International Journal of Data Science and Advanced Analytics (ISSN: 2563-4429)

Gamified Tailored Roleplay Story-based Phishing Awareness Training

William Wijaya[1], Intan Farahana Kamsin[2], Zety Marlia Zainal Abidin[3] and Hemalata Vasudavan[4]
[1'2'3'4]Asia Pacific University of Technology and Innovation, Technology Park Malaysia, Bukit Jalil, Kuala Lumpur, Malaysia
[1]tp059360@mail.apu.edu.my, [2]intan.farahana@staffemail.apu.edu.my, [3]zety@staffemail.apu.edu.my, [4]hemalata@staffemail.apu.edu.my

***Abstract*— Nowadays, the world has transformed into a digitalize era where more and more people are getting access to the internet. Simultaneously, the number of cybercrimes is increasing as well. One of the top number of cases is phishing attack. To tackle this problem, phishing awareness trainings were developed, but they were not effective. This research aims to propose a solution using gamification, AI, and roleplay story element to conduct the training. To carry out the research questions, quota sampling would be used to discover the appropriate game components and context scenarios, and snowball sampling would be used to get suitable data which could be used for learning analytics to craft new scenarios. To sum up, phishing attack is one of the most concerning issues in this digital era and effective solution is yet to be found. Thus, this paper aims to propose a solution to tackle this problem.**

***Index Terms*— Artificial Intelligence, Gamification, Phishing, Security Awareness, Security Training**

1. Introduction

The advancement of technology has changed the way human lives. Humans' access to the internet has been easier than ever. This could be seen from the growth of internet users by year. According to Stanton [1], as of 2018, the total number of internet users globally has reached more than 50% of the world population. Figure 1 shows the internet global usage percentage from 2009–2018. Simultaneously, the number of cybercrimes being reported is growing as well. Based on statistics by the Internet Crime Complaint Center [2], in America alone, there were 847,376 complaints filed in 2021, which include various types of cyber-attacks, namely ransomware, business e-mail compromise (BEC), and frauds. Figure 2 shows the number of cybercrime complaints in America between 2017–2021.

One of the infamous cyber-attacks is social engineering. From a cybersecurity perspective, social engineering is the act of finding a loophole in humans through either passive or active interaction to breach cybersecurity [3]. Social engineering itself could be broken down to various attacks either technical or social. One of them is phishing attack [4]. Further report by IC3 [2] indicated that phishing attack has the highest number of cases for the last 3 years. Figure 3 shows the statistics of cyber-attacks based on various types for the last 5 years. Another report by Anti-Phishing Working Group [5] in 2021, indicated that the number of phishing attacks has tripled from early 2020 whereby December 2021 was marked as the most attacks of all APWG's history in a single month, summing up to 316,747 attacks. Consequently, a lot of solutions has been developed to mitigate this issue. These solutions include giving cybersecurity awareness training and utilizing machines for detection [6].

The rest of the paper is structured as follows. Section 2 reviews the existing literature and similar system. Section 3 discusses the problem with current techniques in mitigating phishing attacks. Section 4, 5, 6, and 7 explain about this research aims, objectives, significance, and methodology. Section 8 provides the overview of the proposed system. Finally, section 9 concludes the paper.

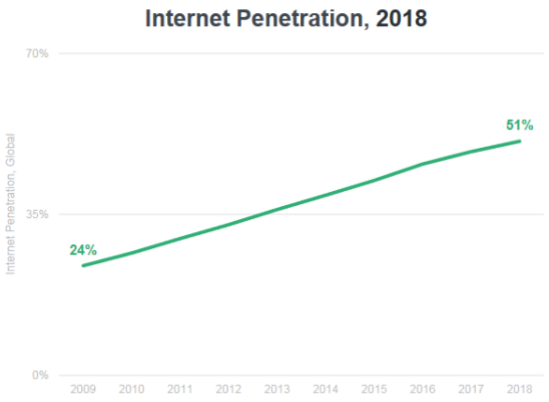

Figure 1 Internet usage percentage globally 2009–2018 [1]

