação maliciosa de conteúdos, propagação de links perigosos e até mesmo a situações de chantagem, tornaram os usuários alvos vulneráveis para cibercriminosos (Shevchuk *et al.*, 2020). O fator humano também é bastante preocupante. Segundo (Jamil *et al.*, 2018) o fator humano é crucial nas falhas de segurança cibernética, e embora o campo cibernético tenha uma natureza técnica, as pessoas continuam a ser o elemento crucial, já que os erros cometidos por seres humanos são responsáveis por 95% dos incidentes.

Nesse cenário, surgem também novas ameaças à segurança dos usuários, incluindo o uso crescente de bots sociais para automatizar interações e disseminar conteúdo malicioso (Shi; Zhang; Choo, 2019). Além de ataques diretos como phishing, engenharia social e disseminação de malwares, práticas como o uso de bots sociais e vazamento de dados pessoais vêm ganhando espaço no cenário digital (Schnebly; Sengupta, 2019; Shukla; Jagtap; Patil, 2021). Tais ameaças comprometem não apenas os dados dos usuários mas também sua privacidade e confiança nas plataformas (Hu *et al.*, 2023). Embora existam diversos estudos sobre temas de forma isolada, percebe-se uma lacuna na literatura quanto à sistematização e integração dessas ameaças e das respectivas medidas de mitigação, o que reforça a necessidade de abordagem integrada e o desenvolvimento de técnicas baseadas em aprendizado profundo para detecção e prevenção (Arin; Kutlu, 2023; Shukla; Jagtap; Patil, 2021).

O crescimento acelerado das redes sociais facilitou a comunicação entre bilhões de pessoas, mas também abriu espaço para novos desafios de segurança, tornando essas plataformas alvos frequentes de ações maliciosas (Thakur; Hayajneh; Tseng, 2019). O *phishing* é uma forma de ataque cibernético que se aproveita da confiança dos usuários e continua sendo um dos meios mais utilizados para obter informações sensíveis, como credenciais de acesso e dados pessoais (Shah *et al.*, 2022). A engenharia social também se destaca por enganar ou manipular e explorar falhas no comportamento humano, induzindo a vitima a fornecer suas informações pessoais e confidenciais pela falta de conhecimento em segurança digital (Salama *et al.*, 2023). Outro vetor relevante de ataque é a disseminação de *malware* por meio de links maliciosos, utilizados por invasores para extrair informações de dispositivos como computadores e celulares, geralmente após induzirem a vítima a baixar e instalar softwares maliciosos (Parthy; Rajendran, 2019). Assim, este trabalho realiza um mapeamento sistemático da literatura entre 2018 e 2024, com o objetivo de identificar os principais tipos de incidentes de segurança enfrentados por usuários de redes sociais, as vulnerabilidades exploradas e as estratégias propostas para prevenção e mitigação desses riscos.

1.1 **Objetivos Gerais**

Este trabalho tem como objetivo geral realizar um mapeamento sistemático da literatura sobre incidentes de segurança em redes sociais, identificando os principais tipos de ataques, vulnerabilidades exploradas e as medidas de prevenção e mitigação que ocorrem contra o usuário.

1.2 **Objetivos Específicos**

1. Identificar e categorizar os diferentes tipos de incidentes de segurança que afetam usuários de redes sociais, incluindo phishing, malware, engenharia social e roubo de identidade.

2. Mapear as medidas de prevenção e mitigação propostas na literatura para proteger usuários e plataformas contra incidentes de segurança.

2 **FUNDAMENTAÇÃO TEÓRICA**

Esta seção apresenta os principais conceitos relacionados a incidentes de segurança em redes sociais, abordando os tipos de ataques, as vulnerabilidades exploradas, as estratégias de prevenção e mitigação, e o impacto social desses eventos, fornecendo a base teórica necessária para o entendimento do problema investigado.

2.1 **Segurança da Informação**

A segurança da informação é definida como o conjunto de práticas voltadas à proteção de sistemas, redes e dados contra ataques digitais, visando evitar o acesso, modificação ou destruição indevida de informações sensíveis (Sun *et al.*, 2019). Seu propósito fundamental é assegurar a confidencialidade, a disponibilidade e a integridade dos sistemas de informação baseados na Internet e das informações e dados dos usuários já que são regularmente comprometidas por um ataque cibernético bem-sucedido (Thakur; Hayajneh; Tseng, 2019).

O crescimento do uso das redes sociais trouxe preocupações significativas quanto à segurança e privacidade dos usuários, devido à disseminação de conteúdos maliciosos como phishing , Engenharia Social, malware e roubo de identidade comprometendo assim tanto os dados dos usuários quanto a confiança nas plataformas(Islam; Latif; Ahmed, 2019; Verma *et al.*, 2023). Portanto o uso de medidas

técnicas como criptografia e autenticação e conscientização é fundamental para proteger os dados e a privacidade do usuário de ataques e evitar ameaças nas redes sociais(Sharma *et al.*, 2018).

2.2 **Detecção de Bots Sociais e Comportamento Malicioso**

Bots sociais são programas automatizados que simulam comportamento humanos nas redes sociais e são usados para diversas finalidades, desde marketing até manipulação da opinião pública e disseminação de desinformação (Shi; Zhang; Choo, 2019; Shukla; Jagtap; Patil, 2021). Tais *bots* podem influenciar debates, espalhar conteúdos enganosos e engajar usuários em fraudes,muitas vezes sem que percebam (Shi; Zhang; Choo, 2019). A identificação desses agentes maliciosos baseiase frequentemente em padrões específicos de atividade, como postagens em horários fixos, seguimento acelerado de contas e replicação de conteúdos de outros usuários, porém a sofisticação crescente desses *bots* torna sua detecção manual cada vez mais difícil (Shukla; Jagtap; Patil, 2021).

Para mitigar essa ameaça, são empregadas técnicas analíticas avançadas focadas no comportamento dos usuários, dados dos perfis e conexões entre contas (Arin; Kutlu, 2023). O uso de algoritmos de aprendizado de máquina que atualizam seus modelos dinamicamente tem se mostrado promissor para preservar a integridade da informação e reduzir o impacto dos *bots* maliciosos (Fazil; Sah; Abulaish, 2021).

Além de padrões comportamentais, abordagens modernas analisam sequências de cliques, tempo de resposta e padrões de interação social para distinguir contas legítimas de contas automatizadas (Beskow; Carley, 2018). Ferramentas baseadas em aprendizado profundo e inteligência artificial possibilitam modelos preditivos mais precisos, facilitando a implementação de filtros eficientes para bloquear ou sinalizar atividades suspeitas antes que causem maiores danos (Schnebly; Sengupta, 2019). Essas soluções também incluem identificação de *spammers* e perfis falsos em redes sociais (Verma *et al.*, 2023).

Recentemente, métodos neuro-fuzzy que são uma técnica híbrida que combina duas áreas da inteligência artificial que são os sistemas fuzzy usados para lidar com

incerteza e imprecisão e redes neurais que são usadas para aprender padrões a partir de dados e ajustar parâmetros automaticamente, têm sido aplicados para identificar anomalias, além da combinação de classificadores para melhorar a precisão na identificação de perfis falsos e bots, ampliando a eficácia dos sistemas de detecção (Sharma *et al.*, 2018). Essas soluções utilizam múltiplos indicadores e modelos híbridos para enfrentar a complexidade crescente da atuação dos *bots* nas redes sociais (Sowmya; Chatterjee, 2020; Latha, P *et al.*, 2022). Também são aplicados métodos de revisão e classificação aprimorados para a detecção eficiente de spam e perfis falsos (Yurtseven; Bagriyanik; Ayvaz, 2021; Wang *et al.*, 2019).

Portanto, o desenvolvimento contínuo de ferramentas automatizadas e inteligentes é essencial para manter os ambientes digitais mais seguros e confiáveis, minimizando o impacto das atividades maliciosas de bots sociais (Harris *et al.*, 2021). Técnicas de deep learning para a detecção automática de usuários não confiáveis nas redes sociais têm sido promissoras nesse sentido (Sansonetti *et al.*, 2020). Além disso, *frameworks* específicos para detectar imagens falsas em tweets ampliam o escopo da segurança digital (Parikh; Khedia; Atrey, 2019).

2.3 **Ataque de Phishing e Técnicas e Técnicas de Detecção**

O *phishing* é uma das formas mais comuns e perigosas de ataque nas redes sociais, visando enganar os usuários para fornecerem informações pessoais por meio de mensagens falsas ou links maliciosos (Tang; Mahmoud, 2022; Abdillah *et al.*, 2022).

