Construção e avaliação de uma matriz de risco para identificação de phishing em mensagens simuladas

Ali Osmar Chehade[1], **André Ricardo Zavan**[1]

[1]Instituto Federal do Paraná (IFPR) – Campus Paranavaí

ali.osmar.chehade@gmail.com, andre.zavan@ifpr.edu.br

Abstract. *The growing use of messaging apps has increased phishing attacks, which exploit social engineering to induce users to share personal data or perform risky actions. This study proposes a simple risk matrix to assist in identifying potentially fraudulent messages. The research involved developing the matrix, creating simulated messages, and applying an online questionnaire to assess participants' risk perception, intended actions, and confidence. The study presents the construction and evaluation of the matrix based on simulated messages and participant responses. The results seek to contribute to greater awareness and security when dealing with suspicious messages.*

Resumo. *O crescimento do uso de aplicativos de mensagens tem aumentado os golpes de phishing, que utilizam engenharia social para induzir usuários a fornecer dados ou realizar ações de risco. Este trabalho propõe uma matriz de risco simples para auxiliar na identificação de mensagens potencialmente fraudulentas. A pesquisa envolveu a construção da matriz, a criação de mensagens simuladas e a aplicação de um questionário para avaliar a percepção de risco, as ações pretendidas e a confiança dos participantes. O estudo apresenta a construção e avaliação da matriz com base nas mensagens simuladas e nas respostas obtidas. Os resultados buscam contribuir para maior atenção e segurança diante de mensagens suspeitas.*

1. Introdução

O crescimento do uso de dispositivos móveis e aplicações de comunicação instantânea tem sido acompanhado pelo aumento significativo de golpes digitais, especialmente aqueles disseminados por meio de mensagens em plataformas como WhatsApp e SMS. Entre essas ameaças, destacam-se os ataques de *phishing*, que buscam enganar usuários para a obtenção de informações sensíveis, como dados pessoais e credenciais de acesso [Adu-Manu et al. 2023]

Esses ataques estão diretamente relacionados ao uso de técnicas de engenharia social, as quais exploram fatores psicológicos e comportamentais dos indivíduos para induzi-los a tomar decisões equivocadas. Conforme discutido por [Mitnick and Simon 2003], a manipulação do fator humano constitui um dos principais vetores de ataque em segurança da informação, tornando os usuários elementos críticos na cadeia de proteção contra ameaças digitais.

Perante esse cenário, observa-se a ausência de instrumentos simples, objetivos e estruturados que auxiliem usuários na avaliação do risco de mensagens recebidas em

seu cotidiano. Nesse sentido, este trabalho propõe o desenvolvimento de uma matriz de risco aplicada ao contexto de mensagens digitais, utilizando critérios mensuráveis para classificar o nível de risco associado a possíveis tentativas de *phishing*.

A partir dessa proposta, este trabalho busca responder à seguinte questão de pesquisa: como uma matriz de risco simples e estruturada pode auxiliar usuários comuns na identificação de mensagens de phishing em ambientes digitais?

2. Justificativa

Embora existam diversas pesquisas sobre detecção automatizada de phishing por aprendizado de máquina, heurísticas e análise de características de mensagens [Jain and Gupta 2021], a maior parte exige recursos técnicos nem sempre disponíveis ao usuário final. Estudos sobre conscientização indicam que o treinamento pode reduzir significativamente a suscetibilidade à engenharia social [de Oliveira et al. 2019], mas ainda são escassos os instrumentos simples e quantitativos que auxiliem pessoas comuns a avaliar o risco de mensagens recebidas em aplicativos de comunicação instantânea. Poucos estudos, ainda, investigam simultaneamente a intenção de clicar em links suspeitos, a confiança do participante em sua decisão e a influência das características das mensagens fraudulentas.

A proposta não busca substituir mecanismos automatizados, mas atuar de forma complementar, oferecendo uma estrutura compreensível para apoiar a avaliação feita pelo próprio usuário. Compara-se a classificação de risco das mensagens com a capacidade de identificação, a intenção de agir e a confiança dos participantes.

