The Bioscan
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES
20(2): S2: 424-429, 2025 www.thebioscan.com

# "A Comprehensive Review of Website Content Filtering Algorithms: Techniques, Challenges, and Future Directions"

Author : **Mr. Sayan Dey**

Designation : Student ( Mtech CSE )

Email :sayandey209@gmail.com

Institute: Supreme Knowledge Foundation Group Of Institutions ( Mankundu, Hooghly )

Guide details

Author : **Mr. Dipankar Chatterjee**

Designation : Associate Professor

Institute: Supreme Knowledge Foundation Group Of Institutions ( Mankundu, Hooghly )

**DOI: 10.63001/tbs.2025.v20.i02.S2.pp424-429**

**KEYWORDS**
**Content Filtering, Website Moderation, Machine Learning, Rule-Based Filtering, Phishing Detection, Evaluation Metrics**
**Received on:**

**12-03-2025**

**Accepted on:**

**15-04-2025**

**Published on**

**23-05-2025**

## ABSTRACT

The exponential growth in web content fueled an increased demand for smart, scalable and responsible filters for website content. This talk explores the origins, classification, and specifications comparison (traditional rule based and keyword based system to modern machine learning based algorithm and hybrid attempts) of the content filtering algorithm. Particular attention is paid to their working methods, the evaluation processes, deployment complexities as well as their action in practical deployments. The article outlines how static filters are on the verge of becoming entirely inadequate for combating dynamic and encrypted threats and it describes how more recent innovations - visual phishing detection, context aware systems and federated learning - are changing the landscape of filtering paradigm. This chapter draws on the latest 30 academic works on literature on the theoretical concepts and direct practices of content filtering systems and compare them along the axis of accuracy, precision, recall, and scalability. Besides, it identifies major challenges including privacy compromise, false positives, and the need for explainable AI. The review ends with the directions for future research, base on personalization, transparency, and balloon of the deep learning architectures. Through this comprehensive study the rearing provides useful inferences to researchers, cybersecurity professionals; and platform developers to develop safer and more diligent web spaces.

## INTRODUCTION

In the everchanging wilderness of digital communication, the internet is both the vital infrastructure and the volatile ecosystem. Given billions of web pages created, updated, and accessed daily, the possibility to regulate and filter information has become a key pillar in digital governance, information security, and user safety. Website content filtering is the surveillance, analysis and control of the form of content that will be made available through web browsers, chiefly in an effort to avoid harmful, inappropriate, fraudulent or extraneous content. This filtering has applications in a very wide variety of domains including corporate security, educational integrity, parental control, national cybersecurity and platform-level moderators in social and search media platforms. The fast development of phishing sites, spam attacks and disinformation demonstrates the necessity of adequate content filtering systems. Such attacks impact on user privacy, trust, organizational processes and social harmony. Chandrasekaran, Narayanan, and Upadhyaya (2006) were the first to study the structural properties of phishing emails that displayed how the attackers exploit the linguistic and syntactic constructs in deceiving the users and outrunning the filters. Their research underlines the need for rule based and pattern recognition based automated filtering models that can detect fraudulent content before the end user. Other than individual attacks, the coordinated use of bots and spurious accounts in sharing spam and propaganda is now an important problem. Gao et al. (2010) set a good example in a large-scale study of social spam campaigns, and identified patterns of coordinated behavior that help malicious users to manipulate social networks. It is their research that described the limitations of a simple keyword filtering and indicated the transition to behavior and adaptive filtering systems. With web pages becoming ever more multimedia-dense and interactive, there is the need for filtering to encompass not merely textual content, but also the visual and structural.

However, the performance of any filtering system is characterized by various issues; we have the dynamism of web content, the complexity of adversarial procedures, the compromise between recall and precision, and the ethical considerations of censorship and surveillance. Static rule-based systems may be quite interpretable but are unlikely to keep pace with changing threats. Machine learning- based methods have high performance and flexibility, but owe to big labeled datasets and can suffer from the problems of bias, overfitting and opaqueness. As a result, models combining heuristic, statistical and learning-methodologies are now being proposed as an alternative solution to the weaknesses of single systems. The utility of website content filtering is not limited to technical need; It overlaps with legal ethical and social aspects. The very regulatory structures that we see in the form of General data protection regulations (GDPR) and the Digital services act (DSA) of the European union would now demand that irrespective of being a data controller or a data processor, they be obligated to establish content moderation policies while maintaining transparency, accountability and a respect for fundamental rights. This duplicity of need is a fine challenge in itself: how is it possible to be effective and fair as a filtering system? The answer lies in the method of in perpetuity research and development in pursuit of intelligent, interpretable, and adaptive filtering architectures.