As redes sociais representam um ambiente propício para tais ataques devido à confiança natural entre contatos próximos e à enorme rapidez e volume de compartilhamento de informações, que dificultam a verificação da autenticidade das mensagens (Oest *et al.*, 2018; Nicholson *et al.*, 2020).

Para mitigar essas ameaças, várias abordagens são empregadas, incluindo o uso de filtros automáticos e sistemas de detecção baseados em aprendizado de máquina, que avaliam

URLs e conteúdos para identificar e bloquear links suspeitos (Ripa; Islam; Arifuzzaman, 2021; Korkmaz; Sahingoz; Diri, 2020). Abordagens complementares, como métodos baseados em decisão multicritério e análise de conteúdo com *machine*

learning, mostram-se eficazes na identificação precoce de sites fraudulentos (Shah *et al.*, 2022; Ozker; Sahingoz, 2020).

Além das soluções técnicas, a conscientização dos usuários é considerada essencial para a prevenção eficaz, pois o sucesso do *phishing* depende significativamente do comportamento humano (Adil; Khan; Ghani, 2020; Arshey; Viji, 2021). Educar os usuários a reconhecerem mensagens fraudulentas, desconfiar de solicitações incomuns e manter softwares atualizados são atitudes fundamentais para evitar prejuízos decorrentes dessas ameaças (Abdillah *et al.*, 2022).

2.4 **Engenharia Social e Falhas Humanas**

A engenharia social é uma técnica que manipula comportamentos humanos para induzir as vítimas a realizar ações que comprometem sua segurança, explorando vulnerabilidades emocionais e cognitivas ao invés de falhas técnicas (Jamil *et al.*, 2018; Hijji; Alam, 2021). Em redes sociais, ataques de engenharia social frequentemente utilizam perfis falsos que simulam amigos, familiares ou figuras de autoridade para influenciar usuários a clicarem em links maliciosos ou fornecerem dados sensíveis além de se aproveitarem de mensagens com temas emocionais para gerar pânico ou empatia, induzindo ações imediatas (Salama *et al.*, 2023; Syafitri *et al.*, 2022).

Para mitigar esses riscos, é fundamental investir em educação digital que capacite os usuários a identificar sinais de manipulação social, promovendo uma postura crítica nas interações online (Parthy; Rajendran, 2019). Estratégias como simulações de ataques, campanhas educativas e a inclusão da cultura de cibersegurança em escolas e ambientes corporativos são importantes para reduzir a eficácia das técnicas de engenharia social (Aldawood; Skinner, 2018).

Além disso, pesquisas recentes demonstram o potencial do uso de aprendizado profundo, como redes neurais que são modelos computacionais inspirados no funcionamento do cérebro humano e convolucionais embeddings de palavras que são vetores que representam palavras numericamente, aproximando termos com significados semelhantes para que algoritmos entendam seu contexto, para o reconhecimento precoce de táticas persuasivas em ataques de engenharia social baseados em chat, contribuindo para a detecção automatizada e prevenção desses ataques (Tsinganos; Mavridis; Gritzalis, 2022). Importante mencionar que cenários

recentes, como a pandemia COVID-19, intensificaram a exploração do medo público através de ataques de *phishing* e engenharia social, reforçando a necessidade de resposta coordenada e contínua contra essas ameaças (Bitaab *et al.*, 2020).

2.5 **Malware e Ameaças Cibernéticas**

Malware são programas maliciosos desenvolvidos para causar danos, roubar informações ou obter acesso não autorizado a sistemas. Em redes sociais, o *malware* geralmente é disseminado por meio de links camuflados, arquivos anexados ou aplicativos de terceiros, uma vez ativados, esses programas podem comprometer seriamente a privacidade e a segurança dos usuários (Thakur; Hayajneh; Tseng, 2019; Chaudhari *et al.*, 2024).

Os tipos de *malware* mais comuns incluem vírus, worms, trojans e spyware, eles podem ser usados para capturar senhas, registrar tudo que é digitado no teclado, ou até mesmo assumir o controle completo de dispositivos. O compartilhamento em massa de conteúdos e a presença de links externos nas redes sociais tornam os usuários alvos fáceis para esse tipo de ameaça, especialmente quando não utilizam medidas de proteção (Cohen; Nissim; Elovici, 2020; Oest *et al.*, 2018).

A prevenção contra *malwares* exige uma combinação de ferramentas tecnológicas, como antivírus e *firewalls*, e práticas conscientes por parte dos usuários. É importante evitar clicar em links desconhecidos, desconfiar de mensagens fora do padrão e manter os dispositivos atualizados. Além disso, plataformas de redes sociais devem reforçar seus mecanismos de detecção automática de ameaças e oferecer respostas rápidas a incidentes reportados pelos usuários (Sun *et al.*, 2019; Cengiz; Kalem; Boluk, 2022; Shivangi *et al.*, 2018).

2.6 **Privacidade e Vazamento de Dados**

Os aspectos relacionados à privacidade digital e vazamento de dados em redes sociais são amplamente discutidos em várias pesquisas recentes. A privacidade digital refere-se ao controle que os usuários possuem sobre as informações que compartilham online, algo que está em constante ameaça devido a vazamentos intencionais, falhas de segurança e políticas pouco transparentes das plataformas (Li *et al.*, 2020). Além disso, dados expostos podem ser explorados por anunciantes,

empresas terceirizadas e cibercriminosos, acarretando consequências graves como roubo de identidade, fraudes financeiras e danos à reputação (Hu *et al.*, 2023).

Os usuários muitas vezes não têm plena consciência da quantidade de informações que disponibilizam ou da forma como esses dados são armazenados e utilizados plenas redes sociais, o que agrava os riscos à privacidade (Lankton; Mcknight; Tripp, 2020). Ademais, os termos de uso dessas plataformas geralmente não são claros ou de fácil compreensão para o público geral, dificultando uma gestão efetiva dos dados pessoais (Shevchuk *et al.*, 2020).

Para mitigar esses riscos, é fundamental que as plataformas adotem políticas de privacidade mais transparentes e disponibilizem controles granulares que facilitem o gerenciamento dos dados pelos usuários, como perfis privados, revisões regulares de permissões de acesso e alertas em casos de atividades suspeitas (He; Cai; Yu, 2018). Além disso, a responsabilidade pela proteção da privacidade deve ser compartilhada entre usuários, desenvolvedores e reguladores, garantindo um ambiente digital mais seguro e respeitoso da privacidade individual (Pietro; Cresci, 2021).

Estas abordagens colaborativas e tecnológicas são cruciais para preservar a privacidade à medida em que as redes sociais evoluem e ampliam o escopo de suas funcionalidades e dados coletados (Li *et al.*, 2020; Shevchuk *et al.*, 2020).

2.7 **Machine Learning e Deep Learning na Segurança Digital**

Machine learning e *Deep learning* são tecnologias importantes na segurança digital, especialmente para identificar ameaças em redes sociais, sendo que técnicas como métodos ensemble ajudam a detectar bots com mais precisão ao analisar padrões de comportamento automatizados (Shukla; Jagtap; Patil, 2021; Arin; Kutlu, 2023). Já o *deep learning* utiliza modelos como *long short-term memory* (LSTM) que é um modelo de rede neural capaz de aprender e lembrar informações por longos períodos para capturar sequências temporais e contextuais nas atividades sociais, melhorando a eficácia na detecção de bots e contas falsas (Arin; Kutlu, 2023). Abordagens multimodais que integram textos, imagens e metadados têm sido aplicadas para tornar mais robusta a identificação de perfis fraudulentos, como demonstrado no uso de redes neurais profundas para analisar múltiplas fontes de dados simultaneamente (Goyal *et al.*, 2023).

Essas técnicas avançadas também têm sido adaptadas para contextos específicos, como a detecção de conteúdos maliciosos em línguas menos representadas, exemplificada pelo uso de *machine learning* para identificar textos perigosos em bengali em redes sociais (Shi; Zhang; Choo, 2019). A segurança digital abrange a análise de imagens, sendo possível identificar arquivos maliciosos, como imagens JPEG infectadas por meio de modelos treinados para diferenciar imagens legítimas de maliciosas (Cohen; Nissim; Elovici, 2020).

3 TRABALHOS RELACIONADOS

A literatura sobre incidentes de segurança em redes sociais tem se expandido bastante nos últimos anos, com foco em ameaças como perfis falsos, ataques baseados em engenharia social e métodos preditivos de cibersegurança. Três estudos se destacam por oferecerem abordagens complementares ao presente trabalho.

