

---

# Journal of Informatics and Web Engineering

Vol. 4 No. 1 (February 2025)

eISSN: 2821-370X

---

## Social Engineering Threat Analysis Using Large-Scale Synthetic Data

Sellappan Palaniappan<sup>1\*</sup>, Rajasvaran Logeswaran<sup>2</sup>, Shapla Khanam<sup>3</sup>, Pulasthi Gunawardhana<sup>4</sup>

<sup>1</sup>Corporate Office, HELP University, No. 15, Jalan Sri Semantan 1, Off Jalan Semantan, Bukit Damansara 50490 Kuala Lumpur, Malaysia.

<sup>2,3</sup>Faculty of Computing and Digital Technology, HELP University, Persiaran Cakerawala, Subang Bestari, 40150 Shah Alam, Selangor, Malaysia.

<sup>4</sup>Department of Information and Communication Technology, Faculty of Technology, University of Sri Jayewardenepura, Sri Soratha Mawatha, Nugegoda, Sri Lanka.

\*corresponding author: (sellappan.p@help.edu.my, ORCID: 0009-0009-1168-2864)

*Abstract* - We frequently hear news about compromised systems, virus attacks, spam emails, stolen bank account numbers, and loss of money. Safeguarding and protecting digital assets against these and other cyber-attacks are extremely important in our digital connected world today. Many organizations spend substantial amounts of money to protect their digital assets. One type of cyber threat that is rampant these days is social engineering attacks that work on human psychology. These attacks typically persuade, convince, trick and threaten naïve and innocent individuals to divulge sensitive information to the attackers. Consequently, traditional approaches have not been effective or successful in preventing these attack types. In this paper, we propose a machine learning model to detect these types of threats. The model is trained using a large synthetic dataset of 10,000 samples to simulate various types of real-world social engineering threats such as phishing, spear phishing, whaling, vishing, smishing, baiting, and pretexting. Our analysis on attack types, patterns, and characteristics revealed interesting insights. Our model achieved an accuracy of 0.8984 and an F1 score of 0.9253, demonstrating its effectiveness in detecting social engineering attacks. The use of synthetic data overcomes the problem of lack of availability of real-world data due to privacy issues, and is demonstrated in this work to be safe, scalable, ethics friendly and effective.

*Keywords*— Social Engineering, Threats, Phishing Attacks, Machine Learning, Synthetic Data

*Received: 01 August 2024; Accepted: 15 November 2024; Published: 16 February 2025*

*This is an open access article under the [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/) license.*



---

### 1. INTRODUCTION

Cyberthreats are rampant today in our digitally connected world. We often hear of compromised systems, virus attacks, spam emails, stolen account numbers, and loss of money. Safeguarding and protecting digital assets against these and other types of cyber-attacks are vitally important in our increasingly connected world. Many organizations spend huge sums of money to safeguard and protect their digital assets. One type of cyber threat is social engineering attacks like

phishing, whaling and pretexting. Social engineering attacks work on human psychology and typically persuade, convince, trick, and even threaten naïve and innocent individuals to divulge sensitive information to the attackers [1]. Organizations and individuals need to be aware and recognize these types of attacks. It is vitally important for cybersecurity professionals to understand how these social engineering threats manifest and how to detect and prevent such attacks.

This paper presents a machine learning model using synthetic data to detect social engineering threats. Specifically, we seek answers to the following research questions: (1) What are social engineering threats? (2) Can machine learning be used to detect social engineering threats? (3) Can synthetic data be used to simulate social engineering threats? (4) Can machine learning and synthetic data be combined to detect social engineering attacks effectively? Consequently, the research objectives to answer the above research questions are: (1) Identify the different social engineering threats; (2) Generate a synthetic dataset to simulate real-world social engineering attacks; (3) Build a machine learning model using a synthetic dataset; and (4) Evaluate the model's accuracy to effectively detect social engineering threats.

## 2. LITERATURE REVIEW

Several studies were conducted to investigate how social engineering threats are launched. The study in [2] reported a sharp rise in COVID-19-related phishing attacks, stressing that attackers today quickly seize opportunities where they can exploit innocent individuals. Another study [3] showed how attackers used advanced tools including artificial intelligence to launch their phishing and other attacks. These studies stress the need for effective approaches to detect and prevent these threats. A more comprehensive survey on social engineering attacks may be found in [4].