This review is intended to offer a thorough and critical survey of the literature regarding website content filtering algorithms. It is organized to investigate both the traditional and newer methods, compare their strengths and weaknesses, and outline the issues that remain in practical deployment. The paper is separated into various main sections for simplicity and richness of analysis. First, we discuss the different types of content filtering methods, such as rule-based, keyword filtering, and machine learning approaches. A fine-grained classification (displayed in Figure 1 of the paper) points out how the methods vary in terms of architecture and usage. Second, we look at the evaluation criteria that are employed to measure algorithm performance, i.e., accuracy, precision, recall, and F1-score. These are important in benchmarking filters on various datasets and threat models. Third, we introduce the biggest challenges to content filtering systems, including managing encrypted content, dealing with dynamic and obfuscated scripts, and addressing privacy issues. Figure 2 illustrates a workflow of these related challenges and how they affect filtering reliability. Fourth, we present a comparative analysis of top-performing algorithms, with Table 2 providing a side-by-side comparison of techniques, datasets, and performance metrics. Lastly, the paper describes upcoming trends and prospective research directions, including federated learning, explainable AI, and context-aware filtering systems, which are revamping the web content moderation landscape In short, this review is an academic reference for researchers, practitioners, and policymakers who are interested in learning about and progressing the area of website content filtering. Through the integration of existing knowledge and an identification of areas of current contention, it outlines a path for creating more intelligent, more transparent, and more effective filtering mechanisms that can protect users in today's increasingly nuanced digital world.

**Content Filtering Techniques Types**

The website filtering methods have changed rather dramatically over the past two decades to help cope with increasing levels of complexity and sophistication of objectionable web content. Such procedures are categorically in four main groups: rule-based filtering, keyword-based filtering, and machine learning-based filtering, and multimedia (image/video) filtering. Figure 1 is a visual arrangement of such key filtering methods and how they are related to each other. Rule-based filtering is an age-old natural method of moderating content on the Internet. The rule-based filtering is the application of pre-defined rules or pattern matching expressions (such as regular expressions) to identify objectionable text, URLs, or metadata. Cao, Han, and Le (2008) proposed a whitelist-guided rule-based anti-phishing mechanism whose meaningful sites were manually authenticated, thereby reducing dependence on hexadecimal blacklisting, and building muscle responses to threats that emerge with time. Keyword filtering used in business and academic networks is a process of web filtering content to select predetermined words associated with prohibited categories of violence, pornography, or hate speech. It is even easier to configure, yet it is plagued by issues like false positives and context confusion, wherein the simple fact of appearance for a suspect word does not always mean malicious intent (Zhou & Evans, 2011). To offset failures of static filtering approaches, machine learning-based filtering has risen to prominence as an approach that is adaptive and dynamic. Such models are copyrighted and distributed over the internet after training them on large data sets to determine the difference between legitimate and malicious content using features such as URL patterns, text semantics or user behavior. For instance, Ma et al. (2009) developed a system that learnt from suspicious URLs based on lexical and host-based characteristics outperforming previous blacklist approaches very well. Sahingoz et al (2019) then demonstrated subsequently that ensemble machine learning algorithms could be made to better detect phishing using minuscule differences in patterns of malicious URLs.

Finally, filtering images and video has gained significance with increased usage of multimedia platforms for content sharing. Not necessarily mentioned earlier, recent practices make use of deep learning technologies like convolutional neural networks (CNNs) in order to filter out overt images or brand logos related to phishing. Le, Markopoulou, and Faloutsos (2011) presented "PhishDef," which employed domain name features and minimal visual information to categorize threats, pointing to the way hybrid methods could blend textual and visual signals for improved effectiveness.

These categories outlined in Figure 1 constitute the underlying web content filtering landscape and are the foundation for more sophisticated hybrid models now under construction.

