

Enhancing AI-driven Cybersecurity Awareness Smart Consultant using RAG Method with Hybrid Knowledge Based

I Putu Agus Eka Pratama¹, I Made Oka Widyantara², Linawati³, Nyoman Gunantara⁴

¹Doctoral Program of Engineering Sciences, Faculty of Engineering, Udayana University, Bali, Indonesia

^{2,3,4}Department of Electrical Engineering, Faculty of Engineering, Udayana University, Bali, Indonesia

Article Info

Article history:

Received Dec 10, 2025

Revised Jan 04, 2026

Accepted Jan 14, 2026

Keywords:

Answer Accuracy Score (AAS)

AI-driven

Cybersecurity awareness

Hybrid knowledge based

Retrieval Augmented

Generation (RAG)

ABSTRACT

The rapid advancement of Artificial Intelligence (AI), particularly in the realm of Large Language Model (LLM), holds significant potential for addressing the escalating issue of cyberattacks that exploit users' insufficient cybersecurity awareness. This study involves the design and development of a prototype cybersecurity awareness smart consultant, leveraging AI through the Retrieval Augmented Generation (RAG) method. This approach integrates hybrid knowledge derived from both user-specific internal cybersecurity documents and internet resources, thereby enhancing the validity of system responses and mitigating the risk of hallucinations. The prototype was evaluated using the Answer Accuracy Score (AAS) method, based on Black Box Testing and human evaluation, across four cybersecurity-related questions, yielding an average score of 0.925, accompanied by comprehensive analysis and discussion. The findings indicate that the system's response accuracy improves when knowledge is synthesized from both internal document resources and internet sources. Future research may focus on incorporating deliberative thinking to further enhance system performance in generating responses.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

I Made Oka Widyantara

Department of Electrical Engineering, Faculty of Engineering,

Udayana University,

Jl. Raya Kampus Unud, Jimbaran, Bali, Indonesia.

Email: oka.widyantara@unud.ac.id

1. INTRODUCTION

The deficiency in cybersecurity awareness among users significantly contributes to the escalation of cyberattacks that exploit vulnerabilities within the end-user environment [1]. These cyberattacks manifest in various forms, including phishing, account hacking and misuse, and personal data theft [2]. This underscores the critical need for cybersecurity education, training, and consultation to enhance user awareness [3]. Although several solutions exist to address this issue, such as training [4], tutorials and guides [5], and discussion forums [6], they are not optimal for increasing user awareness due to a lack of user literacy, gaps in technological understanding, and the absence of prompt and accessible cybersecurity consultation services. This is particularly regrettable given the wealth of available documentation, reports, and guides.

Artificial Intelligence (AI), specifically through Large Language Model (LLM), holds potential for augmenting user awareness via a consultation prototype or chatbot [7], enabling users to pose cybersecurity-related inquiries and receive responses. However, LLM are susceptible to a phenomenon known as hallucinations, where the system's responses appear accurate but are, in fact, erroneous and inconsistent with factual information. Consequently, hallucinations can impede the process of cybersecurity awareness education [8].

In light of the aforementioned context, this study involves the design and development of an AI-based cybersecurity awareness prototype utilizing the Retrieval Augmented Generation

(RAG) method and a hybrid knowledge base. Numerous prior studies have employed hybrid knowledge-based RAG across various case studies and objectives, including the integration of Graph RAG and Vector RAG [9], the development of Secure Multifaceted RAG for enhanced security during the retrieval phase [10], the improvement of precision in the multi-objective mix design of dam concrete with a focus on strength [11], the combination with embedding vector databases [12], and the application of hybrid parameters for AI legality [13]. From these studies, we have identified a research gap: no existing research within the cybersecurity domain has utilized AI in conjunction with the RAG method and a hybrid knowledge base to augment users' cybersecurity awareness. Additionally, we considered the potential of users or organizations possessing extensive cybersecurity document resources, which can be integrated with online resources to furnish more valid system responses related to education, thereby improving users' cybersecurity awareness. Our objective was to evaluate the accuracy of the responses generated by the hybrid knowledge-based system in relation to a set of questions we developed.

2. RESEARCH METHOD

2.1. State of the Art

Numerous prior studies have investigated the application of LLM to enhance cybersecurity awareness among users across various case studies. Research by [14], developed an LLM-based application to enhance cybersecurity awareness among university students, specifically targeting phishing detection and recognition. This research contributes to LLM-based cybersecurity education, yet it does not address hallucinations or improve system response accuracy. Research by [15], developed the CyExec based on LLM, to simulate multiple cyberattack scenarios and assess the system's response capabilities. This research underscores the potential of LLM for cybersecurity awareness in an application format; however, it does not address the issue of hallucinations. Research by [16], introduced a prototype aimed at enhancing cybersecurity awareness, particularly in computer network security, among web users and administrators by utilizing HTTPS data processed through an LLM. This study offers an innovative contribution to the application of LLM for network security awareness; nevertheless, it fails to address hallucinations in LLM. Research by [17], conducted a systematic review and evaluation of various LLM providers, focusing on the strengths and weaknesses of LLM in terms of contextual understanding and their propensity for hallucinations. This study contributes to the existing literature.

To address the challenge of hallucinations in LLM, several studies have implemented RAG within the domain of cybersecurity awareness, given RAG's capability to manage hallucinations [18]. Singh [19], developed a prototype based on RAG to enhance user awareness of cybersecurity, particularly in threat intelligence, utilizing unstructured data from social media. This research contributes to the design and implementation of RAG for cybersecurity awareness, specifically in threat intelligence, but is limited to using unstructured data from social media. Chengshuai [20], developed a RAG-based application incorporating external knowledge and ontology for Question and Answer (QA) systems in cybersecurity education. This research advances the application of RAG and ontology in cybersecurity awareness but is reliant solely on external knowledge. Research by [21], developed a cyberattack classification and reporting application based on RAG and BERT to enhance the capabilities of Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS) and to increase cybersecurity awareness among users, particularly system administrators. This research contributes to the enhancement of AI-based cybersecurity awareness but exhibits limitations in cyberattack detection and knowledge related to cyberattack types, and is restricted to system administrator users. Rajapaksha [22], developed a RAG-based QA system specializing in educating users about cybersecurity awareness related to cyberattack investigation. This research contributes to the utilization of RAG for QA cybersecurity awareness education but is constrained by the limitations in the knowledge employed.