Sob a perspectiva da Segurança da Informação, compreender como usuários avaliam mensagens potencialmente maliciosas pode contribuir para estratégias educacionais mais eficazes e para ferramentas de apoio à decisão. A presente pesquisa busca preencher essa lacuna ao propor uma matriz de risco baseada em critérios objetivos e mensuráveis, aplicada a mensagens digitais simuladas, e ao analisar a relação entre o nível de risco atribuído às mensagens e o comportamento dos participantes. Espera-se verificar se os critérios da matriz se relacionam com a capacidade de identificação das mensagens, com a intenção de realizar a ação solicitada e com o nível de confiança dos usuários em suas próprias decisões.

3. Fundamentação Teórica

3.1. Engenharia Social

A Engenharia Social refere-se a manipulação psicológica de pessoas, para execução de ações ou para a divulgação de informações confidenciais. A engenharia social usa a influência e a persuasão para enganar as pessoas e convencê-las de que o engenheiro social é alguém que na verdade ele não e, ou pela manipulação. Como resultado, o engenheiro social pode aproveitar-se das pessoas para obter as informações com ou sem o uso da tecnologia. [Mitnick and Simon 2003]

Para [Leonov et al. 2021] um dos elementos mais fracos na proteção de sistemas, são os próprios usuários. Ele cita que, o engenheiro social utiliza de e-mail, telefone

celular e outros meios de comunicação, quando uma das partes involuntariamente divulga informações ou executa algum tipo de ação.

Segundo Hijji e Gulzar (2020 apud [Vilela et al. 2022]), a engenharia social está dividida em quatro tipos: fisico - onde os atacantes tem como objetivo principal acumular informações sobre a vítima, a partir de materiais físicos ; social - o mais utilizado, nele os scammers usam da psicologia e da mente humana para convencer o usuário-alvo a construir um tipo de relacionamento de confiança ; técnico - são utilizadas redes sociais e sites, que normalmente são recheados de informações cruciais aos atacantes, para buscar informações relevantes sobre a vítima; e a sóciotécnica - sendo a mais poderosa de todas, onde são combinados os tipos social e técnico. O engenheiro social leva em consideração fatores como cultura, comportamento humano, tecnologias utilizadas e uma utilização de infraestrutura como metas e valores. Isso aumenta a chance de ataques cibernéticos serem bem-sucedidos

3.2. Phishing

Phishing é uma modalidade de fraude digital em que atacantes se passam por entidades legítimas, como instituições financeiras e empresas, para induzir usuários a fornecer informações confidenciais, realizar pagamentos ou acessar páginas maliciosas, por e-mail, SMS, aplicativos de mensagens e redes sociais [Adu-Manu et al. 2023]. De acordo com [Jain and Gupta 2021], destacam-se o spear phishing (direcionado), o smishing (por SMS), o vishing (por chamadas telefônicas) e os ataques baseados em páginas falsas, que exploram tanto vulnerabilidades técnicas quanto o comportamento humano.

Segundo [Vilela et al. 2022], os atacantes empregam estratégias de persuasão como urgência (“Seu PIX será bloqueado em 24 horas”), autoridade (“Banco do Brasil informa...”), recompensa (“Você recebeu um cashback de 500 reais”), escassez (“Oferta válida apenas hoje”) e curiosidade (“Confira as fotos enviadas para você”), a fim de reduzir a capacidade crítica das vítimas e induzi-las a agir impulsivamente. [Florencio and Herley 2008] argumentam que muitos usuários não têm conhecimento técnico para avaliar a autenticidade das mensagens, de modo que a decisão de confiar depende da percepção individual de risco e da interpretação dos indícios presentes no conteúdo.

Nesse contexto, a efetividade dos ataques de phishing está diretamente associada à exploração de fatores psicológicos e comportamentais, o que torna relevante o desenvolvimento de mecanismos simples e objetivos que auxiliem na identificação de possíveis tentativas de fraude. A proposta deste trabalho contribui nesse sentido, por meio de uma matriz de risco aplicada à avaliação de mensagens digitais.

3.3. Golpes em Mensagens (SMS / WhatsApp)

Os ataques de phishing têm se adaptado às novas formas de comunicação utilizadas pelos usuários, tornando aplicativos de mensagens instantâneas e serviços de mensagens curtas (Short Message Service – SMS) importantes para a disseminação de fraudes digitais. Entre essas modalidades destaca-se o smishing, termo utilizado para designar ataques de phishing realizados por meio de mensagens SMS, cujo objetivo é induzir a vítima a acessar páginas maliciosas, fornecer informações pessoais ou realizar transações financeiras indevidas [Jain and Gupta 2021].

