# A new approach to cybersecurity awareness: automated by GPT-4

# ABSTRACT

As our lives become more intertwined with the internet, our reliance on mobile devices continues to grow. With this increasing dependence comes a higher susceptibility to becoming targets of cyberattacks, as a lack of awareness of the risks associated with internet use continues to rise. To combat this, the current study aims to improve ongoing cybersecurity awareness programs by introducing a novel approach. This paper analyzes previous and ongoing cybersecurity awareness initiatives, delves into the advantages of security awareness programs, and outlines the most effective strategy for their implementation. The newly developed program is designed to encompass a variety of cybersecurity topics including password security, phishing awareness, social engineering, data privacy, and social media use, as well as safe internet practices. In this program, OpenAI’s GPT-4 is utilized to create an interactive module whereby users engage with multiple-choice questions, receive feedback based on their responses, and proceed to a new question based on their answers. Given its auto-generated nature, each participant experiences a unique program. The program's effectiveness will be measured by comparing the results of participants who undergo this training with those exposed to cybersecurity video lessons.

**Introduction**

Security awareness remains critical in the ever-evolving landscape of technology, where rapid advancements bring forth new cyber threats. As technology progresses, so does the need to improve our understanding of cybersecurity across all user levels. According to the National Institute of Standards and Technology (NIST), cybersecurity is defined as “the ability to protect or defend the use of cyberspace from cyber-attacks.” **(CITE)** Similarly, cybersecurity awareness can be defined as the degree to which every individual comprehends the significance of information security, aligns with their organization's security levels, understands their personal security duties, and acts in accordance. **(CITE)**This definition suggests that security awareness is intricately linked to behavior. Consequently, an individual who possesses an awareness of information security is expected to exhibit behavior that reflects this understanding. This combination of awareness and behavior underscores the role of education in enlightening individuals about information security and promoting the adoption of secure practices against cyber threats.

Habits in cybersecurity represent learned sequences of actions that become automatic responses, aiding in achieving specific security goals. **(CITE)** The performance of these habits demands minimal mental effort and attention and is therefore performed almost automatically. By extension, effective security behavior arises from consistently practicing these ingrained security habits with minimal cognitive effort. **(CITE)** Current efforts aimed at boosting security awareness encompass diverse approaches such as training platforms like Knowbe4 and GoldPhish, poster campaigns delivering security insights, easily accessible online guides, training through video games, AI-driven initiatives, and informative online videos.

In the development of this comprehensive security awareness program, the utilization of GPT-4, an advanced artificial intelligence model developed by OpenAI, represents a groundbreaking initiative. Capitalizing on the potential of GPT-4, this initiative aims to deliver an exceptional interactive learning experience, personalized to the needs of users. By leveraging GPT-4’s innovative technology, this program is poised to significantly enhance users' understanding of cybersecurity threats and increase their ability to navigate and mitigate potential risks in the digital landscape.

**Background and related work**

**Security anxiety**

Various studies collectively underline the escalating concerns about online privacy and data control, leading to heightened anxiety among internet users. The relationship between perceived threat, anxiety, and coping behaviors underscores the need for an intricate understanding of how individuals respond to internet-related stressors. Understanding these aspects can help tailor interventions to alleviate anxiety and promote safer internet practices.

**Understanding Internet Anxiety and Coping Behaviors:**

Numerous studies explore the anxiety associated with internet usage and its impact on behavior ([1](https://www.sciencedirect.com/science/article/pii/S0747563215300698), [2](https://journal.uii.ac.id/inCAF/article/view/27411/14689), [3](https://www.sciencedirect.com/science/article/pii/S0747563205000397)). In our rapidly evolving technological landscape, the anxiety stemming from the vulnerability of personal data online is a growing concern. It is evident from recent studies that this anxiety is not unusual; surveys and polls reveal a strong sense of concern among individuals regarding their data control and privacy. For instance, a recent Pew Research poll of 607 Americans highlighted that 91% of respondents expressed concerns about losing control over the collection and use of electronic data ([1](https://www.sciencedirect.com/science/article/pii/S0747563215300698)). Similarly, over two-thirds of a global survey sample reported being more worried about online privacy compared to the previous year’s findings ([1](https://www.sciencedirect.com/science/article/pii/S0747563215300698)).

