Human Factors in Cybersecurity, Vol. 127, 2024, 236–246
https://doi.org/10.54941/ahfe1004783

Cracking the Code: How Social Media and Human Behaviour Shape Cybersecurity Challenges

Foteini Markella Petropoulou[1] and Emmanuel Varouchas[2]

[1]Zelus IKE, Greece & Deree – The American College of Greece, Athens, Greece
[2]Deree – The American College of Greece, Athens, Greece

ABSTRACT

In an era dominated by digital connectivity, where people are more connected than ever, understanding how humans can securely interact is crucial. This paper delves into the intricate relationship between social engineering and social media, unravelling the multifaceted dimensions that underscore the human aspects of cybersecurity. As technological defences evolve, adversaries increasingly exploit the vulnerabilities inherent in human behaviour (Wang et al., 2020), making it imperative to dissect the interplay between social engineering tactics and the pervasive influence of social media platforms. The study begins by scrutinizing the psychological underpinnings that make individuals susceptible to social engineering attacks, emphasizing the intricate relationship between trust, curiosity, and social connectivity (Albladi & Weir, 2020). Through a comprehensive critical analysis of real-world examples people encounter in their day-to-day lives, the paper exposes the diverse strategies employed by malicious actors to manipulate human cognition and breach organizational defences. This examination not only dissects the intricacies of phishing, pretexting, and impersonation but also sheds light on the role of emotional triggers and cognitive biases that amplify the effectiveness of these tactics (Wang, Zhu, & Sun, 2021). A significant portion of the paper is dedicated to understanding the role social media plays when it comes to social engineering. The pervasive nature of social media platforms provides a fertile ground for threat actors to extract personal information, exploit social connections, and craft tailored attacks. The paper navigates through the intricate web of privacy erosion, information oversharing, and the amplification of social influence, emphasizing how these factors contribute to the efficacy of social engineering endeavours (Albladi & Weir, 2020). Furthermore, the study explores the role of emerging technologies, such as artificial intelligence and machine learning, in launching social engineering attacks, posing new challenges to the human-centric cybersecurity aspects. To address the ever-changing terrain of social engineering, these emerging technologies advocate for a proactive and flexible strategy that combines technological defences with a solid understanding of human behaviour. In conclusion, this paper elucidates the critical relationship between social engineering, social media, and cybersecurity. By dissecting psychological vulnerabilities and real-world examples, it underscores the intricate tactics employed by adversaries to exploit human behaviour. Emphasizing the role of trust, curiosity, and social connectivity, the study unveils the amplifying effect of emotional triggers and cognitive biases. Focusing on social media's pervasive influence, the paper highlights how platforms contribute to privacy erosion and information exploitation. Acknowledging the challenges posed by emerging technologies, it advocates for a dynamic cybersecurity strategy grounded in both technology and an acute understanding of human behaviour.

Keywords: Social engineering, Social media, Human behaviour, Cybersecurity, Emerging technologies, Artificial intelligence

INTRODUCTION

In the digital age, the intersection of social engineering and social media represents a critical frontier in cybersecurity. As technological defences become more sophisticated, adversaries shift their focus to exploiting human vulnerabilities, a factor that underscores the importance of understanding the human dimensions of cybersecurity (Wang et al., 2020). This paper explores how human behaviour can render individuals susceptible to social engineering tactics, facilitated by the widespread use of social media platforms. By taking into consideration real-world examples that occur daily, such as phishing and pretexting (Krombholz et al., 2015), it is important to highlight how emotional triggers and cognitive biases are manipulated by threat actors to breach security defences (Albladi & Weir, 2020). Furthermore, it considers the implications of emerging technologies like artificial intelligence in enhancing these social engineering attacks, presenting new challenges to cybersecurity measures that must be both technologically advanced and deeply rooted in an understanding of human behaviour. Finally, it explores how we can take advantage of these emerging technologies to identify and prevent social engineering attacks in social networks in the future.

