JOURNAL OF HEALTH CONNECTIONS

ISSN 2594-4606

O IMPACTO DAS NOTÍCIAS FALSAS NO CONFLITO COM OS PRINCÍPIOS NA SEGURANÇA DA INFORMAÇÃO[1]

THE IMPACT OF FAKE NEWS IN CONFLICT WITH INFORMATION SECURITY PRINCIPLES

Gabriel de Jesus França[2]
Rone Marques Santos de Jesus[3]
Marcos Augusto Rodrigues de Menezes[4]
João Gabriel Santana Ferreira[5]
Mariano Florencio Mendonça[6]

Resumo
O principal objetivo das Fake News é espalhar desinformação, seja para afetar um grupo específico ou para um público mais amplo, facilitado pela velocidade da propagação de notícias. Esta revisão sistemática analisa os impactos das fake news nos princípios fundamentais da segurança da informação, observando sua disseminação e explorando métodos de combate. Os resultados mostram como as Fake News afetam negativamente o compartilhamento de informações verídicas, expondo a conteúdos maliciosos que impactam confidencialidade, disponibilidade e integridade. As pesquisas analisadas apresentam estratégias de filtragem e gestão de dados, indicando a contribuição da Inteligência Artificial (I.A) na mitigação do problema.

Palavras-Chave: Notícias Falsas; Segurança da Informação; Princípios da Segurança; Segurança.

Abstract
The main objective of fake news is to spread misinformation, whether to affect a specific group or a broader audience, facilitated by an era in which news spreads extremely quickly. This systematic review analyzes the impacts of fake news on the fundamental principles of information security, exploring methods for dissemination and combat. Results show how fake news negatively affects truthful information sharing, impacting confidentiality, availability, and integrity. Analyzed research highlights various filtering and data management strategies, demonstrating how tools like Artificial Intelligence (AI) can reduce the problem.

Keywords: Fake News; Information Security; Principles Of Security; Security.

[1] Artigo recebido: 24/05/2025; Aceito para publicação: 20/06/2025.
[2] Centro Universitário Estácio, Aracaju, Brasil. Email: gabrieldejesusfc.12@gmail.com
[3] Centro Universitário Estácio, Aracaju, Brasil. E-mail: ronemarquesteci@gmail.com
[4] Centro Universitário Estácio, Aracaju, Brasil. E-mail: marcosaugusto7533@gmail.com
[5] Centro Universitário Estácio, Aracaju, Brasil. E-mail: gabrielsantanaf09@gmail.com
[6] Centro Universitário Estácio, Aracaju, Brasil. E-mail: marianofmendonca@gmail.com

1 INTRODUÇÃO

A ascensão da internet e dos smartphones transformou o consumo de informação, popularizando plataformas digitais e redes sociais como Instagram, TikTok e X. Essa facilidade de compartilhamento, contudo, acelera a propagação de fake news, que confundem e manipulam o público. Essas notícias falsas, muitas vezes com linguagem sensacionalista e “clickbait”, induzem a erros de julgamento e causam danos a indivíduos, empresas e governos, como visto nas campanhas antivacinação durante a pandemia de COVID-19.

Ao contrário do jornalismo tradicional, que verifica fatos, as redes sociais permitem que usuários compartilhem informações sem a mesma responsabilidade. As fakes news misturam dados verdadeiros e falsos para aumentar a credibilidade e o compartilhamento, e são disseminadas por uma miríade de usuários, muitas vezes anônimos. Isso expõe o usuário médio à manipulação e a ameaças à segurança da informação.