Moreover, older adults, in particular, exhibit heightened feelings of vulnerability and discomfort concerning their online data security ([1](https://www.sciencedirect.com/science/article/pii/S0747563215300698)). This prevailing unease has prompted a growing recognition of the importance of safer internet practices. Consequently, there has been a notable surge in individuals expressing interest in protecting the privacy of their personal information online, as highlighted by the Pew poll, where 6 in 10 respondents indicated a desire to take more steps to safeguard their online privacy ([1](https://www.sciencedirect.com/science/article/pii/S0747563215300698)).

**Perceived Threat, Anxiety, and Coping Behaviors:**

Intriguingly, comprehensive studies exploring the relationship between perceived threats, internet anxiety, and coping behaviors among e-banking users ([2](https://journal.uii.ac.id/inCAF/article/view/27411/14689)) reveal influential findings. These findings consistently demonstrate the significant impact of perceived threats and internet-related anxiety on the adoption of protective measures and the willingness to seek assistance ([2](https://journal.uii.ac.id/inCAF/article/view/27411/14689)). However, while these factors strongly influence protective actions and assistance-seeking tendencies, they seem to have a less evident effect on avoidance behaviors in e-banking users ([2](https://journal.uii.ac.id/inCAF/article/view/27411/14689)). Furthermore, the correlation between perceived threats and anxiety significantly shapes users' coping strategies, exerting a noticeable influence on their approaches to safeguarding their online presence ([2](https://journal.uii.ac.id/inCAF/article/view/27411/14689)). The research findings emphasize how users' perceptions and emotional reactions contribute to shaping their behaviors, highlighting the intricate dynamics influencing online security practices within the e-banking community.

This research highlights a critical aspect: despite heightened anxiety, individuals continue to utilize the internet, engaging in online activities despite reduced participation in secure practices, as evidenced by the negative correlation between anxiety and secure behaviors (2). This paradoxical behavior intensifies the pressing need for a comprehensive enhancement in security awareness initiatives, emphasizing the urgency of educating users about the importance of secure online habits, even in the presence of heightened anxiety surrounding online threats.

**Understanding Internet Anxiety:**

A comprehensive study centered on the correlation between internet identification and anxiety aimed to create a specialized measurement tool similar to existing computer anxiety scales ([3](https://www.sciencedirect.com/science/article/pii/S0747563205000397)). This in-depth survey introduced a novel scale comprising six questions to assess varying levels of anxiety associated with internet usage, thereby revealing distinct dimensions related to individuals' anxieties ([3](https://www.sciencedirect.com/science/article/pii/S0747563205000397)). These dimensions encompassed not only the level of anxiety but also explored the behavioral aspects of anxiety avoidance strategies adopted by users ([3](https://www.sciencedirect.com/science/article/pii/S0747563205000397)). Additionally, the survey played a pivotal role in uncovering the significance attributed to internet usage, shedding light on the complex attitudes individuals hold towards their online interactions and activities ([3](https://www.sciencedirect.com/science/article/pii/S0747563205000397)).

The questions utilized in this investigation specifically targeted users' feelings and behaviors regarding internet usage:

1. I always feel anxious when using the Internet.
2. I go out of my way to avoid using the Internet.
3. It is easy for me to use the Internet.
4. It is important for me to be able to use the Internet.
5. My anxiety about using the Internet bothers me.
6. I am more anxious about using the Internet than I should be ([3](https://www.sciencedirect.com/science/article/pii/S0747563205000397)).

