

A EXPLORAÇÃO DO VIÉS DE AUTORIDADE EM ATAQUES DE PHISHING: UMA ANÁLISE PSICOLÓGICA DA ENGENHARIA SOCIAL

André Pinheiro Francimat¹

¹ Ministério do Desenvolvimento, Indústria, Comércio e Serviços, Brasília, DF, Brasil

Professor orientador: Prof. Dr. Rafael Rabelo Nunes

Resumo. Este estudo investiga a intersecção entre ataques de *phishing* e vulnerabilidades psicológicas em ambientes corporativos, demonstrando que apenas defesas tecnológicas são insuficientes contra a engenharia social. Por meio de uma revisão integrativa de literatura, de caráter qualitativo e exploratório, o trabalho avalia os determinantes cognitivos que condicionam a decisão das vítimas, focando no viés de autoridade. A revisão foi conduzida nas bases Scopus, Web of Science e Google Acadêmico, abrangendo publicações em português, inglês e espanhol de 2022 a 2025. Após critérios de inclusão e exclusão, selecionou-se 12 estudos contemporâneos, complementados por clássicos da psicologia social e relatórios institucionais para contextualização empírica do fenômeno. Esse atalho mental inibe o processamento crítico da informação, induzindo respostas automáticas, sendo amplamente explorado em ataques direcionados e potencializado por tecnologias de *deepfakes*. A análise identificou que o mecanismo de obediência opera por meio de microações sequenciais e é amplificado por *deepfakes* que deslocam a validação do campo racional para o sensorial. A principal contribuição do estudo é posicionar a autoridade como eixo estruturante da análise, articulando a teoria clássica da obediência com ataques contemporâneos como *Business Email Compromise* e fraude do CEO, propondo um conjunto integrado de diretrizes psico-organizacionais de mitigação.

Palavras-chave: Viés de autoridade; Fraudes Corporativas; Segurança da Informação; Cultura Organizacional; *Deepfakes*.

Abstract. This study investigates the intersection between phishing attacks and psychological vulnerabilities in corporate environments, demonstrating that technological defenses alone are insufficient against social engineering. Through an integrative literature review with a qualitative and exploratory approach, the study evaluates the cognitive determinants that shape victims' decision-making, focusing on the authority bias. The review was conducted using the Scopus, Web of Science, and Google Scholar databases, covering publications in Portuguese, English, and Spanish published between 2022 and 2025. After applying the inclusion and exclusion criteria, 12 contemporary studies were selected and complemented by seminal works in social psychology and institutional reports to provide an empirical and theoretical context for

the phenomenon. This mental shortcut inhibits critical information processing, inducing automatic responses, and is widely exploited in targeted attacks and further enhanced by deepfake technologies. The analysis identified that the obedience mechanism operates through sequential micro-actions and is amplified by deepfakes, which shift validation from the rational domain to the sensory domain. The main contribution of the study is to position authority as the structuring axis of the analysis, articulating the classical theory of obedience with contemporary attacks such as Business Email Compromise and CEO fraud, while proposing an integrated set of psycho-organizational mitigation guidelines.

Keywords: *Authority Bias; Corporate Fraud; Information Security; Organizational Culture; Deepfakes.*

1. Introdução

A transformação digital tornou tecnologias corporativas indispensáveis, mas, com o aprimoramento das defesas técnicas, os cibercriminosos passaram a focar num dos principais pontos de vulnerabilidade da segurança: o comportamento humano.

Conforme apontam Siddiqi et al. (2022), a engenharia social explora a psicologia humana em vez de recorrer apenas à invasão técnica de sistemas, induzindo erros no processo de tomada de decisão para facilitar o contorno de mecanismos de segurança. Nesse cenário, o *phishing* destaca-se como um dos ataques de engenharia social mais recorrentes, combinando técnicas tecnológicas e manipulação psicológica ancorada em heurísticas e vieses cognitivos (Khadka et al., 2024).

A magnitude dessa ameaça é evidenciada em diversos levantamentos. Segundo relatórios da Deloitte e do Swiss Cyber Institute (apud WANG; LUTCHKUS, 2023), aproximadamente 91% dos ataques cibernéticos analisados teriam início por meio de e-mails de *phishing* voltados à exploração da desatenção dos usuários, ao passo que cerca de 95% das invasões a redes corporativas estariam associadas a ataques direcionados de *spear phishing*. Embora esses percentuais sejam dependentes dos contextos e metodologias adotados nos respectivos estudos, eles ilustram a relevância do *phishing* como vetor de comprometimento inicial.

No Brasil, a tendência de alta se confirma: dados do CERT.br (2025) demonstram que o *phishing* superou amplamente os incidentes envolvendo *malwares*, destacando-se como a principal categoria registrada entre as tentativas de fraude analisadas ao longo de todo o ano de 2025.

A eficácia do *phishing* não reside primariamente na complexidade de códigos maliciosos, mas em sua capacidade de induzir a vítima a agir por impulso. O ataque é desenhado para inibir o chamado processamento mental sistemático, impedindo que o usuário tenha o tempo e o senso crítico necessários para avaliar a credibilidade da

mensagem, conforme explorado por Yang et al. (2022). Ao criar cenários de pressão ou obediência, os atacantes forçam respostas irreflexivas e automáticas, como clicar impulsivamente em um link, baixar um anexo ou autorizar uma transação. Para acionar esses atalhos mentais, os cibercriminosos valem-se dos princípios de influência social descritos por Cialdini (2021), sendo a simulação de "autoridade" o mecanismo predominante em ataques direcionados de alto nível hierárquico, a exemplo do *Business Email Compromise* (BEC) e da fraude do CEO (Lawall, 2025).

Sinais de autoridade tendem a ser particularmente bem-sucedidos em ambientes altamente normatizados e hierárquicos. Isso explica por que o setor financeiro e o de governo (administração pública) figuram entre os principais alvos de páginas falsas e investidas criminosas, de acordo com os dados do CERT.br (2025) e o amplo levantamento de literatura realizado por Khadka et al. (2024). O cenário torna-se ainda mais crítico com a evolução tecnológica alertada por Kaushik et al. (2025) e Folorunsho e Boamah (2025), que evidenciam o crescente uso de inteligência artificial e *deepfakes* para simular, com alto grau de realismo, a voz e o rosto de executivos, elevando significativamente o poder de persuasão dos golpes.

Diante da necessidade de compreender a complexa intersecção entre os ataques cibernéticos e as fragilidades cognitivas em ambientes corporativos, observa-se que a literatura científica atual tende a tratar a autoridade de maneira secundária, como apenas mais um elemento em meio a diversos princípios persuasivos da engenharia social, sem o devido aprofundamento de seu papel como mecanismo central. Este estudo responde a essa lacuna teórica ao isolar o viés de autoridade como o eixo analítico estruturante, buscando responder à seguinte questão de pesquisa: de que modo o princípio de autoridade é explorado em ataques de *phishing* em ambientes corporativos e quais estratégias psico-organizacionais podem reduzir a suscetibilidade dos usuários a esse tipo de engenharia social?

Para responder a essa problemática, o objetivo geral desta pesquisa consiste em analisar como o princípio de autoridade é explorado em ataques de *phishing* e propor diretrizes psico-organizacionais de mitigação. Para tanto, delinearam-se os seguintes objetivos específicos: (a) caracterizar o *phishing* como o principal vetor de exploração das vulnerabilidades humanas no ambiente digital; (b) examinar os fundamentos psicológicos do viés de autoridade e sua influência na tomada de decisão; (c) identificar como sinais hierárquicos são operacionalizados na prática em ameaças direcionadas e

fraudes baseadas em *deepfakes*; e (d) propor estratégias de mitigação e de conscientização voltadas à segurança corporativa.