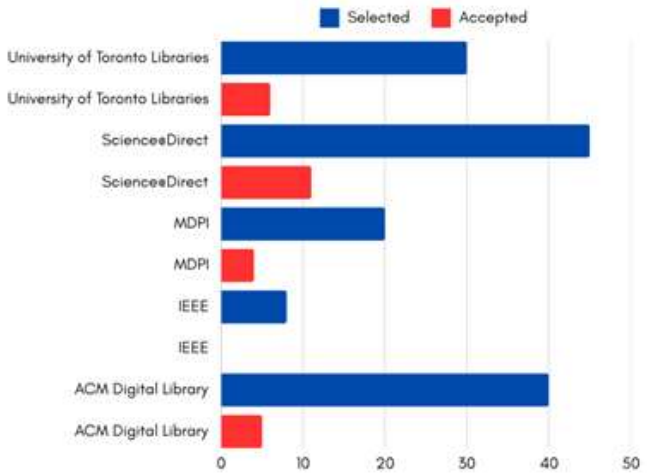
As *fake news* violam os princípios da segurança da informação: confidencialidade (acesso apenas a autorizados), integridade (precisão e proteção contra modificações), disponibilidade (acessibilidade da informação e sistemas) e autenticidade (origem e preservação da integridade da informação, distinta da veracidade). Elas ameaçam a liberdade de expressão, a conscientização pública e as sociedades democráticas, minando a confiança nas instituições.

2 METODOLOGIA

Este estudo utilizou uma revisão sistemática, buscando artigos em bancos de dados científicos como *ACM Digital Library*, *IEEE Xplore*, MDPI, *ScienceDirect* e *University of Toronto Libraries*. Foram encontrados 143 artigos, filtrados por tema, palavras-chave e resumos, e limitados a publicações entre 2018 e 2024, período marcado pela intensificação do debate sobre *fake news* e seu papel em eventos como as eleições legislativas dos EUA de 2018.

ISSN 2594-4606

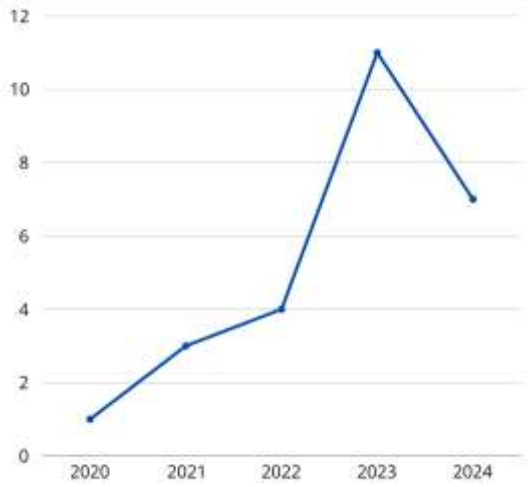
Figura 1 - Quantidade de Artigos Filtrados



Fonte: Autoria Própria (2025).

A Figura 1 e a Figura 2 ilustram o funil de seleção e a distribuição dos artigos por ano de publicação, respectivamente. A escolha de bases de dados híbridas e bibliométricas, comuns em revisões sistemáticas de engenharia de software, visou uma cobertura abrangente. Foram considerados os desafios metodológicos, como a complexidade dos estudos e limitações das bibliotecas digitais, priorizando o rigor.

Figura 2 - Índice de Publicação por Ano



Fonte: Autoria Própria (2025).

Figura 3 - Critérios de Inclusão e Exclusão

CRITÉRIOS DE INCLUSÃO	CRITÉRIOS DE EXCLUSÃO
ARTIGOS CUJO RESUMO ESTEJA DENTRO DO CONTEXTO DA PESQUISA	ARTIGOS PUBLICADOS ANTES DE 2018
ARTIGOS CUJO TÍTULO ESTEJA DENTRO DO CONTEXTO DA PESQUISA	ARTIGOS SEM IDENTIFICAÇÃO DO AUTOR
ARTIGOS PUBLICADOS ENTRE 2018 E 2024	ESTUDOS INACESSÍVEIS PARA DOWNLOAD
ESTUDOS QUE PROPÕEM ANALISAR FAKE NEWS COM BASE NOS PRINCÍPIOS DA SEGURANÇA DA INFORMAÇÃO	ESTUDOS COM RESUMOS FORA DO CONTEXTO DA PESQUISA
	ESTUDOS DISPONÍVEIS COM LEITURA INCOMPLETA
	ESTUDOS DUPLICADOS
	ESTUDOS SECUNDÁRIOS

Fonte: Autoria Própria (2025).