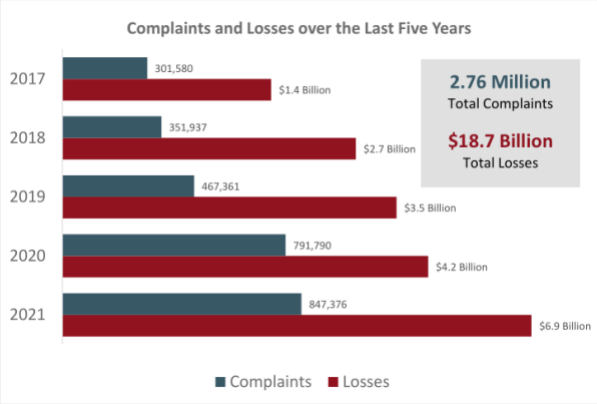

Figure 2 Number of cybercrimes complaint 2017–2021 [2, p. 7]

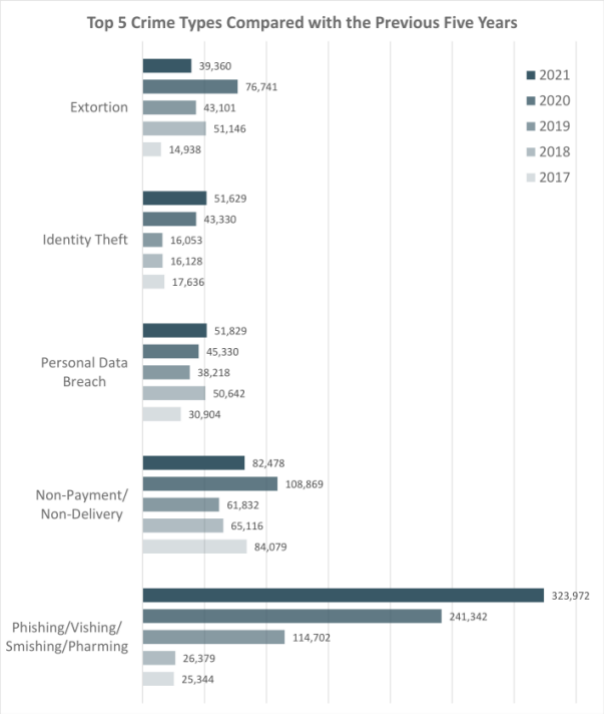

Figure 3 Number of cybercrimes based on types [2, p. 8]

2. Literature Review

2.1 Research Domain

2.1.1 Phishing Awareness Training

Phishing attack is the act of extracting personal information and/or login credentials by sending email, text messages, and phone calls to the victim. claiming to be legitimate party [2]. A common approach, which has been implemented by many companies and considered as industry standard, is "embedded training" in which the training is delivered upon falling for the fake phishing attack [7]. Nonetheless, Wash and Cooper [7] discovered that some users would be encouraged to view any messages thinking that they are harmless and expect to get some insight. The study carried out by them on delivering training in two types: facts-and-advice and stories resulted in lower click rates for facts-and-advice coming from experts, and a lower click rate for stories coming from peers.

A real phishing study done by de Bona and Paci [8] on an Italian company found that attacks which instill sense of urgency led to employees entertaining the attack. Furthermore, their survey on web-based embedded training suggested that it was effective but not enough to ensure the materials being digested by the trainee.

A study carried by Back and Guerette [9], on the effectiveness of awareness training, showed that the target group which took part in the training were more likely to entertain the phishing email. Furthermore, they found that online training platforms were not interactive and limited in content varieties which is essential for effective learning.

Another research which measures the effectiveness of training by frequency of phishing emails was done by Singh *et al.* [10]. A group of users were given emails in which 25% of them were phishing attempt, another group of users had 50%, and the last group had 75% of phishing emails. The result suggested that the training had significant impact on the last group where there is an increase of correctly identified phishing emails. However, this was due to the user's tendency to perceive more emails as phishing attempt.

In short, it could be concluded that the current phishing awareness trainings are not adequate to effectively tackle the growing attacks. However, delivering it by stories and facts-and-advice showed promising outcome.