O artigo está estruturado da seguinte forma: a seção 2 apresenta o referencial teórico; a seção 3 discute os trabalhos correlatos; a seção 4 descreve a metodologia; a seção 5 apresenta a análise e discussão dos resultados; e, por fim, a seção 6 apresenta as conclusões e sugestões para trabalhos futuros.

2. Referencial Teórico: a intersecção entre ataques cibernéticos e vulnerabilidades psicológicas

Siddiqi et al. (2022) argumentam que a engenharia social explora a psicologia humana para contornar controles técnicos, enquanto Khadka et al. (2024) indicam que o *phishing* permanece efetivo porque se aproveita de atalhos mentais (heurísticas cognitivas) e vieses da vítima para induzir uma ação, mesmo na presença de mecanismos tecnológicos de proteção. Sob essa perspectiva, entende-se que o *phishing* não deve ser tratado apenas como uma fraude comunicacional, mas como uma estratégia de conversão comportamental: seu objetivo é criar condições de decisão sob pressão nas quais obedecer pareça mais eficiente do que verificar. Portanto, para explicar sua persistência, é necessário analisar os determinantes cognitivos que degradam o processamento crítico, dinâmica aprofundada nos estudos de Yang et al. (2022), e acabam tornando a própria vítima uma parte ativa do vetor de ataque.

As modalidades de *phishing* variam conforme o canal, o grau de personalização e o tipo de alvo. Mais relevante do que a enumeração dessas variações é compreender o elemento comum entre elas: a criação de credibilidade situacional por meio de sinais de legitimidade, que reduzem o custo cognitivo da decisão (Bezerra; Millian, 2024; Schmitt e Flechais, 2024).

Evidências quantitativas reforçam a centralidade do *phishing* como porta de entrada para incidentes. Relatórios da Deloitte e do Swiss Cyber Institute (apud WANG; LUTCHKUS, 2023), relatam elevada participação do *phishing* no início de ataques e violações, sugerindo que a escalabilidade do engano e a dependência de uma única decisão humana (clicar, responder, autorizar) tornam esse vetor altamente eficiente para atacantes.

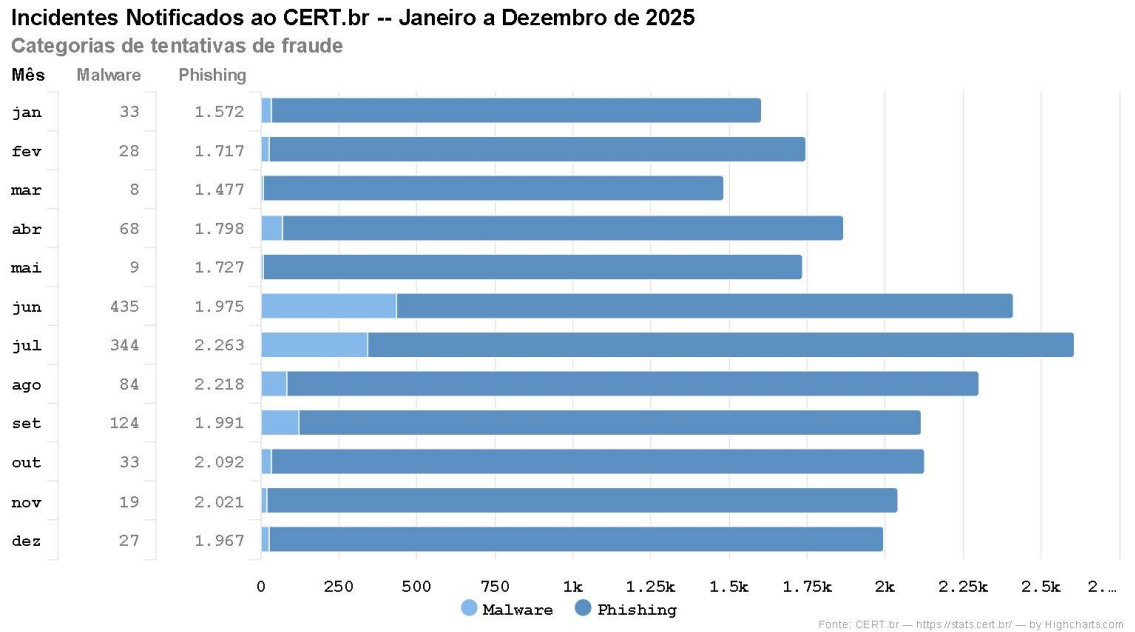
A interpretação analítica desses dados indica que não se trata apenas de volume de incidentes, mas da manifestação de uma vantagem estratégica inerente ao modelo de ataque. Ao priorizar a exploração dos processos decisórios humanos, em detrimento de

abordagens técnicas mais sofisticadas, o agente malicioso é capaz de reduzir custos operacionais, ampliar o alcance das investidas e contornar sucessivas camadas de defesa tecnológica.

Nesse sentido, o problema de pesquisa desloca-se de uma abordagem centrada em “como bloquear mensagens” para uma perspectiva orientada à mitigação comportamental, isto é, à redução da probabilidade de conversão da mensagem em ação. Tal dinâmica torna-se particularmente relevante em contextos nos quais a comunicação se ancora em sinais persuasivos, como autoridade e urgência, que tendem a induzir respostas automáticas e pouco reflexivas.

No contexto brasileiro, a título de ilustração contextual, os dados secundários do CERT.br (2025), compilados nas Figuras 1 e 2, indicam a recorrência do fenômeno e sua distribuição setorial, sem evidenciar associação consistente com ambientes de alta institucionalidade. Embora o *phishing* predomine sobre *malware* no período analisado, a presença do varejo entre os principais alvos relativiza essa relação. Assim, as interpretações devem ser tratadas com cautela, dada a ausência de evidência estatística que comprove padrões estruturais ou causais.

Figura 1 – Tentativas de fraude por categoria em 2025.

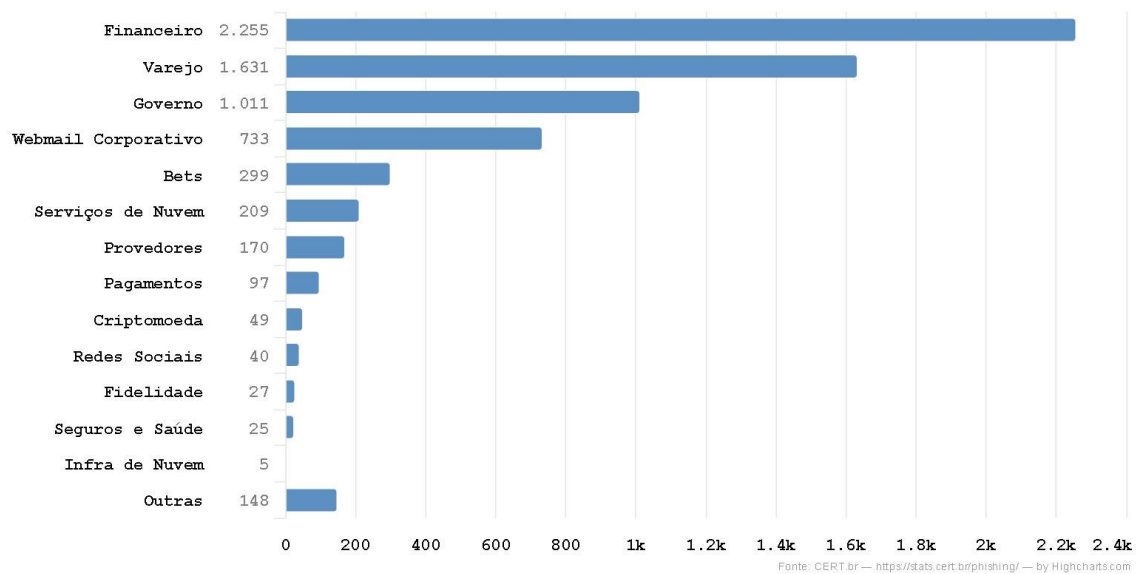


Fonte: Elaborado pelo autor com base em dados do CERT.br, referentes ao ano de 2025, consultados na plataforma do CERT.br em junho de 2026. Foram considerados registros classificados como *phishing* e *malware* no período de janeiro a dezembro de 2025.

