Exploring the Potential Implications of AI-generated Content in Social Engineering Attacks

Yazan Alahmed
Faculty of Engineering
Al Ain University
Abu Dhabi, U.A.E.
yazan.alahmed@aau.ac.ae

Reema Abadla
Faculty of Engineering
Al Ain University
Abu Dhabi, U.A.E.
reemaabadla@gmail.com

Mohammed Jassim Al Ansari
School of Business
University of Central Lancashire
Preston,UK
mjmal-ansari@uclan.ac.uk

***Abstract*— The evolution of artificial intelligence (AI) and machine learning presents both utility and security implications for our digital interactions. This study focuses on the transformative role of generative AI in social engineering attacks, specifically examining three pillars where it significantly amplifies their impact: advanced targeting and personification, genuine content creation, and automated attack infrastructure. The analysis forms a conceptual model named the generative AI social engineering framework. The research delves into human implications and measures to counter social engineering attacks, blending theoretical analysis with practical insights through case studies. Ethical considerations surrounding AI in malicious activities are discussed, emphasizing the importance of safe AI development, and various articles were reviewed to highlight social engineering attacks as a common threat. Two studies were conducted: a user testing study with 48 participants from diverse occupations and social engineering awareness, and an exploratory study collecting qualitative data from 40 social engineering attack victims. The user testing study revealed universal acceptance of the AI-based tool, irrespective of participants' occupations. Victim themes included reasons for falling prey to attacks, methods, prevention advice, and detection. The research concludes by highlighting AI-generated content as a key factor fueling social engineering attacks and bridging the gap between AI development and cybersecurity practices, highlighting the need for interdisciplinary approaches to address evolving challenges.**

***Keywords*-** Machine learning, Chatbot, social engineering, Artificial intelligence, Phishing, ChatGPT.

I. Introduction

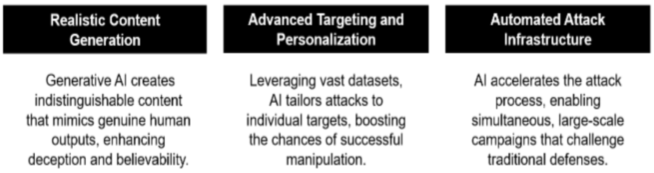
The pervasive growth of technology, catalyzed by the evolution of generative AI systems capable of producing content based on intricate patterns, has introduced a range of threats to our interconnected community [1,2,3]. As our reliance on advanced technology deepens, social engineering has emerged prominently as a major threat, especially with attackers employing increasingly sophisticated approaches facilitated by AI [4]. Social engineering, a term denoting manipulating individuals or groups to acquire confidential information or persuade them to undertake specific actions, relies primarily on psychological and interpersonal skills, distinguishing it from traditional computing-based threats [5]. The surge in social engineering attacks can be attributed to the rise of "powerful" AI. AI models, mirroring human communication and trust signals, present a novel frontier for social engineering and phishing threats [6].

In addressing the broader context, it is crucial to recognize that the increasing prevalence of social engineering attacks not only raises concerns about the structure of these incidents but also delves into the emotional experiences of victims, variations in awareness levels among individuals targeted, and the effectiveness of automated spam detection on social networks. This study aims to explore these multifaceted aspects, shedding light on the intricate interplay between evolving AI technologies and the escalating challenges posed by social engineering in our interconnected society AI-generated content has played a significant role in social engineering attacks by automating convincing and tailored messages, emails, or even deepfakes, making it easier for attackers to manipulate individuals into divulging sensitive information or taking malicious actions [7]. It has also been used by attackers where spam is inflicted through texts, emails or calls. The consequences of such attacks include loss of money and breach of privacy. Moreover, AI-driven tools such as deepfake and AI-generated phishing emails have been used by attackers to create fake identities familiar to that of their target person. Deepfakes utilize AI to create realistic but fabricated audio or video content that mimics someone's appearance and voice, leading to deception and manipulation [8]. Furthermore, just as AI has been used to protect against phishing emails [9], AI-generated phishing emails have also leveraged machine learning to craft messages that closely resemble legitimate communication. These emails mimic official correspondence, hence deceiving recipients. The ability of AI-driven tools to generate spam in the form of calls and texts has led to heavy attacks on organizations and individuals. The power to automate spam through AI-generated content has seen[1] the rise of many social engineering attacks. For instance, IT engineers at organizations have witnessed a rising rate of social engineering attacks of around 140% in the last 2 years [10].

It has been shown that social engineering attacks happen in different categories of interactions: human-to-computer, human-to-human, and computer-to-computer. Exploration is needed to address this issue, to identify ways to curb the attacks when certain technology is used for deceit, and to understand the problems caused [6,11]. The high evolution of AI, together with machine learning, is expected to influence many aspects of life, with significant implications for the security and safety of our digital interactions. Cyber-attacks are gaining scale and impact, causing severe financial harm, breaching customer privacy, and creating chances to tear down critical infrastructure [6]. Global economic forums consistently name such cyberattacks as a key risk worldwide [12]. Machine learning and AI technologies have displayed their capability of assisting in curbing and detecting social engineering and phishing attacks, though, on the other hand, they can be used in malicious activities to amplify the abilities and impact of such attacks. Knowing AI's driven social engineering attacks nature and their approaches is very important to society [13]. Both organizations and individuals can lay out proactive strategies to defend against these cybersecurity attacks. In the progression of AI, the field of social engineering is advancing, with sophisticated algorithms now capable of generating information that mimics the pattern of human communication, leverages psychological triggers with unprecedented accuracy, and evades detection by traditional security measures [13,14,15]. Moreover, the factors that amplify the scope and speed of AI-enabled SE attacks are highlighted in Figure 1. This paper's objective is to explore and analyze the security implications of AI-generated content by examining different attack categories evident in social engineering. It seeks to know ways in which threat actors can harness AI and machine learning to carry out campaigns of social engineering and phishing to educate on threat intelligence and come up with future mitigation approaches [8]. The results of this thesis will be of key importance to researchers, policymakers, and practitioners in guiding them to understand the advancing landscape of AI-generated content used in social engineering and phishing threats.

This paper addresses four research questions:

- RQ1: Which of the AI-driven SE attack cases are obscure to victims?
- RQ2: Which are the social engineering types caused by AI-generative tools and have a significant impact?
- RQ3: How can AI-generative tools be used by cyber attackers to effectively enhance social engineering attacks?
- RQ4: What mechanisms can be employed to enhance easy detection and mitigation of these AI--



Realistic Content Generation

Generative AI creates indistinguishable content that mimics genuine human outputs, enhancing deception and believability.

Advanced Targeting and Personalization

Leveraging vast datasets, AI tailors attacks to individual targets, boosting the chances of successful manipulation.