Além das mensagens SMS, aplicativos de comunicação instantânea, como o WhatsApp, têm sido amplamente explorados por criminosos devido à grande quantidade de usuários ativos e à percepção de confiabilidade associada às conversas privadas. Segundo [Vilela et al. 2022], golpes disseminados por esses meios costumam utilizar links encurtados, solicitações de confirmação de dados pessoais, ofertas promocionais falsas e pedidos de transferência de valores via PIX, dificultando a distinção entre mensagens legítimas e fraudulentas.

Dessa forma, considerando a popularização do uso de smartphones e aplicativos de mensagens instantâneas, torna-se relevante investigar estratégias capazes de apoiar a tomada de decisão dos usuários diante de mensagens potencialmente fraudulentas, motivando a proposta da matriz de risco apresentada neste trabalho.

3.4. Tomada de Decisão e Percepção de Risco

A tomada de decisão consiste no processo onde um indivíduo avalia diferentes alternativas e escolhe uma ação com base nas informações disponíveis, em suas experiências e na percepção dos possíveis riscos envolvidos. Nesse contexto da segurança da informação, esse processo é frequentemente realizado sob condições de incerteza, nas quais o usuário precisa decidir rapidamente se uma mensagem recebida é legítima ou representa uma tentativa de fraude.

Segundo [Kahneman 2012], grande parte das decisões humanas não é resultado de um processo analítico detalhado, mas sim de julgamentos intuitivos e automáticos. O autor propõe a existência de dois modos de pensamento: o Sistema 1, caracterizado por respostas rápidas, intuitivas e baseadas em experiências anteriores, e o Sistema 2, responsável pelo raciocínio analítico, deliberado e mais lento. Em situações cotidianas, como a leitura de mensagens em aplicativos de comunicação, os usuários tendem a utilizar predominantemente o Sistema 1, tornando-se mais suscetíveis a erros de julgamento.

No contexto de ataques de *phishing*, mecanismos cognitivos são frequentemente explorados por criminosos para aumentar a credibilidade de mensagens fraudulentas. Elementos como senso de urgência, autoridade, recompensas financeiras ou ameaças de bloqueio induzem o usuário a agir rapidamente, reduzindo o tempo destinado à análise crítica da mensagem e aumentando a probabilidade de decisões impulsivas.

Considerando esse cenário, compreender como os usuários percebem o risco associado às mensagens digitais torna-se fundamental para o desenvolvimento de mecanismos que auxiliem sua tomada de decisão. Nesse sentido, a matriz de risco proposta neste trabalho busca fornecer critérios objetivos para apoiar a avaliação das mensagens, reduzindo a dependência exclusiva da percepção intuitiva do usuário e contribuindo para uma análise mais estruturada diante de possíveis tentativas de *phishing*.

4. Trabalhos Correlatos

As pesquisas sobre phishing podem ser agrupadas em três categorias: detecção automatizada, comportamento do usuário e gestão de riscos. Na primeira, [Jain and Gupta 2021] apresentam uma revisão de técnicas baseadas em características de URLs, páginas web e conteúdo de mensagens.

Na segunda, [de Oliveira et al. 2019] demonstram que conscientização e treinamento reduzem a suscetibilidade à engenharia social, reforçando que o usuário ainda é considerado o elo mais fraco.

Na terceira, [Konzen et al. 2012] exploram modelos baseados na ISO/IEC 27005, com critérios de probabilidade e impacto; essas abordagens, porém, são usadas em contextos organizacionais, e sua aplicação direta ao usuário final na análise de mensagens ainda é pouco explorada. Observa-se, assim, uma lacuna quanto a instrumentos simples, objetivos e aplicáveis ao cotidiano, que este trabalho busca preencher (Tabela 1).

Em síntese, os trabalhos revisados tratam ou da detecção computacional, ou da conscientização, ou da gestão de riscos organizacional, mas raramente combinam critérios mensuráveis de classificação com a análise do comportamento humano diante de mensagens cotidianas. É nessa combinação, voltada à identificação de phishing por usuários comuns, que o presente trabalho se diferencia.