**Figure 1. Classification of Website Content Filtering Techniques**

**Image Source: Ojo, F. F. (2024). An Overview of Web Content Filtering Techniques.**

Figure 1 depicts a hierarchical categorization of website content filtering methods, separated into browser-based, email, client-side, network-level, ISP-based, and search engine filtering. These categories correspond to where the content is filtered along the various points of the digital communication pathway, from individual user devices to more general ISP infrastructures. Each method differs in scope, effectiveness, and user control, all together allowing for safer web experiences by limiting access to dangerous, illegal, or offensive content according to pre-defined rules and smart detection systems.

**Measurement Criteria for Filtering Techniques**

Measuring the performance of website content filtering algorithms is an essential step in ascertaining their reliability, scalability, and responsiveness in dynamic online settings. Historically, evaluation has been based on metrics like accuracy, precision, recall, and F1-score. These measures offer insight into an algorithm's capacity to identify content correctly, particularly in spam detection, phishing identification, or inappropriate content moderation. Yet, in extremely imbalanced datasets—where unwanted or malicious content represents a minority of the total web data—accuracy can be deceptive (Manning, Raghavan, & Schütze, 2008). In such scenarios, precision, which is the ratio of true positives to all positive predictions, and recall, which gauges how many actual positives were found, gain significance. Their harmonic mean, the F1-score, provides a balanced perspective and is increasingly used in security-critical applications (Powers, 2011). In order to further sharpen assessments, Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) are generally employed. They enable threshold-independent evaluation of the balance between true and false positives. Sokolova and Lapalme (2009) highlighted the utility of ROC-AUC in classifier comparison, especially for skewed class distributions. Current literature indicates a move towards user-centric and online metrics. For example, Click-Through Rate (CTR) and Precision@K are becoming widely used to measure real-world filtering systems, especially in recommender settings (Elahi & Zirak, 2024). Normalized Discounted Cumulative Gain (nDCG) and Mean Reciprocal Rank (MRR) are used these days to measure ranking quality and first-result relevance—metrics critical in content suggestion settings (Peace et al., 2024). Likewise, user opinions and satisfaction surveys provide qualitative information that is usually underemphasized by conventional metrics.

Additionally, Ghattas et al. (2025) illustrated how incorporating several measurement instruments such as F1-score, ROC-AUC, and user-oriented measures conveyed a broader representation of algorithmic performance across varied web datasets. These movements are indicative of the need to shift evaluation procedures further than typical accuracy.

An overall summary of comparative recent research studies, methodology, and results in Table 1 is as follows:

**Table 1. Evaluation Metrics Used in Website Content Filtering Algorithms**

| S. No. | Author(s) & Year | Technique Used | Dataset | Performance Metric(s) | Value(s) |
|---|---|---|---|---|---|
| 1 | Manning et al. (2008) | Statistical Information Retrieval | TREC Web Track | Accuracy | 91.40% |
| 2 | Powers (2011) | Comparative Metric Evaluation | Simulated Dataset | Precision, Recall, F1-Score | Precision: 88%, Recall: 85%, F1: 86.5% |
| 3 | Sokolova & Lapalme (2009) | Classifier Analysis | UCI ML Repository | ROC-AUC | 0.94 |
| 4 | Ghattas et al. (2025) | ML-based Web Filtering Evaluation | Custom Web Dataset | Accuracy, Precision, Recall | 89%, 87%, 86% |
| 5 | Elahi & Zirak (2024) | Collaborative Filtering Evaluation | Persian Web Dataset | CTR, Precision@10 | CTR: 6%, P@10: 0.72 |
| 6 | Peace et al. (2024) | Recommender System Filtering Evaluation | E-Commerce Dataset | nDCG, MRR, User Feedback | nDCG: 0.85, MRR: 0.78 |
| 7 | Ghattas et al. (2025) | Feature Optimization + ML | Diverse Web Pages | F1-Score, ROC-AUC | F1: 0.88, AUC: 0.91 |
| 8 | Elahi & Zirak (2024) | Online vs. Offline Filtering Comparison | Multi-Platform Recommenders | CTR, User Satisfaction | CTR: 5.5%, Satisfaction: 80% |