In our previous research, we conducted a bibliometric analysis concerning the design and development of a prototype aimed at enhancing cybersecurity awareness among users, specifically within the healthcare sector [23]. Furthermore, we undertook preliminary research to develop a RAG-based prototype aimed at enhancing cybersecurity awareness training for users [24].

Additionally, we incorporated a fixed chunking solution into the internal knowledge base [25]. However, this effort was limited by its reliance solely on internal documents as its knowledge base. In this study, we have refined our previous research by employing a hybrid knowledge-based approach, which integrates both internal and external knowledge sources.

2.2. Quantitative Method

This study utilized a quantitative research methodology that integrated both primary and secondary data sources. The primary data were derived from internal user and organizational information pertaining to cybersecurity, specifically in the form of guidelines and policies, which constitute the system's internal knowledge base. Additionally, the primary data in this study included results from prototype testing. The secondary data were obtained from articles in international journals and conference proceedings relevant to the research topic, serving as references and representing the state of the art in the field.

2.3. Research Methodology

The research methodology employed in this study comprised six sequential steps: problem identification and determination of the research objective, literature review, design and implementation, testing, analysis and discussion, and documentation. Figure 1. presents the research methodology in a flowchart format.

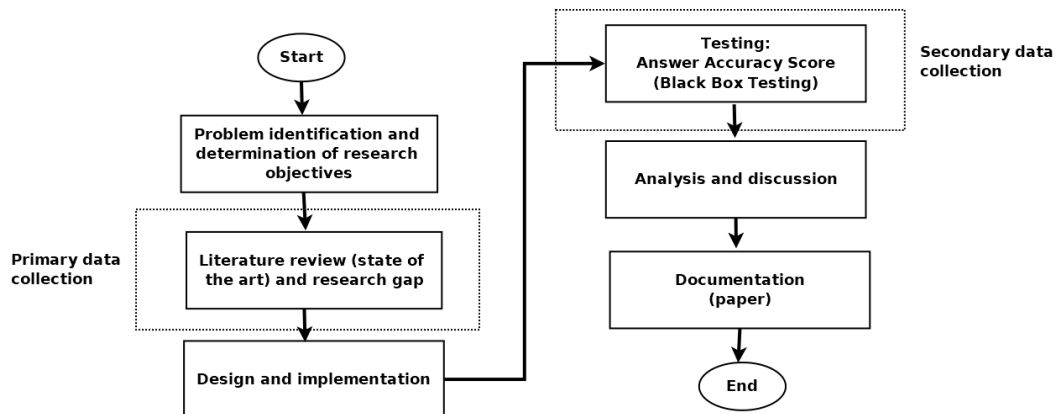


Figure 1. Research methodology (in Flowchart)

As depicted in Figure 1., the problem identification phase involved recognizing existing issues related to insufficient cybersecurity awareness among users, which has led to an increase in cyberattacks targeting users and the suboptimal performance of current solutions. The research objective was established by designing and implementing an AI-based cybersecurity awareness consultant prototype utilizing the RAG method to enhance user cybersecurity awareness by optimizing the use of external and internal knowledge bases. A literature review was conducted, encompassing papers from scientific journal proceedings pertinent to the research topic, to fortify the state of the art and identify research gaps. Based on this review, a prototype was designed and implemented (prototyping).

The prototyping process employed the Python programming language, Node.js, the pg-vector vector database library, YugaByte, and the Pinecone indexing tool, all operating on the Linux system. The AI model utilized was GPT-4o (omni) provided by OpenAI, with a RAG temperature setting of 0 and a maximum token size of 128K. Testing was conducted using the Answer Accuracy Score (AAS) method based on Black Box Testing on the prototype, involving 4 cybersecurity-related questions. The test results were subsequently analyzed and discussed. Finally, the research documentation was compiled.

2.4. Retrieval Augmented Generation (RAG) Method

In this study, we employed the RAG method on an LLM engine to develop a prototype cybersecurity awareness consultant. LLM is a deep learning-based AI that utilizes transformers, large datasets, complex language structures, and semantic relationships between words and phrases to generate responses [26]. However, LLM possesses limited knowledge and is therefore susceptible to hallucinations. The RAG method was developed as a solution to address hallucinations in LLM by integrating retrieval steps from the knowledge base with augmentation and generation [27]. RAG is classified into three types based on pre-retrieval and retrieval steps and module utilization: Naive RAG, Advanced RAG, and Modular RAG [28].

This study utilized Advanced RAG, which involves four steps: Pre-Retrieval, Retrieval, Augmentation, and Generation [29]. Employing a hybrid knowledge-based approach that combines internal knowledge from PDF documents and external knowledge from the internet (documents, URLs), the Pre-Retrieval stage decomposes the internal knowledge into sub-documents (chunks), converts them into vectors (vectorization) using the doc2vec method, and subsequently stores them in a vector database. The storage of these chunks necessitates indexing. Furthermore, each chunk was refined. The Retrieval stage extracts information and knowledge from user queries, embedding them, and determining whether the user's questions are relevant to the available knowledge in the vector database. This process prepares for the augmentation and generation stages, which involve system responses. The Augmentation stage aligns user questions with the available knowledge in the vector database using the cosine similarity method. The Generation stage constructs system responses to be forwarded to the user based on the context of the question posed.

2.5. Hybrid Knowledge Based

The concept of employing hybrid knowledge bases in this research is informed by several prior studies that have utilized multiple knowledge-based sources in Retrieval-Augmented Generation (RAG)-based Large Language Model (LLM) systems across various case studies. The study by [30], implements multi-source knowledge pruning, integrating both internal and external knowledge bases to complement each other. This approach enables RAG to access both knowledge bases to gather information, thereby enhancing the system's ability to compile and complete responses to user inquiries. The research by [31], developed a RAG-based DORAG framework utilizing the Deepseek engine for Question Answering (QA) case studies. This framework employs a combination of knowledge graph and semantic vector retrieval to augment RAG's capabilities during the query process prior to the generation stage, thus improving the precision and relevance of the responses. The study by [32], introduces HybGRAG, a hybrid RAG system that integrates textual and relational knowledge bases with hybrid question-answering capabilities. This system requires both textual and relational information from a semi-structured knowledge base, supported by a retriever bank and critical module, to ensure adaptability and interpretability. The research by [33], conducted a literature review to demonstrate the efficacy of domain-specific RAG, utilizing both external and internal knowledge-based approaches to address the limitations of generic LLM. Meanwhile [34], developed a prototype LLM based on the RAG method, employing a hybrid knowledge base in a manufacturing and industrial case study.