Tabela 1. Comparação entre os trabalhos correlatos e a proposta deste estudo

Autor	Foco do Estudo	Relação com este Trabalho
Jain e Gupta (2021)	Técnicas automatizadas de detecção de phishing baseadas em aprendizado de máquina, heurísticas e análise de similaridade	Apresenta métodos de detecção existentes, porém com foco computacional e automatizado.
de Oliveira et al. (2019)	Conscientização e treinamento de usuários contra phishing	Relaciona phishing ao comportamento do usuário.
Konzen et al. (2012)	Gestão de riscos baseada na norma ISO/IEC 27005	Fundamenta o uso de critérios estruturados para classificação e avaliação de riscos.
Presente Trabalho	Matriz de risco para avaliação de mensagens digitais por usuários comuns	Propõe um instrumento simples e estruturado para apoiar a tomada de decisão frente a possíveis ataques de phishing.

5. Metodologia

5.1. Tipo de Pesquisa e delineamento

Trata-se de pesquisa aplicada, de abordagem predominantemente quantitativa, com análise descritiva, e de caráter exploratório: observa como participantes reagem a mensagens simuladas e se a matriz de risco oferece base útil para interpretar as respostas.

O estudo consistiu em uma única aplicação de formulário estruturado, sem intervenção pedagógica prévia e sem comparação antes/depois. Para cada mensagem foram observados três aspectos: a identificação como golpe ou não golpe, a intenção de realizar ou não a ação sugerida e o nível de confiança na decisão.

5.2. Material utilizado na pesquisa

O material foi composto por dez mensagens simuladas, no formato de comunicações recebidas por WhatsApp e SMS, distribuídas em dois cenários temáticos: entregas e serviços de correspondência e serviços bancários e financeiros (acessos a contas, códigos de verificação, faturas, transferências e PIX). Foram elaboradas cinco mensagens com características de phishing e cinco com características de comunicações legítimas, apresentadas individualmente com elementos visuais semelhantes aos dos aplicativos. Todo o conteúdo é fictício, elaborado exclusivamente para fins acadêmicos, sem envio de mensagens reais.

Esses dois cenários foram escolhidos por representarem situações comuns no cotidiano dos usuários e por serem temas frequentes em tentativas de phishing. No cenário de entregas, as mensagens abordam acompanhamento de encomendas, problemas de entrega e solicitações de regularização; no cenário bancário, tratam de acesso a contas, códigos de verificação, faturas, transferências e comunicações relacionadas ao PIX. A divisão em cinco mensagens suspeitas e cinco legítimas permite comparar a percepção dos participantes diante de cada tipo e verificar possíveis diferenças em suas decisões.

5.3. Matriz de risco das mensagens

Antes da aplicação com os participantes, cada uma das dez mensagens será analisada com base em uma matriz de risco previamente definida. Essa matriz será composta por oito critérios, identificados como C1 a C8 e detalhados na Tabela 3 (Seção 5.4).

Cada critério poderá receber uma pontuação de 0, 1 ou 2, de acordo com a intensidade com que determinado indício de risco estiver presente na mensagem. Ao final, a soma dos oito critérios gerará um escore total para cada mensagem. Com base nesse escore, as mensagens serão classificadas em três faixas de risco:

Tabela 2. Classificação das mensagens segundo o risco

Baixo Risco	0 a 4 pontos
Médio Risco	5 a 9 pontos
Alto Risco	10 a 16 pontos

Essa classificação será utilizada como referência para a análise dos resultados, permitindo verificar se mensagens com maior pontuação de risco tendem a ser identificadas com mais facilidade pelos participantes ou se, em alguns casos, continuam sendo percebidas como confiáveis.

5.4. Participantes e instrumentos de coleta e Critérios da Matriz

Participaram voluntariamente pessoas do público geral, convidadas a responder a um formulário eletrônico no Google Forms. Inicialmente coletou-se o perfil: faixa etária, escolaridade, frequência de uso de aplicativos de mensagens e nível autodeclarado de conhecimento em tecnologia. Em seguida, para cada uma das dez mensagens, o participante indicava (i) se a considerava legítima ou possível golpe; (ii) se realizaria a ação sugerida (clicar em link, responder, confirmar dados, entrar em contato ou pagar); e (iii) seu nível de confiança na decisão, em escala de 0 a 10. Essas três respostas permitem analisar a

capacidade de identificar o phishing, a disposição de agir e a segurança do participante em sua decisão.