Automated Attack Infrastructure

AI accelerates the attack process, enabling simultaneous, large-scale campaigns that challenge traditional defenses.

Fig 1: [16] Three Pillar Framework of Generative AI-enabled social engineering attacks.

As we embark on this examination, it is important to acknowledge the ethical considerations surrounding the malicious application of AI in social engineering attacks. Therefore, by exploring the complexities of AI-generated content in social engineering, this research will help to empower organizations and individuals by displaying valuable insights to fortify their cybersecurity postures and mitigate the advancing threats shown by the convergence of AI and social engineering.

The rest of the paper is organized as follows. Section 2 discusses existing social engineering attacks. The paper's methodology is explained in section 3, followed by the results and discussion in sections 4 and 5, respectively.

II. Literature Review

This section examines the corpus of research on the security implications of artificial intelligence (AI)-generated content about social engineering, encompassing real-world incidents, AI-based social engineering attacks, CEO fraud cases, and the effectiveness of deception in network attacks.

A. Social Engineering Attacks

The psychology of manipulating and taking advantage of human behavior is the foundation for the development of social engineering attacks. [14]. Social engineering attacks symbolize rigid and advancing threats to cybersecurity [15].

Twitter (X) has long been a hotspot for social engineering attacks by cybercriminals. The 2020 Twitter Bitcoin Scam on July 15th is a notable example, where around 130 high-profile accounts, including those of Elon Musk, Barack Obama, and Bill Gates, were compromised. Using social engineering, attackers posed as credible individuals, enticing followers to transfer Bitcoin to a specified address. They claimed to quadruple contributions, citing it as a charitable gesture for COVID-19. The scheme exploited trust, urgency, and the credibility of verified accounts, resulting in over $118,000 in cryptocurrency deposits within a short period.

In 2019, the Akamai organization was subjected to a phishing attack. Frontier Journal from 2020 [15,16] reports that attackers attempted to conceal dubious URLs by prefixing them with the seemingly authentic www.translate.google.com address to trick people into logging in. Once the victims had logged in, the attackers led them to phishing schemes where they were asked to provide their Netflix payment data. The absence of an HTTPS lock and the misspelled URL were two crucial red flags that alerted the organization's members to the phishing attempt.

B. AI-based social engineering attacks

Several studies have attempted to measure the success rate of social engineering attacks augmented by AI-generated content. The research findings reveal that the use of AI can significantly increase the success of phishing campaigns, as AI-generated messages exhibit improved contextual awareness and persuasive language. Analyzing these insights into the practical implications of AI-driven social engineering tactics.

The threat of AI-based cyberattacks, particularly phishing, is emphasized in [17], focusing on the risks posed by social bots for mass phishing attacks. Social bots, advanced AI programs mimicking human interaction [18], offer benefits in social media but serve as potent tools for attackers, as illustrated in [19]. The authors used X's (Twitter's) bots to automate spear phishing attacks, distributing masked phishing URLs through shortened links seamlessly within the regular

Twitter activity. Employing machine learning, they developed SNAP_R, a data-driven system identifying relevant textual patterns in social media spear phishing. Customized messages were then created for high-value or vulnerable X (Twitter) users based on public content. The click-through rates for this extensive phishing effort were among the highest ever recorded, demonstrating the effectiveness of automated social engineering that is well-coordinated and scaled, and therefore highlighting the urgency for addressing such AI-based threats. Moreover, a prominent danger of AI-based social bots is their ability to manipulate public opinion by continuously copying and reposting certain content or hashtags to give the appearance that a presidential candidate, for instance, is more favored [20]. Nonetheless, it is important to note that the study focuses on a specific platform (Twitter) and type of attack (spear phishing). The findings may not fully generalize to other social media platforms or types of AI-based attacks.

In March 2019, a major security breach occurred when the CEO of a UK-based energy firm fell victim to a sophisticated deep fake audio fraud [21]. The executive was deceived by a phone call flawlessly impersonating the voice of the firm's CEO, the chief executive of the company's German parent corporation, and mistakenly transferred about £200,000 to a Hungarian bank account. He instantly transmitted the payment to a Hungarian supplier account, believing this was a legitimate request from his supervisor, unaware that it was a fraudulent scam executed by an individual employing AI speech technology to mimic the CEO's voice.

In a sophisticated CEO fraud campaign in December 2021, a French company incurred a $38 million loss within days [21]. An attacker, posing as the CEO, executed a social engineering scheme, urgently requesting the company's accountant to transfer $300,000 to a bank in Hungary. The fraud went unnoticed initially, leading to an investigation that uncovered not only voice impersonation but also repeated attacks on a real estate developer, resulting in a $38 million transfer. Eight suspects were later arrested. Similarly, in a 2020 deep fake CEO fraud against a Japanese company, fraudsters impersonated the director via phone, directing a $35 million transfer for a supposed acquisition [21]. Despite a later investigation, the money was lost due to the use of deep voice technology to mimic the director's voice. Top of Form Bottom of Form.

In 2018, [22] conducted the Tularosa Study, which involved testing over 130 red team hackers. The research aimed to monitor participants' personalities, psychological intentions, and cognitive abilities while engaging in network attacks. Two different scenarios were presented to the attackers: one involving the use of deception techniques and the other without any deception. The deception strategy was evaluated both with and without the presence of a sample network. The primary method of deception employed in the study was the use of decoys within the network. The authors released theses, research summaries, and academic papers detailing their findings. Given that much of the research surrounding this case study focuses on recent discoveries regarding the effectiveness of deceit as a defense, the insights from [22] are crucial for future studies. Their work demonstrated various aspects of how attackers can be influenced in a decoy-filled environment. Top of Form Bottom of Form

Unlike the traditional social engineering tactics, which often rely on human manipulation and psychological tricks to deceive individuals into divulging sensitive information [26], as per the previews above, augmented AI has leveraged advanced algorithms to create more convincing and personalized deception. These previous studies have extensively explored the technical capabilities of AI-generated content in social engineering attacks. The benefit of these reviews is that they can evaluate the efficacy of awareness campaigns and countermeasures, offering insightful data that can be used to strengthen defenses against social engineering attacks.