O estudo de (Kumar *et al.*, 2023), apresenta uma revisão ampla sobre métodos para detecção de perfis falsos em redes sociais. Os autores classificam os estudos existentes em três categorias principais: abordagens baseadas em atributos de conta, em características textuais e em uma combinação de ambos. São discutidos os principais desafios da detecção de perfis falsos, como a sofisticação dos bots e o uso de técnicas de deep learning para enganar sistemas de verificação. O estudo também detalha os conjuntos de dados e métricas utilizados na área, fornecendo uma base relevante para análises sobre segurança e confiabilidade nas interações online, aspectos diretamente ligados à proteção do usuário em ambientes digitais. Enquanto (Kumar *et al.*, 2023) concentram-se exclusivamente na detecção de perfis falsos, o presente trabalho amplia esse escopo ao abordar também outras formas de incidentes de segurança, como *phishing*, *malware*, vazamentos de dados e engenharia social, por meio de um mapeamento sistemático mais abrangente.

Em um contexto mais amplo de cibersegurança, o estudo de (Sun *et al.*, 2019), apresenta um levantamento abrangente de métodos preditivos para iniciantes de segurança baseados em dados. O artigo destaca a evolução da área de uma abordagem reativa para uma abordagem proativa, enfatizando o uso de aprendizado de máquina e mineração de dados para antecipar ameaças. Entre os principais desafios apontados estão a coleta e rotulagem de dados, a definição de atributos

relevantes e a construção de modelos personalizados de previsão. Embora o estudo aborde diversos domínios, ele também ressalta o uso de dados de redes sociais como uma das fontes para prever comportamentos maliciosos reforçando a relevância de modelos preditivos no combate a incidentes em plataformas digitais. O presente trabalho, além de reconhecer o valor dos modelos preditivos, foca especificamente em incidentes contra usuários em redes sociais, organizando as ameaças em categorias e mapeando também medidas de mitigação técnicas e educativas — algo que não é abordado com profundidade por (Sun *et al.*, 2019).

Por fim, durante a pandemia de COVID-19, o estudo de (Hijji; Alam, 2021), investigou o aumento dos ataques baseados em engenharia social. A revisão multivocal combinou literatura acadêmica com fontes da indústria, revelando que a pandemia intensificou o uso de técnicas manipulativas, como *phishing* emocional e exploração de medo e desinformação. O artigo propõe contramedidas como educação em cibersegurança, autenticação multifatorial e uso de sistemas inteligentes para reconhecimento de padrões persuasivos. Essa análise contribui significativamente para o entendimento do fatos humano como vetor de vulnerabilidade, especialmente em redes sociais, onde os usuários são frequentemente alvos de manipulação psicológica. Ao contrário do estudo de (Hijji; Alam, 2021), que foca exclusivamente em ataques de engenharia social no contexto da pandemia, este trabalho adota uma abordagem mais ampla e longitudinal (2018–2024), categorizando múltiplos tipos de incidentes e analisando suas contramedidas, além de considerar o impacto social e técnico dessas ameaças.

Esses três estudos fornecem bases teóricas e metodológicas valiosas para este trabalho, ao abordarem diferentes facetas dos incidentes de segurança digital em redes sociais — desde perfis falsos e comportamento automatizado, até ataques direcionados à vulnerabilidade humana e estratégias proativas de prevenção.

4 **METODOLOGIA**

Esta seção descreve a metodologia adotada para a pesquisa, delineando os procedimentos e

abordagens utilizados para atingir os objetivos propostos.

4.1 **Tipo de Pesquisa**

A pesquisa adotada neste estudo foi de natureza exploratória e descritiva, com enfoque em um mapeamento sistemático da literatura. Em vez de realizar uma revisão crítica detalhada, o estudo busca identificar, categorizar e mapear as principais tendências e contribuições da literatura relacionada aos incidentes de segurança em redes sociais (Falbo, 2018).

4.2 **População e Amostra**

A população do estudo é composta por artigos científicos e acadêmicos e relatórios técnicos sobre incidentes de segurança contra usuários em redes sociais, com foco em aspectos como tipos de ataques como *phishing*, *malware* e engenharia social, vulnerabilidades exploradas pelos cibercriminosos como fragilidades comportamentais dos usuários, medidas de prevenção e mitigação tanto soluções técnicas, como autenticação de dois fatores e criptografia, quanto estratégias de conscientização e educação digital, e o impacto social desses incidentes, analisando a confiança dos usuários e suas percepções sobre privacidade e segurança nas redes sociais. A amostra será composta por estudos publicados entre 2018 e 2024, selecionados de fontes confiáveis, como periódicos acadêmicos, relatórios de empresas de segurança e outras publicações relevantes.

4.3 **Critérios de Seleção**

Os estudos foram selecionados com base em critérios de inclusão e exclusão para garantir a relevância e qualidade. Os critérios de inclusão exigem busca em plataformas utilizando as palavraschave ’incidentes contra usuários’, ’phishing em redes sociais’, ’malwares em redes sociais’, ’engenharia social em redes sociais’, publicados entre 2018 e 2024, leitura do título e resumo provenientes de fontes confiáveis, como artigos científicos e acadêmicos ou relatórios de segurança, leitura de resultados e conclusões e que estão disponíveis em português ou inglês. Foram excluídos estudos irrelevantes que não fale sobre redes sociais, sem metodologia clara, fora do escopo, como segurança digital genérica, e fontes não verificáveis, como blogs. Também serão descartados estudos em idiomas diferentes de português e inglês.

4.4 **Fontes de Dados e Palavras-Chave**

Para a seleção dos estudos, foi utilizada a seguinte base de dados: IEEE Xplore pois a mesma publica periódicos, conferências e padrões de alta reputação, sujeitos a um rigoroso processo de revisão por pares. Isso garante a qualidade e a credibilidade das publicações. Ao focar na IEEE Xplore, mitigamos o risco de incluir estudos de menor impacto ou relevância técnica, o que é crucial para um mapeamento sistemático que busca consolidar o conhecimento mais robusto e validado na área. As palavras-chave empregadas na busca incluíram incidentes de segurança contra usuários, *phishing* em redes sociais, *malware* em redes sociais e engenharia social em redes sociais. Essas palavras-chave ajudaram a identificar publicações relevantes para a revisão.

A primeira etapa do mapeamento sistemático consistiu na definição das combinações de palavras-chave utilizadas nas buscas. A Tabela 1 apresenta as expressões aplicadas nas buscas, bem como o número inicial de resultados encontrados em cada caso.

Tabela 1 – Combinações de palavras-chave e número de resultados

Combinação de Palavras-chave	Resultados Encontrados
“incidents against users"	278
“phishing"AND "social media"	378
“malware"AND "social media"	567
“social engineering"AND "social media"	17983
Total Inicial	19206

Fonte: Autoria própria (2025).

Em seguida, foi aplicado um filtro temporal para considerar apenas publicações realizadas entre os anos de 2018 e 2024, período definido como recorte temporal deste estudo. A Tabela 2 apresenta o número de artigos resultantes após essa filtragem para cada combinação de palavras-chave.

Tabela 2 – Artigos após filtro por data (2018–2024)

Palavra-chave	Após filtro por data
“social engineering"AND "social media"	12475
“malware"AND "social media"	370
“phishing"AND "social media"	294
“incidents against users"	169
Total	13308

Fonte: Autoria própria (2025).

Após a aplicação do filtro temporal (2018–2024), foram identificados 13.308 artigos. Como forma de priorizar estudos com maior relevância e impacto acadêmico, os resultados foram ordenados em ordem decrescente de citações. Em seguida, selecionaram-se os 70% artigos mais citados, adicionando também todos os que empataram no número de citações com o último artigo incluído nesse percentual. A Tabela 3 mostra os resultados após a aplicação desse filtro.

Tabela 3 – Artigos após filtro por número de citações

Palavra-chave	Após filtro por citações
“social engineering"AND "social media"	1682
“malware"AND "social media"	60
“phishing"AND "social media"	34
“incidents against users"	15
Total	1791

Fonte: Autoria própria (2025).

Após o filtro por citações, foi realizada uma leitura criteriosa dos títulos dos artigos para eliminar aqueles que não apresentavam relação direta com o tema proposto. A Tabela 4 mostra o número de artigos que permaneceram após essa triagem inicial por título.

Tabela 4 – Artigos selecionados após leitura de título

Palavra-chave	Após leitura de título
“social engineering"AND "social media"	217
“malware"AND "social media"	17
“phishing"AND "social media"	24
“incidents against users"	3
Total	261

Fonte: Autoria própria (2025).