2.1.2 Gamification

Gamification is the concept of incorporating gaming experiences into real-world non-game context processes [11]–[13]. Gamification promotes interaction and critical thinking skills which contribute to learning motivation and engagement [14]. The term gamification often mixed up with game-based learning (GBL). However, there is a distinction between the two. In GBL, the learning materials are delivered through a full-fledged game. On the other hand, gamification is more into transforming learning process by adding game elements, such as achievements and daily missions to keep the learners feel

motivated [11], [14]. A review done by Klock *et al*. [12] on current gamification indicated that the most popular game elements used are badges and customization, and suggested user profiles understanding in choosing the suitable gamification process. A study by Sailer and Homner [11] on gamifying learning process revealed that it significantly increases the learning motivation while a moderate effect in cognitive and behavioral learning.

Baral and Arachchilage [15], in their study, suggested that the trainee faced difficulty in digesting education materials and argued the importance of phishing quiz with close resemblance to real life environment. Another research done by Nijland [16] implements time, progress, and feedback mechanic into the phishing training. Upon finishing the training, the users are presented with a fictive ranking based on their score and time taken to complete everything. The result indicated that the scoring and time mechanic were the most favorite component of the game. Nijland suggested for further improvement by incorporating leaderboard and competition between friends.

Kävrestad *et al*. [17] study on comparison between game-based training and contextual-based micro-training (CBMT) yielded that both methods were promising but CBMT had a slight advantage in improving users' accuracy in identifying harmful emails as it has a mechanism to increase the trainee awareness. Moreover, Ek [18] on his research recommended to use stories as a mechanic to give immersive and contextual environment, in which the user could learn from an actual situation of a phishing attack.

A study by Tchakounté *et al*. [19] on various phishing trainings gamification found a tendency towards URL obfuscation as the focus instead of email which was deemed to be ineffective since email is often considered as the starting point of phishing attacks. Several recommendations given by Tchakounté *et al*. towards phishing gamification are multi-steps education which covers three phases of phishing attack, i.e., preparation, attack, and exploitation. This is because existing trainings often tend to put emphasis on attack phase alone. Another suggestion is to include multiplayer elements into the training which could trigger collaborative intelligence.

To conclude, gamification is proven to be effective in keeping the end user engaged and motivated on the training. Although a couple of gamification attempts have been carried out, they are still crude and have room for improvements. These improvements are to create a simulation of a real-life attack situation through stories which helps cover the whole phases of phishing attack as well as a multiplayer system with leaderboards.

2.1.3 Artificial Intelligence

Artificial intelligence (AI) is a field in computing in which a computer simulates how a human brain works to perform tasks as well as the ability to adapt to changes [20], [21]. For the past few years, significant research has been carried out to utilize AI in various domains, namely facial and visual recognition, business decision making, education, and more [22]. AI in education system allows personalized and timely feedback which significantly increases learning experience [21]. This is because everyone's individual differences contribute to one's learning behaviour, motivation, and interest [23].

According to Chen *et al*. [21], the field of learning analytics is currently on the rise to be used in enhancing education domain. As pointed out by Lang *et al*. [24] learning analytics involves data collection and analysis based on the learner contexts. This way, an optimized learning environment could be established. Besides, learning analytics could provide data about the learner's future performance and decide the workaround to ensure that the learning process is beneficial [25]. Kävrestad *et al*. [17] also argued that AI could be used to detect a person susceptibility to a phishing attack which then tailored training could be given to those who need it.

To sum up, AI, particularly learning analytics shows a promising future to be implemented in education domain, especially the tailored training which is valuable to help decrease phishing victims.

2.2 Similar System

2.2.1 Bird's Life

Bird's Life is 2D decision-making game, based on PCs and mobile devices, designed mainly for college students to teach them about phishing and anti-phishing techniques. This game comes in three levels, the first one being introduction to entice the users, the second one allows the users to gather knowledge about phishing attacks by hunting for worms, and on the last one, users' knowledge are tested with phishing emails [26].

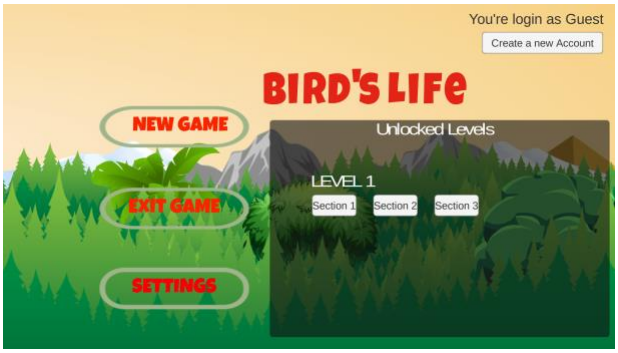

Figure 4 Bird's Life Start Screen [26]