A Figura 3 detalha os critérios de inclusão e exclusão. Para refinar a busca, utilizou-se um "quase-padrão ouro" (Kitchenham e Brereton, 2013) e a string de busca sendo: *("fake news") AND ("security") AND ("information security" OR "principles of security").* Ferramentas de análise textual auxiliaram na seleção inicial.

As perguntas de pesquisa que nortearam este estudo foram:

1. Quais os impactos das fake news nos princípios da segurança da informação (confidencialidade, integridade, disponibilidade e autenticidade)?
2. Quais os impactos das fake news sobre as políticas de privacidade e proteção de dados pessoais?
3. Como as fakes news afetam a tomada de decisões estratégicas em segurança da informação e quais os mecanismos de defesa?
4. Quais os métodos de combate e diminuição das fake news (tecnológicos e educacionais)?
5. Como a legislação e as políticas públicas podem ser ajustadas para lidar com o impacto das *fake news* na segurança da informação?

O artigo está organizado em Trabalhos Relacionados (métodos e desafios na detecção de *fake news*), *Fake News* (formas de disseminação), Identificação e Detecção (estratégias tecnológicas) e Conclusão (resultados e direções futuras).

3 TRABALHOS RELACIONADOS

A detecção de notícias falsas é um campo de pesquisa crescente. Modelos híbridos combinando RNN e SVM (Albahar, 2021) e modelos que integram características emocionais e semânticas com mecanismo de atenção (Jiang et al., 2024) foram propostos para classificar notícias. Para *deepfakes*, Tang et al. (2023) desenvolveram uma estrutura de compressão de modelo.

Sistemas como "AntiFake" usam aprendizado de máquina para identificar informações fraudulentas. Zhang, Yuan e Zhang (2024) propuseram um método de Convolução Extrema Local (BLC) com grafo Bayesiano para detecção de desinformação, destacando a eficácia de métodos baseados em grafos. Para armazenamento seguro de dados de notícias, Xie et al. (2023) introduziram um mecanismo com blockchain e armazenamento em nuvem.

Bayupati, Arsa e Sastrawan (2022) usaram aprendizado profundo (CNN, LSTM bidirecional, ResNet) para detecção de fake news. Huang et al. (2023) propuseram o modelo "ExoFIA" combinando recursos multimodais com atenção para interpretabilidade. Liu et al. (2024) desenvolveram MAGF para detecção multimodal. Schuster et al. (2020) destacaram as limitações da estilometria na detecção de desinformação gerada por Modelos de Linguagem (LM), motivando a melhoria de métodos não estilísticos. Esses estudos demonstram a complexidade e a diversidade de abordagens na detecção de fake news.

3.1 *Fake News*

A detecção de notícias falsas é crucial devido à disseminação generalizada de informações enganosas na internet. A manipulação de mídias digitais, antes complexa, é agora facilitada por aplicativos móveis. Há sobreposição nos tipos de informações falsas; um boato pode usar clickbait, por exemplo. Esta seção detalha as formas de propagação das fake news.

3.2 Propaganda

A propaganda é uma ferramenta antiga de manipulação, usada por Estados para moldar opiniões públicas e afetar políticas. Na guerra Rússia x Ucrânia, a Rússia usa propaganda para criar uma imagem negativa da Ucrânia, causando danos materiais e imateriais e afetando o pensamento crítico. A Ucrânia resiste com inovações em TIC, incluindo IA, e colaboração entre governo e sociedade civil para fortalecer a resiliência digital.

3.3 Clickbait

Clickbait são manchetes enganosas e exageradas para atrair cliques, mesmo que o conteúdo não corresponda. São comuns em redes sociais e plataformas de conteúdo. A detecção exige engenharia de recursos baseada em conteúdo, similaridade textual e informalidade da linguagem. A análise da relação entre vídeo e título é uma área emergente, indicando a expansão do problema para além do texto.