Figura 2 – Distribuição de páginas falsas por setor em 2025.

Páginas falsas - Janeiro a Dezembro de 2025

Totais por categoria



Fonte: Elaborado pelo autor com base em dados do CERT.br, extraídos em 2026. Foram considerados registros classificados como páginas falsas por setor no período de janeiro a dezembro de 2025.

Essa distribuição é coerente com achados de Khadka et al. (2024), que associam maior eficácia da autoridade a contextos de normas rígidas e hierarquias mais marcadas.

Contudo, os dados do CERT.br possuem caráter meramente descritivo e contextual, não permitindo inferir relações causais ou estruturais entre setores específicos e a amplificação do viés de autoridade (manifestado pela obediência à autoridade). A compreensão desses mecanismos decorre da interpretação dos estudos analisados nesta revisão.

A emergência da tecnologia de *deepfakes* intensifica esse quadro porque amplia a capacidade do atacante de simular autoridade de modo hiper-realista, reduzindo ainda mais o espaço de dúvida. Folorunsho e Boamah (2025) discutem o potencial de uso indevido de mídias sintéticas em fraudes e ameaças, enquanto Kaushik et al. (2025) e Jampani (2025) apresentam como voz/imagem de executivos e figuras confiáveis podem ser replicadas para induzir transferências e liberações indevidas.

Analiticamente, *deepfakes* não são apenas “sofisticação do disfarce”: eles elevam a legitimidade perceptiva (voz/rosto) e reforçam o atalho “parece autêntico”, deslocando a validação do campo racional para o sensorial (Kaushik et al., 2025). Consequentemente, a mitigação não pode depender apenas de intuição individual; ela

precisa de processos e cultura que interrompam a obediência automática e instituem verificação sistemática (Sasse; Rashid, 2021).

3. Trabalhos correlatos

Diferentemente da seção anterior, que apresentou os fundamentos conceituais necessários para compreender a relação entre *phishing*, persuasão e viés de autoridade, esta seção tem caráter comparativo. Seu objetivo é analisar criticamente os estudos mais relevantes da literatura recente, identificando convergências, limitações metodológicas e lacunas de pesquisa que justificam a abordagem adotada neste trabalho.

Os trabalhos analisados, a seguir, demonstram que a autoridade aparece de forma consistente na literatura sobre *phishing*, mas, em geral, é tratada como um entre diversos princípios persuasivos, sem receber aprofundamento específico como mecanismo central de decisão. Nesse sentido, observa-se uma lacuna na integração entre fundamentos clássicos da psicologia social, especialmente a teoria da obediência, e sua aplicação em contextos contemporâneos de engenharia social.

Este artigo diferencia-se ao posicionar a autoridade como eixo estruturante da análise, articulando seus fundamentos psicológicos com formas modernas de exploração, como ataques BEC, fraude do CEO, *whaling* e o uso de *deepfakes*. Dessa forma, busca-se não apenas descrever a prevalência do fenômeno, mas explicar o mecanismo pelo qual a autoridade reduz o julgamento crítico e condiciona a ação da vítima em ambientes organizacionais.

A literatura recente tem dedicado esforços para mitigar o problema da suscetibilidade humana a ataques de *phishing* e engenharia social. Contudo, as abordagens divergem quanto ao método e foco de análise. Enquanto alguns trabalhos adotam modelos preditivos ou realizam levantamentos amplos sobre vulnerabilidades gerais, o presente estudo diferencia-se por adotar uma lente metodológica específica: o isolamento do viés de autoridade como eixo estruturante. A seguir, são analisados estudos que buscam resolver o problema do fator humano sob diferentes metodologias, evidenciando a lacuna que esta pesquisa visa preencher.

Siddiqi et al. (2022) realizam uma revisão bibliográfica da psicologia por trás da engenharia social, argumentando que os ataques contornam defesas técnicas ao explorar fragilidades cognitivas. Embora essencial para mapear o problema, o método adotado por esses autores foca na identificação ampla de diversas vulnerabilidades genéricas. Nesse aspecto, o presente trabalho diferencia-se dessa abordagem ao não realizar um

mapeamento geral, mas sim ao aprofundar verticalmente um único mecanismo: o viés de autoridade e sua instrumentalização em ataques corporativos.

De forma complementar, Yang et al. (2022) focam nos processos cognitivos que levam à suscetibilidade ao *phishing*, valendo-se de modelos preditivos. Os autores evidenciam que os atacantes desenham estímulos propositais para inibir o processamento mental sistemático da vítima, forçando respostas irreflexivas e automáticas. Essa constatação é vital para entender como a restrição de tempo favorece a tomada de decisão baseada em atalhos cognitivos.

Avançando para as estratégias de persuasão, a revisão da literatura conduzida por Khadka et al. (2024) investiga como as heurísticas induzem a ação do usuário mesmo diante de proteções tecnológicas ativas. Os resultados desse levantamento indicam a predominância do uso dos princípios de autoridade e prova social, demonstrando que tais táticas são ainda mais bem-sucedidas em ambientes regidos por normas e hierarquias marcadas, como os setores governamentais e financeiros.

No que tange à aplicação prática, Lawall (2025) realiza uma análise comparativa fundamentada nos princípios de influência social descritos por Cialdini. O autor demonstra como os agentes maliciosos articulam a simulação de autoridade com a urgência para elevar as taxas de conversão dos ataques, relacionando diretamente esse mecanismo a golpes de alto nível corporativo, a exemplo do *Business Email Compromise* (BEC) e da fraude do CEO.

Por fim, refletindo a evolução tecnológica das ameaças corporativas, os estudos empíricos de Kaushik et al. (2025) e Jampani (2025) introduzem o forte impacto das mídias sintéticas (*deepfakes*) na engenharia social. Estes pesquisadores descrevem como a simulação hiper-realista da voz e do rosto de executivos eleva significativamente a persuasão, induzindo transferências financeiras indevidas. Tais estudos evidenciam que essa tecnologia atua como um amplificador do viés de autoridade, deslocando a legitimação do campo racional para o sensorial.

Em síntese, embora os trabalhos correlatos busquem solucionar o problema da manipulação humana em ciberataques, suas metodologias tratam a autoridade apenas como um atributo secundário ou como uma variável em modelos preditivos genéricos (como em Yang et al., 2022). O método diferenciador desta pesquisa reside na sua estrutura analítica: diferentemente dos levantamentos citados, este trabalho articula diretamente o mecanismo psicológico de obediência com a evolução para táticas baseadas em tecnologias hiper-realistas (*deepfakes*), culminando não apenas na

descrição do problema, mas na formulação de resultados voltados à mitigação psico-organizacional.