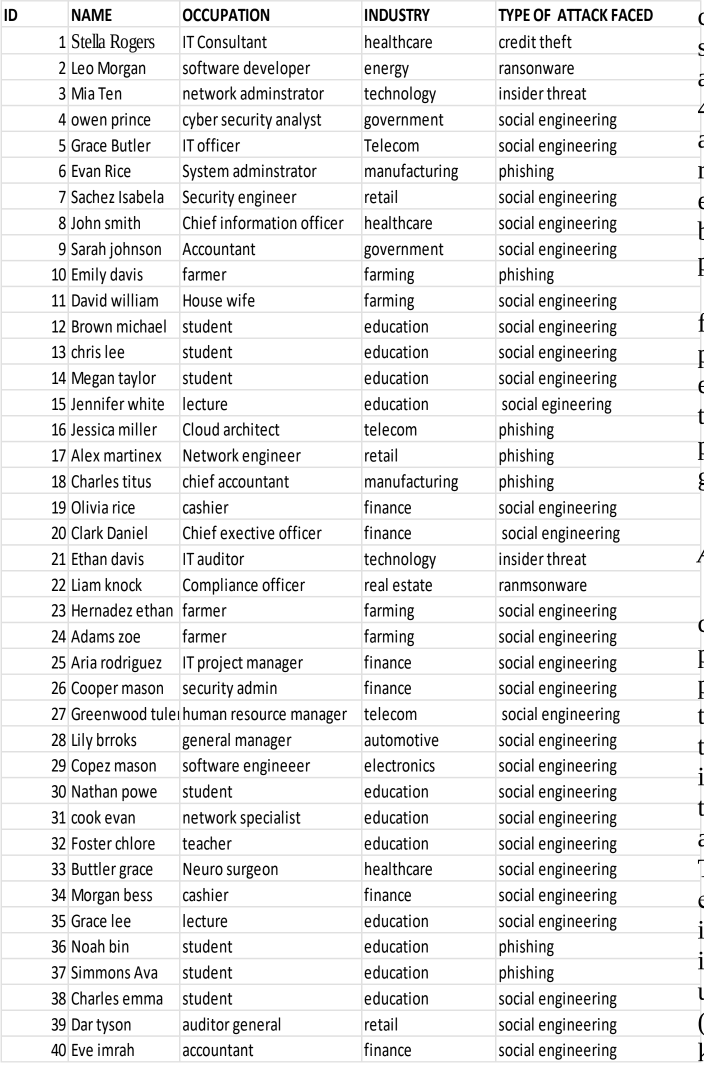
Furthermore, the examinations of studies provide a thorough grasp of the techniques that attackers use. By synthesizing these existing reviews, we were able to identify common patterns and tactics that come up with generative AI content. These research findings reveal that the use of AI can significantly increase the success of phishing campaigns, as AI-generated messages exhibit improved contextual awareness and persuasive language. Analyzing these insights into the practical implications of AI-driven social engineering tactics. Still, there is a clear study vacuum concerning consumers' psychological vulnerabilities. While previous research has contributed valuable insights into the technical aspects of AI-driven social engineering, there remains an unexplored area related to how individuals perceive and respond to information generated by artificial intelligence. This study aims to address this gap by examining various aspects of human interaction with AI-generated content, offering insights that complement existing technical perspectives and advancing our understanding of the intricate interplay between AI and human psychology in the realm of social engineering attacks.

III. METHODOLOGY

This section displays the study procedures and data analysis methods employed to comprehensively explore the security implications of AI-generated work in social engineering attacks. The combination of qualitative and quantitative methods helped us offer a thorough understanding of both technical aspects and human factors involved in these cyber threats. The two approaches (quantitative and qualitative) complement each other by addressing the complexity of our research questions to bring a well-rounded analysis of the subject matter. To ensure the accuracy in our findings, for the two studies carried out, a final validation with the involved participants was done. Every participant was subjected to checking their data review as they responded to the questionnaire.

We used emails and snowball approaches [23] to contact participants who had by any chance experienced any cyber-attack.

TABLE 1: PARTICIPANTS INTERVIEW DATA TABLE



ID	NAME	OCCUPATION	INDUSTRY	TYPE OF ATTACK FACED
1	Stella Rogers	IT Consultant	healthcare	credit theft
2	Leo Morgan	software developer	energy	ransonware
3	Mia Ten	network adminstrator	technology	insider threat
4	owen prince	cyber security analyst	government	social engineering
5	Grace Butler	IT officer	Telecom	social engineering
6	Evan Rice	System adminstrator	manufacturing	phishing
7	Sachez Isabela	Security engineer	retail	social engineering
8	John smith	Chief information officer	healthcare	social engineering
9	Sarah johnson	Accountant	government	social engineering
10	Emily davis	farmer	farming	phishing
11	David william	House wife	farming	social engineering
12	Brown michael	student	education	social engineering
13	chris lee	student	education	social engineering
14	Megan taylor	student	education	social engineering
15	Jennifer white	lecture	education	social egineering
16	Jessica miller	Cloud architect	telecom	phishing
17	Alex martinex	Network engineer	retail	phishing
18	Charles titus	chief accountant	manufacturing	phishing
19	Olivia rice	cashier	finance	social engineering
20	Clark Daniel	Chief exective officer	finance	social engineering
21	Ethan davis	IT auditor	technology	insider threat
22	Liam knock	Compliance officer	real estate	ranmsonware
23	Hernadez ethan	farmer	farming	social engineering
24	Adams zoe	farmer	farming	social engineering
25	Aria rodriguez	IT project manager	finance	social engineering
26	Cooper mason	security admin	finance	social engineering
27	Greenwood tule	human resource manager	telecom	social engineering
28	Lily brroks	general manager	automotive	social engineering
29	Copez mason	software engineeer	electronics	social engineering
30	Nathan powe	student	education	social engineering
31	cook evan	network specialist	education	social engineering
32	Foster chlore	teacher	education	social engineering
33	Buttler grace	Neuro surgeon	healthcare	social engineering
34	Morgan bess	cashier	finance	social engineering
35	Grace lee	lecture	education	social engineering
36	Noah bin	student	education	phishing
37	Simmons Ava	student	education	phishing
38	Charles emma	student	education	social engineering
39	Dar tyson	auditor general	retail	social engineering
40	Eve imrah	accountant	finance	social engineering

The snowball recruitment method is a recruitment technique in which the participants involved in the research are requested to assist the examiners in getting other relevant participants [23]. For the qualitative examination, 40 participants aged between 18 and 60 years were contacted to take part in the interview (N=40; 22 Females and 18 Males).

The participants included: IT undergraduate students, IT security employees, civil servants, and house girls. The participants had different professions with different skills in IT. Participants 18 years old and above of age and who had faced cyber-attack before were picked. This criterion was essential to gather insights from individuals who had encountered different forms of social engineering attacks facilitated by AI-generated content. Such cyber-attacks included: receiving spam messages, receiving clone-voiced client calls, and fraudulent emails. The overall diversity in the participants' ages and professions was chosen to capture a wide spectrum of perspectives and experiences related to AI-driven social engineering attacks. Moreover, participants were selected based on their varying levels of expertise in information technology to ensure a more thorough understanding of the subject matter.