In our cybersecurity awareness case study, the extensive availability of cybersecurity reference sources online, along with internal organizational or user resources (such as documents, reports, and guides), presents an opportunity for optimization to support AI-based cybersecurity education processes. The integration of internal and external knowledge allows the AI-based cybersecurity awareness consultant system developed in this study to employ the RAG method, thereby providing more reliable, relevant, and valid responses. The comprehensive hybrid knowledge-based process flow, encompassing the transition from input to output, as proposed in this study, is depicted in Figure 2.

As illustrated in Figure 2., the process initiates with the acquisition of internal knowledge. Users have the capability to manually incorporate documents, reports, and guidelines pertinent to cybersecurity into the system. During the Pre-Retrieval stage, the system executes chunking, embedding, and document-to-vector (indexing) operations, which are subsequently stored in a

vector database. Users may submit inquiries via a graphical user interface (GUI), and the query is processed by the LLM engine during the retrieval stage, which seeks matches within the internal knowledge-based data housed in the vector database. The system also leverages external knowledge-based data (the internet) to supplement knowledge relevant to the user's query, integrating it with the results from the internal knowledge-based data. Subsequently, in the Augmented stage, the LLM engine processes the amalgamated results of the internal and external knowledge-based data using cosine similarity and reranking to derive the most valid results based on the context of the user's query. This stage then advances to the generation stage, where the system compiles the response, which is subsequently presented to the user via the GUI.

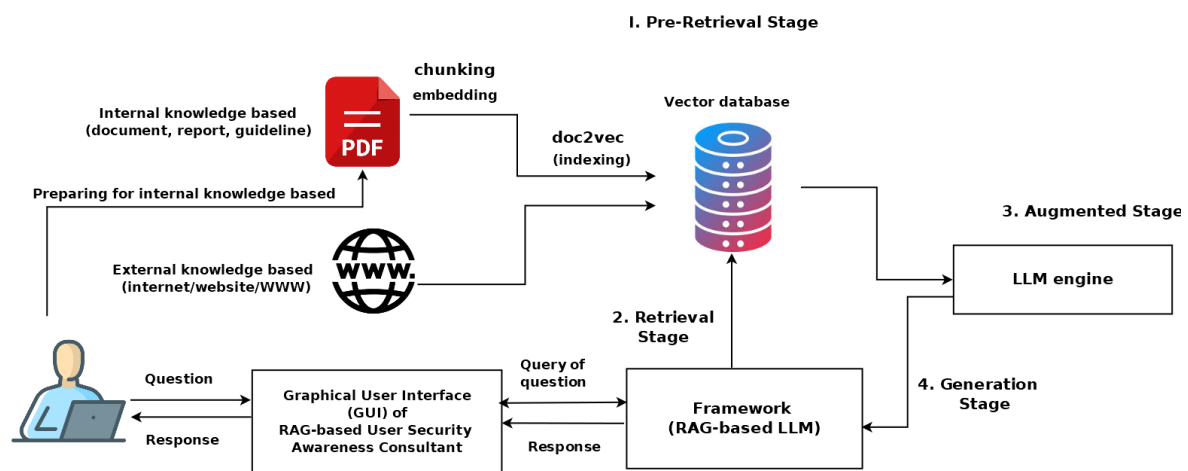


Figure 2. The process of hybrid knowledge based on RAG-based User Security Awareness Consultant

2.6. Testing Method

The methodology employed in this study is Black Box Testing, which is utilized to evaluate the functionality of the constructed prototype [35]. In this study, the primary focus was on evaluating the prototype's capability to generate responses to user inquiries and its accuracy, as determined by a scoring system. Consequently, the Black Box Testing variant employed was the Answer Accuracy Score (AAS). In various scientific literature, AAS is also referred to as answer accuracy, factual accuracy, or truth score, which involves human evaluation. It serves as a testing method and evaluation metric based on Black Box Testing, aimed at assessing the functionality of LLM and RAG-based systems. This was accomplished by measuring the accuracy of the system's responses, which involved calculating the proportion of correct answers relative to the total number of questions asked, based on predetermined criteria to obtain a percentage value [36]. AAS is frequently utilized in LLM and RAG-based systems to evaluate the retrieval ability and accuracy of answers.

In this study, the human evaluation was conducted utilizing Black Box Testing methodologies. As researchers with specialized knowledge in cybersecurity, we assumed the role of internal evaluators to assess the accuracy of the system's responses, deliberately excluding external evaluators. Our evaluation concentrated on determining whether the responses were contextually appropriate to the questions and reference sources, as well as assessing the completeness and accuracy of the answers provided.

Several prior studies have employed different methods related to AAS to evaluate the RAG-based system response. Research by [37], utilized a QA-driven evaluation method based on summaries and references, with an emphasis on factual consistency. Research by [38], employed an automatic metric method and enhanced the summarization model for factual consistency. Research by [39] proposed a scoring model for evaluating factual consistency. Research by [40], introduced a textual entailment method for QA evaluation, particularly for factual correctness. Meanwhile, Huang [41], proposed an alignment and completeness evaluation method for assessing factuality in system responses with extensive character content. Unfortunately, these studies predominantly focus on a singular aspect, particularly factual consistency and correctness.

The AAS evaluation encompasses four key components: 1.) F (factual accuracy), which pertains to the precision of the system's responses based on factual data and reference sources; 2.) S (semantic alignment), which assesses the appropriateness of the response in relation to the question's context; 3.) C (completeness and specificity), which evaluate the level of detail and thoroughness of the response in accordance with the question's context; and 4.) G (groundedness), which refers to the absence of hallucinations or the potential for bias or hallucinations in the response. These factors are assigned different weights, with factual accuracy receiving the highest weight, as well as a range of values and interpretations [42]. From this, three conclusions can be drawn:

1. The weights assigned to each factor are presented in Table 1.

Table 1. The weights assigned to each factor

No	Component	Weight	Score
1	Factual accuracy (F)	0.4	0 (does not meet), 0.5 (partially/ambiguous), 1 (fulfills well)
2	Semantic alignment (S)	0.3	0 (does not meet), 0.5 (partially/ambiguous), 1 (fulfills well)
3	Completeness and specificity (C)	0.2	0 (does not meet), 0.5 (partially/ambiguous), 1 (fulfills well)
4	Groundedness (G)	0.1	0 (does not meet), 0.5 (partially/ambiguous), 1 (fulfills well)