Quadro 1 – Comparativo de Trabalhos Correlatos e Lacuna do Estudo

Autor(es) e Ano	Foco da Pesquisa	Metodologia	Principal Contribuição	Limitação Apontada	Relação com este Artigo (Diferencial)
Siddiqi et al. (2022)	Psicologia por trás da engenharia social e análise das contramedidas existentes.	Revisão bibliográfica assistida por computador e mapeamento conceitual.	Demonstra como ataques de engenharia social contornam defesas técnicas ao explorarem fragilidades cognitivas fundamentais.	Mapeamento abrangente de diversas vulnerabilidades sem aprofundamento vertical em mecanismos específicos de persuasão.	Este artigo diferencia-se ao isolar e aprofundar verticalmente o viés de autoridade e a obediência cega sob a lente de Milgram.
Wang & Lutchkus (2023)	Táticas e gatilhos psicológicos utilizados em e-mails fraudulentos de phishing.	Análise qualitativa e quantitativa de literatura científica e relatórios do setor.	Identifica os principais gatilhos comportamentais (como desatenção) e quantifica o peso do phishing como vetor inicial de intrusão.	Trata as táticas psicológicas de forma ampla e descritiva, sem focar na estruturação hierárquica e na cultura de conformidade organizacional.	Este estudo utiliza os dados estatísticos deles para fundamentar a relevância do tema e foca estritamente na dinâmica da hierarquia corporativa.
Yang et al. (2022)	Predição de suscetibilidade individual ao phishing usando características multidimensionais.	Estudo empírico com modelagem preditiva e testes de simulação comportamental.	Prova empiricamente que a engenharia social opera inibindo o processamento sistemático, forçando decisões sob restrição cognitiva.	Foca estritamente no nível individual e preditivo, sem propor intervenções organizacionais, políticas culturais ou processos institucionais.	Este artigo apoia-se no modelo de processamento sistemático e heurístico de Yang para desenhar diretrizes psico-organizacionais coletivas.
Lawall (2025)	Análise comparativa quantitativa do desempenho dos princípios de Cialdini no phishing.	Estudo empírico quantitativo e simulações com medição de taxas de clique.	Mapeia que a simulação de autoridade, combinada com urgência, apresenta as maiores taxas de conversão de golpes (ex: BEC e fraude do CEO).	Concentra-se na classificação e quantificação de eficácia dos gatilhos de Cialdini, sem articular as bases psicológicas profundas da obediência.	Esta pesquisa avança a partir dos dados de Lawall, integrando a teoria de Milgram para explicar por que a autoridade neutraliza a crítica.
Kaushik et al. (2025) / Jampani (2025)	Impacto do uso adversarial de mídias sintéticas (deepfakes) em fraudes e segurança.	Estudo empírico e exploratório sobre deep learning e ameaças baseadas em IA.	Evidencia que deepfakes de voz/imagem de executivos elevam a persuasão ao deslocarem a legitimação do campo racional para o sensorial.	Foca predominantemente nos desafios tecnológicos de detecção técnica, sem discutir a adaptação de processos de governança interna.	Este estudo aborda os deepfakes como aceleradores do estado de agente (Milgram) e propõe defesas de processo baseadas em fricção planejada

Fonte: Elaborado pelo autor, com base na literatura consultada (2026).

4. Metodologia

Esta pesquisa caracteriza-se como um estudo aplicado de natureza exploratória e qualitativa (Gil, 2002). O procedimento metodológico adotado é a revisão integrativa de literatura (Souza; Silva; Carvalho, 2010), estruturada com base nas etapas propostas por Whittemore e Knafl (2005). A escolha por esta modalidade justifica-se pela sua capacidade de sintetizar de forma rigorosa dados provenientes de pesquisas com diferentes delineamentos metodológicos (empíricos, conceituais e qualitativos), permitindo construir uma compreensão aprofundada e integrada do fenômeno investigado: a exploração do viés de autoridade como um mecanismo de influência em ataques de *phishing* e engenharia social no ambiente corporativo.

As buscas bibliográficas foram executadas sistematicamente entre fevereiro e junho de 2026 nas bases de dados Scopus e Web of Science, recorrendo-se ao Google Acadêmico de forma complementar para identificação de literatura especializada. O recorte temporal para seleção dos estudos compreendeu publicações editadas entre 2022 e 2025, de modo a contemplar evidências científicas recentes sobre *phishing*, engenharia social e vieses cognitivos. Dados estatísticos e relatórios institucionais consultados em 2026 foram utilizados exclusivamente para contextualização do fenômeno analisado. As pesquisas foram aplicadas aos campos de título, resumo e palavras-chave (Title/Abstract/Keywords na Scopus e Topic Search na Web of Science), empregando-se como estratégia de busca a *string* principal: (phishing OR "social engineering") AND ("authority" OR obedience OR persuasion OR "cognitive bias"), adaptada conforme a sintaxe específica de cada base consultada.

Os critérios de inclusão adotados foram: (a) atualidade: artigos publicados entre 2022 e 2025; (b) relevância temática: estudos sobre *phishing*, vieses cognitivos, princípios de persuasão ou comportamento humano na segurança da informação; (c) autoridade científica: periódicos científicos revisados por pares e relatórios institucionais consolidados; e (d) aplicabilidade prática organizacional. Inversamente, foram excluídos: (a) documentos sem revisão por pares, como resumos isolados, editoriais, cartas ao editor ou opiniões (com exceção de obras clássicas e dados oficiais de segurança usados como fontes complementares); e (b) estudos anteriores ao ano de 2022, de modo a garantir a contemporaneidade das ameaças tecnológicas analisadas, excetuando-se as obras clássicas da psicologia social indispensáveis à sustentação conceitual do estudo.

O fluxo de seleção das fontes seguiu de forma adaptada as diretrizes do modelo PRISMA 2020. Inicialmente, a busca nas bases indexadas recuperou 30 registros. Após

a exclusão de 5 documentos duplicados, foram triados 25 registros por meio da leitura de títulos e resumos, o que levou à eliminação de 8 trabalhos que não cumpriam os critérios iniciais. Dos 17 artigos pré-selecionados e avaliados na íntegra, 5 foram excluídos por desalinhamento temático ao escopo específico do estudo (visto que não tratavam diretamente da obediência ou do viés de autoridade). O processo culminou na seleção de 12 estudos contemporâneos que constituíram o corpus científico principal desta revisão.

As fontes complementares compreenderam obras clássicas de psicologia social, referências metodológicas, *frameworks* reconhecidos de cibersegurança, relatórios institucionais e literatura cinzenta pertinente.

A extração e análise dos dados fundamentaram-se em um procedimento estruturado de síntese qualitativa e codificação temática interpretativa (Souza; Silva; Carvalho, 2010). Para organizar as informações, utilizou-se uma matriz de extração contendo o delineamento, objetivos, amostra e principais achados de cada uma das 19 fontes.

Adotando uma abordagem híbrida, os dados foram agrupados por semelhança e codificados em três categorias analíticas estruturantes, alinhadas aos objetivos específicos do estudo: (1) Mecanismos cognitivos e vieses comportamentais, que aborda os determinantes da obediência; (2) Estratégias e evolução dos ataques de *phishing*, que investiga o emprego da autoridade em ataques contemporâneos (incluindo o uso de *deepfakes*); e (3) Implicações organizacionais e estratégias de mitigação, voltada para diretrizes psico-organizacionais de defesa.

Quadro 2 – Estrutura da matriz de extração dos estudos

Campo	Descrição
Autor(es) e ano	Identificação da publicação
Objetivo do estudo	Problema e finalidade da pesquisa
Método	Tipo de estudo, abordagem e procedimentos
Amostra / Corpus	Participantes, documentos ou bases analisadas
Principais achados	Resultados relacionados ao phishing e aos fatores humanos
Relação com o viés de autoridade	Evidências sobre autoridade, obediência ou persuasão
Implicações para mitigação	Recomendações organizacionais e de segurança

Fonte: Elaborado pelo autor.

A análise comparativa dos estudos foi realizada a partir desses campos, permitindo identificar recorrências, divergências e contribuições específicas para cada uma das três categorias analíticas definidas neste trabalho.