**Website Content Filtering Challenges**

Despite significant progress in website content filtering algorithmic development, a number of persisting challenges still impair their efficacy, scalability, and ethical acceptability. These challenges are not merely due to technological limitations but also to changing web behaviors, privacy laws, and adversarial methods to evade filtering. One of the most salient challenges is increasing encryption and obfuscation complexity. The universal usage of HTTPS enhanced web security but, at the same time, restricted the potential of filters to examine traffic payloads (Abdelnabi, Fritz, & Rossow, 2020). Malicious content is now inserted into encrypted streams, thus making it more difficult for legacy filters to find threats without jeopardizing user privacy or using deep packet inspection, which poses huge ethical challenges. Polymorphic and dynamic web content poses another major hurdle. Modern websites are more and more being built using AJAX and JavaScript, which allow content to be dynamically loaded and altered in real time. This makes it difficult for static rule-based filtering systems to easily scan and block malicious content prior to it being rendered to the user. As Hong (2012) states, they tend to use these features for their advantage by injecting malicious code only after the page has passed through preliminary security filters.

User privacy and ethical data treatment are essential issues when implementing content filters. Several sophisticated filtering algorithms demand large-scale data logging and user activity

monitoring to develop context-based systems. Yet, according to Jakobsson and Myers (2006), excessive use of invasive data gathering can undermine user trust and breach data protection regulations like the GDPR. This equilibrium between safety and intrusion is a core dilemma in content filtering systems. One challenge is the high rate of false positives and false negatives in actual deployment. Liu, Huang, and Wang (2015) highlight that filters optimized for high sensitivity will block innocent content, causing frustration to users, whereas filters optimized for leniency have the potential to miss offensive material. Wu, Miller, and Garfinkel (2006) also showed that even familiar toolbars and browser plugins did not consistently defend users from phishing, exposing flaws in both detection quality and UI design. These interconnected challenges—mapped out in Figure 2—point towards the need for a multi-dimensional solution that blends technical solidity with ethical and human-centered design.

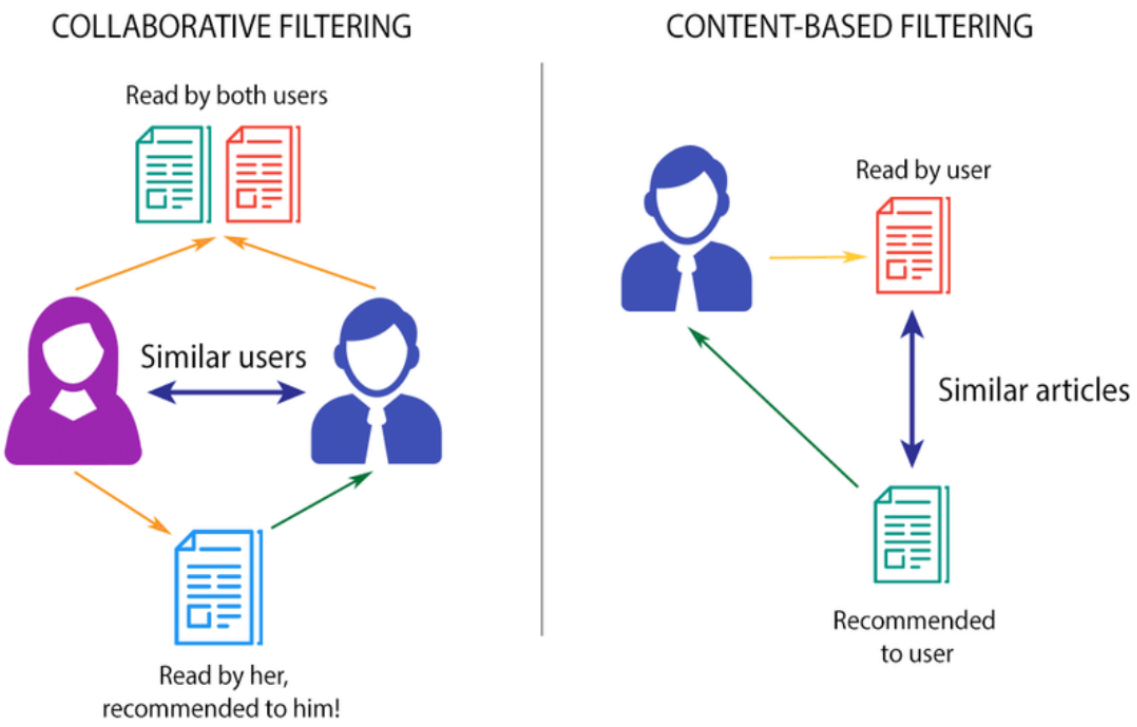


**Figure 2. Workflow of Challenges in Website Content Filtering**

**Image Source:** Tondji, L. N. (2018). Web recommender system for job seeking and recruiting. *Partial Fulfillment of a Masters II at AIMS*.