LITERATURE REVIEW

The burgeoning field of social engineering in cybersecurity, particularly in the context of social media platforms, offers a critical lens through which the dynamics of modern cyber threats can be understood. This review synthesizes the existing literature on the nature, mechanisms, and evolution of social engineering attacks, emphasizing their occurrence on social media and the influence of emerging technologies like artificial intelligence.

Social engineering attacks exploit human vulnerabilities, a concept that forms the core of cybersecurity challenges in the digital era (Wang, Zhu and Sun, 2021). These attacks, as highlighted in the literature, are characterized not by their technical sophistication, but by their psychological manipulation, leveraging trust, curiosity, and social connectivity (Wang, Sun and Zhu, 2020). This human-centric approach to cybersecurity is especially pertinent in the context of social media, where personal information is readily accessible and social interactions are inherently trusted.

According to Jain, threats are divided into three categories i.e. conventional threats, modern threats, and targeted threats. Conventional threats include threats that users have been experiencing from the beginning of the social network. Modern threats are attacks that use advanced techniques to compromise accounts of users and targeted attacks are attacks that are targeted on some particular user which can be committed by any user for varied personal vendettas (Jain et al., 2021).

A significant focus of our research lies in the intersection of social engineering and social media platforms. Social media's pervasive nature not only amplifies traditional social engineering tactics but also introduces new dimensions of threat (Chetioui et al., 2022). The platforms provide a fertile ground for attackers to harvest personal data, exploit social networks, and tailor deceptive strategies, making them a critical area of study (Albladi and Weir, 2020).

Emerging technologies, particularly artificial intelligence, are reshaping the landscape of social engineering attacks. AI's capability to analyse vast amounts of data and mimic human interactions poses new challenges in cybersecurity. The literature emphasizes the need for a proactive and adaptive approach to cybersecurity that combines technological advancements with an acute understanding of human behaviour (Syafitri et al., 2022).

In conclusion, the literature presents a comprehensive view of social engineering attacks, with a particular focus on their manifestation on social media platforms and the emerging challenges posed by technologies like artificial intelligence. This review emphasizes the need for understanding human behaviour in the digital space and the development of sophisticated, human-centric cybersecurity strategies and models to help defend our privacy online.

Defining Social Engineering

From the dawn of recorded history, we see one account after another of humans tricking, duping, conning, or scamming one another. On the surface, there might not be much that is brand new when it comes to social engineering, but that does not mean that nothing ever changes (Hadnagy, 2018). Social engineering in cybersecurity refers to a sophisticated form of attack that differs fundamentally from traditional hacking methods (Wang, Sun and Zhu, 2020). Unlike attacks that exploit technical vulnerabilities, social engineering targets the human element, leveraging psychological manipulation to gain unauthorized access or information (Salahdine and Kaabouch, 2019). This approach to cybersecurity breaches underscores the shift in focus from technological loopholes to human psychology, a trend increasingly relevant in the digital age where interpersonal interactions are often mediated through technology (Chetioui et al., 2022). Influence, persuasion and deception are key elements in the profile of a social engineer. About influence, Cialdini states the "Universal Principles of Influence" which are adopted by social engineers: Reciprocation, Commitment and Consistency, Social Proof, Liking, Authority, Scarcity, Unity (Cialdini, 2021).

Following the methodology adapted from Wang et al. (2021), a domain ontology for social engineering was developed. This ontology serves as a structured framework to understand the multifaceted nature of social engineering attacks. It details the core concepts that constitute this domain, providing a comprehensive understanding of the attacker's motivations, strategies, and the vulnerabilities they exploit (Krombholz et al., 2015). This ontological approach was crucial for dissecting the complexities of social engineering, especially as it can further evolve in the context of social media.