Na etapa seguinte, os resumos e conclusões dos artigos foram analisados com o objetivo de verificar a aderência ao escopo do trabalho e identificar os estudos com maior potencial de contribuição.

A Tabela 5 apresenta a quantidade de artigos que atenderam a esses critérios.

Tabela 5 – Artigos após leitura de resumo e conclusão

Palavra-chave	Após leitura de resumo/conclusão
“social engineering"AND "social media"	30
“malware"AND "social media"	6
“phishing"AND "social media"	10
“incidents against users"	2

Total	48

Fonte: Autoria própria (2025).

Por fim, foi realizada a leitura completa dos artigos previamente selecionados, com base na análise do conteúdo integral. A Tabela 6 apresenta o número final de estudos que foram considerados relevantes e incluídos na análise deste mapeamento sistemático.

Tabela 6 – Artigos selecionados após leitura completa

Palavra-chave	Selecionados na etapa final
“social engineering"AND "social media"	30
“malware"AND "social media"	6
“phishing"AND "social media"	10
“incidents against users"	2
Total	48

Fonte: Autoria própria (2025).

Após a leitura completa foi visto que se manteve os mesmos artigos da etapa anterior. Após a aplicação dos critérios de inclusão e exclusão, foram selecionados 48 artigos para compor a amostra final. A lista completa desses artigos encontra-se no Apêndice I.

4.5 **Coleta e Organização de Dados**

A coleta de dados foi realizada por meio da pesquisa e extração de artigos, conferências, relatórios e outras publicações na base de dados IEEE Xplore, utilizando critérios de busca que incluem termos como "incidentes contra usuários","phishing + redes sociais","malware + redes sociais", "engenharia social + redes sociais".

O processo de seleção foi dividido em etapas para organização e triagem inicial dos resultados, foi utilizado o LibreOffice para gerenciar as listas de publicações e aplicar os filtros necessários:

1. Na triagem inicial: as publicações foram filtradas com base nas palavras chave e no ano de publicação, garantindo que estejam alinhadas aos objetivos da pesquisa.

2. Filtro por número de citações: Em seguida, os estudos foram classificados com base na quantidade de citações, priorizando os mais relevantes academicamente.

3. Leitura dos títulos: Os títulos das publicações foram analisados com o objetivo de excluir aqueles claramente fora do escopo do estudo.

4. Leitura de resumo e conclusões: Foram lidos o resumo e as conclusões das publicações restantes, a fim de verificar a aderência ao tema e a qualidade dos resultados apresentados.

5. Leitura completa: Os artigos aprovados nas etapas anteriores foram lidos na íntegra, possibilitando a extração das informações necessárias à pesquisa.

4.6 **Análise dos Dados**

A análise foi qualitativa e descritiva, seguindo a metodologia de mapeamento sistemático. As publicações foram agrupadas conforme as categorias dos objetivos específicos e analisadas em quatro áreas: tipos de ataques, com a identificação dos mais comuns em redes sociais; vulnerabilidades exploradas, analisando falhas técnicas e comportamentais e medidas de prevenção e mitigação, com foco em soluções técnicas e educativas.

4.7 **Considerações Éticas**

Embora o estudo envolva a análise de publicações acadêmicas e técnicas, não há coleta de dados sensíveis diretamente de indivíduos. No entanto, a pesquisa respeitou as normas de ética científica, garantindo que as fontes e autores sejam devidamente reconhecidos e que os dados sejam utilizados de forma responsável. Essa metodologia permitiu que o estudo alcance seus objetivos, ao mesmo tempo em que oferece uma visão abrangente e detalhada sobre os incidentes de segurança em redes sociais.

5 **RESULTADOS**

Esta seção apresenta os pricipais achados obtidos a partir da análise dos 48 artigos selecionados no mapeamento sistemático. Os resultados foram organizados com base nos objetivos específicos deste estudo: identificação dos tipos de incidentes de segurança e medidas de prevenção e mitigação.

5.1 **Tipos de Incidentes de Segurança Identificados**

A análise permitiu categorizar os principais tipos de incidentes enfrentados por usuários em redes sociais. Os cinco grupos mais frequentes e sua respectiva porcentagem de frequência em relação aos 48 artigos totais utilizados foram:

- Bots Sociais e Comportamento Malicioso (33%): Bots sociais são amplamente utilizados para disseminar desinformação, manipular debates públicos e executar fraudes automatizadas. Segundo (Shukla; Jagtap; Patil, 2021), esses bots simulam interações humanas com grande realismo, dificultando sua detecção por métodos tradicionais.
- Ataques de Phishing (20%): O phishing continua sendo um dos ataques mais frequentes em redes sociais, explorando a confiança dos usuários por meio de mensagens e links fraudulentos. De acordo com (Abdillah *et al.*, 2022), a eficácia dessas fraudes está associada à velocidade de disseminação e ao caráter informal das redes sociais.
- Engenharia Social (18%): Ataques de engenharia social exploram vulnerabilidades emocionais e cognitivas para induzir comportamentos inseguros. (Salama *et al.*, 2023) destacam que esses ataques frequentemente usuam perfis falsos e conteúdos manipulativos para enganar usuários.
- Malware e Ameaças Cibernéticas (15%): Programas maliciosos são disseminados por meio de links camuflados, arquivos anexos e aplicativos de terceiros. (Chaudhari *et al.*, 2024) ressaltam que a ausência de medidas básicas de segurança facilita a propagação dessas ameaças em ambientes digitais.

- Vazamento de dados e privacidade (14%): A exposição indevida de dados pessoais continua sendo uma preocupação crescente. (Li *et al.*, 2020) alertam que muitos usuários desconhecem como suas informações são coletadas e utilizadas, o que agrava os riscos à privacidade nas plataformas.

5.2 **Medidas de Prevenção e Mitigação**

As estratégias de mitigação identificadas nos artigos analisados distribuem-se entre abordagens técnicas, educativas e organizacionais, variando conforme o tipo de incidente de segurança estudado.

- Perfis falsos e bots sociais: Entre os artigos analisados, 93% utilizaram técnicas de *machine Learning* e *Deep Learning*, como *random Forest* e redes neurais, para detectar bots e perfis falsos. Além disso 33% adotaram análise de comportamento para identificar padrões anômalos, enquanto 20% propuseram modelos híbridos ou preditivos. Cerca de 30% desenvolveram ferramentas ou *frameworks* práticos. Evidencia-se que o uso de aprendizado automático e análise comportamental constitui a principal tendência de mitigação nesse tipo de ameaça.
- Phishing: Entre os artigos sobre *phishing*, 80% destacam soluções técnicas como filtros automáticos, análise de URLs maliciosas e sistemas de detecção baseados em aprendizado de máquina. Aproximadamente 30% abordam a educação digital dos usuários, com ênfase na identificação de mensagens fraudulentas, atualização de softwares e cautela com solicitações suspeitas. Observa-se predominância das medidas técnicas, com ações educativas como complemento à prevenção.
- Engenharia Social: Dos artigos sobre engenharia social, 62% enfatizam medidas educativas, como campanhas de conscientização, simulações de ataques e treinamentos em ambientes escolares e corporativos. A cultura de cibersegurança foi apontada em 50% dos estudos como essencial para ambientes organizacionais. Estratégias técnicas mais avançadas, como redes neurais convolucionais e *embeddings* para detectar persuasão em tempo real, apareceram em 13% dos casos. Ações coordenadas durante crises, como a pandemia de COVID-19, foram mencionadas em 25% dos

estudos. Assim, as abordagens predominantes são educativas e organizacionais, com uso crescente de tecnologias preditivas.

- Segurança Geral em Redes Sociais: Em estudos sobre segurança em ambientes virtuais, 100% propõem o uso de ferramentas tecnológicas como antivírus, *firewalls* e IA para detecção automatizada. A responsabilidade das plataformas, com mecanismos automáticos de resposta, foi destacada em 71% dos artigos. já as boas práticas dos usuários foram citadas em 57%, ressaltando a importância do comportamento humano na mitigação.

- Privacidade e Controle de Dados: Nos artigos voltados à privacidade, 83% apontam políticas de privacidade claras e controles granulares como medidas fundamentais adotadas pelas plataformas. A educação dos usuários apareceu em 67% dos estudos, enquanto a responsabilidade compartilhada foi citada em 50%. A segurança da privacidade depende, assim, de uma combinação entre tecnologia, educação e regulação colaborativa.
- Aprendizado profundo e aprendizado de máquina: Todos as artigos sobre este tema apresentam soluções baseadas em aprendizado profundo para detectar *bots*, perfis falsos e conteúdo malicioso. Técnicas como redes neurais profundas e LSTM foram utilizadas em 67% dos estudos. Modelos multimodais, que integram texto, imagens e metadados, surgiram em 17% dos casos, assim como abordagens voltadas à detecção linguística específica. Essas soluções reforçam a importância da adaptação tecnológica aos diferentes contexto sociais e culturais.