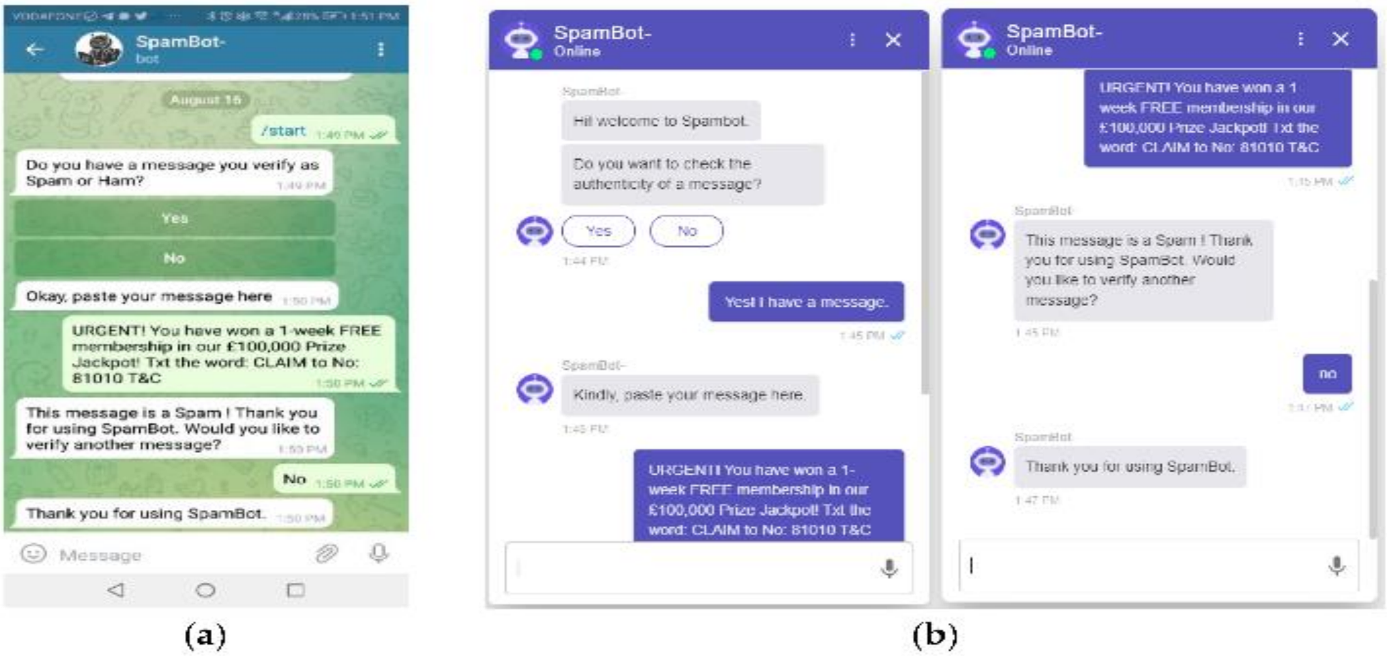
In the user testing examination, 40 participants participated in the exploration of Chatbot (N=40; 30 males and 10 males). The sample size was determined through a combination of practical considerations and statistical significance. Given the specificity of our participant criteria and the focus on in-depth qualitative insights, a sample size of 40 participants for the interviews was deemed sufficient to achieve saturation, where recurring themes and patterns in responses became apparent. This sample size aligns with established guidelines for usability testing, ensuring a balance between obtaining meaningful insights and managing practical constraints.

An earlier pilot study was carried out to confirm the feasibility and suitability of our research design before the primary investigation. The purpose of the pilot study was to evaluate the viability of the data gathering methods, improve the survey tools, and pinpoint any possible difficulties. Only 5 participants took part in the pilot study hence it was noted good to go.

A. *Procedures*

Prior to the semi-structured interview [24], all of the chosen participants were contacted and notified through phone calls, emails, and texts. A low turnout of the contacted participants was a limitation to this methodology. In case 1 of the study, the use of open-ended questions was employed to the participants. Through open-ended questions during interviews and user testing with a chatbot, the research aimed to uncover and explore the intricacies of AI-generated SE attacks that might not be readily apparent to victims (RQ1). The response from the participants was probed further to ensure the interviewees' answers contained clarity. The information from the participants was recorded as audio. The interview lasted for 35 minutes. At last, the audio recordings underwent transcription and coding. In the second case study (usability testing), end users employed a chatbot, sometimes known as a spambot. The analysis of the interview data gathered from victims of social engineering attacks guided the development of the chatbot. To create the chatbot, a deep learning system was employed. Deep learning is a branch of machine learning that was selected due to its capacity to analyze large volumes of data and recognize complex patterns [25]. It is modeled after the structure and operation of the neural networks found in the human brain. By leveraging this advanced algorithm, the chatbot was endowed with enhanced capabilities to simulate human-like interactions, adapt its responses based on user input, and generate contextually relevant content. The dataset used to train the tool was from our previous study. The reason for doing this was to have a benchmark for our spam message exploration. The dataset had 700 spam messages. Telegram was integrated with the deep learning algorithm. One benefit of using a chatbot is that it can be easily integrated with other social media platforms, such as Facebook and WhatsApp, which are visible to end users. These sites are highly targeted media by social engineering attackers for cyber-attacks. The end-users were provided with links of the telegram and chatbot websites to freely use them. Samples of non-spam and spam messages were given to users for testing. Thereafter, the end users used Chatbot and follow-up usability questionnaires. For the Chatbot assessment, a system usability scale was applied. The SUS questionnaire is used to analyze arguments that are raised after post-evaluation.

The usability testing with the AI-generated chatbot specifically targeted social engineering types caused by AI-generative tools. The study aimed to identify and evaluate the impact of such AI-driven social engineering attacks by simulating human-like interactions and leveraging deep learning algorithms. The data collected from participant interactions with the chatbot provided insights into the effectiveness and consequences of these attacks (RQ2). Furthermore, the creation and application of a deep learning algorithm-powered chatbot gave researchers a platform to study how cybercriminals might use AI-generative tools to their advantage in social engineering attempts. The chatbot's interaction with many social media sites such as Facebook, WhatsApp, and Telegram, enabled an exhaustive investigation of potential attack routes and tactics utilized by adversaries (RQ3).



August 15

/start 1:49 PM

Do you have a message you verify as Spam or Ham? 1:49 PM

Yes

No

Okay, paste your message here 1:50 PM

URGENT! You have won a 1-week FREE membership in our £100,000 Prize Jackpot! Txt the word: CLAIM to No: 81010 T&C 1:50 PM

This message is a Spam ! Thank you for using SpamBot. Would you like to verify another message? 1:50 PM

No 1:50 PM

Thank you for using SpamBot. 1:50 PM

Hi! welcome to Spambot.

Do you want to check the authenticity of a message?