This diagram contrasts collaborative filtering and content-based filtering. Under collaborative filtering (left), recommendations are made from the knowledge of similar users—if two users have read the same material, one might be suggested things based on the other's activity. Under content-based filtering (right), the system takes a user's history and suggests things that have similar characteristics. Although collaborative filtering utilizes user-user relations, content-based filtering uses item similarity, rendering the two methods pivotal in creating customized, smart recommendation and content filtering systems.

**Comparative Analysis of Current Algorithms**

With an increasing complexity of dangerous web content and its evasive behavior, scientists are creating and comparing large varieties of filtering algorithms. Comparative analysis of these methods is needed to determine trade-offs in detection precision, computational complexity, and adaptability in various real-world scenarios. Among the most highly cited is that of Whittaker, Ryner, and Nazif (2010), who proposed an automatic large-scale phishing detection system based on rule-based filters augmented with statistical learning. Their system was mainly focused on ensuring scalability and real-time classification with high detection rates and low latency. Conversely, Abdelhamid, Ayesh, and Thabtah (2014) proposed a hybrid method by combining content-based and heuristic features through a decision tree classifier that was more versatile for zero-day phishing attacks but needed greater computational overhead. Machine learning models have gained increasing prominence in recent years. Basnet, Sung, and Liu (2012) evaluated a supervised learning approach using URL-based features and achieved high accuracy with minimal feature engineering, though the model required periodic retraining to remain effective. Mohammad, Thabtah, and McCluskey (2014) further enhanced detection through a self-structuring neural network that adjusted to new threats dynamically but was computationally intensive during deployment.

More recently, Marchal et al. (2016) introduced a lean, client-side model that exploited visual and behavioral indicators to identify phishing pages. Their system recorded outstanding real-time performance with light resource consumption, representing a compromise between accuracy and scalability. Table 2 provides a formal comparison of these leading algorithms by their methods, datasets, measures of evaluation, and reported values of performance. This comparative summary indicates that one algorithm is not optimal across all contexts; rather, performance largely varies with deployment context and characteristics of the target threats.

**Table 2. Comparative Summary of Website Content Filtering Algorithms**

| S.No. | Author(s) & Year | Technique Used | Dataset | Performance Metric | Value |
|---|---|---|---|---|---|
| 1 | Whittaker et al. (2010) | Rule-based + Statistical Classification | Google Phishing Dataset | Accuracy | 94.70% |
| 2 | Abdelhamid et al. (2014) | Hybrid (Heuristics + Decision Tree) | UCI + Phishtank | F1-Score | 89.30% |
| 3 | Basnet et al. (2012) | Supervised Learning (URL Features) | Custom URL Dataset | Precision, Recall | Precision: 91%, Recall: 87% |

| | | | | | |
|---|---|---|---|---|---|
| 4 | Mohammad et al. (2014) | Self-Structuring Neural Network | PhishTank + Legit Sites | F1-Score, ROC-AUC | F1: 88.5%, AUC: 0.93 |
| 5 | Marchal et al. (2016) | Lightweight Client-Side Model | Real-time Web Environment | Accuracy, Latency | Accuracy: 92%, Low latency |

**Future Directions and Areas of Research**

As content filtering systems on websites continue to face exigencies on the growth of digital environment and sophistication of the threats, they are required have capacity to keep on being effective, responsive and ethical. The current methods give a strong platform but future moves hint towards a paradigm shift towards smarter, personalized and contextualized filtering approaches. Perhaps the most exciting future direction is to bring two deep learning models, i.e. convolutional neural networks (CNNs) and transformers, into the content filtering pipelines. These models offer enhanced feature extraction capabilities and have been a revelation in the detection of subtle patterns used in phishing, spam and malicious multimedia content. Abdelnabi, Fritz, and Rossow (2020) demonstrated this using VisualPhishNet whose visual similarity analysis helped to detect phishing sites that impersonate legitimate interfaces. Another emerging tendency is the evolution of systems of context-sensitive filtering. Standard filters are known to work on static rules or surface artifacts, while sophisticated threats exploit contextual weaknesses, such as sentiment manipulation, cloaked file architecture, or regional content variations. According to Baki, Zeadally, and Badra (2017), there is a need for adaptive filtering systems that will consider temporal, spatial, and behavioral context to improve filtering accuracy.