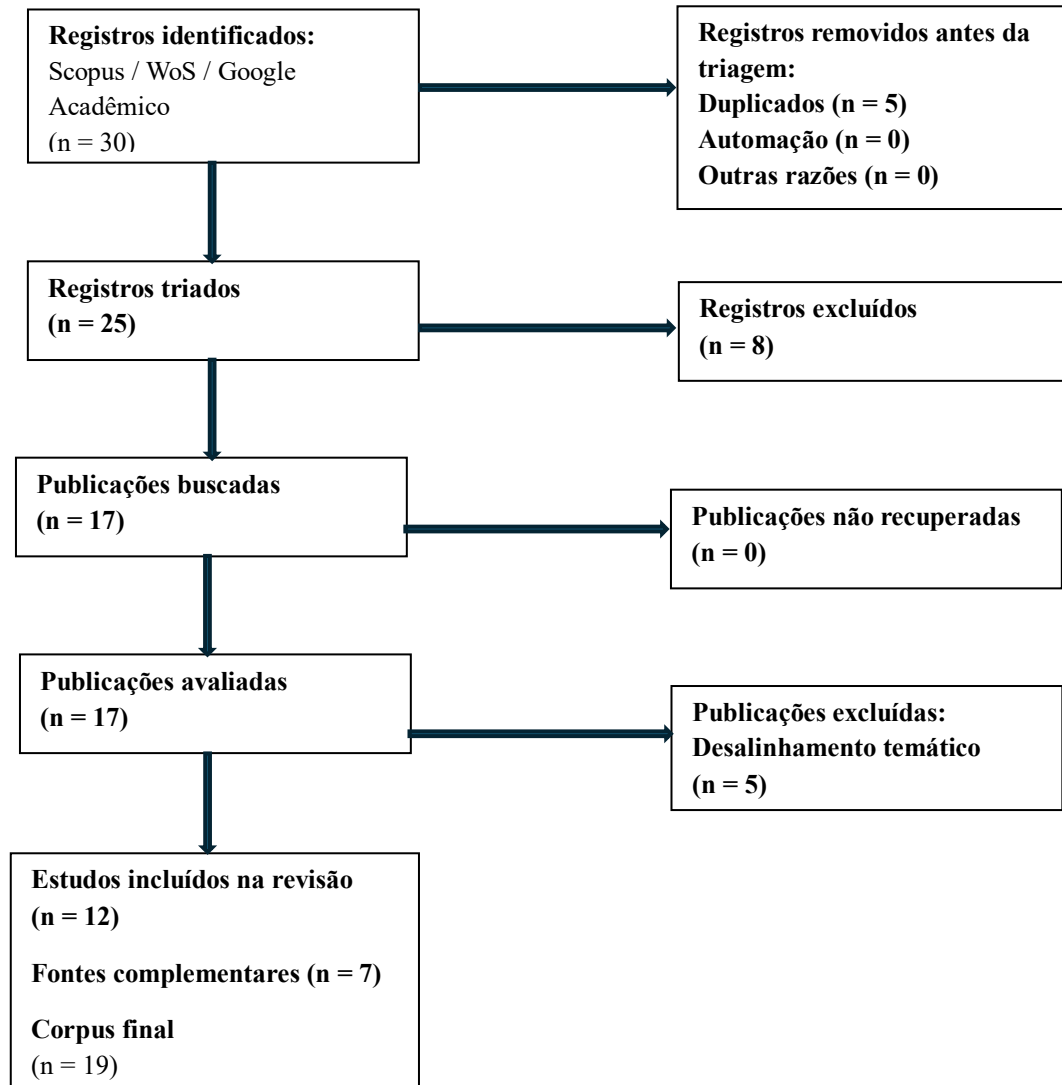


Figura 3. Fluxograma PRISMA 2020 com os números de registros em cada etapa do processo de identificação, triagem, elegibilidade e inclusão.

5. Análise e discussão

Nesta seção, serão apresentados os resultados da revisão integrativa, organizados segundo as três categorias analíticas definidas na metodologia. Para garantir o alinhamento estrutural da pesquisa, a apresentação foi estruturada de modo que cada subseção (5.1 a 5.4) dialogue explicitamente com objetivos específicos diferentes.

Desse modo, as Subseções 5.1 e 5.2 atendem aos objetivos 'a' e 'b', abordando os mecanismos cognitivos e vieses comportamentais inerentes à exploração do viés de autoridade; a Subseção 5.3 relaciona-se ao objetivo 'c', examinando as estratégias e a

evolução das ameaças, com destaque para a amplificação perceptiva promovida pelos *deepfakes*; e, por fim, a Subseção 5.4 responde ao objetivo 'd', consolidando as implicações organizacionais e as estratégias de mitigação. Essa estruturação assegura o encadeamento lógico direto entre a definição do método e a análise dos dados empíricos.

Embora os estudos analisados forneçam evidências relevantes sobre a exploração do viés de autoridade em ataques de *phishing*, seus resultados devem ser interpretados à luz de suas limitações metodológicas. Parte da literatura identificada possui caráter predominantemente conceitual ou de revisão, havendo relativa escassez de investigações empíricas de larga escala voltadas especificamente à mensuração do impacto isolado da autoridade sobre a suscetibilidade dos usuários. Adicionalmente, diferenças de contexto organizacional, setor econômico e recorte geográfico limitam a generalização irrestrita dos achados, recomendando cautela na extrapolação dos resultados para diferentes ambientes institucionais.

5.1 A autoridade como atalho cognitivo

A vulnerabilidade ao *phishing* em contextos corporativos deve ser compreendida como um fenômeno de decisão sob restrição, no qual a vítima raramente dispõe de tempo, motivação e recursos cognitivos para checar a veracidade de cada solicitação recebida.

Nessa condição, a engenharia social opera como um mecanismo de economia cognitiva: em vez de vencer barreiras técnicas, ela força o usuário a decidir rapidamente com base em pistas superficiais de credibilidade. Yang et al. (2022) sustentam que atacantes desenham estímulos para inibir o processamento sistemático e deslocar a avaliação para respostas automáticas, enquanto Siddiqi et al. (2022) destacam que ataques de engenharia social são bem-sucedidos justamente por explorarem traços psicológicos e cognitivos em vez de falhas técnicas.

A implicação é que o *phishing* não “engana” apenas por falsificação, mas por orquestrar condições de julgamento degradado, em que a ação (clicar, responder, autorizar) parece mais eficiente do que verificar. Essa leitura reposiciona o problema de pesquisa: o risco reside no modo como sinais persuasivos capturam a decisão humana, tornando defesas apenas tecnológicas estruturalmente insuficientes (Yang et al., 2022; Siddiqi et al., 2022).

Nesse contexto, os princípios de persuasão funcionam como heurísticas de legitimidade: pistas sociais substituem verificações racionais quando a decisão precisa ser rápida.

Lawall (2025) descreve a aplicação adversarial dos princípios de influência social de Cialdini (2021), mostrando que atacantes combinam sinais que aumentam credibilidade e urgência para elevar a taxa de conversão do golpe. Khadka et al. (2024), por sua vez, sintetizam evidências empíricas indicando que autoridade e prova social aparecem de modo recorrente e com alto desempenho, especialmente em ataques direcionados.

A interpretação crítica é que esses princípios não atuam isoladamente: sua eficácia depende do “encaixe” entre o sinal persuasivo e o ambiente de decisão. Em organizações hierárquicas, por exemplo, obedecer rapidamente pode ser um comportamento funcional (evita atrito, cumpre prazos, preserva relações), o que torna a autoridade um atalho especialmente eficiente. Portanto, compreender a persuasão no *phishing* exige conectar evidência psicológica e estrutura organizacional: o ataque explora heurísticas cognitivas, mas também normas institucionais que punem a hesitação e premiam a responsividade.