Em síntese, os dados relevam uma forte presença de soluções técnicas, especialmente baseadas em aprendizado de máquina, em praticamente todos os tipos de incidentes analisados. No entanto, estratégias educativas e colaborativas se mostram igualmente relevantes em temas como engenharia social, privacidade e comportamento do usuário, evidenciando a necessidade de abordagens integradas para uma mitigação mais eficaz.

6 CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho apresentou um mapeamento sistemático da literatura entre 2018 e 2024 sobre incidentes de segurança contra usuários em redes sociais, com o objetivo de identificar os principais tipos de ataques, as vulnerabilidades exploradas e as medidas de prevenção e mitigação propostas na literatura. A análise permitiu categorizar os incidentes mais frequentes, como o uso de *bots* sociais, ataques de *phishing*, técnicas de engenharia social, disseminação de *malwares* e vazamentos de dados pessoais, além de destacar o papel das tecnologias de aprendizado de máquina na detecção automatizada dessas ameaças.

Os resultados evidenciam que, embora existam diversas soluções técnicas em desenvolvimento, como autenticação de dois fatores, criptografia, filtros inteligentes e modelos preditivos baseados em inteligência artificial, ainda há uma lacuna significativa na conscientização dos usuários quanto aos riscos e às práticas seguras no uso das redes sociais. A engenharia social, por exemplo, continua se mostrando altamente eficaz por explorar falhas comportamentais e emocionais, o que reforça a importância de estratégias educativas contínuas e acessíveis.

Este mapeamento contribui para uma visão integrada dos desafios enfrentados pelos usuários em ambientes digitais, fornecendo ajuda e incentivos para pesquisadores, desenvolvedores e formuladores de políticas públicas na construção de ambientes online mais seguros.

7 LIMITAÇÕES DO TRABALHO

Apesar de ter atingido os objetivos propostos, este estudo apresenta algumas limitações que devem ser consideradas na interpretação dos resultados:

- Base de dados única: A pesquisa foi conduzida apenas na base IEEE Xplore. Embora essa seja uma das principais fontes de publicações na área de tecnologia, o uso de outras bases científicas (como Scopus, Web of Science e ACM) poderia ampliar o escopo dos resultados.

- Filtro por citações: A seleção dos 70% artigos mais citados priorizou publicações de maior impacto científico, mas pode ter excluído estudos recentes e relevantes com baixo número de citações.

7.1 **TRABALHOS FUTUROS**

Como proposta para trabalhos futuros, sugere-se a realização de estudos empíricos que avaliem a eficácia das medidas educativas em diferentes contextos sociais, bem como a criação de ferramentas *open source* voltadas à detecção em tempo real de ameaças emergentes em redes sociais.

REFERÊNCIAS

ABDILLAH, R. *et al.* Phishing classification techniques: A systematic literature review. **Piscatway: IEEE**, v. 10, p. 41574–41591, 2022. Disponível em: https://ieeexplore.ieee.org/document/9755138. Acesso em: 2 março 2025.

ADIL, M.; KHAN, R.; GHANI, M. A. N. U. Preventive techniques of phishing attacks in networks. In: **2020 3rd International Conference on Advancements in Computational Sciences (ICACS).** 2020. p. 1–8. Disponível em: https://ieeexplore.ieee.org/document/9055943. Acesso em: 8 março 2025.

ALDAWOOD, H.; SKINNER, G. Educating and raising awareness on cyber security social engineering: A literature review. In: **2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE).** 2018. p. 62– 68. Disponível em: https://ieeexplore.ieee.org/document/8615162. Acesso em: 2 março 2025.

ARIN, E.; KUTLU, M. Deep learning based social bot detection on twitter. **IEEE Transactions on Information Forensics and Security**, v. 18, p. 1763–1772, 2023. Disponível em: https://ieeexplore.ieee.org/document/10064110. Acesso em: 7 março 2025.

ARSHEY, M.; VIJI, K. S. A. Thwarting cyber crime and phishing attacks with machine learning: A study. In: **2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS).** 2021. v. 1, p. 353–357. Disponível em: https://ieeexplore.ieee.org/document/9441925. Acesso em: 7 março 2025.

BESKOW, D. M.; CARLEY, K. M. Bot conversations are different: Leveraging network metrics for bot detection in twitter. In: **2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).** 2018. p. 825–832. Disponível em: https://ieeexplore.ieee.org/document/8508322. Acesso em: 8 março 2025.

BITAAB, M. *et al.* Scam pandemic: How attackers exploit public fear through phishing.
In:**2020 APWG Symposium on Electronic Crime Research (eCrime).** 2020. p. 1–10. Disponível em: https://ieeexplore.ieee.org/document/9493260. Acesso em: 7 março 2025.

CENGIZ, A. B.; KALEM, G.; BOLUK, P. S. The effect of social media user behaviors on security and privacy threats. **Piscatway: IEEE**, v. 10, p. 57674–57684, 2022. Disponível em: https://ieeexplore.ieee.org/document/9785366. Acesso em: 3 março 2025.

CHAUDHARI, A. *et al.* Cyber security challenges in social meta-verse and mitigation techniques. In: **2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon).** 2024. p. 1–7. Disponível em: https://ieeexplore.ieee.org/document/10575295. Acesso em: 8 março 2025.

COHEN, A.; NISSIM, N.; ELOVICI, Y. Maljpeg: Machine learning based solution for the detection of malicious jpeg images. **Piscatway: IEEE**, v. 8, p. 19997–20011, 2020. Disponível em: https://ieeexplore.ieee.org/document/8967109. Acesso em: 8 março 2025.

FALBO, R. de A. **Mapeamento sistemático**. Retrieved October, v. 7, 2018. Disponível em: http://claudiaboeres.pbworks.com/w/file/fetch/133747116/Mapeamento%20Sistem%C3 %A1t ico%20%20v1.0. Acesso em: 10 fevereiro 2025.

FAZIL, M.; SAH, A. K.; ABULAISH, M. Deepsbd: A deep neural network model with attention mechanism for socialbot detection. **IEEE Transactions on Information Forensics and Security**, v. 16, p. 4211–4223, 2021. Disponível em: https://ieeexplore.ieee.org/document/9505695. Acesso em: 7 março 2025.

GOYAL, B. et al. Detection of Fake Accounts on Social Media Using Multimodal Data With Deep Learning. **IEEE Transactions on Computational Social Systems**, 2023. p. 1–12. Disponível em: https://ieeexplore.ieee.org/document/10210324. Acesso em 3 março 2025.

HARRIS, P. *et al.* Fake instagram profile identification and classification using machine learning. In**: 2021 2nd Global Conference for Advancement in Technology (GCAT).** 2021. p. 1–5. Disponível em: https://ieeexplore.ieee.org/document/9587858. Acesso em: 7 março 2025.

HE, Z.; CAI, Z.; YU, J. Latent-data privacy preserving with customized data utility for social network data. **IEEE Transactions on Vehicular Technology**, v. 67, n. 1, p. 665– 673, 2018. Disponível em: https://ieeexplore.ieee.org/document/8007315. Acesso em: 2 março 2025.

HIJJI, M.; ALAM, G. A multivocal literature review on growing social engineering based cyber-attacks/threats during the covid-19 pandemic: Challenges and prospective solutions. **Piscatway: IEEE**, v. 9, p. 7152–7169, 2021. Disponivel em: https://ieeexplore.ieee.org/document/8615162. Acesso em: 5 março 2025.

HU, X. *et al.* Privacy data propagation and preservation in social media: A real-world case study. **IEEE Transactions on Knowledge and Data Engineering**, v. 35, n. 4,

p. 4137– 4150, 2023. Disponível em: https://ieeexplore.ieee.org/document/9658172. Acesso em: 5 março 2025.

ISLAM, T.; LATIF, S.; AHMED, N. Using social networks to detect malicious bangla text content. In**: 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT).** 2019. p. 1–4. Disponível em: https://ieeexplore.ieee.org/document/8934841. Acesso em: 2 março 2025.

JAMIL, A. *et al.* Mpmpa: A mitigation and prevention model for social engineering based phishing attacks on facebook. In: **2018 IEEE International Conference on Big Data (Big Data).** 2018. p. 5040–5048. Disponível em: https://ieeexplore.ieee.org/document/8622505. Acesso em: 8 março 2025.