Yes No

1:44 PM

Yes! I have a message. 1:45 PM

Kindly, paste your message here. 1:45 PM

URGENT! You have won a 1-week FREE membership in our £100,000 Prize Jackpot! Txt the word: CLAIM to No: 81010 T&C 1:45 PM

This message is a Spam ! Thank you for using SpamBot. Would you like to verify another message? 1:45 PM

no 1:47 PM

Thank you for using SpamBot. 1:47 PM

Fig.2. [26] (a) Telegram Chatbot (b)Website Chatbot

B. *Data Analysis*

The data from the interview was generated through six questions answered as per the experience of the participants about the spam messages they encountered: when was the spam message sent to you, what triggered your conscience to believe the message or call was legitimate, what did you feel after realizing it was just spam, did you ever experience such an incident before, what do you recommend we can do to mitigate such incidents, and finally, a follow-up question was asked: who did you inform first about your case. For the interviewee in this SUS questionnaire, the attackers used Telegram, WhatsApp, Facebook, and X (Twitter).

A grounded theory analysis method was used to analyze the data. Grounded theory is a research technique that aims to generate a theory or conceptual framework that is "grounded" in the data collected during a research process. To achieve this method, a constant comparison was made with the data generated. Theoretical analysis was also done based on the fed data. The approach has three major coding approaches (open, axial, and selective) to categorize and organize the data. The reason for this analysis was to get the victim's experience with social engineering attack mechanisms. The first two coding stages were applied to systematize and outline codes related to the objectives of our study. Three coding rounds were done, after which refining and reviewing of the code followed. After this, the third coding stage was done, where similar codes were noticed and merged.

Using the System Usability Scale (SUS) questionnaire, system usability was assessed as part of the usability testing process. The purpose of this evaluation was to determine how well the chatbot identified and mitigated AI-driven dangers, as well as how well it worked for users. Furthermore, by using interview data, the grounded theory analysis approach was able to discover methods that participants proposed to improve the identification and mitigation of social engineering assaults made possible by AI-generative technologies (RQ4).

In addition, the data analysis methods chosen were strategically aligned with the research questions and the nature of the collected data. Applying the grounded theory analysis approach to qualitative interview data provided a thorough investigation of participants' experiences with AI-driven SE attacks by extracting various insights. Simultaneously, standardized metrics for evaluating the AI-generated chatbot's usability were supplied by the SUS questionnaire when it was applied to quantitative data. This deliberate combination of qualitative and quantitative approaches ensures a holistic examination, effectively addressing both technical intricacies and human responses inherent in the study's research questions.

Ethical issues were crucial in the research on the security implications of AI-generated work in social engineering attempts. Prioritizing informed consent meant that before receiving explicit assent, participants had to be fully educated about the goals, methods, possible dangers, and rewards of the research. All information gathered from interviews and usability tests was anonymized, and strict procedures were followed to securely retain personal data to guarantee participant anonymity. Moreover, transparent research practices were maintained throughout the study to build trust.

Nonetheless, the methodology faced limitations, including a low turnout of participants, potentially introducing selection bias. During interviews, open-ended questions and probing strategies were used to address any biases in participant replies. The chatbot's effectiveness relies on its algorithm's accuracy, introducing a potential limitation, and the controlled environment of usability testing may not fully represent real-world social engineering events. Taking note of these constraints, a thorough and responsible examination of the topic was carried out.

IV. RESULTS

A. Examination Study

For study 1, we coded the transcribed data gathered using open coding, and 33 free nodes were displayed, as seen in Figure 3. A Nvivo-12 tool was used to absorb the 33 free nodes portrayed. As displayed in Figure 3, the diagram displayed by NVivo-12 helped in understanding variations of the transcribed interview data. NVivo-12 supported the constant comparison of the patterns through refinement of the themes and patterns – a capability of the software that allows refining and developing a deep understanding of the emerging patterns.

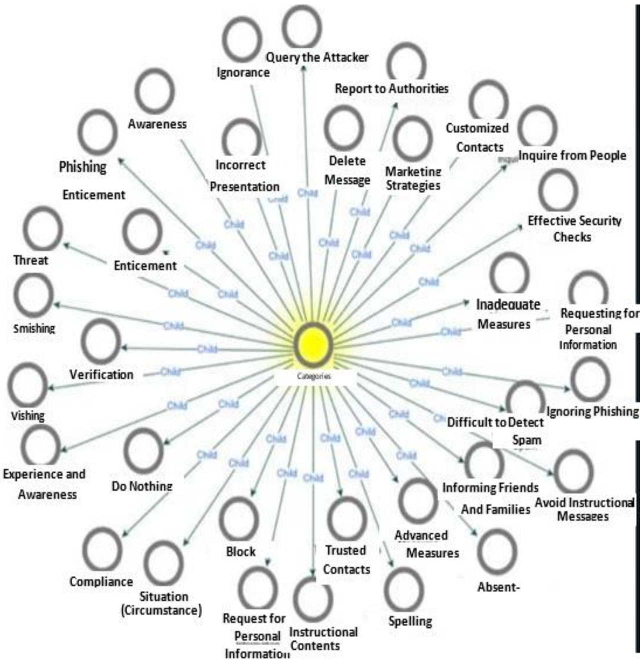


Fig. 3. [27] Abstracted 33 nodes during open coding.

In the other stage of the grounded theory, axial coding, the analysis of free nodes displayed in open coding was grouped into six key groups as seen in Figure 4 (attack context, reasons for falling for attacks, attack-preventing advice, methods of attack, methods of detection, and reaction of the victim). A simple logical relationship between the open codes was used to obtain the six major groups.

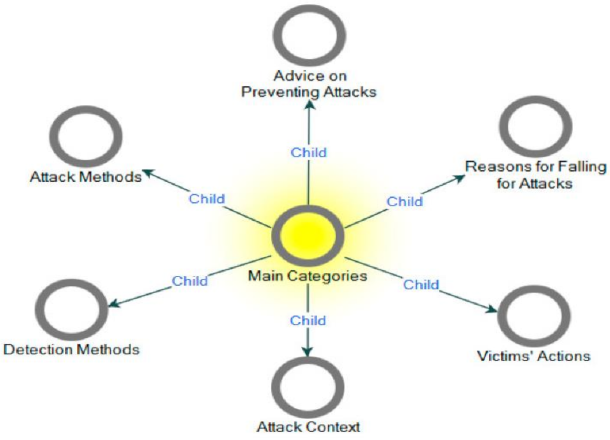


Fig. 4. [28] Major groups (Tree nodes)

The analysis of the open coding was obtained as shown below.