2. The AAS is determined using Equation (1), which involves calculating the ratio of correct responses to the total number of questions posed [43]:

$$\text{AAS} = (\text{total correct answers} / \text{total questions}) \times 100\% \dots\dots\dots (1)$$

Equation (1) can be converted into Equation (2) by taking into account the assigned weights for each dimension and the total number of questions posed, acknowledging that cybersecurity awareness among users constitutes an aspect of the human factor [44]:

$$\text{AAS} = ((F \times w_1) + (S \times w_2) + (C \times w_3) + (G \times w_4)) / \text{total questions} \dots\dots\dots (2)$$

Where: F = factual accuracy, w1 = weight for factual accuracy, S = semantic alignment, w2 = weight for semantic alignment, C = completeness and specificity, w3 = weight for completeness and specificity, G = groundedness, w4 = weight for groundedness.

3. The total score range based on Equation (2) and its interpretation are displayed in Table 2.

Table 2. The total score range, category and its interpretation

No	The total score range	Category	Interpretation
1	0.85-1.00	Valid/Very accurate	The response aligns with the factual evidence or reference sources, exhibits no hallucination, and is pertinent to the context of the inquiry
2	0.70-0.84	Fairly Valid	The response aligns with the context of the question; however, it contains minor inaccuracies or lacks completeness
3	0.50-0.69	Less Valid	The response was insufficiently relevant to the context of the question and lacked completeness or detail
4	< 0.50	Invalid	Incorrect responses do not align with the context of the question and are often the result of cognitive distortions

3. RESULTS AND DISCUSSION

3.1. AI-driven user cybersecurity awareness consultant prototype

We developed a prototype of an AI-driven user cybersecurity awareness consultant utilizing the RAG method. This prototype was hosted on a domain and is accessible via a URL through a web browser, specifically using the Chrome web browser in this study. We prepared 50 PDF documents related to cybersecurity awareness as internal knowledge-based resources, which were incrementally uploaded into the system. Supplementary data for each document was also entered. Subsequently, the prototype underwent processes of chunking, vectorization, and embedding. Figure 3(a) depicts the procedure for data entry into an internal knowledge-based document. Figure 3(b) confirms the successful completion of the insertion process for this document. Figure 4(a) illustrates the input of a question using a prompt to integrate results from the

internal knowledge-based document, which references one of the documents, along with the system's response comprising search results from both internal and external knowledge-based documents, as well as information and knowledge successfully retrieved from the internal knowledge-based document. Figure 4(b) presents the detailed parameters for each process executed by the prototype when addressing user queries to the hybrid knowledge-based document.

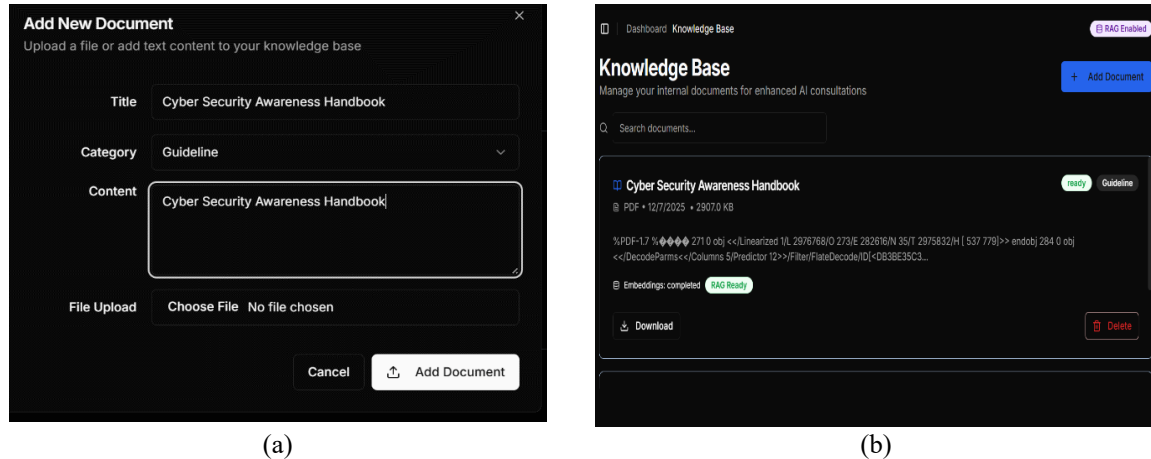


Figure 3. Prototype of AI-driven User Cybersecurity Awareness Consultant using RAG method
(a) The data entry procedures for an internal knowledge-based document (PDF), (b) Confirms the successful completion of the embedding process for an internal knowledge-based document (PDF)

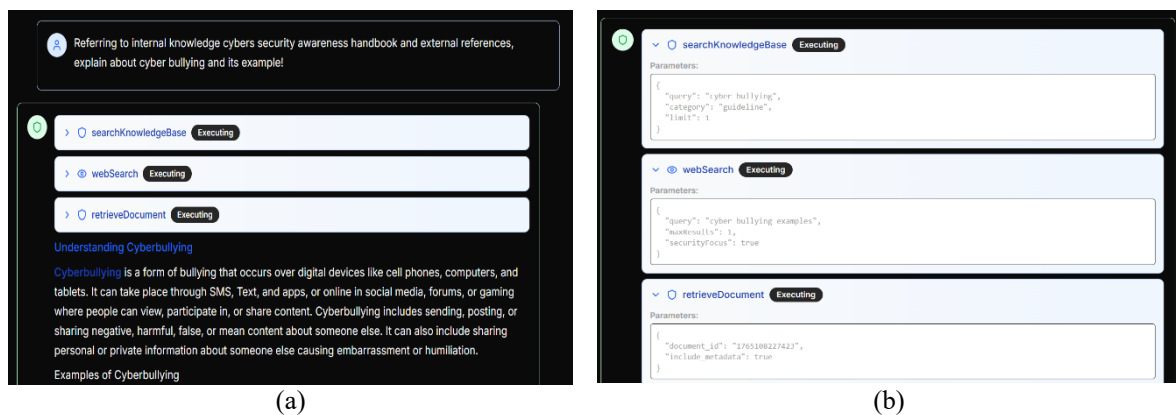


Figure 4. The input of a question and the system's response,
(a) The input of a question using a prompt along with the system's response, (b) The detailed parameters for each process executed