KORKMAZ, M.; SAHINGOZ, O. K.; DIRI, B. Detection of phishing websites by using machine learning-based url analysis. In: **2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT).** 2020. p. 1–7. Disponível em: https://ieeexplore.ieee.org/document/9225561. Acesso em: 3 março 2025.

KUMAR, M. S. *et al.* Fake profile detection on social networking websites using machine learning. In: **2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS).** 2023. p. 119–122. Disponível em: https://ieeexplore.ieee.org/document/10169168. Acesso em: 3 março 2025.

LANKTON, N. K.; MCKNIGHT, D. H.; TRIPP, J. F. Understanding the antecedents and outcomes of facebook privacy behaviors: An integrated model. **IEEE Transactions on Engineering Management**, v. 67, n. 3, p. 697–711, 2020. Disponível em: https://ieeexplore.ieee.org/document/8638541. Acesso em: 7 março 2025.

LI, H. *et al.* Privacy leakage via de-anonymization and aggregation in heterogeneous social networks. **IEEE Transactions on Dependable and Secure Computing**, v. 17, n. 2, p. 350–362, 2020. Disponível em: https://ieeexplore.ieee.org/document/8047477. Acesso em: 7 março 2025.

NICHOLSON, J. *et al.* Investigating teenagers' ability to detect phishing messages. In: **2020 IEEE European Symposium on Security and Privacy Workshops (EuroSPW).** 2020. p. 140–149. Disponível em: https://ieeexplore.ieee.org/document/9229735. Acesso em: 7 março 2025.

OEST, A. *et al.* Inside a phisher's mind: Understanding the anti-phishing ecosystem through phishing kit analysis. In: **2018 APWG Symposium on Electronic Crime Research (eCrime).** 2018. p. 1–12. Disponível em: https://ieeexplore.ieee.org/document/8376206. Acesso em: 3 março 2025.

OZKER, U.; SAHINGOZ, O. K. Content based phishing detection with machine learning. In:
2020 International Conference on Electrical Engineering (ICEE). 2020. p. 1– 6**.** Disponível em: https://ieeexplore.ieee.org/document/9249892**.** Acesso em: 3 março 2025.

LATHA, P. *et al.* Fake profile identification in social network using machine learning and nlp. In: **2022 International Conference on Communication, Computing and Internet of Things (IC3IoT).** 2022. p. 1–4. Disponível em: https://ieeexplore.ieee.org/document/9767958. Acesso em: 5 março 2025.

PARIKH, S. B.; KHEDIA, S. R.; ATREY, P. K. A framework to detect fake tweet images on social media. In: **2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM).** 2019. p. 104–110. Disponível em: https://ieeexplore.ieee.org/document/8919459. Acesso em: 8 março 2025.

PARTHY, P. P.; RAJENDRAN, G. Identification and prevention of social engineering attacks on an enterprise. In: **2019 International Carnahan Conference on Security Technology (ICCST).** 2019. p. 1–5. Disponível em: https://ieeexplore.ieee.org/document/8888441. Acesso em: 8 março 2025.

PIETRO, R. D.; CRESCI, S. Metaverse: Security and privacy issues. In: **2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA).** 2021. p. 281–288. Disponível em: https://ieeexplore.ieee.org/document/9750221. Acesso em: 2 março 2025.

RIPA, S. P.; ISLAM, F.; ARIFUZZAMAN, M. The emergence threat of phishing attack and the detection techniques using machine learning models. In: **2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI).** 2021. p. 1–6. Disponível em: https://ieeexplore.ieee.org/document/9528204. Acesso em: 8 março 2025.

SALAMA, R. *et al.* Social engineering attack types and prevention techniques- a survey. In: **2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN).** 2023. p. 817–820. Disponível em: https://ieeexplore.ieee.org/document/10140957. Acesso em: 7 março 2025.

SANSONETTI, G. *et al.* Unreliable users detection in social media: Deep learning techniques for automatic detection. **Piscatway: IEEE**, v. 8, p. 213154–213167, 2020. Disponível em: https://ieeexplore.ieee.org/document/9269985. Acesso em: 5 março 2025.

SCHNEBLY, J.; SENGUPTA, S. Random forest twitter bot classifier. In: **2019 IEEE 9th Annual Computing and Communication Workshop and Conference**

(CCWC). 2019. p. 0506–0512. Disponível em: https://ieeexplore.ieee.org/document/8666593. Acesso em: 5 março 2025.

SHAH, R. K. *et al.* Detect phishing website by fuzzy multi-criteria decision making. In: **2022 1st International Conference on AI in Cybersecurity (ICAIC).** 2022. p. 1–8.Disponível em: https://ieeexplore.ieee.org/document/9897036. Acesso em 5 março 2025.

SHARMA, V. *et al.* Nhad: Neuro-fuzzy based horizontal anomaly detection in online social networks. **IEEE Transactions on Knowledge and Data Engineering**, v. 30, n. 11, p. 2171–2184, 2018. Disponível em: https://ieeexplore.ieee.org/document/8322278. Acesso em: 5 março 2025.

SHEVCHUK, R. *et al.* Software for automatic estimating security settings of social media accounts. In: **2020 10th International Conference on Advanced Computer Information Technologies (ACIT).** 2020. p. 769–773. Disponível em: https://ieeexplore.ieee.org/document/9208879. Acesso em: 8 março 2025.

SHI, P.; ZHANG, Z.; CHOO, K.-K. R. Detecting malicious social bots based on clickstream sequences. **Piscatway: IEEE**, v. 7, p. 28855–28862, 2019. Disponível em: https://ieeexplore.ieee.org/document/8653381. Acesso em: 7 março 2025.

SHIVANGI, S. *et al.* Chrome extension for malicious urls detection in social media applications using artificial neural networks and long short term memory networks. In: **2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI).** 2018. p. 1993–1997. Disponível em: https://ieeexplore.ieee.org/document/8554647. Acesso em: 7 março 2025.

SHUKLA, H.; JAGTAP, N.; PATIL, B. Enhanced twitter bot detection using ensemble machine learning. In: **2021 6th International Conference on Inventive Computation Technologies (ICICT).** 2021. p. 930–936. Disponível em: https://ieeexplore.ieee.org/document/9358734. Acesso em: 3 março 2025.

SOWMYA, P.; CHATTERJEE, M. Detection of fake and clone accounts in twitter using classification and distance measure algorithms. In: **2020 International Conference on Communication and Signal Processing (ICCSP).** 2020. p. 0067–0070. Disponível em: https://ieeexplore.ieee.org/document/9182353. Acesso em: 5 março 2025.

SUN, N. *et al.* Data-driven cybersecurity incident prediction: A survey. **IEEE Communications Surveys Tutorials**, v. 21, n. 2, p. 1744–1772, 2019. Disponível em: https://ieeexplore.ieee.org/document/8567980. Acesso em: 2 março 2025.

SYAFITRI, W. *et al.* Social engineering attacks prevention: A systematic literature review. **Piscatway: IEEE**, v. 10, p. 39325–39343, 2022. Disponível em: https://ieeexplore.ieee.org/document/9743471. Acesso em: 7 março 2025.

TANG, L.; MAHMOUD, Q. H. A deep learning-based framework for phishing website detection. **Piscatway: IEEE**, v. 10, p. 1509–1521, 2022. Disponível em: https://ieeexplore.ieee.org/document/9661323. Acesso em: 5 março 2025.

THAKUR, K.; HAYAJNEH, T.; TSENG, J. Cyber security in social media: Challenges and the way forward. **IT Professional**, v. 21, n. 2, p. 41–49, 2019. Disponível em: https://ieeexplore.ieee.org/document/8676127. Acesso em: 8 março 2025.

TSINGANOS, N.; MAVRIDIS, I.; GRITZALIS, D. Utilizing convolutional neural networks and word embeddings for early-stage recognition of persuasion in chatbased social engineering attacks**. Piscatway: IEEE** v. 10, p. 108517–108529, 2022. Disponível em: https://ieeexplore.ieee.org/document/9915571. Acesso em: 5 março 2025.

VERMA, A. *et al.* Spammer detection and fake user identification on social networks. In: **2023 IEEE Technology Engineering Management Conference - Asia Pacific (TEMSCON-ASPAC).** 2023. p. 1–7. Disponível em: https://ieeexplore.ieee.org/document/10531323. Acesso em: 7 março 2025.