A caracterização do perfil permite observar a diversidade da amostra e comparar as respostas conforme as características gerais dos participantes, por exemplo, verificando se diferentes faixas etárias, níveis de escolaridade, frequências de uso e níveis de conhecimento em tecnologia levam a avaliações distintas das mensagens simuladas.

A Tabela 3 apresenta os oito critérios (C1 a C8) que compõem a matriz de risco, com a descrição dos indícios que correspondem a cada pontuação (0, 1 ou 2).

Tabela 3. Matriz de risco: critérios e pontuações.

Cód.	Critério	0 – Baixo	1 – Médio	2 – Alto
C1	Urgência/ameaça	Sem pressa	“Hoje” ou pendente	“Bloqueio”, última chance ou poucos minutos
C2	Autoridade invocada	Nenhuma	Instituição citada	Instituição exigindo ação imediata
C3	Pedido de dados sensíveis	Não pede dados	Confirmação de cadastro ou dado simples	Senha, OTP, CPF, cartão ou documento
C4	Link/URL suspeito ou clique	Sem link	Link genérico	Encurtador, domínio estranho ou pedido de login/confirmação
C5	Recompensa/isca	Nenhuma	Benefício plausível	Prêmio, reembolso alto ou PIX a receber
C6	Pressão emocional	Tom neutro	“Evite transtornos” ou “importante”	Fraude, atividade suspeita ou responsabilização
C7	Identidade inconsistente	Identidade coerente	Identidade genérica	“Oficial” de número estranho ou assinatura incoerente
C8	Qualidade textual suspeita	Bem escrito	Poucos erros	Muitos erros, uso excessivo de maiúsculas ou tradução ruim

5.5. Procedimentos, aspectos éticos e análise

Elaboradas as mensagens e feita sua classificação prévia pela matriz, o formulário foi disponibilizado em uma única rodada, sem treinamento anterior e sem reavaliação, para capturar reações mais espontâneas, próximas de situações reais em que é preciso decidir rapidamente. Não houve coleta de senhas, dados bancários ou outras informações sensíveis; a participação foi voluntária e os dados foram tratados de forma confidencial e agregada, sem identificação individual. Os dados foram exportados para planilha eletrônica e analisados por estatística descritiva (frequências, percentuais, médias de confiança e correlação), com os cruzamentos: (A) score × acerto/erro de classificação; (B) score × decisão de clicar ou não; e (C) score × nível de confiança.

Na organização dos dados, cada mensagem ficou associada ao seu escore de risco e ao conjunto de respostas dos participantes, o que permitiu consolidar a quantidade de respostas corretas e incorretas, a quantidade de participantes que realizariam ou não a ação sugerida, os níveis de confiança por mensagem e as comparações entre mensagens legítimas e de phishing e entre as faixas de risco. Com esses cruzamentos, é possível verificar, por exemplo, se mensagens de alto risco são identificadas com maior facilidade, se determinadas mensagens induzem mais participantes ao clique e se há relação entre confiança alta e acerto, ou entre confiança alta e erro. Também é possível observar situações em que mensagens de aparência mais convincente levam a classificações incorretas, mesmo contendo elementos relevantes de risco.

Os principais cruzamentos previstos são:

A)	score da mensagem × acerto ou erro de classificação
B)	score da mensagem × decisão de clicar ou não clicar
C)	score da mensagem × nível de confiança do participante

6. Resultados

6.1. Perfil dos participantes

A pesquisa contou com 125 participantes. Inicialmente, foi realizada a caracterização do perfil dos respondentes considerando a faixa etária, o nível de escolaridade, a frequência de utilização de WhatsApp, SMS ou outros aplicativos de mensagens e o nível de conhecimento na área de tecnologia.

Tabela 4. Distribuição dos participantes por faixa etária

Faixa etária	Participantes	Percentual
18–25	66	52,8%
26–35	29	23,2%
36–45	14	11,2%
46–60	10	8,0%
60 ou mais	6	4,8%
Total	**125**	**100%**

Fonte: Elaborado pelo autor (2026).

Observou-se que todos os 125 participantes (100%) declararam utilizar diariamente o WhatsApp, SMS ou outros aplicativos de mensagensm, como mostrado abaixo. Esse resultado caracteriza a amostra como composta por usuários com contato frequente com os meios de comunicação considerados no estudo.