Figure 5 Bird's Life Gameplay Level 2 [26]

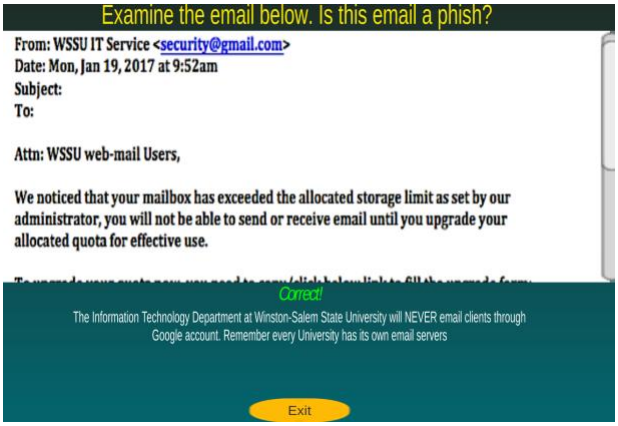

Figure 6 Bird's Life Identify Phishing Attack [26]

2.2.2 Phishy

Phishy is a web-based application which is designed for enterprise employees. This is a single-player game with a story of a character named "Sam" where he got phished of winning millions of dollars and end up in the middle of an ocean together with a tiger in a journey to reach shore to survive. The users are required to fish to feed the tiger and for every fish that the users caught, they are prompted with a question [27].

Figure 7 Phishy [27]

2.2.3 Comparison Table

Table 1 Similar system comparison table

System / Features	Bird's Life	Phishy	Proposed System
Email Phishing	Yes	No	Yes
URL Phishing	No	Yes	Yes
Story	No	Yes	Yes
AI	No	No	Yes
Contextualize Story	No	Yes	Yes
Multiplayer	No	No	Yes
Gamification Component	time, health, points, rewards, tips	points, rewards, narrative	Leaderboard, points, rewards, tips, missions

3. Problem Statement

Given how large the number of phishing attacks has been for the past few years, several countermeasures have been implemented. One of them includes implementing machine learning to detect phishing sites. However, this method still gives high false positive rate and complexity which limit its usability in the near future [28]. Additionally, this technology should not be the sole mean to address cybercrime as human plays a major role as well [9]. This is because social engineering targets the human itself, the weakest link in security chain [3]. Thus, conducting cybersecurity awareness training becomes an additional solution to take care of this weakest link.

Lots of study have been carried out to measure the effectiveness of current awareness training. Some

organizations use services to simulate phishing attacks to their employees and educate them when they fell for it through embedded training, but this increases the click rate instead [7]. Furthermore, a study conducted by Williams *et al*. [29] showed that some employees are complaining about the question variation in security training awareness despite the advancement of cyber threat. Moreover, they thought that including a hypothetical scenario of a phishing attack, in which they could directly witness the consequence of each action, would be beneficial for future reference.

4. Research Aims

This research aims to propose an effective solution in developing phishing awareness training which increase phishing avoidance behavior.

5. Research Objectives

To achieve this research aims, several objectives have been listed as follows.

- To incorporate gamification in training model
- To use artificial intelligence for tailored training delivery
- To include contextual-based learning inside the training model

6. Research Significance

The significance of this research is to ensure that everyone is aware and knowledgeable of phishing attack. This is important due to how phishing attack could take place, namely, email, phone calls, and website popup. Furthermore, phishing attack is difficult to detect since it exploits human thought and emotion by instilling urgency or sense of authority, and victims often take their time to digest that everything said by the attackers are legitimate. Thus, training the users to contextualize everything and getting them motivated throughout the training is essential to ensure that the knowledge is being absorbed by them.

7. Methodology

7.1 Target Users

For the first and last research questions, the target users would be focused on enterprise employees as they are the main victim of current phishing attack. As for the second questions, the target users would be focused on psychological experts.

7.2 Sampling Method

For the first research question, quota sampling method would be used to gather data about user preferences given various set of game components. Quota sampling has been used here to gain a better view of the devised groups. This information would be collected from 200 enterprise employees. These 200 employees would be subdivided into those who play and do not play games and then these groups would be further subdivided into male and female.