WANG, X. *et al.* Drifted twitter spam classification using multiscale detection test on kl divergence. **Piscatway: IEEE**, v. 7, p. 108384–108394, 2019. Disponível em: https://ieeexplore.ieee.org/document/8781937. Acesso em: 8 março 2025.

YURTSEVEN, ; BAGRIYANIK, S.; AYVAZ, S. A review of spam detection in social media. In: **2021 6th International Conference on Computer Science and Engineering (UBMK).** 2021. p. 383–388. Disponível em: https://ieeexplore.ieee.org/document/9558993. Acesso em: 5 março 2025.

1 **APÊNDICE I**

1.1 **48 ARTIGOS UTILIZADOS NO MAPEAMENTO**

SUN, N. et al. Data-Driven Cybersecurity Incident Prediction: A Survey. **IEEE Communications Surveys & Tutorials**, v. 21, n. 2, p. 1744–1772, 2019. Disponível em: https://ieeexplore.ieee.org/document/8567980. Acesso em: 2 março 2025. (261 Citações)

DI PIETRO, R.; CRESCI, S. Metaverse: Security and Privacy Issues. In: **2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)**. [S. l.]: IEEE, 2021. p. 281–288. Disponível em: https://ieeexplore.ieee.org/document/9750221 .
Acesso em: 2 março 2025. (164 Citações)

HE, Z.; CAI, Z.; YU, J. Latent-Data Privacy Preserving With Customized Data Utility for

Social Network Data. **IEEE Transactions on Vehicular Technology**, v. 67, n. 1, p. 665–673, 2018. Disponível em: https://ieeexplore.ieee.org/document/8007315. Acesso em: 2 março 2025. (144 Citações)

ALDAWOOD, H.; SKINNER, G. Educating and Raising Awareness on Cyber Security Social Engineering: A Literature Review. In: **2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)**. [S. l.]: IEEE, 2018. p. 62–68. Disponível em: https://ieeexplore.ieee.org/document/8615162. Acesso em: 2 março 2025. (84 Citações)

KORKMAZ, M.; SAHINGOZ, O. K.; DIRI, B. Detection of Phishing Websites by Using Machine Learning-Based URL Analysis. In: **2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)**. [S. l.]: IEEE, 2020. p. 1–7. Disponível em: https://ieeexplore.ieee.org/document/9225561. Acesso em: 3 março 2025. (80 Citações)

OEST, A. et al. Inside a phisher's mind: Understanding the anti-phishing ecosystem through phishing kit analysis. In: **2018 APWG Symposium on Electronic Crime Research (eCrime)**. [S. l.]: IEEE, 2018. p. 1–12. Disponível em: https://ieeexplore.ieee.org/document/8376206. Acesso em: 3 março 2025. (80 Citações)

HIJJI, M.; ALAM, G. 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. Disponivel em: https://ieeexplore.ieee.org/document/8615162. Acesso em: 5 março 2025. (75 Citações)

SANSONETTI, G. et al. Unreliable Users Detection in Social Media: Deep Learning Techniques for Automatic Detection. **IEEE Access**, v. 8, p. 213154–213167, 2020. Disponível em: https://ieeexplore.ieee.org/document/9269985. Acesso em: 5 março 2025. (66 Citações)

TANG, L.; MAHMOUD, Q. H. A Deep Learning-Based Framework for Phishing Website Detection. **IEEE Access**, v. 10, p. 1509–1521, 2022. Disponível em: https://ieeexplore.ieee.org/document/9661323. Acesso em: 5 março 2025. (59 Citações)

FAZIL, M.; SAH, A. K.; ABULAISH, M. DeepSBD: A Deep Neural Network Model With Attention Mechanism for SocialBot Detection. **IEEE Transactions on Information Forensics and Security**, v. 16, p. 4211–4223, 2021. Disponível em: https://ieeexplore.ieee.org/document/9505695. Acesso em: 7 março 2025. (58 Citações)

SHI, P.; ZHANG, Z.; CHOO, K.-K. R. Detecting Malicious Social Bots Based on Clickstream Sequences. **IEEE Access**, v. 7, p. 28855–28862, 2019. Disponível em: https://ieeexplore.ieee.org/document/8653381.
Acesso em: 7 março 2025. (57 Citações)

SYAFITRI, W. et al. Social Engineering Attacks Prevention: A Systematic Literature Review.
IEEE Access, v. 10, p. 39325–39343, 2022. Disponível em: https://ieeexplore.ieee.org/document/9743471. Acesso em: 7 março 2025. (55 Citações)

SALAMA, R. et al. Social engineering attack types and prevention techniques- A survey. In: **2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)**. [S. l.: s. n.], 2023. p. 817–820. Disponível em: https://ieeexplore.ieee.org/document/10140957.
Acesso em: 7 março 2025. (53 Citações)

LI, H. et al. Privacy Leakage via De-Anonymization and Aggregation in Heterogeneous Social Networks. **IEEE Transactions on Dependable and Secure Computing**, v. 17, n. 2, p. 350–362, 2020. Disponível em: https://ieeexplore.ieee.org/document/8047477. Acesso em: 7 março 2025. (53 Citações)

VERMA, A. et al. Spammer Detection And Fake User Identification On Social Networks. In: **2023 IEEE Technology & Engineering Management Conference - Asia Pacific (TEMSCON-ASPAC)**. [S. l.: s. n.], 2023. p. 1–7. Disponível em: https://ieeexplore.ieee.org/document/10531323.
Acesso em: 7 março 2025. (50 Citações)

ARIN, E.; KUTLU, M. Deep Learning Based Social Bot Detection on Twitter. **IEEE Transactions on Information Forensics and Security**, v. 18, p. 1763–1772, 2023. Disponível em: https://ieeexplore.ieee.org/document/10064110. Acesso em: 7 março 2025. (41 Citações)

CHAUDHARI, A. et al. Cyber Security Challenges in Social Meta-verse and Mitigation Techniques. In: **2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon)**. [S. l.]: IEEE, 2024. p. 1–7. Disponível em: https://ieeexplore.ieee.org/document/10575295. Acesso em: 8 março 2025. (40 Citações)

COHEN, A.; NISSIM, N.; ELOVICI, Y. MalJPEG: Machine Learning Based Solution for the Detection of Malicious JPEG Images. **IEEE Access**, v. 8, p. 19997–20011, 2020. Disponível em: https://ieeexplore.ieee.org/document/8967109. Acesso em: 8 março 2025. (37 Citações)

THAKUR, K.; HAYAJNEH, T.; TSENG, J. Cyber Security in Social Media: Challenges and the Way Forward. **IT Professional**, v. 21, n. 2, p. 41–49, 2019. Disponível em: https://ieeexplore.ieee.org/document/8676127.
Acesso em: 8 março 2025. (34 Citações)

WANG, X. et al. Drifted Twitter Spam Classification Using Multiscale Detection Test on K-L Divergence. **IEEE Access**, v. 7, p. 108384–108394, 2019. Disponível em: https://ieeexplore.ieee.org/document/8781937.
Acesso em: 8 março 2025. (34 Citações)

BESKOW, D. M.; CARLEY, K. M. Bot Conversations are Different: Leveraging Network Metrics for Bot Detection in Twitter. In: **2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)**. [S. l.: s. n.], 2018. p. 825–832. Disponível em: https://ieeexplore.ieee.org/document/8508322. Acesso em: 8 março 2025. (34 Citações)

RIPA, S. P.; ISLAM, F.; ARIFUZZAMAN, M. The Emergence Threat of Phishing Attack and The Detection Techniques Using Machine Learning Models. In: **2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)**.