Sobre o nivel de tecnologia, observou-se que 83 participantes, ou 66,4%, atribuíram a si mesmos nível 7 ou superior de conhecimento em tecnologia.

Tabela 5. Distribuição dos participantes por escolaridade

Escolaridade	Participantes	Percentual
Fundamental – Incompleto	0	0,0%
Fundamental – Completo	1	0,8%
Médio – Incompleto	2	1,6%
Médio – Completo	18	14,4%
Superior – Incompleto	41	32,8%
Superior – Completo	23	18,4%
Pós-graduação – Incompleto	2	1,6%
Pós-graduação – Completo	38	30,4%
Total	**125**	**100,0%**

Fonte: Elaborado pelo autor (2026).

6.2. Efetividade da matriz de risco

A aplicação da matriz de risco às dez mensagens simuladas resultou em diferentes níveis de pontuação, conforme os critérios estabelecidos anteriormente. A partir da soma das pontuações atribuídas aos oito critérios, cada mensagem foi classificada como phishing ou legítima, permitindo observar a correspondência entre o conteúdo das mensagens e o nível de risco identificado pela matriz. A Tabela 6 apresenta a classificação e o respectivo score de risco obtido para cada mensagem.

Tabela 6. Classificação das mensagens segundo a matriz de risco

Mensagem	Classificação	Score de risco
M1	Phishing	9
M2	Phishing	10
M3	Legítima	2
M4	Phishing	4
M5	Legítima	2
M6	Phishing	6
M7	Legítima	2
M8	Phishing	9
M9	Legítima	0
M10	Legítima	2

Fonte: Elaborado pelo autor (2026).

A média dos scores das mensagens de phishing foi: **7,6 pontos** Enquanto a média das mensagens legítimas foi: **1,6 pontos**0

6.2.1. Resultado da identificação

A Tabela 7 apresenta os resultados obtidos na identificação das dez mensagens pelos participantes, indicando a classificação de cada mensagem, a quantidade de respostas corretas e o respectivo percentual de acertos. Esses dados permitem observar o desempenho dos participantes na distinção entre mensagens de phishing e mensagens legítimas.

Tabela 7. Identificação correta das mensagens pelos participantes

Mensagem	Classificação	Acertos	Percentual
M1	Phishing	115	92,0%
M2	Phishing	119	95,2%
M3	Legítima	99	79,2%
M4	Phishing	122	97,6%
M5	Legítima	114	91,2%
M6	Phishing	122	97,6%
M7	Legítima	99	79,2%
M8	Phishing	111	88,8%
M9	Legítima	107	85,6%
M10	Legítima	99	79,2%

Fonte: Elaborado pelo autor (2026).

Phishing: Das cinco mensagens de phishing: 94,24% das avaliações, em média, identificaram corretamente as mensagens como golpe. **Mensagens legítimas:** Das cinco mensagens legítimas: 82,88% das avaliações, em média, classificaram corretamente as mensagens como legítimas.

A matriz de risco foi aplicada previamente às dez mensagens simuladas, sendo cinco classificadas como phishing e cinco como legítimas. Os escores obtidos variaram entre 0 e 10 pontos. As mensagens classificadas como phishing apresentaram scores de 4 a 10 pontos, enquanto as mensagens legítimas apresentaram scores entre 0 e 2 pontos.

A partir das respostas dos participantes, observou-se que as cinco mensagens de phishing foram identificadas corretamente como golpes em média por 94,24% das avaliações. Entre as mensagens legítimas, 82,88% das avaliações foram classificadas corretamente como legítimas. Considerando conjuntamente as dez mensagens, a taxa média de classificações corretas foi de **88,56%**.

6.3. Comportamento dos usuários e padrões identificados

Nas mensagens de phishing, em média apenas 4,16% dos participantes declararam que realizariam a ação proposta (variação de 1,6% a 6,4%), contra 53,76% nas mensagens legítimas (variação de 39,2% a 72,8%). Isso mostra maior disposição a agir quando a mensagem tinha características compatíveis com uma comunicação legítima. Ao relacionar o score da matriz à proporção de participantes que realizariam a ação, obteve-se correlação negativa forte (r = 0,803): quanto maior o risco atribuído, menor a intenção declarada de agir. Esse é um dos principais achados da pesquisa, pois mostra que a

classificação quantitativa da matriz se relaciona com o comportamento declarado (Tabela 9).