For the second research question, snowball sampling method would be used to look for particular people who experts on psychological field. The main information to be extracted from these psychological experts is the cue showed by people when learning as well as people behavior when faced with phishing scenarios.

For the last research question, quota sampling method would be used again to gather information about the user performance between those who take standard training and those who take training mixed with contextual-based scenarios.

7.3 Data Collection and Analysis

Survey would be used to collect data for the first research question and to analyze the result of the survey, descriptive data would be used since the respondents would be choosing from given set of options. As for the second questions, semi-structured interview would be carried out as qualitative data is more valuable.

For the last research question, the users would be tested with questions as well as uninformed phishing attack simulation during the case study before and after the training. However, this has a limitation in which the users may suffer directly from legitimate attack. Thus, this test would be carried out in a contained environment. To reduce the biasness of the participants and possibly replicate the real-life scenario, some of the email content either it is good or bad would take real effect on to the users. For instance, there are chances where they could win real vouchers as well as losing the vouchers. Statistical analysis would be used to make sense of the data gathered.

8. Overview of the Proposed System

The proposed system for phishing awareness training uses gamification component such as point, mission, multiplayer, and roleplaying. Moreover, roleplaying allows contextualization of phishing scenario. Initially, the user would be presented with their role, background stories, and missions. Figure 8 shows the prototype of a roleplay scenario for the user's reference. In the process, users would receive emails which they need to determine whether the emails are legitimate or not to progress (see

Figure 9 and Figure 10). Besides, there is also a social feature which could be used to gain further information from other robots. If played in multiplayer mode, one player could chat with another player to encourage collaborative learning. Figure 11 shows the social tab in which player could interact with one another (multiplayer mode). At the end, users would be able to see their points, review back on the emails, see scoreboard if played in multiplayer mode (see Figure 12).

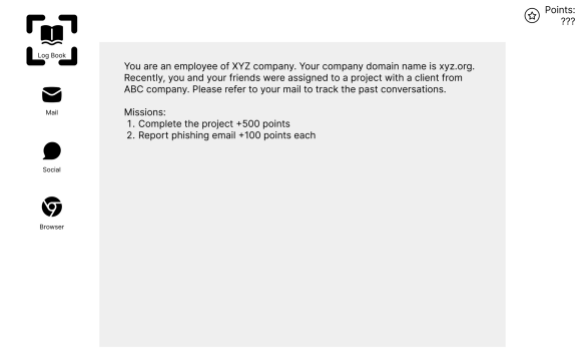

Figure 8 Case scenario

Figure 9 Email list

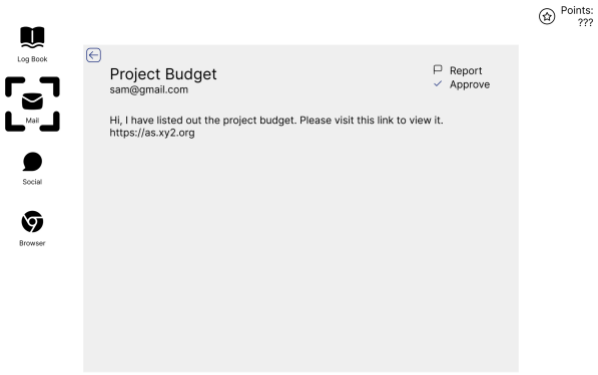

Figure 10 Email preview

Figure 11 Social tab

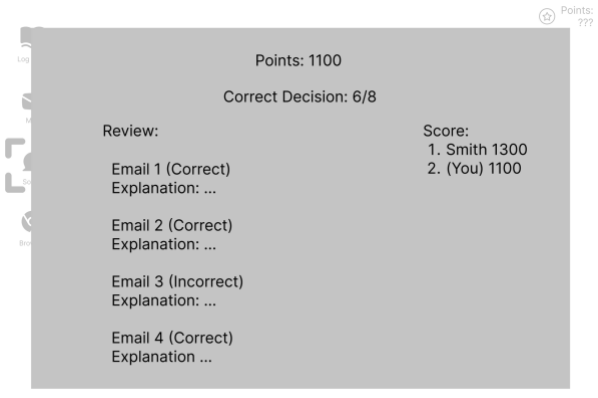

Figure 12 Result

As for tailored learning, the result of each player as well as the time taken to answer each question would be

fed into the learning analytics and then this data would be used to craft new scenarios.