[S. l.]: IEEE, 2021. p. 1–6. Disponível em: https://ieeexplore.ieee.org/document/9528204. Acesso em: 8 março 2025. (28 Citações)

ISLAM, T.; LATIF, S.; AHMED, N. Using Social Networks to Detect Malicious Bangla Text Content. In: **2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)**. [S. l.: s. n.], 2019. p. 1–4. Disponível em: https://ieeexplore.ieee.org/document/8934841.
Acesso em: 2 março 2025. (27 Citações)

ABDILLAH, R. et al. Phishing Classification Techniques: A Systematic Literature Review.
IEEE Access, v. 10, p. 41574–41591, 2022. Disponível em: https://ieeexplore.ieee.org/document/9755138. Acesso em: 2 março 2025. (27 Citações)

SHAH, R. K. et al. Detect Phishing Website by Fuzzy Multi-Criteria Decision Making. In:
2022 1st International Conference on AI in Cybersecurity (ICAIC). [S. l.]: IEEE, 2022. p. 1–8. Disponível em: https://ieeexplore.ieee.org/document/9897036. Acesso em: 5 março 2025. (26 Citações)

GOYAL, B. et al. Detection of Fake Accounts on Social Media Using Multimodal Data With Deep Learning. **IEEE Transactions on Computational Social Systems**, 2023. p. 1–12. Disponível em: https://ieeexplore.ieee.org/document/10210324. Acesso em: 3 março 2025. (25 Citações)

BITAAB, M. et al. Scam Pandemic: How Attackers Exploit Public Fear through Phishing.
In:**2020 APWG Symposium on Electronic Crime Research (eCrime)**. [S. l.]: IEEE, 2020. p. 1–10. Disponível em: https://ieeexplore.ieee.org/document/9493260. Acesso em: 7 março 2025. (24 Citações)

SOWMYA, P.; CHATTERJEE, M. Detection of Fake and Clone accounts in Twitter using
Classification and Distance Measure Algorithms. In: **2020 International Conference on Communication and Signal Processing (ICCSP)**. [S. l.: s. n.], 2020. p. 0067–0070.
Disponível em: https://ieeexplore.ieee.org/document/9182353. Acesso em: 5 março 2025. (23 Citações)

SHUKLA, H.; JAGTAP, N.; PATIL, B. Enhanced Twitter bot detection using ensemble machine learning. In: **2021 6th International Conference on Inventive Computation Technologies (ICICT)**. [S. l.: s. n.], 2021. p. 930–936. Disponível em: https://ieeexplore.ieee.org/document/9358734.
Acesso em: 3 março 2025. (22 Citações)

HARRIS, P. et al. Fake Instagram Profile Identification and Classification using Machine Learning. In: **2021 2nd Global Conference for Advancement in Technology (GCAT)**.
[S. l.]: IEEE, 2021. p. 1–5. Disponível em: https://ieeexplore.ieee.org/document/9587858.
Acesso em: 7 março 2025. (21 Citações)

ADIL, M.; KHAN, R.; GHANI, M. A. N. U. Preventive Techniques of Phishing Attacks in
Networks. In: **2020 3rd International Conference on Advancements in Computational Sciences (ICACS)**. [S. l.]: IEEE, 2020. p. 1–8. Disponível em:

https://ieeexplore.ieee.org/document/9055943.
Acesso em: 8 março 2025. (21 Citações)

NICHOLSON, J. et al. Investigating Teenagers' Ability to Detect Phishing Messages. In:
2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). [S. l.]: IEEE, 2020. p. 140–149. Disponível em: https://ieeexplore.ieee.org/document/9229735.

Acesso em: 7 março 2025. (20 Citações)

PARTHY, P. P.; RAJENDRAN, G. Identification and prevention of social engineering attacks on an enterprise. In:
2019 International Carnahan Conference on Security Technology (ICCST). [S. l.: s. n.], 2019. p. 1–5. Disponível em: https://ieeexplore.ieee.org/document/8888441.

Acesso em: 8 março 2025. (19 Citações)

JAMIL, A. et al. MPMPA: A Mitigation and Prevention Model for Social Engineering Based
Phishing attacks on Facebook. In: **2018 IEEE International Conference on Big Data (Big Data)**. [S. l.: s. n.], 2018. p. 5040–5048. Disponível em:

https://ieeexplore.ieee.org/document/8622505. Acesso em: 8 março 2025. (19 Citações)

SHEVCHUK, R. et al. Software for Automatic Estimating Security Settings of Social Media Accounts. In: **2020 10th International Conference on Advanced Computer Information Technologies (ACIT)**. [S. l.: s. n.], 2020. p. 769–773. Disponível em: https://ieeexplore.ieee.org/document/9208879.

Acesso em: 8 março 2025. (18 Citações)

YURTSEVEN, İ.; BAGRIYANIK, S.; AYVAZ, S. A Review of Spam Detection in Social Media. In: **2021 6th International Conference on Computer Science and Engineering (UBMK)**.[S. l.]: IEEE, 2021. p. 383–388. Disponível em: https://ieeexplore.ieee.org/document/9558993. Acesso em: 5 março 2025. (18 Citações)

LATHA, P. et al. Fake Profile Identification in Social Network using Machine Learning and NLP. In:**2022 International Conference on Communication, Computing and Internet of Things IC3IoT)**. [S. l.]: IEEE, 2022. p. 1–4. Disponível em: https://ieeexplore.ieee.org/document/9767958. Acesso em: 5 março 2025. (17 Citações)

OZKER, U.; SAHINGOZ, O. K. Content Based Phishing Detection with Machine Learning. In: **2020 International Conference on Electrical Engineering (ICEE)**. [S. l.]: IEEE, 2020. p. 1–6. Disponível em: https://ieeexplore.ieee.org/document/9249892. Acesso em: 3 março 2025. (17 Citações)

CENGIZ, A. B.; KALEM, G.; BOLUK, P. S. The Effect of Social Media User Behaviors on Security and Privacy Threats. **IEEE Access**, v. 10, p. 57674–57684, 2022. Disponível em: https://ieeexplore.ieee.org/document/9785366. Acesso em: 3 março 2025. (17 Citações)

SCHNEBLY, J.; SENGUPTA, S. Random Forest Twitter Bot Classifier. In: **2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)**.[S.l.:s.n.],2019.p.0506–0512.Disponível em: https://ieeexplore.ieee.org/document/8666593. Acesso em: 5 março 2025. (15 Citações)

ARSHEY, M.; VIJI, K. S. A. Thwarting Cyber Crime and Phishing Attacks with Machine Learning: A Study. In: **2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)**. [S. l.: s. n.], 2021. v. 1, p. 353–357. Disponível em: https://ieeexplore.ieee.org/document/9441925. Acesso em: 7 março 2025. (14 Citações)

SHARMA, V. et al. NHAD: Neuro-Fuzzy Based Horizontal Anomaly Detection in

Online Social Networks. **IEEE Transactions on Knowledge and Data Engineering**,v.30,n.11,p.2171–2184,2018.Disponíve em: https://ieeexplore.ieee.org/document/8322278. Acesso em: 5 março 2025. (14 Citações)

SHIVANGI, S. et al. Chrome Extension For Malicious URLs detection in Social Media Applications Using Artificial Neural Networks And Long Short Term Memory Networks. In: **2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)**. [S. l.]: IEEE, 2018. p. 1993–1997.

Disponível em: https://ieeexplore.ieee.org/document/8554647.
Acesso em: 7 março 2025. (13 Citações)

HU, X. et al. Privacy Data Propagation and Preservation in Social Media: A RealWorld Case Study. **IEEE Transactions on Knowledge and Data Engineering**, v. 35,n.4,p. 4137–4150, 2023. Disponível em: https://ieeexplore.ieee.org/document/9658172.
Acesso em: 5 março 2025. (13 Citações)

TSINGANOS, N.; MAVRIDIS, I.; GRITZALIS, D. Utilizing Convolutional Neural Networks and Word Embeddings for Early-Stage Recognition of Persuasion in ChatBased Social Engineering Attacks. **IEEE Access**, v. 10, p. 108517–108529, 2022.
Disponível em: https://ieeexplore.ieee.org/document/9915571.
Acesso em: 5 março 2025. (12 Citações)

LANKTON, N. K.; MCKNIGHT, D. H.; TRIPP, J. F. Understanding the Antecedents and Outcomes of Facebook Privacy Behaviors: An Integrated Model. **IEEE Transactions on Engineering Management**, v. 67, n. 3, p. 697–711, 2020.
Disponível em: https://ieeexplore.ieee.org/document/8638541.
Acesso em: 7 março 2025. (12 Citações)

PARIKH, S. B.; KHEDIA, S. R.; ATREY, P. K. A Framework to Detect Fake Tweet Images on Social Media. In: **2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)**.[S.l.]:IEEE,2019.p.104–110.Disponível em: https://ieeexplore.ieee.org/document/8919459. Acesso em: 8 março 2025. (12 Citações)

K. MANASA, V. et al. Fake Profile Detection on Social Networking Websites using Machine Learning. In: **2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)**. [S. l.: s. n.], 2023. p. 119–122. Disponível em: https://ieeexplore.ieee.org/document/10169168.
Acesso em: 8 março 2025. (4 Citações)

DE ALMEIDA FALBO, Ricardo. Mapeamento sistemático. **Retrieved October**, v. 7, 2018.Disponível em: http://claudiaboeres.pbworks.com/w/file/fetch/133747116/Mapeamento%20Sistem%C3 %A1t ico%20%20v1.0 . Acesso em: 10 fevereiro 2025.