Advanced techniques for detecting social engineering attacks were proposed in [5]. Their study compared the effectiveness of different machine algorithms and found that the use of continual learning methods potentially improve the detection accuracy by overcoming the performance deterioration over time of traditional machine learning algorithms. In [6], it was proposed that organizations design effective training programs to help their employees in detecting phishing attacks. Their findings showed that regular training could significantly enhance the user's ability to recognize social engineering attacks.

Synthetic data could be used in cybersecurity to overcome scarcity of real-world data and privacy concerns. An overview of synthetic data applications across several industries, including cybersecurity and how it can overcome the problem of data availability is described in [7]. Artificial Intelligence (AI) techniques are now commonly used in generating synthetic data for cyberattacks [8], and hence can conversely also be used for training models in detecting social engineering attacks. Besides overcoming data availability and privacy issues, synthetic data (if simulated accurately to reflect real-world scenarios), can be effectively used to scale up and simulate various types of cyber threats.

## 3. RESEARCH METHODOLOGY

Our machine learning model uses synthetic data for detecting and analyzing various social engineering threats. The use of synthetic datasets to simulate real-world cybersecurity threats, including social engineering attacks, is becoming popular [7]. We generated a large synthetic dataset to simulate real-world social engineering attacks and analyzed the attack patterns.

### 3.1 Synthetic Data Generation

The synthetic dataset of 10,000 samples was generated to simulate the following types of social engineering threats: Phishing, Spear phishing, Whaling, Vishing (voice phishing), Smishing (SMS phishing), Baiting and Pretexting. Studies have shown that synthetic datasets can be effective in simulating and detecting real-world phishing emails as well as detect other types of cybersecurity threats [9], [10]. It can also simulate features like sender domains, email subjects, and contents [11].

Predefined lists of common and suspicious domains, legitimate and suspicious email subjects, and content templates were used. Randomization techniques were employed to generate diverse, realistic threat scenarios. For example, email-based threats (phishing, spear phishing, whaling) include sender and recipient email addresses, subjects and

content; voice-based threats (vishing) include phone numbers and script content; SMS-based threats (smishing) include phone numbers and message content; and physical threats (baiting, pretexting) include scenario descriptions. To ensure a realistic distribution of these threat types, suitable weights were assigned to each threat category as suggested by [2], [12].

### 3.2 Implementation

Machine learning is currently transforming the field of cybersecurity by enabling advanced detection, prevention and response mechanisms [13]. The machine learning model was implemented in Python, incorporating algorithms for synthetic data generation, threat analysis, model evaluation, and main module. The pseudocode below outlines the logic of these algorithms.

#### 3.2.1 Algorithm for Synthetic Data Generation

This algorithm (see Algorithm 1) generates a diverse set of synthetic social engineering threats, including various types of phishing, vishing, smishing, and physical threats. It uses random selection and specific generation functions for each threat type to create a comprehensive dataset that mimics real-world attack scenarios.

<b>Algorithm 1: GenerateSyntheticData</b>	
Input: num_samples	<i>number of synthetic threats to generate</i>
Output: dataset	<i>list of synthetic threats</i>
1: Function GenerateSyntheticData(num_samples):	
2:   Initialize empty dataset	<i>create empty dataset</i>
3:   For i = 1 to num_samples:	<i>populate the dataset</i>
4:     threat_type = RandomlySelectThreatType()	<i>randomly select threat type</i>
5:     Switch (threat_type):	<i>generate threat based on the threat type</i>
6:       Case 'phishing', 'spear_phishing', 'whaling':	
7:         threat = GenerateEmailThreat(threat_type)	
8:       Case 'vishing':	
9:         threat = GenerateVoiceThreat()	
10:      Case 'smishing':	
11:         threat = GenerateSMSThreat()	
12:      Case 'baiting', 'pretexting':	
13:         threat = GeneratePhysicalThreat()	
14:      Case 'legitimate':	
15:         threat = GenerateLegitimateEmail()	
16:     Add threat to dataset	
17:   Return dataset	<i>dataset generated</i>
18: Function GenerateEmailThreat(threat_type):	
19:   sender = GenerateSender(threat_type)	<i>function to generate email threats</i>
20:   recipient = GenerateRecipient(threat_type)	
21:   subject = SelectSubject(threat_type)	
22:   content = GenerateContent(threat_type)	
23:   Return EmailThreat(sender, recipient, subject, content, threat_type)	
<i>Similar functions created for GenerateVoiceThreat, GenerateSMSThreat, GeneratePhysicalThreat, and GenerateLegitimateEmail</i>	