3.4 Voz Sintética

A manipulação de mídia via voz sintética, impulsionada por IA e deep learning, tornou-se extremamente realista, dificultando a detecção humana (Bestagini et al., 2021). Se usadas maliciosamente, podem espalhar desinformação. A detecção é desafiadora pela variedade de métodos de geração e pela constante evolução das técnicas de síntese de fala. Bestagini et al. (2021) abordaram a fala sintética em três níveis: classificação binária (real/sintético), de conjunto fechado (identifica algoritmo conhecido) e de conjunto aberto (detecta algoritmo conhecido ou desconhecido). Eles propuseram descritores de áudio baseados na evolução temporal do sinal, usando um pipeline de aprendizado supervisionado. Os resultados mostraram que o método supera abordagens anteriores, e a combinação de recursos pode melhorar a precisão. Contudo, detectar precisamente algumas famílias de trilhas de fala sintética em cenários de conjunto aberto ainda é um desafio.

3.5 *Deepfake*

A falsificação facial visual atingiu sofisticação que impede a identificação humana, ameaçando a segurança da informação. A tecnologia deepfake, baseada em IA e GANs, substitui identidades em vídeos para disseminar fake news (falsificando políticos) e distribuir pornografia de vingança. O engano aumenta ao combinar deepfakes com fala sintética, representando uma ameaça à integridade e autenticidade dos vídeos digitais. Métodos de detecção de deepfake buscam expor traços anormais em vídeos, aplicáveis em perícia judicial e moderação de conteúdo. Métodos baseados em Deep Learning (DL) têm alto desempenho, mas carecem de interpretabilidade. A detecção de deepfakes é crucial, sendo o método mais comum a modelagem por aprendizado profundo e classificação binária, embora possa gerar falsos positivos para vídeos reais. Existem seis tipos principais de manipulação facial em deepfakes: síntese facial inteira, troca de identidade, transformação facial, manipulação de atributos, troca de expressão e conversão de áudio/texto.

Um estudo de Schuster et al. (2024) para identificar imagens falsas utiliza o método AltFreezing temporal espacial, alinhando rostos com pontos de referência faciais e usando uma estratégia de aumento de dados guiada por atenção (AGDA). Essa abordagem captura artefatos espaciais e temporais simultaneamente, com resultados promissores em diversos conjuntos de dados como UADFV, FaceForensics++, Celeb-DF e DFDC.

Tabela 1 - Resultados da análise da detecção de conjuntos de dados deepfake existentes.

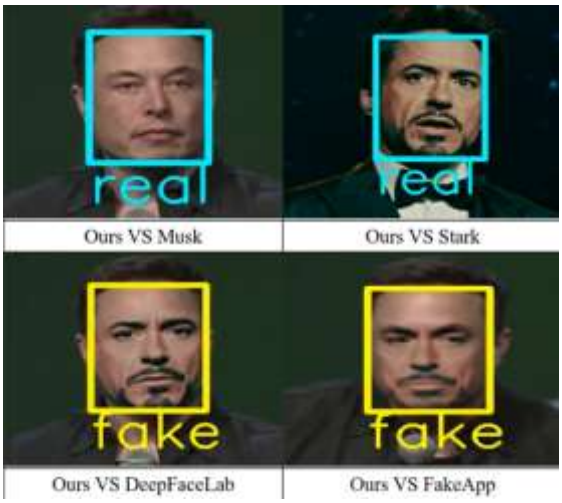
Methods	UADFV	FaceForensics ++	Celeb-DF	DFDC	Avg
FakeVideoForensics	0.9546	0.8637	0.9559	0.9083	0.9206
DeepFakes_FacialRegions	0.2400	0.4000	0.2040	0.2000	0.2610
Improved Xception	0.5625	0.8605	0.6906	0.7755	0.7222

AttFreezing	0.9637	0.9494	0.7390	0.9029	0.8888
Ours	0.9813	0.9794	0.9787	0.9861	0.9814

Fonte: Adaptado de Schuster et al. (2024).