**Motivation**

Within the realm of information security, a comprehensive framework incorporating the Protection Motivation Theory (PMT), social norms, and habitual factors reveals insights into individual behaviors. PMT, coupled with practical studies among university students and government employees, illustrates the dynamics behind security habits. These studies underscore the evolving nature of behaviors and the subtle interplay of motivations and perceptions in fostering strong security practices. (17)

**Protection Motivation Theory (PMT) in Security Behaviors:**

The Protection Motivation Theory integrated with social norms and habitual factors provides a strong framework for examining security behaviors (17). This theory emphasizes the instinctive nature of security habits, defined as learned sequences of acts that become automatic responses, requiring minimal mental effort (17). Constructs within PMT, such as perceived vulnerability, severity, response efficacy, costs, and self-efficacy, constitute vital elements influencing individuals' intentions to engage in information security practices (17). Studies rooted in PMT, notably among university students in South Korea, validate its effectiveness in predicting information security behaviors, emphasizing the pivotal roles of response efficacy and self-efficacy (17).

**Factors Influencing Government Employees' Cybersecurity Behavior:**

A comprehensive exploration into the factors that impact cybersecurity behaviors among government employees in Malaysia reaffirms the profound significance of several critical elements such as motivation, severity, vulnerability, and perceptions of efficacy (18). This in-depth study not only underscores but amplifies the pivotal role played by the perceptions of threat severity and vulnerability in instigating pivotal behavioral shifts towards stronger cybersecurity measures (18). It notably highlights how users' perceptions, specifically regarding vulnerability and severity, act as catalysts for profound changes in behavior, shaping their responses to threats and significantly influencing the prediction of cybersecurity behaviors among employees (18).

**Evolution of Security Habits:**

Other studies present a compelling narrative that sheds light on the dynamic evolution observed in security behaviors, depicting a transformative journey starting from the initial recognition of external threats towards a more profound internalization and habitualization of safeguarding practices over time (17, 18). This evolution underlines a gradual yet discernible shift, indicating the progressive implementation of security habits into daily routines, a transformation rooted in the repetition of behaviors and the accumulation of varied experiences and encounters with security protocols and threats. This shift represents a remarkable transformation from simply recognizing external risks to embedding an internal, habitual, and sophisticated approach to safeguarding information. It signifies a notable developmental progression in individuals' cybersecurity practices and the cultivation of their mindset. (17, 18).

**Need for Nuanced Understanding and Behavioral Adaptation:**

Collectively, these observations distinctly highlight the imperative need for a comprehensive understanding of human behaviors and perceptions regarding cybersecurity (17, 18). They highlight the intricate relationship between motivations, beliefs, and habitual reactions, essential in nurturing resilient information security practices. Moreover, they emphasize the vital role of adaptability in cybersecurity strategies, critical for effectively addressing the ever-evolving landscape of threats and user behaviors.

The combination of PMT, social norms, habitual elements, and user perspectives creates an encompassing framework. This framework not only facilitates understanding but also enables the prediction of information security behaviors across diverse populations. Consequently, it significantly contributes to the ongoing discourse on enhancing security awareness.

**Security awareness**

In examining cybersecurity awareness, it becomes evident that tailoring campaigns to suit diverse audiences and utilizing interactive, motivation-centered approaches is important. This necessitates a continuous educational process and adaptive training methodologies to stimulate favorable behavioral shifts. Acknowledging the evolving landscape of cyber threats and human behavior underscores the imperative nature of these proactive measures in fostering a more resilient cybersecurity culture.

**Creating Effective Security Awareness Campaigns:**

Within the realm of security awareness campaigns, a multitude of studies have explored the essence of security awareness campaigns, addressing various aspects of cyber threats and human behavior (4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16). At its core, security awareness embodies a comprehensive understanding of information security's significance, along with individual responsibilities and behavioral adaptations (4). This distinction becomes pronounced when compared to healthcare awareness initiatives: while ignorance in healthcare campaigns may predominantly affect an individual, overlooking the fundamentals of security awareness extends its consequences beyond personal realms, penetrating entire networks and organizational structures (4). This unique characteristic amplifies the collective impact and societal repercussions of lapses in security consciousness, outlining the critical need for comprehensive security awareness initiatives.