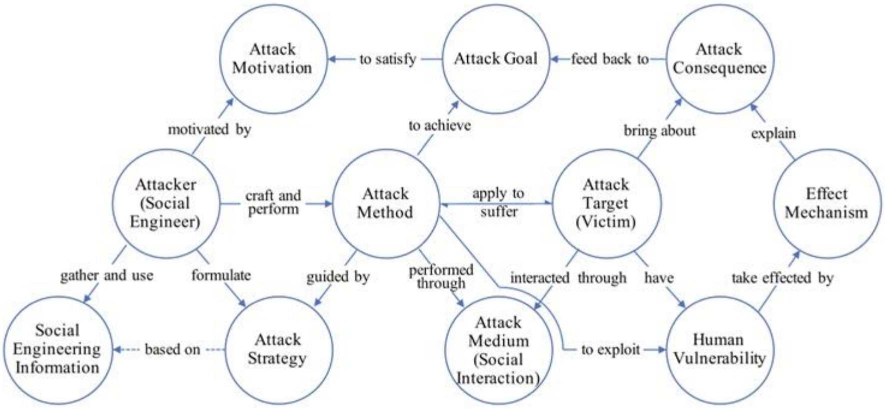

Figure 1: The domain ontology of social engineering in cybersecurity (Wang et al., 2021).

The attack cycle in social engineering is characterized by several stages: the attacker's motivation, information gathering, formulating an attack strategy, execution through a chosen medium, and exploiting targeted vulnerabilities (Wang, Zhu and Sun, 2021). This cycle is not linear but rather an iterative process where the consequences of the attack feedback into refining the attacker's strategies.

The Role of Social Media in Social Engineering Attacks

In the digital landscape, social media platforms have emerged as fertile grounds for social engineering attacks, capitalizing on the unique combination of widespread connectivity and the rich tapestry of personal information available (Chetioui et al., 2022). According to Statista, the number of social media users worldwide will hit 5.85 billion in year 2027.

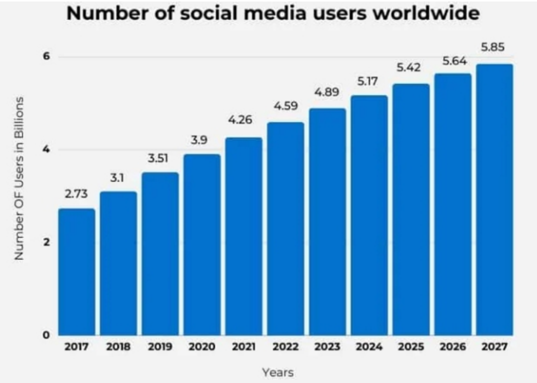

Figure 2: Number of social media users worldwide (Statista, 2023).

Social media inherently fosters a sense of trust and familiarity among its users, making them susceptible to attacks. Attackers exploit this trust, as individuals are more likely to respond to requests or click on links from someone within their social network. This exploitation of trust is a cornerstone in social engineering tactics such as phishing, where malicious actors disguise their true identity under the guise of a familiar contact.

The culture of oversharing on social media provides attackers with a wealth of personal information, which can be used to tailor attacks. Details like employment history, school, location, hobbies, and even family connections give attackers ammunition to craft believable narratives, increasing the likelihood of a successful attack. The sheer volume of personal information available can also lead to information overload, where critical thinking is often overshadowed by the habitual nature of interacting with content online. Social engineering attacks on social media often leverage principles of social proof and authority (Chetioui et al., 2022). Attackers may create fake profiles or hijack existing ones to post endorsements or requests, manipulating users through the influence of apparent authority figures or peers. This tactic preys on the human tendency to comply with requests from authority figures or to follow the actions of a larger group. Additionally, emotional manipulation is another tactic used in these attacks. By crafting messages or scenarios that evoke strong emotions like fear, excitement, or curiosity, attackers can prompt impulsive actions from the target, such as clicking on a malicious link or divulging confidential information (Krombholz et al., 2015). The interactive and visually rich environment of social media platforms makes them particularly effective channels for such emotionally charged attacks. Moreover, the viral nature of social media allows for rapid dissemination of social engineering attacks. A single successful attack can quickly spread, as users unwittingly share malicious links or information with their networks (Chetioui et al., 2022). This amplification effect not only increases the reach of the attack but also lends credibility to it, as the content appears to be endorsed by multiple sources. However, the detection and prevention of social engineering attacks on social media are challenging due to the dynamic and user-generated nature of the content. Traditional cybersecurity measures that focus on technical vulnerabilities are often inadequate in identifying and mitigating these socially engineered threats. Therefore, there is a growing need for cybersecurity strategies that combine technological solutions with user education and awareness to combat social engineering attacks effectively launched through these platforms. Understanding the mechanics of social media is crucial for developing effective countermeasures and fostering a safer digital social environment.