3.2. The list of questions and system responses

In order to conduct the test, we formulated four questions pertaining to cybersecurity. These questions were designed using the prompt “Referring to the internal and external references, explain the topic-what-you-want-to-ask,” or “Referring to the internal knowledge the-title-of document and external references, explain the topic-what-you-want-to-ask,” ensuring that the prototype generates responses that draw upon both internal knowledge from the documents we provided and additional external knowledge sources available on the internet. If the relevant information is absent from one of these knowledge sources (either internal or external), the prototype will present an answer based solely on the available source. In instances where the information is unavailable from both sources, the prototype will indicate its inability to answer the question, thereby avoiding speculative responses. We utilized the Q code to signify questions (e.g., Q1 for question 1) and the R code to indicate system responses (e.g., R1 for the system's response to question 1 or Q1). The questions are enumerated in Table 3., while the system responses are detailed in Table 4.

Table 3. The list of questions

Q	Question
Q1	Referring to internal and external references, explain how to most effectively prevent ransomware infections in small companies with limited IT resources!
Q2	Referring to internal and external references, explain about cybersecurity awareness in the IT sector!
Q3	Referring to internal and external references, explain about cisco cybersecurity awareness!
Q4	Referring to internal knowledge cybers security awareness handbook and external references, explain about cyber bullying and its example!

Table 4. The list of system responses

R	Response
R1	To effectively prevent ransomware infections in small companies with limited IT resources, consider the following strategies based on external references and best practices: Endpoint Protection: Implement strong endpoint protection solutions to detect and block ransomware before it can execute. This includes using antivirus and anti-malware software that is regularly updated to recognize the latest threats. Offline Backups: Maintain regular offline backups of critical data. This ensures that even if ransomware encrypts your files, you can restore them without paying a ransom. Ensure backups are not connected to your network to prevent them from being compromised. Security Awareness Training: Educate employees about the risks of phishing and social engineering attacks, which are common vectors for ransomware. Training should include recognizing suspicious emails and avoiding clicking on unknown links or downloading attachments from untrusted sources. Network Segmentation: Segment your network to limit the spread of ransomware. By isolating critical systems and data, you can prevent ransomware from moving laterally across your network. Regular Software Updates: Keep all software, including operating systems and applications, up to date with the latest security patches. This reduces vulnerabilities that ransomware can exploit. Implement Cyber Threat Detection: Use tools that monitor network traffic and flag unusual behavior, which can indicate a ransomware attack in progress. Access Controls: Limit user permissions to only what is necessary for their role. This minimizes the potential damage if a user's account is compromised.
R2	Cybersecurity awareness in the IT sector is a critical component of maintaining robust security practices and protecting against cyber threats. Here are some key insights from external references: Role of Cybersecurity Awareness: Cybersecurity awareness training is a strategy used by IT and security professionals to prevent and mitigate user risk. It helps employees understand their essential role in securing an organization against cyber incidents and breaches. This training is crucial for building a cyber-aware culture where employees can recognize and avoid falling victim to cyberattacks. Importance of Training Programs: Security awareness training is often considered the number one precaution to improve cyber resilience. It discourages negligent or risky user behaviors by delivering an understanding of cyber risks and incident scenarios. Such programs are integral to many cyber resilience frameworks and help organizations comply with regulatory or industry compliance training requirements. Shared Responsibility: Cybersecurity is a shared responsibility, and everyone has a part to play. National efforts, such as the CISA Cybersecurity Awareness Program, aim to increase understanding of cyber threats and empower individuals to be safer online. This includes providing resources and tools to help make informed decisions when using the Internet. Cybersecurity Awareness Month: Since 2004, October has been declared Cybersecurity Awareness Month in the United States, highlighting the importance of protecting oneself online as threats to technology and confidential data become more common.
R3	Cisco's approach to cybersecurity awareness is multifaceted, focusing on building a strong security culture, enhancing employee accountability, and maintaining compliance with security standards. Here are some key insights from both internal and external references: Internal Knowledge Base Unfortunately, there were no specific internal guidelines or documents found in the knowledge base regarding Cisco's cybersecurity awareness. However, general best practices in cybersecurity awareness typically involve regular training, simulations, and updates on emerging threats. External References Keeping Cisco Safe: Cisco has developed a comprehensive cybersecurity awareness program that emphasizes the importance of understanding and mitigating security risks. The program is designed to keep employees informed about new threats and regulations, ensuring they are equipped to handle potential security breaches. Cisco shares its best practices with customers to help them develop their own cybersecurity awareness programs, thereby building trust and enhancing security culture. Cisco Security Awareness Program: This program includes various tools and strategies such as course builders, phishing simulators, and multilingual user interfaces to enhance security training. The focus is on influencing human behavior, as people are often the weakest link in security. The program aims to make security awareness an integral part of the organizational culture. Cisco Cybersecurity Readiness Index 2025: This report highlights the evolving nature of AI-driven threats and the importance of proactive measures in cybersecurity. It emphasizes the need for organizations to increase their cybersecurity budgets and take decisive actions to transform data security from a vulnerability into a competitive advantage. Learn more about the Cisco Cybersecurity Readiness Index Summary Cisco's cybersecurity awareness initiatives are designed to educate and empower employees, ensuring they are prepared to handle various security threats. By integrating security awareness into the organizational culture and continuously updating

training programs, Cisco aims to stay ahead of emerging threats and maintain a robust security posture.

R4 Understanding Cyberbullying

Cyberbullying is a form of bullying that occurs over digital devices like cell phones, computers, and tablets. It can take place through SMS, Text, and apps, or online in social media, forums, or gaming where people can view, participate in, or share content. Cyberbullying includes sending, posting, or sharing negative, harmful, false, or mean content about someone else. It can also include sharing personal or private information about someone else causing embarrassment or humiliation.

Examples of Cyberbullying

Malware and Hacking: Cyberbullies may use malware to hack into a victim's phone or computer. This can give them access to sensitive personal information, GPS location, or even control over the device's camera, allowing them to stalk the victim.

Identity Theft and Impersonation: Cyberbullies often engage in identity theft by accessing or opening new lines of credit in the victim's name. They may also impersonate the victim by hacking into their user accounts on games, email, or social media to gather information or cause harm.