**Models for Effective Campaigns:**

Extensive research conducted across various studies consistently underscores the effectiveness of models such as the Information-Motivation-Behavioral (IMB) framework in prompting desired behavioral shifts (4, 5). What sets the IMB model apart is its dual focus on knowledge and motivation, proving significantly more effective than campaigns reliant solely on disseminating information (4). Practically implementing the IMB framework involves employing diverse communication channels such as presentations, magazines, and similar mediums to disseminate knowledge. Simultaneously, there's an intentional emphasis on highlighting the benefits associated with adopting secure behavioral practices, thereby nurturing motivation among participants (4). Moreover, harnessing social interactions, informal platforms, and normative feedback emerges as a pivotal driver for inducing behavioral changes (4). Specifically, the incorporation of descriptive and injunctive social norms within these campaigns serves to clarify acceptable behavior patterns and provide tangible examples within the organizational context, catalyzing a commendable shift in employee conduct (4).

**Metrics for Evaluating Security Awareness Programs**

Security awareness campaigns are crucial in cultivating a robust cybersecurity culture. Beyond their implementation, the efficacy of these initiatives relies on the ability to measure their impact and success. Introducing quantifiable metrics becomes imperative within these campaigns to determine their effectiveness in fostering secure behaviors and mitigating cyber threats.

Metrics for evaluating these campaigns can encompass a variety of parameters, ranging from observable behavioral changes among employees or users to the tangible reduction in security incidents post-implementation. For instance, tracking the frequency of password updates, incidents reported, or participation rates in training modules offers insight into the behavioral changes resulting from awareness programs (6, 8, 13). Simultaneously, analyzing incident response times or the decline in successful phishing attempts provides tangible metrics regarding incident prevention and mitigation (11, 13).

Moreover, engagement metrics, including completion rates of training modules and user feedback on the relevance and effectiveness of awareness materials, offer valuable insights into the campaign's reception and efficacy among participants (4, 7, 9, 14).

By integrating these measurable metrics into the framework of security awareness campaigns, organizations can assess, refine, and continually enhance their initiatives to ensure a more resilient cybersecurity posture.

**Targeted Training and Awareness:**

Numerous studies have extensively explored the realm of targeted cybersecurity training tailored for specific populations, including college students (6), government employees (7), individuals in small and medium enterprises (SMEs) (7), and end-users (8, 9, 10, 11, 12). These comprehensive investigations center on critical subjects like strengthening secure password practices, safeguarding against phishing attempts, implementing robust data protection measures, and understanding the intricacies of social engineering (7, 9). What emerges prominently from these studies is the pressing need for a multidisciplinary approach to training, emphasizing continual education and precisely tailored interventions that directly address the ever-evolving landscape of cybersecurity threats (5, 6, 7, 8, 9).

Top of Form

**Knowledge Impact on Security Practices:**

Investigations into the influence of knowledge on security practices consistently underline a direct correlation between a comprehensive understanding of cyber threats and an elevated sense of awareness (8, 13). A deeper understanding significantly contributes to the accurate identification of potential threats, empowering individuals to discern between benign and malicious events with greater precision (8). Nevertheless, these studies also underscore the vital requirement for continual learning and adaptability in response to the constantly evolving and dynamic nature of cyber threats (8).

**Role of AI in Security Awareness:**

AI-driven cybersecurity awareness emerges as a pivotal strategy in strengthening defenses against cyber threats, with its significant efficiency notably demonstrated in reducing the widespread threat posed by phishing attacks (11). The remarkable aspect of AI lies in its capacity to discern intricate patterns present in various cyber threats, empowering it to swiftly detect and respond to evolving attack methods (11). By continuously learning and adapting from past incidents, AI systems exhibit a dynamic learning curve, constantly improving their ability to identify and neutralize threats effectively (11). This process of adaptive learning guarantees a proactive approach to prevent potential vulnerabilities, providing a robust defense against intricate cyber threats. Top of Form

Leveraging AI's learning capabilities in cybersecurity awareness not only enhances threat detection but also improves the overall resilience of digital ecosystems in safeguarding sensitive information and systems from malicious exploits.