#### 3.2.2 Algorithm for Threat Analysis

The threat analysis algorithm as depicted in Algorithm 2, processes the synthetic dataset to identify patterns, trends, and characteristics of the generated threats. It calculates the distribution of different threat types, identifies common elements like suspicious domains and subjects, and analyzes specific scenarios for each threat category.

**Algorithm 2: AnalyzeThreats**

Input: dataset  
Output: threat\_report

*list of synthetic threats*  
*analysis of threat patterns*

1: Function AnalyzeThreats(dataset): *function to analyse the threat*  
 2: Initialize threat\_report *create empty threat report*  
 3: threat\_distribution = CalculateThreatDistribution(dataset) *threat distribution*  
 4: suspicious\_domains = IdentifyTopSuspiciousDomains(dataset) *identify suspicious domains*  
 5: suspicious\_subjects = IdentifyCommonSubjects(dataset) *identify common threat subject lines*  
 6: vishing\_scenarios = AnalyzeVishingScenarios(dataset) *analyse vishing threats*  
 7: smishing\_messages = AnalyzeSmishingMessages(dataset) *analyse smishing threats*  
 8: physical\_scenarios = AnalyzePhysicalScenarios(dataset) *analyse physical threats*  
 9:  
 10: Add all analyses to threat\_report  
 11: Return threat\_report *threat report completed*

12: Function CalculateThreatDistribution(dataset): *function to calculate threat distribution*  
 13: Count occurrences of each threat\_type in dataset  
 14: Calculate percentage for each threat\_type  
 15: Return distribution

*Similar functions created for IdentifyTopSuspiciousDomains, IdentifyCommonSubjects, AnalyzeVishingScenarios, AnalyzeSmishingMessages, AnalyzePhysicalScenarios]*

**3.2.3 Algorithm for Model Evaluation**

This algorithm (see Algorithm 3) evaluates the performance of the machine learning model trained on the synthetic dataset. It splits the data into training and testing sets, trains the model, makes predictions, and calculates various performance metrics including accuracy, precision, recall, and F1 score. The algorithm also generates a confusion matrix to visualize the model's classification performance.

**Algorithm 3: EvaluateModel**

Input: dataset  
Output: performance\_metrics, confusion\_matrix

*list of synthetic threats*  
*performance metrics results*

1: Function EvaluateModel(dataset): *function for performance evaluation*  
 2: X = ExtractFeatures(dataset) *feature extraction*  
 3: y = ExtractLabels(dataset) *label extraction*  
 4: X\_train, X\_test, y\_train, y\_test = SplitData(X, y, test\_size=0.3) *identify training and test data*  
 5:  
 6: model = InitializeModel() *create the model*  
 7: model.Train(X\_train, y\_train) *train the model*  
 8:  
 9: y\_pred = model.Predict(X\_test) *model prediction results*  
 10:  
 11: accuracy = CalculateAccuracy(y\_test, y\_pred) *calculate the performance metrics*  
 12: precision = CalculatePrecision(y\_test, y\_pred)  
 13: recall = CalculateRecall(y\_test, y\_pred)  
 14: f1\_score = CalculateF1Score(precision, recall)  
 15:  
 16: confusion\_matrix = GenerateConfusionMatrix(y\_test, y\_pred) *confusion matrix*

```

17:
18: performance_metrics = {                                consolidate the performance metrics
19:     'Accuracy': accuracy,
20:     'Precision': precision,
21:     'Recall': recall,
22:     'F1 Score': f1_score
23: }
24:
25: Return performance_metrics, confusion_matrix
    
```

### 3.2.4 Algorithm for Main Logic

The main logic algorithm as shown in Algorithm 4 manages the entire process, from data generation to analysis and evaluation. It calls the above algorithms in sequence, generates visualizations based on the results, and outputs the final threat report and model performance metrics.