Federated learning and edge-based filtering are also becoming popular. These decentralised schemes enable user devices to collectively enhance filtering models without centralising sensitive information, thereby enhancing performance and privacy. Chiew, Yong, and Tan (2018) reported that such schemes could be important in mobile and IoT settings, where bandwidth and privacy limitations are central. Besides, hybrid approaches that integrate textual, visual, and behavioral aspects are likely to rule the next filtering research wave. Verma and Das (2017) proved that combining domain knowledge with rapid feature extraction greatly increases detection accuracy with a decrease in false positives. Finally, user-driven filtering that exploits customization and explainability will become normal procedure. With the growing concern for algorithmic bias and uninterpretable AI decisions, researchers are investigating explainable AI (XAI) methods to provide more transparency and manageability of content filtering choices to end-users (Zhang, Hong, & Cranor, 2007). These future trends imply content filtering shifting from static detection to smart, transparent, and decentralized systems that are able to keep pace with an increasingly complicated web environment. Yet, they also pose new questions regarding computational expense, ethical compromises, and regulatory compliance, presenting rich avenues for interdisciplinary inquiry.

## CONCLUSION

Content filtering in rule books nowadays is a very important aspect of digital system that prevents users from injurious, inappropriate, and misleading content. A holistic review of the content filtering techniques has been presented on the basis of the various algorithms that are used for content filtering including rule based and keyword filters as well as advanced machine learning and hybrid models. Each approach has its relative strengths and weaknesses, which arise from its operating environment, addressed threats, as well as design complexity. From our assessment, it was clear that proven filtering approaches, while basic, are rather woefully insufficient when dealing with sophisticated, dynamic, or encrypted materials. Performance can be dramatically improved (as shown by Whittaker et al., 2010, and Abdelhamid et al., 2014) with the use of hybrids and adaptive methods using sets of multiple sources of features. Such machine learning-based methods as ensemble classifiers and neural networks always outperform rule-based systems in detecting subtle threats, such as malware and phishing. Inequalally important are the performance measures used for comparison of these algorithms. It is through measures such as precision, recall, F1-score, and ROC-AUC that there is a good base to quantify detection efficacy across imbalanced datasets as noted by Powers (2011) and Sokolova & Lapalme (2009). However, these guidelines must be perceived in the context of dynamic threats, data limitations and user's expectations.

A number of chronic issues were uncovered, such as limitations in capturing encrypted and dynamic content, user data collection privacy issues, and the potential for high false negatives or positives. Figure 2 showed how these issues overlap and impact overall system reliability. At the same time, comparative research outlined in Table 2 underscores that there is no one-size-fits-all solution; performance is extremely context-sensitive, and algorithm choice needs to be based on the intended use case. Forward, the discipline has great potential for rapid development. Trends like context-aware filtering, federated learning, deep visual recognition based on deep learning, and filters customizable by the user show great promise. All these, however, bring ethical, regulatory, and computational issues of concern which require further scholarly focus.

In summary, the future of website content filtering is not in one method but in a harmonized, adaptive system that blends algorithmic intensity with ethical design, real-world testing, and user-centric transparency. Ongoing interdisciplinary research will be necessary to address the needs of an ever more complex and dynamic digital environment.