Harassment and Threats: Sending threatening messages or emails, posting derogatory comments on social media, or spreading rumors online are common tactics used by cyberbullies.

Exclusion: Intentionally excluding someone from an online group or activity can also be a form of cyberbullying.

Cyberbullying can have severe impacts on the victim's mental health, leading to anxiety, depression, and in extreme cases, suicidal thoughts. It's crucial to address cyberbullying promptly and effectively to protect individuals' well-being and privacy.

3.3. Answer Accuracy Score (AAS)

Following the identification of the list of questions and system responses, the subsequent step involved conducting AAS calculations based on human evaluation. This assessment and calculation employ Equation (2), which is derived from Equation (1) by incorporating four components: factual accuracy (F), semantic alignment (S), completeness and specificity (C), and groundedness (G). The AAS assessment data, which includes the list of questions (Q), system responses (R), and the calculation of four components (F, S, C, and G), are presented in Table 5.

Table 5. Answer Accuracy Score (AAS)

Q	R	F	S	C	G	AAS = (F+S+C+G)
Q1	R1	$0.4 \times 1 = 0.4$	$0.3 \times 1 = 0.3$	$0.2 \times 0.5 = 0.1$	$0.1 \times 0.5 = 0.05$	$0.4 + 0.3 + 0.1 + 0.05$ $= 0.85$
Q2	R2	$0.4 \times 1 = 0.4$	$0.3 \times 1 = 0.3$	$0.2 \times 1 = 0.2$	$0.1 \times 1 = 0.1$	$0.4 + 0.3 + 0.2 + 0.1$ $= 1$
Q3	R3	$0.4 \times 1 = 0.4$	$0.3 \times 1 = 0.3$	$0.2 \times 0.5 = 0.1$	$0.1 \times 0.5 = 0.05$	$0.4 + 0.3 + 0.1 + 0.05$ $= 0.85$
Q4	R4	$0.4 \times 1 = 0.4$	$0.3 \times 1 = 0.3$	$0.2 \times 1 = 0.2$	$0.1 \times 1 = 0.1$	$0.4 + 0.3 + 0.2 + 0.1$ $= 1$

According to the results presented in Table 5, the average AAS value can be determined using Equation (2):

$$\begin{aligned} \text{AAS} &= (0.85 + 1 + 0.85 + 1) / 4 \\ &= 3.7 / 4 \\ &= 0.925 \end{aligned}$$

3.4. Discussion

Upon evaluating the test and analysis results, it was determined that the four questions directed at the prototype (Q1-Q4) and their corresponding responses (R1-R4) each received a score of 1 for Factual Accuracy (F) and Semantic Alignment (S). This outcome signifies that all four responses were accurate, factual, and contextually appropriate. The AI method employed in the prototype, RAG, successfully aligned the queries with the available knowledge, both internal and external. However, regarding Completeness (C), the responses were not deemed complete. Despite the prototype generating detailed and extensive responses, several critical points were omitted. For instance, in the first question (Q1) and its response (R1), although the response was detailed and extensive, it failed to include two significant points: disaster recovery and incident response, or cybersecurity incident response. Consequently, the Completeness (C) score was 0.5, indicating incomplete responses. Concerning Groundedness (G), some responses provided only general knowledge without citing internet reference sources in the form of URLs (when utilizing external knowledge) or document sources (when utilizing internal knowledge). Therefore, the score for each response varied from 0.5 (when a source was not included) to 1 (when a source was included).

With an average AAS score of 0.925 (ranging between 0.85 and 1.00, according to Table 2.), it can be concluded that the prototype functioned as intended. The average response was accurate or valid, minimizing or preventing hallucinations.

Thus, the prototype is deemed suitable for use in cybersecurity education initiatives aimed at enhancing cybersecurity awareness by reducing hallucinations in system responses. The minor deficiencies that persist in the prototype necessitate the incorporation of additional methods to enable the system to include valid sources in its responses, thereby enhancing the Completeness (C) and Groundedness (G) scores to a value of 1. It is anticipated that this improvement will enhance the prototype's efficacy in cybersecurity education efforts, thereby increasing cybersecurity awareness by reducing hallucinations in the system's responses.

4. CONCLUSION

Our assessment of an AI-based cybersecurity awareness consultant prototype, developed utilizing the RAG method and a hybrid knowledge-based approach, revealed its capability to generate contextually accurate and precise responses grounded in factual information and references. The majority of the responses were both accurate and contextually pertinent, with no instances of hallucinations observed. Consequently, the prototype is deemed functionally viable and can be effectively employed to enhance the cybersecurity education process for users, with the objective of augmenting their cybersecurity awareness. In our future research, we intend to refine the prototype by concentrating on the integration of deliberative thinking methods into the RAG framework to further improve the accuracy and contextual relevance of the system's responses. Additionally, a method for evaluating user capabilities before and after educational interventions was incorporated.

ACKNOWLEDGEMENTS

This work was supported by the Directorate General of Research and Development, Ministry of Higher Education, Science and Technology of the Republic of Indonesia, for awarding the 2025 Doctoral Dissertation Research grant, as per research contract Number: B/335-4/UN14.4.A/PT.01.03/2025, dated June 2, 2025.

REFERENCES

- [1] R. Tashfiq, "Human Factors in Cybersecurity: A Scoping Review," *12th Int. Conf. Adv. Inf. Technol. IAIT2021*, 2021, doi: 10.1145/3468784.3468789.
- [2] T. Ostonya, B. M. Muhaimin, M. T. Jean, and R. Mizanur, "Transportation Cyber Incident Awareness through Generative AI-Based Incident Analysis and Retrieval-Augmented Question-Answering Systems," *ArXiv250802523 Cryptogr. Secur.*, vol. 1, 2025, doi: 10.48550/arXiv.2508.02523.
- [3] C. Silvia, C. Francesco, N. Fabio, and P. Giulia, "Transforming threats into opportunities: The role of human factors in enhancing cybersecurity," *J. Innov. Knowl.*, vol. 10, no. 3, 2025, doi: 10.1016/j.jik.2025.100695.
- [4] S. Alahmari, K. Renaud, and I. Omoronyia, "Moving beyond cyber security awareness and training to engendering security knowledge sharing," *Inf Syst E-Bus Manage*, vol. 21, pp. 123–158, 2023, doi: 10.1007/s10257-022-00575-2.
- [5] D. C. Nikitha, L. Anthony, and M. Azab, "Understanding User Behavior for Enhancing Cybersecurity Training with Immersive Gamified Platforms," *Information*, vol. 15, no. 12, p. 814, 2024, doi: 10.3390/info15120814.
- [6] O. A. Temitayo, A. F. Oluwatoyin, S. Kaggwa, and U. U. Prisca, "Cybersecurity Awareness and Education Programs: A Review of Employee Engagement and Accountability," *Comput. Sci. IT Res. J.*, vol. 5, no. 1, 2024, doi: 10.51594/csitrj.v5i.708.
- [7] A. T. Neumann, Y. Yin, S. Sowe, and S. Decker, "An LLM-Driven Chatbot in Higher Education for Databases and Information Systems," *IEEE Trans. Educ.*, vol. 68, no. 1, pp. 103–116, 2025, doi: 10.1109/TE.2024.3467912.