<b>Algorithm 4: Main logic</b>	
1: Function Main():	
2: dataset = GenerateSyntheticData(10000)	<i>generate dataset</i>
3: threat_report = AnalyzeThreats(dataset)	<i>create threat report</i>
4: performance_metrics, confusion_matrix = EvaluateModel(dataset)	<i>evaluate model performance</i>
5: GenerateVisualizations(dataset, threat_report, confusion_matrix)	<i>plot visualization (optional)</i>
6: OutputResults(threat_report, performance_metrics)	
7: Execute Main()	<i>start the simulation</i>

## 4. RESULTS AND DISCUSSION

Analysis of the 10,000 synthetic social engineering threat samples revealed several key findings regarding the distribution of threat types, characteristics of different attacks, and the performance of our simulated threat detection model. These are discussed below.

### 4.1 Threat Type Distribution

Table 1 shows the distribution of different social engineering threat types in the synthetic dataset. The distribution shows that while legitimate communications form the largest category, phishing and its variants (spear phishing and whaling) collectively account for 48.6% of all communications in our dataset. This is significant as almost half of the communications consists of social engineering threats.

Table 1. Threat Type Distribution.

Threat type	Count	Percentage
Legitimate communications	2986	29.86
Phishing	2524	25.24
Spear phishing	1519	15.19
Whaling	817	8.17
Smishing	769	7.69
Vishing	765	7.65
Pretexting	322	3.22
Baiting	298	2.98

#### 4.2 Threat type analysis

**Phishing:** Tables 2 and 3 show the patterns in terms of top suspicious domains and most common suspicious subjects in phishing attempts. The analysis highlights the prevalence of financial and authority-based themes in phishing attempts.

Table 2. Top Suspicious Domains in Phishing Attempts

Domain	Occurrences
account-verify.net	524
bank-update.info	524
prize-claim.com	515

Table 3. Most Common Suspicious Subjects in Phishing Attempts

Subject	Occurrences
CEO Request	540
Claim Your Prize Now	518
Urgent: Account Suspended	507

**Smishing:** From the results obtained, the smishing results showed a uniform distribution of messages, with each unique message appearing only once in the dataset. This indicates high variability in smishing tactics.

**Vishing:** Table 4 shows the most common vishing scenarios. The analysis reveals approximately equal distribution across the three scenarios.

Table 4. Most Common Vishing Scenarios

Scenario	Occurrences
We are calling about your car's extended warranty.	199
You have won a free vacation! Press 1 to claim your prize.	197
This is your bank's fraud department. We have detected suspicious activity on your account.	196

**Physical:** Table 5 shows the most common physical social engineering scenarios. These are mainly in terms of sending of CD-ROMs (or other physical media) either hand-delivered or by post. The former has also evolved nowadays into other means such as providing malicious QR codes.

Table 5. Most Common Physical Social Engineering Scenarios

Scenario	Occurrences
Promotional CD-ROMs mailed to employees	109
Person claiming to be IT support requesting password	107

#### 4.3 Model performance metrics

The performance of the model was analyzed using the performance metrics indicated in Table 6. From the results obtained, there is indication of strong performance in detecting threats, with high precision (approx. 96%) and relatively high recall (approx. 90%). With an accuracy of approximately 90% and F1 Score of approximately 93%, the model performed well, albeit some threats may have been missed.

#### 4.4 Discussion

The analysis using large-scale synthetic data provides important insights into the nature of social engineering attacks. The prevalence of phishing remains a significant concern, accounting for 25.24% of all communications and 35.98% of all threats. This underscores the continued importance of email security and user education in combating these attacks [6]. Traditional phishing's persistent dominance in the threat landscape further emphasizes the need for ongoing vigilance and improved defence mechanisms in email systems.

Table 6. Threat Detection Model Performance Metrics

Metric	Value
Accuracy	0.8984
Precision	0.9552
Recall	0.8972
F1 Score	0.9253

The analysis also shows a growing sophistication in targeted attacks, with spear phishing (15.19%) and whaling (8.17%), together accounting for a substantial portion of threats. This trend highlights the evolving tactics of attackers who are increasingly launching personalized campaigns aimed at high-value targets [14]. The rise of these attacks requires more nuanced and context-aware defence strategies.