Emerging Technologies as an Enabler for Social Engineering Attacks

The integration of emerging technologies such as artificial intelligence (AI) and machine learning (ML) into the domain of social engineering marks a significant evolution in cyber threats. Additionally, it generates a new breed of cyberattacks titled “AI-enhanced social engineering attacks”. AI’s capacity to analyse vast datasets and emulate human behaviour has notably enhanced the complexity and effectiveness of social engineering attacks. Sophisticated AI algorithms now enable attackers to craft personalized and contextually relevant deceptive messages, significantly increasing the likelihood of successful phishing attacks and other forms of social manipulation (King et al., 2019).

Additionally, machine learning techniques are increasingly used to identify potential targets on social media platforms. These technologies analyse online behaviour, preferences, and network connections to pinpoint individuals who may be more susceptible to social engineering tactics. This targeted approach allows attackers to focus their efforts more efficiently, increasing the overall effectiveness of their campaigns. Automation has also expanded the scope of social engineering attacks. Automated bots and scripts can conduct extensive phishing campaigns and create fake social media profiles with minimal human oversight (Huber et al., 2009). This scalability of attacks presents a major threat, reaching a wider audience and increasing the potential for widespread impact. In essence, the rise of emerging technologies in the context of social engineering presents a dual reality. While these technologies offer attackers more sophisticated methods to execute and scale their attacks, they also hold the key to developing advanced, intelligent cybersecurity strategies. The bottom line lies in harnessing these technologies effectively, ensuring that cybersecurity defences evolve alongside the constantly changing threat landscape.

Model Proposal: Using Emerging Technologies to Our Advantage

AI-enhanced social engineering attacks are changing the cybercrime landscape and consequently, the challenge in detecting and mitigating this type of attack is significant. Traditional cybersecurity measures, designed to detect known patterns and signatures, struggle against these continuously evolving and adapting AI-driven threats. This challenge necessitates a shift toward more dynamic and intelligent cybersecurity defences (Zheng et al., 2022), capable of identifying and responding to sophisticated attacks in real-time. However, these emerging technologies also provide potent tools for cybersecurity defence. AI and machine learning can be utilized to detect unusual behaviour patterns, identify communication anomalies, and anticipate potential attacks (Salahdine and Kaabouch, 2019). They can help enable the development of proactive defence mechanisms, including real-time monitoring and automated response systems, thus enhancing the capability to thwart social engineering attacks effectively.

Drawing on the model developed by Aun et al. (2023), through screening and analysing social media posts using sentiment analysis and risk assessments, we can determine the malicious factor behind them.

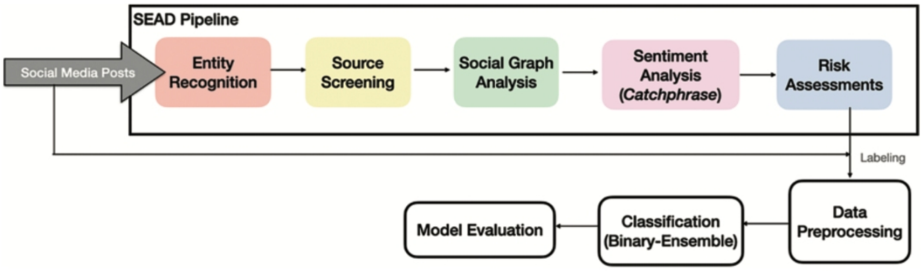

Figure 3: The social engineering attack classifications pipeline (Aun et al., 2023).