TABLE 2. ANALYSIS OF OPEN CODING

[1] **Excerpt Categories**	[2] **Conceptualization**	
[3] Don't open this spam. It will harm you	[4] Ensuring best practices in cyber security	[5] Get away from instruction messages
[6] It is about vigilance: it is about awareness	[7] Getting the legitimate information from institutions	[8] awareness
[9] I will say that people should be conscious of their recipient's number.	[10] Seeking clarity on the sender's details	[11] verification
[12] The messages are only for marketing reasons.	[13] Advertisement of services to lure users	[14] Marketing strategies
[15] This is something (spam message) I get often, mostly through email	[16] Emails as a cyber-attack mean.	[17] phishing
[18] I got the messages two weeks ago	[19] SMS as a method of cyber-attack and identity theft	[20] smishing
[21] It is very hard to restrict these spam messages	[22] Deceptive and obscure attack approaches	[23] Hard to accurately know spam

B.User testing Study

The responses from the recruited SUS questionnaire participants were analyzed. Some rating scores were missed by one participant. Hence, the average SUS score was calculated for the 38 respondents.

A. Third item: For every odd number question, the rating score was less than 1.

B. For even-numbered questions, it was deducted from 5.

C. The total values obtained in steps 1 and 2 were multiplied by 2.5.

D. The SUS score from every respondent was added up to get the average score, and the answer was divided by the respondents'' number.

The average SUS score for every interviewee of the chatbot ranged from 47 to 97, following the steps named above. The findings from the coded data and the SUS questionnaire helped in determining what mechanisms can be employed to enhance easy detection and mitigation of these AI-driven threats. The capabilities of the chatbot in accessing acceptable SE attack instances are recognized.

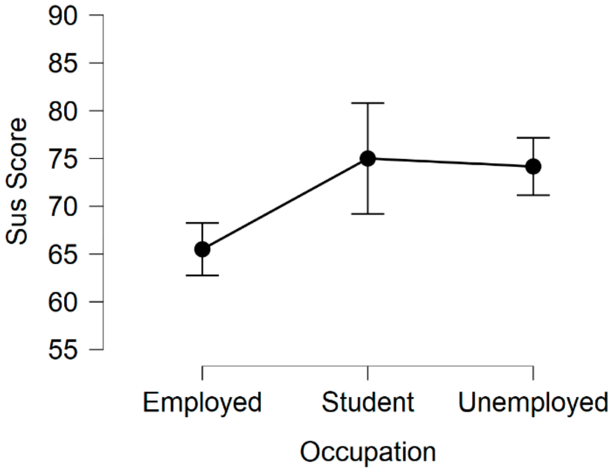


Fig. 5. [29] Descriptive plots of SUS scores and users' occupation.

These ratings show how the users subjectively rated the chatbot's usefulness and capacity to handle SE attack scenarios. A higher SUS score indicates a more favourable perception of usability. The calculated averages provide a quantitative measure of the overall user satisfaction with the chatbot's performance in detecting and mitigating AI-driven threats. While higher scores suggest that users found the chatbot to be intuitive, efficient, and capable of dealing with SE attacks, lower scores may point to areas requiring improvement. A thorough grasp of the chatbot's usefulness in the real world and its ability to improve cybersecurity defenses against AI-driven attacks is made possible by the combination of these SUS scores with the qualitative conclusions from the coded data.

V. DISCUSSION

This research paper explored the broad range of events related to social engineering attacks, particularly those caused by generative AI content. A semi-structured interview and user testing studies were the methods used to examine the exploration. The practical benefit of the methodology studies used in this paper (semi-structured interviews and user testing) is that they helped us reach a human-centric conclusion about problems like psychological impact on victims' need to create social engineering awareness. In addition, the user testing methodology allowed direct observation of participants' interactions with AI-generated content. About 66 participants were involved in the interview. The primary study findings are discussed based on every study result.

Our findings on the general structure of social engineering attack cases display that phishing and smishing methods are the most common cases of SE attack cases (RQ1, RQ2). In phishing, cyber attackers use deceptive emails, messages, or websites to impersonate trusted entities such as reputable companies, agencies, and banks to trick users (RQ3). The social engineering attack-detecting chatbot in this research emerged from the exploratory study part. The chatbot was built on an AI trained on many social engineering malicious texts to detect similar texts. Usability Chatbot evaluation showed that it was effective in detecting every kind of social engineering of users regardless of the level of user exposure (RQ4). For instance, for the three types of users (students, employed, and unemployed), make the chatbot simple and effective after analyzing the results. From the chatbot's descriptive statistics, it was displayed that unemployed people have the highest acceptance rate. This is most likely because this group segment has a higher prevalence of low social engineering expertise. Most employed professions have much awareness due to the many cyberattacks they pass through since they are the most targeted, especially where the attackers do it in search of money [30,31]. In addition, learners display a high acceptance rate as compared to employed people. The awareness of employed people seems to be more paramount as compared to that of students. The assumption might be that the employed persons are mostly IT security specialists.

The reason for a successful social engineering attack is attached to victims' psychological factors like social engineering attack ignorance, absent-mindedness, circumstances of the victim, lack of security, etc.

A.The Practical Benefit of this study

The practical benefit of this research is the social engineering attack management themes arising and how the emerging themes can be integrated into the application and development of superior tools and methods of spam detection, especially on social media platforms. The literature review part of this study displayed how, without knowledge about potential threats related to AI-generated content, organizations and individuals can lose big. The main contribution of this research is the application of advanced mechanisms for detecting spam from different social media platforms. The knowledge gained in this study catalyzes the development of adaptive defense strategies and targeted training programs. By incorporating these insights into organizational cybersecurity practices, we can strive towards creating resilient systems and tools that not only understand the intricacies of AI-generated social engineering attacks but also respond effectively to mitigate risks.

Furthermore, the results have wider ramifications for influencing future cybersecurity practices and legislation in addition to their immediate practical uses. The discovery that phishing and smishing are the two most common ways to use generative AI material emphasizes how attackers are changing their strategies. This research may be used by cybersecurity experts and policymakers to create preventative measures, educational programs, and legislative frameworks that will adjust to the evolving threats of social engineering. Furthermore, the study's human-centric approach highlights the necessity of a cybersecurity plan that addresses both technological flaws and the psychological elements that contribute to successful social engineering attacks. Including these observations in frameworks for policy promotes a stronger cybersecurity posture. To keep ahead of new difficulties in the dynamic field, researchers, industry players, and policymakers must continuously collaborate.

The other contribution is the findings on how to ensure end-user needs are met by different professions, which may impact how they detect and handle spam. Based on the correlation between user groups and awareness levels, organizations can tailor security advice and awareness campaigns based on the specific needs and expertise of different professional groups, therefore ensuring targeted and effective training programs. Moreover, companies developing chatbots for cybersecurity purposes can use the SUS questionnaire results to continually improve the usability of their tools. The feedback from users can guide developers in refining the chatbot's features, making it more intuitive and effective in detecting and mitigating AI-driven threats. In addition, organizations may improve their cybersecurity procedures and incident response plans by utilizing the grounded theory analysis derived from the qualitative data. The research participants' real-world experiences provide insightful information on the human elements involved in social engineering assaults. Through the integration of these insights into their incident response plans, companies may develop more efficient and flexible approaches to address AI-driven risks or threats. A detailed knowledge of the complexities involved in social engineering incidents is provided by the study of open code, which highlights elements like attack context, reasons for falling for attacks, advice on preventing attacks, attack strategies, methods of detection, and victims' reactions. The creation of focused protocols that handle the unique difficulties presented by AI-generated material in social engineering attempts can be guided by this detailed knowledge. The study reveals a changing landscape of social engineering strategies that might inform ongoing updates and improvements to these procedures, therefore providing enterprises with a more resilient and adaptable cybersecurity defense.