The attack vectors show the presence of smishing (7.69%) and vishing (7.65%) attacks. This finding shows that attackers go beyond traditional email-based methods to target multiple communication channels. Thus, there is a pressing need for comprehensive security awareness training that incorporates a broader range of potential threat vectors [15].

There is also an increasing trend in the exploitation of authority figures in social engineering attacks. The prevalence of "CEO Request" as the most common phishing subject demonstrates how attackers leverage employees' tendency to comply with requests from senior management [16]. This tactic exploits organizational hierarchies and human inclination to respond promptly to authority figures, underscoring the need for clear communication protocols and verification processes within organizations.

Financial motivations continue to drive many social engineering attacks, as evidenced by the frequent use of prize claims and bank-related themes in phishing and vishing attempts [17]. This persistent focus on financial gain suggests that attackers view these methods as lucrative, and this needs stricter financial security measures and awareness training specific to financial-themed threats.

While we focus mostly on digital threats, we must not ignore the physical social engineering threats. Our analysis shows these threats still exist. So, organizations must be equally mindful of non-digital threats and must implement comprehensive security strategies and employee training to combat both digital and non-digital threats [18].

Our simulated threat detection model demonstrated strong performance with an F1 Score of 0.9253. However, the presence of false negatives, indicated by a recall of 0.8972, suggests that some threats may still evade detection. This finding further underscores the importance of continuous improvement in detection algorithms and the implementation of multi-layered security approaches to mitigate the risk of undetected threats.

## 5. LIMITATIONS AND FUTURE WORK

While the proposed model provides valuable insights into social engineering threats, several limitations provide directions for future research. The use of purely simulated data is a primary limitation of the study. Although variability was introduced in the synthetic dataset, real-world social engineering attacks often exhibit greater complexity and unpredictability. The controlled environment of the simulation may not fully capture the nuances of actual attack scenarios [7]. The model also does not capture attack patterns over time, which is important in understanding persistent threats [2].

Another limitation is the absence of actual user interaction and user responses [19]. This lack of human behavioural data limits understanding of how individuals might react to these simulated attacks. Additionally, the current approach may not be sufficient for detecting more sophisticated attacks, which might require more advanced machine learning algorithms.

The proposed model does not consider contextual factors such as organizational culture, industry type, or geographical areas [1]. These elements can significantly influence social engineering attacks, and their absence in the model may limit its real-world applicability. There is a need to build organizational resilience to such cybersecurity threats [20].

Future research could address these limitations in several ways. One way is the employment of advanced machine learning techniques to generate more diverse and realistic attack types. This could help bridge the gap between the synthetic data and the complexity of real-world attacks. Some researchers have suggested that deep learning and ensemble methods could improve threat detection accuracy [21]. Exploring these approaches could potentially enhance the model's ability to detect sophisticated attacks. Another way is to generate time-series data to gain greater insights into persistent cyber threats. This dimension could provide valuable insights on attack patterns and evolution over time.

Incorporating psychological and human-computer interaction factors into the model could also enhance the model and predictive power [6], [17]. This would involve simulating user responses or integrating findings from behavioural studies on social engineering susceptibility.

Conducting a comparative analysis between synthetic data and anonymized real-world datasets (where available) could further improve the threat detection accuracy [7]. This comparison could help validate our synthetic approach and identify areas for improvement. Future work could also enhance the model by using a combination of synthetic and real-world datasets.

Adapting the model to industry-specific social engineering scenarios (e.g., finance, healthcare, government) and cross-cultural variations in social engineering tactics could further enhance its practical applicability. This approach would make the model more relevant and effective for diverse organizational contexts.

## 6. CONCLUSION

This paper presents a machine learning model using a large synthetic data set to simulate various social engineering threats. Analysis showed that phishing and its variants are by far the most common threat, accounting for 48.6% of all simulated threats. The rise in targeted attacks like spear phishing and whaling is highlighted. The rise of smishing and vishing threats also emphasizes the need for greater security awareness and training. While the proposed model has the limitation of not using real-world data, with an F1 score of 0.9253, it demonstrates good potential for detecting social engineering attacks.