Entre os princípios persuasivos, a autoridade se destaca não apenas por frequência, mas por mecanismo: ela reduz o julgamento crítico ao transformar a solicitação em “comando legítimo” dentro da cultura organizacional. Ao simular executivos, departamentos internos, instituições financeiras ou órgãos governamentais, o atacante reconfigura a mensagem como obrigação, deslocando a decisão do campo do “avaliar risco” para o campo do “cumprir procedimento”.

Esse é o ponto de transição para a subseção seguinte: se a autoridade opera como mecanismo de obediência automática, então o núcleo analítico do problema é explicar como essa obediência se forma, por que ela persiste mesmo diante de sinais de alerta e como ela é operacionalizada em golpes como BEC, fraude do CEO e *whaling* (Lawall, 2025; Khadka et al., 2024).

O quadro 2 indica que a exploração da autoridade em ataques de *phishing* não depende de um único recurso persuasivo, mas de uma combinação de fatores situacionais, incluindo hierarquia, urgência, aparência institucional e familiaridade contextual. Esses elementos atuam conjuntamente para reduzir o julgamento crítico e induzir respostas automáticas, especialmente em ambientes organizacionais.

Nesse sentido, observa-se que medidas meramente técnicas são insuficientes quando não acompanhadas de protocolos organizacionais que legitimem a verificação e o questionamento de solicitações suspeitas, evidenciando a necessidade de abordagens psico-organizacionais integradas.

Essa interpretação analítica responde diretamente ao objetivo específico 'b' da pesquisa, ao demonstrar como os fundamentos psicológicos da obediência condicionam e degradam os processos de tomada de decisão sob pressão.

Quadro 3 – Síntese dos mecanismos de exploração da autoridade em ataques de *phishing*.

Elemento explorado pelo atacante	Exemplo em <i>phishing</i>	Efeito psicológico esperado	Medida de mitigação
Cargo ou hierarquia	E-mail falso de diretor ou CEO	Obediência e medo de contrariar	Verificação por canal independente
Urgência	“Aprovar pagamento até hoje”	Redução da reflexão crítica	Política de pausa para decisões sensíveis
Linguagem institucional	Logotipo, assinatura institucional e domínio semelhante	Percepção de legitimidade	Treinamento com exemplos reais, checagem de domínio, validação de identidade, canal oficial de comunicação contextualizadas.
Pressão normativa	“Todos já realizaram a atualização”	Conformidade social	Comunicação interna padronizada
Voz ou imagem sintética	<i>Deepfake</i> de executivo	Aumento da confiança aparente	Confirmação multifator para ordens críticas

Fonte: Elaboração própria.

5.2 A simulação de autoridade em ataques de *phishing*

Enquanto a subseção anterior (5.1) dedicou-se a explicar o mecanismo cognitivo do viés de autoridade, o foco desta subseção é examinar como esse viés é operacionalizado em ataques cibernéticos concretos, a exemplo do *Business Email Compromise* (BEC), da fraude do CEO e do *whaling*.

Nesses cenários direcionados, a autoridade opera no *phishing* como um mecanismo de redução do julgamento crítico: em vez de avaliar evidências, a vítima aceita a legitimidade do emissor como atalho para decidir.

Cialdini (2021) descreve o viés de autoridade como uma resposta automatizada socialmente aprendida, útil para a coordenação social, mas perigosa quando símbolos de

autoridade podem ser falsificados. Aronson et al. (2015) aprofundam o mecanismo ao apontar que a obediência se sustenta por influências normativas e informativas, autojustificação e, sobretudo, pela tendência a deslocar responsabilidade pessoal quando a ordem parece vir de uma fonte legítima.

Em termos analíticos, o efeito central é que a autoridade redefine a tarefa da vítima: *não é mais verificar, mas executar*. Isso explica por que o risco do *phishing* não se limita à falsificação técnica do e-mail; ele está na capacidade de produzir legitimidade psicológica suficiente para desencadear conformidade.

Os experimentos de Milgram (1983) oferecem um arcabouço explicativo especialmente útil para compreender ataques corporativos de alto impacto. O ponto relevante para este estudo não é a narrativa histórica do experimento, mas sua lógica: indivíduos podem obedecer a comandos percebidos como legítimos mesmo quando há desconforto, especialmente quando entram em um estado de agente, isto é, quando transferem a responsabilidade moral para a autoridade.

Aronson et al. (2015) reforçam que a dinâmica de obediência é intensificada quando há ritmo acelerado, ambiguidade situacional e comprometimentos incrementais que exigem racionalizações sucessivas. No *phishing* corporativo, particularmente em BEC e fraude do CEO, observa-se mecanismo semelhante: mensagens simulam superioridade hierárquica, criam urgência e induzem microações iniciais (responder, confirmar, abrir documento), que aumentam a probabilidade de ações subsequentes mais críticas (fornecer credenciais, autorizar pagamento). Assim, o ataque não depende de uma única decisão “grande”, mas de uma sequência de pequenas concessões que reduz a chance de ruptura do comportamento.

Sob a perspectiva da teoria da dissonância cognitiva, cada comprometimento inicial tende a aumentar a necessidade de coerência entre comportamento e crenças. Dessa forma, ao aceitar uma primeira solicitação aparentemente legítima, o indivíduo passa a racionalizar suas decisões subsequentes, reduzindo a probabilidade de interromper o processo mesmo diante de sinais de alerta. Em vez de reconhecer a possibilidade de ter cometido um erro, a vítima tende a reinterpretar informações ambíguas de modo compatível com a decisão já tomada, preservando a consistência psicológica de suas ações (Aronson et al., 2015).

Essa leitura ajuda a explicar porque golpes baseados em autoridade são mais eficazes em ambientes altamente normatizados e hierárquicos: a obediência rápida

costuma ser reforçada pela cultura institucional, enquanto questionar pode ser percebido como insubordinação ou atraso do trabalho.

Lawall (2025) descreve a predominância da autoridade em ataques direcionados, e Khadka et al. (2024) indicam que sinais de autoridade se associam a maiores taxas de comprometimento em setores como finanças e administração pública.

A interpretação crítica é que a vulnerabilidade não é apenas individual; ela é sistêmica: quando a estrutura organizacional valoriza velocidade e conformidade, o atacante consegue terceirizar a persuasão para a própria hierarquia. Logo, compreender o viés de autoridade no *phishing* implica reconhecer que a decisão da vítima é coproduzida por mecanismos cognitivos (heurísticas) e por incentivos organizacionais (normas, reputação, medo de consequências).

A partir da consolidação entre as evidências empíricas extraídas da literatura (como os achados de Lawall e Khadka et al.) e a interpretação analítica deste estudo sobre a dimensão sistêmica da vulnerabilidade, o fechamento desta subseção atende diretamente ao objetivo de identificar como sinais de autoridade são empregados em ataques como *spear phishing*, *whaling*, *Business Email Compromise* (BEC) e golpes baseados em *deepfakes*. Fica demonstrado, assim, como os gatilhos psicológicos da obediência são concretamente operacionalizados para forçar ações automatizadas em ataques direcionados, a exemplo do *Business Email Compromise* (BEC), da fraude do CEO e do *whaling*.

5.3 Deepfakes e amplificação da autoridade percebida

A emergência da tecnologia de *deepfakes* intensifica esse mecanismo ao elevar o realismo perceptivo da autoridade. Folorunsho e Boamah (2025) discutem a capacidade de conteúdos sintéticos hiper-realistas ampliarem ameaças e fraudes, e Kaushik et al. (2025) e Jampani (2025) descrevem como voz e imagem de executivos podem ser simuladas para induzir transferências, liberações de acesso ou aprovações indevidas.