A Tabela 1 e a Figura 4 demonstram a eficácia do método de Schuster et al. (2024) na distinção entre vídeos autênticos e *deepfakes*, incluindo aqueles criados por "DeepFaceLab" e "FakeApp".

Figura 4 - Resultados de detecção do método de estudo de Schuster et al. (2024)



Fonte: Retirado de Schuster et al. (2024)

4 IDENTIFICAÇÃO E DETECÇÃO

Esta seção explora o papel crescente da Inteligência Artificial (IA) na identificação e detecção de notícias falsas.

4.1 Inteligência Artificial (IA)

A segurança de dados, que protege informações confidenciais em repouso e em trânsito, utiliza ativamente técnicas de IA. Aplicações incluem prevenção de vazamento de dados, proteção inteligente de e-mail, bloqueio de domínios maliciosos e monitoramento de integridade, visando garantir confidencialidade, integridade e disponibilidade.

A IA, especialmente o Deep Learning, aprende recursos de dados brutos automaticamente, permitindo a criação de sistemas automatizados que executam tarefas de classificação em tempo real com alta precisão, tornando a detecção de fake news mais eficiente.

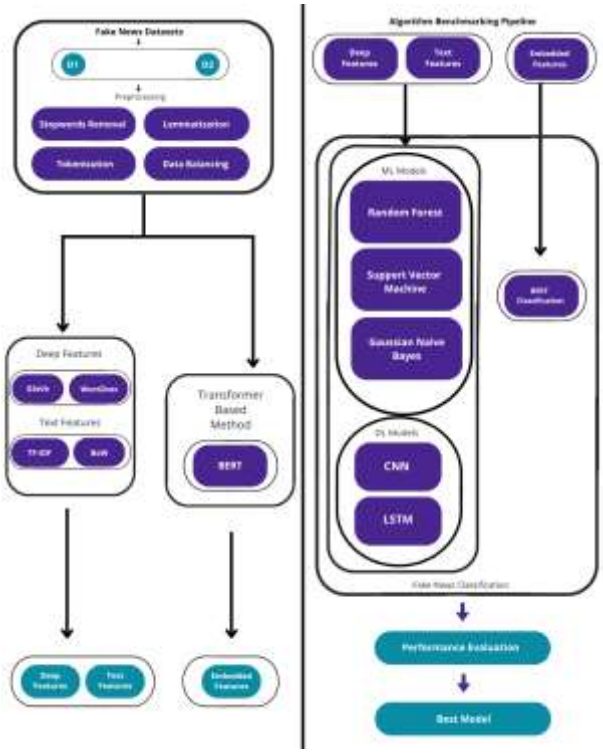
4.2 Benefícios

A verificação de fatos por humanos é precisa, mas lenta e cara. Algoritmos de aprendizado de máquina podem automatizar e acelerar esse processo, eliminando vieses humanos. Embora exijam grandes conjuntos de dados de treinamento e seleção de recursos, eles podem alcançar alta precisão e processar volumes massivos de informações rapidamente. A desvantagem é a dificuldade em obter bancos de dados extensos e de alta qualidade. Detectar fontes de notícias falsas conhecidas é eficaz, mas não se aplica a artigos anônimos ou em redes sociais. Nesses casos, abordagens mais sofisticadas, baseadas no conteúdo da notícia, são necessárias.

4.3 Identificação com IA

O modelo de Al Solami e Saeed (2023) para classificação de notícias falsas utiliza uma arquitetura modular. O pré-processamento inclui remoção de nulos, filtragem de expressões, stopwords, lematização, tokenização e análise contextual. A extração de recursos usa Word2vec, GloVe, N-gram TF-IDF e BoW. Os recursos textuais são classificados por RF, SVM e LR, enquanto os profundos alimentam CNN e LSTM. O BERT também processa e classifica os dados. A avaliação usa precisão, revocação, F1-score e acurácia.