- [8] L. Desta, M. Afifuddin, and A. Adriyanto, "Challenges and Cybersecurity Threats in Digital Economic Transformation," *Int. J. Humanit. Educ. Soc. Sci. IJHESS*, vol. 6, no. 2, 2023, doi: 10.55227/ijhess.v2i6.515.
- [9] B. Sarmah, B. Hall, R. Rao, and S. Patel, "HybridRAG: Integrating Knowledge Graphs and Vector Retrieval Augmented Generation for Efficient Information Extraction," *Comput. Lang. CsCL ArXiv240804948*, vol. 8, 2024, doi: 10.48550/arXiv.2408.04948.
- [10] G. Byun, S. Lee, N. Choi, and J. D. Choi, "Secure Multifaceted-RAG: Hybrid Knowledge Retrieval with Security Filtering," *Information*, vol. 16, p. 804, 2025, doi: 10.3390/info16090804.
- [11] Y. Shang, Z. Ke, P. Lin, and Q. Ren, "Empowering knowledge graphs with hybrid retrieval-augmented generation for the intelligent mix scheme of mass concrete," *Case Stud. Constr. Mater.*, vol. 23, p. 4979, 2025, doi: 10.1016/j.cscm.2025.e04979.
- [12] S. Kiran, "Hybrid Retrieval-Augmented Generation (RAG) Systems with Embedding Vector Databases," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 11, no. 2, pp. 2694–2702, 2025, doi: 10.32628/CSEIT25112702.
- [13] R. Kalra, Z. Wu, A. Gulley, and A. Hilliard, "HyPA-RAG: A Hybrid Parameter Adaptive Retrieval-Augmented Generation System for AI Legal and Policy Applications," *Proc. Ist Workshop Cust. NLP Prog. Chall. Cust. NLP Domain Appl. Group Individ. Cust.*, pp. 237–256, 2024, doi: 10.18653/v1/2024.customnlp4u-1.18.
- [14] Z. Shahbazi, R. Jalali, and M. Maryam, "AI-Based Phishing Detection and Student Cybersecurity Awareness in the Digital Age," *Big Data Cogn Comput*, vol. 9, no. 8, p. 210, 2025, doi: 10.3390/bdcc9080210.
- [15] M. M. Yamin, H. Ehtesham, U. Mohib, and B. Katt, "Applications of LLMs for Generating Cyber Security Exercise Scenarios," *IEEE Access*, vol. 12, pp. 143806–143822, 2024, doi: 10.1109/ACCESS.2024.3468914.
- [16] A. Zineddine, A. Abdeslam, M. Zaoui, and Y. Belfaik, "LLM-based assessment of HTTPS cybersecurity awareness: Dataset from moroccan web users and webmasters," *Data Brief*, vol. 62, 2025, doi: 10.1016/j.dib.2025.112024.
- [17] O. Cohen, G. A. Agmon, S. Asaf, and R. Puzis, "The Information Security Awareness of Large Language Models," *ArXiv Cryptogr. Secur.*, vol. 2411, 2024, doi: 10.48550/arXiv.2411.13207.
- [18] J. Antony, M. Trovati, and S. Bolton, "Retrieval-Augmented Generation to Generate Knowledge Assets and Creation of Action Drivers," *Appl Sci*, vol. 15, no. 11, p. 6247, 2025, doi: 10.3390/app15116247.
- [19] J. P. Singh and S. Agrawal, "Automating Threat Intelligence Analysis with Retrieval Augmented Generation (RAG) for Enhanced Cybersecurity Posture," *Int. J. Sci. Res. IJSR*, vol. 13, no. 5, pp. 251–255, 2024, doi: 10.21275/SR24502103758.
- [20] Z. Chengshuai, G. Agrawal, F. Zhang, and K. Tharindu, "Ontology-Aware RAG for Improved Question-Answering in Cybersecurity Education," *ArXiv241214191 Comput. Sci.*, vol. 1, no. 2, 2024, doi: 10.48550/arXiv.2412.14191.
- [21] F. Blefari, C. Cosentino, F. A. Pironti, and F. Angelo, "CyberRAG: An agentic RAG cyber attack classification and reporting tool," *Future Gener. Comput. Syst.*, vol. 76, p. 108186, 2026, doi: 10.1016/j.future.2025.108186.
- [22] S. Rajapaksha, R. Rani, and E. Karafili, "A RAG-Based Question-Answering Solution for Cyber-Attack Investigation and Attribution," *ArXiv Prepr. ArXiv240806272*, no. Query date: 2024-09-02 23:48:01, 2024, [Online]. Available: <https://arxiv.org/abs/2408.06272>
- [23] I. P. A. E. Pratama, I. M. O. Widyantara, L. Linawati, and N. Gunantara, "Bibliometric Analysis of AI-Based Prototype Proposal for User Security Awareness in Healthcare," *Int. J. Inform. Vis. JOIV*, vol. 9, no. 3, 2025, doi: 10.62527/joiv.9.3.3319.
- [24] I. P. A. E. Pratama, I. M. O. Widyantara, L. Linawati, N. Gunantara, S. Suakanto, and E. T. Irawan, "Technology-Enhanced Learning for User Security Awareness Using AI-based Naive RAG: A Design and Prototype," *2025 Int. Conf. Adv. Data Sci. E-Learn. Inf. Syst. ICADEIS*, vol. 2025, pp. 1–6, 2025, doi: 10.1109/ICADEIS65852.2025.10933283.