O ponto analítico é que *deepfakes* não apenas tornam a falsificação mais convincente; eles reforçam a heurística “parece autêntico”, reduzindo ainda mais o espaço para dúvida e verificação. Isso cria um desafio direto para a mitigação: se a legitimidade passa a ser sensorial, estratégias baseadas apenas em atenção individual ou intuição tornam-se frágeis. Assim, a mitigação precisa responder ao mecanismo, isto é, criar barreiras processuais e culturais que interrompam a obediência automática.

Nesse cenário, cumpre explicitar que o impacto dos *deepfakes* vai além de apenas aumentar a credibilidade do emissor. Essa tecnologia reduz significativamente a eficácia das pistas tradicionais de detecção de fraudes, como a checagem de domínios de e-mail, formatação ou erros ortográficos. Diante de uma simulação hiper-realista, a vítima frequentemente abandona a análise técnica e passa a confiar quase de forma exclusiva em sinais sensoriais imediatos, como a voz, os traços do rosto e a familiaridade aparente da liderança simulada.

Ao explorar essa fronteira tecnológica, esta subseção atende diretamente ao objetivo de identificar como sinais de autoridade são empregados em ataques como *spear phishing*, *whaling*, *Business Email Compromise* (BEC) e golpes baseados em *deepfakes*. Contudo, para manter o rigor metodológico, é preciso separar evidência de tendência. Enquanto a literatura atual (como Kaushik et al. e Jampani) comprova a viabilidade técnica e a ocorrência pontual de fraudes simulando voz e imagem, a tese de que os *deepfakes* poderão alterar de forma relevante os mecanismos tradicionais de validação de autoridade nas organizações consiste em uma projeção analítica deste estudo. Trata-se, portanto, da antecipação de um cenário altamente provável, e não de um dado empírico já massificado.

5.4 Estratégias psico-organizacionais de mitigação

Se o *phishing* baseado em autoridade funciona reduzindo o julgamento crítico e deslocando a responsabilidade para uma fonte percebida como legítima, então a mitigação eficaz deve introduzir fricção deliberada antes de ações sensíveis, não apenas “conscientização” genérica.

Siddiqi et al. (2022) indicam que treinamento e políticas organizacionais são fundamentais, mas que soluções eficazes precisam considerar fatores psicológicos e linguísticos, inclusive em modelos de detecção automatizada. Em paralelo, Cialdini (2021) argumenta que a autoridade não deve ser rejeitada de forma absoluta (pois geralmente é funcional), mas que é essencial desenvolver critérios práticos para distinguir autoridade legítima de autoridade simulada.

A interpretação crítica é que o objetivo do controle não é desconfiar de tudo, mas reduzir a probabilidade de obediência automática quando a ação envolve risco. Portanto, a mitigação deve ser desenhada como resposta direta ao mecanismo: interromper o modo automático, restabelecer o julgamento e deslocar a validação do indivíduo para o processo.

Uma implicação organizacional central é que questionar a autoridade precisa deixar de ser um comportamento excepcional e passar a ser um comportamento legitimado e esperado quando há risco informacional.

Embora o alerta seminal de Mitnick (2003) já apontasse para a necessidade de treinar funcionários a fim de verificarem solicitações suspeitas independentemente da hierarquia, é a literatura acadêmica contemporânea sobre treinamentos de conscientização em segurança que valida empiricamente essa premissa. Estudos focados em contramedidas organizacionais, a exemplo de Siddiqi et al. (2022), sustentam que essa prática só se efetiva quando há apoio explícito da alta gestão. Essa exigência não é apenas cultural, ela é operacional. Sem respaldo institucional, o funcionário percebe que o custo social de checar (parecer lento, desconfiado ou insubordinado) pode ser maior do que o custo imediato de obedecer, o que favorece o atacante.

Logo, protocolos formais, como validação fora de banda, dupla aprovação para pagamentos, confirmação por canal alternativo e segregação de funções, funcionam como substitutos institucionais da coragem individual, reduzindo o espaço para coerção hierárquica e para a exploração do viés de autoridade.

Ao mesmo tempo, torna-se necessário tratar o problema da fadiga de segurança como parte do mecanismo, e não como detalhe periférico. No capítulo sobre Fatores Humanos do CyBOK, Sasse e Rashid (2021) observam que excesso de demandas e comunicações de segurança esgota recursos cognitivos e incentiva atalhos, aumentando comportamentos de risco e a tendência a burlar regras. Analiticamente, isso se conecta diretamente à autoridade: sob sobrecarga, obedecer a uma solicitação “de cima” torna-se o caminho de menor esforço, e a verificação, que exige tempo e energia, tende a ser abandonada. Portanto, mitigação efetiva requer não só mais treinamento, mas melhor desenho de treinamento e processos: mensagens mais enxutas, simulações realistas e repetidas, rotinas simples de checagem, e redução de atrito para reportar suspeitas. Quando a ação segura é a mais fácil, a organização reduz o espaço de atuação do atacante.

A literatura sugere que o ataque explora um mecanismo composto por legitimidade percebida, urgência e sequenciamento de microações (Cialdini, 2021; Aronson et al., 2015; Lawall, 2025), potencializado por realismo tecnológico (Kaushik et al., 2025; Jampani, 2025).

Em resposta, a organização precisa alinhar (i) cultura e liderança (legitimar a checagem), (ii) processos (dupla validação, confirmação fora de banda, segregação de

funções) e (iii) tecnologia (filtros, detecção, monitoramento), além de reduzir a fadiga de segurança (Sasse; Rashid, 2021). Assim, o combate ao *phishing* não é apenas treinar melhor: é redesenhar o ambiente decisório para que a autoridade não opere como gatilho automático de conformidade, mas como sinal a ser verificado sob protocolos claros.

Em síntese, a mitigação efetiva do *phishing* baseado no viés de autoridade exige que as organizações adotem um conjunto claro de diretrizes psico-organizacionais. De forma objetiva, recomenda-se: legitimar o questionamento de solicitações atípicas independentemente da hierarquia do emissor; instituir processos de validação fora de banda; exigir dupla aprovação para a execução de ações sensíveis; reduzir a fadiga de segurança por meio de comunicações enxutas; e, por fim, desenhar treinamentos práticos baseados em cenários de autoridade simulada.

Ao consolidar essas estratégias, esta subseção responde diretamente ao objetivo específico 'd'. Para fins de rigor metodológico, cumpre diferenciar a origem das diretrizes: enquanto os treinamentos contínuos e a mitigação da fadiga de segurança são recomendações consolidadas na literatura, a proposição de instituir 'fricção' processual para atuar como 'substituto institucional da coragem individual' reflete a contribuição analítica deste estudo.

Conclui-se, assim, que mitigar o *phishing* direcionado exige não apenas treinar o usuário, mas redesenhar a própria arquitetura de decisão corporativa.

6. Conclusão e trabalhos futuros

O principal achado desta pesquisa vai além da ideia inicial de que o *phishing* apenas explora falhas humanas. A principal contribuição analítica do estudo é mostrar que o atacante usa a própria cultura da empresa a seu favor. Ao se aproveitar da pressão corporativa por rapidez e obediência, o cibercriminoso transforma a hierarquia da organização em um importante vetor de convencimento, o que explica por que defesas puramente tecnológicas não são suficientes.

A partir dessa perspectiva, entende-se que o objetivo geral de analisar a exploração do viés de autoridade e discutir suas mitigações foi alcançado, assim como os objetivos específicos delineados na introdução do estudo. Primeiramente, (a) caracterizou-se o *phishing* moderno como a principal prática de engenharia social em ambientes digitais, que desloca intencionalmente o ponto de falha da infraestrutura técnica para a tomada de decisão do usuário. Em segundo lugar, (b) o exame dos

fundamentos psicológicos evidenciou que o viés de autoridade inibe o processamento crítico da informação, forçando a vítima a agir de modo automático, dinâmica que se agrava sob condições de urgência e esgotamento cognitivo.