**Challenges and Future Directions:**

These comprehensive studies shed light on the intricate challenges encountered while implementing effective training programs, highlighting the limitations within modern training methodologies and the diverse array of environmental factors at play (14, 15). A key takeaway from these findings highlights the urgent need for holistic programs flawlessly designed to cater to the distinct variations within different organizational and educational settings (14, 15).

**Proposal**

The primary objective of this study is to develop an innovative approach to cybersecurity awareness distinct from existing initiatives found on platforms like Knowbe4 and GoldPhish. Unlike traditional informative videos that often result in limited retention of information, this novel approach integrates GPT-4 technology to create a unique training module for enhancing cybersecurity awareness. Users engage with the program by answering multiple-choice questions and receiving tailored feedback based on their responses. Utilizing GPT-4 enables the program to generate questions aligned with user interactions, ensuring diverse training experiences while focusing on fundamental topics such as password security, social engineering, phishing awareness, data privacy, and safe internet use.

To evaluate the efficacy of this program, the study will analyze pre-defined pre-test and post-test questions, meticulously selected after multiple interactions with OpenAI’s GPT-3.5 to determine the most effective inquiries. These questions will remain unchanged for every user, ensuring uniformity in the assessment process. These assessments will also include an evaluation of users' security anxiety and motivation to practice secure habits before and after the training. Additionally, these outcomes will be compared with those of users assigned to view free online cybersecurity awareness videos sourced from KnowBe4 and/or GoldPhish. Carefully selected videos from these platforms will serve as a benchmark, allowing for an accurate comparison with the program's interactive approach.

Anticipated findings predict significantly improved results among participants engaging in the interactive training compared to those who solely review videos. To encourage participation, the program will be administered under the psychology department's SONA system. This platform offers psychology students an opportunity to gain credit for study participation, as the study involves a psychological component inherent in assessing user responses. Each participant will be randomly assigned to either the module or video group to ensure an unbiased evaluation. The pre-test and post-test evaluations will be administered using the Qualtrics platform. Upon completion of these assessments, participants will be directed to the interactive training program, which will be hosted on a dedicated web page specifically coded and designed for this purpose. This webpage will act as the access point for participants to engage with the training content, offering two distinct options: access to either the video-based materials or the interactive training module.

In conclusion, this study aims to revolutionize cybersecurity awareness by employing an interactive training approach, leveraging GPT-4 technology within a psychology-focused study environment.

**Partial implementation**

Initial research aimed to identify the top five cybersecurity attacks that college students fall for as well as five essential topics that are crucial for education to defend against these specific attacks. Analyzing past research and insights from OpenAI’s GPT-3.5 on this matter resulted in a finalized list of essential security topics: phishing awareness, password security, social engineering, data privacy and social media use, as well as safe internet use. Figures 1 and 2 showcase GPT-3.5 insights for targeted cybersecurity education.

These topics formed the basis for developing pre-test and post-test questions, which were created with the assistance of GPT-3.5. These were designed to assess each participant's learning progress by comparing their knowledge before and after the training sessions. The topics also formed the foundation for selecting KnowBe4 and GoldPhish informative security videos which participants in the control group will view during their training.

Figure 1 showcases an example of a pre-test question generated by GPT-3.5.

Prompt: Generate 4 multiple-choice questions that would give insight into how much knowledge a person has about password security as well as an

idea of their current password creation processes/behaviors.

Answer:

Which of the following is considered a strong password?

a) Password123

b) P@ssw0rd!

c) MyDogL0vesP@ssw0rd$!

d) 12345678

The OpenAI playground was utilized to test GPT-4, assessing its suitability for this program and gaining an understanding of its expected operations behind the scenes throughout the training. However, initial experiments revealed inaccuracies: GPT-4 couldn't reliably indicate correct or incorrect answers and didn't adhere to the system prompt's topic-switching requirements.