- [25] I. P. A. E. Pratama, I. M. O. Widyantara, L. Linawati, N. Gunantara, and E. T. Irawan, "A Proof of Concept for Fixed Chunking in an Aidriven Cybersecurity Awareness Prototype," *Int. Conf. Smart-Green Technol. Electr. Inf. Syst. ICSGTEIS*, pp. 134–139, 2025, doi: 10.1109/ICSGTEIS68532.2025.11284323.
- [26] M. U. Hadi, Q. A. Tashi, and R. Qureshi, "A Survey on Large Language Models: Applications, Challenges, Limitations, and Practical Usage," *TechRxiv*, vol. 10, 2023, doi: 10.36227/techrxiv.23589741.v1.
- [27] P. Lewis, E. Perez, A. Piktus, and F. Petroni, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *Proceeding 34th Conf. Neural Inf. Process. Syst. NeurIPS 2020 Vanc. Can.*, 2020, doi: 10.48550/arXiv.2005.11401.
- [28] Y. Gao, X. Yun, G. Xinyu, and K. Jia, "Retrieval-Augmented Generation for Large Language Models: A Survey," *ArXiv231210997 Comput. Lang.*, vol. 23, 2023, doi: 10.48550/arXiv.2312.10997.
- [29] J. H. A. Alvaro and J. G. Barreda, "An advanced retrieval-augmented generation system for manufacturing quality control," *Adv. Eng. Inform.*, vol. 64, p. 103007, 2025, doi: 10.1016/j.aei.2024.103007.
- [30] Y. Shuo, C. Mingyue, L. Qi, W. Daoyu, and J. Yang, "Multi-Source Knowledge Pruning for Retrieval-Augmented Generation: A Benchmark and Empirical Study," *Comput. Lang. CsCL ArXiv240913694*, vol. 24, 2025, doi: 10.48550/arXiv.2409.13694.
- [31] D. O. Opoku, M. Sheng, and Y. Zhang, "DO-RAG: A Domain-Specific QA Framework Using Knowledge Graph-Enhanced Retrieval-Augmented Generation," *Comput. Lang. CsCL ArXiv250517058*, vol. 25, 2025, doi: 10.48550/arXiv.2505.17058.
- [32] M. C. Lee, Q. Zhu, C. Mavromatis, Z. Han, and S. Adeshina, "HybGRAG: Hybrid Retrieval-Augmented Generation on Textual and Relational Knowledge Bases," *Proc. 63rd Annu. Meet. Assoc. Comput. Linguist.*, vol. 1, pp. 879–893, 2025, doi: 10.18653/v1/2025.acl-long.43.
- [33] M. Murtiyoso, I. Tahyudin, and B. Berlilana, "A Systematic Review of Retrieval-Augmented Generation for Enhancing Domain-Specific Knowledge in Large Language Models," *Sink. J. Dan Penelit. Tek. Inform.*, vol. 9, no. 2, 2025, doi: 10.33395/sinkron.v9i2.14824.
- [34] Y. Wan, Z. Chen, L. Ying, C. Chen, and M. Packianather, "Empowering LLMs by hybrid retrieval-augmented generation for domain-centric Q&A in smart manufacturing," *Adv. Eng. Inform.*, vol. 65, p. 103212, 2025, doi: 10.1016/j.aei.2025.103212.
- [35] E. Z. Selviana, H. E. Wahanani, and E. P. Aditiawan, "Black Box Testing in Community Service Systems Using Boundary Value Analysis and Cause Effect Graph Methods (Case Study of Jombang District Community and Village Empowerment Service)," *East Asian J. Multidiscip. Res.*, vol. 2, no. 10, pp. 4161–4184, 2023, doi: 10.55927/eajmr.v2i10.6435.
- [36] T. Gupta and E. Kumar, "Answer Relevance Score (ARS): Novel Evaluation Metric for Question Answering System," *Int. Conf. Adv. Comput. Commun. Inf. Technol. ICAICIT*, pp. 292–296, 2023, doi: 10.1109/ICAICIT60255.2023.10466080.
- [37] A. Wang, K. Cho, and M. Lewis, "Asking and Answering Questions to Evaluate the Factual Consistency of Summaries," *Proc. 58th Annu. Meet. Assoc. Comput. Linguist.*, pp. 5008–5020, doi: 10.18653/v1/2020.acl-main.450.
- [38] F. Nan, C. N. D. Santos, H. Zhu, P. Ng, and K. M. Keown, "Improving Factual Consistency of Abstractive Summarization via Question Answering," *Proc. 59th Annu. Meet. Assoc. Comput. Linguist. 11th Int. Jt. Conf. Nat. Lang. Process.*, vol. 1, pp. 6881–6894, 2021, doi: 10.18653/v1/2021.acl-long.536.
- [39] A. Fabbri, C. S. Wu, W. Liu, and C. Xiong, "QAFactEval: Improved QA-Based Factual Consistency Evaluation for Summarization," *Proc. 2022 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol.*, pp. 2587–2601, 2022, doi: 10.18653/v1/2022.naacl-main.187.
- [40] P. Yao and D. Barbosa, "Accurate and Nuanced Open-QA Evaluation Through Textual Entailment," *Find. Assoc. Comput. Linguist. ACL 2024*, vol. 151, pp. 2575–2587, 2024, doi: 10.18653/v1/2024.findings-acl.151.

- [41] C. W. Huang and Y. N. Chen, "FactAlign: Long-form Factuality Alignment of Large Language Models," *Find. Assoc. Comput. Linguist. EMNLP 2024*, vol. 955, pp. 16363–16375, 2024, doi: 10.18653/v1/2024.findings-emnlp.955.
- [42] T. Zheng, S. Shen, and C. Zeng, "A Retrieval-Augmented Generation Method for Question Answering on Airworthiness Regulations," *Electronics*, vol. 14, no. 16, p. 3314, 2025, doi: 10.3390/electronics14163314.
- [43] R. B. Permadi and K. Ramli, "Analysis of Measuring Information Security Awareness for Employees at Institution XYZ," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, 2024, [Online]. Available: <https://journal.irpi.or.id/index.php/malcom>
- [44] D. T. J. Jun, A. S. Rafsanjani, S. Aslam, and M. Behjati, "Human Factors in Information Security: A Quantitative Study with Technical Solutions to Prevent Social Engineering Attacks," *Digit. Threats Res. Pract.*, vol. 6, no. 4, pp. 1–35, 2025, doi: 10.1145/37673.