This research underscores the dynamic nature of the cybersecurity landscape, particularly in the realm of social engineering. The prevalence of authority exploitation tactics, such as CEO impersonation, reveals a concerning trend that organizations must address through improved communication protocols and employee education. Furthermore, the persistent financial motivations behind many attacks, coupled with the emergence of physical social engineering scenarios, highlight the need for a comprehensive, multi-faceted approach to security. As threat actors continue to evolve their tactics, this study serves as a foundation for future research and development of more sophisticated detection and prevention strategies. By combining advanced machine learning techniques with a nuanced understanding of human behavior and organizational dynamics, we can work towards creating more resilient defenses against the ever-changing landscape of social engineering threats.

## ACKNOWLEDGEMENT

We thank the anonymous reviewers for the careful review of our manuscript.



## FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

## AUTHOR CONTRIBUTIONS

Sellappan Palaniappan: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation;

Rajasvaran Logeswaran: Project Administration, Writing – Review & Editing;

Shapla Khanam: Writing – Review & Editing.

Pulasthi Gunawardhana: Writing – Review & Editing.

## CONFLICT OF INTERESTS

No conflict of interests were disclosed.

## ETHICS STATEMENTS


No ethical issues. Synthetic data was used in the work.




## REFERENCES

- [1] K. Krombholz, H. Hobel, M. Huber, and E. Weippl, “Advanced social engineering attacks”, *Journal of Information Security and Applications*, vol. 22, 2015, pp. 113-122.
- [2] H. S. Lallie, L. A. Shepherd, J. R. Nurse, A. Erola, G. Epiphaniou, C. Maple, and X. Bellekens, “Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic”, *Computers and Security*, vol. 105, 2021, p. 102248.
- [3] R. Kaur, D. Gabrijelčič, and T. Klobučar, “Artificial intelligence for cybersecurity: Literature review and future research directions”, *Information Fusion*, vol. 97, no. C, 2023, doi: 10.1016/j.inffus.2023.101804.
- [4] S. K. Birthriya, P. Ahlaway, and A. K. Jain, “A comprehensive survey of social engineering attacks: Taxonomy of attacks, prevention, and mitigation strategies”, *Journal of Applied Research*, 2024, pp. 1–49, doi: 10.1080/19361610.2024.2372986.
- [5] A. Ejaz, A. N. Mian, and S. Manzoor, “Life-long phishing attack detection using continual learning”, *Scientific Reports*, vol. 13, 2023, p. 11488. DOI: 10.1038/s41598-023-37552-9.
- [6] P. Kumaraguru, Y. Rhee, A. Acquisti, L. F. Cranor, J. Hong, and E. Nunge, “Protecting people from phishing: the design and evaluation of an embedded training email system”, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2007, pp. 905-914, doi: 10.1145/1240624.1240760.
- [7] E. de Cristofaro, “Synthetic data: Methods, use cases, and risks”. *IEEE Security and Privacy*, vol. 22, no. 3, 2024, pp. 62–67, doi: 10.1109/MSEC.2024.3371505.
- [8] M. Concannon, “AI in social engineering: the next generation of cyber threats”, *Ntiva*, 2024, <https://www.ntiva.com/blog/ai-social-engineering-attacks>.
- [9] S. Gupta, M. Pritwani, A. Shrivastava, M. Mohana, M., Moharir, and A. Kumar, "A comprehensive analysis of social engineering attacks: from phishing to prevention - Tools, techniques and strategies", in *Proceedings of the 2024 Second*