Despite the social engineering attempts occurring via social media posts, a risk always lies within the context of direct messages and communications in social media channels. Phishing attempts and spam accounts are on the rise and can be identified in most users' direct messages. These attempts can be even harder to identify since they can be disguised as seemingly innocent messages from a person you know or an account that portrays another user. To identify such attempts, we are recommending a variation of Aun's model, adapted in such a way that instead of social media posts, take as input the text messages and conversations of users that are exchanged through the messaging service of the platform which also allows the exchange of multiple types of media like audio, photographs, and video. Wang's ontology, focused on behaviour pattern recognition and human dynamic behaviour detection, aligns well with Aun's social engineering attack classification pipeline. Integrating these into the proposed model could significantly enhance our ability to detect and respond to AI-enhanced attacks. Wang's framework provides a structured approach to identify nuanced behavioural cues that signify potential threats, complementing Aun's pipeline which processes textual content from social media for malicious intent. By synthesizing these methodologies, the enhanced model could leverage machine learning to interpret subtle indicators of deception across different communication mediums, ensuring robust defences against sophisticated social engineering tactics. As a result, this integration facilitates a real-time, dynamic cybersecurity defence capable of adapting to the evolving landscape of cyber threats.

Then, through behaviour pattern recognition and human dynamic behaviour detection and classification (Wang et al., 2019), the model could identify and flag potential social engineering attacks. By flagging potentially harmful to the user messages, the user can re-evaluate and take caution when interacting or replying to the message. Utilizing AI to analyse patterns in language and behavior, as detailed by Wang and Aun, helps in pre-emptively identifying risks before they manifest into breaches.

To further refine this approach, integrating natural language processing (NLP) algorithms with machine learning models can offer a sophisticated way to analyse the subtleties in language that may indicate malicious intent. NLP can parse and understand the nuances of human communication, detecting indicators of deception or manipulation that go beyond simple keyword recognition (Sawa et al., 2016). This capability is crucial for identifying sophisticated social engineering attempts that may not use overtly suspicious language.

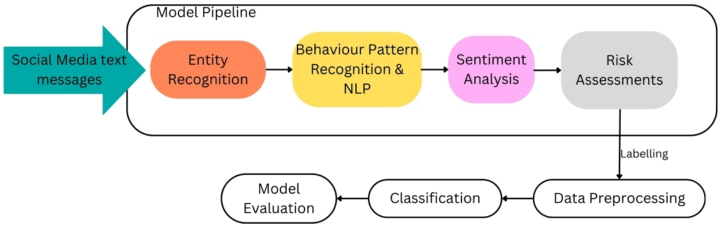

Figure 4: The proposed model pipeline.

The combination of these approaches allows for a layered defence strategy, where behaviour analytics enrich the textual analysis, providing a comprehensive overview of potential threats. This not only enhances the detection capabilities but also refines the response strategies, ensuring that emerging technologies are used to their full advantage to secure digital interactions.

There have been various research projects demonstrating the need for an automated system to recognize social engineering attacks. Many of these projects show proof of concept model using a variety of different techniques, most commonly Natural Language Processing and Machine Learning methods like Artificial Neural Networks (Lansey et al., 2019). Whilst these exist, there is a lack of evidence that the theory has been implemented and proved working, with very little evidence of the rate of success (Tsinganos et al., 2018 and Mouton et al., 2018). As Lansey and his team are proposing in their method, to determine if a dialog between two interlocutors is a Social Engineering attack, certain criteria, or else known as features, need to be chosen. Then, by using advanced techniques such as decision trees or neural networks, the weight of each feature can be calculated programmatically to determine which features carry the most importance regarding whether or not a social engineering attack is taking place (Lansey et al., 2019). The proposed method has been evaluated using a real and a semi-synthetic dataset and can detect social engineering attacks with very high accuracy.

Another approach to detect a social engineering attack in message exchange on social media comes from Sawa and his team. This approach, "uses natural language processing techniques to detect questions and commands, and extract their likely topics. Each extracted topic is compared against a topic blacklist to determine if the question or command is malicious. Our approach is generally applicable to many attack vectors since it relies only on the dialog text. We have applied our approach to analyse the transcripts of several attack dialogs and we have achieved high detection accuracy and low false positive rates in our experiments" (Sawa et al., 2016).