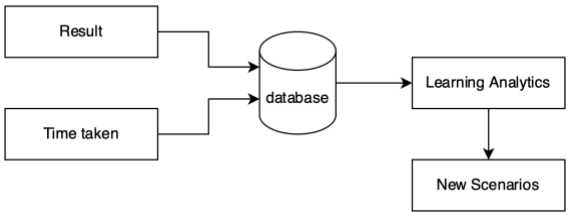

Figure 13 Data flow

9. Conclusion

Phishing attack has always been the main concern in this digitalize era where almost everyone has access to the internet. Although this issue has been there for a long time, an ultimate effective solution is yet to be developed. In this research, gamification, AI, and context-based learning have been studied and used for proposed solution design. Gamification has been gaining popularity lately in promoting educational content. The same goes for AI, specifically learning analytics, which is currently on the rise. Lastly, contextual learning is brought in as well to cover the whole phishing attack life cycle due to the existing solutions which heavily focus on a single part of phishing attack.

References

[1] J. Stanton, "Internet Trends 2019," 2019. Accessed: Apr. 09, 2022. [Online]. Available: https://www.academia.edu/40940037/Internet_Trends_2019_by_Mary_Meeker_I_am_not_the_author_

[2] IC3, "Internet Crime Report 2021," 2021. [Online]. Available: www.ic3.gov,

[3] Z. Wang, L. Sun, and H. Zhu, "Defining Social Engineering in Cybersecurity," *IEEE Access*, vol. 8, pp. 85094–85115, 2020, doi: 10.1109/ACCESS.2020.2992807.

[4] H. Aldawood and G. Skinner, "An Advanced Taxonomy for Social Engineering Attacks," *International Journal of Computer Applications*, vol. 177, no. 30, pp. 1–11, Jan. 2020, doi: 10.5120/ijca2020919744.

[5] APWG, "Phishing Activity Trends Report – 4th Quarter 2021," Feb. 2022. [Online]. Available: http://www.apwg.org,

[6] A. Sumner and X. Yuan, "Mitigating phishing attacks: An overview," in *ACMSE 2019 - Proceedings of the 2019 ACM Southeast Conference*, Apr. 2019, pp. 72–77. doi: 10.1145/3299815.3314437.

[7] R. Wash and M. M. Cooper, "Who provides phishing training? Facts, stories, and people like me," in *Conference on Human Factors in Computing Systems - Proceedings*, Apr. 2018, vol. 2018-April. doi: 10.1145/3173574.3174066.

[8] M. de Bona and F. Paci, "A real world study on employees' susceptibility to phishing attacks," Aug. 2020. doi: 10.1145/3407023.3409179.

[9] S. Back and R. T. Guerette, "Cyber Place Management and Crime Prevention: The Effectiveness of Cybersecurity Awareness Training Against Phishing Attacks," *Journal of Contemporary Criminal Justice*, vol. 37, no. 3, pp. 427–451, Aug. 2021, doi: 10.1177/10439862211001628.

[10] K. Singh, P. Aggarwal, P. Rajivan, and C. Gonzalez, "Training to Detect Phishing Emails: Effects of the Frequency of Experienced Phishing Emails," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, no. 1, pp. 453–457, Nov. 2019, doi: 10.1177/1071181319631355.

[11] M. Sailer and L. Homner, "The Gamification of Learning: a Meta-analysis," *Educational Psychology Review*, vol. 32, no. 1, pp. 77–112, Mar. 2020, doi: 10.1007/s10648-019-09498-w.

[12] A. C. T. Klock, I. Gasparini, M. S. Pimenta, and J. Hamari, "Tailored gamification: A review of literature," *International Journal of Human Computer Studies*, vol. 144, Dec. 2020, doi: 10.1016/j.ijhcs.2020.102495.

[13] R. N. Landers, E. M. Auer, A. B. Collmus, and M. B. Armstrong, "Gamification Science, Its History and Future: Definitions and a Research Agenda," *Simulation and Gaming*, vol. 49, no. 3, pp. 315–337, Jun. 2018, doi: 10.1177/1046878118774385.