The significance of these findings lies in the deep insights into AI-driven social engineering attacks. Identifying phishing and smishing as predominant methods reveals key tactics used by attackers employing generative AI content. Beyond technical vulnerabilities, this research highlights the critical role of psychological factors in successful attacks, emphasizing the need for a multidisciplinary cybersecurity approach. The study not only detects AI-generated content but also emphasizes the importance of addressing human-centric aspects to mitigate social engineering threats effectively. The detailed analysis of user groups and the chatbot's effectiveness provides actionable insights for tailored security advice. Additionally, the research advocates for ongoing cybersecurity updates to adapt to the evolving landscape of social engineering, highlighting the necessity for resilient defenses against AI-driven threats in the digital landscape. Importantly, these findings corroborate existing literature, validate the implications of social engineering, and contribute to new knowledge by advancing theoretical understanding, uncovering novel insights, addressing practical challenges, and stimulating further scholarly inquiry in the field of AI-generated content in social engineering attacks.

In addition to these findings, it is essential to emphasize the importance of using AI-generated tools responsibly maximize their advantages while minimizing possible risks, as they grow more common in domains such as cybersecurity, especially in fighting social engineering attempts. In order for AI systems to successfully identify and counteract social engineering risks, users must have a clear understanding of how these systems make choices and what data they depend on. This requires transparency and explainability. To avoid misuse, such as the use of AI-generated material in dishonest social engineering schemes, ethical policies must direct the application of AI. It is important to maintain ongoing oversight and address biases in AI results to guarantee that these tools do not unintentionally worsen vulnerabilities. Furthermore, to prevent hackers from using sensitive data, it is essential to secure data privacy and put strong security measures in place. Maintaining human oversight is important to make sure AI technologies are used in conjunction with human judgment when detecting and responding to social engineering attempts. AI systems must be updated and improved on a regular basis to remain ahead of social engineering techniques that are always changing. Programs for education and awareness can provide users with additional tools to identify and control the risks and limits of artificial intelligence, particularly when it comes to social engineering attacks as it has constantly been shown that such attacks are not coming to an end any time soon.

B.Recommendations for future research

The following suggestions were reached for future work based on the findings of this research:

- **Enhanced Chatbot capabilities & Real-time interaction:**

A general Chatbot that can be used in detecting social engineering (SE) attacks by identifying suspicious requests, analyzing communication patterns, raising alerts, etc. through NLP and ML techniques. The chatbot should not only be able to observe real time interactions during email exchanges, messaging, or social media conversations, but should also be context-aware and able to identify SE cues.

- **Criminal Data Integration:**

Feed detection algorithms with information on known SE attackers. Through pattern analysis of criminal activity, the system can become more accurate in detecting spam and harmful intent. This will also keep the system updated on evolving threats.

- **Continuous Awareness Campaigns**

Open-access continuous awareness programs to be shared in educational and organizational institutions, as well as social media platforms. The more engaging the campaign’s content is (e.g. short ads, interactive quizzes, infographics, etc.) the further the audience reach.

- **User-Centric Approaches**

Verbal Training programs in the forms of presentation can often be tedious and lacking of practical showcases. The programs should be more user-driven where people can practice recognizing SE attempts via practical scenarios and interactive exercises. Moreover, feedback collection from users using security tools can be used to improve accuracy and user-friendliness.

- **Detection Enhancement Research**

Technology-based companies need to explore more techniques for identifying potential SE attacks such as considering behavioral biometrics (e.g., mouse and keystroke dynamics), contextual hints (e.g., daytime, location, device type), and combining text, image, and metadata for enhanced detection.

C.Limitation

The interviewees in this study were recruited from the same region (Al-Shahaniya, Qatar), and therefore, due to different cultures and traditions in different countries, the methods of SE attacks may differ. In addition, our age recruitment for the participants covered a few options, especially for the elderly who might have little awareness of SE attacks. Furthermore, extrapolating our findings to all users may be hampered by the small sample size of participants in our study. Once more, the spectrum of professions held by those in employment is unrestricted, as the training methods for SE threats may vary depending on the industry.

VI. Conclusion

Recently, AI-generated content has eased the mode of social engineering threats in society. AI-driven "power" has made social engineering attacks more successful than in past times. As artificial intelligence continues to advance, so does its capacity to create deceptive and tailored content through social engineering. Phishing and smishing are social engineering types that are significantly impacted by AI-generative tools. Certain AI-driven social engineering attack cases remain obscure to victims, underscoring the sophistication and deceptive capabilities of these emerging threats. Such tricks include phishing and smishing. These days, rather than unemployed users, users who are either employed or students frequently witness advancements in social engineering processes. The occupation classification of users can be associated with the degree of awareness and strategies employed by SE attackers. As a result, it's important to highlight the necessity of cyber security awareness in social media channels and enhance the way that automated applications detect social engineering assaults so that users of all stripes can benefit. Both organizations and individuals should be educated on information security awareness since it's an essential component in their lives. By mitigating this, we can work towards fortifying our digital defenses and mitigating the potential risks associated with the dynamic landscape of AI-driven social engineering attacks. The psychology of victims plays a key role in the success of social engineering attacks. Understanding the factors that contribute to successful social engineering attacks can help in developing more effective prevention and mitigation strategies. While chatbots and other automated systems have made great strides in identifying and mitigating different types of malicious activities, there is always room for improvement. Addressing the source of social engineering attacks can provide a more comprehensive defense strategy. For instance, we can incorporate advanced behavioral analysis techniques to detect anomalies in user interactions. For instance, in this case, if a user typically interacts with the system during certain hours but suddenly starts engaging at odd times, it could be flagged for further investigation. Integration of AI technologies into security operation centers enhances real-time analysis, threat identification, and response capabilities, fostering a more robust defense against AI-driven social engineering attacks.

VII. References

[1] Wu, J., Gan, W., Chen, Z., Wan, S., & Lin, H. (2023). Ai-generated content (aigc): A survey. arXiv preprint arXiv:2304.06632.

[2] Kaloudi, N., & Li, J. (2020). The ai-based cyber threat landscape: A survey. ACM Computing Surveys (CSUR), 53(1), 1-34.