Based on the above, we can say that academic research and projects are conducted aiming at developing novel approaches and models which need to be more thoroughly tested and eventually serve as the basis for improving the security and preventing social engineering attacks on social media.

In conclusion, the battle against social engineering attacks, especially those facilitated through social media, requires a dynamic and multifaceted approach. By leveraging emerging technologies such as AI, machine learning, and NLP, we can develop more effective defence mechanisms that are adaptive, intelligent, and capable of countering sophisticated threats in real time. However, the success of these technologies is maximized when coupled with ongoing user education and awareness. Together, these strategies form a comprehensive defense model that can significantly mitigate the risk of social engineering attacks, protecting users and organizations in the increasingly interconnected digital world.

Future Research & Development Prospects

The findings of this paper have triggered innovative directions for the future which are grouped into two categories: those related to implementation and practical application of the proposed model, and those related to future research initiatives beyond the scope of this study. Concerning the first category, the model could inform and recommend additional functionality to existing Threat Intelligence Platforms like, IBM X-Force Exchange or Anomali ThreatStream. The additional functionality will assist social media users in the evaluation of potentially harmful text messages and in responding accordingly. We envision that if social media platforms implement the above for the benefit of users, then this would be a significant strike on social engineers' devious intentions. Concerning the second category, we would complement the literature review with a structured questionnaire as an instrument for quantitative survey and analysis to be distributed to undergraduate students in higher education. The results would be used to further enhance the proposed model of this paper.

CONCLUSION

The conclusion of the research paper emphasizes the crucial interplay between human behaviour, social media, and cybersecurity challenges. It highlights the importance of understanding psychological vulnerabilities to bolster cybersecurity measures. The paper suggests integrating technological solutions with insights into human behaviour to create a more effective defence against cyber threats lurking in the world of social media. The major contribution of this paper is a model combining an existing social engineering attack classifications pipeline with a pattern recognition ontology. This multifaceted approach, combining a proposed model with emerging as well as more advanced security technologies, is deemed essential for addressing attacks in the evolving landscape of cybersecurity in a more scientific manner.

ACKNOWLEDGMENT

The authors would like to acknowledge the financial support provided for the following projects: 'Collaborative, Multi-modal and Agile Professional Cybersecurity Training Program for a Skilled Workforce In the European Digital Single Market and Industries' (CyberSecPro) project, which has received funding from the European Union's Digital Europe Programme (DEP) under grant agreement No. 101083594. The views expressed in this paper represent only the views of the authors and not of the European Commission or the partners in the above-mentioned project. Additionally, we acknowledge Deree - American College of Greece, School of Business and Economics, for the financial support of this conference paper presentation.

REFERENCES

Albladi, S. M. and Weir, G. R. S. (2020). Predicting individuals' vulnerability to social engineering in social networks. Cybersecurity, 3(1). doi: https://doi.org/10.1186/s42400-020-00047-5.

Aun, Y., Gan, M.-L., Haliza Binti Abdul Wahab, N. and Hock Guan, G. (2023). Social engineering attack classifications on social media using deep learning. Computers, Materials & Continua, 74(3), pp. 4917–4931. doi: https://doi.org/10.32604/cmc.2023.032373.

Chetioui, K., Bah, B., Alami, A. O. and Bahnasse, A. (2022). Overview of Social Engineering Attacks on Social Networks. Procedia Computer Science, [online] 198, pp. 656–661. doi: https://doi.org/10.1016/j.procs.2021.12.302.

Cialdini, R. B. (2007). Influence : The psychology of persuasion. S. L.: Harperbusiness.

Hadnagy, C. (2018). Social engineering : The science of human hacking. [online] Indianapolis, In Wiley. Available at: https://www.wiley.com/en-us/Social+Engineering%3A+The+Science+of+Human+Hacking%2C + 2nd+Edition-p-9781119433385.