- International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, 2024, pp. 1-8, doi: 10.1109/ICoICI62503.2024.10696444.
- [10] V. Kumar, and D. Sinha, "Synthetic attack data generation model applying generative adversarial network for intrusion detection", *Computers and Security*, vol. 125, 2023, p. 103054, doi: 10.1016/j.cose.2022.103054.
- [11] F. Salahdine, and N. Kaabouch, "Social engineering attacks: A survey", *Future Internet*, vol. 11, no. 4, 2019, p. 89.
- [12] A. Aleroud, and L. Zhou, "Phishing environments, techniques, and countermeasures: A survey", *Computers and Security*, vol. 68, 2017, pp. 160-196.
- [13] N. K. Thawait, "Machine learning in cybersecurity: Applications, challenges and future directions", *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, vol. 10, no. 3, 2024, pp. 16-27, doi: 10.32628/CSEIT24102125.
- [14] F. P. E. Putra, Ubaidi, A. Zulfikri, G. Arifin, and R. M. Ilhamsyah, "Analysis of phishing attack trends, impacts and prevention methods: literature study", *Brilliance: Research of Artificial Intelligence*, vol. 4, no. 1, 2024, pp. 413-421, doi: 10.47709/brilliance.v4i1.435.
- [15] H. C. Pham, D. D. Pham, L. Brennan, and J. Richardson, "Information security and people: A conundrum for compliance", *Australasian Journal of Information Systems*, vol. 21, 2017, doi: 10.3127/ajis.v21i0.1321.
- [16] R. Bhakta, and I. G. Harris, "Semantic analysis of dialogs to detect social engineering attacks", in *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)*, 2015, pp. 424-427, doi: 10.1109/ICOSC.2015.7050843.
- [17] Huang, K., Siegel, M., and Madnick, S. (2018). "Systematically understanding the cyber attack business: a survey". *ACM Computing Surveys (CSUR)*, 51(4), 1-36. DOI: 10.1145/31996.
- [18] R. Heartfield, and G. Loukas, "A taxonomy of attacks and a survey of defence mechanisms for semantic social engineering attacks", *ACM Computing Surveys (CSUR)*, vol. 48, no. 3, 2015, pp. 1-39, doi: /10.1145/2835375.
- [19] A. Vishwanath, B. Harrison, and Y. J. Ng, "Suspicion, cognition, and automaticity model of phishing susceptibility", *Communication Research*, vol. 45, no. 8, 2018, pp. 1146-1166, doi: 10.1177/0093650215627483.
- [20] T. Munusamy, and T. Khodadi, "Building cyber resilience: Key factors for enhancing organizational cyber security", *Journal of Informatics and Web Engineering*, vol. 2, no. 2, 2023, pp. 59–71, doi: 10.33093/jiwe.2023.2.2.5.
- [21] S. Mushtaq, T. Javed, and M. Mohd Su'ud, "Ensemble learning-powered URL phishing detection: A performance driven approach", *Journal of Informatics and Web Engineering*, vol. 3, no. 2, pp. 134–145, doi: 10.33093/jiwe.2024.3.2.10.

## BIOGRAPHIES OF AUTHORS

	<p><b>Sellappan Palaniappan</b> is a Professor of Information Technology at HELP University, Malaysia. With over 30 years of academic experience, his current research interests are in the application of artificial intelligence and machine learning in diverse domains like cybersecurity, data analytics, healthcare, biotechnology, retail, education, agriculture, logistics, and sustainable development initiatives. He is also interested in quantum physics, DNA, neuroscience, and energy, frequency, and vibration for health and wholeness. He has published more than 100 scholarly research papers and authored several widely-used IT textbooks for college and university students. He may be contacted at <a href="mailto:sellappan.p@help.edu.my">sellappan.p@help.edu.my</a>.</p>

	<p><b>Rajasvaran Logeswaran, SMIEEE</b> is a Professor and Dean of Computing and Digital Technology at HELP University, Malaysia. With over 25 years of academic experience, his research interests are in medical image processing, artificial intelligence, data science and cybersecurity, with over 180 publications in books, peer-reviewed journals and international conference proceedings. He actively serves as a speaker at many international conferences, as well as volunteers as a judge in STEM and innovation competitions for schools at the local and international levels. He may be contacted at logeswaran.nr@help.edu.my.</p>
	<p><b>Shapla Khanam</b> is a Senior Lecturer at HELP University, Malaysia. She received her PhD from University of Malaya, where she was a researcher and helped students to develop Cybersecurity solutions using Deep Learning and Machine Learning for Internet of Things (IoT). Her research interests include Intrusion Detection, Network Security, Machine Learning, Deep Learning, IoT and WSNs. She has published several articles in journals and presented at international conferences. She may be contacted at shapla.k@help.edu.my.</p>
	<p><b>Pulasthi Gunawardhana</b> is a Senior Lecturer at the University of Sri Jayewardenepura. His research focuses on gamification/educational games, web technologies, video technologies, and multimedia technologies in diverse domains. He has written many publications in the domain of gamification, video technologies, games, and educational psychology. Apart from that he actively serves as a speaker at many international conferences and has written many reviews on digital economies as well. He can be contacted at email: pulasthi@sjp.ac.lk</p>