Figura 5 - Diagrama de estrutura para o modelo de detecção de notícias falsas



Fonte: Tabela reformulada de AL SOLAMI; SAEED (2023)

A Figura 5 apresenta o fluxo de trabalho do modelo. Dois conjuntos de dados abertos foram utilizados: "Fake News Detection" (4009 instâncias, 2137 reais, 1872 falsas) e "Real and Fake news" (6335 instâncias, 3164 falsas, 3171 reais). A Tabela 2 e a Tabela 3 detalham as estatísticas desses conjuntos.

Tabela 2 - Estatísticas do conjunto de dados de notícias falsas

Attribute	**Value**
Dataset attributes	URL, headline, body text, label
Total news	4009
Real news	2137
Fake news	1872
Word count	2,326,583
Character count	10,111,421
Sentence count	4273

Fonte: Adaptado de AL SOLAMI; SAEED (2023)

Tabela 3 - Estatísticas do conjunto de dados de notícias reais e falsas

Attribute	**Value**
Dataset attributes	URL, headline, body text, label
Total news	6,335
Real news	3,171
Fake news	3,164
Word count	5,365,684
Character count	23,842,804
Sentence count	6,341

Fonte: Adaptado de AL SOLAMI; SAEED (2023)

O modelo de Al Solami e Saeed (2023) aplica pré-processamento, extração de características e classifica os dados com SVM, RF, GNB, CNN, LSTM e BERT. O SVM demonstrou melhor precisão para características textuais de TF-IDF (98,83%) e BoW (97,66%), superando RF e GNB, e apresentando melhores métricas gerais, como mostrado na Tabela 4.

Tabela 4 - Resultados da classificação de recursos textuais do Conjunto de Dados 1 com algoritmos de ML

PEM	SVM-TF-IDF (%)	SVM-BoW (%)	RF-TF-IDF (%)	RF-BoW (%)	GNB-TF-IDF (%)	GNB-BoW (%)
Accuracy	98.83	97.66	95.7	94.3	93.4	92.5
Precision	99	98	96	94	94	93
F1-score	99	98	96	94	93	93
Recall	99	98	96	94	93	93

Fonte: Autoria Própria (2025)

5. CONCLUSÃO

Este artigo revisou sistematicamente os perigos das fake news para os princípios da segurança da informação, detalhando como elas ameaçam a confidencialidade, integridade, disponibilidade e autenticidade. Exploramos as diversas formas de propagação (propaganda, clickbait, vozes sintéticas, deepfakes) e analisamos pesquisas sobre métodos de combate, incluindo o uso de blockchain e Inteligência Artificial.

A pesquisa aponta que as fakes news são um problema complexo e multifacetado, com impactos e complicações crescentes, especialmente devido ao aumento de notícias geradas por máquinas, que amplificam a credibilidade de informações falsas. Apesar das iniciativas de combate, ainda falta um plano de ação global claro, pois o problema não foi suficientemente definido.

Este documento serve como diretriz para futuras pesquisas no combate à desinformação e aos riscos que ela representa para a segurança da informação, visando um ecossistema informacional mais resiliente.

6. REFERÊNCIAS

ALBAHAR, M. A hybrid model for fake news detection: Leveraging news content and user comments in fake news. **IET Information Security,** v. 15, 2021.

BORRELLI, C.; BESTAGINI, P.; ANTONACCI, F.; SARTI, A.; TUBARO, S. Synthetic speech detection through short-term and long-term prediction traces. **EURASIP Journal on Information Security,** 2021.

CASSAR, D. The misinformation threat: A techno-governance approach for curbing the fake news of tomorrow. **Digit. Gov.: Res. Pract.,** v. 4, 2023.

SYEROV, Y.; KRYVINSKA, N.; FEDUSHKO, S. Antifake system: Machine learning-based system for verification of fake news. **Procedia Computer Science,** v. 238, 2024.

JACOBS, B. The authenticity crisis. **The Computer Law and Security Report,** v. 53, 2024.

KAUR, R.; GABRIJELČIČ, D.; KLOBUČAR, T. Artificial intelligence for cybersecurity: Literature review and future research directions. **Information Fusion,** v. 97, 2023.