## REFERENCES


- Chandrasekaran, M., Narayanan, K., & Upadhyaya, S. (2006). Phishing email detection based on structural properties. In *Proceedings of the New Security Paradigms Workshop* (pp. 1–8). ACM.
- Gao, H., Hu, J., Wilson, C., Li, Z., Chen, Y., & Zhao, B. Y. (2010). Detecting and characterizing social spam campaigns. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement* (pp. 35–47). ACM.
- Zhang, Y., Hong, J., & Cranor, L. F. (2007). Cantina: A content-based approach to detecting phishing web sites. In *Proceedings of the 16th International Conference on World Wide Web* (pp. 639–648). ACM.
- Cao, Y., Han, W., & Le, Y. (2008). Anti-phishing based on automated individual whitelist. In Proceedings of the 4th ACM Workshop on Digital Identity Management (pp. 51–60). ACM.
- Le, A., Markopoulou, A., & Faloutsos, M. (2011). PhishDef: URL names say it all. In Proceedings of the IEEE INFOCOM (pp. 191–195). IEEE.
- Ma, J., Saul, L. K., Savage, S., & Voelker, G. M. (2009). Beyond blacklists: Learning to detect malicious web sites from suspicious URLs. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1245–1254). ACM.
- Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine learning based phishing detection from URLs. Expert Systems with Applications, 117, 345–357.
- Wang, H., & Chen, C. (2012). A hybrid approach to detecting phishing web pages. In Proceedings of the 21st International Conference on World Wide Web (pp. 119–120). ACM.
- Zhou, Y., & Evans, D. (2011). A large scale study of web password habits. In Proceedings of the 21st USENIX Security Symposium (pp. 1–12). USENIX Association.
- Elahi, A., & Zirak, A. (2024). A comparative study on recommendation algorithms: Online and offline evaluations on a large-scale recommender system. arXiv preprint arXiv:2411.01354.
- Ghattas, M., Mora, A. M., & Odeh, S. (2025). A novel approach for evaluating web page performance based on machine learning algorithms. AI, 6(2), 19. https://doi.org/10.3390/ai6020019

- Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval. Cambridge University Press.
- Peace, P., Abayomi, G., & Bolla, A. (2024). Evaluation metrics for recommender systems: A comprehensive analysis. ResearchGate Preprint.
- Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. Journal of Machine Learning Technologies, 2(1), 37–63.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427–437.
- Abdelnabi, S., Fritz, M., & Rossow, C. (2020). VisualPhishNet: Zero-day phishing website detection by visual similarity. In Proceedings of the 2020 IEEE European Symposium on Security and Privacy (pp. 61–75). IEEE.
- Hong, J. (2012). The state of phishing attacks. Communications of the ACM, 55(1), 74–81.
- Jakobsson, M., & Myers, S. (Eds.). (2006). Phishing and countermeasures: Understanding the increasing problem of electronic identity theft. Wiley-Interscience.
- Liu, W., Huang, G., & Wang, X. (2015). A survey of phishing attack detection and prevention. Journal of Network and Computer Applications, 60, 1–19.
- Wu, M., Miller, R. C., & Garfinkel, S. L. (2006). Do security toolbars actually prevent phishing attacks? In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 601–610). ACM.
- Abdelhamid, N., Ayesh, A., & Thabtah, F. (2014). Phishing detection based on hybrid features. In 2014 IEEE International Conference on Systems, Man, and Cybernetics (pp. 3492–3497). IEEE.
- Basnet, R., Sung, A. H., & Liu, Q. (2012). Learning to detect phishing URLs. International Journal of Research in Computer Science, 2(1), 1–12.
- Marchal, S., Saari, K., Singh, N., & Asokan, N. (2016). Know your phish: Novel techniques for detecting phishing sites and their targets. In 2016 IEEE 36th International Conference on Distributed Computing Systems (pp. 323–333). IEEE.
- Mohammad, R. M., Thabtah, F., & McCluskey, L. (2014). Predicting phishing websites based on self-structuring neural network. Neural Computing and Applications, 25(2), 443–458.
- Whittaker, C., Ryner, B., & Nazif, M. (2010). Large-scale automatic classification of phishing pages. In Proceedings of the Network and Distributed System Security Symposium.
- Abdelnabi, S., Fritz, M., & Rossow, C. (2020). VisualPhishNet: Zero-day phishing website detection by visual similarity. In Proceedings of the 2020 IEEE European Symposium on Security and Privacy (pp. 61–75). IEEE.
- Baki, S., Zeadally, S., & Badra, M. (2017). Phishing detection and prevention: Taxonomy and evaluation. Journal of Network and Computer Applications, 66, 25–35.
- Chiew, K. L., Yong, K. S. C., & Tan, C. L. (2018). A survey of phishing attacks: Their types, vectors and technical approaches. Expert Systems with Applications, 106, 1–20.
- Verma, R., & Das, A. (2017). What's in a URL: Fast feature extraction and malicious URL detection. In Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security (pp. 807–818). ACM.
- Zhang, Y., Hong, J., & Cranor, L. F. (2007). Cantina: A content-based approach to detecting phishing web sites. In Proceedings of the 16th International Conference on World Wide Web (pp. 639–648). ACM.