An example of where the OpenAI Playground loses state and provides incorrect results is as follows:

System: You are a cybersecurity specialist educating college students on cybersecurity awareness. You will ask them multiple choice (A, B, C, D) questions on 5 different topics (Phishing Awareness, Password Security, Social Engineering, Safe Internet Use and Data Privacy and Social Media Awareness). You will wait for the user to answer the question (they will answer with A B C or D) then create new questions based on their response. You will start a new topic only if they have answered 3 questions correctly for that topic. You will provide feedback after each response.

Assistant:

11. Which of the following is not a safe internet practice?

A. Ignoring software updates

B. Using strong and unique passwords

C. Regularly backing up critical data

D. Limiting the amount of personal information shared online

\*\*\*\*\*\*\*\*\*\*\*LET’S ANSWER QUESTION 3 WRONG TO SEE IF IT STAYS ON THE SAME TOPIC\*\*\*\*\*\*\*\*\*\*\*

User:

B

Assistant:

Good job! The correct answer is (A). Ignoring software updates is not a safe internet practice. Updates often contain important security patches that help protect your device and your information from new threats.

\*\*\*\*\*\*\*\*\*\*\*SAID THAT THE ANSWER WAS RIGHT WHEN IT WAS WRONG, STARTED A NEW TOPIC FOLLOWING THIS\*\*\*\*\*\*\*\*\*\*\*

These initial experiments highlighted that GPT-4 cannot be treated as an agent and lacks the ability to maintain state. Subsequent trials involved treating GPT-4 as a function by breaking down system prompts to consult GPT-4 for single steps, assigning it one straightforward task at a time while manually managing state. Troubleshooting was conducted while experimenting with various temperatures which regulate GPT-4 creativity, with a temperature of 0.5 proving to be of optimal efficiency. Upon achieving satisfaction, efforts transitioned to coding Python program to mirror the OpenAI GPT-4 playground experiments. A comprehensive code was developed to generate questions, receive user responses, provide feedback, and initiate a new topic after either three correctly answered questions or the completion of five questions within that topic. The code maintains records of correct answers and total questions answered to ensure optimal functionality. Figure 4 showcases a screenshot of the program’s functionality in the back-end.

**Conclusion**

The extensive exploration into cybersecurity awareness reveals a diverse landscape requiring tailored interventions. Numerous studies emphasize the necessity of comprehensive approaches, highlighting the pivotal role of motivation-centered strategies in prompting behavioral changes. The efficacy of models like the Information-Motivation-Behavioral framework stands out, underscoring the critical role of knowledge and motivation in fostering desired behavioral changes. Targeted training programs, ranging from specialized interventions for distinct populations to comprehensive cybersecurity education, emerge as vital foundations in increasing cybersecurity consciousness.

Moreover, the integration of AI-driven initiatives, exemplified by the promising strides in mitigating phishing attacks, underscores the potential of adaptive learning processes in improving cybersecurity defenses. The pivotal role of knowledge in influencing security practices underscores the need for continual education to navigate the ever-evolving threat landscape effectively. Challenges within modern training methodologies and environmental factors necessitate the design of holistic, adaptable programs aligned with diverse organizational settings.

The proposed study presents a novel approach to cybersecurity awareness, leveraging GPT-4 to craft unique training modules aimed at enhancing users' security consciousness. This pioneering program, diverging from conventional video-based platforms, focuses on interactive modules, and personalized feedback, offering a promising pathway for improving cybersecurity preparedness. By employing carefully selected questions, uniform for every user, the study seeks to unravel the program's effectiveness, drawing comparisons against established platforms like KnowBe4 and GoldPhish.

In essence, the convergence of innovative methods, ranging from AI-driven interventions to tailored educational programs, signifies a crucial turning point in shaping a resilient cybersecurity culture. The collective insights gained from these diverse studies form a robust foundation for future endeavors, propelling towards a more skillful, knowledgeable, and watchful stance in cybersecurity awareness.

Top of Form