[14] R. S. Alsawaier, "The effect of gamification on motivation and engagement," *International Journal of Information and Learning Technology*, vol. 35, no. 1. Emerald Group Publishing Ltd., pp. 56–79, 2018. doi: 10.1108/IJILT-02-2017-0009.

[15] G. Baral and N. A. G. Arachchilage, "Building Confidence not to be Phished Through a Gamified Approach: Conceptualising User's Self-Efficacy in Phishing Threat Avoidance Behaviour," in *2019 Cybersecurity and Cyberforensics Conference (CCC)*, May 2019, pp. 102–110. doi: 10.1109/CCC.2019.000-1.

[16] J. Nijland, "Gamification of Cyber Security Awareness Training for Phishing against University Students," 2022.

[17] J. Kävrestad, A. Hagberg, M. Nohlberg, J. Rambusch, R. Roos, and S. Furnell, "Evaluation of Contextual and Game-Based Training for Phishing Detection," *Future Internet*, vol. 14, no. 4, p. 104, Mar. 2022, doi: 10.3390/fi14040104.

[18] P. Ek, "Teach phishing awareness with games Comparing the effects of Gamification and Learning to read on Phishing Awareness," Feb. 2021.

[19] F. Tchakounté, L. K. Wabo, and M. Atemkeng, "A Review of Gamification Applied to Phishing," 2020, doi: 10.20944/preprints202003.0139.v1.

[20] Y. Duan, J. S. Edwards, and Y. K. Dwivedi, "Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda," *International Journal of Information Management*, vol. 48, pp. 63–71, Oct. 2019, doi: 10.1016/j.ijinfomgt.2019.01.021.

[21] L. Chen, P. Chen, and Z. Lin, "Artificial Intelligence in Education: A Review," *IEEE Access*, vol. 8, pp. 75264–75278, 2020, doi: 10.1109/ACCESS.2020.2988510.

[22] G. J. Hwang, H. Xie, B. W. Wah, and D. Gašević, "Vision, challenges, roles and research issues of Artificial Intelligence in Education," *Computers and Education: Artificial Intelligence*, vol. 1. Elsevier B.V., Jan. 01, 2020. doi: 10.1016/j.caeai.2020.100001.

[23] K. Mangaroska and M. Giannakos, "Learning Analytics for Learning Design: A Systematic Literature Review of Analytics-Driven Design to Enhance Learning," *IEEE Transactions on Learning Technologies*, vol. 12, no. 4, pp. 516–534, Oct. 2019, doi: 10.1109/TLT.2018.2868673.

[24] C. Lang, G. Siemens, A. Wise, and D. Gašević, *Handbook of Learning Analytics*. Society for Learning Analytics Research (SoLAR), 2017. doi: 10.18608/hla17.

[25] C. Troussas, A. Krouska, and M. Virvou, "Using a multi module model for learning analytics to predict learners' cognitive states and provide tailored learning pathways and assessment," in *Intelligent Systems Reference Library*, vol. 158, Springer Science and Business Media Deutschland GmbH, 2020, pp. 9–22. doi: 10.1007/978-3-030-13743-4_2.

[26] P. Weanquoi, J. Johnson, J. Zhang, P. ; Weanquoi, and J. ; Johnson, "Using a Game to Improve Phishing Awareness," 2018. [Online]. Available: https://digitalcommons.kennesaw.edu/jcerpAvailableat:https://digitalcommons.kennesaw.edu/jcerp/vol2018/iss2/2

[27] C. J. Gokul, S. Pandit, S. Vaddepalli, H. Tupsamudre, V. Banahatti, and S. Lodha, "Phishy - A serious game to train enterprise users on phishing awareness," in *CHI PLAY 2018 - Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, Oct. 2018, pp. 169–181. doi: 10.1145/3270316.3273042.

[28] H. Abroshan, J. Devos, G. Poels, and E. Laermans, "Phishing attacks root causes," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 10694 LNCS, pp. 187–202. doi: 10.1007/978-3-319-76687-4_13.

[29] E. J. Williams, J. Hinds, and A. N. Joinson, "Exploring susceptibility to phishing in the workplace," *International Journal of Human Computer Studies*, vol. 120, pp. 1–13, Dec. 2018, doi: 10.1016/j.ijhcs.2018.06.004.