Tabela 8. Comportamento dos participantes diante das mensagens

Mensagem	Classificação	Realizaria ação (%)
M1	Phishing	6,4%
M2	Phishing	4,0%
M3	Legítima	39,2%
M4	Phishing	2,4%
M5	Legítima	72,8%
M6	Phishing	1,6%
M7	Legítima	60,0%
M8	Phishing	6,4%
M9	Legítima	52,0%
M10	Legítima	44,8%

Fonte: Elaborado pelo autor (2026).

6.4. Tabela Final

A Tabela 9 apresenta uma síntese dos principais resultados obtidos na pesquisa. São apresentados os dados referentes à quantidade de participantes e mensagens analisadas, à classificação das mensagens, aos escores médios de risco, aos percentuais de identificação e realização da ação, além da confiança dos participantes e da correlação entre o score de risco e a ação realizada.

Tabela 9. Síntese dos principais resultados da pesquisa

Indicador	Resultado
Participantes	125
Mensagens analisadas	10
Mensagens classificadas como phishing	5
Mensagens classificadas como legítimas	5
Média do score – phishing	7,6
Média do score – legítimas	1,6
Identificação correta – phishing	94,24%
Identificação correta – legítimas	82,88%
Acurácia geral das classificações	88,56%
Realização da ação – phishing	4,16%
Realização da ação – legítimas	53,76%
Confiança média dos participantes	9,07/10
Correlação score $\times$ ação	$r = -0,803$

Fonte: Elaborado pelo autor (2026).

7. Conclusão

Os resultados indicam que a matriz de risco apresentou relação consistente com a percepção dos participantes diante das mensagens simuladas. Os scores diferenciaram claramente as mensagens de phishing (média 7,6) das legítimas (média 1,6); 94,24% das avaliações identificaram corretamente o phishing, e a taxa geral de acerto foi de 88,56%.

Observou-se ainda que a intenção de agir foi de apenas 4,16% nas mensagens de phishing, contra 53,76% nas legítimas, com correlação negativa forte entre score e intenção de ação ($r = 0{,}803$). Quanto à confiança, a média foi alta (9,07), mas a diferença entre respostas corretas e incorretas foi pequena, de modo que confiança elevada não garantiu decisão correta.

Em resposta ao problema de pesquisa, a matriz mostrou potencial para auxiliar na identificação de phishing e esteve associada às decisões declaradas dos participantes. Os resultados devem, contudo, ser interpretados com cautela, considerando o número reduzido de mensagens, o caráter simulado dos conteúdos e o fato de as respostas representarem comportamentos declarados, e não ações efetivas. Como contribuição, o trabalho apresenta uma abordagem quantitativa e simples que transforma características observáveis (urgência, autoridade, solicitação de dados, links suspeitos e pressão emocional) em uma pontuação de risco, e relaciona essa pontuação à percepção, à intenção de ação e à confiança dos usuários.

Referências

Adu-Manu, K. S., Ahiable, R. K., Appati, J. K., and Mensah, E. E. (2023). Phishing attacks in social engineering: A review. *Journal of Cyber Security*.

Jain, A. K. and Gupta, B. B. (2021). Phishing detection: Analysis of techniques and countermeasures. *Computer Networks*.

Kahneman, D. (2012). *Rápido e Devagar: Duas Formas de Pensar*. Objetiva, Rio de Janeiro.

Leonov, P. Y. et al. (2021). The main social engineering techniques aimed at hacking information systems. In *Proceedings of the 2021 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*, pages 471–473. IEEE.

Mitnick, K. D. and Simon, W. L. (2003). *A Arte de Enganar: Ataques de Hackers Controlando o Fator Humano na Segurança da Informação*. Makron Books, São Paulo.

Vilela, E., Ueda, E. T., and Gava, V. L. (2022). Phishing and social engineering: Concept, modalities, techniques of detection and prevention of fraud. a systematic review of the literature. In *19th CONTECSI - International Conference on Information Systems and Technology Management*, São Paulo, Brazil. TECSI, FEA USP. Proceedings of the 19th CONTECSI.