Avançando na análise estrutural das ameaças, (c) constatou-se que os sinais persuasivos de autoridade são o motor de ataques direcionados, como *spear phishing*, *whaling* e *Business Email Compromise* (BEC), cujo risco é severamente amplificado pelo advento dos *deepfakes*, que transferem a validação de legitimidade do campo racional para o sensorial. Por fim, (d) o estudo propõe que a mitigação efetiva demanda diretrizes psico-organizacionais: as corporações precisam redesenhar seus processos para combater a fadiga de segurança e, sobretudo, instaurar uma cultura que legitime e encoraje a verificação de solicitações atípicas, independentemente da posição hierárquica do emissor.

Ressalta-se que os resultados deste estudo possuem caráter interpretativo e estão sujeitos a limitações metodológicas que devem ser consideradas. Em primeiro lugar, por se apoiar exclusivamente em uma revisão de literatura, a pesquisa não inclui coleta de dados primários ou validação empírica direta dos mecanismos psicológicos em ambientes corporativos reais. Em segundo lugar, a restrição do protocolo de busca a bases de dados específicas pode ter excluído trabalhos relevantes não indexados nessas plataformas, além de expor a análise a um possível viés de publicação, cenário no qual estudos com resultados estatisticamente significativos têm maior probabilidade de serem publicados e selecionados. Por fim, as fontes analisadas apresentam limitações geográficas, ancorando-se majoritariamente em métricas do cenário brasileiro (por meio dos relatórios do CERT.br) e em estudos ocidentais sobre comportamento organizacional, o que exige cautela na generalização irrestrita desses achados para diferentes contextos culturais.

Assim, como limitação, destaca-se que este estudo possui natureza exclusivamente bibliográfica e qualitativa, baseada em fontes secundárias e sem coleta de dados empíricos, estando sujeito a possíveis vieses de seleção e publicação inerentes à estratégia de busca adotada; além disso, as recomendações propostas devem ser interpretadas considerando as especificidades organizacionais, institucionais e socioculturais de cada contexto, não sendo passíveis de aplicação uniforme a todas as organizações.

Como trabalhos futuros, sugere-se a estruturação de treinamentos e simulações de *phishing* focados especificamente em explorar o viés de autoridade, abordando de

forma incisiva a nova fronteira tecnológica dos ataques: o uso de inteligência artificial e *deepfakes* para simular a voz e a imagem de executivos de alto escalão, bem como aprofundar a análise de desempenho de diferentes perfis de usuários frente a essas ameaças hiper-realistas.

7. Referências

ARONSON, Elliot; WILSON, Timothy D.; AKERT, Robin M. **Psicologia social**. 8. ed. Rio de Janeiro: LTC, 2015.

BEZERRA, P.; MILLIAN, L. **Estudo de técnicas de phishing: métodos de ataque e estratégias de defesa**. Trabalho de Conclusão de Curso. São Paulo: Universidade Presbiteriana Mackenzie, 2024. Disponível em: <https://dspace.mackenzie.br/items/00491918-2042-4338-a1a6-dab251f5b297>. Acesso em: 10 jun. 2026.

CERT.br. **Estatísticas mantidas pelo CERT.br**. São Paulo: NIC.br, 2025. Disponível em: <https://stats.cert.br/>. Acesso em: 10 jun. 2026.

CIALDINI, Robert B. **As armas da persuasão**. Rio de Janeiro: Sextante, 2021.

FOLORUNSHO, F.; BOAMAH, B. Deepfake technology and its impact: ethical considerations, societal disruptions, and security threats in AI-generated media. **International Journal of Information Technology Management and Information Systems**, v. 16, n. 1, 2025. Disponível em: https://iaeme.com/MasterAdmin/Journal_uploads/IJITMIS/VOLUME_16_ISSUE_1/IJITMIS_16_01_076.pdf. Acesso em: 10 jun. 2026.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

JAMPANI, S. Social engineering 2.0: deepfake and deep learning-based cyber-attacks (phishing). **International Journal for Multidisciplinary Research**, v. 7, n. 1, 2025. Disponível em: <https://www.ijfmr.com/papers/2025/1/35527.pdf>. Acesso em: 10 jun. 2026.

KAUSHIK, P. et al. Financial fraud and manipulation: the malicious use of deepfakes in business. In: KAUSHIK, Pooja et al. (org.). **Deepfakes and their impact on business**. Hershey: IGI Global, 2025. cap. 8. Disponível em: <https://www.igi-global.com/chapter/financial-fraud-and-manipulation/364352>. Acesso em: 10 jun. 2026.

KHADKA, K. et al. **A survey on the principles of persuasion as a social engineering strategy in phishing**. Preprint, 2024. Disponível em: <https://researchprofiles.canberra.edu.au/en/publications/a-survey-on-the-principles-of-persuasion-as-a-social-engineering-/>. Acesso em: 10 jun. 2026.

LAWALL, A. Quantifying persuasion: a comparative analysis of Cialdini's principles in phishing attacks. In: SECURWARE 2025, 2025. **Anais [...]**. [S. l.]: [s. n.], 2025. p. 180-184. Disponível em:

https://personales.upv.es/thinkmind/SECURWARE/SECURWARE_2025/securware_2025_1_180_30071.html. Acesso em: 10 jun. 2026.

MILGRAM, Stanley. **Obediência à autoridade: uma visão experimental**. Tradução de Luiz Orlando Coutinho Lemos. Rio de Janeiro: F. Alves, 1983. Obra original publicada em 1974.

MITNICK, Kevin D.; SIMON, William L. **A arte de enganar: ataques de hackers: controlando o fator humano na segurança da informação**. São Paulo: Pearson Education do Brasil, 2003.

SASSE, M. A.; RASHID, A. Human factors. In: RASHID, A. et al. (ed.). **The Cyber Security Body of Knowledge (CyBOK)**. Versão 1.1.0. Bristol: University of Bristol, 2021. Disponível em: <https://www.cybok.org/>. Acesso em: 10 jun. 2026.

SCHMITT, M.; FLÉCHAIS, I. Digital deception: generative artificial intelligence in social engineering and phishing. **Artificial Intelligence Review**, v. 57, art. 324, 2024. DOI: 10.1007/s10462-024-10973-2. Disponível em: <https://link.springer.com/article/10.1007/s10462-024-10973-2>. Acesso em: 10 jun. 2026.

SIDDIQI, M. et al. A study on the psychology of social engineering-based cyberattacks and existing countermeasures. **Applied Sciences**, v. 12, n. 12, art. 6042, 2022. DOI: 10.3390/app12126042. Disponível em: <https://www.mdpi.com/2076-3417/12/12/6042>. Acesso em: 10 jun. 2026.

SOUZA, Marcela Tavares de; SILVA, Michelly Dias da; CARVALHO, Rachel de. Revisão integrativa: o que é e como fazer. **Einstein**, São Paulo, v. 8, n. 1, p. 102-106, 2010.

WANG, P.; LUTCHKUS, P. Psychological tactics of phishing emails. **Issues in Information Systems**, v. 24, n. 2, p. 71-83, 2023. Disponível em: https://iacis.org/iis/2023/2_iis_2023_71-83.pdf. Acesso em: 10 jun. 2026.

WHITTEMORE, R.; KNAFL, K. The integrative review: updated methodology. **Journal of Advanced Nursing**, v. 52, n. 5, p. 546-553, 2005.

YANG, R. et al. Predicting user susceptibility to phishing based on multidimensional features. **Frontiers in Psychology**, v. 12, art. 782888, 2022. DOI: 10.3389/fpsyg.2021.782888. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8786481/>. Acesso em: 10 jun. 2026.