Huber, M., Kowalski, S., Nohlberg, M. and Tjoa, S. (2009). Towards Automating Social Engineering Using Social Networking Sites.2009 International Conference on Computational Science and Engineering. doi: https://doi.org/10.1109/cse.2009.205.

Jain, A. K., Sahoo, S. R. and Kaubiyal, J. (2021). Online social networks security and privacy: comprehensive review and analysis. Complex & Intelligent Systems, [online] 7(5). doi: https://doi.org/10.1007/s40747-021-00409-7.

King, T. C., Aggarwal, N., Taddeo, M. and Floridi, L. (2019). Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions. Science and Engineering Ethics, 26, pp. 89–120. doi: https://doi.org/10.1007/s11948-018-00081-0.

Krombholz, K., Hobel, H., Huber, M. and Weippl, E. (2015). Advanced social engineering attacks. Journal of Information Security and Applications, [online] 22, pp. 113–122. doi: https://doi.org/10.1016/j.jisa.2014.09.005.

Merton Lansley, N. Pliatsikas and Kapetanakis, S. (2019). SEADer: A Social Engineering Attack Detection Method Based on Natural Language Processing and Artificial Neural Networks. pp. 686–696. doi: https://doi.org/10.1007/978-3-030-28377-3_57.

Mouton, F., Nottingham, A., Leenen, L. and Venter, H. S. (2018). Finite State Machine for the Social Engineering Attack Detection Model: SEADM. SAIEE Africa Research Journal, 109(2), pp. 133–148. doi: https://doi.org/10.23919/saiee.2018.8531953.

Salahdine, F. and Kaabouch, N. (2019). Social Engineering Attacks: A Survey. Future Internet, 11(4), p. 89. doi: https://doi.org/10.3390/fi11040089.

Sawa, Y., Bhakta, R., Harris, I. G. and Hadnagy, C. (2016). Detection of Social Engineering Attacks Through Natural Language Processing of Conversations. [online] IEEE Xplore. doi: https://doi.org/10.1109/ICSC.2016.95.

Syafitri, W., Shukur, Z., Mokhtar, U. A., Sulaiman, R. and Ibrahim, M. A. (2022). Social Engineering Attacks Prevention: A Systematic Literature Review. IEEE Access, [online] 10(5), pp. 39325–39343. doi: https://doi.org/10.1109/ACCESS.2022.3162594.

Tsinganos, N., Sakellariou, G., Fouliras, P. and Mavridis, I. (2018). Towards an Automated Recognition System for Chat-based Social Engineering Attacks in Enterprise Environments. Proceedings of the 13th International Conference on Availability, Reliability and Security. doi: https://doi.org/10.1145/3230833.3233277.

Wang, S., Gao, J. Z., Lin, H., Shitole, M., Reza, L. and Zhou, S. (2019). Dynamic Human Behavior Pattern Detection and Classification. 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService). doi: https://doi.org/10.1109/bigdataservice.2019.00028.

Wang, Z., Sun, L. and Zhu, H. (2020). Defining Social Engineering in Cybersecurity. IEEE Access, 8, pp. 85094–85115. doi: https://doi.org/10.1109/access.2020.2992807.

Wang, Z., Zhu, H., Liu, P. and Sun, L. (2021). Social engineering in cybersecurity: a domain ontology and knowledge graph application examples. Cybersecurity, 4(1). doi: https://doi.org/10.1186/s42400-021-00094-6.

Wang, Z., Zhu, H. and Sun, L. (2021). Social Engineering in Cybersecurity: Effect Mechanisms, Human Vulnerabilities and Attack Methods. IEEE Access, 9, pp. 11895–11910. doi: https://doi.org/10.1109/access.2021.3051633.

Zheng, Y., Li, Z., Xu, X. and Zhao, Q. (2021). Dynamic defenses in cyber security: Techniques, methods and challenges. Digital Communications and Networks. doi: https://doi.org/10.1016/j.dcan.2021.07.006.