KITCHENHAM, B.; BRERETON, P. A systematic review of systematic review process research in software engineering. **Information and Software Technology,** v. 55, 2013.

HUANG, Y.-Y.; SHEI, C.; HSIEH, H.-P.; LI, P.-X. Exofia: Deep exogenous assistance in the prediction of the influence of fake news with social media explainability. **Applied Sciences,** v. 13, 2023.

PHAM, M. T. et al. A dual benchmarking study of facial forgery and facial forensics. **CAAI Transactions on Intelligence Technology,** 2024.

BANSAL, D.; RASTOGI, S. A review on fake news detection 3T's: typology, time of detection, taxonomies. **International Journal of Information Security,** v. 22, 2022.

KHARCHENKO, A.; ROMTSIV, O. I**. Methods of Russian information propaganda and their influence on the image of Ukraine in the world.** 2023.

AL SOLAMI, E.; SAEED, A. Fake news detection using machine learning and deep learning methods. **Computers, Materials and Continua,** v. 77, 2023.

BAYUPATI, I. P. A.; ARSA, D. M. S.; SASTRAWAN, I. K. Detection of fake news using deep learning CNN–RNN based methods. **ICT Express,** v. 8, 2022.

SCHUSTER, R.; SHAH, D. J.; BARZILAY, R. The limitations of stylometry for detecting machine-generated fake news. **Computational Linguistics,** v. 46, 2020.

SCHUSTER, R.; SHAH, D. J.; BARZILAY, R. Video detection method based on temporal and spatial foundations for accurate verification of authenticity. **Electronics,** v. 13, 2024.

LIU, M.; ZHANG, M.; WANG, Z.; SHAN, F. Fake news detection based on cross-modal message aggregation and gated fusion network. **Computers, Materials and Continua,** v. 80, 2024.

TAHERDOOST, H. Cybersecurity vs. information security. **Procedia Computer Science,** v. 215, 2022.

[CERONI, E. G.; MARZIALI, S.; PANCINO, N.; MAGGINI, M.; BIANCHINI, M.; TANFONI, M. Generated or not generated (GNG): The importance of background in the detection of fake images. **Electronics,** v. 13, 2024.

TARANENKO, A. Ensuring information security: Countering Russian disinformation in Ukrainian speeches at the United Nations. **Social Sciences Humanities Open,** v. 10, 2024.

JAMES, T. L.; LOWRY, P. B.; VILLACIS CALDERON, E. D. How Facebook's Newsfeed algorithm shapes childhood vaccine hesitancy: An algorithmic fairness, accountability, and transparency (FAT) perspective. **Data and Information Management,** v. 7, 2023.

XIE, H.; JI, S.; LIU, L.; HUANG, D.; WANG, X. Blockchain-based fake news traceability and verification mechanism. **Heliyon,** v. 9, 2023.

WANG, M.; ZHANG, X.; REN, K.; YAN, F.; JIANG, W. A model for detecting fake news by integrating domain-specific emotional and semantic features. **Computers, Materials and Continua,** v. 80, 2024.

TANG, S.; ZHU, M.; HE, P.; LI, S.; CAO, Y.; XU, X. A novel model compression method based on joint distillation for deepfake video detection. **Journal of King Saud University - Computer and Information Sciences,** v. 35, 2023.

ZHANG, S.; YUAN, G.; ZHANG, G. Bayesian graph local extrema convolution with long-tail strategy for misinformation detection. **ACM Transactions on Knowledge Discovery from Data,** v. 18, 2024.

ŁUCZUK, P. Awareness and following of information security policies as the main rule to protect against threats in digital communication processes. **Cybersecurity as the Arena of Modern Warfare. Social Communication,** v. 7, 2021.

ZHENG, N.; GUO, C. (J.); GUO, C. Seeing is not believing: A nuanced view of misinformation warning efficacy on video-sharing social media platforms. **Proceedings of the ACM on Human-Computer Interaction,** v. 7, 2023.