[3] Wu, X., Duan, R., & Ni, J. (2023). Unveiling security, privacy, and ethical concerns of chatgpt. Journal of Information and Intelligence.

[4] Qi, Y., Shi, G., Yu, X., & Li, Y. (2015, June). Visualization in media big data analysis. In 2015 IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS) (pp. 571-574). IEEE.

[5] Sandeep, K. S., & Patil, N. (2018). A multidimensional approach to blog mining. In Progress in Intelligent Computing Techniques: Theory, Practice, and Applications: Proceedings of ICACNI 2016, Volume 2 (pp. 51-58). Springer Singapore.

[6] Tsirakis, N., Poulopoulos, V., Tsantilas, P., & Varlamis, I. (2017). Large scale opinion mining for social, news and blog data. Journal of Systems and Software, 127, 237-248.

[7] Delipetrev, B., Tsinaraki, C., & Kostic, U. (2020). Historical evolution of artificial intelligence.

[8] Jatobá, M., Santos, J., Gutierriz, I., Moscon, D., Fernandes, P. O., & Teixeira, J. P. (2019). Evolution of artificial intelligence research in human resources. Procedia Computer Science, 164, 137-142.

[9] Lu, Y. (2019). Artificial intelligence: a survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1-29.

[10] Mijwil, M. M., & Abttan, R. A. (2021). Artificial intelligence: a survey on evolution and future trends. Asian Journal of Applied Sciences, 9(2).

[11] Waltz, D. L. (2006). Evolution, sociobiology, and the future of artificial intelligence. IEEE Intelligent Systems, 21(3), 66-69.

[12] Binsaeed, K.; Stringhini, G.; Youssef, A.E. Detecting Spam in Twitter Microblogging Services: A Novel Machine Learning Approach based on Domain Popularity. Int. J. Adv. Comput. Sci. Appl. 2020, 11. [Google Scholar] [CrossRef]

[13] Canham, M., & Tuthill, J. (2022, June). Planting a Poison SEAD: Using Social Engineering Active Defense (SEAD) to Counter Cybercriminals. In International Conference on Human-Computer Interaction (pp. 48-57). Cham: Springer International Publishing.

[14] Basyoni, L., & Qadir, J. (2023, October). AI Generated Content in the Metaverse: Risks and Mitigation Strategies. In 2023 International Symposium on Networks, Computers and Communications (ISNCC) (pp. 1-4). IEEE.

[15] Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt. arXiv preprint arXiv:2303.04226.

[16] Naderifar, M., Goli, H., & Ghaljaie, F. (2017). Snowball sampling: A purposeful method of sampling in qualitative research. Strides in development of medical education, 14(3).

[17] Cridland, E.K.; Jones, S.C.; Caputi, P.; Magee, C.A. Qualitative research with families living with autism spectrum disorder: Recommendations for conducting semistructured interviews. J. Intellect. Dev. Disabil. 2015, 40, 78–91. [CrossRef]

[18] "Research reveals a rise in novel social engineering attacks," Digitalisation World, Apr. 06, 2023. https://m.digitalisationworld.com/news/65255/research-reveals-a-rise-in-novel-social-engineering-attacks (accessed Dec. 25, 2023).

[19] IBM, "What is Deep Learning?," www.ibm.com, 2023. https://www.ibm.com/topics/deep-learning

[20] "World Economic Forum Names Cybercrime and Cyber Insecurity Among Top 10 Global Risks for 2023," Tenable®, Feb. 17, 2023. https://www.tenable.com/blog/world-economic-forum-names-cybercrime-and-cyber-insecurity-among-top-10-global-risks-for-2023

[21] imperva, "What is Social Engineering | Attack Techniques & Prevention Methods | Imperva," Learning Center, 2019. https://www.imperva.com/learn/application-security/social-engineering-attack/

[22] "What are AI Generated Attacks?," https://mixmode.ai/. https://mixmode.ai/what-is/ai-generated-attacks/#:~:text=security%20at%20risk.-

[23] K. Merola, "Why are Social Engineering Attacks on the Rise? [Infographic]," Elevate Security, Oct. 27, 2022.

https://elevatesecurity.com/why-are-social-engineering-attacks-on-the-rise-infographic/

[24] K. Ferguson-Walter et al., “The Tularosa Study: An Experimental Design and Implementation to Quantify the Effectiveness of Cyber Deception.,” www.osti.gov, May 01, 2018. https://www.osti.gov/servlets/purl/1524844

[25] N. Kaloudi and J. Li, “The AI-Based Cyber Threat Landscape,” ACM Computing Surveys (CSUR), vol. 53, no. 1, pp. 1–34, Feb. 2020, doi: https://doi.org/10.1145/3372823.

[26] “What is a Social Media Bot? | Social Media Bot Definition | Cloudflare,” Cloudflare. Available: https://www.cloudflare.com/learning/bots/what-is-a-social-media-bot/#:~:text=Broadly%20speaking%2C%20social%20media%20bots

[27] Y. Alahmed, R. Abadla, N. Ameen, and A. Shteiwi, “Bridging the gap between ethical AI implementations,” International Journal of Membrane Science and Technology, vol. 10, no. 3, pp. 3034–3046, Oct. 2023, doi: 10.15379/ijmst.v10i3.2953.

[28] Y. Alahmed, R. Abadla, A. A. Badri and N. Ameen, "“How Does ChatGPT Work” Examining Functionality To The Creative AI CHATGPT on X's (Twitter) Platform," 2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS), Abu Dhabi, United Arab Emirates, 2023, pp. 1-7, doi: 10.1109/SNAMS60348.2023.10375450.Y. A. Ahmed and A. Sharo, “On the Education Effect of CHATGPT: Is AI CHATGPT to Dominate Education Career Profession?,” IEEE Xplore, Jun. 01, 2023. https://ieeexplore.ieee.org/document/10192993

[29] Y. A. Ahmed and A. Sharo, "On the Education Effect of CHATGPT: Is AI CHATGPT to Dominate Education Career Profession?," 2023 International Conference on Intelligent Computing, Communication, Networking and Services (ICCNS), Valencia, Spain, 2023, pp. 79-84, doi: 10.1109/ICCNS58795.2023.10192993.

[30] R. Abadla, A. Alseiari, A. Alheili, M. Sh. Daoud, and H. Al-Mimi, “Intelligent Phishing Email Detection with Multi-Feature Analysis (IPED-MFA),” ICCNS, Jun. 2023, doi: 10.1109/iccns58795.2023.10193714.

[31] H. Hesham, Y. Al Ahmed, B. Wael and M. Saleh, "Solar-Powered Smart Bin: Revolutionizing Waste Classification for a Sustainable Future," *2023 24th International Arab Conference on Information Technology (ACIT),* Ajman, United Arab Emirates, 2023, pp. 1-8, doi: 10.1109/ACIT58888.2023.10453850.